PROCEDURES AND PERSPECTIVES IN FORENSIC PHONETICS

Angelika Braun
Speaker Identification and Tape Analysis Section, Bundeskriminalamt, Wiesbaden, Germany

ABSTRACT

In this contribution, it is argued that the forensic applications of their field should no longer be ignored or denied by phoneticians. The development of forensic phonetics in the last decade including the increased importance of computerized procedures is outlined. Owing to the degradations introduced to the signal by the conditions under which forensic recordings are typically made, however, there is serious doubt that a fully automatic voice identification device will emerge in the near future. Topics for further research are indicated.

1. INTRODUCTION

The forensic application of phonetic sciences is one of the most controversial issues within the phonetics community. The extreme standpoints are probably represented by the successors of the so-called voice print technique in the United States, i.e. the Voice Identification and Acoustic Analysis Subcommittee (VIAAS) of the International Association for Identification (IAI) on the one hand and groups like the British Association of Academic Phoneticians (BAAP) or the Bureau du Groupe *Communication Parlée* de la Société Française d'Acoustique on the other hand. Whereas the former group not only advocates forensic speaker identification unconditionally but also basically holds the view that anyone with a high school diploma can do it after having undergone a two-week training course[1], the latter have taken rather strong positions against forensic phonetics in general and forensic speaker identification in particular by adopting motions to the effect that phoneticians should not engage in such tasks[2][1]. A third view on the subject is represented by the International Association for Forensic Phonetics (IAFP), which was formed in York in 1989. This organization aims to provide a forum for discussion among those who either work in the field of forensic phonetics and/or have an academic interest in it as well as to define and ensure professional standards in this area.

Discussion about the forensic application of phonetics has focussed on two principal issues: (1) Is it ethical for *anyone* to undertake forensic case work at all as long as scientific/empirical proof for the notion of "one speaker - one voice" has not been established?; (2) Are *phoneticians* more qualified than others to do forensic speaker identification?.

This contribution will address these issues which are controversial among phoneticians as well as - in line with the theme of this session - the question of what can be realistically expected to come from the laboratory within the next few decades.

Since a good part of the reservations that many phoneticians have about the forensic application of their field seem to stem from misconceptions about the exact nature of that work and the conditions under which it is done, a brief account of what forensic phoneticians actually do as well as the methods employed will be given. Although virtually any question related to speech or sound may be put to the phonetician in a particular case, this contribution will largely focus on speaker identification.

[1] There are indications to the effect that several members of BAAP now take a different view of forensic applications of phonetics than they did 15 years ago when the motion was passed.

2. WHAT FORENSIC PHONETICIANS DO

Some of the misconceptions of phoneticians about the forensic applications of their field may be due to the voiceprint legacy or other rather rash accounts of cases which are not representative of the state of the art in forensic phonetic work [3]. Specifically, many phoneticians do not seem to be aware of the fact that speaker identification, i.e. the comparison of a speech sample produced by an unknown speaker involved in the commission of a crime to that of one or more suspects, forms an important task for the forensic phonetician but by no means the only one. Other activities include speaker profiling or characterization, the analysis of disputed utterances, the analysis of background noise, the design of voice line-ups as well as interpretation of their results, intelligibility enhancement of noisy tape recordings, and tape authentication. The most relevant of these is speaker profiling, a task which is regularly requested in the early stages of e.g. kidnappings when a recording of the criminal's voice is available. Most of the time the voice forms the only lead at this stage of the investigation, and its analysis with respect to sex, age group, regional accent or dialect, peculiarities or defects in the pronunciation of certain speech sounds, sociolect, mannerisms etc. is of paramount importance for the investigation and thus, eventually, for the victim's life.

Not every forensic phonetician should or would engage in all of the above activities; what people are ready to take on largely depends on their specialization during their education and - as in other fields of expertise - on the amount of insight in the limits of their knowledge. The International Association for Forensic Phonetics has established a Code of Practice in order to ensure that its members will not exceed the limits of their expertise.[4]

In this context it is important to mention that one of the foremost duties of phoneticians is to explain to various groups of people what forensic phonetics **cannot** do, e.g. point to the limitations induced by telephone transmission or by speech samples in a language of which the phonetician does not have perfect command or the impossibility to judge a speaker's sincerity based on phonetic evidence alone.

