

SOME ASPECTS OF CRIMINAL INVESTIGATION OF PHONOGRAPHS FOR SPEAKER IDENTIFICATION

Brendel O.

*National Scientific Center "Hon. Prof. M. S. Bokarius Forensic Science Institute" of the
Ministry of Justice of Ukraine,
Kharkiv, Ukraine*

The article pays attention to issues that are brought to the expert's decision in the forensic investigation of oral speech of persons for the purpose of their identification. Objective and subjective research methods are analyzed, including linguistic and instrumental analysis, auditory, visual research methods using software and technical tools. The specific features of the variability of a person's voice and speech are considered, and the influence of external environmental and other factors on the formation of recordings is analyzed. It is highlighted that there are various possibilities for conducting identification research, both using instrumental and linguistic methods. An analysis of speech diagnostics was carried out, which according to research results can provide information about the personality of the unknown, situational, emotional, physiological, psychological, physical, and other characteristics of the speaker.

Keywords: *analysis of audio and video recordings, expert conclusions, identification of individuals, physical parameters of voice, linguistic features of spoken language, identificatory and diagnostic studies.*

The problem of forensic research of phonograms with oral speech for the purpose of identifying the speaker by voice and speech arose at the beginning of the 20th century, and has not lost its relevance until now and is discussed in special legal and technical literature. Due to specific features, the task of speaker identification is quite complex, the complexity of the problem is due, first of all, to the very object of research - the speech phonogram, which is influenced by a wide variety of factors, which are both speech and non-speech in nature and distort the natural picture of what is recorded on phonogram of the speaker's voice. What is voice? The voice is more than just a string of sounds. Voices are inherently complex. In addition to the intended linguistic message, they carry a large amount of information about the speaker: for example, about his gender, emotional state or state of health, etc. Some of this information clearly has potential forensic value. The voice is an extremely complex object, a significant part of this complexity lies in its natural variability and the significant influence of external environmental factors on the characteristics of the voice, including in the process of recording on a certain information medium. When comparing voices, you need to know what exactly is being compared. It should be noted that in addition to the content side, the recorded speech message may contain information about personal or situational moments related to the speaker, for example, his emotional or psychological state, socio-biological features of the speaker (origin, profession, intelligence, etc.), physical characteristics (age, gender, external features) and many other information related to the surrounding sound environment.

A large number of scientific works by both Soviet and foreign specialists, such as: Kaganov O. V. Sh. [1], Zharikov Y. F., Mohnev S. P. [2], Saltevsky M. V., Sharshunsky V. L., Lozhkevich O. O.,

Correspondence address: Brendel Olha Ivanivna, Head of the video and audio recording research sector of the engineering, technical and military research laboratory of the National Scientific Center "Hon. Prof. M. S. Bokarius Forensic Science Institute" of the Ministry of Justice of Ukraine. Kharkiv, Ukraine, 8-A, Zolochivska St., Kharkiv, UA-61177 (057) 372-28-29, e-mail: nncise@nncise.org.ua, olgabrenda@ukr.net

Snetkov V. O., Chivanov V. O. [3], Ramishvili G. S., Chikoidze G. B. [4], Flanagan J. L. [5], Fant G. [6] and others. One of the first methods of identifying the identity of the speaker was the «Voiceprint» method, which was carried out using the spectral analysis of spoken words and their presentation in the form of contour spectrograms. This method found limited use in judicial and investigative practice, since the positive results of its results did not meet the expectations that were placed on it. Despite certain objections, this method was adopted in many states and federal legal bodies of the United States, Canada, and was being prepared for distribution in a number of Western European countries, Israel, Japan, and South Africa [7]. A detailed and qualified review of the works related to the identification studies of a person by voice and speech have been published at various times on the pages of authoritative journals JASA and Proceedings of the IEEE [8-9].

Subsequently, in various scientific centers of the world, forensic scientists continued the development of various approaches and constructive description of individual features of voices, based on a number of stationary and dynamic properties of voices, which led to the development of automated speaker identification systems. It should be noted that an important component of the forensic examination of phonograms for the purpose of identifying the identity of the speaker is the linguistic analysis of speech features. In these directions, the development of research is entrusted to the work of such scientists as Granovsky G. L., Regbun E. K. [10], Chikoidze G. B. [11], Galyashina O. I. [12] and others. Modern research in these areas is carried out by the following scientists: Hollin H. [13], Neustein A., Patil H. A. [14], Solovyov V., Rybalskiy O., Zhuravel V., Shablya A., Tymko Ye. [15] and others.

