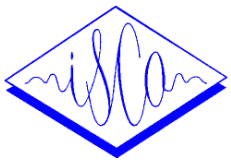

ESCA Workshop on Automatic Speaker Recognition, Identification, and Verification
Martigny, Switzerland
April 5-7, 1994

Current Approaches to Forensic Speaker Recognition

Hermann J. Künzel

Abstract --- **Forensic speaker recognition has become an important tool in crime contravention, since the use of the human voice as an instrument in the commission of crime is ever-increasing. To a certain degree, this is undoubtedly a consequence of the highly-developed and fully automated telephone networks of the industrialized countries, which may safeguard a perpetrator's anonymity almost perfectly. To date, different approaches to forensic speaker identification have been developed by phoneticians, linguists, speech scientists and engineers. In this review, the three main types of procedures currently used under various legal systems will be presented. It is argued that although computer-based tools for acoustical analysis and new statistical data about some speaker-specific features may aid the expert in drawing his conclusions, a fully automatic voice identification device is certainly not in sight, due to the vast number of imponderables and distortions of the speech signal which are typical to the forensic situation.**

Keywords --- **speaker recognition, forensic phonetics, law enforcement**

1. INTRODUCTION

In most criminal cases involving speaker recognition (SR) tasks, an expert is supposed to define the possibilities and limitations of this forensic discipline. Quite often he will face questions by solicitors or attorneys such as: "Isn't there a simple solution to this problem: after all, everyone seems capable of safely recognizing people by their voices, even on the phone! And why can't you use one of those computerized devices they have in detective films, which can verify voices in just a couple of seconds?" Quite obviously most individuals do not realize how difficult and complex the problem of forensic SR really is. So before elaborating on different methodological approaches, it is possibly useful for the purposes of this tutorial to clarify the fundamental differences between forensic and other SR applications.

Today non-forensic SR, which for the sake of simplicity will be called 'commercial', is not much of a problem either scientifically or technologically. Quite a number of solid-state devices were or still are on the market for control of access to high security installations such as certain military facilities, nuclear power stations, research laboratories and computer centers. More recently, telephone banking and particularly tele-monitoring of individuals on probation or on parole have become lucrative applications, since huge expenditures may be saved if conventional control and surveillance procedures become obsolete. One such system has been claimed capable of verifying up to 20,000 voices of clients via normal telephone channels. Systems developed by Texas Instruments and Bell Laboratories during the seventies are of special historical interest, since they showed the way to later developments, as far as acoustic parameters and statistical evaluation procedures are concerned.

Commercial systems typically operate according to the following principle [5,32,33]: after identifying him/herself with a personal number code, a person entitled to enter a controlled zone, a so-called customer, is required to pronounce a test phrase chosen from a more or less limited set of possible phrases or combinations of words. After some sort of acoustical analysis, mostly based on LTAS parameters, a feature vector is derived from the test signal and matched to vectors gained from earlier access claims of the person in question. A similarity index is then calculated. Recognition is affirmed if a certain threshold is exceeded; if not, the procedure may be repeated or the person will be regarded as a so-called impostor or intruder. Manufacturers claim that the performance of such systems comes close to a one per cent error rate for both false rejections and false identifications, or, in terms of the signal detection theory, false alarms and missed hits.

The preceding remarks will help to reinforce the principal differences between commercial and forensic SR:

1. In commercial applications an individual desires to be recognized: he is thus a *co-operative speaker*. In a forensic situation, however, one of the very reasons for oral communication is to conceal a person's identity. Therefore a so-called *non co-operative speaker* has no interest in natural, loud and clear speaking. As a matter of fact, in about fifteen per cent of the cases worked on at Germany's Federal Criminal Police Office (BKA) Speaker Identification Department there is evidence of deliberate voice disguise.

2. In a forensic case there is no pre-selected or pre-arranged text to be produced which could be used for powerful word-to-word comparisons. In many cases there is not a single aspect of the speech sample - including its overall duration - on which the expert can exert an influence. Either, suspected persons bluntly refuse to produce any speech sample at all, which of course they may do under most western legal systems, or it would be risky to openly address a person, for ex-

H.J. Künzel is Head of the Department of Speaker Identification, Tape Authentication and Linguistic Text Analysis of the Bundeskriminalamt (BKA) in Wiesbaden, Germany

ample in a live kidnapping case. In as many as 20 per cent of the BKA cases, questionable and/or known samples contain less than 20 seconds of what may be called 'net' speech, i.e. what remains of the original signal after removing speech portions of dialogue partners and badly disturbed portions. Another 'advantage' of commercial SR is the severe limitation on the number of possible test words/phrases, mostly down to a few dozen; and, facilitating the task even further, these words would contain carefully selected speech sounds such as open vowels or nasal consonants, which are known to contain highly speaker-specific information.

