# Building of a Speaker's Identification System Based on Deep Learning Neural Networks

Victor Solovyov [1], Oleg Rybalskiy [2], Vadim Zhuravel [3], Aleksandr Shablya [4], Yevhen Tymko [5]

[1] *"Silentium Systems", Hornby Str 950 – 777, BC V6Z 1S4, Vancouver, Canada,*
[2] *National Academy of Internal Affairs, Solom'yanska Are. 1, Kyiv, 03035, Ukraine*
[3] *Kyiv scientifically-research expertly-criminalistics center Ministries of Internal Affairs of Ukraine, Vladimirska Str. 15, Kyiv, 01001, Ukraine*
[4] *Odessa Research Institute of Forensic Expertise of the Ministry of Justice of Ukraine, Uspenska Str. 83/85, Odesa, 65011, Ukraine*
[5] *Kyiv Scientific Research Institute of Forensic Expertise of the Ministry of Justice of Ukraine, Smolenska Str. 6, Kyiv, Ukraine*

**Abstract**

Paper discusses the main methodological and technological features of studying and building of speaker's identification systems built on the basis of deep learning neural networks.
The problems and tasks arising in the process of creation of such systems are considered. The purpose of the research article is to show the suggested ways and methods for their elimination and solutions, which were found in the process of creation of an automated system for speaker identification and verification, built due to the use of such networks.
In the process of its creation, the method of spectral analysis of speech signals at short time intervals was determined, which ensures high resolution. In addition, a solution of the problem of invariance of the system to different language groups and the duration of phonograms is proposed. This ensured the generality and efficiency of the obtained results of speaker's identification.
As a result, an automated system for forensic identification and speaker's verification was developed on the basis of deep learning neural networks. In the process of the development of a system based on the comparison of the spectral characteristics of speech signals, several methods have been proposed and tested providing the possibility of identification (verification) of the speaker by speech messages of short duration.

**Keywords**

phonogram duration, forensic identification of speaker, deep learning neural network, spectral analysis, frequency domain, efficiency.

## 1. Introduction

The use of modern technologies of neural networks for the examination of materials and digital sound recording equipment allows, as a rule, to obtain a more higher level of its efficiency [1,2]. Usually, the efficiency of the system is taken as a quantity determined by the probability of errors of the first and second kinds, inherent in mentioned type of examination.

According to the materials of the SRE NIST tests carried out recently, the point of cross of the graphs of errors of both the first and second kinds for the best automated speaker identification systems is on average (3–10)%. At the same time, several tests of systems on messages of less than 10 s duration are carried out relatively rarely [3].

*ISIT 2021: II International Scientific and Practical Conference «Intellectual Systems and Information Technologies», September 13–19, 2021, Odesa, Ukraine*
EMAIL: edemsvi@gmail.com (A. 1); rov_1946@ukr.net (A. 2); fonoscopia@ukr.net (A. 3); alik.shablya@gmail.com (A. 4); e.tymko@kndise.gov.ua (A. 5)
ORCID: 0000-0003-3541-4740 (A. 1); 0000-0002-1440-8344 (A. 2); 0000-0002-2777-2102 (A. 4);

Similar data are provided by the materials of the working group of forensic speech and audio analyzes of the European Network of Forensic and Scientific Institutions [4].

It is generally accepted that the effectiveness of automated speaker's identification systems is significantly lower than the effectiveness of, for example, fingerprinting, video or DNA recognition.

Paper presents the results of research and development of a speaker identification system based on spectral characteristics of voice signals and on deep learning neural networks, which, we believe, can change this point of view.

During past two decades, the number of noteworthy publications related to technical systems for voice identification numbers in the thousands. Therefore, we will consider only the main methodological and technological features of research and design of such systems that are directly related to the problems and tasks solved by the conducted developments.

Despite the hundreds of different methods and algorithms for speaker's identification based on the physical characteristics of voice signals various spectral parameters are generally recognized as physical bases. This applies to both classical methods and methods based on neural networks. But the use of various characteristics and parameters of signals with the use of classical spectral methods allows only to improve slightly the efficiency of identification, since any of them is based on a discrete orthogonal Fourier transform.