The purpose of the article is to highlight the peculiarities of the identification studies of the announcer by voice and speech. The work considers both the variable features of the voice and speech, which must be taken into account by the expert when conducting the examination of video and audio recordings, as well as the factors influencing the voice characteristics of the elements of the external environment, the influence of audio recording equipment on the quality of the original phonogram, etc.

From the point of view of the theory of forensic identification, important factors are the selection of essential and non-essential, general and individual features, the totality of which can be considered sufficient to establish the identity of a person, in this case - the speaker, by using the properties of his voice and speech reflected on the phonogram. The current state of expertise in video and audio recording involves the concept of the expert's use of systems of objective and subjective methods. Subjective methods usually include auditory, visual and linguistic analyses. Objective methods include methods of automated identification and verification of the speaker based on the physical parameters of the voice using modern capabilities of software and technical means. That is, according to the procedure of expert analysis, a list of hypotheses is put forward, which are checked by an expert, a set of individual characteristics of the speaker is formed, which are resistant to the influence of the factors of a specific sound environment, as well as additional information about the means of recording, the environment, recording media and the person of the speaker. Auditory, visual, technical, trasological, linguistic and other methods of phonogram analysis are involved in the system of analysis methods. It should be noted that, in most cases, the conditions of production of the compared records can differ dramatically. Thus, the task of identifying a complex of individualizing features is one of the main stages of examination of video and audio recordings. During the expert examination, both the variability of the direct speech signal and the influence of factors of the external sound environment, as well as the characteristics of the recording equipment, should be taken into account, since the crime event must always be determined as an event that exists in a specific environment determined by the characteristics of space and time.

The phonogram of speech, as a carrier of information about the personality of the speaker, is a complex object of interconnected complex multifactor processes of a speech and non-speech nature. The variability of speech signal parameters consists of various components of the speech production process. These components can be factors related to the anatomical and physiological features of the articulator organs of the speech tract, as well as situational programs of the central nervous system, as a result of which speech can be natural, pathological, intentionally distorted, emotionally colored, etc. All these factors require special attention, since any phonogram is the result of recording a certain speech message on the information carrier, as well as various accompanying factors of a non-speech nature.

The variability of the speech signal in the conditions of a person's natural speech activity includes the pitch of the voice, the frequency of the main tone, the pace of speech, intonation coloring, voice fatigue, as well as the connection with the influence of such factors as a cold of the vocal cords, a hangover, the influence of other psychotropic agents etc. – i.e., the physical picture of the speech signal, which is related to the speech-forming tract of a person. The variability of the speech signal, which is caused by pathological changes in the structure of the speech tract, includes not only stationary pathological changes in the structure of certain parts of the speech-forming tract, but also temporary processes that are associated with a specific human condition, for example, voice breaking during puberty, changes in the voice in old age or in the climacteric period in women. It should be noted that the limits of physiological changes are individual and depend on each specific case. Random pathological voice changes that occur under the influence of subjective or objective factors should also include changes caused by colds, hangovers, the influence of narcotics or medicines, the influence of external disturbances and obstacles, vibrations, extreme stressful events, as well as the influence of the voice in the process of pronouncing simultaneous smoking, eating, participating in a group conversation and much more. The list of such factors is essential in the study of individual features of the voice and the possible assessment of the limits of their changes. Congenital physical defects of the speech tract primarily emphasize the individuality of the speaker and are an unmistakable sign of this voice. Thus, in contrast to situational factors, stationary pathological variations of voice characteristics should be considered by the expert as the most significant signs for identification.