3. THE FORENSIC ENVIRONMENT

At first glance, any discussion about forensic voice comparison methodology may seem quite dated in view of the fact that very powerful speaker recognition algorithms are available for commercial purposes, i.e. access control. But all of these systems require *cooperative* speakers in the sense that the speaker makes an effort to articulate clearly, that she or he agrees to pronounce a preselected phrase particularly suitable for comparison purposes, and that she or he is prepared to repeat an utterance if necessary. Needless to say, none of these prerequisites are met by forensic recordings.

Furthermore, in commercial speaker *verification* the number of speakers with whom the actual sample has to be compared is by definition finite, whereas forensic speaker *identification* is almost always an open-set task. Thus, even if as many as 20 recordings of suspects are submitted, there is no reason to assume that the offender is among them.

Aside from these principal issues there are some technical factors which preclude the use of commercial speaker recognition techniques in the forensic domain. The most frequent and also the most salient one is telephone transmission, which implies a bandwidth limitation to 300-3400 Hz and a restriction of the dynamic range to 30dB, if the line is good. This loss of frequency and amplitude information can obviously not be compensated for and leaves the phonetician with a limited basis for judgement. Specifically, formants outside the frequency range of the telephone line cannot

be measured, and misarticulations of fricatives like lisps may no longer be detectable.

Finally, the quantity and the quality of the material available for analysis is to a large extent controlled by the offender. Thus, even if as much as one minute of net speech (i.e. not counting pauses, hesitations etc.) is available, it may not be assumed that the material will fully represent the range of that person's verbal behavior.

One way of reacting to adverse conditions like the ones outlined above is, of course, to refrain from doing any forensic phonetic work at all. This seems to be the course of action suggested by BAAP as well as the GCP who have taken a rather strong view against phoneticians' engaging in forensic work by adopting motions to the effect that "phoneticians should not consider themselves expert in speaker identification until they have demonstrated themselves to be so" and "the GCP Bureau affirms that, in its opinion, speaker identification experts have yet to furnish any verifyable proof of their abilities"[2], respectively. On the other hand, there can be no doubt that phoneticians do possess specific knowledge about the human voice and its analysis, and it seems difficult to argue that the knowledge there is should be withheld from the legal community just because it is limited. To put it drastically: If a child has been kidnapped and a recording of the kidnapper's voice as well as that of a suspect were available, it would be absurd to outright refuse to do a phonetic voice comparison for lack of theoretical foundation.

4. METHODOLOGICAL APPROACHES

4.1. Auditory vs. spectrographic

As far as methods employed in forensic speaker identification are concerned, the history of forensic phonetics is a history of extremes. On the one hand, there used to be a very strong aural-perceptual phonetic tradition in Britain [3,5]. The conclusions reached by this method are largely based on a minute dialectological description of the samples in question, along with judgements of segment articulations as well as pitch and intonation. Although auditory phonetic procedures still form an important part of forensic speaker identification and obviously the most powerful tool in speaker profiling, voice comparison reports which are based on listening techniques alone are no longer considered state of the art [6,7].

The other extreme is represented by those who sought to reduce the human factor by applying various partly or fully automatic procedures. The worst facet of this is what has become known as the "voiceprint" technique, first introduced in the United States by Lawrence Kersta in the 1960s [8]. The obviously untenable analogy to the evidential value of fingerprints as well as the lack of theoretical foundation and poor training of the so-called "experts" [1] have done severe damage to forensic speaker identification as a whole. The visual inspection and comparison of spectrograms is obviously neither objective nor superior to aural-perceptual methods - the subjective judgement is merely shifted to the visual domain, and considering the sensitivity of the human ear as compared to the crude resolution of a spectrogram easily reveals the severe shortcomings of the technique as a whole. No claims are made concerning the theoretical validation of the procedures used beyond the - unvalidated - assumption that formant structures and other spectral characteristics which are evident from a spectrogram are different for each individual. Even though this assumption has been shown to be incorrect [9, 10], and voice identifications based on spectrograms were found to be much less reliable than those based on aural-perceptual judgements [11], it has taken decades to convince judges in most, though not all, States of the US to no longer admit "voiceprint" evidence, and it still seems to be practised in some countries including Israel and Italy. A slightly modified form of the voiceprint technique is still adhered to by the VIAAS of the IAI, but the "Voice Comparison Standards" as published by that organization [1] cannot be considered as a basis for serious discussion, as is indicated by the list of required reading for all of its examiners, which consists of 11 titles followed by the suggestion to read the manuals for any equipment used in the examination.

