DOI: https://doi.org/10.32353/khrife.1.2021.14
UDC 343.982.32

Kh. Lutsenko,
Researcher at NSC «Hon. Prof. M. S. Bokarius FSI»,
Kharkiv, Ukraine,
ORCID: https://orcid.org/0000-0001-7883-3764, e-mail: about-work@ukr.net
A. Roman,
Leading Forensic Expert at NSC «Hon. Prof. M. S. Bokarius FSI»,
Kharkiv, Ukraine,
ORCID: https://orcid.org/0000–0002–8406–7979, e-mail: anna_roman@ukr.net
S. Grigoryan,
Head of Computer-Technical Expertises Department of National Bureau of Expertises (NBE — SNPO), Yerevan, Armenia,
e-mail: s.grigoryan@nbe.am
V. Kocharyan,
Forensic Expert of the Department of Photo-Technical, Portrait, Audio and Video Expertises at the National Bureau of Expertises of the National Academy of Sciences of the Republic of Armenia, Yerevan, Republic of Armenia,
e-mail: qocharyan86@gmail.com

Research on a voice changed by distortion

The concept and essence of distortion are considered from a technical standpoint and criminal law perspective. The most common ways of distorting a voice and a speech are provided, as well as certain methods of detecting an intentional change in a voice and human speech by computer tools and linguistic analysis.

Some software and hardware tools changing a speech signal both in real time and in a pre-prepared recording are analyzed.

To solve diagnostic and identification tasks, a pressing issue in forensic video and audio analysis is studied which is addressed to forensic experts in this field more often.

Keywords: *simulation; distortion; forensic video and audio analysis; editing; person identification by voice and speech; instrumental analysis; linguistic analysis.*

Formulation of Research Problem. In recent years, the number of various anonymous messages (in particular, fraudulent ones, threats, as well as reports on mining administrative buildings, shopping centers, etc.) has grown. With necessary skills, you can simulate the speech of an opposite sex, another age and/or ethnic or territorial affiliation, the speech of a person in an excited state, in a state of alcohol or narcotic intoxication and other speech characteristics of a person.

Research on methods of concealing a crime is a multidimensional and complex problem that possesses not only forensic but also criminological, criminal-law and procedural significance. Some of the mentioned aspects of the problem have already been considered by forensic experts in scientific papers,

and ways to hide traces of a crime and concealing can be interpreted differently from the standpoint of various legal sciences. Thus, from the point of view of the criminal law, crime concealing is a way to avoid criminal liability, from the standpoint of criminology: an element of the method of committing a crime; and for forensic psychology, the way to conceal a crime is the embodiment of peculiarities of behavior and traits of a criminal personality and crime victim.

Specific nature of the forensic aspect of the issue of concealing is to study the laws of crime concealment, gain and handle information to investigate and prevent crimes.

Objections arise as to accepted in criminalistics identification in the structure of the method of committing crimes of the following components: preparation for crime commission → direct commission of a crime → concealing traces of a crime. Concealing (hiding, way of avoiding responsibility) may take place both at the stage of preparation for the commission of a crime and while its commission, as well as after its implementation. In individual cases concealing may not be present. Sometimes concealing can be absent while crime commission (for example, when it is committed as a result of negligence) or after its commission (for example, if a criminal voluntary and honestly declares what he has done). Consequently, concealing is always associated with the way of committing a crime.

Concealing should be interpreted as actions or inactivity of criminals or other persons aimed at full or partial destruction, change, hiding of traces of a crime which is being committed or has already been committed, misleading of both persons carrying out investigations and victims, witnesses on individual accomplices, a method of committing a crime, location of a corpse, objects obtained by criminal means, crime weapons and the rest of traces. Concealing is always a structural element of a crime commission. Study of concealing in the system of forensic science should be inextricably linked to the teaching on the way of committing crimes.

Knowledge of laws as to formation of certain concealing methods of criminal activity will allow not only to identify the method of concealing but also criminals by studying certain sings of these methods.

Concealing actions in criminal activity are conditioned by:

- a sense of fear, fear of exposure;
- impossibility of criminal intend realization without relevant concealing;
- imitation of certain patterns of criminal behavior taken from stories of other criminals, literature, movies, etc.;
- the desire to determine *originality,* impress a victim, witnesses, accomplices and other persons.