Modern research in the field of neurophysiology of hearing indicates the feasibility of spectral analysis of speech signals at short time intervals [5]. In particular, a number of important applications of processing audio information at short time intervals have already become classics, for example, when compressing audio files. Thus, the basic basis for the majority of audio file compression formats is the transformation of signals from the time domain to the frequency domain at short time intervals (16–20) ms [5]. But the discrete orthogonal Fourier transform at small time intervals (about 20 ms) has insufficient frequency resolution from the point of view of the neurophysiology of hearing. So, for a time interval of 20 ms, the use of such a transformation for the transition from the time to the frequency domain, provides a frequency resolution of 50 Hz. However, it is known from the neurophysiology of hearing that the resolution of human hearing is approximately 1 Hz [5].

As it will be described below, such a low frequency resolution at short time intervals significantly reduces the efficiency of any speaker identification systems and is one of the determining factors in the spectral analysis of audio information. And the modern practice of expert examination points to serious problems of speaker identification for phonograms of short duration [3,4].

Another important problem, in our opinion, is that any methods and algorithms for conducting an examination within the framework of this methodology relate to certain parameters of sound signals in the frequency domain, from which some, as the most important, are selected by an expert. For example, the frequency of the main tone, the spectrum of specific sounds, etc. are compared. In this case, all comparisons are made for integrative assessments obtained as a result of averaging the spectral parameters over the entire duration of the phonogram. This approach, taking into account numerous accompanying factors and their variability, often does not provide a high generality and efficiency of the obtained results. An important property of almost all known approaches, including those based on the use of deep learning neural networks, is also the difficulty in achieving common results for large, gradually growing databases. Such databases form the bases of Big Data arrays used to train neural networks. But in most cases, the plots of errors of the first and second kinds, used to check the quality of network training, will be bound to the DataSet obtained from a specific training material. As a result, a large degree of data generalization requires, as a rule, repetition of research and calculations for a new training set.

This raises the problem of the correct quantitative assessment of the effectiveness of the forensic identification of the investigated object. The construction of graphs of the probability of errors of the first and second kinds, in our opinion, is the most informative and therefore the most preferable option for determination of the magnitude of such errors.

The use of deep learning neural networks allows us to consider the possibility of developing an effective automated universal system designed for forensic identification of a speaker. Under the universality of the system, we consider its suitability for work with the speech of speakers speaking different languages, belonging to different sexes and with phonograms of different duration (including several seconds).

Thus, in order to solve the problems that exist in the construction of an automated system for forensic identification of a speaker, we define the following tasks:

to determine the method of spectral analysis of speech signals at short time intervals, providing high resolution;

to propose a solution to the problem of invariance of the system to different language groups and the duration of phonograms. This will ensure the generality and effectiveness of the results obtained for the identification of the speaker.

The purpose of this paper is to show the ways and methods of solving these problems by use of the example of the results obtained during research and development of an automated system for identification and verification of a speaker (the "Avatar" system [6]).

## 2. Ways and methods of solving the tasks

Let us consider a discrete non-orthogonal time-frequency conversion for a 20 ms time window with a signal of the audiofrequency range. To be specific, we will use the Morlet wavelet with the basis

$$C_{mor}(t)=\pi F_b^{-1/2}\times e^{j2\pi F_c t}\times e^{-t^2/F_b}, \quad (1)$$

where

$j=\sqrt{-1}$ – imaginary unit,

t – time,

$F_b$ – wavelet width parameter,

$F_c$ – wavelet center frequency [7].

In this case, we will consider redundant transformations, in which the number of samples in the time domain falling on the selected area is less than the number of samples in the frequency domain. So, for example, let's take an arbitrary 20 ms fragment of the speech signal of sound [A] with a sampling rate of 44100 Hz. Then the number of discrete samples falling on a 20 ms segment is $N$ = 882. Let us construct and compare two types of time-frequency conversion in the frequency range from 0 to 2500 Hz for the same signal segment. The first of them is non-orthogonal based on the Morlet wavelet with a frequency step $D_{Fc}$ = 1 Hz [7]. The total maximum possible number of frequency steps in the selected range is 2500. The second is orthogonal with a discreteness of $D_{Fc}$ = 50 Hz (in accordance with the duration of the time window). The total maximum possible number of frequency steps in the selected range is 50.

