LESLI
Linguistic Evidence in Security, Law and Intelligence

Volume 1, No. 1 (2013) | ISSN 2327-5596 (online) | DOI 10.5195/lesli.2013.3 | http://lesli-journal.org

Barriers to Progress in Speaker Identification

with Comments on the Trayvon Martin Case

Harry Hollien PhD
Institute for the Advanced Study of Communicative Processes, University of Florida
hollien@ufl.edu

Abstract

Linguistics and phonetics overlap in many areas. The essay to follow reviews some of the problems experienced by phoneticians in one of these regions. It may provide some insight for linguists when they are confronted by barriers in their own field. The present example involves individuals who are attempting to identify speakers from voice analysis. The *fundamental* challenge they face is, of course, caused by the thousands of variables associated with that task. Included here are differences among speakers' gender, age, size, physiology, language, dialect, psychological/health states, background/education, reason for speaking, situation, environment, configuration of the acoustic channel -- plus many others. Many formal assessment procedures -- both aural-perceptual ones conducted by humans or machine/computer based systems -- have been proposed and/or used for the cited analyses. Unfortunately, however, few have enjoyed particularly high levels of success. Worse yet, reasonable progress has suffered from external impedances; the report to follow will outline some of them. Among the problems considered are: 1) competition (verification vs. identification, from voiceprints), 2) concept disputes 3) the continued undervaluation of relevant evidence and 4) markedly dissimilar philosophies of professionals from different disciplines. A response in the form of a short review of the data and concepts which clearly support the possibility of robust speaker identification is presented. Also included are suggestions as to how to enhance the effectiveness of disciplines such as ours.

Keywords: speaker identification, automated speech processing, expert witnesses, Trayvon Martin

Introduction

Although linguists and phoneticians inhabit somewhat different domains, their efforts and interests functionally overlap, and do so at many boundaries. Likewise, they both have experienced a number of successes as well as conditions which have interfered with their optimal progress. This essay will not be used to explore multiples of these suggested commonalities, but rather only a single instance – one which is related to forensics. Specifically, its focus will be on problems experienced by phoneticians in an area which parallels a similar one in linguistics.

Linguists are often called upon to apply their analytic techniques to identify the author of a text or otherwise identify a person from his or her language usage. Decisions made in this area of forensics can be of substantial importance to an investigation or a trial. This particular activity relates directly to one in a forensic-based area of phonetics. That is, phoneticians are sometimes asked to apply available techniques to identify individuals from their speech and voice – an issue which, in many ways, parallels the linguists' authorship task. In any event, the review to follow will be focused on speaker identification (within speaker recognition). While it is recognized that the unique obstacles confronting linguists in this area may differ somewhat from those to be discussed, it is just possible that the frustration and discomfort experienced by phoneticians in *their* struggle may reverberate within the linguistic psyche.

The Problem of Competition

Speaker Identification (SI) vs. Speaker Verification (SV)

It first must be pointed out that speaker identification (SI) is but one of two related divisions subsumed under *speaker recognition* (Figure 1). The other is speaker verification (SV). **Speaker verification** (SV) is 1) where a talker wishes to be correctly identified (examples: to gain access to a bank account, to attain entry to a restricted area) or 2) when an individual's identity needs to be validated for some reason (examples: a speaker at a security outpost or as to which astronaut is talking in the space capsule). Only the most sophisticated procedures and equipment are employed for speaker verification. Moreover, the SV examiner can establish a library of many and varied referent speech samples to compare with the questioned utterance(s). On the other hand, **Speaker identification** (SI) constitutes a much greater challenge. Here an often uncooperative and unknown speaker, residing within a population of unknown size and composition, must be identified by analyzing samples of his or her speech and voice. The examiner must do so under conditions which, in many cases, range from poor to minimally adequate. As is obvious, a different approach to successful recognition is needed for SI than is the one for SV. Research has been carried out on both issues and for over a century. But, can it be said that progress in these areas has advanced in parallel? No. It cannot.

SPEAKER VERIFICATION (SV)	SPEAKER IDENTIFICATION (SI)
The talker is usually cooperative	The talker is usually uncooperative
Mimicry can occur; it has a negative impact on SV systems	Voice disguise can exist; it has a negative impact on SI systems
Latency and processing time is usually short	Latency and processing time can be lengthy
SV is text dependant; vocabulary is limited, and controlled	SV is text independent; vocabulary may be quite large
Environmental issues rarely occur	The SI environments rarely are similar
Channel characteristics and the signal-to-noise ratio are usually favorable and can be controlled	Channel characteristics may be poor or differing and the signal-to-noise ratio may be difficult to control
Speech samples can be updated continuously with parallel content	Speech samples usually vary and can be difficult to equalize
The test set is closed	The test set is open

Figure 1: Some of features that demonstrate the differences between speaker verification and speaker identification.

Volume 1, No. 1 2013 | DOI 10.5195/lesli.2013.3 | http:// lesli-journal.org

Rather, the presence of, and competition by, SV has seriously impeded good progress in the development of effective SI procedures. It has done so for two reasons. First, SV is commercially viable. It can be useful in many situations -- in the military, at correctional institutions, for security purposes and, especially, in commerce. As a result, the SV developer's financial rewards – i.e., those resulting from the purchase and use of his SV system – are substantial indeed. As you might expect, this situation attracts many professionals to organizations that need to verify the identity of their customers, clients, associates, personnel and/or others. In turn, these relationships have served to shift most of the financial and research support from SI to SV.

Secondly, the relatively benign environment associated with SV and the *relatively* modest number of variables encountered here are attractive to professionals and draw them to this area (rather than to SI). Ironically, since SV is many magnitudes a lesser challenge than is SI, if a good SI system could be developed, the problem of verification would be solved also[24]. Finally, engineers tend to be comfortable with verification. Their response often is to develop algorithms which permit them to build a device that can be directed at the problem (even though it may not directly focus on speech or voice) and then use it. If the system does not operate effectively, it is discarded and replaced by a different one.

Thus, it is phoneticians who are left to develop speaker-specific ways to identify talkers. When they do, they then must subject their systems to a series of experiments (sometimes a long series) in order to see if they work. Of course, this latter approach is costly and results in a delay between the system's concept and its use. In short, the glamour and financial rewards of SV have seriously overshadowed the slow, grinding pathway to successful SI. Worse yet, the differences between the SI and SV domains have tended to force the phoneticians and engineers to work apart from each other. Only a few of them presently collaborate and this lack of extensive cooperation serves to impede progress in the development of effective speaker identification systems.

Unfortunately, much of the SI-SV "confusion" has persisted even to the present. A large portion of the SV research which has been completed is not helpful to SI development and urban legends have built up around the "impossibility" of developing effective SI. Nonetheless, the need here is one of critical importance to the law enforcement and intelligence communities.

But will it be possible to devise successful SI systems at all or is doing so just a dream? Well, after World War II ended, a number of events occurred which suggested that this goal might be attainable. Indeed, strong evidence emerged that SI could become a reality. Then "Voiceprints" happened.

The Serious Distraction Created by Voiceprints

Of course, it must be understood that interest in speech recognition is not a new phenomenon. It began many years – indeed, many millennia – before the identification-verification issue discussed above arrived on the scene. Occasional references to it can be found over the centuries – especially in the legal and pedagogical literature. However, this issue probably did not reach the courts until the 17th century[18]. One early example had a judge in England becoming convinced that one of the witnesses at a trial "knew" the defendant well enough to recognize his voice. The judge thereby allowed the witness to testify to that effect. A similar event occurred in 1907 in the United States[42]. And, only a few years later Charles Lindbergh testified in the Hauptmann trial[19] that he recognized the defendant's voice as that of his child's kidnapper. In turn, that particular event led to one of the earliest of the modern experiments in speaker identification[43]. And, that effort appeared to create something of a "breakthrough". However, World War II was imminent and, even though the SI issue was of interest to the intelligence services and the military, not a great deal of follow-up research seems to have been carried out during the war years. That is,

Volume 1, No. 1 2013 | DOI 10.5195/lesli.2013.3 | http:// lesli-journal.org/

about all we "know" about work in this area was from newspaper stories. An example; when an assassination attempt was made on the life of Adolf Hitler, "Mac" Steer (a professor at Purdue) was retained by the U.S. Government to determine if it was the Nazi leader -- or his stand-in -- who was making the subsequent radio broadcasts (Steer was quoted as saying that his analysis pointed to Hitler rather than a double).

In any event, interest in all facets of speaker recognition increased rapidly after the war, even to the point where some major resources were assigned to the study of SI. A number of top-level engineers with research backgrounds began to experiment with the basic recognition relationships at Bell Telephone Labs, MIT, Haskins Labs and several U.S. universities (Kenneth Stevens, Gordon Peterson, "Mac" Pickett and Franklin Cooper were among the leaders here). They were joined by several of the leading U.S. phoneticians (including Grant Fairbanks, John Black, James Curtis, and "Mac" Steer). Moreover, even though they were slowed by the devastation left by the war, research of this type was echoed by individuals and groups in Europe and the Pacific Rim countries. Better yet, the SI-SV "competition" had not yet developed so, with growing funding, improved research apparatus and scientific cooperation across disciplines, it appeared that speaker identification was on the brink of enjoying significant progress.