Analysis of Essential Researches and Publications. There is a great number of ways to create a human voice sample. Usually these are different combinations of frequency and statistical characteristics of a voice, such as intonation or tone height. Let’s consider some modern technologies created on the basis of artificial intelligence, namely neural networks.

The *Google DeepMind* British company presented a new algorithm for the synthesis of human voice called *Wavenet.* It is based on the use of neural networks helping to reach a realistic voice simulation. The program distinguishes sounds and assembles them on the basis of entered data that allows you to get a rather natural sound: for example, in this case *SIRI* or *Google Assistant* [1] may serve this purpose.

Also, there is an alternative method: parametric synthesis which uses a fully computer-generated voice and does not necessitate to use the collection of "live" speech. Its work is grounded on already set parameters corresponding to the rules of grammar and principles of sounds pronunciation.

In the field of creating tools for voice changing, there is a huge number of commands aimed at software products commercialization, for example *Resemble.AI* (provides demo version of a program), *iSpeech* (there is a demo version for 27 languages), *Lyrebird AI* (you can download demo version for 3 hours of speech), as well as *Vera Voice* created by *ScreenLife Technologies* company of Timur Bekmambetov and developers of "*Robot Vera" Stafory*[2] *HR* service. The technology of voice synthesis introduced by *Facebook*: authors of this project taught the neural network to speak the voice of Bill Gates. *Amazon* reported about a similar development. A voice assistant of the *Alexa* company can talk to its owners with the voice of the actor Samuel Leroy Jackson, and it will be able to master voices and speech of other celebrities over time.

Researchers presented a program based on artificial intelligence called *Vocodes*. This tool may simulate voices of more than 100 celebrities (politicians, scientists and artists: Donald Trump, Barack Obama, Bryan Cranston, Danny DeVito, Elon Musk, etc.). To simulate a voice, the user has to enter a text and select a voice. The *ThenExtWeb* edition notes that there are several such applications, but *Vocodes* impresses with a huge number of voices available for reproduction. However, it has its own disadvantages: every time the application faces a word that cannot read, it simply misses it. Among these words: *Tesla* and *Elon Mask* [3].

That is, the neural network has learned to simulate the tone, intonation and other nuances of a voice and speech. At the same time, creators of projects assure that technically it is impossible to identify whether the recording is synthesized or original, but how they do not specify. It is known that the system possesses the ability to detect synthesized human speech and distinguish it from natural.

[1] Siri або Google Assistant: хто кращий, розумніший і крутіше [Електронний ресурс]. URL: https://razborka-pc.com.ua/siri-abo-google-assistant-khto-krashchyj-rozumnishyj-i-kruti.html (date accessed: 13.03.2021).

[2] Проект Vera Voice научил нейросеть говорить голосом знаменитостей [Электронный ресурс]. URL: https://screenlifer.com/projects/screenlife-technologies-i-robot-vera-nauchili-nejroset-govorit-golosom-znamenitostej (date accessed: 14.03.2021).

[3] Нейросеть научилась реалистично имитировать речь человека [Электронный ресурс]. URL: https://nplus1.ru/news/2016/09/14/google-ai-voice (date accessed: 14.03.2021).

The **Article Purpose** is to analyze certain methods of detecting intentional voice change by computer tools and linguistic analysis.

Main Content Presentation. The main task of forensic video and audio analysis is to identify a person by a voice and speech. The speaker recognition while forensic video and audio analysis is a process of determining the belonging of a particular statement to an individual based on the characteristics of speech signals. As it is known, voices of different people usually are not similar to each other, besides (excluding the linguistic component) a human voice conveys information on a geographical and social status of the speaker, on his physiological features and psychological traits, emotional state, etc.

Referring to forensic video and audio analysis, new available methods for concealing and distorting a speech signal appear, which complicate the identification of certain characteristics.

According to the distortion nature, they can be ***intentional*** (artificial) and ***random*** (natural). *Natural distortions* have a "situationally, technically, technologically conditioned nature".

The *intentionality* of voice and speech *distortion* can be traced through peculiarities of a provided voice message, inconsistency of identified signs of a voice and speech in a studied sound record and in specimens of oral speech, which can be detected at the level of auditory perception. If a forensic expert identifies signs of any distortion in a phonogram, then he has to establish a *"mechanism, methods and reasons for introduction of changes into a sound record, their artificial or natural character".*

The concepts of *distortion* and *simulation*, at first glance, are extremely similar and aimed at: *"in any way to hide the real voice of the speaker within the framework of forensic audio and video analysis, however, simulation and distortion are implemented in various ways and have a different degree of performance complexity".*

Simulation is primarily imitation of a voice and speech of another person by analyzing peculiarities of the voice of a speaker (tempo, articulation, intonation, etc.). Possessing needed skills, you can simulate the speech of an opposite sex, another age and/or ethnic or territorial affiliation, the speech of a person in an excited state, in a state of alcohol or narcotic intoxication and other speech characteristics of a person.