3. In more than 95 per cent of the cases worked on at the BKA, telephone-transmitted material has to be dealt with; this implies a multitude of potential degradations of the speech signal (see [26] for a detailed discussion). To name only a few of them: The signal is bandpassed at 300 - 3400 Hertz; thus, low-frequency components containing the first formant of some sounds and, more importantly, fundamental frequency, are no longer audible; also, some highly speaker-specific features such as a peculiar or defective pronunciation of high-frequency obstruents like /s/ or /sh/ become unobservable. The dynamic range of a recording is reduced to 30 dB at best; there are distortions of different degree and mathematical complexity, and last but not least, additive noise. In most commercial systems, however, the acoustical environment is carefully controlled; some use sound-treated rooms, high fidelity recording equipment and high quality lines to the computer system.

4. Perhaps the most salient theoretical difference between the two kinds of SR is the fact that for many though not all commercial applications the number of customers is relatively small, and, even more importantly, finite. Such a closed-set recognition task consisting of only one paired comparison is therefore called *verification*. In a typical forensic situation, however, it cannot be taken for granted or even assumed that the unknown speaker is among the two or perhaps even twenty suspects' voices handed in for comparison. In fact, the set of potential speakers is open, i.e. practically unlimited: it may for example consist of all adult male native speakers of a certain language. This type of voice comparison is called *identification*.

As was mentioned earlier, commercial systems use similarity thresholds on which the decisions are based. Threshold values are fixed on a cost/benefit rationale: a balance is struck between the damage which might be caused by a potential impostor and the delay and loss of working time caused by falsely rejecting a customer. Signal detection theory has made it clear that one type of error cannot be reduced without increasing the other [7]; see also [2]. However, the better the overall performance of the system, the lower the point of intersection of the probability density functions of both errors (the so-called equal error rate, EER).

In a forensic SR environment, the concept of one (general) similarity threshold value is out of the question for two reasons. Firstly, an error here will not result in a financial loss. Rather, it will lead to the acquittal of a guilty person or to the condemnation of an innocent one. Whereas the former error is deplorable, the latter is a catastrophe to any judicial system and must be avoided by all possible means; in terms of the signal detection theory: even at the cost of increasing the probability of a missed hit to near-certainty. Also, there is no way of having the result of a verification checked and possibly revoked by an independent controlling mechanism. If, for example, a customer is rejected by the access control system because his similarity index is affected by a sore throat or acute abuse of alcohol on "the night before", he can still be visually identified by security staff. Secondly, in the forensic domain we are still very far from constructing a single feature vector powerful enough to be used for standard identification tasks. This is without even considering a procedure or an algorithm for the statistical evaluation of the various speaker-specific parameters, some of which may be exploited in one case but not in another. Strictly speaking: we do not even know the totality of parameters that might be of speaker-specific value under a given set of circumstances.

To the background of this rather disenchanting perspective let us look at how forensic SR is being carried out today. Basically there are three different approaches:

1. auditory SR, consisting of two completely different variants:
   (a) performed by non-experts, i.e. mostly victims or witnesses of an offence.
   (b) performed by a phonetician or speech scientist on scientific principles;
2. visual inspection of broad-band spectrograms;
3. semi-automatic, computer-aided SR systems.

2. CURRENT PROCEDURES

2.1 *Speaker recognition by non-experts*

Auditory SR has long been used and accepted in forensics as part of the testimony of a victim or witness. Prior to the inventions of the telephone and sound recording equipment, it could be the key evidence on behalf of which a suspected individual could be identified or excluded from an offence committed in the dark, or when a victim had been blindfolded. Ever since the telephone has become popular as an instrument in committing crimes, SR has become relevant in nearly all the kinds of offences listed in the code of law [20].