Then a serious obstruction suddenly appeared. **Voiceprints**!

This problem was triggered by a number of events and conditions. First, not long after the war ended, civil unrest developed over a variety of issues and at a number of places throughout the world. Law enforcement organizations had to respond to an ever expanding number of incidents. In addition, the rapid spread, and increasing sophistication, of electronic communication and recording systems amplified the negative aspect of this situation. As a result, law enforcement (plus other agencies) became aware that they badly needed ways to remotely identify suspects and perpetrators from their speech – especially from their recorded utterances. At this point in time, a few phoneticians attempted to develop and use various sorts of aural-perceptual speaker identification (AP SI) techniques and systems. In addition, certain engineers tried to adapt a few of the procedures they had been using in speaker recognition research. Neither of these two approaches was particularly successful. Thus, when a technician at Bell Telephone Labs described[33] an approach he labeled "voiceprints", his efforts attracted interest.

The technician in question had, over a period of time, made many time-frequency-amplitude type spectrograms (often called Sonagrams) for the BTL scientists and engineers (see Figure 2). It was during this period that he noticed that, even though they often were blurry and the frequency scale was routinely misapplied, these printouts seemed to provide interesting "pictures" of human speech, animal calls, and other acoustic events. He speculated that, since humans sometimes could be identified from what *might* be "unique" speech features, he should be able to use spectrograms to identify some of them. Later, he claimed that he actually was able to do so.

Voiceprints became a sensation almost overnight. The police began to use them in investigations; they were accepted in some courts. Their "inventor" subsequently set up a school and "trained" individuals (mostly police and ex-law enforcement personnel) to carry out the procedures he proposed. This training, plus a part-time apprenticeship, resulted in the establishment of a number of "voiceprint examiners".

Volume 1, No. 1 2013 | DOI 10.5195/lesli.2013.3 | http:// lesli-journal.org

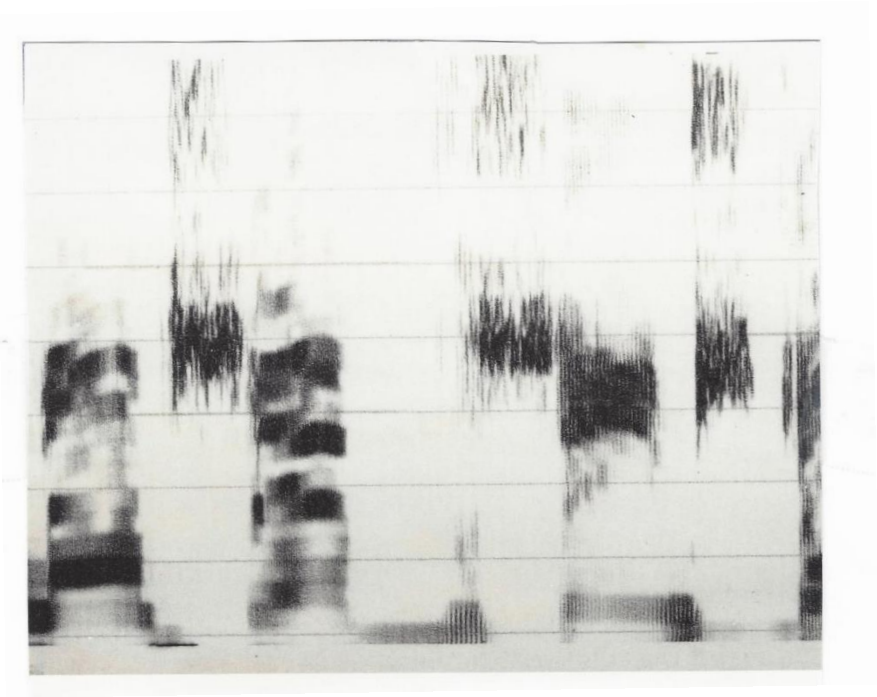


Figure 2: A printout of a time-frequency-amplitude spectrogram (Sonogram) of the type used as a basis for "voiceprints". As can be seen, the bars result from vowels (their formats) and the vertical noise bands from consonants (fricatives); time is along the horizontal axis; frequency the vertical.

To the dismay of many professionals, the notoriety surrounding voiceprints operated to block many of the legitimate SI research programs that had been developing. Moreover, controversy ensued as more and more phoneticians and engineers began to recognize just how crude the T-F-A spectrograms were – and how inappropriate they were for SI purposes. Indeed, most of the scientific community was skeptical as to their value and a few members even testified in court relative to their lack of validity[41].

But what about research? While only a few studies on voiceprints were carried out by their proponents, more and more of them were being completed by groups of neutral scientists. And, even though the "system's" proponents 1) added an AP SI procedure to their protocol, 2) stiffened operator training, plus the "apprenticeship," a little and 3) attempted to employ some available data in support of their system, it quickly became evident that this approach to SI was not a valid one. Nevertheless, the death knell for voiceprints was indeed very slow in coming and for a number of reasons. For one thing, it took time to fund and conduct appropriate research; hence, there were serious latencies between the enthusiastic adoption of the cited procedure and the emergence of data showing its inadequacy. This situation was exacerbated by media/public support as well as the misinformation television and motion pictures provided (and still provide to some extent). In short, the presence of voiceprints fostered a well over 20-year delay in identifying many of the speaker specific parameters that actually could be employed for SI purposes as well as the development of systems that could be tested for use. So too were there other losses which occurred as a result of the negative publicity about SI in general and the waste of resources in the particular. Indeed, only a tiny portion of this deprivation was compensated for by the suggestion (from research on voiceprints) that, if properly handled, some speaker specific elements could be expected to exist within spoken vowels. In any event, what could have been the "golden years" of SI progress were pretty much lost in the voiceprint altercation.

Volume 1, No. 1 2013 | DOI 10.5195/lesli.2013.3 | http:// lesli-journal.org/

Current Problems

Logically, it could be expected that, following the demise of voiceprints, the SI area would be reorganized and reasonable progress would then occur. Actually, such has not been the case. The lingering effect of voiceprints delayed the structuring of effective SI programs just long enough to permit other conditions and impediments to be established. Moreover, because of the new problems, the pathways taken by engineers and phoneticians continued to diverge.

The Engineers

This group of professionals has been the happy recipient of an almost unbelievable technological revolution – a revolution that they individually and collectively had a hand in creating. They also were swept up by an operational model which, while somewhat wasteful, has proved to be wildly successful as a basis for a large number of their endeavors. It is the one that was referred to earlier in this review. That is, to make progress in an area, they would find or develop an algorithm which could possibly meet the needs of the target issue; then build a system or design a device to meet those needs. If it worked; fine: if it didn't, the engineers would simply find/develop another algorithm and proceed in the same manner. Ordinarily, they judged, there was little need to apply an integrated or extensive research program. Just a little preliminary testing should suffice. Thus, most engineers have never learned to be scientists as; actually, they do not need to. The concepts they require are provided by physical scientists and those individuals among their ranks who actually conduct research. It is difficult to fault them because they have been enormously successful in many areas. Not convinced? For but a small sample of their success, simply check out your TV, iPhone or computer. Indeed, these, and other, advances in engineering now appear to be poised to make much of science fiction a reality.

As would be expected, the changes cited above also have led to a second trend. For those who are interested in the general area of speaker *recognition*, the logical shift is to speaker *verification* and its commercial promise. In turn, this trend has led to the development of a number of automatic speaker verification systems. For example, some of the early ones showed promise[51], as do several of the more recent ones[5, 50, 59]. Unfortunately, however, very little confirmatory data – hard data, that is – about either their success or failure is available. There are a few professionals and organizations who have volunteered to test systems of this type, but their success in attracting relevant "customers" (and, hence, the generation of data about these systems) is not known to the undersigned.

Finally, and as you might expect, engineers tend not to have much of a background in the behavioral sciences -- in the anatomy and physiology of speech, or in psychoacoustics. Yet, these parameters all play important roles when a person attempts to understand, and compensate for, the many variables that impact the speaker identification process. Inadequacies of this type make it difficult for engineers to develop systems which can identify specific speakers unless they are in a sharply controlled environment and their utterances vary but little from the speech contained in the reference sets. Of course, this situation can be rectified to some extent simply by having the engineers cooperate with phoneticians and other behaviorists. An upgrade of this type also is critical to success when engineers attempt to build fully automatic SV or SI systems.

The Phoneticians

This type of professional does exhibit some advantages when speaker identification (SI) is under consideration. For one thing, no matter how electronically sophisticated an SI system is, humans must make most of the decisions relative to its use. That is, most forensic phoneticians are able to recognize when SI should be applied, which system or approach should be selected for use, identify the needed materials and samples, select the particular processes to be employed, monitor them, determine the validity of the results and interpret them. Moreover, to successfully perform these (and related) tasks, the phoneticians must (and ordinarily do) understand a great deal about the basic processes involved. About: 1) the nature and sensitivity of the analysis system, 2) the nature of motor speech mechanics, 3) the sensitivity of the perceptual and auditory systems, 4) the myriad of psychological, acoustical and environmental states that affect SI, 5) language and dialect and 6) many of the other variables that can operate to support, modify or degrade the process. What they do not understand is how to *design and build* semi-automatic and automatic systems. In some instances, they do not really understand how a particular electronic system works in the first place.

