Vjatcheslav V. KOVTUN[1], Maria S. YUKHIMCHUK[1], Piotr KISAŁA[2], Akmaral ABISHEVA[3], Saule RAKHMETULLINA[4]

Vinnytsia National Technical University (1), Lublin University of Technology (2), Al-Farabi Kazakh National University(3), East Kazakhstan State Technical University named after D.Serikbayev (4)

doi:10.15199/48.2019.04.32

Integration of hidden markov models in the automated speaker recognition system for critical use

***Abstract**. In this article, the author theoretically substantiated the possibility of integration of hidden Markov models (IHMM) in the structure of the automated speaker recognition system for critical use (ASRSCU) for analysis of speech information from a plurality of independent input channels, which allowed within the statistical conception of pattern recognition to combine the accuracy of the approximation of input signals inherent the apparatus of GMM models. The authors proposed a mathematical apparatus for the integration of hidden Markov models, which allows us to adequately describe the set of interacting processes in the Markov paradigm with the preservation of temporal, asymmetric conditional probabilities between the chains*

Streszczenie.** W tym artykule autorzy teoretycznie uzasadnili możliwość integracji ukrytych modeli Markowa (IHMM) w strukturze zautomatyzowanego systemu rozpoznawania głosu osoby mówiącej do zastosowań krytycznych (ASRSCU) do analizy informacji o mowie z wielu niezależnych kanałów wejściowych, które dopuszczają wewnątrz statystyczna koncepcję rozpoznawania wzorców w celu połączenia dokładności aproksymacji sygnałów wejściowych z aparatem modeli GMM. Autorzy zaproponowali aparat matematyczny do integracji ukrytych modeli Markowa, który pozwala odpowiednio opisać zestaw oddziałujących procesów w paradygmacie Markowa z zachowaniem czasowych, asymetrycznych warunkowych prawdopodobieństw między łańcuchami. **(Integracja ukrytych modeli Markowa w zautomatyzowanym systemie rozpoznawania głosu do zastosowań krytycznych).

Keywords: microphone set, Gaussian mixures models, hidden Markov models.
Słowa kluczowe: zestaw mikrofon, modele miksów Gaussa, ukryte modele Markowa.

Introduction

In speaker recognition systems (SRS) as a whole and in the automated speaker recognition system for critical use (ASRSCU) in particular, use the classical methods of pattern recognition theory, namely statistical simulation methods for describing the vectors of individual features of speech signals. Most often, in models of Gaussian mixtures [1, 2], artificial neural networks [3, 4] or support vector machines [5, 6, 7]. Less used are hidden Markov models [5, 6, 8, 9], which, however, together with Gaussian mixtures models, are very often used as part of speech recognition systems.

Gaussian mixture models (GMM) are used in SRS to estimate the density of the probabilities of the variability of speech data due to moderately low computational cost of analysis and convergent adaptation algorithms [9, 10], in particular, the Expectancy-Maximization (EM) algorithm, the Maximum a Posteriori Probability (MAP) algorithm or Maximum Likelihood Linear Regression Maximization (MLLR) algorithm. However, the GMM has a low sensitivity to the variability of speech signal over time, which is usually compensated by detail for an adequate description of the individual features of speech, which leads to an increase in the sensitivity of the received features space to the presence in a phonogram of a speech signal a noises of the surrounding space.