The use of formants as a sole basis for forensic voice comparisons has fairly recently been advocated in a different context by some scientists whose background seems to be in engineering rather than phonetics [12, 13, 14]. They propose to compare formant values and sometimes also pitch [sic]. They argue that all it takes to arrive at exact percentage values for the probability of identity or non-identity of two voice samples is the right statistical procedures. This, however, would only be true if it could safely be assumed that the within-speaker variation with respect to formants and formant-related acoustic parameters is under all circumstances smaller than the between-speaker variation. That this is precisely not the case has been demonstrated in the course of the voiceprint controversy (see above). Thus, approaches like those described so far do not only lack theoretical foundation but run counter to established phonetic knowledge.

4.2. The current approach

Since the early 1980s, an approach to speaker identification which combines traditional aural-perceptual and acoustic phonetic techniques has become increasingly widespread. It emerged from a research project at the German Bundeskriminalamt and has been used in thousands of cases at that institution alone [15]. The first stage in the examination consists in a detailed auditory analysis of the voice samples involved. Much like the profiling of anonymous voices, this part of the analysis pertains to parameters like voice quality, dialect or regional accent, speech defects, misarticulations of sounds, speech rate, intonation, rhythm, but it also includes observations on syntactic, idiomatic, and even paralinguistic features like breathing patterns. The main results of this analysis are documented in a transcript using IPA symbols in order to facilitate a comparison of the results with those of other experts. This aural-perceptual analysis is complemented by an acoustic phonetic examination of the recordings. Thus, several of the parameters used in the report can be quantified or described more precisely than by auditory analysis alone. A good example is formed by the set of parameters concerning voice. A "high-pitched" voice in auditory phonetic terms can thus be described as exhibiting an average fundamental frequency of, say, 158 Hz. What the auditory phonetician might call a well-modulated voice can be characterized as having a standard deviation from the average F0 of, say, 28 Hz. An intonation contour which strikes the auditory phonetician as "unusually stylized", can be described as involving steps of, say, 87 Hz. In the area of articulation, formants as well as e.g. the frequency of a "sharpened" /s/ or a strikingly long aspiration can be measured. Thus, of the parameters studied, as many as possible are documented using the whole set of techniques which are currently available in acoustic phonetics. Some of the algorithms were tailored to the specific needs of forensic material. All analyses are carried out bearing in mind the communicative context and the emotional state of a speaker. Of particular interest are features like those mentioned above, which deviate from the usual. The difficult part for the forensic phonetician is, of course, to define what is "usual" or "norm" and what is "deviation". This is partly done on the basis of statistics showing the distribution of features like average F0 in the relevant population or, if such are not available, on the basis of experience.

Much forensic phonetic research is directed at establishing distributional data for as many parameters as possible. For the time being, however, the subjective element in the formulation of the conclusion cannot be completely eliminated. For the same reason, conclusions are phrased in terms of probability scales instead of percentages. The phrasing of the probability in a particular case will depend on the amount, quality and phonetic-linguistic yield of the material and the rarity of the features which are contained in the voice(s) involved.

