Is forensic speaker identification unethical – or can it be unethical *not* to do it?

Angelika Braun and Hermann J. Künzel

Bundeskriminalamt (BKA), Wiesbaden, Germany

ABSTRACT This paper makes a case for forensic speaker-identification, but only if the practitioner is properly trained and carries out the task conscientiously. It could be argued (and has been argued) that it is unethical to engage in forensic speaker-identification until there is a well-established and fully automatic (i.e. machine-based) approach available: in other words phoneticians should not practise in this field at all until the subjective element of their task has been removed. The present contribution will focus on forensic speaker-profiling and identification. First, the specifics of the forensic task as opposed to the commercial speakeridentification (SI) task will be summarized, followed by a brief outline of the methods currently employed by forensic phoneticians. The applicability of automatic SI procedures will then be examined. It transpires that only in a small proportion of forensic cases does the material which is available from either the plaintiff or the investigating agency lend itself to the application of automatic methods. Therefore, it might seem unethical to apply these methods uncritically. However, in the vast majority of cases, other non-automatic methods have to be pursued. It is contended that the forensic phonetician has a moral obligation to aid the course of justice within the limitations which are imposed by the quantity and quality of the speech samples in question.

KEYWORDS forensic speaker-identification, ethics

INTRODUCTION

Ethical issues have always been closely linked to forensic speaker-identification. The reason is that this forensic application of phonetics, unlike other applications such as foreign-language teaching, dialect study or speech technology development, may have an enormous and immediate impact on the lives of individuals, whether they be alleged criminals or victims. In some countries convictions may be based on voice evidence alone and defendants may be sentenced to years in prison. On the other hand it is no exaggeration to say that the lives of people who have been kidnapped, or who could buy deliberately poisoned food when stores are being blackmailed, often depend on the work of phoneticians. Rape victims may draw a feeling of relief from a voice line-up carried out by a phonetician, which allows them to demonstrate their memory of the assailant's voice under scientifically controlled conditions.

Given this situation the question arises of whether the current sum of our knowledge in phonetics, or what some might prefer to call speech science, justifies the involvement of many phoneticians throughout the world in forensic work, trying to aid the course of justice. In other words: is there really enough knowledge today about the individuality of vocal behaviour to fulfil the often urgent demands made by police authorities and the judiciary?

HISTORICAL BACKGROUND

One of the first spectacular cases involving voice evidence was the trial of Richard Hauptmann in 1935 for the kidnapping and murder of Charles Lindbergh's baby boy. German-born Hauptmann was convicted partly on the grounds that Colonel Lindbergh had identified him as the person to whom he had briefly talked two and a half years before the abduction. There was no phonetician involved in the case, but several years later the psychologist Frances McGehee analysed a series of crucial aspects of auditory speaker recognition in two pioneering articles (1937; 1944). In particular, she examined the decay over time of long-term memory for voices. According to her results, it seems hardly justifiable to accept a positive identification of a speaker's voice on the basis of one brief encounter. It should be noted that in this case psychophonetic research produced an argument against using the testimony in this particular setting – an example of auditory speaker recognition which did not involve any tape-recorded samples. (For details of this forensic identification task see Broeders and Rietveld 1995; Hollien et al. 1995; Künzel

In the 1960s the so-called voiceprint technique of speaker-identification gained widespread acceptance in US courts, mainly because proponents claimed that it was objective and highly reliable. Basically the technique consists of a visual comparison of broad-band spectrograms by people professing to be 'experts', most of whom have no scientific education in speech science or related areas. In fact, practitioners are sometimes former police officers, electrical engineers or just private investigators who have little more to offer than a high-school graduation certificate and a set of training courses provided by a 'professional' organization. Voiceprint identification can be contested – and in fact has been in the past – on theoretical and technical grounds. The most fundamental issue is that it has never been shown that the inter-individual variation of spectral and/or temporal features of a spectrogram is greater than intraindividual variation (Bolt *et al.* 1979: 2, 32). There are, however, examples to the contrary (cf. the figures on pp. 216–19 in Hollien 1990).