In their turn, the hidden Markov models (HMM) are statistical models that describe the analyzed system as a Markov process with unknown parameters [6] in order to determine the most probable state of the sequence of units of acoustic elements of speech signals based on pre-trained models. For SRS, each state of the HMM is represented by different stable elements of speech (for example, the phonemes), and the time information is encoded by the permitted transitions between states. Thus, the speaker recognition using HMM is to determine for each speaker the optimal position between the sequence of the test speech vector represented by the phonogram and the GMM associated with a certain word or a passphrase. In current speaker recognition systems, HMM-GMM methods are used jointly, where the description of the structure of the speech signal in time is carried out by means of the HMM, and registration of individual features of speech is carried out by means of GMM in the form of indicators of abnormalities of patterns of linguistic units from their reference representations received at the educational stage. To simulate complex systems, HMM of higher order are used, in particular, the factorial HMM for the modeling of independent processes within the system, connected HMM [11, 16] for simulating the symmetric effects without taking into account cause-effect relationships, hidden Markov decision trees [12], which describe the cascade of impacts of basic and subordinate HMM without taking into account causal relationships. However, the mathematical apparatus of HMM, in its classical form, is not suitable for the simulation of systems consisting of completed component elements, which, however, interact in space and time within a system of higher order. So, in order to increase the robustness of the ASRSCU, the recording of a speech signal can occur by a set of microphones in different places in space relative to the speaker to obtain a refined noise model of the environment. The simulation of such a system within the framework of the HMM should describe the relationship between processes that have different structure and degrees of influence on each other within a single recognition system, that is, to account for causative relationships. We illustrate the expediency of such an approach to the description of the ASRSCU with a set of microphones for recording password signals in an example of describing the behavior of opponents in a sporting event: the integration of HMM without regard to cause-effect relationships will describe the game of each opponent separately, while the integration of IHMM (IHMM) with taking into account causal relationships will take into account the reaction of each player to the opponent's action with a limited set of strategies.

Problem statement

The author suggests the method of integration of HMM, which allows to account for cause-effect (temporal) and symmetric and asymmetric factors between the basic elements of the described system, which, in particular, is

ASRSCU with a set of microphones for recording passwords. The phonograms with the recordings of identical passwords speech signals received in different acoustic conditions of the speaker surrounding environment, described in the framework of the HMM-GMM models, have to be effectively integrated into the complex speaker recognition system with the aim of increasing the reliability of the speaker recognition. The proposed IHMM method is based on integrated projections of integrated HMM and allows to synthesize an approximating learning algorithm for HMM within the boundary field based on the junction tree (JT) [13] algorithm.

The method of integration of hmm and description of their learning process

The hidden Markov model quantizes the space of system configurations to a finite set of discrete states. Let the discrete variable *s* show the current state of the system, then changes in the states describing the dynamics of the system will be described by the matrix of transition probabilities $P_{s(t)=i|s(t-1)=j}$. Full description of the system within the Markov model includes a set of parameters {S, $P_{i|j}$, P_i, $P_i(o)$}, where S={s_1, s_2,..., s_N} - finite set of discrete states of the system, $P_{i|j}=P_{s(t)=i|s(t-1)=j}$ - transition probability between states 1<i and $j<N$, $P_i=P_s(0)=i$ - a priori probability of the initial state of the system, $P_i(0)=P_{s(t)=i}(0(t))$ - output probability for each state of the system. Graphically Markov models are time ordered probabilistic-independent network-graph (see Fig.1), using structural elements are square units representing observation $o(t)$, round nodes that represent hidden state variables $s(t)\in S$, horizontal links representing transition matrix $P_{s(t)|s(t-1)}$, and vertical nodes that represent the probabilistic features of observation of the current state in the form of $P_{s(t)}(o(t))$,for example, averages or covariates of poly-factor Gaussian. The values of the state and output variables change in time, and at any time t, the memory is limited to the value of the state variable $s(t-1)$.

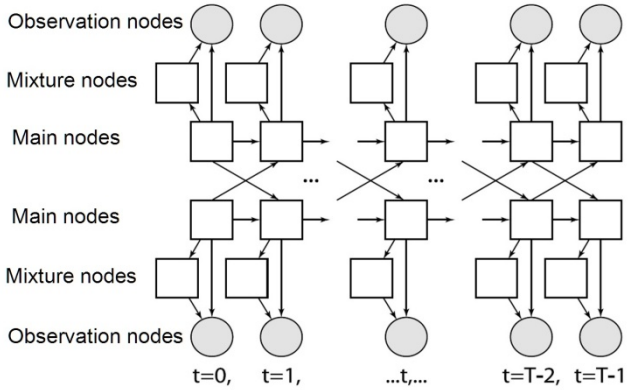

Fig.1. Generalized image of the IHMM in the time space

We obtain IHMM C as a result of the integration of separate HMM A i B and by performing the Cartesian product of their states and transition parameters:

(1) $\{C\}=\{A\}\times\{B\}$, $c_{ij}=a_i\wedge b_j$, $P_{c_{ik}|c_{jl}}=P_{a_i|a_j}P_{b_k|b_l}$

Taking into account that the sum of probabilities is equal to one, we will consider the feedback of the parameters of the HMM components:

(2) $P_{a_i|a_j}=\sum_l P_{b_l}\sum_k P_{c_{ik}|c_{jl}}$, $P_{b_k|b_l}=\sum_j P_{a_i}\sum_i P_{c_{ik}|c_{jl}}$

where P_{bl}=1/|{B}, P_{aj}=1/|{A} provided there is no a priori probability. These projections form (|{A}|·|{B}|)^2- dimensional IHMM transformation matrix with |{A}|^2- and |{B}|^2 - dimensional transformation matrices that characterize the input HMM.

Using the same reasoning, we will express the relationship of output HMM in the form of such expressions:

(3) $P_{a_i|b_l}=\sum_j P_{a_i}\sum_k P_{c_{ik}|c_{jl}}$, $P_{b_k|a_j}=\sum_l P_{b_l}\sum_i P_{c_{ik}|c_{jl}}$

and, as a result:

(4) $$P_{c_{ik}|c_{jl}}=P_{a_i|b_l}P_{b_k|a_j}$$

Expressions (1)-(4) form the basis of a method that allows us to teach the IHMM with standard learning methods, taking into account only the features that affect both the input HMMs, that is, factoring will be carried out after the re-evaluation of the IHMM. However, factoring can also be carried out after a direct-inverse analysis of initial HMM:

(5) $$P_{a(t)=i,a(t-1)=j|O}=\frac{\sum_k\sum_l C_{jl,t-1}\cdot P_{ik|jl}\cdot P_{c(t)=ik}(o(t))\cdot C'_{ik,t}}{P(O)}\sim$$
$$\sim\frac{P_{a(t)=i}(o(t))}{P(O)}\sum_k C'_{ik,t}\sum_l\left(C_{jl,t-1}\cdot P_{ik|jl}\right)\sim\frac{P_{a(t)=i}(o(t))\cdot P_{i|j}}{P(O)}\sum_k C'_{ik,t}\sum_l C_{jl,t-1}$$

where the variables C and C' characterize the direct and inverse stages of the learning process of the IHMM respectively. Note that the equivalent components (5) are formulated in order to accelerate the learning process with, respectively, the loss of ever greater part of information. It is probable that the factoring and recovery procedures may prevent the convergence of the learning process, but empirical studies, the formulation and results of which are described below, proved the capacity of the proposed IHMM training in the form of (5) under the conditions (1).

Note that the results of the Cartesian products of the initial parameters of the IHMM are not used in the training - they are evaluated directly in the initial HMMs and when integrated, they are taken into account as a posteriori probabilities of the states of the initial HMMs, which provides the following advantages: at each iteration of the learning process $O(2N)$ of the initial parameters are overestimated instead of $O(N^2)$; it is possible to vary the reliability of statistical parameters used by choosing the desired version of the learning process (5); the procedures of direct-inverse analysis and epy Viterbi analysis are done in $O(N)$ times faster, because the bulk of the computation falls on the calculation of multidimensional Gaussian. Consequently, the author's proposed (1) - (5) procedure for the formation and training of the IHMM is not only adapted to describe the pattern recognition in general and for the ASRSCU in particular, but also is computationally effective.

Applied application of the ihmm method to description of processes in the asrscu

Consequently, the proposed mathematical model of the integration of the HMM proposed by the author makes it possible to use in the recognition system an efficient set of HMMs, each of which describes an independent flow of data. For example, when the system processes speech signals from a set of microphones, each of which is separate, not linked to another source of information. For such a system, the hidden structure of the basic nodes of each HMM at the time t is due to the information of the basic nodes of other HMM at the time t-1, provided that there is a connection between the structural elements of the networks concerned. Generalized parameters (1) for the case of combining two HMMs in the ASRSCU as follows:

(6) $$\pi_o^c(i)=P\left(q_1^c=i\right)\, b_t^c(i)=P\left(O_t^c\middle|q_t^c=i\right)$$

$$a_{i|j,k}^c=P\left(q_t^c=i\middle|q_{t-1}^w=j,q_{t-1}^v=k\right)$$

where $\pi_o^c(i)$- the probability of the initial state i, $a_{i|j,k}^c$ - the probability of transition $i\to j|k$, $b_c^t(i)$- the probability of O_t^c belonging to i, the variable $c\in\{w,v\}$ describes the set of independent channels of input information, and q_c^t- describes the condition of the combined nodes within the system at the time t. In the combined Gaussian mixtures the probability of the observation nodes will describe the expression

(7) $$b_t^c(i)=\sum_{m=1}^{M_i^c}\omega_{i,m}^c N\left(O_t^c,\mu_{i,m}^c,U_{i,m}^c\right)$$

where O_t^c - the vector of observations at the time t within the system c, $\mu_{i,m}^c$, $U_{i,m}^c$and $\omega_{i,m}^c$ - the mathematical expectation, the covariance matrix and the weights of the GMM of the state i, with associated with the mixture within the system c, M_i^c - the number of mixtures describing the state within the system c. For ASRSCU, each HMM describes one of the possible phoneme pairs for each speaker, the information about which is generalized in the base of the system references.

We formalize the process of learning of the IHMM in the ASRSCU (see Fig.2), taking into account (5) in two stages. At the first stage, we obtain a speaker-independent background model for each IHMM corresponding to the phoneme pair of the passphrase. At the second stage, we will adapt the IHMM parameters to the speaker individual features described by the speech model, using the MAP algorithm. Also, in accordance with the ideology of SRS architecture, we teach two additional IHMMs to identify the pause between words and sentences in the password phonogram respectively.

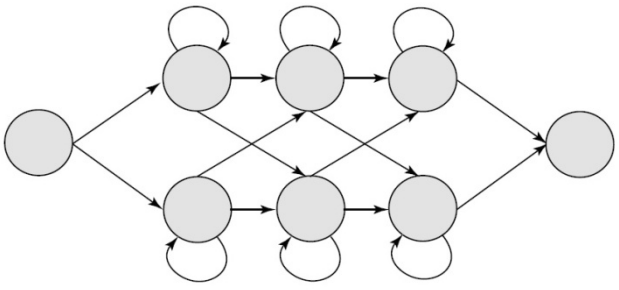

Fig.2. Diagram of states of the IHMM where $c\in\{w,v\}$

To perform the first stage, we initialize the IHMM for isolated pairs of phonemes using the Viterbi algorithm [14], after which we use the EM algorithm [5], [6]. Each created for the description of the isolated pair of phonemes HMM has one input, a set of non- emitting states and one output. The use of non-emitting states, in particular, provides synchronization of phonemes within the model (Fig. 2). The IHMM parameters are refined by the built-in learning method based on continuous linguistic information, which is provided to the inputs of all created IHMM. At this stage, the tags of the initial sequences contain only phonemic sequences, and all individual information is ignored. To describe the second stage, which regulates the adaptation of the basic parameters of the background model to the speaker's individual features of the speech signals that will be analyzed by the ASRSCU, denote $(\mu_{i,m}^c)$ - epy mathematical expectations, $(U_{i,m}^c)_{IHMM}$ - the covariance matrices and $(\omega_{i,m}^c)_{\text{IHMM}}$ - weights of the mixture m of the state i of the channel c of the trained IHMM. Apply Bayesian Adaptation algorithm [15] to update the state parameters $\hat{\mu}_{i,m}^c$, $\hat{U}_{i,m}^c$, $\hat{\omega}_{i,m}^c$ for all of the IHMM:

$$\hat{\mu}_{i,m}^c=\mu_{i,m}^c\theta_{i,m}^c+\left(\mu_{i,m}^c\right)_{IHMM}\left(1-\theta_{i,m}^c\right)$$

(8) $$\hat{U}_{i,m}^c=U_{i,m}^c\theta_{i,m}^c-\left(\mu_{i,m}^c\right)^2+\left(\mu_{i,m}^c\right)_{IHMM}^2+\left(U_{i,m}^c\right)_{IHMM}\left(1-\theta_{i,m}^c\right)$$

$$\hat{\omega}_{i,m}^c=\omega_{i,m}^c\theta_{i,m}^c+\left(\omega_{i,m}^c\right)_{IHMM}\left(1-\theta_{i,m}^c\right)$$

where $\theta_{i,m}^c$ - is the MAP-adaptation degree parameter of the mixture m of the state i of the channel c.