It also must be said that phoneticians face yet another problem. While most of them are trained experimentalists, they tend to experience difficulty when attempting to establish SI research programs. They must overcome at least two impedances when they try to do so. The first is of lesser importance than the second but can be debilitating nevertheless. It is that nearly all of them are associated with universities. Thus, they have to meet their teaching, student mentoring, administrative and service responsibilities before engaging in research. Moreover, if they maintain an AP SI practice (as part of their "service"), their non-research workload is further increased. Second, they must face the difficulties in obtaining funds for the research. Of course, if they do win relevant grants/contracts, they can 1) obtain released time, 2) hire assistants and 3) fund other research costs that are routinely unavailable from departmental or center budgets. Unfortunately, SI grant/contract funds are in short supply primarily due to the fact that most organizations want to ***verify*** the identity of a speaker rather than identify an unknown talker. Moreover, even the small amounts (provided for SI support) are reluctantly awarded because nearly all contracting agencies desire rather prompt submission of the project's "deliverables". That is, while it takes years to carry out the research that can support an SI system, they seem to want it available "yesterday".

A second problem. While phoneticians are aware of most of the variables surrounding effective speaker identification and some of the possible ways to control them, they tend not to fully appreciate the significance of the veritable explosion of sophisticated engineering systems and competencies that is occurring currently. Hence, both they -- and the engineers also -- can be reluctant to rely on speaker-recognition "systems" – whether interactive or automatic – for actual use in the forensic/judicial domains[9, 26]. Thus, it is most unfortunate that, at the present time, neither group is appropriately employing the others' many skills, ideas and/or products in their attempts to carry out valid/reliable speaker identification. But what strategies could be established in order to mitigate these problems? Active -- and extensive -- cross-disciplinary cooperation would be a start. But first, it may be helpful to clear up some of the misunderstandings about speaker identification that exist currently, and lay some groundwork.

An Undervalued Discipline

As indicated, concerns have been expressed about the validity of *any* approach to speaker identification. To some extent anyway, these concerns appear justified. Even the most knowledgeable of our professionals have travelled down blocked SI pathways or have been overwhelmed by its sheer complexity. However, much of the confusion actually appears to have been caused by a lack of understanding – and clear descriptions – of the positives that do exist. In turn, good perspectives suffer from an inadequate structuring of relevant concepts, relationships and

Volume 1, No. 1 2013 | DOI 10.5195/lesli.2013.3 | http:// lesli-journal.org/

strategies. Moreover, many groups interested in SI have fallen into a trap similar to the one currently being experienced by psychologists. As you would expect, members of that discipline have developed a large number of theories about various aspects of human behavior and, over the years, most of these theories have generated a cadre of proponents. Unfortunately, these individuals have tended to cluster themselves into generally isolated communities where little communication or interaction exists with other psychologists or groups outside their hardened "compounds". Mischel [44] has referred to this as psychology's "toothbrush problem". He writes that "psychologists treat other psychologists' positions like toothbrushes – as no self-respecting person would want to use anyone else's". On the other hand, Bower[6] simply identifies the problem as one resulting from "closed thinking". These two psychologists, plus others[12, 15, 55], decry this situation. That is, that too many of the mainline psychological theorists dismiss or ignore the work of others as being a threat to their own[56]. Further, they stress that, without scientific competition and debate, much of the current research in psychology goes nowhere.

Unfortunately, much the same thing is happening in the SI area, especially with respect to the current approaches and systems. Just as with psychology, progress here will require a thrust to counteract these insulating trends. Until this is done, we will not be taken seriously by members of the other forensic disciplines – nor, more specifically -- by law enforcement, the military or the courts.

Of course, one of our problems is that legitimate knowledge about SI is not at all well-organized or presented. That is, proper conceptual structuring has not been provided about the sensitivity and acoustic processing of the human auditory mechanism and how it can be adapted to 1) accurately process acoustic signals for SI purposes and more importantly 2) provide a model upon which to pattern electronic SI systems. It is of vital importance that these "basics" are articulated and organized properly if any real progress in speaker identification is to be realized. One way to start this upgrade would be to focus on the data that is available.

First, the three elements of subjective SI, as described above, must be put into perspective. They are 1) the identification of a person from memory, 2) the nature of earwitness identification and 3) speaker identification by trained, experienced professionals. The first two of these elements are of little consequence; the large corpus of research to be discussed provides the base for the third. That is, the act of court testimony by a lay person provides little quantitative information about SI; all it furnishes is some understanding about listener talent, plus memory for voices and other acoustical signals[11]. Second, memory for speech/voice also is important for conceptualizing earwitness lineups – i.e., the process where a person who has heard, but not seen, a perpetrator then attempts to remember his or her voice and identify it when mixed with a group of foil voices[20, 61]. If the limited information generated from the study of these first two areas were all that was available for identifying speakers, but little success could be expected. Worse yet, many who interface with SI in some manner seem convinced that this is the way all SI "systems" operate even when they are employed by the third group – i.e., by the trained, experienced professionals and their highly organized protocol. But, of course, this is **not** the case at all. Rather, there is a wealth of information and data available which provides a robust basis for the structured speaker identification approaches carried out by phoneticians or by computer programs. The chaos that surrounds the view that the model for speaker identification comes from how the lay public carry out SI tasks is misleading at best. Rather, it should be one based on the substantial body of data about speaker-specific elements, relationships and procedures. This is the extensive and sophisticated corpus of information that is undervalued. It will be briefly reviewed before the abilities of the practitioners or the possibility of machine processing is considered.

To reiterate; what is not appreciated – even by many phoneticians – is the existence of the relevant SI databases, the tools that are available and the effective SI procedures which have been developed. Also not appreciated is the fact that the phonetician or examiner is free to take the time necessary, plus apply all procedures desired, in order to permit intelligent decisions. Thus, and as stated, the value of this entire area is badly underestimated as a foundation either for practice by phoneticians or as the model for machine-based systems. Nor has the relevant information

been properly integrated into a predictive model. It is hoped that the next four sections can function as a brief outline of the material that can counter these misconceptions about SI.

Early Research

A good place to initiate this review is with a quick look at the early SI studies. The earliest of these efforts involved analyses of lay listener responses to speech samples (later, more sophisticated auditors were used also). Investigators employed either 1) talkers who were known to the auditors or 2) experimental subjects who were familiarized with the speaker's utterances by some sort of training. As stated above, McGeehee[43] carried out a study of the latter type way back in 1937. She presented listeners, who had received some training, with live utterances of sentence-length material, spoken by the target subjects. The samples were embedded in sets with the others provided by distractor speakers. She found the identification rates to be quite high (over 80% in most cases) when the trials were conducted immediately after training but that these levels decayed over time. Later she replicated these experiments with recorded samples, the results were similar. Even though these studies were rather primitive, her basic findings of the level[7] and speed of decay[48] have been validated.

When using familiar voices, Bricker and Pruzansky [7]obtained even higher correct identification scores (i.e., 98%) for sentence length stimuli. While their correct response levels were lower (mean = 56%) when the samples were very short, both data sets were statistically significant. Please note also that, as early as 1954[49], data became available which suggested that modest identification accuracy could be attained for even *extremely* short stimuli. In that case, the authors demonstrated that scores would systematically increase for samples of up to about 1200ms in length. Further increases resulted not from greater duration but from the presence of a more extensive phonetic repertoire. Since that time, many dozens of similar studies have been carried out and reported. Different speakers, auditors, utterances, dialects, recording fidelity environments, and so on were studied. They found that, overall, humans demonstrated rather good innate SI abilities. Of course, many factors enhance or degrade accuracy[18, 19]. On the plus side are those of auditor familiarity with the target speaker, reinforcement, a good acoustic environment, good quality samples, extensive samples, enhancement of listener capabilities, professionally trained examiners and so on. On the other hand, accuracy is degraded if some of the above conditions are not extant, if noise and distortion are present, if the target population contains "sound-alikes" or large numbers of talkers, etc. In any event, a substantial amount of evidence about speaker-specific elements is available from the hundreds of now available studies. The data here suggests that humans can make remarkably good identifications under many different conditions and that these skills can be augmented. These primary relationships provide a number of core concepts both for testing and as a basis upon which more complex relationships can be studied. They certainly aid in the building of platforms that permit deployment of useful SI procedures.