Fig. 1 shows an illustration of a comparison of the spectra of one signal fragment obtained for two types of transformations.

Visually, these graphs are very close. However, the difference in the positions of the local maxima of the spectra for applied examination problems is very significant, since in most methods for identifying speakers an important factor is the value of the frequencies of such maxima [8]. As it is seen in Fig. 1, the values of the frequencies of the local maxima for orthogonal and non-orthogonal transformations of the same signal differ by more than 20 Hz. This circumstance significantly affects the accuracy of the assessment of the spectral parameters of speech.

It is known that when averaging any function over a large number of time windows with a duration of T = 20 ms, the calculation accuracy is proportional to the square root of the number of window transformations [7].

But this means that to achieve an accuracy of 1 Hz, obtained with a non-orthogonal transformation on an interval of 20 ms, with orthogonal transformations taking into account averaging, 400 × 20 ms = 8 sec are required.

This shows the practical impossibility of analyzing phonograms of short duration (several seconds) by use of conventional methods of time-frequency transformations, which is confirmed by the modern practice of expertise [3,4].

At the same time, as it will be shown below, the use of non-orthogonal time-frequency transformations with a higher resolution in the frequency of localization of maxima (of the order of 1 Hz) significantly increases the accuracy and efficiency of speaker's identification by use of phonograms of short duration.

The general concept of the developed speaker identification system is based on the data of classical studies in the neurophysiology of hearing [5]. One of the important factors in the identification of a speaker by human auditory analyzers are the individual characteristics of vowel sounds [8]. Therefore, when designing the system, a separate basic module was developed for automatic extraction of vowel sounds from speech phonograms. The methodology for its development is based on deep study of neural networks.

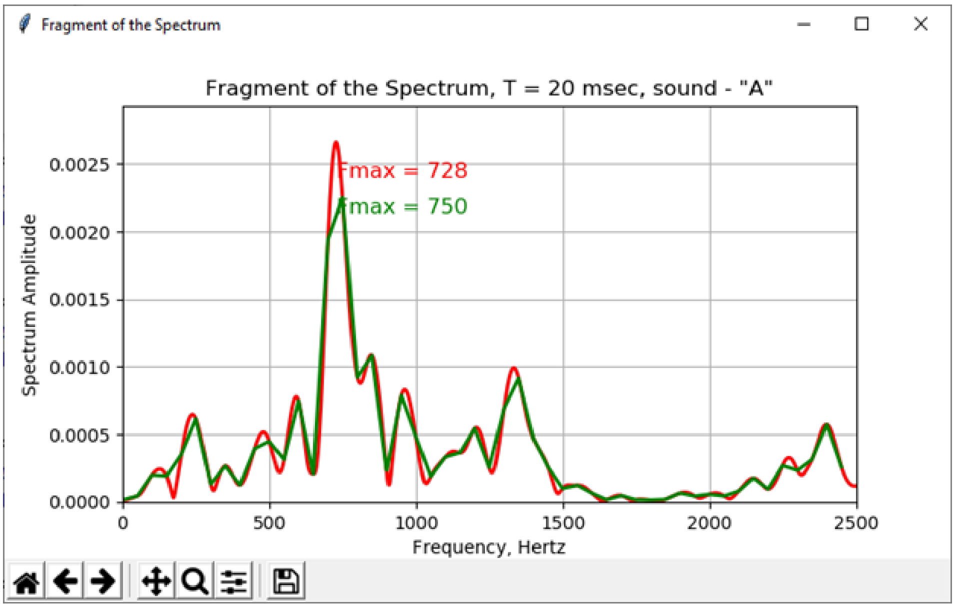


**Figure 1**: Compared spectra of one fragment of sound [A] (T = 20 ms)

The speaker's identification technology used in the system is based on the automatic determination of the proximity of the spectral characteristics of two vowel sounds – [A] and [I], isolated from two different phonograms. At the same time, the proximity of the characteristics of two fragments of vowel sounds in phonograms is determined on the basis of a special model created on the basis of a deep learning neural network. Let's consider some fundamental features of this technology.

