ASSESSMENT OF THE POSSIBILITIES OF IMPLEMENTATION OF EMOTIONAL STATE RECOGNITION TECHNOLOGIES IN THE DISTANCE EDUCATION SYSTEM

Liudmyla Tereikovska[1]
Ihor Tereikovskyi[2]

DOI: https://doi.org/10.30525/978-9934-26-364-4-4

Abstract. One of the most promising directions in the development of distance learning systems is the use of technologies in them that allow to effectively organize the cognitive activity of students in an automatic mode for their assimilation of educational materials. For this, the current emotional state of the students should be taken into account, which indicates the expediency of introducing appropriate means of recognition into the distance learning system. *The subject of the study* are models, methods and means of recognizing the emotional state of members of the distance learning system by a person's face and voice. *The purpose of the work* is to determine the list of parameters for evaluating the effectiveness of means of recognizing the emotional state of members of the distance learning system. *The research methodology* is based on the theory of complex systems and system analysis and involves the determination of an approach to the development of technology for harmonizing the emotional state of a student of a higher educational institution, the analysis of scientific and practical works in the field of operational recognition of the face and emotional state of a user of a computer system based on the image of a face and voice, and as well as the definition of the list of parameters for evaluating the effectiveness of the relevant means of recognition. *As a result of the conducted research*, an approach to the development of a technology for harmonizing the emotional state of a student of a higher educational institution was determined, which

[1] Doctor of Technical Sciences, Associate Professor,
Professor of the Department of Information Technologies Design and Applied Mathematics,
Kyiv National University of Construction and Architecture, Ukraine
[2] Doctor of Technical Sciences, Professor,
Professor of the Department of System Programming and Specialized Computer Systems,
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine

involves a neural network analysis of the biometric parameters of a student of a distance learning system to recognize his current emotional state, which should be used to adjust the type and volume of educational materials when they are submitted in an automated mode. The expediency of using neural network analysis tools based on biometric parameters, which can be registered using common computer tools, is shown. It was determined that in the distance learning system, the specified biometric parameters should primarily include a person's face and voice. It is justified that the following should be included in the list of basic parameters for evaluating the effectiveness of means of recognizing the emotional state of a member of a distance learning system based on neural network analysis of a person's face and voice: the completeness of recognition of basic emotions; ability to recognize complex emotions; the ability to recognize an emotional state; recognition accuracy; computational complexity of recognition; the ability to recognize emotions without prior recognition of a person. As a result of the research of well-known technologies for recognizing emotions and persons by face and voice, a list of additional parameters was determined, which provide the possibility of evaluating the effectiveness of the specified technologies when they are used in the distance learning system. Taking into account the proposed performance evaluation parameters, it is shown that from the point of view of using modern emotion recognition tools in the distance learning system, their critical shortcomings are related to: low accuracy of recognition of the nomenclature of both basic and complex emotions; high computational resource intensity; inability to recognize emotions based on a complex analysis of several biometric parameters; low integration with means of identity recognition; the difficulty of recognizing an emotional state. It is substantiated that the prospects of further research in the field of the application of emotional state recognition tools in the distance learning system should be correlated with their adaptation to the expected conditions of use and their integration with the word recognition system in the member's voice signal.

1. Introduction

Over the past 10-15 years, one of the main trends in the development of higher education is the increasingly widespread introduction of distance learning. At the same time, the use of distance education as the

main form of education increases the need to increase the efficiency of the mechanisms for its implementation. The specified problem is related to the fact that conducting online broadcasts of remote lectures and practical classes significantly increases the load on the teacher, and when teaching materials are presented exclusively in the form of illustrated text or multimedia presentation, the quality of perception significantly deteriorates. According to a number of authoritative researchers [2, p. 52], the deterioration of the quality of perception of educational materials is largely related to the lack of monitoring of the emotional component of the educational process. This is confirmed by the "Yerkes and Dodson" law, which shows the dependence of a person's work performance on the degree of his emotional tension. On the basis of experimental data, it has been proven that with an increase in emotional arousal, labor productivity first increases rapidly, then this growth slows down a bit, and after reaching a certain critical level, it begins to decrease at first smoothly, and then sharply. As evidenced by the results [10, p. 6], in the process of learning, a student can feel different emotions that have different effects on the assimilation of information. Students can be surprised, excited, excited and even angry. All these emotions can have a positive effect on the learning process, despite the fact that not all emotions are positive in themselves. It is noted that the main thing is to avoid such emotions as sadness and boredom, because if the student is mastered by these emotions, then he will not be able to absorb the information well. And neutral events in education are practically not postponed in the student's memory. The appearance of sadness, boredom or apathy is possible, for example, with an uninteresting presentation of the material. An even more negative emotion in the learning process is fear, the occurrence of which not only nullifies the desire for knowledge, but also generally blocks the ability to think.

Thus, according to the ideas of modern science, one of the most promising directions for the development of distance learning systems (DLS) is the use of technologies in them that allow to effectively organize the cognitive activity of members for their assimilation of educational materials. At the same time, the emotional state of the members plays an extremely important role, which indicates the feasibility of introducing appropriate means of recognizing the emotional state in the DLS. The relevance of

the research in this direction is confirmed by the fact that emotional state recognition tools are not used in known DLS, although the possibility of their implementation is indicated by the known facts of the use of relevant technologies and software applications, which are based on neural network analysis of the image of a person's face and voice.