The introduction of voice line-ups analogous to the well-known visual line-ups was the first formalisation of the recognition task, providing an objective, controllable, even statistically analysable response from a subject. In the United States, procedures such as those described by Huntley [14] have been widely accepted in court. Another well-developed, empirical variety of voice line-up is applied at the request of German courts and police authorities by scientists who are expected to explain in detail the underlying principles, the test format used, and in particular the error matrices gained from a subject's responses. In most cases a so-called direct identifica-

tion or naming test will be conducted, because it complies better with the conditions of a forensic application than discrimination or rating tests. For reasons of space this issue cannot be discussed in detail here (for details see [21,23,24]). Fundamentally, the experimental setup consists of identical text material gained from the unknown and at least four known dummy speakers ('foils') which are selected according to the results of a thorough auditory phonetic and linguistic analysis of the unknown voice. Thus it is possible to obtain speakers which are closely similar to the unknown speaker as far as age, dialect, voice quality, hesitation phenomena, speech rate etc. are concerned. The reason for this procedure is to render the subjects' task of picking out the unknown voice more difficult. Furthermore, the stimuli obtained from the dummy speakers are treated acoustically in order to reconstruct the telephone channel characteristics of the unknown recording. Finally the set of randomised stimuli is re-recorded on a test tape. Theoretically, the most fatal mistake in this type of auditory SR - which is beyond the control of the expert - would be to use the voice of a 'suspect' who himself is not the criminal although he or she may be 'identified' consistently by one or more of the subjects as the 'voice in question'. The reason would be that the 'voice in question', pre-selected as a result of prior investigations, exhibits a *purely accidental* similarity to one or more features of the anonymous person's voice, e.g. very high or low voice pitch, unusual dialect, systematic mispronunciations of speech sounds (e.g. lisping), etc. In such a case, which of course illustrates well the open set condition in forensic SR, it will be virtually impossible to discover an actual non-identity. To avoid this danger, the legal authorities must bestow great care upon the preselection of speakers to be nominated for a line-up procedure. Other factors that have to be accounted for by the expert conducting the test are:

- differences in the classification strategies of subjects,
- distinctiveness of the unknown voice,
- acoustical quality and duration of the taped material,
- hearing ability of subjects,
- elapsed time between first contact with the unknown voice and the recognition task,
- subjects' familiarity with the unknown voice,
- general circumstances of the criminal situation.

Another advanced type of formalised auditory identification test or rather test series with paired comparisons and 'blind' panels of listeners has been described by Hollien [10, p.204f]. It consists essentially of paired comparisons (ABX stimulus patterns). However, listeners do not have to respond to 'experimental' samples only, i.e. samples containing the known and unknown voices: they will also have to respond to special 'reference' material in order to enable the expert to assess their general ability to discriminate between voices auditorily. If performance does not meet a pre-fixed threshold for this task, a subject's responses to the 'experimental' material are discarded from the test.

Although much research is still needed in order to understand the principles underlying auditory SR - particularly under forensic circumstances - it is probably justified to state that these types of SR experiments, if properly performed, may yield reliable and objective results. In principle SR by non-expert individuals (witnesses, victims) may thus be regarded as a valuable tool in a subset of cases without live recordings of a crime.

2.2 *Speaker recognition by experts*

Whenever both questioned and known speech materials have been recorded in a case, voice comparison by expert is the method of choice. Unlike the layman, the expert will be able to make his observations and analyses explicit and state the underlying scientific categories and principles. Thus it is possible for himself as well as for other experts appointed to the case, and last but not least for the judge and the jury, to control and reconstruct every link in his chain of arguments. An expert is able to analyse all the features of a speech signal which may carry speaker-specific information, that is linguistic, phonetic and acoustic. Doddington [4, p.1653] uses the terms "high level" and "low level information", the former comprising "dialect, subject matter or context, and style of speech, " etc; the latter "spectral amplitude, voice pitch frequency, formant frequencies and bandwidths" and other acoustic features.

Whether or not an expert equipped with such tools will achieve a better performance than naive listeners in all types of experiments and under all possible circumstances of a real-world case is a another question. The experiment by Shirt [35] which is often referred to by opponents of the auditory approach, does not offer conclusive evidence for the claim that there is no essential difference between recognition rates of phoneticians (without further specifications of their native languages, duration of (auditory) phonetic training, specialization, etc.) and naive listeners. Although the original speech material had been provided by a police authority, the experimental setup contained several characteristics completely untypical of forensic situations, particularly the duration of the test items of only four seconds. Unless a speaker exhibits several rare features such as speech defects, hyponasality, certain types of hoarseness etc., and unless all these features happen to occur in such a brief sample, any responsible phonetician, particularly one with forensic experience, would altogether refuse to give a formal opinion on identity, because the minimal base for applying his analytical tools is lacking. Such a situation would remind one of a surgeon who is urged to remove an appendix in a couple of minutes, a time span he will normally need just to prepare the operational site! Speech samples are not like holograms which contain the entire picture in every little fragment. The BKA Speaker Identification Department demands a minimum of 20 to 30 seconds of 'net' speech material from each unknown and known speaker for computer-aided voice comparisons, but even this may be too small a base in a number of cases.