Other factors are related to the unpredictable effects of telephone transmission (which has to be dealt with in the vast majority of cases handled

at the speaker identification laboratory of the Bundeskriminalamt). In our experience, it would also be highly unlikely that another pre-requisite of voiceprint examination could be met – i.e. that a reference sample could be obtained which contained the same words and phrases as the sample under examination. Moreover, the individual ability of different experts, especially their personal thresholds for identity vs. non-identity judgements, comes into play. In summary, it is probably not unfair to say that the visual interpretation of spectrograms does not solve the problem of identification but simply shifts it from the auditory to the visual domain (see Künzel 1994).

After severe problems with voiceprint identification in the 1970s the US Ministry of Justice commissioned a group of experts in acoustic engineering and speech science to analyse and evaluate the spectrogram comparison technique. In their extensive report the assessors, Bolt *et al.*, caution against its use for legal purposes and describe it as 'an empirical art' rather than a science (1979: 11f.). The opinion of the present authors on this matter is that superior techniques for speaker-identification are now available. Therefore, it would be indeed perverse and unethical to ignore these and to continue to employ the visual comparison of spectrograms.

On the other hand, there was a long tradition, particularly in the UK, of carrying out speaker-identification based on aural-perceptual analysis alone. This type of analysis largely consisted of a detailed auditory assessment of regional accent and distinguishing vowel and consonant features, combined with an impressionistic evaluation of voice pitch, voice quality and other prosodic features (Baldwin and French 1990). While traditional listening techniques still form an important element in both speaker-profiling and speaker-identification, their exclusive application can no longer be considered sufficient for these purposes. Nolan (1983; 1994) takes a rather critical view, addressing specifically the shortcomings of using auditory or spectrographic methods alone. While making it clear that 'speaker recognition as a discipline can develop successfully only as an integral element of linguistic phonetics' (1983; 206), he emphasizes the necessity for extensive testing of any method to be used in forensic speaker-identification, as well as stating the limitations of current approaches. In 1990, a group of French speech scientists, most of them with a background in engineering, took up Nolan's argument (Chollet 1991). They concluded that all methods for speaker-identification available at that time were far from perfect and suggested that experts should be allowed to carry out forensic phonetic casework only after having provided scientific proof of their competence. In 1980, the British Association of Academic Phoneticians (BAAP) took a similar view. However, several of those who were opposed to carrying out forensic work have since changed their minds in view of recent developments e.g. Nolan, who, in 1991, emphasized the responsibility of phoneticians

in this area. BAAP has taken no stance on the ethics of phoneticians carrying out forensic work for the past nineteen years.

The International Association for Forensic Phonetics (IAFP) was formed in 1989, with the explicit aim of providing a forum for discussion among those who are active as forensic phonetics experts and/or have an academic interest in the field. It has established a professional Code of Practice, which is binding upon its members, as well as Guidelines for Keeping a Record of Analysis.

This brief account contains in essence the three current attitudes towards forensic phonetic activities (or at least towards forensic speakeridentification), ranging from almost unconditional approval, through a guarded approval taking into account the possibilities and limitations of techniques currently available, to categorical rejection. Furthermore, some scientists propound the somewhat irrational approach of working for one side (i.e. the defence) only. It will be argued below that this attitude is by no means more ethical than the others.

CENTRAL ISSUES IN FORENSIC SPEAKER-IDENTIFICATION

Looking at the issue from the perspective of forensic phonetics experts with many years of experience, categorical views such as those just mentioned would seem rather short-sighted and do not take the complexity of the whole discipline into account. They focus too much on voice comparison, which is, of course, a central forensic phonetic task. However, there are quite a few other forensic tasks, some of which are likely to be uncontested on ethical grounds. Categorical views also fail to acknowledge the numerous non-linguistic constraints which are often imposed by the specifics of a case. In an attempt to clarify the issue the main areas of forensic phonetics - or what we prefer to call 'forensic speech-processing' - will briefly be outlined (cf. also Braun 1995; French 1994; Künzel 1998).

