Kh. V. Lutsenko,
Researcher of KhRIFE
Kharkiv, Ukraine
ORCID:https://orcid.org/0000–0001–7883–3764
e-mail: about-work@ukr.net

K. V. Nikulin,
Researcher of KhRIFEKharkiv, Ukraine,
ORCID:https://orcid.org/0000–0002–0665–6313
e-mail: knikulin87@gmail.com

VOICE IDENTIFICATION OF ANNOUNCER AS ONE OF THE MODERN PERSONAL BIOMETRIC IDENTIFICATION METHODS

The article deals with the most widespread biometric identification systems of individuals, including voice recognition of the speaker on video and sound recordings. The urgency of the topic of identification of a person is due to the active informatization of modern society and the increase of flows of confidential information.

The branches of the use of biometric technologies and their general characteristics are given. Here is an overview of the use of identification groups that characterize the voice. Also in the article the division of voice identification systems into the corresponding classes is given.

The main advantages of voice biometrics such as simplicity of system realization are considered; low cost (the lowest among all biometric methods); No need for contact, the voice biometry allows for long-range verification, unlike other biometric technologies.

The analysis of existing methods of speech recognition recognition identifying a person by a combination of unique voice characteristics, determining their weak and strong points, on the basis of which the choice of the most appropriate method for solving the problem of text-independent recognition, Namely the model of Gaussian mixtures, was carried out.

The prerequisite for the development of speech technologies is a significant increase in computing capabilities, memory capacity with a significant reduction in the size of computer systems. It should also be Noted the development of mathematical methods that make it possible to perform the Necessary processing of an audio signal by isolating informative features from it.

It has been established that the development of information technologies, and the set of practical applications, which use voice recognition technologies, make this area relevant for further theoretical and practical research.

Key words: biometric identification technologies, Voice Identification Speaker, person verification, automatic speech recognition methods, examination of video, audio recordings.

Formulation of Research Problem.

Biometric technologies are technologies based on measuring the unique characteristics of a person.

These can be both unique birth characteristics (DNA, fingerprints, iris, etc.) as well as characteristics acquired over time or may change with age or external influence (handwriting, voice, gait, etc.).

Biometric technologies are widely used in many industries that are related to the security of access to information and material objects, as well as the unique identification of a person.

These can be access to jobs and network resources, law enforcement, social services to information protection, banking and other financial security.

Analysis of recent researches and publications.

The first international patent for the voice identification system was filed in 1983 by the CSELT Research Telecommunications Center (Italian Republic), by Michele Cavazza and Alberto Ciaramella. In May 2013, Barclays banking units began using the customer identification system over the first 30 seconds of a normal call. The system was developed by Nuance.

Currently known are the following developers of voice identification systems: Nuance, USA; Nok Nok Labs; VoiceVault, a US based research and development center in the United Kingdom of Great Britain and Northern Ireland; Sensory, Inc., United States; CMT companies, Russian Federation; Sistema-Sarov Innovation Technology Center, Russian Federation; BioLink, Russian Federation; AFM Solutions, Russian Federation; ValidSoft; Auraya Systems; Authentify; KeyLemon; Verint Systems; VoiceTrust.

A widely recognized market leader is Nuance, their solution is powered by Aeroflot, Speech Interpretation and Recognition Interface (Siri) is a personal assistant and answer and answer system developed for iOS.

This application uses natural language processing to answer questions and make recommendations. Siri adapts to each user individually exploring its benefits over time.

However, since a person's voice may vary depending on a number of factors, this method is not completely accurate[1].

The Article Purpose is to provide an overview and analysis of existing methods for biometric identification of a person, in particular, voice identification of a speaker.

Main Research Content Presentation.

The basis of the science of personality identification is the idea of measuring the body of a person and its parts.

These ideas were first formulated by Alphonse Bertillon (1853–1914), French forensic scientist, an employee of the Paris prefecture, who was engaged in the registration of criminals.

---

[1] Биомет Biometrija ot «A» do «Ja» [Biometrics from "A" to "I"]: polnoe rukovodstvo biometricheskoj identifikacii i autentifikacii. URL : https://securityrussia.com/ blog/biometriya.html (data zvernennja 04.06.2019) [in Russian].рия от «А» до «Я»: полное руководство биометрической идентификации и аутентификации. URL : https://securityrussia.com/ blog/biometriya.html (data zvernennia 04.06.2019).