Based on the results [12, p. 56], the starting point for the development of emotional state recognition tools, adapted for use in DLS, should be a list of parameters that allow evaluating the effectiveness of the specified tools, which is the goal of this study. To achieve the goal of the research, the following tasks should be solved:

– To form the basic principles of technologies for recognizing the emotional state of DLS members, which allows to determine the list of basic parameters for evaluating the effectiveness of recognition tools.

– Conduct a study of well-known technologies for recognizing emotions and persons by face and voice, which allows to evaluate the effectiveness of the specified technologies when they are used in DLS.

To solve the specified tasks, the research methodology provides for the definition of an approach to the development of technology for harmonizing the emotional state of a student of a higher educational institution, the analysis of scientific and practical works in the field of fast recognition of a person and the emotional state of a computer system user based on the image of a face and voice, as well as the definition of a list parameters for evaluating the effectiveness of the corresponding means of recognition.

2. Basic principles of emotional state recognition technologies

Emotions reflect the attitude of a person to the events of the surrounding world in connection with the possibility of meeting his needs. An emotional state is a generalization of emotions over a period of time.

Emotions have two sides – objective (a person who feels emotions can smile, frown, cry, tremble, their pulse rate, breathing rhythm, etc. change), and subjective – the inner thoughts and experiences of a person with occasion of certain events. In modern psychology, four to eight basic emotions are distinguished, the manifestations of which do not depend on a person's race and culture. Basic emotions are elementary emotions that can no longer be split into anything and can themselves be components of other complex emotions. According to [7, p. 673], there are seven basic emotions:

– Joy is a positive emotional state associated with the possibility of satisfying the actual needs of a person.

– Anger is a negative emotional state caused by the sudden occurrence of a serious obstacle to the satisfaction of important needs.

– Disgust – a negative emotional state, caused by an event that sharply contradicts the moral or aesthetic principles and attitudes of a person.

– Sadness is a negative emotion associated with experiencing a negative fact.

– Fear is a negative emotional state that appears when the subject receives information about a real or imagined danger.

– Surprise – does not have a clearly expressed positive or negative sign of an emotional reaction to suddenly arising circumstances.

– Contempt – a negative emotional state that arises in relation to a negative event that is considered unworthy to be reckoned with.

A neutral emotional state (neutrality) is also distinguished, which is associated with events that do not affect a person's mood and health.

In the case of selection of four basic emotions, they include: happiness (joy), sadness (sadness), fright (fear), anger (disgust). At the same time, the number and nomenclature of complex emotions is still not clearly defined, and the corresponding theory is far from perfect.

All emotions are divided into positive and negative, as well as sthenic and asthenic. Positive emotions are associated with the satisfaction of needs, while negative emotions cause feelings of dissatisfaction and demand a change in the situation. Sthenic emotions activate the body and raise the mood (anger, rage, delight), and asthenic emotions (longing, sadness, sadness, shame) relax a person and suppress the activity of the body. At the same time, the task of recognizing emotions is poorly formalized, and its solution is complicated by the uncertainty of the characteristics of the manifestations of emotions. It is emotions that stimulate the centers of the corresponding parts of the brain, which receive signals. This emotional catalyst is necessary for the learning situation to be successful, since the member can learn the material only when emotional centers are activated in the brain. The emotional spectrum of a person is quite wide, and if the current emotional state of the member is taken into account in the DLS, then this fact can be used to influence the speed of delivery of educational materials and their volume, which will ensure an increase in the effectiveness of the

learning process. Based on the data [3, p. 51], the following functions of the emotional component of pedagogical support of the distance learning process can be determined:

1. Emotion-generative – improvement of mood, awakening of interest in educational activities and development of educational discipline.

2. Diagnostic – revealing one's own reserves of internal activity, the possibility of self-expression and self-discovery.

3. Relaxing – removal or reduction of anxiety, physical and intellectual tension, restoration of internal forces and reserves.

4. Therapeutic – overcoming learning difficulties.

As evidenced by the results of research in the field of pedagogy, psychology and project management, comprehensive coverage of the emotional component of the educational discipline can be ensured by including emotional concepts in the teaching content. Actual emotional concepts can be represented by various means of both verbal and non-verbal character [17, p. 278].

Today, the most tested approach to harmonizing the emotional state of a student of a higher educational institution is to adjust the mode of delivery of educational materials. For example, the presentation of lecture material can be carried out by the method of ready knowledge, which involves passive perception and memorization of the information provided by the teacher. Also, lecture materials can be presented by the research method, which in turn involves independent processing of the proposed lecture material: analysis of phenomena, formulation of the problem, formation and verification of hypotheses, independent formulation of conclusions.

Corrective influences on the mode of delivery of educational materials can include: the duration of the lesson, the duration of the break, the type, content, and volume of educational materials. It is obvious that the decision about the need to implement corrective influences should be made on the basis of the current emotional state of the DLS member. It is advisable to implement emotional state monitoring on the basis of BP analysis, which reflect the member's emotions and can be registered using common computer tools. This statement is based on the results of the analysis of literary works [11, p. 1141], which indicate that the vast majority of known automated and automatic emotional state recognition systems are based on emotion recognition technology based on BP analysis results.

The generalized scheme of the interaction of the components of the adjustment contour of the mode of providing educational materials in the DLS is shown in Figure 1.

