

# Conference report

## Annual Meeting of the International Association for Forensic Phonetics, Paris, 4–6 July 2001<sup>1</sup>

*Michael Jessen*

*Department of Speaker Identification and Tape Analysis  
German Federal Police  
Michael.Jessen@bka.bund.de*

This year's meeting of the International Association for Forensic Phonetics (IAFP) was organized in conjunction with the meeting of the ENFSI (European Network of Forensic Science Institutes) Expert Working Group for Forensic Speech and Audio Analysis. These events took place from 2–3 July (ENFSI) and 4–6 July (IAFP) at the Institut de Recherche Criminelle of the Gendarmerie Nationale (IRCGN) in Rosny Sous Bois, near Paris.

Following the official opening of the IAFP conference and a presentation of the IRCGN, Andrea Paoloni (Fondazione Ugo Bordononi, Rome) presented an overview of different approaches to speaker recognition. In his lab, both automatic and semi-automatic systems are used regularly, and the efficiency of these methods is tested on the basis of the FOrensic CorpUS – a database specifically designed to replicate forensic conditions. Automatic procedures include pattern-matching algorithms, vector quantization, and Gaussian mixtures. The semi-automatic IDEM system, among other features, offers formant candidates that can either be accepted or corrected according to the expert judgment of the forensic phonetician.

The remaining presentations of the conference were grouped into sessions on 'features for speaker recognition and speech analysis,' 'forensic

audio and tape,' 'semi-automatic speaker recognition,' 'case report', and 'studies and experiments'.

### FEATURES FOR SPEAKER RECOGNITION AND SPEECH ANALYSIS

Francis Nolan (Cambridge University, Cambridge UK) presented an experiment in which the hypothesis was tested that the detailed alignment of pitch targets with the segmental string differs between speakers. He was able to show that the temporal alignment of the midpoint of an F0 fall has a substantial speaker-discriminating value. This effect, however, was statistically less robust than the F3-of-/r/ data presented in his 1983 book, which were used as the basis of comparison in this experiment.

Peggy Bonnet and Franck Marescal (IRCGN, Rosny Sous Bois) investigated the contribution to speaker identification of parameters based on pitch and the structure of the harmonics. These parameters, which were explained in detail, had been modelled using a Gaussian mixture mode (GMM). Using a database of 100 speakers, 50 of whom were also recorded over the telephone, Bonnet and Marescal showed that the inclusion of pitch- and harmonics-based information increases the speaker identification performance as compared to a system that is based on mel-cestral coefficients alone.

As part of a PhD project in which the prosody of French and Somali were compared, David Le Gac (IRCGN, Rosny Sous Bois) presented data on the F0 correlates of focalization in French. He carefully distinguished within- and across-speaker variability in the phonetic implementation of focus and pointed out implications for speaker recognition.

### FORENSIC AUDIO AND TAPE

Philip Harrison (J P French Associates, York) demonstrated how mobile telephone transmission can be degraded acoustically and auditorily by electromagnetic radiation leakage. The goal of his study was to develop an algorithm for the automatic cancellation of the pulses that result from GSM interference. Since notch filtering alone cannot remove the pulses, he based his procedure on single channel adaptive filtering. The audio result of his speech enhancement algorithm was convincing. The contribution by Harrison was voted best student paper by the IAFP members.

Lorique Noglotte and Franck Marescal (IRCGN, Rosny Sous Bois) presented the results of a comparison between different time-frequency diagrams for the purpose of speaker identification and the analysis of gun shots. Advantages and limitations of diagrams including wavelets, Wigner distribution, and classical spectrograms were discussed.

Dagmar Boss (Bavarian LKA, Munich), Stefan Gfroerer (Bundeskriminalamt, Wiesbaden), and Nikolai Neoustroev (Research Institute for Material Science and Technology, Moscow) presented a new method for

the visualization of magnetic features on tapes. This method is based on crystals that change their optical properties according to external magnetic fields. The images resulting from this method have higher resolution than traditional methods based on ferrofluids etc. and they have no destructive effects on tapes. It was shown how this procedure can be applied to tape authentication and the detection of unique signatures of different tape recorders.