In 1879, he introduced a system for identifying criminals, called anthropometry, and included: measurements of height, length and volume of the head, length of arms, fingers, feet, etc., as well as a verbal portrait of the criminal, a photo portrait in full face and profile, as well as description of special features.

Modern criminalistics still uses such a system, supplementing it with anthropometry, fingerprinting, photographic work, new methods for describing special features on a person's face or body and technologies for their implementation.

However, the concept of biometrics was formed a decade later.

In the origins of early biometrics was Francis Galton, English researcher.

The term *biometrics* first appeared around 1980.

In the encyclopedia of modern Ukraine, there is such an interpretation of the term biometrics (from *bio* and *metrics*): the science of applying the methods of mathematical statistics to study the phenomena of life. Another title is variation statistics.

The **ISO / IEC 2382-37: 2012** international standard: Information technology — Vocabulary — Part 37: Biometrics provides the definition of biometric systems.

Biometric system is a system designed to automatically recognize an individual (human person) based on his or her behavioral and biological characteristics[3].

In the field of information technology, the following interpretation applies: *biometrics* is a set of automated methods and means of identifying a person based on his physiological or behavioral characteristics.

In the field of information technology, the following interpretation applies:

In recent biometrics based systems and technologies, biometric verification, identification and authentication should be distinguished.

In accordance with **ISO / IEC 2382-37–2016 37.08.03 biometric verification**: the process of conforming a biometric statement by comparison. The use of the term *authentication*" instead of the term *biometric verification* is unacceptable[4].

In accordance with the same **ISO / IEC 2382-37–2016 37.08.02** biometric identification: a process of searching a biometric retrieval database aimed at finding and returning an identifier (s) of a biometric control pattern associated with a single individual.

The use of the **authentication** term instead of **biometric identification** is unacceptable.

---

[3] ISO/IEC 2382–37:2012. («Information technology – Vocabulary – Part 37:Biometrics», IDT). URL : https://www.iso.org/standard/55194.html (data zvernennia 13.05.2019).

[4] ISO/IEC 2382–37–2016. [ISO/IEC 2382–37:2012] («Information technology – Vocabulary – Part 37:Biometrics», IDT). URL : https://meganorm.ru/Data2/1/4293747/4293747566.pdf (data zvernennia 13.05.2019).

In the same international standard **ISO / IEC 2382-37–2016 37.08.01** authentication– action that proves or demonstrates undeniable origin or authenticity.

The term is used in biometrics as a synonym for biometric verification application, **biometric verification** function used as a synonym for biometric identification supplements and biometric identification function.

Thus, *biometric identification* is a way of identifying a person by specific specific biometric features (identifiers) that are inherent in a particular person.

Biometric authentication is the recognition of an individual based on physiological characteristics and behavior. Authentication is carried out by checking the identity of the person presented with the identification mark.

Table 1 provides examples of the use of biometric technologies in various applications[5].

Table 1

Areas of use of biometric technologies and their general characteristics.

№	Field of use	Key features
1	Computer Security	In this field, biometrics are used to replace (sometimes intensify) the standard logon procedure for a variety of passwords, startup cards, touch-memory tablets, and the like. The most common solution based on biometric technology is the identification (or verification) of biometric characteristics on a corporate network or when entering a workstation (personal computer, laptop, etc.).
2	Trade	Main Destinations: – in stores, restaurants, coffee bars, biometric identifiers are used either directly as a means of identifying the buyer to withdraw money from the account, or to confirm the buyer's right to any discounts or other benefits; – in vending machines and ATMs as a means of identifying a person instead of magnetic cards or supplementing them; – In e-commerce, biometric identifiers are used as a means of remote identification via the Internet, which is much more secure than

---

5 Biometricheskie sistemy identifikacii i autentifikacii [Biometric Identification and Authentication Systems]. URL : http://mirznanii.com/a/309729/biometricheskie-sistemy-identifikatsii-i-autentifikatsii (data zvernennja 04.06.2019) [in Russian].