It is assumed that the trainee processes the educational materials presented to him during the implementation of the educational program. The delivery mode of educational materials is adjusted by a specialized module – a controller that can change the parameters of the delivery mode taking into account the schedule of the educational program and the current emotional state of the DLS member. The current emotional state of the DLS member is recognized on the basis of the results of the analysis of his BP, which are registered with the help of computer tools that are part of the DLS software and hardware client.

Thus, the solution to the task of automating the pedagogical support of the distance learning process, taking into account the emotional component,

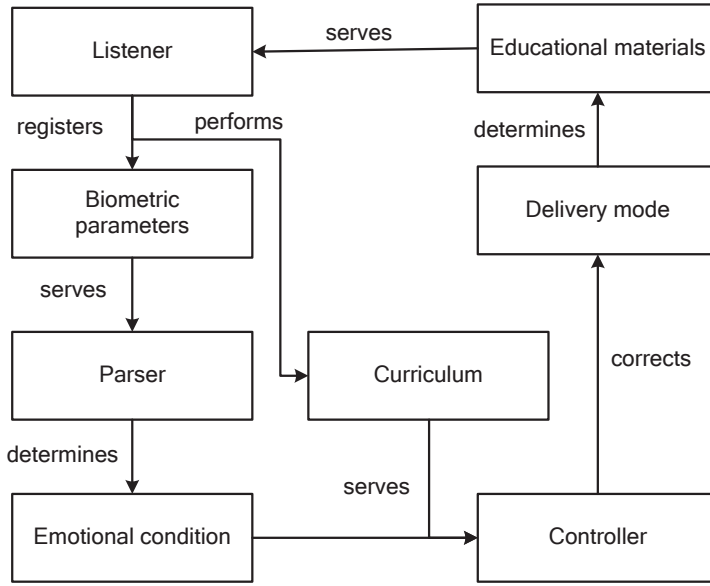

Figure 1. A generalized diagram of the interaction of the components of the adjustment contour of the mode of providing educational materials in the DLS

can lie in the development of electronic educational materials, the mode of delivery of which will be adapted to the emotional state of each specific student of the DLS. It is obvious that such developments should be based on the MARES of DLS members. At the same time, as already mentioned above, only the experimental application of the specified means of emotional state recognition in digital support systems of the educational process are known. Similar MARES are used in some schools of the People's Republic of China to recognize students who are distracted from completing educational tasks.

In common DLSs such as Moodle, ATutor, Blackboard, Lotus LearningSpace, Canvas, Sakai LMS, the mentioned emotional state monitoring tools are not available. At the same time, the possibility of using MARES in DLS is confirmed by successful examples in other applied areas. Thus, there are known attempts to use MARES in systems for recognizing the impact on the subconscious in multimedia messages of the mass media during information wars, for diagnosing mental illnesses, in the entertainment industry, and in smart home management systems.

It can also be reasonably asserted that in the basic version, as for most universal ISs, the accuracy of recognizing the emotions of the DLS member is sufficient at the level of 0.8, and the accuracy of recognizing the identity of the member during the training process is 0.9. This brings to the fore the issue of development and introduction of instrumental means of recognizing the emotional state of the member of the DLS based on the results of BP analysis, the registration of which can be implemented with the help of a common hardware and software [13, p. 1425]. First of all, these BPs include facial images and voice.

Registration of such BPs involves the use of only a video camera and a microphone, which de facto are always part of a typical DLS member's hardware and software. It should be noted that MARES based on the indicated BPs have certain shortcomings. A significant limitation to the use of facial image analysis tools is the need to eliminate characteristic obstacles associated with the angle of video recording, uneven lighting, and the resolution of the video camera. At the same time, practical experience and the results of scientific and applied works [14, p. 69] indicate the need for significant modernization of voice analyzers in the direction of reducing resource consumption, increasing recognition accuracy, expanding the nomenclature of emotions that must be recognized, reducing the

development period and increasing the level of adaptation to conditions of use. It is possible to eliminate the indicated shortcomings both due to the comprehensive analysis of all indicated BPs, and due to the improvement of existing theoretical solutions in the field of analysis of each of the BPs. At the same time, in most known cases [14, p. 68], the expected result of BP analysis is the integral recognition of the individual's face and emotional state. It should also be noted that the task of recognizing an individual's emotional state is significantly simplified in the case of preliminary recognition of the individual's face. Thus, recognition of the identity of the member is one of the important components of the task of recognizing the emotional state of the member of DLS.

In addition to the expected increase in the accuracy of emotional state recognition, the addition of face recognition tools to the DLS can cause another positive effect related to an increase in the level of protection of such a system and an increase in the level of reliability of training results. Due to biometrics, it becomes possible to realize the automatic recognition of the student's identity not only in the process of entering the DLS, but also during the educational process, in particular, during the student's processing of educational materials and during the performance of test tasks. завдань. It should be noted that the use of well-known solutions in the field of BP analysis for person and emotions recognition of the member of DLS requires their adaptation to the expected conditions of use. In order to determine the ways to improve MARES, at the first stage of research, an analysis of scientific works in this area was carried out [14, p. 70; 15, p. 325]. As a result of such an analysis, a group of basic parameters was determined, which allow to estimate the efficiency of MARES in the conditions of DLS in a first approximation. The list and explanation of the specified basic parameters are given in the Table 1.

Given in the Table 1 list of basic performance evaluation parameters (group A) is common to all types of MARES, regardless of the type of BP on the basis of which emotion recognition is implemented. In addition, the parameter related to the evaluation of the ability to recognize emotions in the conditions of interference can serve as an efficiency parameter common to all types of MARES. However, due to the significant affinity of interference with the nature of the BP, the specified parameter was specified, according to the type of BP.