Distortion of a voice is associated with changing linguistic and acoustic characteristics in a person's voice and speech to intentionally hide their own personality or personality of another person. When simulating the voice of a person, simulators try (intentionally or not) to reach those values of formant frequencies which correspond to a voice of a person that should be copied by reproducing certain characteristics of a voice and speech. However, as research demonstrates, it is almost impossible for simulators to reach values of formant parameters. At the same time, by means of comparative research using microanalysis of sounds and analysis of ratio between formant frequencies, it is

possible to determine with a certain degree of probability whose voice is recorded in a studied sound record.

To some extent, simulation of a voice and speech is an integral part of the concept of *distortion* which is more comprehensive in content, that is distortion can be implemented both by simulating a person's voice and speech.

According to authors, ***the most common methods of distorting voice and speech are:***

• *linguistic:* a change in linguistic parameters of oral speech;

• *acoustic:* a change in acoustic parameters of a speech signal;

• *mixed:* a change of both certain linguistic features, and acoustic characteristics of oral speech.

Linguistic distortion manifests itself in distortion of certain linguistic features of a voice and speech which leads to a change in some personality traits. For example, deliberate use of profanity (if it was not characteristic of a speaker), increase in the number of violations in grammatical rules (government, agreement, etc.), improper stress, a great number of jargons lowering the level of a speaker's speech culture. However, this type of distortion can be detected by both acoustic and linguistic analysis; while comparative research, concurrence of signs that have not been distorted could be traced, and appropriate evaluation of identified inconsistencies in features will enable to establish facts as to linguistic distortion of personality traits.

Mixed distortion is associated with distortion of certain linguistic features of a voice and speech, acoustic parameters, but without using computer tools. In this case, many features belonging to a linguistic group can be intentionally altered (for example, timbre, articulation, prosodic, lexical features and signs of speech culture, etc.), acoustic parameters are also changing (for example, the frequency of the main tone). Thanks to using mixed distortion, you can change the gender, age and dialect of a person. It can be connected with the need to simulate the voice of another person, and to hide their identity. Thus, a voice and speech are distorted, for example, with the help of clamping nostrils, jaw clenching, retracting cheeks, lips protrusion, etc.[1].

A criminal may resort to acoustic distortion in situations when there is an intention to remain unrecognizable, that is when his main goal is to ensure the impossibility of his further identification.

The majority of computer tools for altering voice focus on changing primarily features indicating the gender and age of a speaker. It should be noted that when changing the voice in real time, there is no change in a sound record. Changes are introduced at the stage of developing a speech signal entering the recording channel, as a result (by means of digital-to-analog transformation) a new, modified speech signal is formed, but without any signs of processing.

[1] Лебедева А. К. Судебно-экспертное исследование обликовых характеристик личности по фонограммам речи: правовые и методические аспекты : дис. … канд. юрид. наук. Москва, 2017. 231 с.

To understand how a voice is changed and which changes it undergoes, first, it is needed to study existing techniques and methods of intentional voice change using computer tools. There is a variety of programs for changing a speaker's voice, such as *AV Voice Changer Diamond,* Voxal *Voice Changer, MorphVOX JR, MorphVOX Pro, Scramby, Fake Voice, Funny Voice, Clownfish Voice Changer, VoiceMod,* etc.[1].

The above-mentioned programs are created particularly for a PC, however, there are also applications for altering voice which are made for modern phones and smart phones. It is possible to download them on the Internet. These programs enable to change the sex and age of a speaker and to help a criminal hide his personality this way.

Software and hardware tools for changing a voice can be classified into two groups:

- modifying a speech signal directly before it was sent to a recording channel, that is in real time;
- changing a speech signal already recorded in a sound record using post processing.

Diagnostic research on determining voice and speech characteristics is performed with the use of a set of methods:

- audio-linguistic analysis is intended to study a voice message, verbal behavior by analyzing and evaluating features of a speaker's speech and voice: features of the speech flow, phrases, words, sounds.
- acoustic comprising of phased calculation and evaluation of acoustic diagnostic features and comparison of these features vectors reflecting physical parameters of a speech signal.