Other evidence does indeed suggest that the performance of trained observers is superiour to that of naive listeners. In one series of experiments with familiar and unfamiliar speakers Köster [18] obtained recognition rates of 100 per cent twice from a group of five phoneticians, as against 94 and 89 per cent from thirty naive listeners. Personal experience from several hundred court sessions shows that judges, attorneys or solicitors may at first have great difficulty in perceiving such obvious speaker-specific features as creaky voice, sigmatism and even instances of tonic or clonic stuttering. In fact, some individuals would have a demonstration tape enhancing such features replayed several times before making up their minds. Since all relevant features are also demonstrated visually through computer-based spectrograms or other forms of documentation, an expert is sometimes doubtful whether the original auditory phenomenon or perhaps the pertinent red or green spot on a multi-coloured computer printout were essential to the individuals' decisions.

Auditory expert opinions by linguists and phoneticians have long been accepted by courts in many countries. After several decades of progress in acoustics, phonetics, signal processing and pattern recognition techniques, scientists have begun to disagree as to whether or not this method is still up to date. A comprehensive discussion can be found in Baldwin & French's monograph [1] which contains many case reports from the UK. Quite naturally, this issue has first arisen in countries like the United States or Great Britain, where criminal jurisdiction is based on the adversarial principle, which implies that there are two opposing experts in court, each appointed to one party. Many continental European jurisdictions, however, require only one 'independent' expert appointed to a case by the court. In Germany, for instance, a second opinion will be called for only if the court or one party officially declare that they consider the first expert to be biased or incompetent.

A practitioner who is frequently being asked by courts to express his opinion on auditory, auditory-acoustic, and/or computer-aided investigations on speaker identification, performed by colleagues from universities, State Laboratories, or private companies, will probably come to view the problem as follows: Whenever a speaker's dialect, sociolect, jargon (i.e. a linguistic subsystem typical of a group of individuals with identical occupations, such as railwaymen, professional criminals, journalists, etc.), individual hesitation phenomena, speech errors and the like have to be analysed, the trained ear of a phonetician or dialectologist is certainly the most adequate tool, superiour to any technical procedure, if such is applicable at all. This is particularly important in the initial stages of a case when there are only (one or more) unknown voices which are to be described in as much detail as possible. In fact, a careful dialectological analysis often proves to be the decisive clue in the search for a criminal, because it leads to a close limitation of his/her origin. An excellent example involving the English language is the dialectological and sociolectal analysis by two English phoneticians of the verbal behaviour of the hoaxer in the famous Yorkshire Ripper case [6]. There is no doubt either that certain auditory findings may be called objective, at least in the sense that so-called narrow phonetic transcriptions made by different individuals represent acoustical events which have been interpreted according to the same well-established categories as provided by general phonetic theory; see for example [3,25,19]. In this author's practice there have been numerous examples of identical transcriptions performed by two or more phoneticians working independently on the same forensic case, even when the speech samples in question were severely degraded [20]. On the other hand it is clear that the phonetic and linguistic parameters just mentioned are only a small subset of speaker-specific parameters, although they might be critically important in some cases [18,29,30,40,41]. Therefore a view such as that expressed by Baldwin cannot be supported that an exclusively "auditory approach" to forensic speaker identification is "fully adequate for the task" [1, p.9]. Stating the question in polemical terms: nobody would acknowledge the diagnostic strategy of a medical doctor who uses only his stethoscope for the diagnosis of a potentially life-threatening heart disease, ignoring modern achievements like ECG, computer tomography, echo cardiography, and blood tests.

