Digital Content Processing Method for Biometric Identification of Personality Based on Artificial Intelligence Approaches

Eugene Fedorov[1][0000-0003-3841-7373], Tetyana Utkina[1][0000-0002-6614-4133], Kostiantyn Rudakov[1][0000-0003-0000-6077], Andriy Lukashenko[2][0000-0002-6016-1899], Serhii Mitsenko[1][0000-0002-9582-7486], Maryna Chychuzhko[1][0000-0001-5329-7897], Valentyna Lukashenko[1][0000-0002-6749-9040]

[1] Cherkasy State Technological University, Cherkasy, Ukraine
{t.utkina, ckc, k.rudakov, s.mitsenko, m.chychuzhko}@chdtu.edu.ua, fedorovee75@ukr.net
[2] E. O. Paton Electric Welding Institute, Kyiv, Ukraine
ineks-kiev@ukr.net

Abstract. The paper suggests a method for processing digital content for biometric identification based on artificial intelligence approaches. To get the goal the methods of forming digital content characteristics, creating a structure model of a system for processing digital content, the method of selecting the structure determination of parameter values of the mathematical model of digital content processing system are suggested. The suggested characterization of digital content automates the processing of digital content which increases the accuracy and speed of determining the values of signs. The suggested creation of a model structure of a digital content processing system provides knowledge in the form of easily accessible for human understanding rules that simplifies the process of determining the structure of the system and also allows parallel processing of information that allows increasing the learning speed. The suggested selection of structure method of determining values of model parameters of the processing system of the digital content based on the genetic algorithm uses a combination of directed and random search that decreases the probability of a hit in local extremum and provides an acceptable speed of determining values of the model parameters. The suggested method of digital content processing for biometric identification of a personality by voice can be used in various intelligent digital content processing systems.

Keywords: digital content processing, biometric identification of personality, artificial neural network, fuzzy inference systems, genetic algorithm.

1 Introduction

Human-machine interfaces are one of the directions of digital content processing. For these interfaces, biometric identification of a person is important.

Automated biometric identification of a person means decision making based on acoustic and visual information, which improves the quality of recognition of the person being studied [1-3]. Unlike the traditional approach, computer biometric identification speeds up and improves the accuracy of the recognition process, which is especially critical in limited time conditions.

A special class of biometric identification of a person is formed by methods based on the analysis of acoustic information [4-8].

The methods of biometric identification of a person by voice include: dynamic programming [9, 10]; vector quantization [11, 12]; artificial neural networks [13, 14]; decision tree [15]; Gaussian mixture models (GMM) [16-19]; their combination [20].

Artificial neural networks are the most popular methods.

The advantages of neural networks consist in: the possibility of their training and adaptation; the ability to identify patterns in the data, their generalization, i.e. extracting knowledge from data, therefore, knowledge about the object is not required (for example, its mathematical model); parallel processing of information, which increases the computing power.

The disadvantages of neural networks include: the difficulty of determining the network structure, since there are no algorithms for calculating the number of layers and neurons in each layer for specific applications; the difficulty of forming a representative sample; a high probability of a learning method and adaptation getting into a local extremum; inaccessibility for human understanding of knowledge accumulated by the network (it is impossible to present the relationship between output and output in the form of rules), since it is distributed between all elements of the neural network and is presented in the form of its weighting coefficients.

Recently, neural networks have been combined with fuzzy inference systems.

The advantages of fuzzy inference systems are the following: presentation of knowledge in the form of rules that are easily accessible for human understanding; no accurate assessment of variable objects is needed (incomplete and inaccurate data).

The disadvantages of fuzzy inference systems include: the impossibility of their training and adaptation (parameters of the membership functions cannot be automatically configured); the impossibility of parallel processing of information, which increases the computing power.

Since genetic algorithms can be used instead of neural network learning algorithms for training of membership function parameters, we note their advantages and disadvantages.

The advantages of genetic algorithms for neural networks training are the following: the probability of getting into a local extremum decreases.