Among the factors that consciously or subconsciously affect the variability of the main vocal characteristics of speech, the emotional state of the speaker occupies a significant place. By the color of the voice, intonation, and the content of conversations, it is possible to establish the emotional state of a person at the time of the utterances: calm, even, irritable, depressed, joyful. According to C. Izard [16], fundamental emotions include: interest – excitement; joy; surprise; grief – suffering; anger; abomination; contempt; fear; shame; fault. As for the features of the speech signal, anger, anxiety, fear, depression, persuasion have a particularly strong effect on it. So, for example, the voice can be apathetic, calm, or vice versa, excited, strong-willed, etc., that is, when the emotional state of the speaker changes, the intelligibility of speech, its fluency, tempo, intensity, frequency of the main tone changes in direct proportion. Analyzing the listed factors characteristic of certain types of emotions, when conducting identification studies, those properties of the speech signal that are significant for the greatest deviations from a neutral psychological state: excited or inhibited should be singled out.

The next variable factor that affects the characteristics of a certain person's voice is whispering. When transitioning from normal speech to whispering, changes occur in the physiological picture of the speech signal, due to which identification of whispered speech by instrumental methods becomes almost impossible. Qualitative characteristics of the acoustic parameters of speech acquire significant changes in all indicators, from intelligibility, recognizability of voices, to deformation of the formant structure, absence of fundamental tone parameters, weakening of the frequency range, etc. In this case, it should be noted that a number of prosodic (mainly temporal) characteristics of pronunciation, such as rhythm, tempo, pause, are preserved, which are used in linguistic identification studies. In addition, the informational value of the broadcast content is completely preserved. That is, in whispering, despite the fact that deep transformations occur with the acoustic parameters, in some favorable cases a certain number of features essential for linguistic research methods are preserved, which allows not only a separate group differentiation of the speaker, but also in some cases, identification persons.

Undoubtedly, an important factor in the study of speech is the possible presence of external noise, a concomitant process of the speech activity of the speaker, which affects not only the intelligibility of speech, but also has a masking effect, distorts not only the threshold of audibility,

but also the objective characteristics of the voice. It is generally known that in a noisy environment, the speech activity of the speaker increases primarily by increasing the voice level. As a rule, the pitch of the voice increases and, accordingly, the frequency of the main tone and the energy of the speech spectrum (shifts to the high frequencies). The listed processes are closely related to each other and occur involuntarily in the speaker.

Among the various factors that affect the variability of the output speech signal, on the one hand, the factors of the sound environment should be noted, and on the other hand, the variability of the characteristics of the technical means, which are used to fix the speech message on a certain information medium. Sharshunskyi V. L. [17] considered the sound environment as a texture of the surrounding environment, which consists of a complex of different sounds. At the same time, the peculiarities of such a set of sounds will characterize the acoustic conditions in which the recorded phonogram was recorded. Issues of forensic diagnostics of the sound environment require separate consideration, since the sound environment of a crime event is a single system of sound traces of natural and artificial sources, subject-communicative activity of the criminal, tools, instruments and technical means used by him, united by the unity of purpose. Therefore, when conducting speaker identification studies, the characteristics of the external environment and their influence on the variability of voice characteristics must be taken into account. For example, to reduce the factors directly affecting the recording equipment on the quality of the output signal, to preserve the naturalness of a person's speech, you should use high-quality recording equipment. The first factor influencing the environment on a natural signal is the acoustic properties of the room in which the recording is conducted. The main characteristic of the room, which most affects the timbre of the voice, is the loudness of the sound and the feeling of space - reverberation - the process of repeated reflection of the acoustic signal from the surfaces of the room and the objects located in it. One of the effects of reverberation is an echo, which is perceived by ear as an independent, time-dispersed signal. The acoustic properties of the room can significantly affect the spectral pattern of the output speech signal, mainly deforming its high-frequency region [4, p. 60]. Irreversible distortions in the speech signal can also introduce the sounds of an open acoustic environment - the voices of animals or birds; sounds of natural phenomena - wind, thunder, rain, etc.; sounds of municipal origin - noise of the street, transport, production, etc. The list of such sounds is quite wide, most of such accompanying sound signals cannot be filtered without significant loss of useful individual speech signal, since the physical characteristics of such interference are close to the speech signal.