Voice comparison has already been mentioned – i.e. the comparison of a speech sample produced by an unknown speaker involved in committing a criminal offence with that of one or more reference speakers. Other activities include speaker-profiling (see below); the analysis and documentation of disputed utterances; the identification of background sounds; the design, execution and interpretation of voice line-ups; speech enhancement – i.e. enhancement of the intelligibility of noisy tape recordings using (mostly digital) signal processing techniques; and tape authentication. Of these additional tasks, speaker-profiling is probably the one which raises most ethical concerns. At the same time, it is the one which can have the most immediate relevance to a victim's life, since the recorded voice of a kidnapper or blackmailer often forms the only lead in the early stage of an investigation. Literally hundreds of

cases have shown that a detailed speaker profile may be of paramount importance to investigators in their attempt to narrow down the number of potential suspects. Typically, an anonymous voice sample is analysed in terms of sex, age group, regional and social accent, or foreign-language influence, peculiarities or defects in the pronunciation of certain speech sounds, mannerisms etc.¹ It is particularly important for the investigators to have this type of information (and to have it quickly), because these investigations are normally carried out under extreme time pressure, and being able to exclude certain possibilities – i.e. individuals on the basis of a voice profile – saves time and resources.

This last point touches upon another ethically relevant issue in forensic phonetics: the time factor. Under non-forensic conditions, scientists normally take their time to carry out a project. Most often a concrete time frame will be set at the very start of a project. It goes without saying that unforeseen events may delay the course of a study and create new and more immediate deadlines. In principle, however, a scientist will come forward with results only when he/she deems them reliable and final. In the forensic setting, the time scale is often dictated by the criminal, particularly in cases of kidnapping or extortion, where ransom money has to be delivered. In these cases, quick results are needed and thus expected from the phonetics expert within, say, a few hours, which is often quite unacceptable from a strictly scientific point of view. However, if a suspect held in custody can be convinced that his or her cause is lost and that revealing the location where the victim is being held will have a favourable effect on his own trial, his timely confession may save the victim's life. There are many cases in the practice of the Bundeskriminalamt Laboratory where voice comparison reports carried out speedily did have exactly that effect.

CURRENT APPROACHES TO FORENSIC SPEAKER-IDENTIFICATION AND THEIR ETHICAL IMPLICATIONS

It is evident from the above considerations that the issue of ethics in forensic phonetics takes into account both the necessity and urgency of the task and the methods available for carrying it out. Since the parameters analysed in speaker-profiling are largely sociographic and linguistic, there is no opportunity to use any kind of automatic approach. Expert systems like DRUGS (Datenbank regionaler Umgangssprachen des Deutschen; Künzel and Köster 1995) may be used to aid the identification and analysis of regional accents;² F0 (i.e. fundamental frequency, the acoustic correlate of perceived pitch) may be measured in order to distinguish between the voice of a child and that of a high-pitched female, but the bulk of the work, however, will always have to be carried

out by the linguistically trained phonetician. With regard to voice comparison, a number of automatic or semi-automatic procedures, which may be applicable for forensic purposes, are currently available or are being tested. Here the parameters used – such as cepstral coefficients – are no longer 'natural' in the sense that they have close auditory or articulatory phonetic correlates (Hollien 1990: 238; Hollien, Gelfer, Huntley 1990: 71). It is therefore all the more important to explore the possibilities and limitations of any and all procedures which are to be applied. This is true for automatic and what may be called 'hybrid' procedures, such as the one employed by the BKA and many other forensic laboratories throughout the world. At first glance, hybrid methods might seem to be more error-prone than automatic devices because they involve a subjective component. For more than one reason, however, this is a rather short-sighted view. In principle, the former also offer more possibilities for correcting errors. With regard to wholly automatic procedures, the first and most important decision to be made by the expert is whether they are at all applicable to a given case. Once this question has been answered in the affirmative, a larger or smaller number of variables including threshold values, filter coefficients, the size of the relevant population³ etc. have to be set – by the expert (Broeders 1995). Mistakes in any of these settings will put the ensuing automatic analysis on the wrong track. Another problem of automatic procedures is that, owing to their complexity, each individual step and the relative weight thereof for the calculation of the distance measure or similarity index cannot easily be reconstructed. Furthermore, it is difficult to interpret the meaning of individual distance measures. Therefore, extensive testing which involves systematic variation of all variables is needed in order to assess their possibilities and limitations.