		passwords, and in combination with cryptography, gives electronic transactions a very high level of protection.
3	Access Control and Management Systems	In a network architecture, where multiple entrances are equipped with biometric locks in the building, all staff members' biometric characteristics templates are stored centrally, with information about who and where (and possibly when) the entry is allowed. The following recognition technologies are implemented in such systems: fingerprint, face, arm shape, iris, voice, etc.
4	Automated fingerprint information systems	The main purpose of such a system is to manage the rights granted by the state to citizens and foreigners. Citizens' rights, voting, places of residence, work of foreigners, the right to receive social security, etc. are recognized and confirmed through documents and various cards. Nowadays, such systems have become widespread because some countries have begun to use them to verify the identity of those who drive.
5	Complex systems Systems	Systems of this type include solutions that combine the systems of the first three classes. The company employee registers with the system administrator only once, in the future he is automatically assigned all the necessary privileges both to enter the premises and to work in the corporate network with its resources.

In addition to these major sectors of biometric technology, biometrics is now being actively used in other areas, such as:

– Gaming.

Biometrics are used in two ways: checking all of the "black lists" (analogue of mass identification by persons used in airports), as well as the identification system and payment of regular customers;

– identification on mobile devices such as mobile phones, tablets, laptops, etc;

– in the transport industry as a means of payment;

– electronic voting systems (used instead of cards);

– medicine.

Biometrics are used to identify health care providers upon access to classified information and to receive electronic signatures in medical records.

Modern biometric authentication is based on two main methods:

– static method of authentication — recognizes the physical parameters of the person, which he possesses throughout life: from his birth to his death (fingerprints, characteristics of the iris, thermogram, face geometry, hand geometry, fragment of genetic code, fragment of genetic code);

– dynamic method of authentication — analyzes the characteristics, features of user behavior, which are displayed at the time of performing any ordinary daily activity (signature, keyboard handwriting, voice, etc.).

Static method has always been a major player in the global biometric protection market.

Dynamic authentication and combined information security systems accounted for only 20% of the market.

However, in recent years there has been an active development of dynamic protection methods[6].

Voice identification has been known for a long time, and in the absence of any technical means, the person identified the other person by three possible features — voice, signature and appearance.

Therefore, voice identification is one of the oldest methods of biometric identification.

Each person has a unique voice that is different from any other specific trait.

Voice identification is one of the most attractive systems for identification, but current problems in this type of biometric system should be taken into account at least in working systems.

For example, voice recognition can be effectively used as an additional method, for example, to face recognition, since the probability of self-recognition error in voice is 2-5 %.

Currently, the areas of identification by voice are actively developing.

The advantage of voice biometrics is the ease of implementation of the system, which usually consists of a voice receiver, voice recorder, voice modulator, biometric software and a voice database.

Unlike other biometric technologies, voice biometrics allows for long-range verification.

One of the promising ways to increase the reliability of voice identification is to integrate the characteristics of the dynamics of subliminal movements, which is actively used in signature identification.

On the other hand, there are applications where voice identification is most convenient, such as remote access to telecommunication channels for voice data analysis[7].

---

[6] Biometrija ot «A» do «Ja»…[Biometrics from "A" to "I"]   [in Russian].
[7] Tsarov R. Yu., Lemekha T. M. Biometrychni tekhnolohii. [Biometric Technolonies] Navch. posib. Odesa, 2016. S. 140 [in Ukrainian].

The voice recognition method identifies a person by a set of unique voice characteristics.

The algorithms analyze the main features that make a decision about the speaker's identity: voice source, resonant frequencies of the speech tract, their attenuation, and the dynamics of articulation management.

Based on age experience, modern scientists use two groups of features to characterize the voice when identifying by voice.

The first group is the physiological (anatomical) features that are related to the features of the mechanism of human speech formation.

The second group is the so-called articulatory features, which are based on the peculiarities of the human nervous system, which determines the nature of the use of physiological features.

Physiological features.

Physiological traits are based on the model of the speech tract. In this case, the main characteristics are several parameters that characterize the voice:

– energy of the speech signal;

– frequency range of the speech signal;

– fundamental frequency determines the length of the speech tract;

– formants determine the concentration of the speech signal by frequency and characterize the loud sounds.

Articulation features.

If the physiological features reflect the statistical properties of the language apparatus, then the articulatory features make it possible to describe the behavior of the language apparatus over time, that is, to reflect the articulatory dynamics of the language.

The main factor influencing this group of attributes are the socially conditioned speech skills of a person, his / her individual experience, temperament and personality traits.

Articulatory signs take into account the intonation of speech, rhythm, stress, volume. To obtain these characteristics, the concept of syntagm is used.