Auditory Capability

Some of the results presented in the previous section may appear vulnerable to challenge. Can the human auditory system *really* permit judgments as precise as those reported, especially when conditions are somewhat unfavorable? Without question, a position in the affirmative can be argued. The research corpus to follow will provide data about this very – and somewhat unexpected – sensitivity of the human auditory mechanism. Specifically psychoacousticians and neurophysiologists have long studied how the hearing process operates (from the external auditory meatus all the way to the cortex and, then, to the brainstem) and just how sensitive it may be. Even though but little of this research on auditory capability has focused specifically on speaker identification, strong evidence of the human auditory mechanism's sophistication can be demonstrated – and in a number of ways. For example,

Volume 1, No. 1 2013 | DOI 10.5195/lesli.2013.3 | http:// lesli-journal.org/

Kraus and Nicol[35] provide a general review of some of the auditory differences found between musicians and untrained subjects. They describe certain of the absolutely remarkable auditory processing abilities possessed by both groups (musicians are better, of course) – and, hence, by humans in general. In doing so, they (and others) describe how well speech embedded in noise can be processed (Bronkhorst[8]), the speed by which complex sounds can be decoded[37] (i.e., in only 10-12ms), the enhancement of speech processing by training and/or experience[37, 58] (a very important set of concepts), how emotions can be coded[58], how language[60] (including tonal), plus speech and speakers[36, 38], can be assessed and so on. Other research in this area supports the validity of their positions -- and ours. In short, the hypothesis that the auditory system is capable of very rapid -- and accurate -- processing of the type necessary for accurate speaker identification is supported by these, and related, experiments.

Research on Speech Features

The third area of research is focused on identifying those speech features which serve to support identification accuracy. That is, a substantial number of studies have been carried out investigating which speech/voice features lend themselves to auditory processing for identification purposes. Many of these investigators focused on suprasegmental elements such as SFF, i.e., speaking fundamental frequency[4, 31], voice quality[14, 25] (long term specta), temporal speech features[29, 32] and related. Also researched for SI purposes were vowels[40, 46], (especially vowel formats), fricatives[53], nasality and related segmentals[45, 47]. The observed relationships have been revisited again and again over the years; they have tended to confirm and expand Steven's[57] early review of the segmental and suprasegmental elements that can be employed in the speaker-identification process. The power of this extensive research corpus demonstrating how speech/voice parameters fit into the identification process should not be underestimated. Related clusters of them can be most useful in the development of speech vectors. Of course, the reader should be reminded that none of these speech units (i.e., parameters or vectors) can, when applied alone, provide acceptably high identification levels. On the other hand, as with speaker specific elements from other domains, clusters of them prove to be identity sensitive. Especially effective are multi-vector profiles.

Forensic Experiments

Perhaps a particularly relevant set of illustrations can be found among those SI studies directly related to forensics. First, it must be pointed out that most of the experiments of this type were based on subjects having but brief encounters with very limited stimuli – sometimes ones with an utterance duration of only a vowel or a word -- under experimental procedures that were quite exacting. Consider how difficult it would be for a listener to identify a person "known" to him or her only from a fairly brief exposure to their speech – followed by samples which are presented in noise or embedded among utterances produced by other individuals. Many of the experimental tasks upon which research of that type was based are as challenging to the auditor as is the above example; few are markedly easier. Yet it is clear that, even under such circumstances, several of the experiments to follow provide yet another kind of confirmatory support for the SI hypothesis articulated above. That is, they serve to yet further demonstrate just how discriminative the auditory mechanism is and how well "it" can perform for SI purposes -- even within the sharp limitations of the forensic model.

Three examples (of the many investigations of this type) should provide sufficient evidence here. In the first of these, Shirt[54] obtained short speech samples from a large number of talkers (suspected perpetrators) which were provided to her by the British Home Office. After some exposure to these utterances, her listeners were requested to identify specific talkers (from among the others) when they were heard in sets. Two groups of listeners were used: 1) phoneticians and 2) university students. Although some students did as well as the phoneticians, the success of

Volume 1, No. 1 2013 | DOI 10.5195/lesli.2013.3 | http:// lesli-journal.org

the professional group (i.e., the means for) was higher. A remarkable aspect of this research was that both groups did quite well even though the speaker recognition training was limited and the stimuli relatively brief. Second, Koester[34] and his associates carried out similar studies but with known voices. Both his student listeners and the Phoneticians did quite well in identifying the speakers. Remarkably, not one of the Phoneticians made as much as a single error in the identification process.

One experiment[28] in particular can be said to provide pivotal information in the area. Three groups of listeners were studied; 1) individuals who knew the talkers very well, 2) those who did not know them but who received about two hours of training in recognizing their voices and 3) a group who neither knew the speakers nor the language spoken (but who also were briefly trained). The talkers were 10 adult males who produced a number of 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 (two different samples of each of the three speaking conditions) and had to identify each by name. The results are best understood by consideration of Figure 3. As may be seen, the accuracy of the listeners who knew the speakers approached 100% correct identification for both normal and stressed speech; further, they could often tell (80% accuracy) who the talker was even when disguise was permitted. The university students did not do as well, of course, but their native processing abilities, and the limited training they received, allowed them to succeed from double chance to four times better, depending on the speaking condition. Even a general population of native speakers of Polish (unfamiliar with English) exceeded chance – and for all conditions. These data can be considered of yet greater import when it is remembered that the presentation procedure was an extremely challenging one. That is, all 60 samples were presented in a single trial and listeners had to rapidly identify the heard talker, find him on the check sheet and, then, place a corresponding identification number on an answer form before the next sample in the sequence was presented.

The sampling of the SI and related investigations provided in these four sections were drawn from hundreds of research reports. Collectively, they provide a brief summary of much of what is known about how well humans can function to identify individuals from their speech. The data also provide robust support of a hypothesis, proffered 60 years ago by Hecker[17], namely that the human ear (i.e., the auditory mechanism, including the brain and hind-brain) is the most powerful speaker identification system possible. In turn, this position supports this author's contention that it would be best if computer systems were designed, constructed and then "trained" to replicate the SI processing abilities of the human "ear". Enough is known already to permit a workable system of this type to be developed and tested. The advancing power of the computer would also permit programming control of variables outside the speech/voice domain (which are already known) or as they are identified by new research and/or by system operation in real life situations. So too could the speaker-specific speech parameters and vectors be upgraded by this same process.

