ACADEMICIA
An International Multidisciplinary Research Journal
(Double Blind Refereed & Peer Reviewed Journal)

DOI: 10.5958/2249-7137.2021.01656.6

SOME ASPECTS OF FORENSIC PERSONAL VOICE AND SPEECH IDENTIFICATION

Lyudmila Yuriyevna Yugai*

*Doctor of Philosophy (Ph.D.) in Juridical Science,
Academy of the Ministry of Internal Affairs of the Republic of UZBEKISTAN
Email id: yugai.lyudmila@mail.ru

ABSTRACT

The article is devoted to the phonoscopic research issue. The historical review of the development of forensic voice and speech identification, starting with the first cases of using "aural-perceptual" testimony of a witness to the modern period of the introduction of artificial intelligence and neural networks, is covered in the article. To date, active developments are being carried out and automated systems are being introduced into practice for conducting phonoscopic studies. Modern problems of personal voice identification on digital phonograms are also covered. Appropriate recommendations and suggestions are proposed based on the analysis of specialized literature and expert practice.

KEYWORDS: *Phonoscope, Phonogram Authenticity, Artificial Intelligence, Sound, Personality Identification*

INTRODUCTION

The identification issues using biometric technologies are of particular relevance. One of the most promising areas of biometric personality identification is the use of phonoscopic face recognition systems based on a voice signal. The widespread use of technical devices for fixing audio information, their widespread use and availability entail an increase in the use of phono evidence in legal proceedings, as well as the need for improving modern methods of identification by voice signal and speech.

Phonoscopy is a relatively new branch of science. "Phonoscopy" comes from the Greek "phone" – sound and "logos" – teaching.

Phonoscopic studies cover many issues that arise in the study of any signals that have a sound nature or are closely related to sound. However, the main tasks of forensic phonoscopic research remain the search for signs of installation (or technical research) and the identification of a

person by oral speech (identification research) [1, P. 151-153]. The identification of a person by his voice is of particular relevance today.

Research on personality identification by speech signal began in the middle of the twentieth century, when the tape recorder began to be used first for military purposes, and then in other spheres of human information activity.

Main part. The case of the first recorded mention of the admissibility of using "aural-perceptual" (i.e. by ear) testimony was rooted in medieval England, where in 1660 the case of a certain William Hewlett was considered, who was identified by his voice [2].

Starting in 1907, some courts in the United States and the United Kingdom allowed subjective identification by voice-based on testimony as a source of evidence. At the same time, the development of scientifically grounded methods of identifying a speaker by voice is noted by Soviet scientists, whose work in the 40s of the XX century was first described in the novel by A.I. Solzhenitsyn "In the first circle" [3, P. 181-195].

Phonoscopy was used for the first time in the USSR, in the summer of 1949, to promptly identify the caller and prove his guilt during the investigation of a criminal case initiated upon the fact of a telephone call to the U.S. Embassy regarding the sale of information about Soviet foreign agents [4, P. 110-145].

The issues of personal identification by speech signal began to be dealt with in the middle of the twentieth century, the means of sound recording began to be used first for military purposes, and then in all spheres of human information activity.

The first inventor of the mechanical sound recording device is considered to be the French poet and inventor Charles Cros, who in April of 1877 presented an invention, which he called the "phonograph", to the French Academy of Sciences. The device made it possible to mechanically record sound on a rotating glass disk covered with soot, and then transfer the soundtracks to a photosensitive chrome plate using a photo method. In December 1877, the American inventor Thomas Edison demonstrated his phonograph in action, in which sound was recorded on a cylinder wrapped in tin foil.

In 1887, the German inventor Emile Berliner received a patent for a phonograph, the principle of which was based on the developments of Charles Cros. Sound is recorded on a zinc disc covered with a thin layer of wax, and replication is carried out by chemical etching.

Due to the very laborious sound recording technology, the above inventions of phonographs have not been used in forensic identification. For the same reason, optical phonograms were practically not used, which gave sound to silent films. Magnetic phonograms served as a powerful impetus in the development of forensic phonoscopy. And it should be noted that not only magnetic tapes have met (and are met) in expert practice. It should be noted that 20 years ago, magnetic records on steel wires were encountered in expert practice.

The first studies of speech signals recorded using mechanical methods – phonographs, and after electromagnetic ones – tape recorders were based on auditory or aural-perceptual analysis (by ear). The above was based on the subjective perception of the sound signal and, therefore, caused distrust on the part of the judiciary.