Syntagma is a rhythmic-melodic unit of language, grammatically worded and defined within a more complex whole structure (such as a sentence) with a finite thought.

Within the syntagm, the segments distinguish between language characteristics and intonation characteristics of speech, namely: voice intensity; melody of voice; stress system; temporal characteristics: length of segments and pauses; rate of speech; tone of speech.

It should be noted that modern voice identification systems can use both physiological and articulatory features at the same time.

Voice identification systems can be divided into the following classes: text-dependent, text-independent, voice-dependent, voice-independent.

Voice-dependent systems are systems that focus on the characteristics of a particular person or group of people, so they can be used to identify only that person (group of persons).

If you change the speaker (the person who is identified by the system), you need to adjust the system again using the voice features of the new speaker.

Dictator systems are systems that are not tied to a person's voice tags and can be used to identify any person.

Such systems themselves identify the necessary voice features and compare them with the standard from the base.

Text-dependent systems are voice identification systems that identify a person by a particular keyword or key phrase that an identifying person, for example, announce a password that is randomly generated.

The use of individual features and the coherence of phrases generated and recognized increases reliability.

Non-text-based systems are systems that identify a person by voice without being tied to any keywords.

In this case, the articulatory features of the human voice are important, they are used as the main features, and the physiological features act as secondary ones.

Non-textual identification implies the use of only individual features.

An important characteristic of the voice identification system is the speed (speed) of identification.

Improved performance can be achieved through the use of new fast processing algorithms.

Adweek estimates that in 2019, the voice recognition platform market will grow to $ 601 million, and by the end of 2011 to $ 40 billion.

All because it is easier for people to talk than to type, they need voice assistants that support habitual communication.

There are already a lot of helpers on the market: Amazon Alexa, Google Assistant, Cortana, Bixby, Alice, SoundHound, Apple Siri, X.ai and others.

The introduction of voice control devices in the car is one of the trends that will lead to global changes in the automotive sector.

Such devices will be able to centrally control most of the car's functions through human speech, eliminating the need to use buttons, dials, switches and more. Using voice recognition devices, users will be able to easily control a range of car features, making it more comfortable for the person and avoiding direct driving while focusing on driving.

The introduction of such technologies will increase in the near and medium term[8].

Improving the reliability of voice identification is important not only in the areas such as the separation of access to physical and information

---

[8] Biometrija ot «A» do «Ja»…[Biometrics from "A" to "I"]   [in Russian].

objects, such as access to the operating system of a personal computer or remote access to telecommunication channels for voice data analysis.

There is some interest in related areas of language technology: speech recognition, voice command management, and more. Electronic digital signature for the protection of confidential documents in the form of a secure electronic device (token) has become widespread today and in this regard the development of the protection of confidential documents based on a language signature is in view.

In addition, the practical applications of such research are useful for law enforcement agencies, such as the identification of a person by the physical parameters of the voice.

However, the development of technology and technological advances not only bring about positive points but also empower criminals.

One of the major offenses in the telecommunications environment is telephone fraud, which is rapidly gaining in popularity and is becoming a real epidemic.

Victims of criminals are all without exception — these are businessmen, officials, stars of show business, and ordinary citizens.

Here are some basic types of phone fraud and how to combat them.

The most common type of telephone fraud is the so-called "Roddy in Need".

How is it organized?

A person is calling from an unknown number.

The offender is introduced by a relative or acquaintance and in an agitated voice informs him that he has been detained by police officers and is accused of committing a crime.

It can be a road accident, or storage of weapons, drugs, bodily harm, or even murder.

Next, the so-called police officer comes into the conversation, who tells in a confident tone that he has repeatedly assisted such people. To resolve the issue requires a certain amount of money, which should be brought to the designated place and handed to some man.

The asking price is usually between one and several thousand dollars.

Similar voices, one group voices that do not differ can be encoded in the cellular communication systems approximately the same, so they will appear similar to the degree of mixing in auditory perception and comparison of integral acoustic parameters.

It is in this particular feature of cell broadcasting that the prerequisites for the implementation of telephone fraud lie when the likelihood of mistakenly recognizing someone else's voice as an acquaintance is high when accessing a cellular phone.

As it is known, emotional state of a person significantly affects the characteristics of voice, conversation style, etc.