Volume 1, No. 1 2013 | DOI 10.5195/lesli.2013.3 | http:// lesli-journal.org/

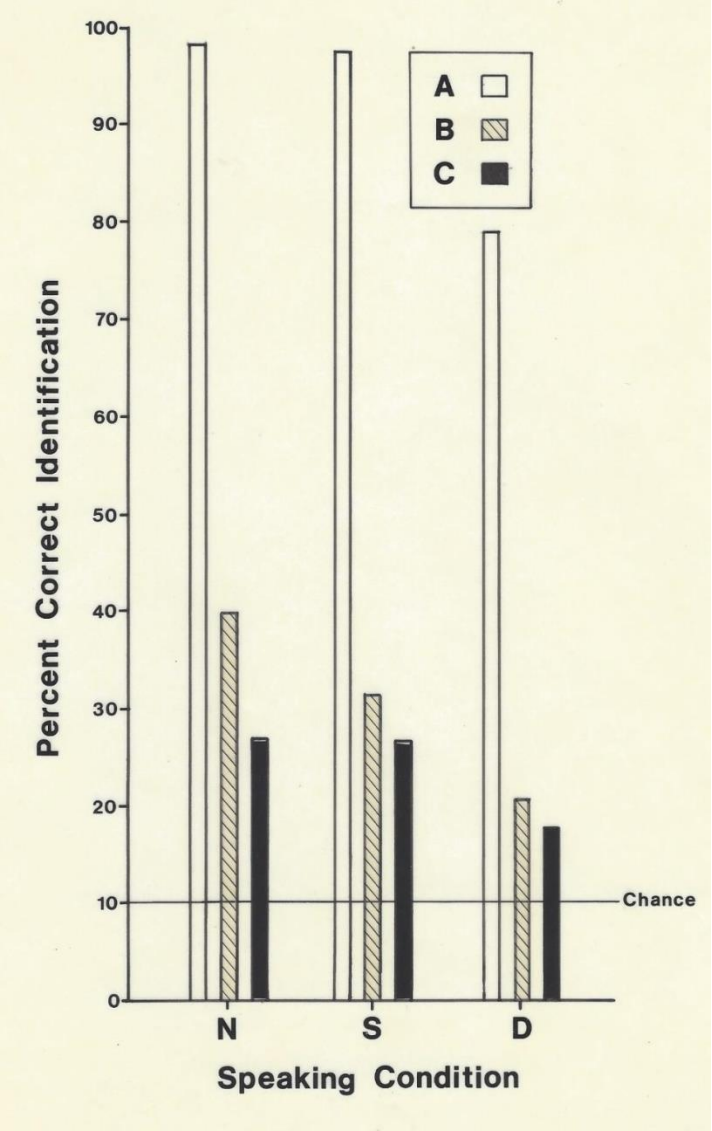


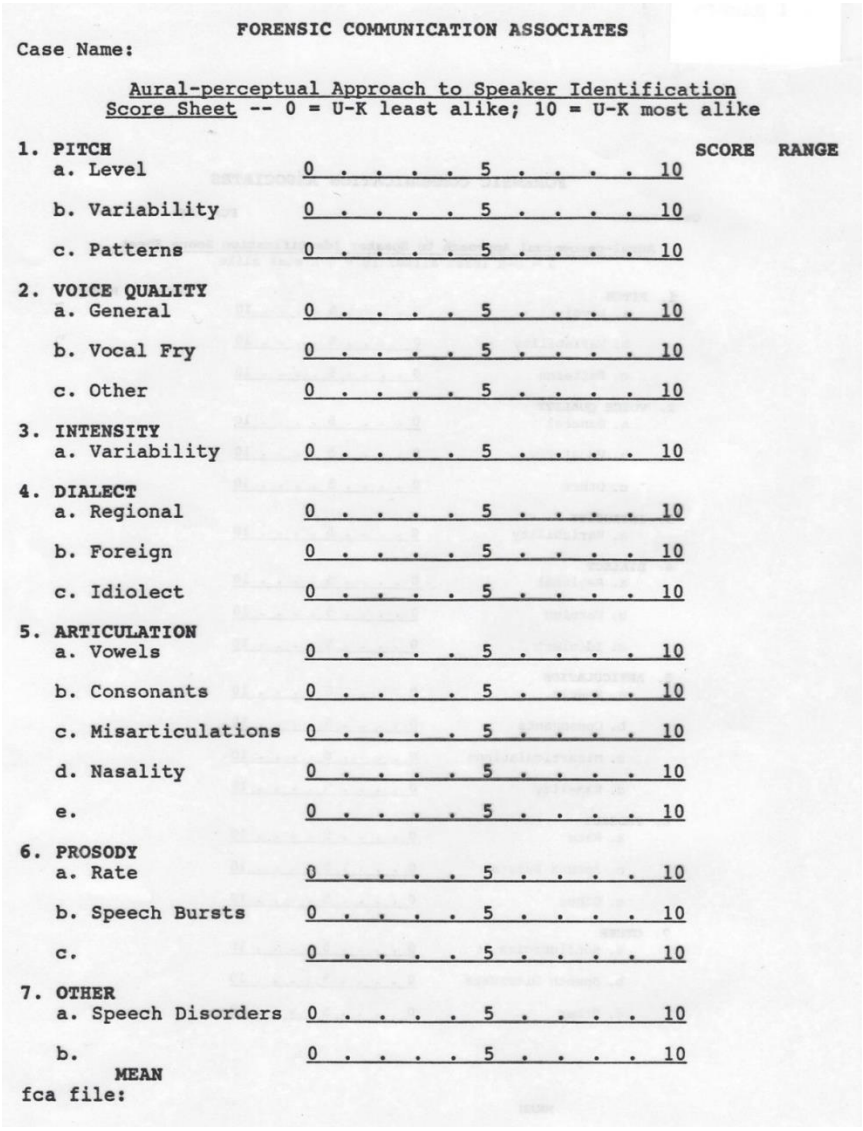
Figure 3: The means of the correct identifications of 10 talkers speaking under the conditions of N = normal, S = stress and D = disguise. The three listener groups consisted of: (A) individuals who knew the talkers very well, (B) listeners who did not know the talkers but were trained to recognize them and (C) those that were trained similarly but did not know either English nor the talkers. The task was to name each speaker for all six of his presentations out of a total of 60.

The Present

The discussion found above (i.e., in the section titled "An Undervalued Discipline") might sound like an advertisement published by SI practitioners. It should not be taken as such. What is actually provided is a model for individuals who are engaged in developing AP SI. Not only that, but it also provides one for those who are designing semi-automatic or automatic SI systems. So too does it feature a brief review of an extensive and cohesive body of research which can provide the information necessary to understand 1) the identification processes employed by human speakers and 2) many of the variables that impact these processes. Most present-day forensic

Volume 1, No. 1 2013 | DOI 10.5195/lesli.2013.3 | http:// lesli-journal.org

phoneticians appreciate these basics plus the variables they must use, mitigate or compensate for in SI. Indeed, they have tended to model their approaches to AP SI on these elements, vectors and interactions. A number of these professionals, including the present author, have used them in real life SI situations. The author's[19, 23] approach is described as follows.



FORENSIC COMMUNICATION ASSOCIATES

Case Name:

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		
e.	0 . . . . . 5 10		
6. PROSODY			
a. Rate	0 . . . . . 5 10		
b. Speech Bursts	0 . . . . . 5 10		
c.	0 . . . . . 5 10		
7. OTHER			
a. Speech Disorders	0 . . . . . 5 10		
b.	0 . . . . . 5 10		
MEAN			

fca file:

Figure 4: The AP SI form developed for use with the aural-perceptual speaker identification approach described in the text.

As can be seen in Figure 4, a number of parameters (not quite vectors as yet) have been organized into groups along low-to-high continua – for which a 0-10 range provides a simple metric scale. Note that the first parameter to be judged is fundamental frequency or heard "pitch". In this case, a simple vector actually can be seen being established as "pitch" is made up of level, variability and patterns. These three parameters can each be detected aurally and -- just as important -- they can be physically measured. While many of the parameters that follow pitch are suprasegmental in nature (i.e., voice quality, vocal intensity, speech timing, etc.), this approach also includes a number of segmentals (that is: vowels, diphthongs, consonants, consonant clusters, nasality, etc.). In any event, the protocol here would be to compare speech samples uttered by an unknown talker (U) with those of a known one (K).

Volume 1, No. 1 2013 | DOI 10.5195/lesli.2013.3 | http:// lesli-journal.org/

That is, to *repeatedly* assess pairs of U vs. K voice samples, parameter by parameter (but one parameter at a time), until every one of them (i.e., those that can be judged) has been rated on the basis of their U-K similarity. After a number of trials, a judgment can be made as to how likely it is that the two voices being assessed were produced either by a single individual or by two individuals. Judgments falling in the 0-3 region suggest two different talkers; whereas scoring in the 7-10 range suggests but a single one. As might be expected, structured approaches such as this one (although labor intensive) have served forensic phoneticians quite well and for many years. Some data on validity assessments of these systems (both experimental and field) have been published with good accuracy (and reliability) rates reported. So too were they satisfactory in those cases where verification experiments were carried out or external confirmation was available. It also should be noted that these data further confirm the material previously cited, i.e., that many of the elements existing within the speech signal can be clustered into a reliable identity profile. So too have they met the Daubert[10] standard. The procedure described has been accepted well over 50 times in the United States, and in both state and federal courts. Lastly, this SI system has been paralleled at the University of Florida by a computer-based semiautomatic SI system which will be described in a latter section. It also provides a reasonable basis for other approaches.

Collaboration

After the steady flow of complaints about events and conditions which have impeded or blocked the orderly development of effective speaker identification, one might expect to encounter yet more gloom and despair. Or, on the other hand, a change of pace might be expected; say, one promising an "easy fix". Actually, neither will be forthcoming. That is, it should appear obvious by now that good progress should result from groups of phoneticians and engineers forming teams and going to work. But, is there any evidence that cooperation of this kind actually would be helpful? Happily, at least a few thrusts of this nature can be described as having exhibited moderate success. While they have not been earth shattering, they, at least, should provide some guidance. Four examples are offered in support of this contention.

The first would be a review of how the AP SI system described immediately above came into being. It took a team, led by the undersigned, several years to develop and test it, and then several more years to use it in the field and further refine it. While it was developed primarily by phoneticians, a strong assist was provided by bright, young graduate students and cooperative engineers. In practice, it has been shown to exhibit rather high "hit rates" and very few false positives. That is, evidence is available that if U=K, that particular finding will be forthcoming, but if U≠K, the suspect will not be erroneously identified. However, while it is relatively easy for a practitioner to transfer this system from its home base to a new location, it has been found to be a little difficult for them to learn to use it – at least, without direct study with its authors. In addition, even though it is highly structured and controlled, the effect of its subjective components cannot be discounted. Accordingly, it now would appear desirable to develop supplemental computer programs that will mimic the operations here – as well as the judgments now being made by the phonetician. Finally, it must be said that this procedure is work intensive and validation research on it is very much so.

The second example shifts the procedural responsibilities more toward the engineers. That is, in this particular instance, the main component of the developmental team consisted of engineers[3, 52], with the phoneticians (it appears anyway) serving as consultants. In this case, the research program is being carried out by an investigative team spread throughout Australia and Malaysia. As you would expect, their thrust is on verification rather than on SI. Nevertheless, it is possible to learn from their research and how they test the systems they have fabricated. Basically, they describe their program as one where they are building and evaluating the performance of a series of automatic speaker recognition (ASR) systems. When one is ready, a typical SV procedure is carried out. It is first tested

directly and, then, by comparing its accuracy to that of human listeners. In most past instances, the hit rate obtained from the humans has exceeded that of the particular ASR system being evaluated. However, in the latest of their experiments[16], their newest system appears to be catching up and this trend has been reversed. That is, the authors found that their new "human assisted" ARS system achieved an 80% accuracy rate whereas the listeners ("operating in fusion") achieved the 78% level. In this instance, however, the phoneticians should have provided a stronger input to the human recognition tasks. Specifically, the SI samples for all subjects (who were drawn from three countries) were presented in English even though none of them were native speakers of that language. Indeed, five different (other) languages were listed by the auditors as their "first language". Moreover, since the task was one of verification, not identification, they were allowed to directly compare the speech samples. Nevertheless, these investigations 1) demonstrate innovative approaches to speech recognition research 2) provide yet further validation of the hypothesis that the human auditory mechanism can operate as a formidable identification system and especially 3) demonstrate the effectiveness of mixed discipline teams. On the other hand, it also can be suggested that SR teams should have a better phonetician-engineer balance and that difficulties should be expected if the team members are not in reasonable proximity to each other.

