JSC “Civil Aviation Academy”

UDC 629.7.066 Manuscript copyright

ANAYATOVA RAZIYAM KURVANZHANOVNA

The methodology for reducing the impact of the human factor on flight safety

6D071400 – Aviation Engineering and Technology

Thesis for the degree
of Doctor of Philosophy (PhD)

Academic adviser
Doctor of Technical Sciences, Professor
Koshekov K.T.

Foreign academic adviser:
Doctor of Philosophy in Engineering,
Professor Streltsov A.V.

Republic of Kazakhstan
Almaty, 2021

CONTENTS

REGULATORY REFERENCES

This dissertation uses references to the following standards:

Human Factors Training Manual, Doc 9683-AN/950, first edition - 1998.

Decree of the Government of the Republic of Kazakhstan. On the state program “Digital Kazakhstan”: approved on December 12, 2017, No. 827.

Global Flight Assurance Plan. June 2007. International Civil Aviation Organization.

Law of the Republic of Kazakhstan No. 339-IV of July 15, 2010 “On the use of airspace of the Republic of Kazakhstan and aviation activities”.

Doc 9806: Basic Principles of Human Factors in Flight Safety Inspection Guidelines. First edition, 2002. International Civil Aviation Organization.

Order of the acting Minister of Transport and Communications of the Republic of Kazakhstan dated May 16, 2011 No. 279. On approval of the Instruction on the organization and maintenance of air traffic. Registered with the Ministry of Justice of the Republic of Kazakhstan on June 13, 2011 No. 7006.

Doc. 9859 AN/474: Safety management guidelines. – Ed. 2nd – Montreal: ICAO, 2010.

Doc. 9835 AN/453: Guidelines for implementing ICAO language proficiency requirements. – Ed. 2nd. – Montreal: ICAO, 2010.

TERMS AND ABBREVIATIONS

IATA	– International air transport association
HF	– Human factor
FS	– Flight safety
APP	– Atomic Power Plant
ICAO	– International civil aviation organization
PES	– psycho-emotional state
CRM	– crew resource management
LOFT	– line oriented flight training
SHELL	– software, hardware, environment, liveware
ACT	– aircraft crew team
ATC	– air traffic control
AA	– aviation accident
AC	– aircraft crash
AC	– aircraft
DSP	– digital signal processing
CAL	– computer assisted learning
ADC	– analog-to-digital conversion
VAD	– voice activity detector
STFT	– Short Time Fourier Transform
WDFT	– windowed-discrete Fourier transform
LPC	– linear prediction coefficient
MFCC	– mel-frequency cepstral coefficient
DCT	– discrete cosine transform
CNN	– convolution neural network
DCNN	– deep convolutional neural network
UNO	– united nations organization
RTC	– radiotelephony communication
ESL	– English as a second language
AAIB	– Air Accidents Investigation Branch
CRM	– crew resource management
NS	– native speaker
NNS	– non-native speaker
LPR	– language proficiency requirements
RW	– runway
IMC	– instrumental meteorological conditions

INTRODUCTION

Relevance of the study

It is generally recognized that safety of transport is one of the most urgent problems, and more than 80% occurs under the influence of the human factor.

The concept of “human factor” in civil aviation is extremely multifaceted and complex, and is associated with the problem of accidents, since it is the speed and correctness of human decision-making that determines people’s life and health [1].

How acute this problem is in the aviation industry of the Republic of Kazakhstan can be judged at least from the sources on the analysis of the state of flight safety, where it is indicated that the reasons in more than 90% of cases of aviation accidents were incorrect, including violations by the crews of the established rules for performing flights, in-flight decisions and poor-quality actions in the operation of aviation equipment, insufficient training of aviation specialists (pilots and air traffic controllers) in English, errors in piloting techniques, etc.

The analysis of works devoted to the present topic of research has revealed the existence of several solutions to the above problem by applying new methods of psychodiagnostics and training [2], increasing the responsibility of the person himself, his psychophysiological, psychological and behavioural capabilities.