The disadvantages of genetic algorithms for neural networks training are the following: the speed of the solution search method is lower than that of neural network training methods; in the case of binary genes, an increase in the search space reduces the accuracy of the solution with a constant chromosome length; in the case of binary genes, there are encoding/decoding operations that reduce the speed of the algorithm.

In this regard, it is relevant to create a method of digital content processing for biometric identification of a person, which will eliminate these drawbacks.

The aim of the work is to increase the efficiency of digital content processing system due to the artificial neuro-fuzzy network, which is trained on the basis of the genetic algorithm.

To achieve this goal, it is necessary to solve the following tasks:

1. Generation of digital content attributes.
2. Creation of a model of digital content processing system.
3. Choice of the structure of the method for determining the parameter values of the mathematical model of digital content processing system.

2 Generation of digital content attributes

The generation of digital content attributes in the case of biometric identification of a person by voice provides for the following steps:

- determination of vocal segments of a speech signal based on statistical estimation of short-term energies;
- definition of formants of the central frame of the vocal segment;
- choice of vocal speech sound attributes based on formants of the central frame of the vocal segment.

2.1 Determination of vocal segments of a speech signal based on statistical estimation of short-term energies

The paper proposes a method for determining vocal segments of a speech signal based on statistical estimation of short-term energies, which includes the following steps:

1. Set a speech signal with one vocal sound $y(n)$, $n \in \overline{1, N^f}$. Set the number of quantization levels of a speech signal L (for an 8-bit sound sample $L = 256$). Set the length of the frame N, on which the short-term energy is calculated, $N = 2^b + 1$, where the integer parameter b is selected from the inequality $b-1 < \log_2\left(f_s / f_{\min}\right) < b$, f_s is the sampling frequency of the speech signal in Hz, $8000 \le f_s \le 22050$, $f_{\min}$ is the minimum frequency of the fundamental human tone in Hz, $f_{\min} = 50$. Set the parameter for adaptive threshold β, $0 < \beta < 1$.
2. Calculate short-term energies

$$E(n) = \sum_{m=-N/2}^{N/2} y^2(m+n), \ n \in \overline{N/2+1, N^f - N/2 - 1}.$$

3. Calculate the mathematical expectation of short-term energies

$$\mu = \frac{1}{N^f - N - 1} \sum_{n=N/2+1}^{N^f - N/2 - 1} E(n).$$

4. Calculate the standard deviation of short-term energies

$$\sigma = \sqrt{\frac{1}{N^f - N - 1} \sum_{n=N/2+1}^{N^f - N/2-1} E^2(n) - \mu^2} \ .$$

5. Calculate the adaptive threshold $T = \mu - \beta\sigma$.
6. Determine the left and right borders of the vocal segment:
   6.1. Set the sample number $n = 1$;
   6.2. If $E(n) < T \wedge E(n+1) \ge T$, then $N^l = n+1$, go to step 6.1;
   6.3. If $E(n) \ge T \wedge E(n+1) < T$, then $N^r = n$, proceed to completion;
   6.4. If $n < N^f - N - 1$, then go to the next sample, i.e. $n = n+1$, go to step 6.2, else $N^r = n$, proceed to completion.

As a result, the left and right boundaries of the vocal segment are determined. For the method of formants determining, the frame with the center in the sample with the number $N^c = round\left(\left(N^l + N^r\right)/2\right)$ is selected as the central frame.

2.2 Definition of formants of the central frame of the vocal segment

The paper proposes a method for determining the formants of the central frame of the vocal segment based on linear prediction coding, which includes the following steps:

1. Perform through the low-pass filter the balancing of the spectrum having a steep decline in the high frequency region

$$\breve{s}(m) = s(m+1) - \alpha s(m), \ m \in \overline{N^c - N/2, N^c + N/2} \ ,$$

where α is the filtration parameter, $0 < \alpha < 1$.