The third example demonstrates one of the better approaches; it is one which balances project responsibility between the relevant disciplines – especially if the teams are able to sustain the effort over a period of time[24]. The system described here is a semi-automatic SI approach developed by groups led by the undersigned[18, 19]. Specifically, the research program was initiated with an assist by a second phonetician, an engineer and two student assistants. Through the years, several other faculty-level investigators (of both types), assisted by post-doctorals, and several graduate assistants, were associated with the project -- and for various lengths of time. The research program employed may be best understood by consideration of Figure 5. Using this model, a large number of parameters and vectors were generated and tested at various levels of difficulty. Research was sometimes informal but, usually, it was not (i.e., it involved experiments); also included in the development and testing phases were several theses and dissertations[11, 30]. While the development was lengthy (often tedious), the resulting process can be summarized as follows.

The mechanism was named SAUSI (Semi-Automatic Speaker Identification) system[27]. Since it was discovered early-on that traditional signal processing approaches were lacking, a number of speech/voice features were identified and evaluated[21]. This decision was further supported by 1) the early experiments, 2) the AP SI research literature and 3) the realization that the system combined acoustic parameters in such a way that they emulated how human listeners processed speech for everyday SI. This research eventually led to four vectors, each made up of a number of related speech/voice parameters. They are SFF – speaking fundamental frequency, LTS – long term spectra (voice quality), VFT – vowel formant analysis and TED – temporal patterning (prosody). A full description of these vectors is available[18, 21]. Since no single SAUSI vector, by itself, was found to provide appropriately high levels of correct identification for all of the different situations encountered, the multi-vector profile cited above was developed. That is, the results from the four vectors were normalized, combined into a single unit and organized as a two-dimensional continuum or digital profile (Figure 6). 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 entire process is replicated. The final continuum usually consists of the data from 2-3 complete replications and includes a summation of all vectors. Hence, any decision made about identity is based on a very large number of individual comparisons (factors, parameters, vectors, summations, replications). The results of a number of studies (high fidelity, noise, restricted bandpass, field) have been published. High hit rates and extremely low false positive rates have been reported, especially for the more recent experiments. The product of these experiments is in a form which should permit the results to be compared to those from other systems by Basian likelihood procedures.

Volume 1, No. 1 2013 | DOI 10.5195/lesli.2013.3 | http:// lesli-journal.org/

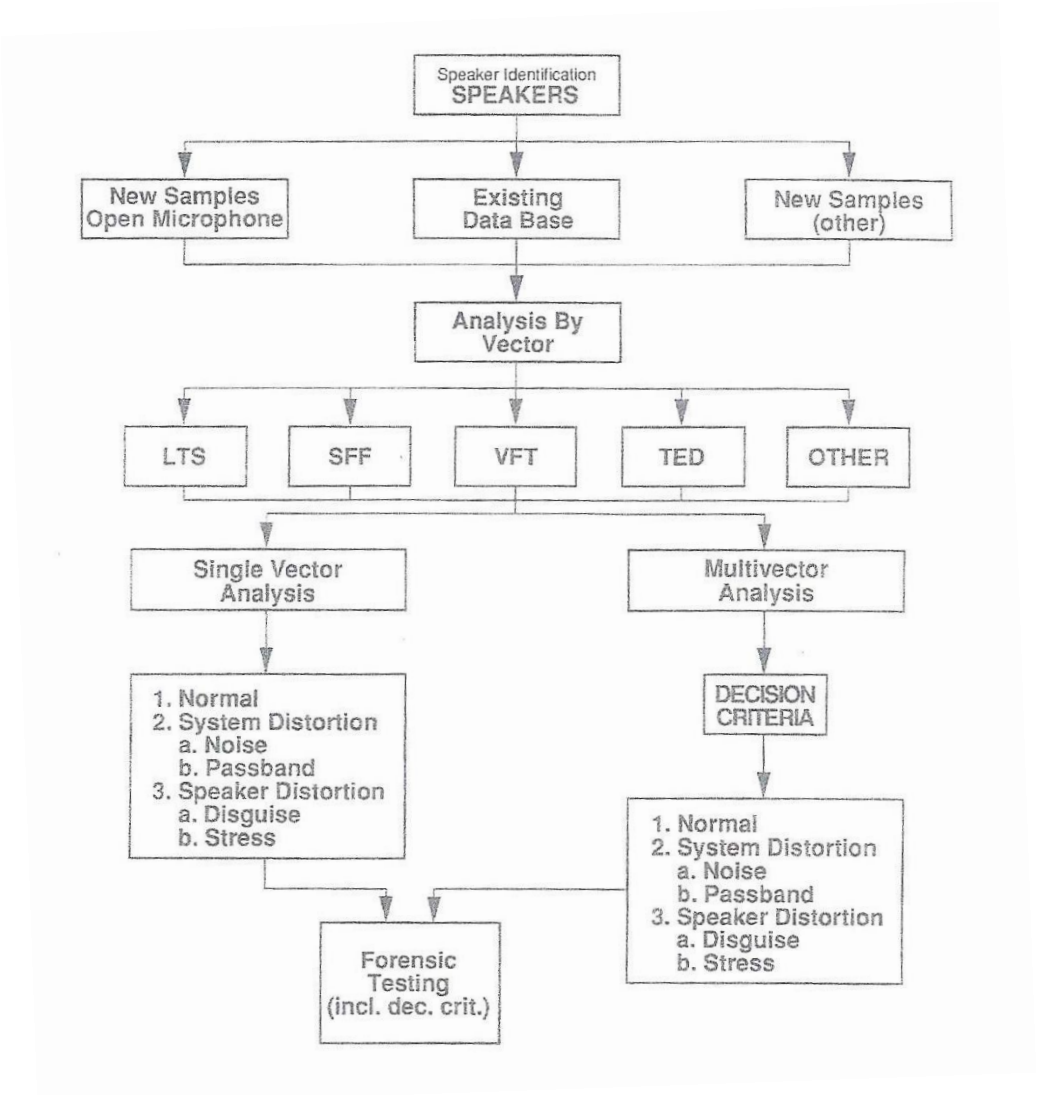


Figure 5: Model providing the basis for comprehensive development of semi-automatic speaker identification systems. The basic model has been modified to include field and forensic research.

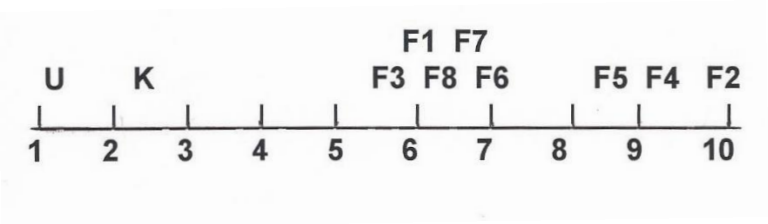


Figure 6: A continuum providing normalized data for the unknown, the known and 8 foil speakers. As can be seen K = U. Note that data for U is included for validation purposes.

Volume 1, No. 1 2013 | DOI 10.5195/lesli.2013.3 | http:// lesli-journal.org

The fourth example is one where a forensic phonetician has carried out a number of experiments on a SI system that has been developed, is now on the market and is being used (at least by some groups). The manufacturers provide a rather detailed handbook[2] and report some system analyses. However, it is a phonetician[39] who has rigorously evaluated the Batvox system in a number of ways. He has recently completed/reported a pair of these experiments. The first of them tests the SI sensitivity of the system as a function of six different languages. Then, in a related experiment, its ability is evaluated to accurately identify the target speakers as a function of several different types of telephones -- i.e., the level of correct SI for different channels. The cited research was well designed and rigorous. The fact that low to very low error rates (depending on the experimental condition) were found is of credit to the system. As a critique, it must be said that the research cited was appropriate and useful; it certainly added to the legitimacy of this device. On the other hand, it must also be said that this type of research -- plus many other studies -- should have been routinely conducted and reported during its developmental years.

As can be seen, each of the four programs reviewed above exhibited some strengths; however, none of them proved to be ideal nor even were appropriately complete. For any SI approach to be fully validated, research teams must first find support and then either take one of the available systems through a rigorous evaluation program or, starting with a new algorithm, follow a similar developmental/evaluative pathway. What the team(s) need to do and what competencies the device needs to exhibit must be established. However, there yet remains a problem that may be difficult to overcome. It is the problem of proprietary rights. That is, once a system of this type has been completed, the developers ordinarily do not wish to reveal just how it works (for fear it will be pirated). Of course, in some cases, limitations of this type may be reduced; moreover patents can have a positive effect here. In any event, it is hoped that this issue will not prove to be too restrictive.