Apply an EM algorithm for obtaining statistical parameters $\mu_{i,m}^c$, $U_{i,m}^c$ and $\omega_{i,m}^c$ for personify the states of the IHMM for the task of speaker recognition for ASRSCU application:

$$\mu_{i,m}^c=\frac{\sum_{r,t}\gamma_{r,t}^c(i,m)O_{r,t}}{\sum_{r,t}\gamma_{r,t}^c(i,m)}$$

(9) $$U_{i,m}^c=\frac{\sum_{r,t}\gamma_{r,t}^c(i,m)\left(O_{r,t}-\mu_{i,m}^c\right)\left(O_{r,t}-\mu_{i,m}^c\right)^T}{\sum_{r,t}\gamma_{r,t}^c(i,m)}$$

$$\omega_{i,m}^c=\frac{\sum_{r,t}\gamma_{r,t}^c(i,m)}{\sum_{r,t}\sum_{k}\gamma_{r,t}^c(i,k)}$$

where:

$$\gamma_{r,t}^c(i,m)=\frac{\sum_j P_r^{-1}\alpha_{r,t}(i,j)\beta_{r,t}(i,j)}{\sum_{i,j}P_r^{-1}\alpha_{r,t}(i,j)\beta_{r,t}(i,j)}\times\frac{\omega_{i,m}^c N\left(O_{r,t}^c\middle|\mu_{i,m}^c,U_{i,m}^c\right)}{\sum_k\omega_{i,m}^c N\left(O_{r,t}^c\middle|\mu_{i,k}^c,U_{i,k}^c\right)},$$

$\alpha_{r,t}(i,j)=P\left(O_{r,1},\ldots,O_{r,t}\middle|q_{r,t}^w=i,q_{r,t}^v=j\right)$ and

$\beta_{r,t}(i,j)=P\left(O_{r,t+1},\ldots,O_{r,T_r}\middle|q_{r,t}^w=i,q_{r,t}^v=j\right)$- anterior and posterior variables [9] respectively which are calculated for r of the observed sequences $O_{r,t}=[(O_{r,t}^w)^T,(O_{r,t}^v)^T]^T$, $\theta_{i,m}^c=\frac{\sum_{r,t}\gamma_{r,t}^c(i,m)}{\sum_{r,t}\gamma_{r,t}^c(i,m)+\delta}$ the degree of adaptation, and δ - the relevance factor (for further research used value δ=16), which establishes a direct relationship between the volume of speech-dependent data used for learning mixture m state i and channel c and reliability of statistics characterizing the personality of the speaker, in the set of basic parameters of the MAP-algorithm (8). If the data for mixtures learning is not enough, the basic parameters of the MAP algorithm will follow the parameters of the background model.

The decision procedure for the person speaking the ASRSCU with IHMM will be implemented in two stages. At the first stage, the decision on the speaker's person is taken separately for each incoming data channel by the corresponding HMM. Usually these results will be characterized by different levels of relative reliability, since the level and nature of noise in different input channels is different. To take into account this circumstance we will replace the probability of observation used in the recognition on $\tilde{b}_t^c(i)=|b_t^c|^{\lambda_c}$, $c\in\{w,v\}$, where $\lambda_c\in\{\lambda_w,\lambda_v\}$ - a complex parameter that satisfies the conditions $\lambda_w,\lambda_v\geq 0$ and $\lambda_w+\lambda_v=1$, and express the second stage of the recognition procedure in the form of this rule:

(10) $$L\left(O^w, O^v \middle| k\right) = \lambda_w L\left(O^w \middle| k\right) + \lambda_v L\left(O^v \middle| k\right)$$

where O^w, O^v - the sequences of the results of the recording sessions of the speech information in the corresponding channels, operation $L(*|k)$ - describes the decision procedure for the k person speaking based on the corresponding input data, parameters λ_w, λ_v are the quality of weighting factors that characterize the reliability of the corresponding channel of input information.