Difficulty while research on sound records having undergone changes in real time is that none of the features of audio editing are formed: a completely new continuous sound record with a speech signal recorded in it, changed directly while its recording is created. In case of sound record change by means of post processing, a forensic expert should pay attention (excluding search for signs of change in the acoustic characteristics of the speech signal) to the presence of audio editing signs. The voice in a sound record changed both in the first and second way can be studied by analyzing file service-based data, within the framework of which the binary structure of a file, its metadata and other service characteristics are studied. A similar analysis is carried out by implementing such programs as *X-Ways Forensics, DUMP, Exiftool, WinHEX,* etc.

The majority of computer tools for changing a speaker's voice or his age work according to a similar algorithm: they automatically, according to a prelaid program, change those voice parameters that make up a functional-dynamic complex of human acoustic skills, especially the frequency of the main voice

[1] 19 лучших программ для изменения голоса онлайн и офлайн [Электронный ресурс]. URL: https://lifehacker.ru/programmy-dlya-izmeneniya-golosa (date accessed: 16.03.2021).

tone. The above-mentioned programs can change both the previously recorded sound record as well as modify the speech signal in real time (for example, when communicating by IP telephony, such as *Skype, WhatsApp, Viber, Telegram,* etc.). This method is used when it is necessary to hide your personality, but there is no chance to avoid a conversation in real time. The scheme for this algorithm is as follows: at the beginning, the speech signal is converted to a digital code with the help of analog-to-digital converter, and then processor converts this code according to the algorithm recorded in the processor. As a result, the digital code is converted to an electrical signal by means of analog-to-digital converter that the addressee receives at the other end of the telephone line.

In case of using the mentioned programs, delays appear while transmission of already altered speech signal. Before voice recording or telephone conversation, the user independently chooses how he wants to change his voice, how to lower or raise the frequency of the main tone, and only after begins to record a conversation accordingly.

So, let's consider programs processing a speech signal, that is when a sound record itself undergoes changes by means of any device. When changing the gender and age of a speaker this way, a situation is possible when a sound record with a recorded voice of one person is divided into several parts and each part is processed individually in a different way by software tools. As a result of changes, you can get several sound records with perfect voices from one sound record with a recording of the voice of one person. In such cases, the forensic expert should draw attention (except for the search for signs of change in acoustic characteristics of a speech signal) to the possibility of the presence of audio editing signs. Both types of the above programs changing sex characteristics of the speaker are based on the algorithm according to which the main tone changes by lengthening or shortening short speech segments. To preserve a natural speech tempo in case of a tone lowering (lengthening segments), some segments are deleted; in case of tone raise (shortening segments), certain segments are duplicated. In a foreign literature, the process associated with modification of frequency of the main tone without changing its speed is called *Pitch-Shifter* (word per word translation: *the main tone change*) is a sound effect or a corresponding device that adds to a signal its copy falling behind the main tone at any interval within two octaves up or down[1]. Pit-cipher is a way to alter the basic tone of the speech signal without changing its length. As it is known, the voice coloration depends on the values of formants frequency: if we change the frequency of the main tone of the speech signal with the help of specialized tools, then in this way, we will also change the value of formants, resulting in change in the voice nature, specifically in its timbre.

In the United States of America, over 10 companies produce telephone devices and add-on devices which have a voice change feature (*voice changer*)

[1] Питч-шифтер [Электронный ресурс]. URL: https://ru.wikipedia.org/wiki/Питч-шифтер (date accessed: 16.03.2021).

and are implemented by using the latest specialized chip. Add-on and telephone devices provide the user with an opportunity to speak one of 16 versions of voices: from a low bass to a high soprano, at the same time maintaining the speech tempo and sound naturalness. It enables to talk and not be recognized even by your closest people. Such products are of demand mostly for parents, minors who often stay home alone[1].

The Americans conducted research on the effect occurring as a result of increasing the frequency of the main tone called *Mickey Mouse effect*. It was called that way since due to increase in the frequency of the main tone the voice pitch also raises, and the listener hears a little " squeaky" timbre: similar to characters of cartoons about *Mikki Mouse.* To avoid this effect and to make a voice sound more natural, a formant correction algorithm is implemented thanks to which the position of formant frequencies updates after or at the time of pitch shifting, in other words the main tone formant frequencies do not change with a change in a frequency.