2. Calculate the autocorrelation function $R(k)$

$$\hat{s}(m) = \breve{s}(m) w(m), \ w(m) = 0.54 + 0.46 \cos \frac{2\pi m}{N} \ ,$$

$$R(k) = \sum_{m=N^c - N/2}^{N^c + N/2 - 1 - k} \hat{s}(m) \hat{s}(m+k), \ k \in \overline{0, p} \ ,$$

where $w(m)$ is the Hamming window, p is the linear prediction order, $ceil(f_d/1000) \le p \le 5 + ceil(f_d/1000)$, $ceil(f)$ is the function that rounds *f* to the nearest integer.

3. Calculate the LPC coefficients a_j in accordance with the Durbin procedure [21, 22]:
   3.1. $E^{(0)} = R(0)$;

3.2. $k_i = \left[R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j) \right] \Big/ E^{(i-1)}$;

3.3. $\alpha_i^{(i)} = k_i$;

3.4. $\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}, 1 \le j \le i-1$;

3.5. $E^{(i)} = (1-k_i^2) E^{(i-1)}$;

3.6. $i = i+1$;

3.7. if $i \le p$, then go to step 2;

3.8. $a_j = \alpha_j^{(p)}, 1 \le j \le p$.

4. Calculate the gain coefficient G

$$G = \sqrt{E} = \sqrt{R(0) - \sum_{k=1}^{p} a_k R(k)} \, .$$

5. Calculate the logarithmic energy spectrum using the gain and LPC coefficients

$$10\lg W(k) = 10\lg \frac{G^2}{\left(1 - \sum_{m=1}^{p} a_m \cos\left(\frac{2\pi}{\Delta N} km\right)\right)^2 + \left(\sum_{m=1}^{p} a_m \sin\left(\frac{2\pi}{\Delta N} km\right)\right)^2}, \ k \in \overline{0, N-1}$$

6. Calculate the frequency and amplitude of the formant in the logarithmic energy spectrum of the central frame:

6.1. Set frequency number $k = 0$. Set the number of formants $i = 0$;

6.2. If $10\lg W(k) > 10\lg W(k-1) \wedge 10\lg W(k) > 10\lg W(k+1) \wedge 10\lg W(k) > 0$, then fix the formant frequency, i.e. $F_{i+1} = k$, and the formant amplitude, i.e. $A_{i+1} = 10\lg W(k)$, increase the number of local extremums, i.e. $i = i+1$;

6.3. If $i < 3$, then go to the next frequency, i.e. $k = k+1$, go to step 6.2.

2.3 Choice of vocal speech sound features based on formants of the central frame of the vocal segment

The following vocal speech sound features have been chosen:

- - the frequency of the first formant $x_1 = F_1$;
- - the frequency of the second formant $x_2 = F_2$;
- - the frequency of the third formant $x_3 = F_3$;
- - the amplitude of the first anti-formant $x_4 = A_1$;
- - the amplitude of the second anti-formant $x_5 = A_2$;
- - the amplitude of the third anti-formant $x_6 = A_3$.

The total number of features is denoted as $Q = 6$.

3 Creation of a model of digital content processing system

The proposed digital content processing system that performs biometric identification of a person by voice is the artificial neuro-fuzzy network, a graph model of which is shown in Fig. 1.

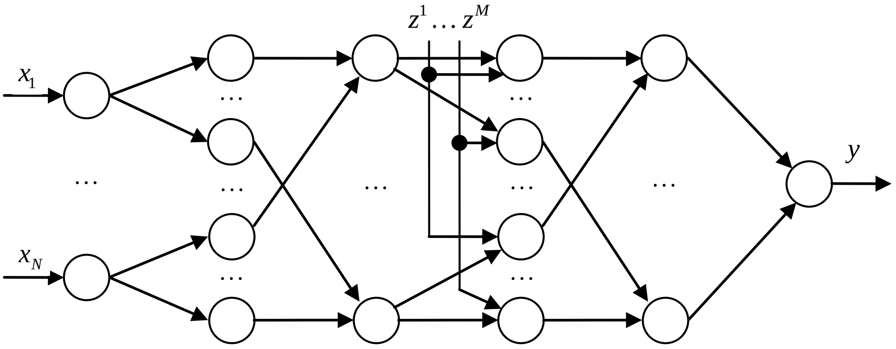

Fig. 1. A graph model of digital content processing system.