Thus, in the process of fixing the phonogram of spoken speech on a certain medium, all components of this process take part and influence the characteristics of the voice: the natural variability of the speech signal, sound sources, the environment, features of sound recording devices, etc. As a result, speech features recorded on the original phonogram may be distorted. Depending on the degree of their distortion, a phonogram with a recording of spoken speech can be satisfactory for conducting identification studies - suitable for correction, or unsatisfactory, due to the presence of distortions that cannot be corrected. In the latter case, conducting an examination for the purpose of identification research of the speaker will be impossible.

For a successful examination of a video and audio recording for the purpose of speaker identification, not only the researched (disputed) recording, but also a voice recording sample (standard) created for comparative research has a significant impact. Special attention should be paid to the process of selecting voice and oral speech samples, since the possibility to take into

account and compare the entire set of individual characteristics of the speaker during expert research, to check and compare them, depends on their quality. First of all, the person who conducts the selection of the reference recording of the voice needs to carefully study the researched recording. When preparing comparative material, all the available variable factors in the disputed record should be taken into account. In order to create a high-quality reference recording of the person being tested, it is desirable and necessary, in addition to using high-quality equipment, to create an exemplary recording that will be as close as possible to the disputed recording, both in terms of acoustic conditions and the emotional state of the speaker. Of course, when recording the reference speech of a certain person, it is necessary to exclude as much as possible parallel conversations, noises, external acoustic signals, even if they are present on the recorded recording. In the process of selecting speech samples and preparing materials for examination, the following basic rules should be observed: the selection of voice and speech samples cannot be approached formally; the person whose speech samples are being selected should be given the opportunity to get used to the environment, to establish psychological contact with him; it is necessary to remember that the main purpose of obtaining samples is to record the oral speech of the person from whom they are taken, so there is no need to interrupt her, speak at the same time as her [18].

In order to carry out a full-fledged identification study of the speaker's personality, it is necessary to select features (preferably of an acoustic and linguistic nature) from the disputed and reference records, and then compare them with each other. The decision-making process in the task of forensic identification of the person of the speaker based on phonograms of oral speech is very responsible, as it requires the use of various means of examination (systems of objective and subjective methods), which are based on the peculiarities of the nature of the voice and are manifested in its various characteristics. Basically, all of them are based on various approaches to studying the nature of a person's voice and speech, reflect its various properties and are easily amenable to measurement and instrumental processing. For example, auditory analysis is based on the broad capabilities of auditory and visual perception of the expert, and automated instrumental methods rely mainly on the spectral, temporal, and amplitude characteristics of the speech signal. Accordingly, different methods use different decision-making mechanisms. Only linguistic analysis is significantly different from others, since the foundations of such analysis are based on the peculiarities of the construction of speech, this method practically does not overlap with any of the other instrumental methods of identifying a person. Undoubtedly, the complex application of various methods of forensic identification of the speaker based on phonograms of oral speech will provide a more significant and complementary generalized result than the use of individual instrumental or linguistic research methods. But at the same time, from the point of view of the significance of the obtained results, linguistic and instrumental research methods are independent and autonomous.

While analyzing videos, especially if the audio part is accompanied by interference, additional information is provided by video channel and if there is significant noise in the audio recording: visual information (articulation) can be the only source to determine the event content.

Traditionally, the task of forensic examination of video and sound recording is the identification research on speaker by the physical and/or linguistic characteristics of the voice. Naturally, the issue arises as to the possibility of using visual information either as a stand-alone research object (video as a source of information about the sound part of recorded events) or as an auxiliary, additional to the auditory analysis of a particular message. While investigation,

investigative bodies increasingly need to obtain information about the spoken text only from visual and auditory information (lip reading) to determine the conversation content [19].

As for possibility of using visual information (lip reading) as an independent research object within forensic video and audio analysis in order to determine the content of audio information, it has many disadvantages. In view of the above, to use visual information as an independent research object is impractical according to the author, because the quantity and quality of text reproduction obtained by lip reading do not guarantee the result accuracy, because they depend on many factors including skills a person who interprets the image (its subjective perception) and can testify only to the conversation topic (it is impossible to obtain a literal full textual reproduction). However, the use of lip reading skills as an additional auxiliary tool in compiling the conversation content in combination with regular listening, will significantly improve the results of this work[20].