Marketers have long realized that the use of media characters in projects always draws audience attention to a product and, as a result, increases profits. The Americans and Europeans get used to listening to rules of passengers' behavior on board the aircraft voiced by celebrities.

However, expenditures for a project from attracting different celebrities are also growing: you have to pay a honorarium, to adapt to a celebrity's schedule, in addition, some of them can refuse to participate in it. As we noted above, creators of the *Vera Voice* project, made by using a neural network, succeeded in doing so because they decided not to invite celebrities at all: why they should pay them if you can simulate a speech acoustic signal with the help of the program.

Changing the sex of the speaker voice is possible by deleting frequency bands from 100 Hz to 500 Hz from the spectrum of a male voice; as a result of such changes, listeners identified a male voice in a suggested sound record as female[2].

Research on a voice altered by computer tools is a complex and controversial task for forensic video and audio analysis. Due to the fact that instrumental study of a changed voice recorded in a sound record still has certain difficulties, forensic experts use linguistic methods for identifying changes made to a sound record in practice. As it is known, the content and semantic parts of the speech material characterize a number of personality traits, and in the case of intentional

[1] Кринов С. Н. Современные методы изменения голоса и идентификация говорящего. *Информатизация правоохранительных систем* : мат-лы VIII Междунар. конф. Москва, 1999. С. 328—330.

[2] Потапова Р. К. Сексолект как комплексное гендерное понятие (об усложненном характере задачи определения сексолекта в судебной фонетике). *Информатизация правоохранительных систем* : мат-лы X Междунар. конф. Москва, 2001. С. 314—320.

change in a voice using computer tools, there is no change in the content of statements: only acoustic characteristics change.

For example, when oral speech in a sound record at the auditory level is perceived as a female voice of an elderly person, and the conversation is characterized by a low level of stylistic and speech competence (for example, with the use of curse words, low degree of development of speech and stylistic skills) and there are no signs of conscious distortion in oral speech, it can become the basis for drawing a possible conclusion on the fact that a speaker is a person with elementary / incomplete secondary education who has a low level of speech culture in communication.

It should be stressed that men and women speak with different intonations, use various means of non-verbal communication, pronounce words in a different way; also, their so-called vocabulary varies. In speech, women more often use interjections ("Ouch!", "Oh"", "Hey", "My God", etc.) helping to express emotions, diminutive words, phrases that convey feelings and emotions ("my sweetheart", "scary", "cool", "it is such a pleasure", etc.).

Conducting of forensic video and audio analysis comprises of a complex of various methods of not only linguistic analysis but also instrumental. Certain linguistic voice characteristics may not change in case of using computer and technical tools for distortion of the speaker's sex and age: the forensic expert must know how to establish the fact of changing a person's voice by means of instrumental analysis methods.

As far as it is known, methods of research on an altered voice were not developed until today. Thus, we will analyze which of the existing methods of forensic video and audio analysis can be applied to study a voice changed with the help of computer tools.

When conducting experimental researches on a voice and speech of males and females (20—35 years old), their characteristics were changed using *Adobe Audition* and *AV Voice Changer Diamond.* It should be noted that these programs are publicly exposed: they can be downloaded from the Internet.

As it is known, the sound pitch depends on the frequency of the main tone. The higher frequency, the higher a sound. To change the voice of a woman-speaker, we used *lower pitch* function, to change the voice of a man-speaker: *raise pitch* function. Using a certain algorithm laid down in the program, the voice pitch in the first case automatically lowers as a result of frequency decrease of the main voice tone, the female voice turns into a male, in the second case: on the contrary, raises due to increase in the main tone frequency.

It is possible to change the voice pitch in *Adobe Audition* program manually by changing certain settings. For example, on the *Effects* tab in the *Time and Pitch* section, select *Stretch and Pitch*, in the dialogue box: *Pitch Shift* parameter, reducing its value by 5 halftones: as a result, the voice tone pitch immediately lowers (for example, a female voice turns into a male). It should be noted that the main advantage of the *Stretch and pitch* effect is that it is possible to change

the tone of the sound signal without changing the tempo of the voice recording. In *Adobe Audition* program, you can also select the *Advanced Settings* panel (*Advanced* tab). By selecting the *Izotope Radius* algorithm, the user gets advanced features, for example, setting *Preserve Speech Characteristics* check box maintains the realism of speech in case of changing in the frequency of the main voice tone.

