

CRIMINALISTIC PERSON IDENTIFICATION BY VOICE SYSTEM

Bernardas SALNA

Lithuanian Institute of Forensic Examination, Vilnius, Lithuania

ABSTRACT: Person recognition by voice system of the Lithuanian Institute of Forensic Examination is designed. Main application of this system is criminalistic person identification by voice. Identification features are presented by means of statistical distribution diagrams of specific parameters and correlation coefficients between these diagrams. Thus an expert can motivate his decision by the help of diagrams and specific numbers. This enables, to compare with traditional sonographic approach, better motivate phonoscopic examination, accelerate the investigation and reduce requirements for investigative and comparative speech records. The investigation becomes independent from the speech text recorded in the investigative and comparative speech records.

KEY WORDS: Forensic speaker recognition; Acoustic analysis.

Problems of Forensic Sciences, vol. XLVII, 2001, 268–272

Received 6 October 2000; accepted 15 September 2001

INTRODUCTION

Automated person identification by voice system is used in the Lithuanian Institute of Forensic Examination. This system consists of hardware, software and corresponding criminalistic person identification methodology. A block diagram of the system is presented in Figure 1. Software consists of two special software packages – SIS and SIVE. This software is devoted to investigation of criminalistic voice records (phonoscopic examination). The software package SIS (STC, Saint Petersburg, Russia) is devoted to investigation of criminalistic voice records. It consists of a number of programs for displaying and transformations of voice records. Also programs for filtering and automated converting to text of voice records (transcriber) are included. The main application of the software package SIVE is person identification by voice. This package was developed in collaboration with the Institute of Mathematics and Informatics (Vilnius) and the Lithuanian firm Technogama Ltd.

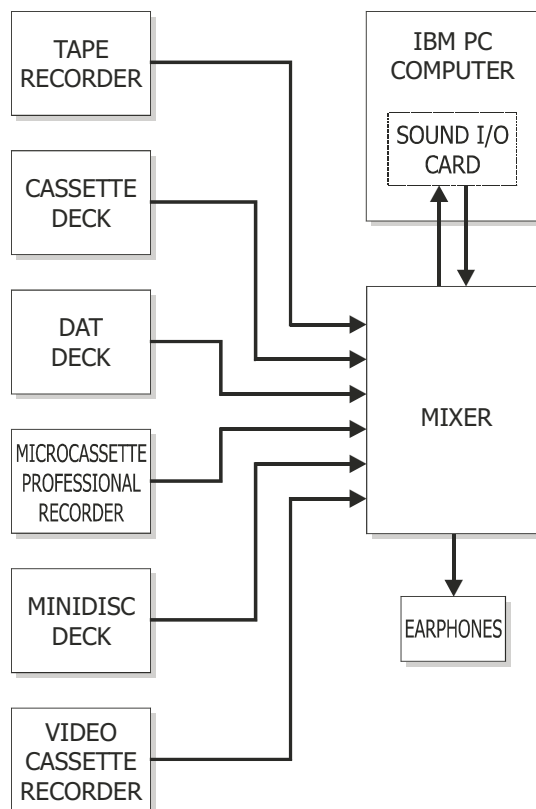


Fig. 1. A block diagram of the computerised working place of phonoscopic examination.

Since 1995 for speaker identification we use the combined method:

1. Auditory-perceptive analysis (we call it auditive analysis);
2. Phonetic-linguistic analysis;
3. Acoustic analysis.

Auditory-perceptive and phonetic-linguistic analysis is based on e.g. pronunciation manner, general voice quality, accent characteristic, lexicon and etc. These methods are described in the phonoscopic examination literature.

For acoustic analysis we use the semiautomatic system SIVE. SIVE is devoted to extraction of identification features from speech signals and their comparison, namely, calculation of pitch and its derivative parameters, calculation of relative distance between phonemes of the same type, calculation of relative distance between voiced stationary segments of speech and statistical evaluation of the obtained results. At the final investigation stage, identification features are presented by means of statistical distribution diagrams of specific parameters and correlation coefficients between these di-

agrams. Thus, after investigation, expert can motivate his decision by the help of diagrams and specific numbers. These enables to compare with traditional sonographic approach better motivate phonoscopic examination, accelerate the investigation and reduce requirements for investigative and comparative speech records. The investigation becomes independent from the speech text recorded in the investigative and comparative speech records.