An important stage in the development of phonoscopic research was the invention of a device that made it possible to visualize a speech signal in a form reflecting both articulation and the speaker's voice. Many scientists in the period from the 20s to the 30s of the last century presented various forms of visualization of speech signals. The device was developed by the end of the Second World War, after a series of experiments that were carried out in closed mode. In 1945, in the weekly journal Science, Ralph Potter published an article "Visible Sound" [5], in which he first described a new device developed by employees of Bell Telephone Laboratories. At first, the device was called the "sonic spectrograph", later another name was fixed – "sonograph" [6, P. 16].

The advent of the sonograph, which made it possible to see both articulatory and vocal features of the speaker at the same time, already by 1944 pushed specialists to research the identification of a person by a speech signal. In this connection, the term "voiceprint" (by analogy with the term "fingerprint") appeared, which meant just one of the types of narrow-band sonograms of speech signals.

After unsuccessful experiments that were undertaken during the war, to address the issue of personal speech identification, a specialist began to research the nature of speech production. In the late 1950s, G. Fant, in close collaboration with the Massachusetts Institute of Technology Acoustics Laboratory, carried out fundamental research, based on which the acoustic theory of speech production was created [7]. The transition to acoustic methods and the creation of a solid theoretical base have revived the study of individual signs of speech.

In the early 60s, the Soviet scientist G.S. Ramishvili researched the comprehensive study of individual characteristics of speech and voice. The acoustic parameters of speech signals, reflecting certain individual features of speech production, were established, their information content and the reliability of determination were assessed [8].

At the same time, in the United States, Lawrence Kerst is researching the development of a method for identifying a person using sonograms of a speech signal. Sonograms were used as evidence in the proceedings. In 1972, criteria for assessing the reliability of recognition were developed, supported by the Supreme Court. Even though identification studies using sonograms have been criticized many times, this technique has nevertheless found its application in the forensic practice of some countries. Based on the results of the analysis of this type of forensic research, the International Association for Identification (IAI) put forward the requirements for the special training of forensic experts in the theoretical and practical part of identification with the obligatory passing of an exam to obtain a certificate for the right to produce phonoscopic examinations. In 1975-1977, after several judicial errors, this technique was severely criticized by leading scientists and specialists in speech technology.

In the mid-1980s, research in speech technology intensified at the FBI. In the scientific journal "The Journale of the Acoustic Society of America" there was an article containing an overview of the FBI's achievements in identifying individuals by speech. It was noted in the article that errors of identification of the first kind (excluding the identity of voices of the same person) amount to 0.53%, and errors of the second kind (identification of votes of different persons) – 0.31%.

From the point of view of the validity of the conclusions of the study, these indicators are considered high reliability of identification. However, these results raised doubts among leading experts in speech technologies, after which they published an article in the same journal, calling into question the previously published results. The critical article was scientifically and practically grounded enough that the authors from the FBI had to publish a response article in the aforementioned journal explaining what they meant. It turned out that not real indicators of the reliability of the entire identification system were presented, but indicators concerning "obvious cases". In other cases, (which turned out to be the majority), experts usually gave answers of the type: "the material is not suitable for research" or "it is not possible to answer the question".

In the Soviet Union, G.L. Granovsky was the first to propose the method of spectrography in forensic science, calling it "vocalography". The first studies of speech technologies were carried out in the scientific departments of the People's Commissariat for Internal Affairs (NKVD) and the Ministry for the Preservation of Public Order (MOOP) of the USSR and were of a closed nature, which made it difficult to use them procedurally by specialists and criminologists [3, P. 181-195].

In 1962-1963, E.I. Abalmazov and A.I. Kugushev developed a method for identifying a person using phonograms of free speech. In this work, the main attention was paid to the theoretical study of the possibility of implementing the so-called "optimal identifiers", carrying out the identification of the speaker based on the analysis of the physical characteristics of the speech signal.

In the 70s of the 20th century, methods were developed with the help of which it became possible to present sound vibrations in the form of a visual image [9, P. 28-31]. Subsequent works were devoted to the automated process of identifying a person by voice.

Results and Discussions. Relatively recently, compared to other biometric systems, automated voice identification systems were in many ways inferior in terms of identification accuracy, model size and other parameters. Yu.N. Matveyev notes that in recent years, voice recognition systems have made relatively significant advances, which can improve efficiency, especially when creating multimodal systems that contain several types of biometric parameters [10, P. 46-61].

