L'identificatione della persona per mezzo della voce. Atti della tavola rotonda, Padova 14-15 Sept. 1978. F. E. Ferrero (ed.). ESA (Edizioni scientifiche associate) Roma.

Speech being an intra-species mode of communication of the biogenetic entity homo sapiens, it is only natural that ethnic language should be, and has been, for about 2 1/2 thousand years now, the object of scientific or quasi-scientific study. The spoken form of language contains, however, not only linguistic information — which until recently was the only object of language study — but also information about the speaker, which only became the object of scientific interest about 20 years ago, although the development of telephony and radiotelephony should have made the problem topical long before that. Personal voice characteristics are not irrelevant for psycholinguistics. Still, a serious motivation for wide-scale research did not arise before the forensic implications were realized. The human voice turned out of be not only a means of communication, but also a vehicle of crime, one of the commonest modes to which is blackmail by telephone. Some agencies of public security and state defence may also be concerned with voice recognition.

Scientific papers dealing with voice identification and verification appear sporadically in a few acoustical, communications, phonetic and criminological journals. The following monographs devoted to the subject should be mentioned: HECKER [2], KIRSTEIN [3], RAMISHVILI [4], TOSI [5] and BOLT et al. [1]. Though some of the work in this area is reported at such acoustic conferences as the International Congress on Acoustics or the International Conference on Acoustics, Speech and Signal Processing, yet the first large-scale meeting exclusively devoted to the subject was the round-table conference on "The Identification of the Speaker by his Voice" organized by the Italian phonetician Franco E. Ferrero (now director of the Institute of Glottology and Phonetics, University of Padua), under the auspices of the Italian Acoustical Society in the spring of 1979. The same year saw the appearance of the carefully edited proceedings of the conference, which are reviewed here. It was a national meeting, so almost all the participants were affiliated to Italian institutions, and represented the following areas: physics, phonetics, applied mathematics, electronics, electrical engineering, communications, computer science, linguistics, dialectology and psychology. Also present were representatives of the Defense Ministry.

The individual Sections included: (1) methodology and technology, (2) linguistic analysis, (3) recognition by temporal parameters, (4) recognition by spectral parameters, (5) identification based on the inspection of spectrograms and on auditory judgment, and (6) general problems of recognition.

The papers in Section 1 were mainly concerned with two problems, viz. the effect of disturbance and distortion of the signal on voice recognition rate and the alleged parallelism between identification by finger prints and by "voice prints", erroneously maintained by some American specialists. Particular note was taken of the invariability of the former compared with the substantial variability of the speech signal produced by a given voice, and of the relative ease of imposture in speech. Much attention was directed towards distortions and disturbances in telephone channels.

The leitmotif of Section 2 was the importance of dialect and linguistic-phonetic studies. There is probably no country interested in the forensic aspects of voice studies that lays so much emphasis on dialectal and linguistic-phonetic features as Italy. This accounts for an involvement of linguists in research and expertise that is stronger there than elsewhere.

The papers read in Section 3 pointed out the significance of such parameters as the duration of consonantal occlusion, relative phonation time and speech tempo for the procedures of grouping voices into classes and, in some cases, for the identification of the speaker.

Section 4 was devoted to tutorials of new techniques of acoustic parameter extraction, such as the Fast Fourier Transform and cepstral analysis. An application of this type of analysis is exemplified by forming a trajectory in the F_1 F_2 plane. The final decisions are taken either by the experimenter on the basis of the results of his analysis or – with the aid of an identifying algorithm – by a computer. Mathematical-statistical methods of multivariate analysis are frequently used.

Section 5 presented auditory and visual methods of voice recognition. The latter is based on the inspection of "visible speech" pictures (sonagrams). In the early sixties L. G. Kersta secured himself a sparkling though somewhat short-lived career in the USA promoting the sonagrams as fully reliable evidence in cases of voice identification, said to be as convincing as the finger prints (hence the term "voice print"). A large group of phoneticians, both from the USA and other countries launched a concentrated campaign against Kersta – a fact hardly noted in the proceedings – showing by scientific argumentation and counter-evidence that he was working under self-deception. It transpired with time (after a number of American acousticians cautiously joined the "anti-voice-print" front) that American courts of justice have passed several wrongful sentences on the basis of evidence given for the prosecution by Kersta or his disciples. Although, after a ten-year-long battle, Kersta had to suffer moral and financial defeat (his firm went bankrupt), American courts even now take recourse occasionally to the visual analysis of sonagrams. The problem seems at present to be scientifically settled. In this volume F. E. Ferrero and his associates once again prove thoroughly and comprehensively that sonagrams should not be used in court evidence and with the utmost caution in investigation, especially if the voice was transmitted over a telephone line. Ferrero supports his view by concrete evidence which he presents in another paper. Only a sketchy article is devoted to auditory recognition.

The reports included in Section 6 are general in character and present some legal aspects of voice recognition as well as some basic statistical methods which are, or can be, used in this area. However, such methods as statistical clustering or multiple scaling, which can also be applied, are not discussed.

The closing paper contains some proposals for the main lines of attack in the field under discussion, namely : (a) improvement of the methods of extracting such acoustical parameters from the speech signal as are most effective for characterizing individual voices, (b) development of statistical methods leading to correct identifications, (c) optimization of the decision-making procedures, and (d) normalization of the expertise.

The volume under review contains, on 305 pages, information which, though only coming from Italian sources, includes by virtue of its comprehensiveness, combined with a critical comparison of the methods described with those used elsewhere, valuable scientific and practical material. Research workers involved in the study of human voice and experts engaged in law-enforcing agencies should not fail to acquaint themselves with the results of the conference.

References

[1] R. H. Bolt et al., *On the theory and practice of voice identification*, National Academy of Sciences, Washington DC 1979.

[2] H. L. Hecker, *Speaker recognition. An interpretive survey of the literature*, American Speech and Hearing Association, Washington DC 1971.

[3] M. KIRSTEIN, *Akustische Untersuchungen zur automatischen Sprecherklassifikation*, H. Buske Verl. Hamburg 1971.

[4] G. S. RAMISHVILI, *The speech signal and the individuality of the voice* (in Russian), Izd. Mecniereba, Tbilisi 1976.

[5] O. TOSI, *Voice identification, theory amd legal applications*, Univ. Park Press, Baltimore 1979.

Wiktor Jassem (*Poznań*)