Table 1

Basic parameters for evaluating the effectiveness of emotion recognition tools

№	Description of basic parameters
$A_{i,1}$	Completeness of basic emotions recognition
$A_{i,2}$	Ability to recognize complex emotions
$A_{i,3}$	Ability to recognize an emotional state
$A_{i,4}$	Accuracy of the recognition
$A_{i,5}$	Computational complexity of recognition
$A_{i,6}$	Ability to recognize emotions without prior recognition of a person

Also, as a result of the analysis of known MARES, in addition to the basic parameters allocated to group A, the need to use additional parameters that allow evaluating the effectiveness of the specified means was determined. Additional performance parameters are divided into three groups. Parameter group B characterizes features of MARES recognition, parameter group C – features of technical implementation, and group D – service (additional) capabilities of MARES.

For the basic parameters, the notation of the type $A_{i,j}$ is adopted, where i is the BP number, on the basis of which emotion recognition is implemented, j is the number of the basic efficiency parameter.

The following numbering of the BP is used: face image – 1, VS – 2. The list of parameters related to groups B-D depends on the type of BP on the basis of which recognition is implemented. For the indicated parameters, designations of the type $B_{i,j}$, $C_{i,j}$, $D_{i,j}$ are adopted, where i is the number of BP, on the basis of which emotion recognition is implemented, j is the number of the efficiency parameter within the corresponding group.

It is assumed that the application of the specified parameters will allow to implement a comprehensive assessment of the efficiency of MARES. In the basic version, each of the ratings is proposed to be implemented in the range from 0 to 1. For some i-th parameter, a rating of 0 means that the i-th requirement is not met in MARES, a rating of 0.5 means partial coverage, and a rating of 1 – full coverage such a requirement. The formation of the list of basic parameters and the approach to the definition of additional parameters made it possible to move to the next stage of research related to the analysis of emotion and personality recognition technologies to determine the ways of their improvement and adaptation to the conditions of DLS.

3. Technologies of emotion and person recognition by face

У In most cases, modern MARES are based on the recognition of facial expressions, which is defined as a set of movements of parts of a person's face that expresses his state or attitude to what he perceives. For this purpose, MARES uses a facial expression recognition module, the functioning of which is based on modern methods of computer vision, which in most cases are based on the use of different types of NNM.

The easiest way to determine an emotion from a face image is based on the classification of key points (facial landmarks), the coordinates of which can be obtained using various algorithms: PDM, CML, AAM, DPM or CNN. Usually, from 5 to 68 points are marked, tying them to the position of the eyebrows, eyes, lips, nose, jaw, which allows to partially capture facial expressions. The normalized coordinates of the points can be directly fed to a classifier (such as SVM or Random Forest) and a basic solution can be obtained. The position of the face image must be aligned. The simple use of coordinates without a visual component leads to a significant loss of useful information, therefore, to improve the system, various descriptors are calculated at these points: LBP, HOG, SIFT, LATCH, etc. After concatenation of descriptors and dimensionality reduction, the obtained feature vector can be used for emotion classification. However, such an approach is already considered outdated, as it is known that in the field of visual data analysis, deep CNNs are the most effective solution. Today there is a fairly large number of hardware and software tools designed for emotion recognition.

The **Affectiva Affdex** complex (www.affectiva.com) is designed for tracking emotions with mobile applications. Unlike most similar programs, Affdex allows data processing directly on the mobile device, and not on a remote server. To recognize 7 basic emotions, the complex uses its own database consisting of more than 5 million facial images. The work of the complex consists in recording an image of a person with subsequent wavelet transformation and comparison with samples based on SIFT descriptors.

Similar solutions are used in the **Pupeteer** software complex (www.hikvision.com), designed to assess the behavior and emotional state of students. In addition to the specified functionality, the complex allows you to recognize such movements of the head as nodding, swaying, turning from side to side. The Bayesian filtering method is used to recognize mixed

states, when it is not possible to say exactly which emotion prevails. During the experiment, the Pupeteer complex made it possible to determine 6 basic emotions with 96 percent accuracy.

The **CLMtrackr** online service (www.auduno.com/clmtrackr) allows you to identify 7 basic emotions based on facial images in real time. A webcam is used as a sensor. The functioning of the service is based on tracking 70 points on the face image, drawing lines between them and comparing these lines with samples of various basic emotions.

Microsoft's **Oxford** project offers a catalog of API functions (azure. microsoft.com/ru-ru/services/cognitive-services) based on computer vision algorithms for recognizing 7 basic emotions and a neutral state. Recognition can be carried out both on a single and on a group photo, which corresponds to a file size of about four megapixels. Photo sizes from 36×36 to 4096×4096 pixels. The number of images that can be simultaneously analyzed is up to 64. An emotion, in Oxford terms, is an 8-dimensional vector whose normalized components add up to 1. According to independent testing, the recognition accuracy is about 50%. The quality of recognition largely depends on the angle of inclination of the face, lighting and the presence of obstacles. Face images that are full-face or half-face are best recognized.