Experiments and result analysis

In order to assess the adequacy of the above described theoretical concepts, the author synthesized the ASRSCU with the IHMM, the recording of the speech information for which was carried out by two microphones, located linearly at a distance of 2 cm and 50 cm from the source of the speech signal. In the formation of vectors of features that compactly describe the speaker's individual features of the speech signal to be analized by the ASRSCU, phonograms with speech signals were clasped on 25 ms frames with a 15 ms overlap, each of which was described by a vector of 13 MFC coefficients and their first and second derivatives. To summarize information from different input channels, a three-state IHMM for each channel was used without feedback (see Fig.2). Each IMSM state was described by 32 components of the mixture in the form of a diagonal covariance matrix. In the training and testing of the created ASRSCU with IHMM data from the database of records of speech signals NOIZEUS were used. To study the system, 1450 phonograms with speech signals from 200 speakers with known signal to noise ratios (SNR) were used. To test the ASRSCU with the IHMM on the basis of the materials of the training base NOIZEUS created two sets of test data, based on 700 records of 95 speakers. The first set included phonograms with the levels of SNR = 5, 10, 15, ..., 30 dB. When receiving phonograms for the second test set of phonograms with noise, they merged into a common file, which was subsequently broken into 1800 phonograms with a randomly selected level of SNR. Note that the data from the second set are closer to the real ones, since they do not contain meaningful.

The results of speaker recognition by the created system are shown on Fig.3 and Fig.4. To evaluate the results of the experiments, the Equal Error Rate (EER) criterion, which is characterized by the equality point of errors of the first and second kind, is determined by the point of intersection of the distribution curves of the probabilities of these errors. The EER value allows you to evaluate the informality of the features used by the system to identify the speaker. The smaller the EER value, the less the overlap between the error curves of the first and the second kind and the more compact the features space obtained using the author's method. Fig.3 shows the recognition results based on the first test set using information from only one channel (λ_w, λ_v)=(1,0) and from only another channel (λ_w, λ_v)=(0,1), with the recognition of both channels equally reliable (λ_w, λ_v)=(0.5,0.5) and with the recognition of a more reliable first channel (λ_w, λ_v)=(0.7,0.3) or vice versa - (λ_w, λ_v)=(0.3,0.7). Fig. 4 shows the results of the speaker recognition according to the data of the second test set when using information from only one channel (λ_w, λ_v)=(1,0) and from only another channel (λ_w, λ_v)=(0,1) and with the recognition of both channels equally reliable (λ_w, λ_v)=(0.5,0.5).

The results presented in Fig.3 show that the integration of HMMs in the ASRSCU enabled to achieve a stable qualitative speaker recognition, even with the growth of SNR. It is interesting to find the result of recognition when using information exclusively from the first channel, where the dependence of the recognition quality on the SNR is not followed. This can be explained by the fact that the first narrow-gauge microphone is located at a distance of 2 cm from the source of the speech signal, which limits the presence of a noise in the phonogram, but this location causes reverb for the explosive segments of the speech signal, which leads to a decrease in the quality of speaker recognition for the investigated configuration of the features space. Finally, comparing results with established channel priorities in the form (λ_w, λ_v)=(0.5,0.5) and (λ_w, λ_v)=(0.7,0.3) it can be stated that the second option provides more stable results, which correlates with the previous explanation regarding the quality of speaker recognition quality according to the data of the first channel.

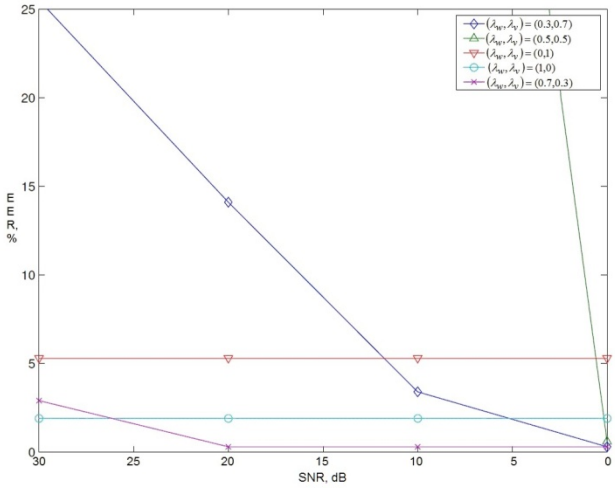

Fig.3. Results of testing of ASRSCU with IHMM based on the data of the first test set