It goes without saying that other, 'less automatic', methods have to be tested as well. Here, not only is the ultimate result – essentially a yes–no decision – relevant, but also all the individual steps in the comparison will have to be scrutinized as well - for instance, analysis of the regional accent, measurement of F0 and its variation etc. The question of which parameters should for one reason or another *not* be considered in a given case also has to be examined – for instance average F0 in the case of certain types of voice disguise or speaking style; syllable or articulation rates if the sample is too short etc. A suitable test format for interactive speaker-identification procedures seems to be the classical double-blind approach which is widely used in medicine, particularly for the comparison of the therapeutic performance of different drugs. In fact, the BKA method had to undergo such tests before it was cleared for forensic use. Double-blind tests will certainly have to be considered as an element of quality-control assessment in forensic speaker-identification (see below).

CONCLUSIONS

So far, at least two ethically relevant dimensions to forensic phonetic casework have been identified – one related to the principal issue of whether or not there is sufficient information on the individual voice to recognize it (under all circumstances) and one related to the amount of time which is available to perform this task. With regard to the first issue it has to be stated unambiguously that phonetic science cannot at the present time offer a universally applicable theory or automatic procedure which will provide a reliable identification of any speaker under each and every circumstance, and it is perhaps not trivial to add that the same holds true for the kind of auditory speaker-identification which we carry out in everyday life. The reason is simply that even if in principle the voice of an individual could be regarded as unique, forensic conditions, such as telephone transmission, voice disguise, short speech sample etc., can degrade or destroy many or even most of the speaker-specific features to the extent that individuality is lost – just as a partial or blurred fingerprint is not sufficient for the purposes of identification. But does this really imply that unless there is a perfect procedure - which, in the opinion of the present authors may perhaps never be available – there is a moral obligation to refrain from doing forensic phonetic casework altogether? Surely not, because forensic phonetics is arguably in the same position as, for instance, the medical sciences with respect to the treatment of cancer: at the present time, there is no absolutely guaranteed treatment at hand, nor does one seem imminent. Yet nobody, whether doctor or patient, would refuse a treatment which could at least alleviate the illness and might even cure it. The question could well be asked: can the withholding of treatment in such cases be ethically justified? In such a situation a doctor is always bound to the principle *nihil nocere*, i.e. never to inflict harm upon his patient. This principle should apply to the forensic phonetician as well, replacing 'patient' with 'speaker in question'. One point should be made: refusing to do any forensic work at all also means refusing to exonerate the innocent. In fact, since there is often more than one suspect, those cleared by the phonetician are - in certain jurisdictions at least - far more numerous than those implicated by their reports.

The argument that one should work for one 'side' only (i.e. for the defendant) deserves further comment. When considering this issue, two very different judicial settings have to be taken into account: the adversarial vs. the inquisitorial system. The latter, which exists in those European countries whose judicial system follows the tradition of Napoleonic law, means that the expert is commissioned by the court and that a full report of his or her findings will enter the court proceedings no matter what. The former relies on case law and, more importantly, may have an expert on each side. This can, however, involve expert testimo-

ny being heard only from one side, as the defence has the option of ignoring any unfavourable reports from an expert. Also, an expert may be denied the opportunity to present his or her findings in full, because the evidence is essentially 'lawyer-led' and thereby is selective. It is for these reasons that both the present authors, who have carried out casework in both legal frameworks, feel that the constraints imposed on the expert by the adversarial system may result in an ethical conflict which is unrelated to the science of forensic phonetics: the expert may know the 'whole truth', but he or she may not have the chance to state it.

Irrespective of the legal system, some have argued in favour of working for the defence only, on the grounds that falsely acquitting a guilty defendant is a lesser evil than a false conviction. While this is certainly true from a strictly legalistic point of view (*in dubio pro reo*), it may equally well have catastrophic consequences in individual cases. There are, unfortunately, many instances of falsely acquitted individuals committing new crimes, often of the same type, once they have been released. This is especially true for rape, child abuse and other sexual offences. One might therefore argue that any expert who contributed to the acquittal is at least morally responsible for such events. Forensic psychiatrists advising the court on matters of diminished responsibility or early release from prison are probably most affected by this dilemma, but in principle phoneticians giving evidence on speaker-identification are in the same position.