The input (zero) layer contains $N^{(0)} = Q$ neurons (corresponds to the number of features). The first hidden layer implements the fuzzification and contains $N^{(1)} = MQ$ neurons (corresponds to the number of values of linguistic variables). The second hidden layer implements the aggregation of subconditions and contains $N^{(2)} = M$ neurons (corresponds to the number of rules M). The third hidden layer implements the activation of conclusions and contains $N^{(3)} = M^2$ neurons. The fourth hidden layer implements the aggregation of conclusions and contains $N^{(4)} = M$ neurons. The output layer implements the defuzzification and contains $N^{(5)} = 1$ neuron.

All weighting coefficients are equal to 1.

The creation of the mathematical model of digital content processing system involves the following steps:

- formation of a fuzzy rule base;
- fuzzification;
- aggregation of subconditions;
- activation of conclusions;
- aggregation of conclusions;
- defuzzification.

3.1 Formation of a fuzzy rule base

Imagine the j -th fuzzy rule in the form

$$R^j : \text{IF } \tilde{x}_1 \text{ is } \tilde{\alpha}_1^j \text{ AND ... AND } \tilde{x}_Q \text{ is } \tilde{\alpha}_N^j \text{ THEN } \tilde{y} \text{ is } \tilde{\beta}^j ,$$

where $\tilde{x}_i$ is the name of the input linguistic variable, $i \in \overline{1,N}$; $\tilde{y}$ is the name of the output linguistic variable; $\tilde{\alpha}_i^j$ is the fuzzy variable (the value of the linguistic variable $\tilde{x}_i$), $j \in \overline{1,M}$, $i \in \overline{1,Q}$; $\tilde{\beta}^j$ is the fuzzy variable (the value of the linguistic variable $\tilde{y}$), $j \in \overline{1,M}$.

The fuzzy set $\tilde{A}_i^j$ is the range of values of the fuzzy variable $\tilde{\alpha}_i^j$, the fuzzy set $\tilde{B}^j$ is the range of values of the fuzzy variable $\tilde{\beta}^j$.

3.2 Fuzzification

Let's determine the degree of truth of the i -th subcondition, i.e. let's establish the correspondence between the input variables x_i of the j -th rule and the values of the membership function $\mu_{\tilde{A}_i^j}(x_i)$.

Since a number of methods related to person identification by voice use the Gauss function, we choose this function as $\mu_{\tilde{A}_i^j}(x_i)$, i.e.

$$\mu_{\tilde{A}_i^j}(x_i) = \exp\left[-\frac{1}{2}\left(\frac{x_i - m_i^j}{\sigma_i^j}\right)^2\right],$$

where m_i^j is the mathematical expectation, σ_i^j is the standard deviation.

3.3 Aggregation of subconditions

The membership function of the condition for the j -th rule is defined as

$$\mu_{\tilde{A}^j}(\overline{x}) = \mu_{\tilde{A}_1^j}(x_1)...\mu_{\tilde{A}_n^j}(x_n)\,,\ j \in \overline{1,M}\ .$$

3.4 Activation of conclusions

The membership function of the conclusion for the j -th rule is defined as

$$\mu_{\tilde{C}^j}(y) = \mu_{\tilde{A}^j}(\overline{x})\mu_{\tilde{B}^j}(y)\,,\ j \in \overline{1,M}\ ,$$

$$\mu_{\tilde{B}^j}(y) = \begin{cases} 0, & x \le j - 0.5 \\ \left(x - (j - 0.5)\right)/0.5, & j - 0.5 \le x \le j \\ \left((j + 0.5) - x\right)/0.5, & j \le x \le j + 0.5 \\ 0, & x \ge j + 0.5 \end{cases}$$ is a triangular function.