The **FaceReader** service of the Dutch company Noldus Information Technology (www.noldus.com/human-behavior-research/products/facereader) uses the Active Template method to superimpose a deformable template on the face image and the Active Appearance Model to create a model of person identification based on control points. Neural networks trained on a training corpus of 10,000 frames are used for classification. 7 basic emotions and a neutral state are defined. In addition, FaceReader is able to determine people's age, gender and ethnicity based on the image of their face. Recognition accuracy is about 75-80%. A slight tilt of the face practically does not affect the accuracy of recognition. At the same time, FaceReader does not allow recognizing the emotions of children under the age of five, classifying the image of a face in glasses, or detecting a turned face. There are also fairly strict requirements for the illumination of the object to be recognized. It is noted that FaceReader is able to recognize only the face image of one person at a time. If there are several faces in the image, they are recognized sequentially, while their position should not change.

The **Emovu** service (www.eyeris.ai) of the Eyeris (USA) company is used to determine a person's interest in content in a video stream. For this, computer vision methods based on the use of deep neural networks are used. Promising developments of this company are related to the implementation of SARES in autonomous unmanned vehicles, where it is necessary to determine the sensations, needs of passengers, as well as the dynamics of human states over time.

The University of Toronto project **NuraLogix** (www.nuralogix.ai) allows recognition of hidden emotions based on the analysis of color changes in the subcutaneous pattern of blood vessels in facial images.

The **eMotion Software** system (www.visual-recognition.nl/eMotion.html), developed by Visual Recognition (Netherlands), allows you to recognize 6 basic emotions (anger, sadness, fear, surprise, disgust and happiness) and calculate their integral value. The system creates a 3D model of the face with 12 key areas highlighted on it, such as the corners of the eyes and mouth. The recognition algorithm is not described.

The **MMER_FEASy** system – the FacE Analysis System (www.mmer-systems.eu/en/application-areas). Development company MMER-Systems (Germany). The development uses the Active Appearance Model methodology of superimposing a certain deforming mask on the face image, which allows you to calculate the necessary parameters in real time. The program recognizes 6 basic emotions, and also allows you to determine age, gender and ethnicity. In addition, the system can identify a person.

Similar characteristics are known for MARES **FaceSecurity** (www.cognitec.com) of Cognitec (Germany). Declared system capabilities:

– high speed of comparisons of the analyzed image with existing samples (about 900,000 comparisons per second on hardware of average power);

– use of a wide range of registration tools, image formats;

– integration with modern database management systems of families Oracle, IBM DB2, MSSQL Server, which allows efficient use of large image databases.

Disadvantages of the system include the ability to recognize only frontally located faces (possible deviations up to 15 degrees). Another disadvantage is high sensitivity to lighting. In addition to web applications and SDKs for mobile phones, Cognitec provides an API for digital billboards to display ads to target audiences. In addition, Cognitec systems are used in cars to analyze the face image of drivers and passengers, as well as to detect the

position of the head, detect distracted gaze, and detect closed eyes. As a result of the analysis of known MARES based on facial image analysis, a list of additional parameters was formed that characterize the features of the recognition process, features of their technical implementation and service (additional) capabilities. A brief description of additional parameters that were obtained is presented in the Table 2.

Table 2

Additional parameters for evaluating the effectiveness of MARES based on the image of the face

№	Description of parameter
1	**2**
Features of recognition	
$B_{1,1}$	Ability to recognize emotions from a fragment of a face image
$B_{1,2}$	Resistance to changes in lighting
$B_{1,3}$	Resistance to interference in the face image
$B_{1,4}$	Ability to recognize a distorted face
$B_{1,5}$	Resistance to image rotation
$B_{1,6}$	Resistance to image tilt
$B_{1,7}$	Ability to analyze a group image
$B_{1,8}$	Ability to analyze a color image
$B_{1,9}$	The possibility of automatic selection of the contour of the image
$B_{1,10}$	Ability to analyze a black and white image
$B_{1,11}$	Ability to analyze the image in shades of gray
$B_{1,12}$	Ability to analyze a two-dimensional image
$B_{1,13}$	Ability to analyze a three-dimensional image
$B_{1,14}$	Resistance to image resizing
Technical implementation	
$C_{1,1}$	The possibility of integration with DBMS
$C_{1,2}$	Possibility of integration with different types of video cameras
$C_{1,3}$	Ability to analyze various image formats
$C_{1,4}$	Ability to analyze the video stream
Service capabilities	
$D_{1,1}$	Recognition of person of the user
$D_{1,2}$	Conversion of a two-dimensional image of a person into a 3D model
$D_{1,3}$	Definition of race
$D_{1,4}$	Determination of gender

The results of the analysis made it possible to determine the values given in the Table 1. efficiency parameters for each of the analyzed MARES operating on the basis of facial image analysis.

The obtained estimates for the first group of characteristics are shown in the Table 3.

Table 3

Evaluation of the first group of parameters determining the effectiveness of MARES based on the image of the face

Name of MARES	Parameter evaluation					
	$A_{1,1}$	$A_{1,2}$	$A_{1,3}$	$A_{1,4}$	$A_{1,5}$	$A_{1,6}$
1	**2**	**3**	**4**	**5**	**6**	**7**
Affectiva Affdex	1	0	0	0,7	0,5	0,9
Pupeteer	1	1	1	0,9	0,5	1
CLMtrackr	1	0	0	0,9	0,5	1
Oxford	1	0	0	0,9	0,7	1
FaceReader	1	0	0	0,9	0,7	0,9
Emovu	1	1	1	0,8	0,7	1
NuraLogix	1	1	0	0,7	0,7	1
eMotion Software	1	1	1	0,9	0,8	1
MMER_FEASy	1	0	0	0,8	0,7	1
FaceSecurity	1	1	1	0,9	0,8	1

As follows from the analysis of the data in the Table 3, all analyzed MARES are not adapted to the recognition of emotions based on the analysis of several BPs, which makes it impossible to use them in the case of the absence or low-quality video registration of the face image of the DLS member. At the same time, most of the known MARES do not recognize complex emotions, and also do not allow determining the current emotional state. In addition, the analysis of the identified MARES revealed their low adaptability to changes in the level of illumination and to the possibility of recognizing emotions on partially noisy facial images. It should be noted that regardless of the nature of the interference, their negative impact is that areas appear on the image that actually distort the image.