5. WHY PHONETICIANS

There are laboratory studies [16, 17] which suggest that trained phoneticians are not significantly (though marginally) better at certain perceptual tasks related to speaker recognition than phonetically naive subjects. Those studies specifically deal with (closed-set) speaker identification and pairing [16], and age estimation [17]. The relevance of these findings with respect to forensic speaker identification, however, is not quite clear, because the experimental design of neither study represents forensic conditions. Furthermore, in an experiment reported by Köster [18], recognition and identification rates were higher for the expert than for the non-phoneticians. On the other hand, there may be perceptual tasks at which phoneticians are not necessarily much better than phonetically naive listeners. One should look very closely whether any experiments carried out in this area test genuinely phonetic skills - let alone *forensic* phonetic skills [6] - or involve intuitive tasks whose underlying mechanisms have not even been fully explained as yet. Certainly, the results of tests like those cited above should not tempt us into thinking that phoneticians are no more competent to describe and analyze voices than non-phoneticians and that therefore forensic speaker identification can be done by anyone. This would almost amount to a denial of phonetics as a scientific field.

Forensic phoneticians have been criticized for not having come up with their own experiments which would demonstrate that they have speaker identification skills which are superior to those of ordinary people [19]. On the other hand, the question is whether there is a fair (to the non-phoneticians) way of comparatively testing genuinely phonetic skills like doing a narrow transcription, describing the laryngeal setting of a certain speaker or explaining why a lisp cannot be detected in a telephone call. Particularly in the courtroom situation, it is of paramount importance that any opinion about voice identity be made explicit in terms of descriptive phonetic parameters. In order to do this, phonetic training is mandatory.

As Bolt et al. point out [10, p.99] there are "two kinds of experience [which] provide knowledge about the problems inherent in voice identification as well as some indication of possible success. The first is the experience of those who have attempted the task in real-life situations. The second is that of laboratory experimenters [...]".The position outlined above is strongly supported by the first type of experience cited. Although no exact account was kept, within the BKA laboratory alone there are literally hundreds of cases in which non-phoneticians have made very strong claims about speaker identity, while the phonetician indicated that the samples originated from different speakers. A typical example occurred in the course of the investigation of a kidnapping. A Turkish boy had been abducted and was still held by the kidnapper(s). Two police officers who had been listening to telephone taps of a particular person implicated in another crime for several months were absolutely convinced that he was also the kidnapper who had phoned to demand ransom. The voices were indeed very similar, but there was also phonetic evidence suggesting that the samples came from two different people. Later in the investigation, another suspect was recorded because he had been identified by witnesses as having made the anonymous phone call, but again there was strong phonetic evidence against identity (i.e. the suspect had a stutter whereas the offender did not).Thus, even without formal testing, there is a lot of evidence from everyday work for the superior performance of phoneticians.

This example can also be used to demonstrate the implications of forensic phonetic work: If the phonetician fails to recognize speaker identity, the kidnapper goes free, and the victim may be killed. If the scientist falsely identifies the wrong person, that person might be physically harmed by members of the special squad trying to make an arrest and free a kidnapped child. In the present author's view, this kind of responsibility should make anyone involved in forensic phonetic work very cautious, but it can hardly be used as an argument against providing expertise to the legal community.

Another reason why it seems difficult for phoneticians to refrain from forensic case work altogether is a political one. With so-called speech analysis packages available for any home computer for less than £100, even people with no specific training in phonetics may set out to do forensic work. French [3, pp. 58-59] mentions two cases from England in which sound engineers failed to distinguish between letters and sounds in their reports. In another country, two former members of the police force set out to do speech enhancement using commercially available signal manipulation software, having to admit that they were not sure what was actually happening when they operated certain controls.

There is an imminent danger that this will happen much more often in the future, particularly in countries like England and the United States whose judicial system is adversarial, i.e. where usually both sides hire their own experts. Under these circumstances, it would seem almost like a moral obligation to speak up against charlatans working for the other side. It should be added at this point that in Germany as well as the Netherlands the conditions under which any forensic expert works are quite different: The judicial systems in these countries can be described as inquisitorial rather than adversarial, this term implying that any expert is appointed by the court rather than by one side. The rôle of an expert within these systems is to supply the court with expertise pertaining to specific areas in which the judges themselves[2] do not feel sufficiently competent. The expert is to be impartial, and she or he *has* to present a full report of her or his findings irrespective of the implications for the trial. Thus, it is extremely uncommon to have more than one expert in a trial, and some of the problems specifically related to the fact that phoneticians may act as "hired guns" simply do not occur. The author would like to add at this point that she is extremely grateful to be working in this kind of framework since she would find it difficult, if not impossible, to be restricted in what she says by either prosecution or defense strategy.