Today, active developments are being carried out and automated systems are being introduced into practice for conducting phonoscopic studies, which include voice identification systems SIS (Center for Speech Technologies in Saint Petersburg), AIS Dialect (Russia) with a complex for processing and analyzing speech signals CSL (KAY company, the United States) [11, P. 6-15]. Based on this development, it became possible to carry out forensic identification of a person by recorded negotiations, which is especially important in the disclosure and investigation of crimes related to kidnapping, extortion, false reports about the preparation and commission of terrorist acts [12, P. 387-394].

In foreign banking systems, multimodal biometric systems are actively introduced, among which voice technologies are used [13].

Phonoscopic studies cover many issues that arise in the study of audio signals that are sound or closely related to sound. However, the main tasks of forensic phonoscopic research remain the

search for signs of installation (or technical research) and the identification of a person by oral speech (identification research) [1, P. 151-153].

In the period of digital transformation of society, the possibility of violating the authenticity of a speech event can be identified as a separate problem. These include distortion of the speech signal, soundtrack editing, synthesis of voice and speech, imitation of a speech event, various characteristics of the speaker's voice and speech, masking and modification of voice and speech [14, P. 62-71]. The above technologies can be used for criminal purposes and certainly complicate the formulation of a conclusion by a forensic expert in the production of phonoscopic examinations and research.

In connection with the above, it becomes necessary to identify, suppress and prevent falsification of phonograms, which will be presented as evidence. A phonogram that adequately reflects an acoustic event that took place without making any changes is usually called authentic. An authentic phonogram must be continuous, recorded from the source of a sound, the recording of speech and other signals from various sources must be carried out simultaneously, the time and linear sequence of signal recording must be observed, the speech event must be fully displayed.

The authenticity of the phonogram is carried out by establishing the degree of correspondence of the displayed sound trace to the occurring acoustic event, taking into account those natural distortions that are introduced into the signal when it is generated by the sound source, transmitted through communication channels, and recorded on a material carrier.

Distortion of an audio recording means making changes to the phonogram. Modification is a deliberate or natural impact on a phonogram, altering or distorting its properties. Imitation is an imitation of the voice and speech of another person, by reproducing timbre, articulation, intonation, etc. Voice masking is the concealment of the speaker's identity by altering his linguistic and acoustic characteristics.

In modern conditions of the development of universal digitalization, it becomes possible to deliberately change the information recorded on the phonogram without leaving any traces of the manipulations performed, which certainly complicates the assessment of the reliability of phonograms as evidence.

It should be noted that digital sound recording expands the scope of its application due to ease of use, rewriting, duplication of copies, storage, small dimensions, absence of noise typical of analog media, etc.

Digital sound recording is used to record conversations using dictaphones, cell phones, video cameras, as well as by law enforcement agencies for documenting procedural actions and operational-search activities, including monitoring and listening to telephone conversations. A digital recording of voice signals, which is enshrined by procedural legislation, is recognized as material evidence or other documents and is attached to the materials of the criminal case.

The digital form of a phonogram often causes mistrust in the reliability of phonograms as evidence. Yu.L. Belykh believes that at present, experts from state expert departments do not have proper scientifically grounded methods and means of identifying signs of electronic installation [15, P. 24].

Opponents of digital evidence believe that analog phonograms are more credible as evidence, because when they are falsified, there are "obvious traces" that are easily detected by expert methods. However, experts note that analog sound recording is not a guarantee of the reliability of the information. Duplication on an analog medium is often deliberately used to hide traces of manipulations with the original digital phonogram (it happens that an investigator or a court under the guise of an original is provided with an analog phonogram obtained by recording an edited digital phonogram through an acoustic channel, reproduced through acoustic systems) [16, P. 15-26].

The high level of development of information technology makes it possible to synthesize voice signals using neural networks and artificial intelligence. With the help of them, any typed text is converted into a voice signal. These programs help users of software products and the Internet as voice reminders, reading lectures, recording audiobooks, sounding telephone robots, etc. These technologies allow simulating any speech event that will contain false information about any person or fact, as well as hide identity as the author of the text.

Certain programs allow changing voice using computer technology. With the help of them, according to the set parameters, you can correct some characteristics of the voice, the frequency of the main tone. In modern phones, the network has applications such as Voice Recorder and Changer, Voice Changer and Disguise, Celebrity Voice Cloning, etc.

One of the options for identifying individual signs of speech synthesis are methods of auditory perception and acoustic-instrumental methods. The expert should pay attention to the peculiarities of intonation structures and the presence of artifacts in the controversial phonogram. Artifacts are an effect arising during the gluing of phonemes, which is audibly perceived as a sharp change in the amplitude and/or frequency of sound, associated with insufficient requirements for the smoothness of the stitching [14, P. 69].