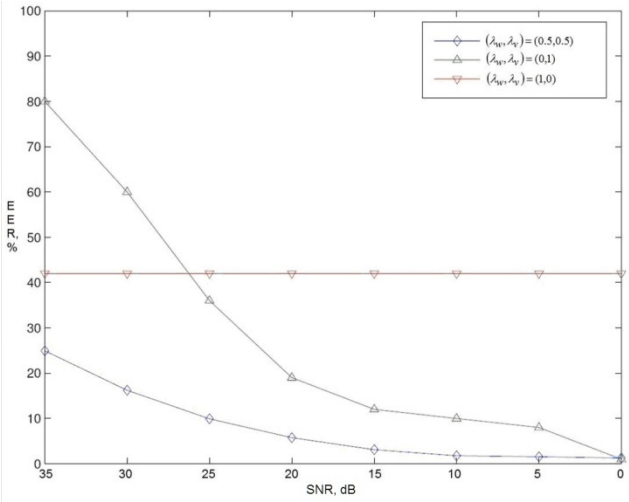

Fig.4. Results of testing of ASRSCU with IHMM based on data from the second test set

The results presented on Fig.4 supplement the above, the established system showed a stable result when operating under the conditions of the SNR>15 dB in the input phonograms. Note that the second test sample consisted of fragments of speech signals with different levels of SNR, which is typical for the operation of the system in a complex technological environment.

Conclusions

In this article the author substantiates the possibility of integration of HMMs within the unified automated speaker recognition system of critical use for the analysis of speech

information from a set of independent input channels, which allowed within the statistical conception of pattern recognition to combine the accuracy of the approximation of input signals inherent in the apparatus of Gaussian mixtures models, the individual features of speech representation due to the possibilities of hidden Markov models and robustness which is a key characteristic of critical recognition systems, through the author's method of integrating receiving information from independent input channels.

The productivity of classical HMM when applied in the perceptual decision-making systems for the classification of non-stationary input signals, which can be attributed to a speaker's person, is reduced, in particular, by the limited Markov states. The author proposed a mathematical apparatus for the integration of hidden Markov models, which allows us to adequately describe the set of interacting processes in the Markov paradigm with the preservation of the temporal, asymmetric conditional probabilities between the chains. The method of adaptation of the Viterbi algorithm for the training of the HMM is proposed, which allows for each iteration of the learning process to overestimate $O(2N)$ original parameters instead of $O(2N^2)$ providing the variability of the reliability of the statistical parameters used by choosing the desired learning process option (5) and performing the procedures of direct-inverse analysis and Viterbi analysis in $O(2N)$ times more quickly, since the bulk of the calculus falls on the calculation of multidimensional Gaussian.

The proposed scientific results are empirically tested, in the course of experiments on the speakers recognition by the ASRSCU with IHMM in conditions of the variable level and type of noise in the input speech signals. The experiments results are shown on Fig.3 and Fig.4. Note that the IHMM revealed less sensitivity to the operating conditions of the recognition system that allows you to recommend their use as part of the automated speaker recognition system of critical use. In further research, the author plans to evaluate the stability of the designed system to different types and levels of noise in the incoming speech signal by changing the number and parameters of the channels of speech information receipt and consider the possibility of ensuring the ASRSCU with IHMM becomes text-independent, since the presented HMM system is text-dependent.