However, the proposed solutions are not effective enough, since only professional training and compliance with some medical indicators are taken into account when admitting aviation personnel to the operation and maintenance of aviation equipment. At the same time, there is no external adequate assessment of moral and psychological qualities in the absence of perfect techniques for recognising emotional states.

The formulated problem and established conclusions formed the basis for theoretical research on the evaluation of statistical information of aviation accidents, on summarizing the achievements of the world aviation powers and domestic researchers in reducing the influence of the “human factor”, on providing a highly efficient flight safety management system based on the application of artificial intelligence.

As a result, it has been found that a possible reduction of a human factor influence on flight safety can be achieved by applying two approaches. Firstly, by improving not only aviation English, but also plain English for pilots and dispatchers as well as for engineering and technical personnel and flight attendants; Secondly, by applying methods of psycho-emotional state recognition based on digital processing of speech signals and intelligent data analysis.

Thus, based on the current state of the problem, dissertation research aimed at developing a methodology to reduce the impact of the human factor on flight safety and aviation accidents is currently very important and relevant.

The aim of this thesis is to develop theoretical and methodological approaches, scientific and practical recommendations to improve the safety of flying on the basis of reducing the impact of the human factor, by creating and optimizing methods for digital processing and intelligent analysis of speech signals of phrasal verbs in aviation and plain English.

To achieve this goal, the following **tasks** are solved within the framework of the dissertation work:

1. Study of the psychological safety of workplace on production plant.
2. Research and analysis of the impact of the human factor (HF) on flight safety (FS), determination of human errors in aircraft maintenance and development of a safety system at aircraft enterprises based on the assessment and analysis of the risks of psychological safety of workplace.
3. Investigation of the possibility of assessing the psycho-emotional state (PES) of aviation personnel using fixed phraseological units of aviation and plain English.
4. Study of the processes and phenomena underlying the mechanism of speech formation. Study of the structure of a speech signal in the time and frequency domains. Analysis of the peculiarities of acoustic perception waves by the human hearing organs from the position of the theory of signals and systems.
5. Development of an integrated approach to intelligent automatic speaker-independent recognition of a person's emotional state by a speech signal when using aviation English.
6. Formation of a representative corpus of speech signals for seven archetypic types of PES in English, characterized by a variety of utterances used, the number of speakers presented, and the variability of the degree of PES manifestation.
7. Development of the structure and substantiation of the main stages of the preliminary digital processing of speech signals process (preprocessing) for the selection of informative features in automatic recognition of speaker’s emotional state tasks.
8. Search and determination of the most informative signs of speech signals that can provide the maximum information increase in the emotional state classifying tasks of a person by voice.
9. Development and experimental study of intelligent algorithms for the classification of speech signals for automatic detection of PES without the need to recognize the semantic load of the analyzed statements.
10. Effectiveness evaluation of the developed algorithms and methods for intelligent analysis of speech signals in comparison with existing solutions for speaker-independent recognition of PES by a voice.
11. Development of a methodology that includes the rules for the formation of fixed phraseology, an integral technique and algorithm for reducing the impact of HF on flight safety based on aviation personnel psycho-emotional state recognition by a speech signal.

Research methods

To solve the problems posed in the dissertation research there were widely used such methods as analytical research and mathematical statistics, signals and systems theory, digital signal processing, spectral short-term, as well as correlation and probability-analysis method. Методы исследования

To build a mathematical model classifier, we used intelligent methods of pattern recognition based on machine learning algorithms, an apparatus for synthesizing deep convolutional neural networks, tools for computer and mathematical modeling, linear algebra, and a set of optimization methods. In the experimental part of the work there were actively used mathematical calculations automating means and results visualizing tools based on the Python 3 language programming.

The methodology was developed using the expertise and advice of leading civil aviation experts.

The subject of research focuses on PES recognition technologies to improve flight safety.

The object of the research is speech signals for the assessment of psycho-emotional state with issuing expert opinions.