The procedures discussed in the next chapter may be considered as a way out of this dead-end approach since they include the valuable classical phonetic tools. But first, a brief comment should be made on another technique which was widely used in the United States, parts of Europe, Israel and other countries in the late sixties and seventies as an allegedly objective method for forensic voice comparison, namely the visual interpretation of ordinary broad-band spectrograms. Whereas in the past ten years this procedure has been losing ground in the States and was abandoned completely in Germany and the Netherlands, it is still being used in countries like Israel, Italy, Spain and Colombia; the FBI are using it for investigative purposes [16]. A vast number of studies have exposed the substantial shortcomings of spectrography as a forensic tool, which are quite obvious even in the latest so-called voice comparison standards issued by a "Voice Identification and Acoustic Analysis Subcommittee" of the "International Association for Identification" [42]. Here is just one characteristic example: To be an examiner, "a minimum of high school diploma is required, but a college degree is desirable" (p.373). In other words: a thorough scientific education is by no means regarded as a pre-requisite. Furthermore, a total of 12 (!) publications are the "required reading" for the would-be examiner, and another 13 publications are "suggested reading" (p. 374-376). The most comprehensive and extremely critical report on voice identification by spectrographic analysis was published by the "Committee on Evaluation of Sound Spectrograms" which had been installed by the U.S. Department of Justice ([2]; see also [9], [4], and the controversy between Koenig and collaborators [16,17] and Shipp et al. [34]. In short, it would seem that intra-speaker variation in voice spectrograms is not consistently smaller than inter-speaker variation. Thus, the main condition for probability statements on identity of speakers is not fulfilled (cf. also [4, p. 1654]).

It is difficult to state exactly which patterns or salient features of spectrograms have been used by its proponents, at least there is no exhaustive checklist to be used in casework by all the examiners. Obviously, the most widely used features are formant bandwidths, center frequencies, hubs, the spectral composition of fricatives and plosives for individual sounds (segments) and transitions, and something which is termed "peculiar spectrographic 'gestalt'" without being defined or even described [38, p.116]. An individual observer's internal threshold for a decision on identity/non-identity relates to the visual similarity of these parameters, tacitly assuming that variability within subjects is smaller than between them. Many critics have shown this to be wrong. Consider, for example, spectrograms of same and different speakers in [10, p. 216-219], or [5, p. 1654]. It is probably not unfair to say that the visual interpretation of spectrograms does nothing but shift the problems and pitfalls of the comparison task from the auditory to the visual domain, that is: from the ears to the eyes of a subjective human observer. The allegation by some authors [15,36,38,39] that their technique typically attains up to 99 per cent correct results has not been supported by any independent scientist, i.e. by one who did not himself belong to the inner circle of spectrograph analysts. Rather, error rates of more than 20 per cent are reported [2]. As a matter of fact, auditory analysis of speech samples has been shown to be more effective than spectrogram interpretation [37]. Another fact to be noted is that people with commercial interest in the propagation of the technique introduced the term "voiceprint" analysis [15], thus consciously alluding to the high identificational power of fingerprints which, contrary to spectrographic patterns, have indeed proved to be unique and unalterable. In summary: the visual interpretation of spectrograms cannot be considered an adequate technique for forensic SR. Therefore it can only be deplored that some countries not only tolerate but have even recently introduced this technique.

2.3 The modern phonetic approach

It is clear from the preceding discussion that an adequate procedure for forensic SR must reduce the human factor as far as possible and augment the number of and speaker-specific power of objective parameters. It is also clear, however, that even the most advanced SR systems require the interaction of an expert in many ways, starting off with the selection of appropriate speech material, the setting of threshold values for filtering and other pre-processing measures, the selection or suppression of certain acoustic or linguistic parameters, the interpretation of numerical results, and, last but not least, the formulation of the final statement on identity or non-identity. The word 'expert' used here is of special significance, as can be seen from this example: in the early seventies a sophisticated semi-automatic speaker identification system (SASIS) for forensic use was developed in the United States. It showed the way to several later projects as far as interactive segmentation of the speech signal, nature of acoustic parameters and statistical evaluation procedures are concerned [31]. Yet, ultimately the system failed, one of the main reasons being the condition that it would be operated by police officers after some training: imagine how well or how poorly such personnel may have performed in a highly sensitive task like the phonetic segmentation of fluent, telephone-transmitted, degraded speech from an oscillogram and/or spectrogram! If a phonetician had been consulted on this issue he would certainly have predicted such an outcome.