To Sum

Finally, phoneticians (and linguists also) must face the fact that they sometimes have played a role in intensifying the severity of the barriers they have encountered. Some of these effects can be deduced from the above narrative; others are not so obvious. On the other hand, it appears right now that the best course of action might not be to intensify the discomfort being felt. Rather, it would appear useful to make a few positive suggestions; ones which might ease some of the more negative feelings that continue to exist. Accordingly, let us consider a few of these positives.

First, please consider coming out of your respective cocoons - - as warm and cozy as they may be. Check out what other people are doing, what they have accomplished, what they are saying, and what you can learn from them. Refuse to remain completely anchored within your own group. There is a lot to be learned from other forensic specialists, of course. But it is even more important to tap into the knowledge and product -- and even the opinions and attitudes -- of engineers and other legitimate professionals working in ASR, SV and SI areas. You just may be surprised as to how much you can learn from these many and varied sources.

Second, publish, publish, publish! Spread your ideas, your research, and your case reports out for others to see and react to. Check out, especially some of the new on-line journals. One example is a publication which considers articles on case reports, technical notes, procedural descriptions, position papers and the like. It is the fairly new "Investigational Sciences Journal", J.M. Adkins[1], Editor. And how about this journal (Carole Chaski, Editor). Further, besides "The Journal of Speech, Language and the Law" (R. Butters, D. Eades, P. Foulkes and P. French, Eds.), there are several Phonetics and/or Linguistics publications for those of us who are conducting research in Forensic Phonetics or Forensic Linguistics, The American Academy of Forensic Sciences' publication, the "Forensic Sciences Journal" is also a key platform. Your work will be noticed there. Indeed, it will be noticed across

Volume 1, No. 1 2013 | DOI 10.5195/lesli.2013.3 | http:// lesli-journal.org/

all forensic disciplines -- as it also will when placed in "Forensic Sciences International", "Acustica" or any of the many engineering and IEEE journals. Finally, if you think your research will pass muster, "The Journal of The Acoustical Society of America" is one of the world's top rated journals, with respect to **all** fields/disciplines. And, "no" you do *not* have to pay to be published in JASA. All you have to do is get approved by the reviewers.

Perhaps this tirade has gone on too long. Actually, however, there are a couple of "last points" that should be made 1) we should organize our field better. We need to describe it, to define it and to communicate it. First, to ourselves, and then to all other individuals and groups with whom we interface or *may* interface. In this regard, we have been a little "light" in our effort to define what we are doing, and where we want to go. We especially need to get better at establishing our standards[26]. For example, what level must be reached if one is to conclude that an SI match has been made? What are the standards which must be met before it can be said that a recording or document has been authenticated?

Finally, it is important that we organize and structure our procedures either before, or as, we use them. We should do this before we describe them to others – especially to the professionals in allied fields. To do so will assist us in being taken seriously by other forensics specialists. Another way to do this -- and to validate what you do -- is to *test* your approaches and procedures. And then test them again. If you are conducting AP SI analyses, should be able to provide your hit rate *and* your false positive rate -- both directly and by means of tests external to your procedures. If you are using machine/computer approaches, you should do the same both in the laboratory then and in the field. In short, evaluate yourself as an evaluator.

Finally, the best of luck to you.

A Postscript on the Trayvon Martin Case:

It is important that you take a guarded position when confronted by pronouncements or articles made by others relative to the forensic areas in which you may be working. Your response to, or analysis of, "reports" found in the media should be fair but rigorous. It also is important to be especially wary of persons who claim to be "experts" but who do not seem to actually operate at that level. Be cautious also of individuals who are "wannabes". The "announcements" they make can be both misleading and harmful. A good example of the above is the experience I and my colleague (J.D. Harnsberger) had with the recent Zimmerman[13] murder trial. This case involved a "neighborhood watch" individual who approached, fought with and then (when losing) shot a young man he had been observing. The entire incident and what followed are quite complex, however, they have been extensively reported. What is important here is that a woman residing in the house next to where the altercation took place heard it and called 9-11. It was by this link that a recording – a rather poor one unfortunately – was made of about 20 utterances made by the two men; and then the sound of a firearm being discharged. We were retained as consultants by the first group of district attorneys and investigators assigned the case – and later by the second group (after the first had been removed by the governor of Florida). When the incident is considered, it first must be remembered that the complete series of utterances were not all captured (the call was made *during* the fight), and that most of them were distorted and/or, masked either by noise or the speech of others. Only about eight rather faint calls could be extracted from the recording for processing and they were very short (less than 10 seconds total).

We advised the prosecutors that little in the way of speaker identification could be successfully carried out. However, during the delay caused by the case reassignment, two individuals reported (in the media) that they had heard the 9-11 recording and determined what type of speech was available (all cries for help), what actually was said (viz "I'm begging you" "help me, please help me" etc.) and who made the cries. In all cases, they said it was the

"victim" (T. Martin). One of these two "experts" was a phonetician who had carried out some SI research as a student but who had not otherwise exhibited much activity in forensic phonetics. The other individual was clearly a "wannabe" as he made his living as a private detective. Neither seemed to be aware that 1) the recording was not complete due to the earlier exchanges not being recorded, 2) much of the speech was mixed with other voices (the caller, 911 operator and another male) and 3) clearly some of the earlier calls heard on the recording were the type of utterances (grunts, gasps, etc.) associated with a fight. Yet these two "experts" assigned them all to the victim and indicated that they were all calls for help. Nevertheless, at the request of the State's attorney (as stated, we had indicated that little if any decoding or speaker identification could be carried out), we agreed to see if we could generate some useful data.

They provided us with a large number of "exemplar" type recordings of both principals. Included was a reenactment of the incident by G. Zimmerman. These recordings were not available to the above cited "experts". Needless to say, it certainly was a challenge; it took time and every standard procedure we could apply (plus a couple of new ones) to generate any information at all. We were only able to isolate eight out of the 20 possible utterances as being "clean" enough for analysis, and only generated useful information on half of those remaining[22]. When all the data had been obtained, synthesized, and evaluated we found – at about a 60% confidence level – that each of two of the early calls were vocalizations (not cries for help) by Martin. We also judged that the last two utterances were what could have been cries for help (or of pain), and were by Zimmerman. This second set of judgments could be made at only about a 65% level of confidence. As can be seen, none of these judgments reached the match category. Intelligibility also continued to be a problem mostly because the utterances were very soft, masked by noise and were only grunts or calls with but a very few words of any kind produced (much less intelligibly).

We realized, at the end, that the prosecution was not keen about our findings – they much preferred the pronouncements of the "wannabes". But, the data were what the data were – no more, no less. Nor were the defendant's lawyers enthusiastic about our product – which hardly was grist for their case. So they found two witnesses (G. Doddington, P. French) who testified that no judgments, at all, could be made. Our position was that we had not made a match nor could we tell who was vocalizing beyond the modest suggestion articulated above. In short, we were pleased not to have to try to defend what we found as it was not really relevant. But the fact remains; you have to be wary of "wannabes". While this pair didn't do any real harm to the trial, they materially complicated things. Worse yet, they caused the families of the principals, considerable grief and confusion.

Volume 1, No. 1 2013 | DOI 10.5195/lesli.2013.3 | http:// lesli-journal.org/

References

Note: This article reviews so many events and experiments -- those occurring over such a long period of time -- that over 300 references would be needed to fully document them. However, in order to reduce their number to a manageable level, certain steps were taken. First, the well-known "rule of three" was imposed. In addition, a reference was included only when 1) identification of an event or project was absolutely necessary or 2) further explanation of a concept was considered desirable. Finally, when any of many dozens of references would be relevant, only the best or most important was included.

1) Adcock, J.M., (Editor) Investigative Sciencs Journal, Contact: jmadcock@jma-forensics.org or www.investigativesciencejournal.org
2) Agnitio, (2009) Batvox 3.0, Basic User Manual, Madrid, Spain
3) Alexander, A., Botti, F., Dessimoz, D. and Drygajlo, A. (2005) The Effect of Mismatched Recording Conditions on Human and Automatic Speaker Recognition in Forensic Application, Forensic Sci. Internat., S95-99.
4) Atal, B.S. (1972) Automatic Speaker Recognition Based on Pitch Contours, J. Acoust. Soc. Amer., 52: 1678-7697.
5) Beigi, H. (2011) Fundamentals of Speaker Recognition, Secausus, NJ, Springer.
6) Bower, B. (2013) Closed Thinking, Science News, 183: 26-29.
7) Bricker, P. and Pruzanzky, S. (1966) Effects of Stimulus Content and Duration on Talker Identification, J. Acoust. Soc. Amer., 40: 1441-1450.
8) Bronkhorst, A.W. (2000) The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-talker Conditions, Acustica, 86: 117-128.
9) Campbell, J., Shen, W., Campbell, W. Schwartz, R., Bonastre, J.F. and Matrouf, D. (2009) Forensic Speaker Recognition, IEEE Signal Processing Mag., March: 95-103.
10) Daubert vs. Merrel Dow Pharms Inc., (1992) 509 U.S. 579, 113S. CT 2786.
11) DeJong, G. (1998) Earwitness Characteristics and Speaker Identification Accuracy, PhD dissertation, Univ. of Florida.
12) Fiedler, K., Kutzner, F. and Krueger, J. (2012) The Long Way from -Error Control to Validity Proper, Perspect. Psychol. Sci., 7: 661-669.
13) Florida vs. Zimmerman, (2013) No. 1712F4573 Circuit Court, Seminole County, Florida.
14) Gelfer, M.P., Massey K.P., and Hollien, H. (1989) The Effects of Sample Duration and Timing of Speaker Identification Accuracy by Means of Long-term Spectra, J. Phonet; 17: 327-338
15) Gigeenzer, G. (2010) Personal Reflections on Theory and Psychology, Theory and Psychology, 20: 733-743.