The *Formant Shift* function automatically determines how formants adapt in the event of a tonality shift. By default, the zero value adjusts formants along with a tonality shift, at the same time maintaining timbre and realism. Setting values above zero raises the timbre enabling you to convert a male voice to a female one, below zero: creates the opposite effect: converts a female voice to a male one.

We deem it important to stress that due to the fact that low frequencies possess more energy than high ones, and some sounds are clearly recognized by one first formant (for example, sounds [a], [o], [y]), if you cut the spectrum auxiliary formants of the speech signal with the help of low-pass filter, the individual coloration of uttered sounds for each person will disappear, but the very semantic information embedded in the speech message will not be altered and remain clear.

Let's consider, for example, the *AV Voice Changer Diamond* program, which belongs to methods of changing a speech signal in real time, that is in the process of its formation during entry into the recording channel. This program has a *Voice morpher* function, which is responsible for changing a voice. The algorithm for changing a voice is the same as in the previous program: we can raise or decrease the frequency of the main voice tone while recording. To start recording a voice, click on the *Recorder* icon and turn on the recording *(rec).* Now, when you speak into the microphone, the speech signal will change according to the parameters of the *Voice morpher* field. These settings for changing a voice can be changed while recording: for example, we can record one half of a speech at a low frequency and the other half: at a high frequency. Pressing the *stop* button stops recording and automatically saves it *in MPEG-1 Layer 3(*.mp3)* format.

The program also includes certain algorithms, enabling which, we can choose, for example, the change of a female voice to male, the voice of a child, an elderly person, etc. You can select a specific algorithm in the *Nickvoices* tab, individual algorithms are introduced for female and male voices.

This program allows you to change a voice when communicating through various programs such as *Skype, WhatsApp, Telegram* via voice chats and voice messages, and can also change a voice from the streaming audio that will be played on the computer when using the program.

However, we cannot exclude the fact that voice change systems are implemented by criminals to commit criminal offenses. As practice demonstrates, even an experienced forensic expert may not identify changes in a voice, which led to the following conclusion: perhaps in some unsolved criminal cases involving anonymous phone calls, criminals used technical means of voice

change; however, due to the reasons mentioned above the fact of change could remain unnoticed by forensic experts.

To solve this task, it is needed to consider the properties of possible algorithms for changing a voice. The above-mentioned *pitch shift* and *voice changer* algorithms alter the main tone by lengthening or shortening short sections of speech. To maintain the natural tempo of speech in case of lowering a tone (lengthening of sections), some sections are deleted, in case of raising a tone (shortening of sections), some sections are repeated twice. For this purpose, there is no need to perform any analysis of the input signal: the algorithm operates formally and equally at pauses, vowels and consonants.

In case of a tone raise, there should be repetitive sections of speech in the signal. The duration of such a section is a parameter by which you can determine the degree of shift and restore the original speech.

In case of a tone lowering certain sections are distinguished in a signal. Since the algorithm functions formally, without being connected to specific sounds, it may result in loss of some consonant and transition sections. Such deformations can be detected by ears at raised and even normal tempo of an original speech. To avoid it, a speaker delivering a speech should speak more slowly, maintaining prolong pauses in a speech, comparing to the norm. At the same time, a listener can perceive disappearance of some short consonant and transition sections as slip of a tongue. Identification of signs of tone lowering algorithm application in such conditions becomes a complex task which can be solved in availability of sufficient duration of recorded message in a studied sound record, when the quantity of detected slips of tongue (in case of listening sections) is statistically significant. In this case, the evaluation of probability as to disappearance of short speech segments in a studied sound record must coincide with a calculated probability under a certain parameter of the tone shift down the frequency[1].

Functionalities of modern software include the task of destroying signs of changes in a sound record. For example, in the *Sound Forge* program you can create a desired sound wave shape, adjust a tempo and change spectral characteristics, timbre and sound quality of an original sound record, remove signs of crossfade, which considerably complicates the process of forensic video and audio analysis. By using *hex*-editors, you can remove signs of post-processing even in the binary structure of a file. In such cases, the only way to identify the set of basic personality traits on the basis of the analysis of a person's voice and speech is to conduct a linguistic analysis allowing to identify signs of editing even in well-edited records, which will help a forensic expert to make a well-grounded and informed decision.

[1] Лебедева А. К. Особенности судебно-экспертного исследования голоса, изменённого при помощи компьютерно-технических средств. *Известия Тульского государственного университета. Экономические и юридические науки*. Тула, 2016. № 3—2. С. 323—329.