4. Emotion and person recognition technologies by voice

The analysis of the scientific and practical works devoted to the development of means of analysis of VS indicates that the tasks traditionally performed by them are reduced to the recognition of individual words and the personality of the announcer. Recently, the range of these problems has significantly expanded. First of all, the need to recognize the emotional and functional state of the announcer was added. In addition, the directions of determining mental abnormalities, diagnosing cardiovascular diseases, and assessing the reliability of the announcer's information are considered promising.

Various aspects of the problem of analysis of human VS were considered by a number of domestic and foreign scientists [4, p. 25; 5, p. 508]. However, in well-known works in the analysis of VS, emphasis was placed on increasing the effectiveness of recognizing words and the speaker's person [16, p. 4052]. Other issues, due to their novelty, have been less well researched. At the same time, the data [6, p. 812] indicate that the recognition of emotions in the voice is associated with the analysis of the intonation and timbre of the voice. It should be noted that the intonation of GS is a collective concept that includes tonality, strength of sound, pace of speech, accents and pauses. From the point of view of emotion recognition, the most important components are the tonality of speech and speech tempo. The tonality of the VS reflects the emotional and volitional attitude of a person when achieving a specific goal. According to the type of influence, three types of tonality are distinguished:

1. Major, which is characterized by positive emotions, active energy, general optimistic mood.

2. Minor, which is characterized by negative emotions, weak energy, general pessimistic mood.

3. Neutral – characterized by the objectivity of the presentation, the minimal importance of the psychological component of language communication.

Such characteristics of tonality as calmness and tension are distinguished. At the same time, calm is a relatively even state, and tension is an unstable state of a certain tonality, characterized by the accumulation of a certain number of signals of another tonality, the further increase of which can lead to a qualitative jump and transition to another tonality.

The pace of speech reflects the speed of pronouncing various elements of speech (sounds, syllables, words).

The timbre of the voice is a psychoacoustic characteristic determined by the formant structure of the VS spectrum, the degree of expressiveness of noise components, inharmonic overtones, modulation processes, growth and fading of the VS. Like the tonality of the VS, the timbre of the voice reflects the brightness of the sound, its individuality and is largely determined by the current psychological state of a person. In most of the known MARES, the tonality and timbre of the VS are evaluated based on the pitch of the fundamental tone and overtones.

As evidenced by the data [9, p. 306], the operation of common systems of analysis of VS is based on mathematical models based on the Bayesian approach, hidden Markov processes, support vector machines, dynamic programming methods, as well as NN theory. At the same time, the data of modern literary sources [11, p. 1141], and the results of the research of the most well-known tools for a similar purpose (Google+, Microsoft Office, VoiceNavigator, Siri) allow us to claim that the neural network methodology for the analysis of VS has the broadest prospects.

At the same time, in the analyzed literature there is no consensus on which type of NNM is appropriate to use to solve the problem of recognizing a person and recognizing his emotions. The results of the analysis of literary works allow to state that the vast majority of known systems for automated and automatic recognition of a person's emotional state are based on emotions recognition technology, which revolves around the results of processing the spectrum of VS. The same results are used to recognize a person's face by VS. In most cases, we are talking about the analysis of mel-cepstral coefficients of the VS [11, p. 1143]. An analysis of the most well-known means of recognizing human emotions by voice was also carried out.

Computer system of the Higher School of Economics (perm.hse.ru) allows to recognize 8 basic emotions based on the analysis of VS – neutrality, calmness, happiness, sadness, anger, fear, disgust and surprise. The claimed average recognition accuracy reaches 71%. At the same time, the accuracy of recognition of neutrality and calmness is higher, and the accuracy of recognition of fear, sadness, surprise and disgust is below average. The system is experimental in nature. The basis of the speaker-

independent recognition block is a recurrent NN of the LSTM type, which is used for the analysis of the VS presented in the form of a set of mel-cepstral coefficients. The system provides filtering of white noise and equalization of the volume of the VS.

The real-time emotion tracking technology Beyond Verbal (www.beyondverbal.com), created in Israel, provides the ability to determine about 400 variants of the announcer's emotional state. The technology is used by a number of companies and involves the use of various VS processing mechanisms. The final classification of emotions is implemented with the help of special templates, for the creation of which 2,300,000 audio recordings of VS from 170 countries of the world were analyzed. The technology provides the possibility of both speaker-dependent and speaker-independent recognition of emotions.

The Xpression mobile application developed by EI Technologies from the UK (www.ei-technologies.com) is able to recognize 5 human emotions: calmness, happiness, sadness, anger and anxiety. The application has a client-server architecture. The client module is used for registration of VS and its preliminary processing. On the server, fragments of VS with a duration of 200 milliseconds are analyzed. For this purpose, a NN is used, the input of which is fed with parameters describing the volume of the VS, its tonality and speech tempo. The recognition process is speaker-dependent, which involves the creation of a specific user template. In the future, the application monitors deviations in the timbre of the voice, its volume, changes in the time intervals between sounds and words, and compares all this with the defined template. Classification is carried out on the basis of calculated deviations. During the development process, it was determined that representatives of different nations speak at different speeds and volumes. In addition, the VS of people of different ages and different sexes has different intonations.