The process of speech production is a complex and multifactor process, in which both the anatomical speech-forming apparatus of a person, which creates its own individual picture of the speech signal characteristic of a certain person, and the speech behavior of the speaker, skills, individual experience, mental state, temperament, intelligence, etc., are used. Variability of speech characteristics simultaneously constitutes both individual and variable features of the voice. Estimating the limits of such variations, their ratio, is the main priority task that is solved by an expert in the process of forensic identification of a speaker. The use of several approaches in the research allows the expert to determine various characteristics of the speaker, which complement each other in the complex analysis of the phonogram.

Thus, on the one hand, the variability of the factors affecting the characteristics of the voice recorded in the sound recording prevent the conduct of an instant full-fledged identification study, and on the other hand, indicate ways to overcome obstacles to conducting a qualitative identification study of the speaker. Expert research should be conducted by experienced specialists. Conducting identification procedures should be preceded by a thorough examination of the materials received for research. The conclusion of an expert study should summarize individual decisions, taking into account the competence of each individual approach or method in the conditions of a specific study.

References

1. Каганов А.Ш. Кибернетические методы идентификации по голосу. Вопросы судебно-фоноскопической экспертизы: Материалы научно-практического семинара. Тбилиси, 1988.
2. Жариков Ю.Ф., Мохнев С.П. Идентификация дикторов по параметрам спектральных переходов речевых сигналов. Автоматическое распознавание слуховых образов. Тбилиси, 1978. С. 34–36.
4. Салтевский М. В., Жариков Ю. Ф. Современные возможности исследования голоса человека в криминалистической практике. Киев. НИ и РИО КВШ МВД. 1982. 46 с.
5. Рамишвили Г.С., Чикоидзе Г.Б. Криминалистическое исследование фонограмм речи и идентификация личности говорящего. Тбилиси: Мецниереба, 1991. 265 С.
6. Flanagan J. L. Speech Analysis Synthesis and Perception. Berlin. 1965.
7. Fant G. Acoustic Theory of Speech Production. The Hague. 1960.
8. Hollien, H. Status Report on «Voiceprint» Identification in the United States, *Proceed., Internat. Conf., Crime Countermeas.* Oxford, UK, 1977. 9-20.
9. Bolt R.H., Cooper F.S., David E.E. et al. Speaker identification by speech spectrograms: some further observations. *JASA*, 1973. V. 54, № 2. p.2.
10. Розенберг А. Автоматическая верификация диктора. Обзор. *ТИИЭР*, 1976, т. 64, № 4.
11. Грановский Г.Л., Ребгун Э.К. О теоретических основах судебной фоноскопии. Современные проблемы экспертизы и пути повышения эффективности деятельности судебно-экспертных учреждений в борьбе с преступностью: Тез. респ. научн. конф. Киев : КНИИСЭ, 1983. С. 222–224.

12. Чикоидзе Г.Б. Лингвистические аспекты проблемы распознавания личности. Сб. Языковые процессоры и распознавание речи. Тбилиси: Мецниереба, 1985.
13. Галяшина Е. И. Актуальные проблемы экспертизы цифровых фонограмм. *Теорія та практика судової експертизи і криміналістики* : зб. наук. праць. Вип. 8. Харків. Право. 2008. 8. С. 248-256.
14. Hollien, H. The Future of Speaker Identification: A Model, Proc. 13th Int. Congr. Phonetic Sci., Stockholm, 3. 1995. P. 138-145.
15. A. Neustein, H. A. Patil (eds.), Forensic Speaker Recognition, DOI 10.1007/978-1-4614-0263-3, Springer Science+Business Media, LLC, 2012.
16. Solovyov V., Rybalskiy O., Zhuravel V., Shablya A., Tymko Ye. Building of a Speaker's Identification System Based on Deep Learning Neural Networks, pp. 107-116. Міжнародна науково-практична конференція «Інтелектуальні системи та інформаційні технології»: Матеріали конференції / Одеський державний екологічний університет. Одеса, 2021, 360 с. **Scopus**
17. Izard, Carroll E.; Kagan, Jerome; Zajonc, Robert B. Emotions, Cognition, and Behavior (англ.). Cambridge: Cambridge University Press, 1984. P. 17. ISBN 0-521-25601-1
18. Brendel O. (2021). Modern possibilities of diagnostic research in the field of forensic video and audio analysis. *Theory and Practice of Forensic Science and Criminalistics*. 23rd Issue, p. 203—211. DOI: 10.32353/khrife.1.2021.15.
19. Брендель О. І. Організаційні проблеми під час призначення судової експертизи відео-, звукозапису. *Науковий вісник Дніпропетровського державного університету внутрішніх справ: Науковий журнал*. 2020. Вип. 4 (107). С. 284-290.
20. Brendel O. I. Some aspects of visual and auditory perception of oral language in forensic video and audio analysis. *Theory and Practice of Forensic Science and Criminalistics* : Collection of Scientific Papers. Kharkiv: Pravo Publishing House, 2020. 21st Issue, p. 349—358. DOI: <https://doi.org/10.32353/khrife.1.2020.24>.