### SEMI-AUTOMATIC SPEAKER RECOGNITION

Peggy Bonnet and Franck Marescal introduced this section by presenting an overview of the Speaker Recognition Workshop that took place in Crete from 18–22 June 2001. Following this introduction, Joaquin Gonzalez-Rodriguez, Javier Ortega-Garcia (Universidad Politecnica, Madrid), and J. J. Lucena-Molina (Guardia Civil, Madrid) presented their 'IdentiVox' software, a Windows-based tool for speaker recognition within a forensic context. This is a GMM-based text-independent speaker recognition system that provides a Bayesian evaluation for legal purposes. The system was tested on a NIST-selected database of Spanish, and its success was presented in the form of Tippet plots. The system scored high in the NIST 2001 evaluation at the Workshop in Crete.

Franck Marescal presented a speaker recognition platform that was developed in his lab. This platform incorporates recent improvements in speech processing (GMM etc.) and evidence interpretations (Bayesian framework). For the system to work properly, the suspect voice has to be recorded under the same conditions as the samples in the large reference database of French speech. The approach presented here is recognized in the French legal system.

Ying-Li Wong (GuangDong Provincial Public Security Department, GuangZhou) and Fu-Chiu Kwok (Hong Kong Government Laboratory, Hong Kong) gave an overview of the situation on speaker recognition in Hong Kong and the province of Guangdong. The range of criminal activities involving speech communication was explained and some spectrographic speaker identification procedures by the forensic expert were illustrated. Since a new lab is being developed in Hong Kong, inquiries to the audience were made regarding the acquisition of state-of-the-art technology in forensic phonetics.

### CASE REPORT

Ton Broeders, Tina Cambier-Langeveld, and Jos Vermeulen (Netherlands Forensic Institute, Rijswijk) presented a case in which in the disputed material a native speaker of Irish English spoke to a Dutch man. The conversation took place in English, and there was evidence that both speakers adapted their speech according to the communicative needs of the other. For this reason it was necessary to simulate the communicative situation as

closely as possible. This was performed by having an Irish and a Dutch suspect talk to each other over the phone in a Map-Task-like setting where the participants had to cooperate in identifying certain events and locations in a picture that was visible to both. The speech material obtained in this manner was very useful for the voice comparison process. (See above, pp. 85–91.)

The case discussed by Peter French (J P French Associates, York) was similar to one discussed in his publication in 1990 (in H. Kniffka (ed.) *Texte zur Theorie und Praxis forensicher Linguistik*, Tübingen: Niemeyer). In both cases the task was to decide whether an incriminating ‘can’ or a non-incriminating ‘can’t’ was spoken by a non-native speaker of English. The problem was that in both cases the native language has no vowel distinction corresponding to these words (presence/absence of the final ‘t’ was no reliable cue either). The chosen method was to gather natural conversation data showing whether the speaker makes a consistent phonetic difference between ‘can’ and ‘can’t’ and to use this material as reference in the analysis of the incriminating speech sample. The result of the recent case was that the speaker made no consistent distinction, hence the question of whether an incriminating ‘can’ or a non-incriminating ‘can’t’ was spoken could not be answered.

J. J. Lucena-Molina (Guardia Civil, Madrid) discussed the status of closed-set speaker identification. While it is clear that the identification of a speaker among a group of suspects (closed set) is important for investigative purposes, there is no agreement as to whether closed-set speaker identification has any formal status in the judicial context. In this context only open-set tests are proposed because the identification problem turns into a verification task. Cases were presented to illustrate this problem.