There are times when, upon receiving such a message, a person is deceived, even if the person referred to in the message is nearby.

This is how parents are arranged: they always worry about their children, and the reaction to a possible threat to them is very strong.

Simple, effective and brazen scheme that uses strong feelings and basic instincts.

The same principle is based on a type of telephone fraud — "Mom, I have problems, do not ring, transfer money to this account."

Like fingerprints in forensic science, criminalistics (fingerprinting), forensic examination uses video and audio objects, namely video, recordings recorded on storage media.

In this regard, forensic forensics use their specific methods and techniques.

The physical basis of voice verification is the anatomy of the speech tract, the properties of the articulation management system, and features of the voice source.

The anatomy of the tract determines the spectral characteristics of the speech sounds, the articulation control system influences the speech rate, the speed of transients, and the duration of the speech segments, and the voice source determines the pitch frequency and the timbral characteristics of the speech signal.

Only those features that can be directly measured in a speech signal are investigated.

However, as the results of the research show, speaker verification in the acoustic parameter space provides characteristics that are most satisfying and best suited to the real world environment.

Diagnosis uses known features of a class or group of objects and compares them with features of a particular object, which determines its belonging to that class or group.

To achieve this goal, experts solve different types of tasks.

An important characteristic of the voice identification system is stability.

Obstacles are distortions, noises, impulses, etc.

Current classification methods used in voice identification systems are very sensitive to noise, resulting in reduced reliability when exposed to noise.

An example of channel distortion is the reverberation of sound, ie, the sound is repeatedly reflected from objects in the room.

The background, in some cases being an obstacle, can have a significant impact on voice identification.

They may have a *significant* signal level (for example, vehicle sounds; sounds produced by devices and household appliances; sounds inherent in the mechanisms, devices, apparatus, devices and means that accompany the operation of these sources; sounds of wildlife; sounds of phenomena of nature ; mechanism sounds; sounds of devices, apparatus and devices specially designed to create, amplify and radiate sound, etc.) and *overlap* the range of the speech signal.

In a telephone channel, there may be obstacles such as clicks, overloads, music, beeps (tones), etc.

Since modern digital mobile devices typically have a built-in microphone and productive hardware, creating a voice authentication system with more cost-effective computing methods is quite a challenge for mobile platforms.

However, it is still necessary to provide minimal computational costs while maintaining accuracy, noise immunity to various types of interference, and sufficient reliability with common hardware.

Speaker identification is the process by which the system can determine who is the speaker based on the information from the speech signal.

Recently, sound recording has also been widely used in integrated security systems, video surveillance systems (CCTV) and security systems.

The information we receive is used to identify violators and to analyze the state of the audio environment in order to monitor personnel actions and security.

Audio information is also used in telephone call transfer information systems (TIS), alert systems, alarm calls, and the like.

In this regard, it becomes urgent to solve problems related to the analysis of audio information obtained during recording in security systems and which can be used for further analysis in specialized laboratories of law enforcement agencies, laboratories and centers of forensics, research and training centers to:

– identification of the person by means of the recorded phonogram;

– analysis of the noise background, diagnostics of the acoustic environment and the conditions of recording;

– identification of sound recording equipment;

– improving the quality and legibility of existing records;

– protection of the speech signal from unauthorized access;

– compression of speech messages;

– establish the literal content of low quality records.

When solving the problems of protecting physical objects and information resources from criminal and terrorist threats, it is very interesting to use audio information (audio voice recordings) in access control (AC) and management systems (ACS).

The peculiarity of such systems is that they allow remote (by telephone) and hidden authentication, which is sometimes the only possible means of establishing the interlocutor's identity.

Consider the most common modern automatic speaker recognition techniques that underlie voice biometric systems.

A method for comparing fundamental statistics[9].

---

[9] Matveev Ju. N., Simonchik K. K. Sistema identifikacii diktorov po golosu dlja konkursa NIST SRE 2010 [Sistema identifikacii diktorov po golosu dlja

The fundamental tone is one of the basic parameters of the speech signal, and does not depend very much on the recording conditions and the type of channel.

У зв'язку з цим, метод розпізнавання дикторів, заснований на параметричній статистики основного тону (ОТ) є одним з базових серед автоматичних методів розпізнавання диктора.

The advantage of the spectral method used to calculate the pitch frequency is that it allows the pitch frequency to be estimated using all the available signal bandwidth.