16) Hautamäki, V., Kinnunen, T., Nosratighods, M., Lee, K.A., Ma, B. and Li, H. (2010) Approaching Human Listener Accuracy with Modern Speaker Verification, In INTERSPEECH-2010, 1473-1476.

17) Hecker, M.H.L. (1971) Speaker Recognition: An Interpretive Survey of the Literature, ASHA, Monograph #16, Washington, D.C.

18) Hollien, H. (1990) Acoustics of Crime, New York, Plenum Press.

19) Hollien, H. (2002) Forensic Voice Identification, London, Academic Press Forensics.

20) Hollien, H. (2012) On Earwitness Lineups, Investigat. Sci. J., 4: 1-17.

21) Hollien, H. and Harnsberger, J. (2010) Speaker Identification: The Case for Speech Vector Analysis, J. Acoust. Soc. Amer., 128: 2394A (and submitted)

22) Hollien, H. and Harnsberger, J. (2013) Attempted Speaker Identification: Florida vs. Zimmerman (1712F4573), submitted to the Office of the State Attorney, Fourth Judicial Circuit, Jacksonville, FL.

23) Hollien, H. and Hollien, P.A. (1995) Improving Aural-Perceptual Speaker Identification Techniques, Stud. Forensic Phonet., (A. Braun and J-P Köster, Eds.) Wissenschaftlicher Verlag, 64: 87-97

24) Hollien, H. and Jiang, M. (1998) The Challenge of Effective Speaker Identification, RLA2C, Avignon, France, 1: 2-9.

25) Hollien, H. and Majewski, W. (1977) Speaker Identification Using Long-term Spectra Under Normal and Distorted Speech Conditions, J. Acoust. Soc. Amer., 62: 975-980.

26) Hollien, H. and Majewski, W. (2009) Unintended Consequences: Due to Lack of Standards for Speaker Identification and Other Forensic Procedures, Proceed. 16th Internat. Congr. Sound/Vib., Krakow, Poland, July 866: 1-6.

27) Hollien, H., Hicks, J.W. and Oliver, L.H. (1990) A Semi-Automatic System for Speaker Identification, Neue Tend. Angewandten Phon. III (V.A. Borowski and J.P. Koester, Eds.), Hamburg, Helmut Buske Verlag, 62: 89-106.

28) Hollien, H., Majewski, W. and Doherty, E.T. (1982) Perceptual Identification of Voices Under Normal, Stress and Disguise Speaker Conditions, J. Phonetics, 10: 139-148.

29) Jacewicz, E., Fox, R.A., and Wei, L. (2010) Between-speaker and Within-speaker Variation in Speech Tempo of American English, J. Acoust. Soc. Am., 128: 839-850

30) Jiang, M. (1995) Experiments on a Speaker Identification System (PhD dissertation, Univ. of Florida)

31) Jiang, M. (1996) Fundamental Frequency Vector for a Speaker Identification System, Forensic Ling., 3: 95-106

32) Johnson, C.C., Hollien, H. and Hicks, J.W. (1984) Speaker Identification Utilizing Selected Temporal Speech Features, J. Phonet., 12: 319-327.

33) Kersta, L. (1962) Voiceprint Identification, Nature, 196: 1253-1257

34) Koester, J.P. (1987) Performance of Experts and Naïve Listeners in Auditory Speaker Recognition, in German, Festschrift fiir H. Wangler (R. Weiss, Ed.) Hamburg: Buske, 171-180.

35) Kraus, N. and Nicol, T. (2010) The Musician's Auditory World, Acoustics Today, 3: 15-27.

36) Kraus, N., McGee, T., Carrell, T.D. and Sharma, A. (1995) Neurophysiologic Bases of Speech Discrimination, Ear and Hear., 16: 19-37.

37) Kraus, N., Skoe, E., Parberry-Clarke, A. and Ashley, R. (2009) Experience-induced Malleability in Neural Encoding of Pitch, Timbre and Timing: Implications for Language and Music, Annals New York Acad. Sci., Neurosci. and Music III, 1169: 543-557.

38) Krishnan, A, Xu, Y.S., Gandour, J. and Cariani, P. (2005) Encoding of Pitch in the Human Brainstem is Sensitive to Language Experience, Cognitive Brain Res., 25: 161-168.

39) Künzel, H. (2013) Automatic Speaker Recognition with Cross-language Speech Material, Journal of Speech, Lang. and Law, Vol. 20-1: 21-44.

40) LaRiviere, C.L. (1975) Contributions of Fundamental Frequency and Formant Frequencies to Speaker Identification, Phonetica, 31: 185-197.

41) Lea, W. (1981) Voice Analysis on Trial, Springfield Il, Thomas, Charles C.

42) Mack vs. State of Florida, 54, Fla. 55 44 50 706 (1907) citing 5, Howell's State Trials 1186

43) McGehee, F. (1937) The Reliability of the Identification of the Human Voice, J. Gen. Psychol., 17: 249-271.

44) Mischel, W. (2008) The Toothbrush Problem, The Observer, Assn. Psychol. Sci., 21: 1-3.

45) Morrison, G.S. (2002) Liklihood-ratio Forensic Voice Comparison Using Parametric Representations of the Formant Trajectories of Diphthongs, J. Acoust. Soc. Amer., 125: 2387-2397.

46) Morrison, G.S. (2006) Vowel Inherent Spectral Change in Forensic Voice Comparison, J. Acoust. Soc. Am., 125: 2695A

47) Orchard, T., and Yarmey, A. (1995) The Effects of Whispers, Voice-sample Duration and Voice Distinctives on Criminal Speaker Identification, Appt. Cogn. Psychol., 9: 249-260

48) Papcun, G., Kreiman, J. and Davis, A. (1989) Long-term Memory for Unfamiliar Voices, J. Acoust. Soc. Amer., 85: 913-925.

49) Pollack, I., Pickett, J.M. and Sumby, W.H. (1954) On the Identification of Speakers by Voice, J. Acoust. Soc. Amer., 26: 403-412.

50) Reynolds, D.A. (1995) Speaker Identification and Verification Using Gaussian Mixture Speaker Models, Speech Comm., 17: 91-108.

51) Sambur, M.R. (1976) Speaker Recognition Using Orthogonal Linear Prediction, IEEE Trans., ASSP, 24: 283-287.

52) Schmidt-Nielson, A and Crystal, T.H. (2000) Speaker Verification by Human Listeners: Experiments Comparing Human and Machine Performance Using the NIST Speaker Evaluation Data, Digit. Sign. Proc., 10: 249-266.

53) Schuartz, M.F. (1986) Identification of Speaker Sex from Isolated Voice Fricatives, J. Acoust. Soc. Am., 43: 1178-1179

54) Shirt, M. (1984) An Auditory Speaker Recognition Experiment, Proceed., Conf. Police Appli. Speech, Tape Record. Evidence, London, Instit. Acoust., 71-74.

55) Siegfried, T. (2010) Odds Are, It's Wrong, Science News, 177: 26-35.

56) Simons, J., Nelson, L. and Simonsohn, U. (2011) Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant, Psychol. Sci., 22: 1359-1366.

57) Stevens, K.N. (1971) Sources of Inter- and Intra-speaker Variability in the Acoustic Properties of Speech Sounds, Proceed. 7th Int. Cong. Phonetic Sci., Montreal, 206-232.

58) Strait, D., Skoe, E., Kraus, N. and Ashley, R. (2009) Musical Experience and Neural Efficiency: Effects of Training on Subcortical Processing of Vocal Expressions of Emotion, Europ. J. Neurosci., 29: 661-668.

59) Tsai, W.H. and Wang, H.M. (2006) Speech Utterance Clustering Based on the Maximization of Within-clustering Homogeneity of Speaker Voice Characteristics, J. Acoust. Soc. Amer., 120: 1631-1645.

60) Wong, P., Skoe , E., Russo, N., Dees, T. and Kraus, N. (2007) Musical Experience Shapes Human Brainstem Encoding of Linguistic Pitch Patterns, Nature Neurosci, 10: 420-422

61) Yarmey, A.D. (1995) Earwitness Speaker Identification, Psychol. Public Policy Law, 1: 792-816.