SIVE

At the initial investigation stage (auditory-perceptive analysis) an expert is listening to the investigative and comparative speech records. Speech segments, which mostly represent person identity, are prescribed for computer investigation. In such way corresponding investigative and comparative speech records files are obtained.

For reliable results it is necessary to create a file for each investigative and comparative speaker, consisting at least 20÷30 s of a speech signal. If the signal is of poor quality (with noises or disturbed), it would be desirable to create file consisting of 50÷120 s of a speech signal.

Pitch and derivative from pitch features

The pitch frequency is one of parameters of voiced signals, which is least dependent upon the quality of recording, conditions and the channel. It is also important for speaker identification. SIVE package uses a frequency-autocorrelation method for pitch estimation. Due to physical differences in specific features of human speech tract there are many harmonics of the pitch and their amplitudes also are decreasing sooner or later. That is why additionally to the pitch (PGT) estimation it is calculated such pitch derivatives as the highest harmonic of the pitch (PGMTH), voice clearness (BS) and timbre (T). The results of the analysis are presented as a list of minimum, average and maximum values of PGT, PGMTH, BS and T parameters, their variance and variation coefficients, distribution diagrams and correlation coefficients, and final coincidence coefficient of the pitch parameters.

Relative distances between the phonemes

This method is based on assumption that by having two phonemes spoken by the same person, e.g. "A", and performing identification according to the first four formants, depending on the pronunciation of the sound (espe-

cially first three) and special features of speakers voice tract (especially third, fourth, fifth) the relative distance should be the smallest.

First of all both speech signals comparative and investigative are segmented manually in order to make a full set of vocalized phonemes. It is advisable to make segment in a way that total length of one single phoneme would be at least $0.5 \div 0.7$ s. In the next stage the matrix of identification parameters (features) is calculated for the phonemes segmented for both records. In this matrix for each phoneme a frame 25 ms length is allocated and 36 parameters are calculated. These parameters are made from parameters of different combinations of formants and spectral pairs. That forms the matrix of $N \times 36$ parameters, there N – number of frames in a given signal. Next, the identification of matrices vectors is done according to the frequencies of the first three formants, that is the comparison of the three elements from each vector of the matrix corresponding to the frequencies of the first three formants, and then a search for closest vector from the matrix under investigation is performed. Then the vector with the smallest distance according the first three formants is found, the absolute difference between each element of the vector is calculated. The total distance between given investigative and comparative phonemes is calculated for the final decision of the speaker identification.

In order to guarantee the reliability of the results achieved it is highly recommended to select at least two different phonemes from both records.

Relative distance between the investigative and comparative voice records

Every pseudostationary interval of voiced sounds from comparative and investigative record is described by linear prediction parameters (LPC) or cepstral coefficients, calculated from parameters of linear prediction model (LPCC), corresponding to the vocal tract and excitation signal. In that way we have two sets of parameter vectors corresponding to the speech signals: one, for investigative speech record and another for comparative. Then likelihood ratio distances between vectors of vocal tract parameters and between vectors of parameters corresponding to excitation signal are calculated for comparative and investigative speech records. Further the average minimal distance between parameters of investigative speech record and comparative speech record is calculated. This distance depends on the weight, assigned to influence of vocal tract and excitation signal parameters to the average distance.

Module of speaker verification is based on comparison of distributions of intra-individual and inter-individual distortions. This analysis allows answering a question if the same person as comparative one utters the investigative record or not. If speech records belong to the same speaker then the

distributions of intra-individual and inter-individual distortions should be the similar. By calculation estimates of these distributions – histograms it is possible to evaluate the degree (level) of coincidence and make a decision if the speech record belongs to the same person or not.

This method practically is fully automatic that is why analysis may be carried out very fast. Nevertheless, that is effective enough only in the case when investigative and comparative speech records are of high quality and made in the same recording conditions. Also it is necessary to have a lot of spoken material in investigative and comparative speech records, because this method is based on the assumption that both speech signals (investigative and comparative speech records) have equivalent and full sets of phonemes.

CONCLUSION

At the final investigation stage, identification features are presented by means of statistical distribution diagrams of specific parameter and correlation coefficients between these diagrams. Thus, after the investigation, an expert can motivate his decision by the help of diagrams and specific numbers. This enables to compare with traditional sonographic approach, better motivate phonoscopic examination, accelerate the investigation and reduce requirements for investigative and comparative speech records. The investigation becomes independent from the speech text recorded in the investigative and comparative speech records.

The system is realised in the form of software package (SIVE) and can work with any type of IBM PC computer, supplied with professional sound input/output card.