In the vowel extraction module for each of the two phonograms, arrays of fragments of sounds [A] and [I] with a duration of 20 ms are formed. Further, for all fragments, a non-orthogonal Morlet wavelet transform is implemented with a frequency resolution of local maxima of 1 Hz. The fragments are converted in the frequency range from 0 to 12000 Hz. Changing the value of the upper frequency of the range makes it possible to study the performance of neural networks for their various configurations and structures.

The obtained spectra are normalized by dividing the amplitude of each spectral peak by the sum of the amplitudes of all spectral peaks for the entire frequency range

$$A_{iN} = \frac{A_i(f_i)}{\sum A_i(f_i)}, \qquad (2)$$

where

$A_{iN}$ – normalized amplitude of the spectral component selected at the i-th scanning step,

$A_i(f_i)$ – the amplitude of the spectral component selected at the *i*-th scanning step.

Then one of the classic technological approaches is used for further forming of the DataSet. Arrays of fragments consisting of various combinations of two spectra are formed from a set of phonograms with the speech of different speakers. In this case, for fragments of the spectra of the same speaker, marking "he" is used, and for different speakers – "not he". Arrays of a combination of spectra and a separate array with a priori known labeling are the basis of the DataSet for training the neural network.

Thus, to transform fragments in the frequency range from 0 to 12000 Hz, the training DataSet will be composed of fragments containing 24000 spectral amplitudes in various combinations. It should be noted that for orthogonal transformations, the number of frequencies is only 480, and in this version, training of a neural network is a more simple task. Our studies of the learning process of neural networks with different structures for a DataSet with fragments containing 24000 frequency amplitudes, in practice, showed the problematicness of obtaining effective results for most of the known structures of neural networks. As the analysis showed, the main reasons are known factors – gradient attenuation and retraining. In this the most general statement the structure of a deep learning neural network based on convolutional networks was used to solve the problem in effective way. It should be noted that parallel studies are conducted on the basis of fully connected networks. For these structures a preliminary selection of 50 local maxima of the spectrum (normalized and then ranked according to the magnitude of the amplitude) was applied. For this variant of neural structures, results were obtained with less efficiency.

The keras library (bakend tensorflow) was used to train the neural network. The created speaker identification model is actually intended to solve the binary classification problem. Fragments of the neural network training process are shown in Fig. 2

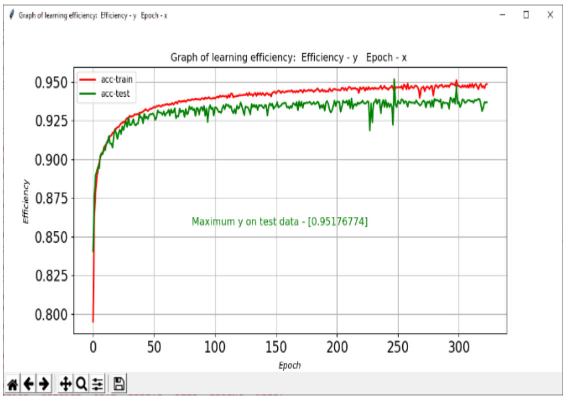


**Figure 2**: Graph of the effectiveness of training a neural network

By use of the developed model, it was possible to obtain a high efficiency of speaker identification from the point of view of expert examination practice. But the practical implementation of the identification process requires much time to calculate phonograms. In addition, the identification model built by the neural network is a “black-box”. It is not possible to establish the causal relationships that determine the speaker's identification. At the same time, the implementation of software systems into the practice of examination requires an “internal conviction of an expert” in the correctness of made decisions. This conviction can only be based on well-known classical ideas about the characteristics of sounds and speech.

Therefore, in the process of designing a speaker’s identification system on the basis of the parameters of voice signals (the “Avatar” system [6]), an approach was formed on the basis of classical concepts associated with a neural network model.

Analysis of various phonograms showed that for fragments of spectra with a high probability of speaker’s identification (above 0.999), the spectra of 20 ms fragments are very close. So, Fig. 3 shows the spectra in the range 0 – 2500 Hz for the same speaker (sound – [A]) with the probability of correct identification 0.9991.