For example, if a voice in a sound record is perceived by a listener as the voice of a child, it is possible to determine, for example, discrepancy between the level of speech culture and the level of education of a person and so on. In case a forensic expert has samples of a person's voice and speech whose voice is presumably recorded in a sound record, it is possible to conduct a comparative study and determine the presence of coincidences / differences by certain features of the linguistic group. In the course of instrumental analysis, statistical, dynamic, spectral speech characteristics are studied. Main provisions of the instrumental analysis of a voice are based on the individuality of the spectral content of a voice for each person, despite general mechanisms of specific sounds formation.

Conclusions. Summarizing the above, let's highlight key moments that a forensic expert should pay attention to in the presence of signs characteristic for the use of computer software to change a voice.

1. To determine the presence/absence of signs of changes in a sound record and their cause which may be the result of the use of computer tools to change a voice, editing of a sound record or peculiarities of devices operation, etc. During the instrumental analysis, to identify, if possible, the implemented software to record a studied sound record by analyzing the file structure, spectral representation of a signal, etc.
2. The use of computer tools to change a voice can be evidenced by: the presence in the file metadata of information on the use of specific software (being used to alter a voice), inconsistency between linguistic and acoustic features and characteristics of the speaker's speech. For example, a voice that a listener perceives as male at the auditory level in a studied sound record corresponds to characteristics of a woman's speech or a person in an excited state (alcohol or drug intoxication, etc.).

Thus, despite the lack of developed methodological approaches to the study of a voice altered by computer tools, it is crucial to establish the fact of the use of software tools, and (in some cases) to detect altered voice and speech characteristics that individualize a person, and later, at the next stage: to identify a speaker.

References

19 luchshikh programm dlia izmeneniia golosa onlain i oflain [Ehlektronnyi resurs]. URL: https://lifehacker.ru/programmy-dlya-izmeneniya-golosa (data zvernennia: 16.03.2021) [in Russian].

Krinov, S. N. (1999). Sovremennye metody izmeneniia golosa i identifikatsiia govoriashchego. *Informatizatsiia pravookhranitelnykh sistem* : mat-ly VIII Mezhdunar. konf. Moskva [in Russian].

Lebedeva, A. K. (2016). Osobennosti sudebno-ehkspertnogo issledovaniia golosa, izmenennogo pri pomoshchi kompiuterno-tekhnicheskikh sredstv. *Izvestiia Tulskogo gosudarstvennogo universiteta.* Ehkonomicheskie i iuridicheskie nauki. № 3—2 [in Russian].

Lebedeva, A. K. (2017). *Sudebno-ehkspertnoe issledovanie oblikovykh kharakteristik lichnosti po fonogrammam rechi: pravovye i metodicheskie aspekty* : dis. … kand. iurid. nauk. Moskva [in Russian].

Neiroset nauchilas realistichno imitirovat rech cheloveka [Ehlektronnyi resurs]. URL: https://nplus1.ru/news/2016/09/14/google-ai-voice (data zvernennia: 14.03.2021) [in Russian].

Pitch-shifter [Ehlektronnyi resurs]. URL: https://ru.wikipedia.org/wiki/Pitch-shifter (data zvernennia: 16.03.2021) [in Russian].

Potapova, R. K. (2001). Seksolekt kak kompleksnoe gendernoe poniatie (ob uslozhnennom kharaktere zadachi opredeleniia seksolekta v sudebnoi fonetike). *Informatizatsiia pravookhranitelnykh sistem* : mat-ly X Mezhdunar. konf. Moskva [in Russian].

Proekt Vera Voice nauchil neiroset govorit golosom znamenitostei [Ehlektronnyi resurs]. URL: https://screenlifer.com/projects/screenlife-technologies-i-robot-vera-nauchili-nejroset-govorit-golosom-znamenitostej (data zvernennia: 14.03.2021) [in Russian].

Siri abo Google Assistant: khto krashchyi, rozumnishyi i krutishe [Elektronnyi resurs]. URL: https://razborka-pc.com.ua/siri-abo-google-assistant-khto-krashchyj-rozumnishyj-i-kruti.html (data zvernennia: 13.03.2021) [in Ukrainian].