Scientific novelty. The most significant new scientific results contained in the thesis are:

1. Justification of applicability of the approach to reducing the impact of PF on flight safety on the basis of recognizing PES of aviation personnel on the steady phraseological phrases of aviation English, which differs from the known ones in that it allows objectively, by informative parameters, to determine the levels of stress and fatigue, recognize depressed states, prevent fatigue.

2. A general process and methods of sequential digital pre-processing of speech signals have been developed and justified for the classification of emotional state of the speaker based on said signals.

3. Important informative features are proposed, as well as a method for extracting and presenting them for automatic classification of PES by voice signals using data mining methods.

4. A general approach to solving the problem of automatic classification of human PES on the basis of voice signals is proposed by the methods of intelligent information analysis.

5. A mathematical model of the PES classifier based on speech signal has been developed, which is based on the use of complex information from two deep convolutional neural networks trained on various informative features.

6. A comprehensive methodology is proposed for increasing flight safety by taking expert corrective actions based on emotional condition of aviation personnel assessments as pilots, dispatchers, engineers and flight attendants.

Practical relevance.

The practical significance of the work consists in the possibility of applying the results obtained in the study to build automatic systems for the aviation industry, allowing to perform recognition of the emotional state of a person by a speech signal. This includes determining the level of stress and fatigue, recognizing depressive states, and keeping from tiredness.

The results of the presented study allow us to develop new effective decision support systems for human personnel, aimed at reducing the risk of errors and lowering attention.

The preprocessing structure proposed in the paper allows to effectively implement in practice the process of digital processing of speech signals for the subsequent application of intelligent methods of information processing.

The comprehensive approach to intelligent automatic PES recognition by speech signal presented in the thesis research can be applied in other spheres of human activities related to the operation of complex technical systems with man-machine interfaces.

The proposed intelligent system for the recognition of PES of aviation personnel, allows, among other things, to provide noise-resistant recognition of speech signals of complex form, to build up databases, to give various quantitative and qualitative characteristics.

Relationship with government programs

In the Address of the President of the Republic of Kazakhstan - the Nation Leader Nursultan Nazarbayev to the people of Kazakhstan “The Strategy “Kazakhstan-2050”: the new political course of the established state” (Astana, Akorda, 2012) a special role is allocated to the development of transport and logistics potential. In this regard, this dissertation work is fully consistent with the formulated requirements for the modern direction of air transport development.

In the Head of the State’s Message “Constructive public dialogue is the basis of stability and prosperity of Kazakhstan” Astana, Akorda, 2019, to the people of Kazakhstan the President of the Republic of Kazakhstan Kassym-Zhomart Tokayev drew the attention of the executive power to the full and high-quality implementation of Infrastructure Development State Program “Nurly Zhol” for 2020 - 2025 years, approved by Government Decree of the Republic of Kazakhstan dated December 31, 2019 No. 1055. This strategic project states that effective modernization will affect the entire transport infrastructure.

State program “Digital Kazakhstan” approved by the Decree of Kazakhstan Government No. 827 dated December 12, 2017, implies accelerating the pace of economy development of the Republic of Kazakhstan and improving the quality of population life through the use of digital technologies in the medium term, as well as creating conditions for the transition of Kazakhstan economy to a fundamentally new trajectory of development, ensuring the creation of the digital economy of the future in the long term.

The proposed methodology for reducing the impact of HF on air safety and the intelligent voice PES recognition method are consistent with the aviation concept for air safety and to meet the needs of the economy of the Republic of Kazakhstan, Individuals and entities in aviation services.

Scientific research presented in the dissertation work was carried out within the framework of grant financing of MES RK on the topic “AP08857126 - Development of a complex of interactive training programs on technological processes of aircraft repair”.

Thesis provisions to be defended:

1. A comprehensive approach to intelligent automatic recognition of PES by a speech signal to reduce the impact of HF on flight safety.

2. The process of speech signal preprocessing at the stage of extraction of informative features for automatic classification of human emotional state.

3. The method of using informative features and the form of their representation to build a model of multi-class classifier in the task of PES recognition by speech signal.

4. A mathematical model of a multiclass classifier for determining the emotional state of a speaker by his speech signal based on synthesized deep convolutional neural networks trained on various types of informative features.