Frequency response (AFC) of a sound recording channel always significantly affects the shape of the spectrum.

This effect should be taken into account, because otherwise the frequency response of the channel, on the one hand, may mask the individual parameters of the voice of the speaker, and on the other hand, mask some of the spectrum of the speech signal and make it inaccessible for further biometric recognition.

The key factors that affect the effectiveness of authentication by a method based on analysis of pitch statistics are the following:

–speech volume of the material under research and the volume of speech material in the voice and speaker samples (system performance is noticeably higher if a voice duration recording is sufficient to obtain reliable pitch statistics);

– the presence (both for the subjects and for video, phonograms with samples of oral speech) obtained with the same emotional state of the speaker);

– the presence (both for the investigated and for video, phonograms with samples of oral speech) obtained with the same background environment (when the influence of the environment is so large that the speaker's speaking style changes);

– signal-to-noise ratio;

– no reverb on the records.

Method of spectral-formant analysis The method of identification of speakers based on spectral-formant features compares the studied natural speech recordings with similar samples from the speech database of standards by analyzing the position of formants.

Main stages of work:

– reprocessing of an incoming audio file, including: deleting pauses; normalization per channel; constructing speaker models using formant vectors as input features; construction of SVM (Support Vector Machine) model of speakers;

– Comparison of speaker models.

konkursa NIST SRE 2010] // Trudy 20-j mezhd. konf. po komp'juternoj grafike i zreniju «GrafiKon'2010». Sankt-Peterburg, 2010 [in Russian].

The algorithm of the speaker identification method based on the comparison of spectral-formant representations consists of the following blocks:

– normalization per channel;

– calculation of speaker identification models;

– calculation of individual decision thresholds "own / alien" by comparing the obtained identification model with a standard set of reference models of consciously "alien" speakers;

- direct comparison of the obtained identification model with the model from the database and decision *friend or foe* according to individual thresholds and given probabilities of false alarm and omission of the target.

The method of analysis of a mixture of Gaussian distributions.

Currently mixtures of Gaussian distributions (MGD) are one of the main approaches in solving the problem of voice biometric authentication.

The speaker's Gaussian blend model provides a plausible model of the main sounds contained in the speaker's speech.

Linear combination of Gaussian functions is used to represent the large volume of experimental distributions as a basis. The main advantage of this method is the possibility of forming smooth approximations of the experimental distributions of the acoustic space components, the shape of which can be arbitrary.

The main difficulty in solving the problem of biometric authentication for systems based on the SGR method is to offset the impact of the mismatch introduced by the interference contained in the channel used during recording.

The reasons for this discrepancy may be: environmental noise during recording, distortion of the recording and transmission channels of speech signal, variability of voice of the speaker over time.

The channel, in this case, means the following set of effects: distortions made by the recording equipment; the influence of the microphone of the device used to obtain the individual voice biometric characteristics of the speaker; the influence of the frequency response of the connection channel.

The speaker's Gaussian blend model is convenient for modeling the speaker's voice characteristics, sound recording channel, and environment.

Each of the components of the model reflects some common, but unique for each speaker features of voice. That is why this approach can be successfully used to address the problem of speaker identification.

In order to build a model of the speaker, it is necessary to accurately estimate its parameters, which most closely correspond to the distribution of the vectors of features of the educational expression.

There are a number of methods for estimating model parameters. One of the most popular and well-established is the maximum likelihood estimation method.

The purpose of estimation by this method is to determine the model parameters that will maximize the likelihood of the model when given data for training.

The model is an efficient algorithm that enables identification with high accuracy of recognition. However, there are a number of problems associated with selecting the number of components of the model and initializing its initial parameters.

Joint factor analysis method[12].

This solution is to use Joint Factor Analysis (JFA), which allows a separate announcement of the speaker to effectively separate the channel information from the speaker information.

This allows you to build a channel-independent GMM model of the speaker's speech and suppress the effects of the channel in the audio data, which builds the speaker model.

Full Variability Matrix Method.

One solution to the large size of speaker models is to use low-dimensional feature vectors.

Thus, in the modified version of JFA, the Total Variability (TV) matrix is used to generate feature vectors.

This modified version of JFA is often referred to as the TV's auto speaker recognition method.

This method is widely accepted as the most promising method of speaker recognition, which allows the large-scale input data to be reduced to a low-dimensional feature vector, while preserving most of the useful information.