3.5 Aggregation of subconditions

The membership function of the final conclusion is defined as

$$\mu_{\tilde{C}}(y) = \max(\mu_{\tilde{C}^1}(y), ..., \mu_{\tilde{C}^M}(y)) .$$

3.6 Defuzzification

To obtain the class number, the membership function maximum method is used.

$$y = \arg\max_{z^j} \mu_{\tilde{C}}(z^j) \text{ ; } z^j \text{ is the center of the fuzzy set } \tilde{C}^j .$$

Thus, the mathematical model of digital content processing system (Fig. 1) can be represented as

$$y = \arg\max_{z^k} \max_{j \in \overline{1,M}} \mu_{\tilde{B}^j}(z^k) \prod_{i=1}^{Q} \mu_{\tilde{A}_i^j}(x_i) , \ k \in \overline{1,M} .$$

The determination of the parameters of this system is carried out on the basis of the genetic algorithm.

4 Choice of the structure of the method for determining parameter values of the mathematical model of digital content processing system

The choice of the structure of the genetic algorithm, which allows to determine parameter values of the mathematical model of digital content processing system, involves the following steps:

- identification of individuals of the initial population;
- definition of fitness function;
- choice of reproduction (selection) operator;
- choice of crossing-over operator;
- choice of mutation operator;
- choice of reduction operator;
- definition of a stop condition.

4.1 Identification of individuals of the initial population

Material genes have been selected for the following reasons:

- - the ability to search in large spaces, which is difficult to do in the case of binary genes, when an increase in the search space reduces the accuracy of the solution with a constant chromosome length;
- - the ability to configure solutions locally;
- - the lack of encoding / decoding operations that are necessary for binary genes increases the speed of the algorithm;
- - proximity to the formulation of the most applied problems (each material gene is responsible for one variable or parameter, which is impossible in the case of binary genes).

An ordered vector of parameters (mathematical expectations and standard deviations) acts as the chromosome, which represents the i -th individual of the population $H = \{h_i\}$

$$h_i = (lx_1^1 + i * \Delta m_1^1, lx_1^2 + i * \Delta m_1^2, ..., lx_n^1 + i * \Delta m_n^1, lx_n^2 + i * \Delta m_n^2,$$

$$lx_1^1 + i * \Delta \sigma_1^1, lx_1^2 + i * \Delta \sigma_1^2, ..., lx_n^1 + i * \Delta \sigma_n^1, lx_n^2 + i * \Delta \sigma_n^2),\ i \in \overline{1, |H|},$$

$$\Delta m_k^j = \frac{rx_k^j - lx_k^j}{|H|},\ \Delta \sigma_k^j = \frac{rx_k^j - lx_k^j}{|H|},\ j \in \overline{1, M},$$

where $|H|$ is the population power, lx_k^j , rx_k^j are the left and right boundaries of the values of the k -th feature, calculated experimentally.

4.2 Definition of fitness function

In the paper the following fitness function, which corresponds to the probability of correct identification of a person by voice, is proposed

$$F = \frac{1}{P} \sum_{p=1}^{P} I(y_p - d_p) \to \max_{m_i^j, \sigma_i^j},\ I(a) = \begin{cases} 1, & a = 0 \\ 0, & a \neq 0 \end{cases},$$

where d_p is the response received from the object (person), y_p is the response obtained by the model, P is the number of test implementations.

4.3 Choice of reproduction (selection) operator

The following effective combination is used to select parameter vectors for crossing and mutation as a reproduction operator

$$P(h_i) = \frac{1}{|H|} \exp(-1/g(t)) + \frac{1}{|H|} \left(a - (2a - 2) \frac{i-1}{|H| - 1} \right) (1 - \exp(-1/g(t))).$$

Thus, in the early stages of the genetic algorithm, an uniform selection is used to ensure that the entire search space is studied (random selection of chromosomes), and in the final stages, linearly ordered selection is used to make the search directed (the current best chromosomes are preserved). This combination does not require scaling and can be used to minimize fitness function.