Articles [1, p. 305] describe an experimental **system for recognizing emotions in human speech** using wavelets. The system is designed to recognize 4 basic emotions in audio recordings: anger, fear, joy and norm. A detailed description of the mathematical support of the system is given. Recognition is implemented on the basis of the values of correlation coefficients, for the calculation of which spectral analysis of VS based on the Morlet wavelet is used.

The article [8, p. 45] describes **a program for measuring the emotions of a mobile phone subscriber**. The result of the program is a graph of the subscriber's emotional curve, which can be used to recognize his emotions by comparing them with the emotional curves of basic and complex emotions.

The **EmoWatch** program from the Japanese company Smartmedical (eng.smartmedical.jp) is designed to identify and recognize the emotions of Apple mobile device users. The program is based on Empath technology, which analyzes the main tone, volume and speed of the VS. Thanks to this information, the program allows you to analyze the user's energy levels and recognize 4 types of emotions – anger, calm, joy and sadness. A feature of the program is its speaker independence and the possibility of recognition regardless of the user's language. In addition to the considered emotion recognition systems, a number of complexes are known, the functionality of which is also based on VS analysis. These complexes are designed to solve the following tasks: user verification, determination of a person's functional state, lie detection, creation of various voice effects, determination of race, age, and gender.

As a result of the analysis of known emotion recognition tools based on VS, a list of additional parameters was formed that characterize the features of the recognition process, features of technical implementation, and service capabilities. A brief description of the obtained additional parameters is presented in the Table 4.

The results of the analysis made it possible to evaluate the capabilities of the analyzed MARES by voice using the selected parameters. The assessment was carried out similarly to the assessment of MARES based on the image of the face. Focusing attention on the first group of characteristics is dictated by the fact that the subject of research is related to the development of effective models and methods for recognizing emotions expressed in VS. The obtained estimates for the first group of characteristics are shown in the Table 5.

As follows from the analysis of the data in the table. 5, all analyzed MARES are focused on recognition of emotions exclusively by VS, which significantly narrows the scope of their application in DLS. At the same time, most of the known MARES do not allow recognition of the entire range of basic emotions and are not adapted to recognize complex emotions.

Table 4

Additional parameters for evaluating the effectiveness of MARES by voice

№	Parameter description
1	2
Features of recognition	
$B_{2,1}$	Adaptability to the announcer's language
$B_{2,2}$	Adaptability to the announcer's gender
$B_{2,3}$	Adaptability to the announcer's age
$B_{2,4}$	Ability to filter noise
$B_{2,5}$	The possibility of normalizing the volume of the VS
$B_{2,6}$	Possibility of recognition in polyphonic conditions
$B_{2,7}$	Narrator independence of the recognition process
$B_{2,8}$	Ability to analyze spontaneous speech
$B_{2,9}$	The need to use educational databases containing examples of VS with different emotions
Technical implementation	
$C_{2,1}$	The possibility of integration with DBMS
$C_{2,2}$	The possibility of integration with various audio stream registration systems
$C_{2,3}$	Ability to analyze various audio file formats
Service capabilities	
$D_{2,1}$	Recognition of the user's identity
$D_{2,2}$	Determination of age
$D_{2,3}$	Determining the speaker's language
$D_{2,4}$	Determination of gender

Table 5

Evaluation of the first group of parameters determining the effectiveness of MARES by voice

Name of MARES by voice	Evaluation of parameter					
	$A_{2,1}$	$A_{2,2}$	$A_{2,3}$	$A_{2,4}$	$A_{2,5}$	$A_{2,6}$
Computer system of the Higher School of Economics	0,9	0	1	0,7	0,8	1
Beyond Verbal	0,9	1	1	0,8	0,7	0,7
Xpression	0,7	0	1	0,8	0,8	0
System for recognizing emotions in human speech	0,5	0	1	0,8	0,7	1
Program for measuring the emotions of a mobile phone subscriber	0,5	0	1	0,8	0,7	1
EmoWatch	0,5	0	1	0,8	0,9	1

In addition, practical experience indicates the indicates the need for a significant modernization of MARES in the direction of reducing resource consumption, increasing the accuracy of recognition, expanding the nomenclature of both basic and complex emotions that must be recognized, shortening the development period and adaptability to use in the conditions DLS It was determined that the possibilities of modernization are closely related to the use in MARES by voice of modern solutions in the field of neural network analysis of VS, based on the application of DLS and modern types of recurrent NN.

5. Conclusions

As a result of the conducted research, an approach to the development of a technology for harmonizing the emotional state of a student of a higher educational institution was determined, which involves a neural network analysis of the biometric parameters of a student of a distance learning system to recognize his current emotional state, which should be used to adjust the type and volume of educational materials when they are submitted in an automated mode

The expediency of using neural network analysis tools based on biometric parameters, which can be registered using common computer tools, is shown. It was determined that in the distance learning system, the specified biometric parameters should primarily include a person's face and voice.