This reduces the size of the speaker model to 2-3 kbytes, which is often acceptable when solving the problem of speaker identification for a large collection or construction of a system that requires the transmission of speaker models via slow communication channels.

Probabilistic linear discriminant analysis method[14].

The second solution to the mismatch problem is to modify the JFA method, which contains an additional operation based on probabilistic linear discriminative analysis (PLDA).

This allows you to more effectively offset the effect of channel distortions when solving the problem of voice recognition of the speaker.

---

[12] Shhemelinin V. L. Metodika i kompleks sredstv ocenki jeffektivnosti autentifikacii golosovymi biometricheskimi sistemami [Methods and a set of means for assessing the effectiveness of authentication with voice biometric systems] : dis. … kand. tehn. nauk. Sankt-Peterburg, 2015. S. 22 [in Russian].ации голосовыми биометрическими системами : дис. … канд. техн. наук. Санкт-Петербург, 2015. С. 22.

[14] Zakharov V. P., Rudeshko V. I. Biometrychni tekhnolohii v KhKhI stolitti ta yikh vykorystannia pravookhoronnymy orhanamy [Biometric technologies in the 21st century and their use by law enforcement agencies] : posib. Lviv, 2015. S. 492 [in Ukrainian].

Dynamic Time Warping (DTW) Timeline Transformation Method[15].

This method allows you to find the closeness between two measurement sequences over a period of time.

This method was first used in speech recognition to determine how two language signals represent the same original spoken phrase.

In the general case, these sequences can be of different lengths, and measurements can be made at different speeds. The main advantage of the DTW algorithm is its ease of implementation.

However, this algorithm is not suitable for solving the problem of text-independent speaker identification.

A special place of identification of a person by voice occupies at investigation of crimes.

Defining such characteristics by speaker's voice as gender, age, nationality, dialect, emotional coloring of speech are also important in the field of forensic and anti-terrorist activities.

Identification results are important when conducting video, audio recording, expert forensic research based on the theory of forensic identification.

Much research has been done in the area of speech recognition, but so far, speech recognition systems are not one hundred percent accurate.

In the future, speech recognition systems should be free from these restrictions.

The growth of the global voice recognition market depends on many factors.

One of the main factors is the increasing demand for voice biometrics services.

The high demand for voice biometrics, which is unique to any person, is crucial in establishing a person's identity.

Biometrics will continue to be increasingly used in access control systems, in particular in various educational and medical institutions, and biometrics will increasingly become identification cards, various documents and voting systems.

Biometrics will make you feel safer because its main task is to fight crime, terrorism, a way of preserving personal information, and improving the comfort of citizens.

User-friendliness, simplicity, ability to integrate with other methods of identifying individuals are also significant factors that confirm the feasibility of using voice technologies in biometric systems, as separate

---

[15] Zb. nauk.prats Tsentru voienno-stratehichnykh doslidzhen Natsionalnoho universytetu oborony Ukrainy imeni Ivana Cherniakhovskoho [Coll. Scientific Work of the Center for Military-Strategic Studies of Ivan Chernyakhovsky National University of Defense of Ukraine]. 2018. № 3(64). S. 146. URL : https://nuou.org.ua/assets/journals/zbirnyk_cvsd/zb–3–64–2018.pdf (data zvernennia 08.05.2019 [in Ukrainian].

from other methods of verification and identification, and in combination with them.

Conclusions.

This article provides an overview and analysis of existing methods for biometric identification of a person, in particular, voice identification of a speaker.

Based on the above, we can conclude that the development of biometric systems is another way to keep up with the times.

There are currently a number of methods available to solve the speaker's voice identification task, each with its own advantages and disadvantages.

However, the most common method is the Gaussian mixture model. Gaussian mixture models have proven themselves to be a stochastic model for building recognition systems. They are convenient not only for modeling the characteristics of the voice of the speaker, but also for the recording channel, the environment.

The individual components of the model can model a single set of acoustic features.

Each of the components of the model reflects both the general and the voice-specific features of each speaker.

That is why this approach can be successfully applied to solve the problem of text independent speaker identification.

Thus, voice identification of a person in certain circumstances has significant advantages that need to be developed, especially in Ukraine.

Biometrics will make you feel safer because its main task is to fight crime, terrorism, a way of preserving personal information, and improving the comfort of citizens.