НЕКОТОРЫЕ АСПЕКТЫ УГОЛОВНОГО РАССЛЕДОВАНИЯ ФОНОГРАФИИ С ЦЕЛЬЮ ИДЕНТИФИКАЦИИ ГОВОРЯЩЕГО

Брендель О.И.

В статье освещены вопросы, стоящие перед экспертом при идентификации лиц в ходе судебного расследования устной речи. Анализируются объективные и субъективные методы исследования, включая лингвистический и инструментальный анализ, слуховые и визуальные методы исследования с использованием программного обеспечения и технического оборудования. Рассматриваются специфические особенности вариативности голоса и речи человека, а также влияние внешней среды и других факторов на формирование записей. Подчеркивается, что существуют различные возможности проведения идентификационных исследований как с использованием инструментальных, так и лингвистических методов. Проведен анализ речевой диагностики, который, согласно результатам исследований, может предоставить информацию о личности неизвестного, ситуационных, эмоциональных, физиологических, психологических, физических и других характеристиках говорящего.

Ключевые слова: анализ аудио- и видеозаписей, заключение судебного эксперта, идентификация личности, физические параметры голоса, лингвистические особенности устной речи, идентификационные и диагностические исследования.

ՁԱՅՆԱԳՐՈՒԹՅԱՆ ՔՐԵԱԿԱՆ ՀԵՏԱԶՈՆՈՒԹՅԱՆ ՈՐՈՇ ԱՍՊԵԿՏՆԵՐ ԽՈՍՈՂԻ ՆՈՒՅՆԱԿՆԱՑՄԱՆ ՆՊԱՏԱԿՈՎ

Քրենդել Օ.Ի.

Հոդվածում լուսաբանվում են դատական քննության ընթացքում անձանց բանավոր խոսքի նույնականացման նպատակով փորձագետին առաջադրվող հարցերը: Վերլուծվում են հետազոտության օբյեկտիվ և սուբյեկտիվ մեթոդները, ներառյալ՝ լեզվական և գործիքային վերլուծությունը, ինչպես նաև լսողական և տեսողական հետազոտման մեթոդները՝ ծրագրային ապահովման և տեխնիկական սարքավորումների կիրառմամբ: Դիտարկվում են մարդու ձայնի և խոսքի փոփոխականության առանձնահատկությունները, ինչպես նաև ձայնագրությունների ձևավորման վրա արտաքին միջավայրի և այլ գործոնների ազդեցությունը: Ընդգծվում է, որ կան ինչպես գործիքային, այնպես էլ լեզվական մեթոդներով նույնականացման հետազոտություններ իրականացնելու տարբեր հնարավորություններ: Կատարվել է խոսքի դիագնոստիկ վերլուծություն, որն, ըստ հետազոտության արդյունքների, կարող է տեղեկատվություն տրամադրել անհայտ անձի իրավիճակային, հուզական, ֆիզիոլոգիական, հոգեբանական, ֆիզիկական և այլ բնութագրերի մասին:

Բանալի բառեր. աուդիո և վիդեո ձայնագրությունների վերլուծություն, դատական փորձագետի եզրակացություն, անձնական նույնականացում, ձայնի ֆիզիկական պարամետրեր, բանավոր խոսքի լեզվական առանձնահատկություններ, նույնականացում և դիագնոստիկ ուսումնասիրություններ:

The article was submitted: 08.10.2024

The article was accepted: 12.11.2024