In view of these difficulties, it is all the more important that the expert exercises the utmost caution when carrying out the analyses as well as in the phrasing of his or her conclusions. This basically involves tailoring the scope of analysis and the formulation of conclusions to the individual case in a scientifically justified and responsible manner. In order for this to happen, certain conditions have to be met. First of all, the expert really has to be an expert in speech science/phonetics. This is by no means trivial, as is exemplified by the 'voiceprinters', but there have also been cases more recently of audio lab owners⁴ and other charlatans⁵ appearing in court as 'expert' witnesses. This is clearly a development which the whole scientific community needs to curb. Secondly, the expert needs to familiarize himself with the specifics of the forensic setting, i.e. the possibility of degraded and/or disguised material in order to be aware of the limitations which this may impose on the analysis. Thirdly, the expert needs to be impartial. Leaving aside any additional difficulties caused by the adversarial system, he or she may have to disappoint his or her clients by a non liquet decision or a very low probability rating if there is very little and/or badly degraded material. Those who have done actual casework will know that this is often much more difficult to do than living up to the role of an expert by drawing firm conclusions.⁶ The International Association for Forensic Phonetics, through its Code of Practice, by which its members are bound, has taken a first step towards creating standards. The IAFP has the establishment of an accreditation procedure for experts in this field well underway.

SUMMARY AND OUTLOOK

Forensic phonetics involves a number of tasks, most of which are ethically uncontested. With respect to the most controversial one, namely speaker-identification, the present authors feel that while there is certainly no perfect solution available at present, the scientific community is under a moral obligation to contribute whatever it can to aid the course of justice. The emphasis here is clearly on 'whatever it can', because this is a decision which has to be made by the forensically experienced phonetician. A total refusal to carry out any forensic work will not only result in others, who are likely to be less qualified, doing it; it will also result in a loss of credibility for the phonetic sciences.

What is clearly needed in the future is to try and expand the 'whatever it can' by joint research initiatives involving forensic phoneticians and their colleagues in academia in order to study questions arising from practical work and gain a more complete insight into the concept of the individuality of voice. Despite recent advances in automatic speaker-verification technology, especially with regard to the robustness of parameters, enlarged sizes of speaker groups and new statistical algorithms, high-risk applications such as those found in the forensic domain are at present out of the question. The ultimate goal of interdisciplinary research would be a more or less automated procedure which could be used under clearly defined conditions on the speech material in question. However, owing to the reasons discussed earlier it would seem that – at least for the foreseeable future – a universally applicable automatic procedure is at best unrealistic (cf. the more pessimistic view of Boves 1998:158).

We must do what we can, whilst treading carefully and acknowledging our limitations.

NOTES

- It ought to be emphasized at this point that the 'etc.' expressly does not include any personality judgements or an assessment of the sincerity of threats uttered. If this is at all feasible (of which there is serious doubt based on the relevant research), it is clearly beyond the scope of phoneticians. Therefore, the International Association for Forensic Phonetics (IAFP) requires its members to refrain from this type of work.
- The currently available algorithms for *foreign* accent spotting are in no

- way sufficient in the forensic context.
- The size of the population in question forms one of the many fundamental differences between commercial speaker-verification and forensic speakeridentification. Whereas there is always a closed set containing a relatively small number of potential customers in the first application, an open set of speakers has to be assumed in the vast majority of forensic cases. This has obvious consequences for stating the results: whereas exact error rates can be indicated in the commercial setting, conclusions are expressed on a relative probability scale in forensic reports.
- Personal communication from Dr. J. P. French (1996).
- A flagrant case in Germany has recently called for official action by the authors. The individual in question is a professor of mineralogy who has repeatedly testified in North German courts. He justifies his expertise with the argument that speech is an acoustical signal much like the seismic waves studied in mineralogy and therefore may be analysed by the same methods. In this case, the 'method' consists - horribile dictu - of an interpretation of the similarity of simple long-term average spectra – derived from telephone-transmitted speech!
- This is particularly true if the client is paying and the expert depends on forensic casework for a living. The danger of saying more than is scientifically justifiable is much smaller for those who work for government institutions since they have nothing to gain from 'pleasing their customers' whereas on the other hand they have a reputation to lose (by stating rash conclusions which will not hold up in court). Thus, the view which is sometimes aired by defence lawyers that government experts are potentially less impartial than private experts may be pure courtroom strategy - at least in countries adhering to democratic principles.
- This should not be taken to imply an obligation for each individual phonetician. If someone has a personal disinclination to work for courts, this is of course a perfectly respectable attitude. But those who do feel that they can be of assistance should not be prevented from carrying out forensic work.