6. ANSWERS FROM THE LAB

Nolan has pointed to the shortcomings with respect to the theoretical foundation of forensic speaker identification 12 years ago [20]. Defining the speaker under laboratory (HiFi) conditions seems to be a vastly different (and in many respects: easier) task than defining what is left of a speaker in terms of information contained in the signal under forensic conditions. In view of the limitations outlined above, there is a possibility that we may never be able to come up with an exhaustive list of speaker-characterizing features at all. Even if it could be demonstrated experimentally that each speaker has a voice which is distinct from

[2]In Germany, there are no jury trials. Instead, for major crimes there is a panel of five judges, two of whom are lay persons. They decide on the question of guilt as well as the sentence by majority vote.

those of all other speakers of that speech community, this does by no means imply that the distinction can always be discovered in the forensic material that happens to be available. On the other hand, there are many areas of empirical research which can help to widen the basis for judgement under forensic conditions.

Thus, a two-way approach is suggested here. Obviously, any laboratory experiment addressing the one-speaker-one-voice issue will be of great interest to anyone involved in forensic phonetics, even though the findings may have no immediate bearing on forensic work, e.g. if articulatory parameters are measured. Of particular interest from the forensic point of view would be attempts to describe the full range of a person's verbal behavior, i.e. changes introduced to the "neutral" way of speaking by psychological (stress, emotion) or physiological (fatigue, smoking, alcohol, medication) factors. Some of these factors have been studied in detail, often with the forensic application in mind [examples are 21-24], but the need for this kind of "top-down" research providing basic data will probably not be met for decades.

On the other hand, there is the necessity to start at the other end, i.e. to ask how, in view of the forensic environment, the procedures currently used in speaker identification can be improved. This "bottom-up" research starts out at the parameters which can still be assessed in degraded recordings and seeks to either quantify parameters which could not be quantified before or to gain information concerning the statistical distribution of certain features in order to be in a better position to assess the frequency of their occurrence. One example for this kind of research is a project currently under way at the Bundeskriminalamt, involving the quantification of a certain type of hoarseness from running speech [25]. A matter of great interest not only to forensic phoneticians would be an exhaustive phonetic description of hesitation markers including questions of intrapersonal vs. interpersonal variability. Another area for research would be the distribution of phonetic and linguistic characteristics in the population, which would enable the expert to weigh that parameter more precisely. An example of this kind of research is currently being carried out as a joint project between the Bundeskriminalamt and the Universities of Marburg and Trier [26]. It consists in establishing a data base of regional varieties of German and will enable the forensic phonetician to listen to samples of up to five min. duration from 450 locations. Narrow phonetic transcriptions of the samples are available. It is also possible to search for specific segments, morphs and words (in different phonetic contexts). A thesaurus component has been built into it which will display the phonological system of the accent or dialect in question. It would certainly be desirable to have similar data bases for e.g. speech defects.

7. CONCLUSIONS

Speaking is such a complex type of behavior that I tend to be sceptical that we may expect an answer to the question asked in the theme of this session any time soon. I am not even sure that the answer is going to be positive, particularly with the complicating factors induced by the forensic setting in mind. No matter how good a definition of a speaker will come from the lab, the forensic application of these findings will always be limited by the amount of information about that speaker which is contained in a recording. This applies both to the technical side, i.e. the amount of frequency and amplitude information available, and the representativeness of the material in terms of the speaker's "normal" voice. There is no doubt, however, that any step that is taken towards the definition of each speaker will make forensic speaker identification an easier task.

8. REFERENCES

[1] Voice Identification and Acoustic Analysis Subcommittee (1991), "Voice comparison standards", *J. Forensic Identification*, vol.41, pp. 373-392.

[2] Bureau du Groupe *Communication Parlée* de la Société Française d'Acoustique (1990), "Motion adopted Sept.7", *NESCA - The ESCA Newsletter*, no. 4, p 39.