5. Methods for reducing the impact of HF on flight safety based on the recognition of the emotional state of aviation personnel using steady phraseology of aviation English.

The approbation of results.

The results of the thesis research have been implemented in the training process of the Civil Aviation Academy and in the “Sunkar Air” LLP industry to improve flight safety as well as taken into account in the development of the Strategic Development Plan of the company until 2025, which aims to improve the profiling technology based on artificial intelligence in the aviation security system.

The main results of the thesis are reported and discussed at: the XIV-International Scientific and Technical Conference “Dynamics of Systems, Mechanisms and Machines” (Omsk, 2020); VII-International Scientific and Practical Conference “Science and Education in the Modern World: Challenges of the 21st Century” (Nur-Sultan, 2020); International Scientific Conference “V International Scientific-Practical Conference “Integration of the Scientific Community to the Global Challenge so four Time” (Tokyo, 2020); IV-International scientific-practical conference “Scientific and technical aspects of innovative development of the transport complex” (Donetsk, 2018); International scientific-theoretical conference of students and young scientists “Rukhani zhangyru-the choice of the President, supported by the society” and the World Cosmonautics Day (Almaty, 2018); International scientific-theoretical conference of students and young scientists of the Civil Aviation Academy (Almaty, 2017); International conference “1st international pre-service teachers conference “Teaching and learning English in transition to tri-lingual education: research, challenges and success” (Shymkent, 2018); III International Scientific and Practical Youth Conference “Creative potential of youth in solving aerospace problems” (Baku, 2018).

Publications.

The main results of the dissertation research are reflected in 24 scientific papers, including 8 articles published in the editions recommended by the Committee for Quality Assurance in Education and Science of MES, 3 articles are in the international scientific journals indexed in the Scopus database, 8 papers reflected in the proceedings of international scientific conferences, 5-in international and national scientific peer-reviewed journals, including specialized in the field of aviation technology and engineering.

The author's personal contribution.

The author independently obtained the main results of theoretical and experimental studies. In published scientific works as part of the team of co-authors, the applicant is the main contributor in receiving, summarizing and analysing the achieved results.

The structure of the thesis.

The dissertation has a classical structure: the introductory part, main part (four chapters), a conclusion, the list of cited references and applications. The work is presented on 120 pages of computer text, includes 36 figures, 10 tables and 130 titles of bibliographic sources.

The main research findings.

The thesis study provides a theoretical basis and proposes a solution to the current scientific problem of reducing the influence of the human factor (HF) on the safety of air transport system based on the application of artificial intelligence. As a result of statistical analysis of accidents and incidents the professional group of aviation personnel, insufficient level of knowledge of aviation or plain English, as well as their improper actions, is defined HF, which affect the flight safety. These are pilots, dispatchers, engineering staff and flight attendants.

According to the results of the study, it was found that there are certain characteristics of the behavior of aviation personnel, which are reflected in the speech signal.

A new approach to solving the problem of reducing the number of accidents and incidents by determining the psycho-emotional states (PES) of aviation personnel on the basis of speech signal recognition, as this characteristic is individual, easy to measure, and hardware and software implementation of analysis and processing has low cost and is applicable to a wide range of tasks.

Machine learning theory methods have been found to be an effective intellectual technology for automatically classifying PES by speech, as they allow the detection of hidden patterns in data, including some uncertainty. In order to create a representative set of educational data, an emotionally colored speech box was created for the seven classes in aviation and plain English, characterized by a diversity of narrators of both sexes spoken in a set of phrases, emotional color.

A new discrete model of speech formation is proposed for the selection of informative features in the preprocessing structure. The special procedures of digital signal processing for prefiltering and pause removal are proposed. This allowed establishing the features of the objects for training the mathematical model of the classifier, as they contain information about emotional coloration of speech.