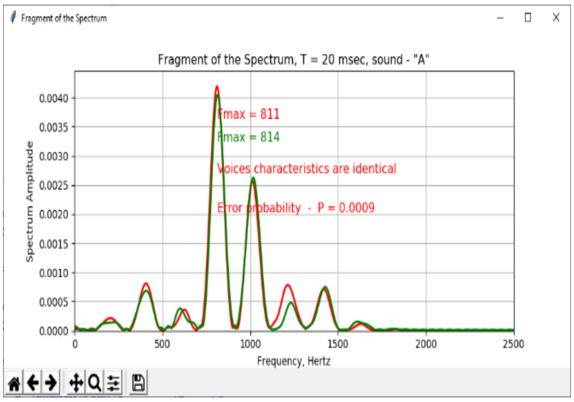


**Figure 3**: Spectra of similar characteristics of voices

As it can be seen from the spectra of the same two fragments, considered in the frequency band from 0 to 4000 Hz, both local maxima and formant features of the [A] sound practically coincide (Fig. 4).

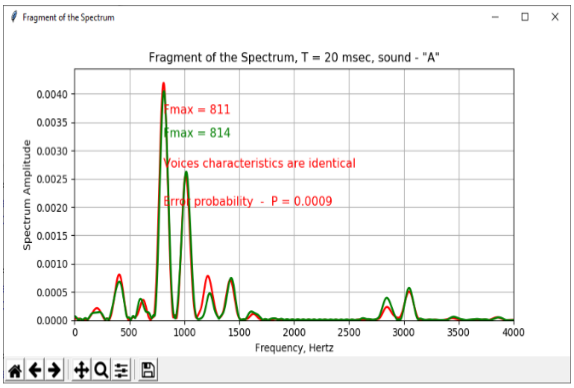


**Figure 4**: Spectrum of similar characteristics of voices with formant signs

In accordance with the concept of the adopted approach to the practical implementation of the “Avatar” system with a good approximation in terms of efficiency, a classical heuristic criterion for the proximity of two spectra was introduced - the sum of the absolute values of the difference between the normalized amplitudes of the spectra of the sounds [A] and [I].

$$B = \sum_{i=1}^{N} |A_1 - A_2|, \tag{3}$$

where

$A_1$ – the value of the averaged amplitude of the spectrum of the first phonogram,

$A_2$ – the value of the averaged amplitude of the spectrum of the second phonogram,

$N$ – the number of frequencies in a given frequency range.

The smaller this value, the closer the characteristics of the voice of the speaker and almost any function of the spectrum accordingly. It includes the frequency of the main tone and formant features.

The classical approach to the assessement of the effectiveness of forensic identification of any object, including the speaker, is to determine the magnitude of errors of the first and second kinds. At the same time, plotting such errors is the most preferable option for finding out the levels of such errors. From the physical prerequisites of the tasks of identification (and verification) of a speaker, it is known that identification errors significantly depend on the duration of sound phonograms. In addition, it is generally accepted that the specific characteristics of a language and language groups should affect the effectiveness of identification. Both of these factors were taken into account in the studies and in the implementation of the system under consideration. In particular, the curves of errors of the first and second kind were built for a mixture of speech messages of different speakers, made in different languages – English, Chinese, Russian and Ukrainian. As well as messages made separately in English, Russian and Ukrainian. The dependence of the magnitude of errors on the duration of phonograms was also determined. Thus in Fig. 5 and Fig. 6 graphs of errors of the first and second kinds for several variants of their construction have been shown.

Important "technological" factors should be noted, which, due to the insufficient mass use of systems for automatic identification of speakers, are practically not represented in scientific publications today. At the same time, these factors play a very significant role in the practice of expertise (including in its legal aspects).

The first factor is that the probability of identification error in the studies under consideration can be arbitrarily small. Including less than 0.00001 (0.001%). It is possible under provided that the fragments of the spectra of sounds in two phonograms are very close on average. And this can be observed in real practice of examination. So, in Fig. 7 shows an illustration of identification based on two phonograms with a duration of approximately 60 seconds from the same speaker.