One of the first phonetic-acoustic approaches to forensic SR was developed at the German Federal Criminal Police Office. It was introduced in 1980 and has since been employed in about two thousand cases. The procedure which has been fully documented elsewhere [20] consists of the following main steps: After pre-processing the data, a phonetician isolates speaker-specific features from the speech signal in each of the three areas of the verbal behaviour, i.e. voice, speech and what is called 'manner of speaking', the latter containing among other features articulation rate, rhythm, hesitation phenomena, and intonation contours. These features, part of which are initially perceived by the trained ear of the expert, may be put on an objective physical basis through the use of modern instrumental techniques such as LTAS, traditional spectrograms with a choice of one out of seven filter bandwidths according to the nature of the acoustic features of interest (focus on temporal or frequency aspects), three-dimensional and waterfall spectra, oscillograms, cepstra and others. In a voice comparison report, the final opinion is determined by a synopsis of the results from each parameter, that is about eight to twelve on average, according to the nature of the speech material. In order to optimize the procedure, a computer-based 32-colors signal manipulation and analysis system was developed and is constantly being extended. Thanks to a very specialized hardware and firmware configuration almost all analyses can be performed in real time except for some of the more complicated filtering and jitter extraction algorithms. Currently a module is being tested which provides an algorithm for the quantitative assessment of certain types of hoarseness. The key feature of this module is that it works on fluent speech, which is of course the *conditio sine qua non* for its use in forensics: an unknown speaker would hardly be as cooperative as an ENT-patient who is required to sustain an open vowel for several seconds on a monotone for an automatic measurement of hoarseness.

A number of speaker-specific features such as fundamental frequency (Fo), Fo variation and acoustical characteristics of all sorts of pathological events, mainly mispronunciations and speech errors, can be evaluated using statistical background data gained from large numbers of subjects and/or real cases. Figure 1 shows the cumulative distribution of the average voice fundamental frequency (Fo) for a group of 266 adult males, a voice parameter which may be of considerable speaker-specific importance under certain speaking conditions. In court, such material may be of considerable value to the expert, for example when he states that both an anonymous caller and a defendant have about the same low Fo of say 90 Hz, because he can in fact tell the jury that there is only a three per cent chance of two males having such a low or lower

voice. Of course, this figure represents this parameter only and must not be considered as a measure for identity/ non identity in general. The aim of the BKA's team of scientists is to increase the number of parameters which can be expressed on such quantitative scales and thus broaden the objective basis for SR. In view of the tremendous amount of experimental work needed, cooperation from other researchers is welcome.

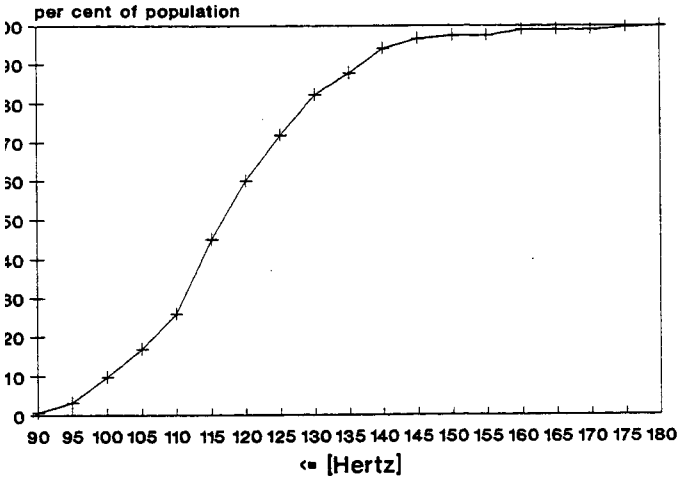

Fig. 1. Cumulative distribution of mean voice fundamental frequency in 266 male adults

Another SR system for forensic use which relies even more on acoustic parameters has been developed at the Los Angeles County Sheriff's Office [27,28]. After pre-processing, the speech signal is analysed for up to 14 text-independent parameters from both the time and spectrum domains. The so-called intensity deviation spectrum (IDS), a sort of normalised spectrum clear of transmission distortions is particularly important. After the feature vector is established, a sophisticated statistical analysis and weighting procedure is applied, resulting in what is called a-proximity index. On a preliminary background data base of 50 male speakers the system is claimed to reach a 98 per cent correct identification rate. Unfortunately, the project has been discontinued recently.

A semi-automatic SR system called SAUSI was developed by Hollien and associates at the University of Florida [10,12]. It is all the more interesting since, similar to the German concept, it basically uses 'natural speech' parameters such as Fo, number and duration of silent intervals, speech rate, vowel durations and the like ([11]; see also the early approach by Holmgren [13]). Such features may be considered to parallel in part the mechanism of auditory SR to a certain degree and are also much easier to explain to a court than abstract parameters such as the nth cepstral coefficient. SAUSI currently works on a background data base of about 400 speakers to which questioned and known samples may be matched. However, the applicabilty of the system to ordinary forensic cases remains to be established.