The architecture of the deep-convolutional neural network (DCNN) and the algorithm of its training on selected informative attributes, allowing obtaining high results of classification of PES of aviation personnel for seven classes of objects only by acoustic data of the studied samples were defined. To improve PES classification parameters, a method is proposed that combines the results of classification from two DCNN trained on different types of informative attributes as mel-spectrograms and mel-frequency cepstral coefficients. The result is an average of the probabilities that the test sample belongs to each of the seven PES classes, equal to 0.9007 in the deferred test sub-sample, which confirms the superiority of the proposed method of quality metrics over existing models.

The scientific-theoretical foundations of phraseology formation in radio communication for pilots and controllers, as well as standard phrases for engineering and technical personnel and flight attendants were proposed. This subsequently allowed proposing a new methodology to reduce the HF impact on flight safety on the basis of PES recognition by speech signal for aviation personnel by seven archetypal classes, including the following elements: rules of formation of fixed phraseological units and phrases, integral methodology and algorithm, intelligent system with the issuance of recommendations to experts.

An intelligent system has been developed that allows for additional noise-resistant recognition of speech signals of complex form, building up an information feature database, the ability to provide various quantitative and qualitative (linguistic) characteristics, intelligent scheduling etc.

1 HUMAN FACTOR IN ENSURING FLIGHT SAFETY

1.1 Psychological safety of work

Air transport is one of the fastest growing sectors of the world economy, and the flight safety support is one of the key elements of successful operation [3, 4].

Both a transport plane and an atomic power plant (APP) are man-made weapons, and persons with mental, physiological and physical defects cannot be pilots and operators of an atomic power plant, since a mistake can lead to a great catastrophe.

At the same time, the relationship between professional activity and the state of the operating staff has certain features [5]. In particular, people with a strong nervous system, resistant to psychophysical stress are able to be calmer and more productive in extreme conditions than people with a weak nervous system. At the same time, people with a weak nervous system act more reliably and safer in a minimal and optimal mode than people of a strong type, due to higher discretion, increased sensitivity and diligence.

According to D. Klebelsberg [6], people with the following biographical data are predisposed to car accidents:

- origin from a family where there are frequent conflicts;
- abnormal behavior at school;
- difficulties in production activities (conflicts with bosses, team);
- problems in the family (an unhappy marriage, domestic difficulties and others).

As follows from, the problems of accidents and injuries at work cannot be solved only by engineering methods. A common cause of injury is not hazardous working conditions, but dangerous actions of operator. In particular, as follows from, accidents are often based not on engineering and design defects, but on organizational and psychological reasons: poor attitude development of discipline in the workplace, inadequate focus on occupational health and safety compliance, access to hazardous types of work for persons with an increased risk of injuries, the presence of people in a state of lassitude or a specific mental condition.

Based on the foregoing, and analyzing the reasons for the increase in injuries due to the influence of the human factor, we can conclude that not only accidents with the loss of large material values, but also accidents associated with injuries and deaths occur from the operator's erroneous actions [7]. It also follows that the development of technology is ahead of psychological measures to protect against its dangerous and harmful effects, and, therefore, using modern info communication technologies, it is possible to significantly reduce the influence of the human factor on industrial accidents.

Considering the reasons for the growth of the rates of injuries in connection with human factors, we can conclude that the development of technology moves ahead of psychological measures for protection against its dangerous and harmful effects.

Currently, a new scientific direction, that investigates the problems of the influence of psychological and psychophysical aspects of human activity on the efficiency of the labor of personnel and the prevention of their erroneous operations with the application of technical means in the work environment, has been formed.

Based on the analysis of a number of sources [8], the new scientific direction can be called psychological safety on production plant with the following areas of research:

– educational psychology in profession and safety techniques;

– psychology of attitude development of professional caution;

– psychology of management in occupational health and safety.

The applied aspects of this scientific direction make it possible to solve the organizational problems in production using the following operations:

– to carry out professional selection according to medical, medico-psychological and psychophysiological indicators;

– to develop optimal work regimes by time periods, by the nature of the activity and the features of the technology;

– to evaluate the technological reliability of specialists;

– to control the mental state of personnel by a manager, a medical worker and a technical specialist;

– generalization of experience in the analysis of accidents and psychological stability.

Work placement and attitude development, including instructions, training and control measures, are tools for increasing the psychological safety of workplace.