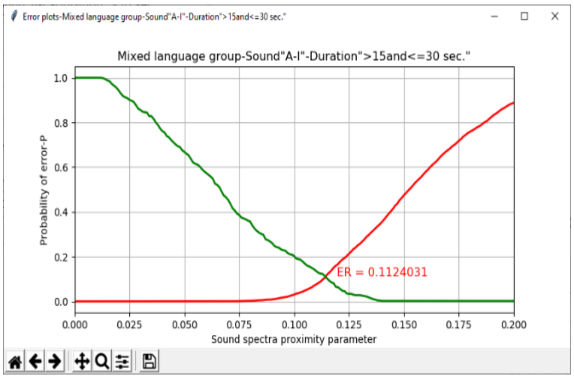


**Figure 5**: Error graphs for mixed language group

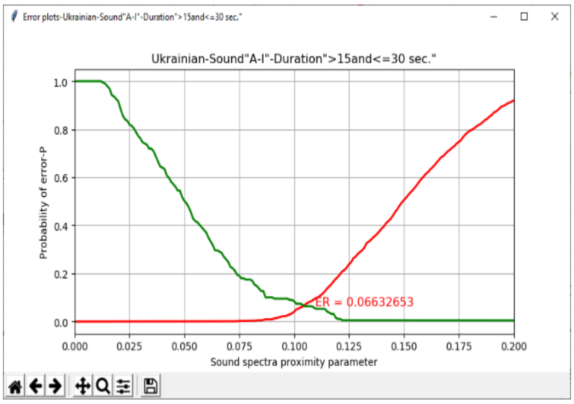


**Figure 6**: Graphs of errors for the Ukrainian language

The most problematic from the point of view of making decisions on identification (for any systems) are phonograms, in which the analysis results give close values for errors of the first and second kinds. So in Fig. 8 a similar illustration is shown. The decision made on the basis of information about errors of the first and second kinds when the calculated values are close to the point of cross of the curves is less justified.

Obviously, at the point of cross of the curves of errors of the first and second kinds, the hypotheses "he" and "not he" are equally probable. At the same time, the probabilities of errors of the first and second kinds on the graphs, although the same, are not great. Such a contradiction is a consequence of the incorrect methodology for evaluating errors from the point of view of the practice of interpreting the probabilities of errors in these variants. But, due to the experience of experts, it should be noted that the expert practice is the area of most common errors.

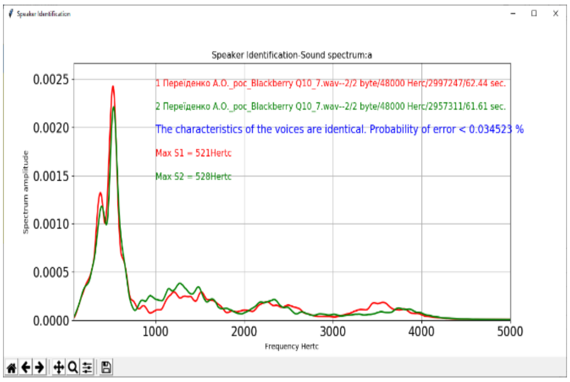


**Figure 7**: Illustration of speaker's identification by the total spectrum of sound [A].

When using integral estimates (for example, when comparing the averaged spectra of the fundamental tone for the sound [A], allocated along the entire length of each of the phonograms), the method of setting the error probability threshold for decision making is usually used. As a rule, this threshold is determined by the assessment of the probability at the point of cross of the graphs of errors.

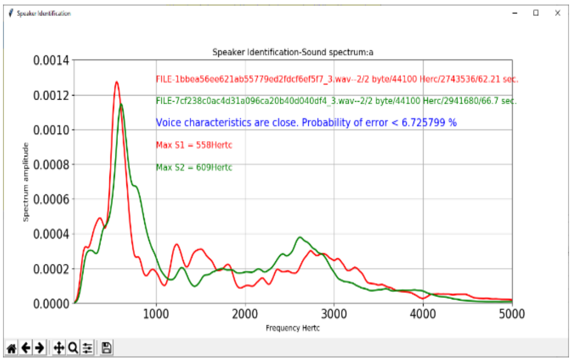


**Figure 8**: Illustration of speaker's identification provided that the calculated values are close to the intersection point of the error curves of the first and second kind

The peculiarity of its application is that, regardless of the characteristics of such graphs obtained in this case, the decisions made on the basis of a given threshold are always subjective. However, this methodology is generally recognized in the practice of the probabilistic approach to decision making [9]. We believe that the use of such a threshold, if the obtained value of the probability of an identification error is close to it, has a too high degree of subjectivity. It is advisable to use it in practice only for contrasting values of the error probability. There is a significant difference between the probability of error for a specific measure of the proximity of two spectra, and the probability of errors of the first and second kinds at the point of intersection of their graphs.