References

Biometricheskie sistemyi identifikatsii i autentifikatsii. URL: http://mirznanii.com/a/309729/biometricheskie-sistemy-identifikatsii-i-autentifikatsii (data zvernennia 04.06.2019) [in Russian].

Biometriya ot «A» do «Ya»: polnoe rukovodstvo biometricheskoy identifikatsii i autentifikatsii. URL: https://securityrussia.com/blog/ biometriya.html (data zvernennia 23.05.2019) [in Russian].

Building Synthetic Voices, Festvox, Carnegie Mellon University. URL: http://www.festvox.org/flite/doc/index.html (data zvernennia 08.05.2019).

Entsyklopediia suchasnoi Ukrainy. URL: http://esu.com.ua/ search_articles.php?id=35327 (data zvernennia 06.05.2019) [in Ukranian].

Huang, X., Acero, A., Hon, H. (2001). *Spoken language processing: a guide to theory, algorithm, and system development.* Prentice Hall PTR, 2001.

ISO/IEC 2382–37:2012. (Information technology — Vocabulary — Part 37:Biometrics, IDT) URL: https://www.iso.org/standard/55194.html (data zvernennia 13.05.2019).

ISO/IEC 2382–37–2016 [ISO/IEC 2382–37:2012] (Information technology — Vocabulary–Part37:Biometrics, IDT) URL: https://meganorm.ru/ Data2/1/4293747/ 4293747566.pdf (data zvernennia 13.05.2019).

Koval, S. L. (2006). *Formants matching as a robust method for forensic speaker identification.* SPECOM'2006 Proceedings. XI International Conference «Speech and Computer», (SPECOM'2006), 25–29 June 2006. St. Petersburg, Russia, 125–128.

Matveev, Yu.N., Simonchik, K.K. (2010). *Sistema identifikatsii diktorov po golosu dlya konkursa NIST SRE 2010.* Trudyi 20-y mezhd. konf. po kompyuternoy grafike i zreniyu GrafiKon'2010' Sankt-Peterburg [in Russian].

Shchemelynyn, V. L. (2015). *Metodika i kompleks sredstv otsenki effektivnosti autentifikatsii golosovyimi biometricheskimi sistemami*: diss. na soisk. nauch. stepeni kand. tehn. nauk), (18–23) Sankt-Peterburg [in Russian].

Tsarov, R. Iu., Lemekha, T. M. (2016). *Biometrychni tekhnolohii'*: navch. posib. Odessa [in Ukranian].

Zakharov, V. P., Rudeshko, V. I. (2015). *Biometrychni tekhnolohii v XXI stolitti ta yikh vykorystannia pravookhoronnymu orhanamy*: posib, Lviv [in Ukranian].

Zbirnyk Naukovykh prats Tsentru voienno-stratehichnykh doslidzhen Natsionalnoho universytetu oborony Ukrainy imeni Ivana Cherniakhovskoho. URL: https://nuou.org.ua/assets/journals/zbirnyk_cvsd/zb–3–64–2018.pdf (data zvernennia 12.06.2019) [in Ukranian].

Х. В. Луценко, К. В. Нікулін

ГОЛОСОВА ІДЕНТИФІКАЦІЯ ДИКТОРА ЯК ОДИН ІЗ СУЧАСНИХ БІОМЕТРИЧНИХ МЕТОДІВ ІДЕНТИФІКАЦІЇ ОСОБИ

У статті розглянуто найбільш розповсюджені біометричні системи ідентифікації осіб, у тому числі голосової ідентифікації диктора на відео-, звукозаписах.

Приведені галузі використання біометричних технологій та їх загальні характеристики. Наведений огляд використання груп ознак при ідентифікації, які характеризують голос.

Розглянуто основні переваги голосової біометрії, такі як простота реалізації системи; низька вартість (найнижча серед усіх біометричних методів); вона не потребує контакту і дозволяє здійснювати верифікацію на великій відстані на відміну від інших біометричних технологій.

Проведено аналіз існуючих методів розпізнавання мовної інформації, які ідентифікують особу за сукупністю унікальних характеристик голосу.

Ключові слова: біометричні технології ідентифікації, голосова ідентифікація диктора, верифікація особи, автоматичні методи розпізнавання диктора, експертиза відео-, звукозапису.