[3] Baldwin, J. & French, P. (1990), *Forensic phonetics*. London: Pinter.

[4] "Announcement", in *JIPA*, vol. 22, pp. 80-81.

[5] Ellis, S. (1994), "The Yorkshire Ripper enquiry: part I", *Forensic Linguistics*, vol. 1, 197-206..

[6] Künzel, H.J. (1994), "Current approaches to forensic speaker recognition". *Proc. ESCA Workshop on automatic speaker recognition, identification, verification*, Martigny, pp. 135-141.

[7] Nolan, F. (1990), "The limitations of auditory-phonetic speaker identification". In: H. Kniffka (ed.), *Texte zu Theorie und Praxis forensischer Linguistik*, Tübingen: Niemeyer, pp.457- 479.

[8] Kersta, L. (1962), "Voiceprint identification", *Nature*, vol. 196, pp. 1253-1257.

[9] Hollien, H. (1990), *The acoustics of crime*. New York: Plenum Press.

[10] Bolt, R.H. et al.(1979), *On the theory and practice of voice identification*. Washington, D.C.: Natl. Academy of Sciences.

[11] Stevens, K.N., C.E. Williams, J.R. Carbonell, and B.Woods (1968), "Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material", *JASA* vol.44, pp.1596-1607.

[12] Maturi, P. (1990), "Speaker identification in forensics: a simulation experiment". *Proc. ESCA Workshop on speaker characterization in speech technology*, Edinburgh: CSTR, pp.155-160.

[13] Federico, A. & Paoloni, A. (1993), "Bayesian decision in the speaker recognition by acoustic parametrization of voice samples over telephone lines". *Proc. 3rd European Conference on Speech, Communication, and Technology, Berlin*, vol. 3, pp.2307-2310.

[14] Lipeika, A. & Lipeikiene, J. (1993), "The use of pseudostationary segments for speaker identification". *Proc. 3rd European Conference on Speech, Communication, and Technology, Berlin*, vol. 3, pp.2303-2306.

[15] Künzel, H.J. (1987), *Sprechererkennung. Grundlagen forensischer Sprachverarbeitung*. Heidelberg: Kriminalistik-Verlag.

[16] Shirt, M. (1984), "An auditory speaker-recognition experiment", *Proc. Inst. Acoustics*, vol. 6, pp.101-104.

[17] Braun, A. & Rietveld, A. (1995), "The influence of smoking habits on perceived age". *Proc. XIIIth Intl. Congr. Phon. Sc.*, Stockholm, 4 p.

[18] Köster, J.-P. (1987), "Auditive Sprechererkennung bei Experten und Naiven". In: R. Weiss (ed.), *Festschrift für H. Wängler*, Hamburg: Buske, pp.171-180.

[19] Nolan, F. (1991), "Forensic phonetics", *J. Linguistics*, vol. 27, pp.483-493.

[20] Nolan, F. (1983), *The phonetic bases of speaker recognition*, Cambridge: CUP.

[21] Künzel, H.J., Braun, A., Eysholdt, U. (1992), *Einfluß von Alkohol auf Stimme und Sprache*, Heidelberg: Kriminalistik-Verlag.

[22] Hollien, H. (1993), "An oilspill, alcohol and the captain: a possible misapplication of forensic science". *Forensic Science International*, vol. 60, pp. 97-105.

[23] Johnson, K., Pisoni, D.B., Bernacki, R.H. (1990), "Do voice recordings reveal whether a person is intoxicated? A case study, *Phonetica*, vol. 41, pp.215-237.

[24] Braun, A. (1994), "The effect of cigarette smoking on vocal parameters".*Proc. ESCA Workshop on automatic speaker recognition, identification, verification*, Martigny, pp. 161-164.

[25] Wagner, I. (1995), "Jitter-measurements from telephone-transmitted speech". *Proc. XIIIth Intl. Congr. Phon. Sc.*, Stockholm, 4 p.

[26] Künzel, H.J. & Köster, J.-P. (1995), "Forensic Data Base System of Regional Accents of German", *Proc. AAFS Annual Meeting*, Seattle, Wa., p.88.