At the same time, in the presence of large arrays of fragments with a duration of 20 ms, used to identify the speaker, provided that the above mentioned binary approach to decision-making is applied ("he" is "not he"), it is possible to build a less subjective approach.

In the developed system, a slightly different approach is applied to the calculation of the probabilities of identification errors for phonograms of short duration. It is due to the technological features of the adopted model.

An array of fragments of speaker identification by two phonograms is considered. For phonograms with a duration from 1 sec. to several minutes these are arrays of spectra ranging in number from several hundred to tens of thousands.

It should be mentioned, that in the model under consideration, identification is carried out by separate fragments, consisting of combinations of vowel sound spectra extracted from two phonograms. The output of such a model is the probabilities of correct identification ("he" – "not he") for each pair of compared fragments, determined by the measure of the proximity of their spectra. These probabilities are the collections of discrete random variables.

Then the statistical average of the probability of correct identification can be calculated as the average over the entire array using the well-known formula

$$m_x = \frac{1}{N}\sum_{i=0}^{N} x_i \,, \qquad (4)$$

where

$x_i$ – the assessment of the value of the probability of correct identification for each *i*-th fragment,

*N* – the number of averaged fragments [10].

Due to this approach, the probability of an error in decision-making is determined by the statistical averaged mx and the statistical standard deviation (RMS) from the averaged one, defined as

$$S = \sqrt{\frac{\left(x_i - m_x\right)^2}{N-1}} \,, \qquad (5)$$

Since averaging uses a huge number of fragments, which are used to determine the statistical averaged and statistical standard deviation when comparing two phonograms, they are subject to the normal distribution law, and we can use the corresponding probability density distribution graph to determine the probability of errors when making a decision.

The decision error for the totality of the array of fragments with this approach is the sum over the probability density distribution in the range from

$0 \leq P \leq 0.5$ (Fig. 9). The illustration in Fig. 9 is given for the variant of the statistical average number of the probability of correct identification mx > 0.5. For mx < 0.5, the decision error is the sum over the probability density distribution in the range from $0.5 \leq P \leq 1$.

In Fig. 9 the average probability of identification is $P = 0.6$, the number of fragments is 894, the standard deviation of the probability of identification by fragments is $S = 0.21$.

The accepted approach to the assessment the probabilities of identification errors contains only calculated (experimental) parameters of fragments of the identification array for two phonograms. In particular, this $m_x$ is the average value of the identification probability for the entire array of identification fragments, $N$ is the number of fragments over which the averaging was carried out, $S$ is the standard deviation of the statistical mean of the probabilities of correct identification. These parameters completely determine the identification errors. In this case, the dependence of the identification efficiency on the duration of the phonograms is automatically taken into account, which is determined by the value of the parameter $N$.

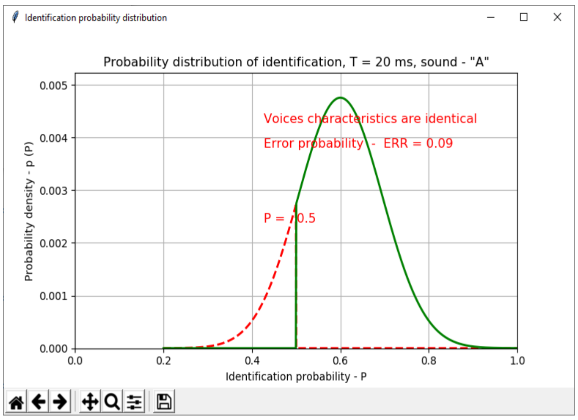


**Figure 9**: The density of the probability distribution of correct speaker’s identification for an array of fragments

Depending on the language of the speaker's speech recorded on the phono-gram, the parameters $m_x$ and $S$ will change. So, for tonal languages (for example, Chinese), due to the greater variability of the characteristics of vowel sounds, $S$ will increase, which, in turn, will increase the errors identification (this statement is true for the characteristics of any tonal speech).