3. THE FUTURE OF FORENSIC SPEAKER RECOGNITION

In future SR systems for forensic use, the amount of objective parameters will constantly increase. There will also be refined algorithms for statistical evaluation of individual parameter results according to their speaker-specific significance under the circumstances of a given case. However, the limits for progress are staked out by the forensic real-world conditions. To quote George Doddington, an outstanding expert on automatic SR: "...it is not reasonable to expect that *any* level of performance is possible, limited only by improvements in feature extraction and algorithm development. Rather [...] the performance [...] is dependent on the amount of control that can be exerted on the operational conditions" [4, p. 1663; my emphasis]. As was shown earlier, however, there is almost no such control in a normal forensic environment. Thus it is inevitable that even in the long run an automatic procedure will not be available, and for this reason experts will still be needed to perform the tasks mentioned earlier. Thus the question of who is to be regarded as competent is all the more important. The SASIS example has shown that it is certainly not sufficient to have law enforcement personnel trained. There should be general agreement also as to the unfitness for such work of signal processing or sound engineers, or acousticians, who are able to assess only some acoustic but probably not phonetic and linguistic features. Incidentally, a British colleague was recently appointed to a case where his opponent regarded himself as qualified because he was the owner of a recording studio. Eventually, it became obvious that this individual was totally unaware of (if the matter were not so serious one should rather say 'uncompromised by') the difference between letters and sounds! Qualification as an expert in forensic SR requires a thorough education in modern speech science so that all speaker-specific features of the speech signal, in Doddington's terms both "high" and "low level information", may be covered. Consequently, phoneticians and speech scientists are probably best qualified to meet this requirement. But even they should always be aware of the abyss between carefully controlled experiments in the laboratory and the intricacies of the forensic world.

An elaborated version of this report will appear in the *Forensic Science Review.*

REFERENCES

[1] J. Baldwin, P. French: *Forensic Phonetics,* Pinter: London/New York; 1990.

[2] R.H. Bolt, F.S. Cooper, D.M. Green, et al.: *On the Theory and Practice of Voice Identification;* National Academy of Sciences:Washington DC; 1979.

[3] J.C. Catford: *A practical introduction to phonetics,* Clarendon Press: Oxford UK; 1988.

[4] G.R. Doddington: Speaker Recognition - Identifying People by Their Voices, *IEEE-ASSP-Transactions 73*:1651;