However, experts emphasize the difficulty of verifying and assessing the reliability of information captured in an oral speech on a digital phonogram, taking into account the possibility of editing a phonogram without displaying visible traces of manipulations [17, P. 214].

To solve problems of identification by voice, complex integral scientific and expert studies are proposed, combining the achievements of various fields of scientific activity (mathematics, physics, cybernetics, jurisprudence, linguistics). A.Sh. Kaganov notes that the main subjects in solving the problem of forensic speaker recognition are expert researchers who have undergone specialized training. As a rule, these are acoustic engineers and applied linguists who have devoted themselves to the issues of processing and research of speech signals, specialists in phoniatrics, psychology of sound speech, etc. [18, P. 246-250]

CONCLUSIONS

Identification by speech signal and voice is often the only biometric parameter that allows one to establish an identity over the phone when reporting a planned terrorist act, extortion of a reward, etc. In a relatively short period, phonoscopic studies have come a long way from the creation of the first device for fixing a sound signal to the analysis of the artificial synthesis of speech signals using neural networks. During the introduction of digital technologies into all spheres of

public life, it becomes necessary to develop new complex methodological approaches to personal identification by speech signal and voice.

REFERENCES

1. *Vasiliev A.V.* Current issues of identification studies of the human voice and speech // Bulletin of the Academy of the Investigative Committee of the Russian Federation. – 2019. – No. 1. – P. 151-153.
2. *Hollien H.* The Acoustics of Crime: The New Science of Forensic Phonetics. – New York: Plenum Press, A Division of Plenum Publishing Corporation, 1990.
3. *Galyashina E.I. T*he history of forensic phonoscopic examination // Bulletin of the University named after O.E. Kutafina. – 2014. – No. 3. – P. 181-195.
4. *Kopelev L.* Ease my sorrows. Memoirs. – M.: Slovo, 1991. – P. 110-145.
5. *Potter R.K.* Visible Patterns of Sound // Science 1945. No. 2654. November 9.
6. *Kaganov A.Sh.* Forensic examination of sound recordings. Monograph. – M.: Jurlitinform, 2005. P. 16 (272 p)
7. *Fant G.* Acoustic Theory of Speech Production / Translated from English L.A. Varshavsky, V. I. Medvedev / Ed. by V.S. Grigorieva. – M.: Nauka, 1964.
8. *Ramishvili G.S.* Speech signal and voice personality. – Tbilisi: Mitsniereba, 1976. – 183 p.
9. *Barkovskaya E.G.* Historical background and practical needs for the integration of biometric and forensic knowledge. – Philosophy of law. – 2010. - No. 6. – P. 28-31.
10. *Matveev Yu.N.* Technologies of biometric personal identification by voice and other modalities. Bulletin of the Bauman Moscow State Technical University. – 2012. – No. 3. – P. 46-61.
11. *Galyashina E.I.* Theoretical and applied foundations of forensic phonoscopic examination: Abstract. dis. ... doctor of juridical sciences. Voronezh, 2002. – P. 6-15.
12. Expert. Guide for experts of internal affairs bodies / Ed. by T.V. Averyanova, V.F. Statkusa. – M., 2003. – P. 387-394.
13. *Chernyshova E.* Biometrics in banks: what is it, why and what it will lead to [Electronic resource] – Access mode: https://trends.rbc.ru/trends/industry/5fd3ac6a9a79475333bfcb4f. – Date of access: 20.02.2021.
14. *Lebedeva A.K.* Problems of the production of forensic phonoscopic examination in the light of the development of digital technologies // Bulletin of the Kutafin Moscow State Law University. – 2020. – No. 6. – P. 62-71.
15. *Belykh Yu.L.* Forensic investigation of audio documents (procedural and forensic aspect) // Forensic expert. – 2006. – No. 4. – P. 24.
16. *Galyashina E.I.* Forensic phonoscopic examination: problems of diagnosing the authenticity of phonograms // Bulletin of the Kutafin Moscow State Law University. – 2014. – No. 3. – P. 15-26.

17. *Galyashina E.I.* Forensic speech: Textbook. – M., 2020. – P.214.

18. *Kaganov A.Sh.* Forensic personal voice and speech identification as an integral scientific and expert study // Philological Sciences. Theory and practice. – 2019. – Volume 12. – Issue 6. – P. 246-250.