Ton Broeders, Tina Cambier-Langeveld, and Jos Vermeulen presented a difficult voice line-up case. Not only had the line-up been conducted in an uncommon language (Sranan Tongo) and not only was there a long time lapse between language exposure and line-up, but most importantly, the witness had already ‘recognized’ (in his judgment) the speaker in a conversation one day after exposure to the incriminating speech sample. If the witness were to identify the suspect in the line-up it would be possible that the identification was based on the recognition of the voice heard one day after the incident and not on the recognition of the voice heard during the incriminating incident. Despite these unfavourable conditions the line-up was carried out nevertheless on demand of the Court of Appeal (using the help of a native speaker to gather samples from other comparable Tongo speakers). It turned out that the witness was unable to identify the suspect at all. Probably due to the long time lapse, the ability to recall the voice both from the day of the incident and from the conversation one day later must have become too weak to enable an identification. General implications of this case were discussed. (see above, pp. 102–10.)

## STUDIES AND EXPERIMENTS

Gea de Jong and Terry Honess (City University, London) presented an experiment on the perceptibility of tape speed manipulations. This experiment constitutes an elaboration of a pilot study presented at last year's IAFP conference in Rome. Some of the earlier results were replicated, including those showing that incorrect speed was easier to detect when the sample was too fast rather than too slow, and those showing that expert phoneticians did not have consistently higher perception scores than naive or trained listeners. The revised study, which added confidence ratings to the experimental design, has shown that expert phoneticians – though not better in their overall perception score than others – have better performance for cases with a high confidence score. (see above, pp. 44–55.)

Franck Marescal and Olivier Cappe (IRCGN, Rosny Sous Bois) investigated the perceptual effect and the impact on automatic speaker recognition scores of pitch transformation procedures available in the Soundforge software. Pilot perception tests showed that modifications of the pitch in the order of a few per cent are usually not perceptible. A formal speaker identification experiment has shown that even with moderate factors (in the present experiment a 10% increase) the automatic speaker identification performance was severely reduced.

Gea de Jong, P. R. Newis, and J. Hunt (City University, London) (see above, pp. 56–71) addressed problems associated with the use of MiniDisc technology in applications such as audio harvesting by law enforcement agencies. The use of bit reduction techniques in MiniDisc technology combined with compression algorithms leads to speech audibility problems, especially at the low-amplitude speech and high-amplitude noise levels commonly found in forensic material. In addition, each copy from MiniDisc to MiniDisc leads to a change of information. An experiment has been conducted to investigate the impact on speech intelligibility in background speech babble of MiniDisc-recorded speech and multiple generations of MiniDisc-to-MiniDisc copies. Surprisingly, later generations of MiniDisc copies were sometimes more intelligible than earlier ones. Possible reasons for this effect were discussed.

Outside the scope of the five sessions, S. Keith McElveen (Signalscape Corporation, Raleigh) discussed problems concerning the use of digital-format recordings and their processing in the US legal system and presented solutions based on software by Signalscape and other companies. Increasingly, evidence is recorded digitally or digitized for processing and for presentation in court. In order to ensure authenticity under these circumstances it is recommended to assign a digital signature to this material. Furthermore, given increasing standards in quality management it would be useful if the speech processing software left a 'trail' of all the steps that have been carried out in the preparation and evaluation of forensic material. A demonstration of a current version of Signalscape's 'Digital Signatures and Audit Trail' was given.

At the IAFP Annual General Meeting topics and issues concerning members of IAFP were discussed. The IAFP homepage has been updated and expanded (<http://www.iafp.net>).

In all, the conference was very well organized by this year's host, the French Gendarmerie Nationale, and a rich cultural and culinary programme was offered, culminating in the Conference Dinner in the Sky of Paris Restaurant and a bus round-trip through Paris.

#### NOTE

- 1 This report was first published in *The Phonetician* 83, 2001–2. The editors of *Forensic Linguistics* are grateful to the editors of *The Phonetician* for their permission to reprint it here. The present edition of *Forensic Linguistics* includes written versions of the oral presentations given at the conference by Nolan; de Jong and Honess; de Jong, Newis and Hunt; and both case reports by Broeders, Cambier-Langeveld and Vermeulen. A written version of Harrison's presentation was included in Volume 8(2) of *Forensic Linguistics* (2001).