[5] G.R. Doddington, R.E. Helms, B.M. Hydrick: Speaker Verification III, *Texas Instruments Inc. Report for RDAC, Rome*: New York; 1976.
[6] S. Ellis, J.W. Lewis: The Yorkshire Ripper - a case history, *Forensic Linguistics 2* (forthcoming May 1994).
[7] D.M. Green, J.A. Swets: *Signal detection theory and psychophysics*, Wiley & Sons: New York/London; 1966.
[8] M.H. Hecker: *Speaker Recognition: An Interpretive Survey of the Literature*; Monograph No. 16 of the American Speech and Hearing Association, 1969.
[9] H. Hollien: Peculiar case of "voiceprints", *JASA 56*:210; 1974.
[10] H. Hollien: *The acoustics of crime*, Plenum Press: New York; 1990.
[11] H. Hollien, M.P. Gelfer, R. Huntley: The natural speech vector concept in speaker identification, in J.P. Köster (Ed): *Neue Tendenzen in der Angewandten Phonetik III*; Buske: Hamburg, p.71; 1990.
[12] H. Hollien, J.W. Hicks, L.H. Oliver: A semi-automatic system for speaker identification, in J.P. Köster (Ed): *Neue Tendenzen in der Angewandten Phonetik III*; Buske: Hamburg, p.89; 1990.
[13] G. Holmgren: Physical and psychological correlates of speaker recognition, *Journal of Speech and Hearing Research* 10:57; 1967.
[14] R. Huntley, K.J. Pass: Assessment of voice lineup procedures, in *Program and Abstracts of the 45th Annual Meeting of the American Academy of Forensic Sciences, Boston*; Colorado Springs; p. 102; 1993.
[15] L.G. Kersta: Voiceprint identification, *Nature* 196:1253; 1962.
[16] B.E. Koenig: Spectrographic voice identification: A forensic survey, *JASA* 79:2088;1986.
[17] B.E. Koenig, D.S. Ritenour, B.A. Kohus, A.S. Kelly: Reply to 'Some fundamental considerations regarding voice identifications', *JASA* 82:688; 1987.
[18] J.P. Köster: Leistung von Experten und Naiven in der auditiven Sprechererkennung, in R. Weiss (Ed): *Festschrift für H. Wängler*; Buske: Hamburg, p.171; 1987.
[19] K.J. Kohler: *Einführung in die Phonetik des Deutschen*, Schmidt: Berlin; 1977.
[20] H.J. Künzel: *Sprechererkennung: Grundzüge forensischer Sprachverarbeitung*, Kriminalistik-Verlag: Heidelberg; 1987.
[21] H.J. Künzel: Zum Problem der Sprecheridentifizierung durch Opfer und Zeugen, *Goltdammer's Archiv für Strafrecht* 5:215; 1988.
[22] H.J. Künzel: How well does average fundamental frequency correlate with speaker height and weight? *Phonetica* 46:117; 1989.
[23] H.J. Künzel: *Phonetische Untersuchungen zur Sprecher-Erkennung durch linguistisch naive Personen*, Steiner: Stuttgart; 1990.
[24] H.J. Künzel: On the problem of speaker identification by victims and witnesses, *Forensic Linguistics 1*, (in press).
[25] P. Lieberman, S.E. Blumstein: *Speech physiology, speech perception, and acoustic phonetics*, Cambridge University Press: Cambridge; 1988.
[26] L.S. Moye: *Study of the effects on speech analysis of the types of degradation occurring in telephony*, Standard Telecommunication Laboratories: Harlow UK; 1979.
[27] H. Nakasone, C. Melvin: Computer assisted voice identification system. *IEEE-ASSP-Transactions* 36:587; 1988.
[28] H. Nakasone, C. Melvin: *Project C.A.V.I.S. Final Report*. Los Angeles County Sheriff's Department:Whittier; 1989.
[29] F. Nolan: *The Phonetic Bases of Speaker Recognition*, Cambridge University Press: Cambridge UK; 1983.
[30] F. Nolan: The limitations of auditory-phonetic speaker identification, in X. Kniffka (Ed): *Texte zu Theorie und Praxis forensischer Linguistik*, Niemeyer: Tübingen, p. 457; 1990.
[31] J.E. Paul, A.S. Rabinowitz, J.P. Riganati, J.M. Richardson: *Semi-automatic speaker identification system (SASIS). Analytical Studies Final Report*, Rockwell Int'l: Anaheim; 1974.
[32] A.E. Rosenberg: Evaluation of an automatic speaker verification system over telephone lines, *Bell Technical Journal* 55:723; 1976.
[33] A.E. Rosenberg, M.R. Sambur: New techniques for automatic speaker verification, *IEEE-ASSP- Transactions* 23:169; 1975.
[34] T. Shipp, E.T. Doherty, H. Hollien: Some fundamental considerations regarding voice identification; *JASA* 82:687; 1987.
[35] M. Shirt: An auditory speaker-recognition experiment; *Proc. Inst. Acoustics* 6:101; 1984.
[36] L.L. Smrkovski: *Voice Identificaton*, Michigan Dept. of State Police: East Lansing;1977.
[37] K.N. Stevens, C.E. Williams, J.R. Carbonell, B. Woods: Speaker authentication and identification : A comparison of spectrographic and auditory presentations of speech material, *JASA 44*:1596; 1968.
[38] O.I. Tosi, H. Oyers, W. Lashbrook, C. Pedrey, J. Nicol, E. Nash: Experiment on voice identification, *JASA* 51:2030; 1972.
[39] O.I. Tosi: *Voice Identification. Theory and Legal Applications*, University Park Press: Baltimore; 1979.
[40] D. Van Lancker, J. Kreiman, K. Emmorey: Familiar voice recognition. Patterns and parameters. Part I: Recognition of backward voices, *J Phonetics* 13:19; 1985.
[41] D. Van Lancker, J. Kreiman, K. Emmorey: Familiar voice recognition: Patterns and parameters. Part II: Recognition of rate-altered voices, *J Phonetics* 13:39; 1985.
[42] VIAAS (Voice Identification and Acoustic Analysis Subcommittee of the International Association for Identification): Voice Comparison Standards; *J Forensic Identification* 5:373; 1992.