Х. В. Луценко, А. І. Роман, С. С. Григорян, В. В. Кочарян

Дослідження голосу, зміненого маскуванням

Розглянуто поняття й сутність маскування з технічного погляду та з погляду кримінального права. Наведено найпоширеніші способи маскування голосу й мовлення, деякі методи виявлення навмисного змінення голосу та мовлення людини за допомогою комп'ютерно-технічних засобів і лінгвістичного аналізу.

Проаналізовано останні публікації, а також дослідження в галузі імітування голосу й мовлення людини за допомогою новітніх комп'ютерно-технічних засобів — деяких сучасних технологій, створених на базі штучного інтелекту, а саме — нейронних мереж. Розглянуто найпоширеніші способи маскування голосу й мовлення диктора. Наведено деякі методи виявлення навмисного змінення голосу за допомогою комп'ютерно-технічних засобів і лінгвістичного аналізу.

Проаналізовано деякі наявні програмно-технічні засоби, що змінюють мовленнєвий сигнал як у реальному часі (а саме — у процесі його формування під час надходження у канал запису), так і на заздалегідь підготовленому записі (за допомогою засобів оброблення мовленнєвого сигналу після запису на будь-якому пристрої). Констатовано (зважаючи на експертну практику), що в тих випадках, коли виявити ознаки маскування голосу/мовлення не виявляється можливим інструментальним способом (що, зі свого боку, суттєво ускладнює процес виконання судової експертизи відео-, звукозапису), то єдиною можливістю встановити індивідуальні характеристики голосу й мовлення залишається лінгвістичний аналіз.

Зроблено висновок: незважаючи на відсутність розроблених методичних підходів щодо дослідження голосу, зміненого за допомогою комп'ютерно-технічних засобів, є змога як визначити факт використання програмних засобів, так і (у деяких випадках) діагностувати змінені характеристики голосу й мовлення, що індивідуалізують особу, а згодом, на наступному етапі,— провести ідентифікацію диктора.

Із метою вирішити діагностичні й ідентифікаційні завдання досліджено проблемне питання в експертизі відео-, звукозапису, яке все частіше постає перед експертами у цій галузі.

Ключові слова: *імітація; маскування; експертиза відео-, звукозапису; монтаж; ідентифікація особи за голосом і мовленням; інструментальний аналіз; лінгвістичний аналіз.*

К. В. Луценко, А. И. Роман, С. С. Григорян, В. В. Кочарян

Исследование голоса, изменённого маскировкой

Рассмотрены понятие и сущность маскировки с технической точки зрения и с точки зрения уголовного права. Приведены наиболее распространённые способы маскировки голоса и речи.

Проанализированы последние публикации, а также исследования в области имитации голоса и речи человека с помощью современных компьютерно-технических средств, а именно обученного искусственного интеллекта на основе нейронных сетей. Рассмотрены наиболее распространённые способы маскировки голоса и речи диктора. Приведены некоторые методы выявления умышленного изменения голоса и речи человека с помощью компьютерно-технических средств и лингвистического анализа.

Проанализированы некоторые существующие программно-технические средства, изменяющие речевой сигнал как в реальном времени (а именно — в процессе его формирования при поступлении в канал записи), так и на заранее подготовленной записи (с помощью средств обработки речевого сигнала после записи на любом устройстве). Констатировано (исходя из экспертной практики), что в тех случаях, когда выявить признаки маскировки голоса/речи не представляется возможным инструментальным способом (что, в свою очередь, существенно затрудняет процесс производства судебной экспертизы видео-, звукозаписи), то единственной возможностью установить индивидуальные характеристики голоса и речи является проведение лингвистического анализа. Несмотря на отсутствие в настоящее время разработанных методических подходов к исследованию голоса, изменённого с помощью современных технологий, установить факт использования вышеуказанных средств возможно.

С целью решить диагностические и идентификационные задачи исследован проблемный вопрос в экспертизе видео-, звукозаписи, который всё чаще встаёт перед экспертами в этой области.

Ключевые слова: *имитация; маскировка; экспертиза видео-, звукозаписи; монтаж; идентификация личности по голосу и речи; инструментальный анализ; лингвистический анализ.*

Надійшла до редакції 31.03.2021 р.

Lutsenko Kh. V., Roman A. I., Grigoryan S. S., Kocharyan V. V. Research on a voice changed by distortion. *Теорія та практика судової експертизи і криміналістики* : зб. наук. пр. / редкол.: О. М. Клюєв, В. Ю. Шепітько та ін. Харків : Право, 2021. Вип. 23. С. 188—202. DOI: https://doi.org/10.32353/khrife.1.2021.14.