***Authors**: Ph.D., Associate professor of the Computer Control Systems Department Vjatcheslav V. Kovtun, Vinnytsia National Technical University, Khmelnytsky Hwy, 95, 21021 Vinnytsia, Ukraine, e-mail: kovtun_v_v@vntu.edu.ua; Ph.D., Associate professor of the Computer Control Systems Department Maria S. Yukhimchuk, Vinnytsia National Technical University, Khmelnytsky Hwy, 95, 21021 Vinnytsia, Ukraine, e-mail: umcmasha@gmail.com; PhD., Prof. Piotr Kisała, Lublin University of Technology, Institute of Electronics and Information Technology, Nadbystrzycka 38A, 20-618 Lublin, Poland, e-mail: p.kisala@pollub.pl; M.Sc. Akmaral Abisheva, Al-Farabi Kazakh National University, Almaty, Kazakhstan, email: ak_maral@mail.ru; M.Sc. Saule Rakhmetullina, East Kazakhstan State Technical University named after D.Serikbayev, email: rakhmetulinas@mail.ru*

REFERENCES

[1] Sadyihov R.H., Rakush V.V., Modeli gausovyih smesey dlya verifikatsii diktora po proizvolnoy rechi, *Dokl. Belorusskogo gos. un-ta inform. i radioel*, 4 (2003), 95–103

[2] Reynolds D. A., Quatieri T. F., Dunn R. B., Speaker verification using adapted Gaussian mixture models, *Digital Signal Processing*, 10 (2000),1–3, 19–41

[3] Bykov M. M., Kovtun V. V., Vikoristannya mnozhini mikrofoniv u avtomatizovaniy sistemi rozpiznavannya movtsya kritichnogo zastosuvannya, *Visnik Vinnitskogo politehnichnogo institutu*, 3 (2017), 84-91.

[4] Kovtun V.V., Bykov M.M., Kovtun V.V., Smolarz A., Junisbekov M., Targeusizova A., Satymbekov M., Research of neural network classifier in speaker recognition module for automated system of critical use. SPIE 10445*, Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments,* (2017), 1044521

[5] Kruger S. E., Schaffoner M., Katz M., et al., Mixture of Support Vector Machines for HMM based Speech Recognition, *The 18th International Conference of Pattern Recognition ICPR'06*

[6] Golowich S. E., Sun D. X., A support vector/hidden markov model approach, *Proc. Of the Statistical Computing Section.* (1998).

[7] Collobert R., Bengio S., Bengio Y., Parallel mixture of SVMs for very large scale problems, *Proc. the Statistical Computing Section,* (2002)

[8] Ghahramani Z., Jordan M.I., Factorial hidden Markov models, *Advances in Neural Information Processing Systems,* 29 (1997), 2–3, 245–273

[9] Romanowski A., Grudzień K., Garbaa H., Jackowska-Strumiłło L., Parametric methods for ect inverse problem solution in solid flow monitoring, *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska,* 7 (2017), nr. 1, 50-54

[10] Zhang W., Liu J., Discriminative Universal Background Model Training for Speaker Recognition, *Speech and Language Technologies*, 6 (2011), 241–256

[11] Bartlett P., Shawe-Taylor J., Generalization performance of support vector machines and other pattern classifiers, *Advances in Kernel Methods* Cambridge: MIT Press, (1998), 43–54

[12] Smyth P., Heckerman D., Jordan M., Probabilistic independence networks for hidden Markov probability models, *Neural Computation*, 9 (1997), 2, 227-269

[13] Jordan M.I., Ghahramani Z., Saul Z. K., Hidden Markov decision trees, *Advances in Neural Information Processing Systems*, 8 (1998)

[14] Jensen F., Lauritzen S., Olsen K., Bayesian updating in recursive graphical models by local computation, *Computational Statistics Quarterly*, 4 (1990), 269-282

[15] Bartlett P., Shawe-Taylor J., Generalization performance of support vector machines and other pattern classifiers, *Advances in Kernel Methods*, (1998), 43–54

[16] Lach Z., Smolarz A., Wojcik W., et al., Optically powered system for automatic protection of a fiber segment, *Przegląd Elektrotechniczny*, 84 (2008), n.3, 259-262

[17] Bykov M.M., Gafurova A.D., Kovtun V.V., Doslidzhennya komItetu neyromerezh u avtomatizovanly sistemI rozpiznavannya movtsiv kritichnogo zastosuvannya, *Visnik Hmelnitskogo natsionalnogo universitetu, seriya: Tehnichni nauki*, 247 (2017), nr. 2, 144-150