4.4 Choice of crossing-over (crossover, recombination) operator

To combine the two options of the vector of parameters selected by the reproduction operator, an uniform crossing-over is used as the crossing-over operator.

Parents are selected through the following effective combination – in the early stages of the genetic algorithm, outbreeding is used to provide an investigation of the entire search space, and in the final stages, inbreeding is used to make the search directed. This combination does not require scaling and can be used to minimize fitness function.

After the selection of parents, a cross is carried out, and two descendants are produced.

For a global search for the optimal vector of parameters, it is necessary to increase the variety of options.

4.5 Choice of mutation operator

To ensure the variety of options for the vector of parameters after crossing-over, an non-uniform mutation is used.

The mutation step is defined as

$$\Delta = \begin{cases} (Max_j - h_{ij})r\left(1-\dfrac{t}{T}\right)^b, & r < 0.5 \\ (h_{ij} - Min_j)r\left(1-\dfrac{t}{T}\right)^b, & r \geq 0.5 \end{cases},$$

where Max_j, Min_j are the maximum and minimum values of the j-th gene; t is the iteration number; T is the maximum number of iterations; r is the random number, $r \in [0,1]$; b is the parameter controlling the speed of step decrease, $b > 0$.

To simulate annealing, the probability of mutation is defined as

$$P_m = P_0 \exp(-1/g(t)),\ g(t) = \beta g(t-1),\ 0 < \beta < 1,\ g(0) = T_0,\ T_0 > 0,$$

where P_0 is the initial probability of mutation.

Thus, in the early stages of the genetic algorithm, a large step mutation occurs with high probability, which provides an investigation of the entire search space, and in the final stages, the probability of mutation and its step tend to zero, which makes the search directed.

4.6 Choice of reduction operator

The reduction operator allows to create a new population based on the previous population and parameter vectors obtained by crossing-over and mutation. As a reduction operator, a scheme $(\mu + \lambda)$ is applied that does not require scaling and can be used to minimize fitness function.

4.7 Definition of a stop condition

The following condition is proposed in the work

$$1 - \max_i F(h_i) < \varepsilon \vee t \geq T.$$

The values of ε and T are calculated experimentally.

5 Numerical research

Table 1 presents the probabilities of a person identification by voice obtained on the basis of TIMIT based on the artificial neural network of the multilayer perceptron type and the proposed method. At the same time, the artificial neural network has had two hidden layers (each has consisted of six neurons, like the input layer).

According to Table 1, the proposed method gives the best results.

Table 1. The probability of biometric identification of a person by voice.

Method	Identification probability
Artificial neural network	0.8
Proposed method	0.98

6 Conclusions

1. To solve the problem of increasing the efficiency of digital content processing system for biometric identification of a person by voice, the corresponding speaker recognition methods have been investigated. These studies have shown that today the use of artificial neural networks in combination with the fuzzy inference system and the genetic algorithm is the most effective method.
2. The proposed method of digital content processing for biometric identification of a person by voice automates the process of generation of digital content features, provides a representation of knowledge in the form of rules that are easily accessible for human understanding, and simplifies the determination of the structure of the model due to the fuzzy inference system; reduces the probability of falling into a local extremum and provides an acceptable speed for determining the parameter values of the model by choosing the effective structure of the genetic algorithm; allows parallel processing of information due to the artificial neural network.
3. As a result of a numerical study, it has been found that the proposed method of digital content processing provides 0,98 probability of biometric identification of a person by voice, which exceeds the probability obtained by the artificial neural network such as a multilayer perceptron.
4. The proposed method of digital content processing for biometric identification of a person by voice can be used in various intelligent systems for digital content processing.