REFERENCES

Baldwin, J. and French, J. P. (1990), Forensic Phonetics. London/New York: Pinter.

Bimbot, F., Chollet, G. and Paoloni, A. (1994), 'Assessment methodology for speaker identification and verification systems', in Proceedings ESCA Workshop on automatic speaker recognition, identification, verification. Martigny, 75-82.

Boe, L.-J. (1998), 'L'identification juridique de la voix: le cas français. Historique, problématique et propositions', in Proceedings RLA2C Workshop on Speaker Recognition and its Commercial and Forensic Applications, Avignon, France, 222-39.

- Bolt, R. H. et al. (1979), On the theory and practice of voice identification. Washington, D.C.: National Academy of Sciences.
- Boves, L. (1998), 'Commercial applications of speaker verification: overview and critical success factors', in Proceedings RLA2C Workshop on Speaker Recognition and its Commercial and Forensic Applications, Avignon, France, 150-9.
- Braun, A. (1995), 'Procedures and perspectives in forensic phonetics', in Elenius, K. and Branderud, P. (eds), *Proceedings of The XIIIth Interna*tional Congress of Phonetic Sciences, Stockholm: KTH, vol.3, 146-53.
- Broeders, A. P. A. (1995), 'The role of automatic speaker recognition techniques in forensic investigations', in Elenius, K. and Branderud, P. (eds), Proceedings of The XIIIth International Congress of Phonetic Sciences, Stockholm: KTH, vol.3, 154-61.
- Broeders, A. P. A. and Rietveld, A. (1995), 'Speaker identification by earwitnesses', in A.
- Braun and J. P. Köster (eds), Beiträge zur Phonetik und Linguistik, Band 64, 24–40.
- Chollet, G. (1991), 'About the Acoustics of Speaker Identification', in Proceedings of The XIIth International Congress of Phonetic Sciences, Aix-en-Provence: Université de Provence, vol.1, 397.
- French, P. (1994), 'An overview of forensic phonetics with particular reference to speaker identification', Forensic Linguistics, 1(2): 169– 81.
- Hollien, H. (1990), The acoustics of crime. The new science of forensic phonetics. New York: Plenum Press.
- Hollien, H., Gelfer, M. P., and Huntley, R. (1990), 'The natural speech vector concept in speaker identification', in Borowski, V. A. and Köster, J. P (eds), Neue Tendenzen in der Angewandten Phonetik III, Hamburg: Buske, 71–87.
- Hollien, H., Huntley, R., Künzel, H. J., and Hollien, P. (1995), 'Criteria for earwitness lineups', Forensic Linguistics 2 (2): 143-53.
- Künzel, H. J. (1987), Sprecher-Erkennung. Grundzüge forensischer Sprachverarbeitung. Heidelberg: Kriminalistik-Verlag.
- Künzel, H. J. (1994) 'Current approaches to forensic speaker recognition', in Proceedings ESCA Workshop on automatic speaker recognition, identification, verification. Martigny, 135–41.
- Künzel, H. J. (1994), 'On the problem of speaker identification by victims and witnesses', Forensic Linguistics 1 (1): 45-58.
- Künzel, H. J. and Köster, J. P. (1995), 'Forensic Data Bank System of Regional Accents of German', in Proceedings Annual Meeting of the American Academy of Forensic Sciences. Seattle WA, 88.
- Künzel, H. J. (1998) 'Forensic speaker identification: A View from the Crime Lab', in Proceedings COST250 – Workshop on Speaker Recognition by Man and by Machine: Directions for Forensic Applications, Ankara, Turkey, April 1998, 4-8.

- McGehee, F. (1937), 'The reliability of the identification of the human voice', Journal of General Psychology 17: 249-71.
- McGehee, F. (1944), 'An experimental study in voice recognition', Journal of General Psychology 31: 53-65.
- Nolan, F. J. (1983), The phonetic bases of speaker recognition. Cambridge: CUP.
- Nolan, F. J. (1991), 'Phonetics in the next ten years', in Proceedings XIIth International Congress of Phonetic Sciences, Aix-en-Provence: Université de Provence, vol. 1, 125-9.
- Nolan, F. J. (1994), 'Auditory and acoustic analysis in speaker recognition', in Gibbons, J. (ed.), Language and the law, London/New York: Longman, 326–45.