Significant computational complexity is the disadvantage from the point of view of the implementation of this probabilistic approach in an automatic identification system is. This approach requires, for example, twice more computational time than the described above heuristic approach based on the comparisons of the averaged spectra of vowel sounds. The second important factor is the “non-standard approach” when making expert decisions on speaker identification.

## 2.1. Results and discussion

In the process of the development of any speaker’s identification system, the issue arises of the applicability of the system and the corresponding methodology to various language groups. The account of the dependence of the identification efficiency on the duration of phonograms, as well as the dependence of the identification efficiency on specific algorithms.

From the point of view of eliminating dependence on various algorithms, the approach based on models of deep learning neural networks is the most general one. But it works under the condition that all information is supplied to the input of the neural network. At the same time, an issue arises that determines the completeness of the model's coverage of various factors. In particular, whether a particular DataSet can cover most of the listed above factors.

Our studies indicate a high probability of covering most of the factors in the developed approach. In particular, with the number of speakers over 15 (male, female voices) and several language groups, the results practically do not change with an increase in the number of speakers in the DataSet.

Another important factor is the methodology for work with a set of 20 ms phonogram fragments, which ensures that there are practically no parameters in the system that are selected by an expert and therefore have a subjective character. This makes it possible to uniformly solve identification (and verification) problems regardless of the duration of phonograms and language groups.

We believe that the developed identification methodology has high versatility in the above mentioned sense.

## 3. Conclusions

An automated system for forensic identification and verification of the speaker based on deep learning neural networks has been developed. In the process of developing a system based on the comparison of the spectral characteristics of speech signals, methods have been suggested and tested that provide the possibility of identification (verification) of the speaker by voice messages of short duration.

## 4. References


[1] Solovyov V. I., Rybalskiy O. V., Zhuravel V. V. Verification of fun-damental fitness of neuron networks of the deep educating for the construction of the system of exposure of editing of digital phonograms. Cybernetics and Systems Analysis, Vol. 56, No. 2, March, 2020, pp. 326¬–330. doi: 10.1615 / 10.1007 / s10559-020-00249-2.

[2] Solovyov V.I., Rybalskiy O.V., Zhuravel V.V. Method of exposure of signs of the digital editing in phonograms with the use of neuron networks of the deep learning. Journal of Automation and Information Sciences, 2020, Vol. 52, No. 1, pp. 22–28. doi: 10.1615 / JAutomatInfScien.v 52.i1.30.

[3] NIST USA, SRE. Available at: https://www.nist.gov/itl/iad/mig/nist-2019-speaker-recognition-evaluation (accessed 24 August 2020).

[4] ENFSI working group for forensic speech and audio analysis. Availa-ble at: http: //www.Enfsi.eu/about-enfsi/structure/working-groups/speech-and-audio (accessed 24 August 2020)

[5] Alexandrova Yu.I. Psychophysiology. M.-S.P .: Nauka, 2006, 463 p.

[6] Solovyov V.I., Rybalskyi O. V., Shablya A. N., Zhuravel V. V. System of automated search for votes. Informatics and Mathematical Methods in the Model. 2015, Vol. 5, No. 4, pp. 302–307.

[7] Mallat S. A wevlet tour of signal processing. Academic Press, New York. 1999, 671 p.

[8] Bondarenko M.F., Driuchenko A.Ya., Shabanov-Kushnarenko Yu.P. Vowel sounds in theory and experiment. Kh .: Khark. nat. un-t radioelektroniki, 2002, 348 p.

[9] Rybalskiy O.V., Solovyov V.I., Cherniavskyi S.S., Zhuravel V.V., Zheleznyak V.K. A probabilistic approach to making expert decisions on the analysis of complex objects. Bulletin of the National Academy of Sciences of Belarus. A Series of Physical and Technical Sciences. 2019, Vol. 64, No. 3, pp. 346–352. doi: 10. 29235 / 15-8358-2019-64-3-346-352.

[10] Handbucher der industriellen Messtechnik / 3 Auflage/ Herasgeber prof., Dr., P. Profos / 1 Auflage. Vulkan-Verlag, Essen. 1984, 491p.