It is justified that the following should be included in the list of basic parameters for evaluating the effectiveness of means of recognizing the emotional state of a member of a distance learning system based on neural network analysis of a person's face and voice: the completeness of recognition of basic emotions; ability to recognize complex emotions; the ability to recognize an emotional state; recognition accuracy; computational complexity of recognition; the ability to recognize emotions without prior recognition of a person.

As a result of the research of well-known technologies for recognizing emotions and persons by face and voice, a list of additional parameters was determined, which provide the possibility of evaluating the effectiveness of the specified technologies when they are used in the distance learning system.

Taking into account the proposed performance evaluation parameters, it is shown that from the point of view of using modern emotion recognition tools in the distance learning system, their critical shortcomings are related to: low accuracy of recognition of the nomenclature of both basic and complex emotions; high computational resource intensity; inability to recognize emotions based on a complex analysis of several biometric parameters; low integration with means of identity recognition; the difficulty of recognizing an emotional state.

It is justified that the prospects of further research in the field of the application of means of emotional state recognition in the distance learning system should be correlated with their adaptation to the expected conditions of use and to the leveling of typical disturbances, which are characteristic when registering the image of the face and voice of the member. It also seems appropriate to develop a system for recognizing words in the voice signal of a member of a distance learning system, which is an extremely important factor in recognizing the member's emotions by voice.

References:

1. Bello J., Pickens J. (September 11-15, 2005). A Robust Mid-Level Representation for Harmonic Content in Music Signals. *6th International Conference on Music Information Retrieval*. London, UK, pp. 304–311.

2. Bushuev S. D., Rusan N. I. (2016). Emotional intelligence of the project manager. Management of the development of complex systems, is. 28, pp. 47–56.

3. Bykov V., Burov O., Gurzhii A., Zhaldak M., Leshchenko M., Lytvynova S., Lugovoi V., Oliynyk V., Spirin O., Shishkina M. (2019). *Theoretical and methodological principles of informatization of education and practical implementation of information and communication technologies in the educational sphere of Ukraine*. Kyiv: Comprint, 214 p.

4. Hinton G., Sabour S., Frosst N. (2018). Matrix capsules with EM routing. *ICLR Conference*. Vancouver, BC, Canada.

5. Hu Z., Tereikovskyi I., Korystin O., Mihaylenko V., Tereikovska L. (2021). Two-Layer Perceptron for Voice Recognition of Speaker's Identity. *Advances in Intelligent Systems and Computing*. Springer, Cham., vol. 1247, pp. 508–517.

6. Juslin P., Laukaa P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code. *Psychological bulletin,* vol. 129, no. 5, pp. 770–814.

7. Matsumoto D., Hwang H., Ekman P. (2021). Encyclopedia of Behavioral Neuroscience: Second Edition, pp. 671–677.

8. Mykhaylov V. O. (2012). Software implementation of measuring the emotions of a mobile phone subscriber. *Modern problems of science and education*, no. 5, pp. 45–62.

9. Penagarikano M., Bordel G. (2004). Layered Markov models: a new architectural approach to automatic speech recognition. Proceedings of the IEEE Signal Processing Society Workshop Machine Learning for Signal Processing, pp. 305–314.

10. Rusan N., Bushuev S. (2021). Application of empathy and emotional intelligence for different types of teams in pandemic conditions. *Bulletin of the National Technical University "KhPI". Series: Strategic management, management of portfolios, programs and projects*, no. 1(3), pp. 3–8.

11. Tereikovska L., Tereikovskyi I., Beketova A., Karaman G., Makovetska N. (March 2021). Recognition of Speaker's Emotion by Squeezenet Convolutional Neural Network. *Journal of Theoretical and Applied Information Technology*, vol. 99, no. 5, pp. 1139–1148.

12. Tereykovskaya L., Tereykovskiy I., Aytkhozhaeva E., Tynymbayev S., Imanbayev A. (2017). Improvement of learning efficiency of the neural networks, intended for recognition of graphic images in systems of biometric authentication. *Computer Modelling & New Technologies*, vol. 21, is. 2, pp. 54–57.

13. Tereikovska L., Tereikovskyi I., Mussiraliyeva S., Akhmed G. (2019). Recognition of emotions by facial geometry using a capsule neural network. *International Journal of Civil Engineering and Technology*, vol. 10, is. 03, pp. 1424–1434.

14. Tereikovskyi I., Mussiraliyeva S., Tereikovska L., Chernyshev D., Nyussupov A., Abaiuly Y. (28-30 April 2022). The Procedure of Adapting the Design Parameters of the Convolutional Neural Network During the Speaker's Emotions Recognition. *International Conference on Smart Information Systems and Technologies.* Nur-Sultan, pp. 68–73.

15. Toliupa S., Tereikovskiy I., Dychka I., Tereikovska L., Trush A. (2019). The Method of Using Production Rules in Neural Network Recognition of Emotions by Facial Geometry. *3rd International Conference on Advanced Information and Communications Technologies*, pp. 323–327.

16. Variani E., Lei X., McDermott E., Lopez Moreno I., Gonzalez-Dominguez J. (2014). Deep neural networks for small footprint text-dependent speaker verification. *IEEE International Conference on Acoustic, Speech and Signal Processing*, pp. 4052–4056.

17. Yudin O., Toliupa S., Korchenko O., Tereikovska L., Tereikovskyi I., Tereikovskyi O. (2020). Determination of Signs of Information and Psychological Influence in the Tone of Sound Sequences. *IEEE 2nd International Conference on Advanced Trends in Information Theory,* pp. 276–280.