References

1. Bolle, R.M., Connell, J., Pankanti, S., Ratha, N.K., Senior, A.W.: Guide to biometrics. Springer, New York (2004).

2. Jain, A.K., Flynn, P., Ross, A. (Eds.): Handbook of biometrics. Springer, New York, NY (2008).
3. Dunstone, T., Yager, N.: Biometric system and data analysis: design, evaluation, and data mining. Springer, New York (2009).
4. Singh, N., Khan, R., Shree, R.: Applications of Speaker Recognition. Procedia Engineering. 38, 3122–3126 (2012). doi: 10.1016/j.proeng.2012.06.363
5. Li, Q.: Speaker authentication. Springer-Verlag Berlin Heidelberg, Heidelberg (2012).
6. Keshet, J., Bengio, S.: Automatic speech and speaker recognition: large margin and kernel methods. John Wiley & Sons, Chichester (2009).
7. Herbig, T., Gerl, F., Minker, W.: Self-learning speaker identification: a system for enhanced speech recognition. Springer, Berlin (2013).
8. Campbell, J.: Speaker recognition: a tutorial. Proceedings of the IEEE. 85, 1437–1462 (1997). doi: 10.1109/5.628714
9. Togneri, R., Pullella, D.: An Overview of Speaker Identification: Accuracy and Robustness Issues. IEEE Circuits and Systems Magazine. 11, 23–61 (2011). doi: 10.1109/MCAS.2011.941079
10. Beigi, H.: Fundamentals of speaker recognition. Springer, New York (2011).
11. Reynolds, D.A.: An overview of automatic speaker recognition technology. IEEE International Conference on Acoustics Speech and Signal Processing. 4, 4072–4075 (2002). doi: 10.1109/ICASSP.2002.5745552
12. Kinnunen, T., Li, H.: An overview of text-independent speaker recognition: From features to supervectors. Speech Communication. 52, 12–40 (2010). doi: 10.1016/j.specom.2009.08.009
13. Reynolds, D., Rose, R.: Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Transactions on Speech and Audio Processing. 3, 72–83 (1995). doi: 10.1109/89.365379
14. Zeng, F.-Z., Zhou, H.: Speaker Recognition based on a Novel Hybrid Algorithm. Procedia Engineering. 61, 220–226 (2013). doi: 10.1016/j.proeng.2013.08.007
15. Jeyalakshmi, C., Krishnamurthi., V., Revathi, A.: Speech recognition of deaf and hard of hearing people using hybrid neural network. 2010 2nd International Conference on Mechanical and Electronics Engineering. (2010). doi: 10.1109/ICMEE.2010.5558589
16. Nayana, P., Mathew, D., Thomas, A.: Comparison of Text Independent Speaker Identification Systems using GMM and i-Vector Methods. Procedia Computer Science. 115, 47–54 (2017). doi: 10.1016/j.procs.2017.09.075
17. Chauhan, V., Dwivedi, Sh., Karale, P., Potdar, S.M.: Speech to text converter using Gaussian mixture model (GMM). International Research Journal of Engineering and Technology (IRJET). 3, 160–164 (2016).
18. Reynolds, D.A.: Automatic speaker recognition using Gaussian mixture speaker models. IEEE Transactions on Speech and Audio Processing. 3, 1738–1752 (1995).
19. Fedorov, E., Lukashenko, V., Utkina, T., Rudakov, K., Lukashenko, A.: Method for parametric identification of Gaussian mixture model based on clonal selection algorithm. CEUR Workshop Proceedings. 2353, 41–55 (2019).
20. Larin, V.J., Fedorov, E.E.: Combination of PNN network and DTW method for identification of reserved words, used in aviation during radio negotiation. Radioelectronics and Communications Systems. 57, 362–368 (2014). doi: 10.3103/S0735272714080044
21. Rabiner, L.R., Juang, B.-H.: Fundamentals of speech recognition. Pearson Education, Delhi (2005).
22. Markel, J.D., Gray, A.H.: Linear prediction of speech. Springer-Verlag, Berlin (1976).