Consequently, the safety psychology studies the application of psychological knowledge to ensure the safety of human activity, considering mental processes, properties and various forms of mental states observed in the process of work.

In the psychological activity of personnel, there are three main groups of components: thinking processes, properties and states.

Thinking processes determine the qualities of professionalism characteristic of a person: cognitive, emotional, volitional, emotions, sensations, perceptions, memory and thinking. Consequently, they form the basis of psychological activity and are a dynamic reflection of reality. Without them, the formation of knowledge and the acquisition of life experience are impossible.

The properties of the human’s psychic setup are the qualities of the personality or its essential features: intellectual, emotional, volitional, labor, moral, etc.

A psychological state is a relatively stable structural organization of all components of the mental health, which performs the function of active interaction of a person with the external environment and is characterized by two states:

1. Industrial - stress and fatigue.

2. Specific - alcoholic, medicinal, transcendental stress, paroxysmal, asthenia, severe fatigue.

There are certain signs of behavior that allow you to assess the psychological state. For example, with severe fatigue, vitality is lost, coordination of actions decreases, and stiffness and slowness appear.

With the out-of-limit distress, tremors of hands, voices, verboseness, hyperactivity are observed.

Among the special psychological conditions that are important for the mental reliability of the operator, it is necessary to highlight paroxysmal attention disorders, psychogenic disorders of consciousness, affective states and conditions associated with the use of mentally active drugs (stimulants, tranquilizers, alcohol) [9].

Psychological conditions arising in the process of working practice are classified into the following main groups:

1. Relatively stable and long-term, arising from reasons of satisfaction or dissatisfaction with work, interest, indifference, etc.
2. Temporary, situational, fast-moving, arising under the influence of various kinds of malfunctions in the production process or in the relationship of workers.
3. Constant, arising periodically for the reasons of boredom, somnolence, apathy, production, lassitude, etc. in the process of working practice.

On the basis of the predominance of one of the sides of the psyche, the following production psychological states are distinguished as:

- emotional volitional, accompanied by volitional effort;
- perception and sensation;
- change in attention, leading to distraction and decreased focus;
- change in mental activity.

As follows from, the most significant sign from the point of psychological safety is the state of distress, which is divided into the following groups:

1. Moderate distress - normal operating condition optimally.
2. Increased distress - accompanies activities in extreme conditions.

In this case, the factors causing increased distress are also divided into two subgroups:

- the first, includes physiological discomfort, lack of time, increased significance of erroneous actions, conflict conditions;
- The second includes biological fear, increased difficulty of the task, information gap, information overload, failure.

The increased distress is divided into the following types:

- monotony;
- physical;
- expectations;
- fatigue;
- sensory;
- polytony;
- emotional;
- motivational.

As follows from [9], lassitude is one of the most common factors that have a significant impact on the safety of activities. Lassitude components can have the following consequences as:

- inability to concentrate;
- weakness;

- motor disorders;
- weakening of will;
- sensity disorders;
- somnolence;
- retention defects;
- thinking defects.

Most importantly, severe forms of lassitude create injury and are a common cause of accidents.

The evaluation criterion of the activity of the operating personnel is determined by the reliability of work, as the need to successfully complete the work or the assigned task at a given stage of the system's functioning within a given time interval with certain requirements for the duration of the conducting operations [10].

Human error is defined as failure to complete the assigned task or perform a prohibited action that could cause damage to equipment or property, or disrupt the normal course of planned operations [11].

Errors arising on production plant due to the influence of psychological factors are divided into the following groups:

- design planning - due to the imperfection of technological processes and equipment for the creation of design solutions;
- operating personnel - occur when the employee does not properly follow the established procedures;
- manufacturing - formed at the production stage due to: unsatisfactory quality of work, incorrect choice of material and manufacture of a product with deviations from design documentation;
- maintenance - occur during operation due to poor-quality repair and installation, unsatisfactory equipping with tools and devices;
- control - associated with erroneous acceptance, as a suitable element or device, the characteristics of which are outside the tolerances;
- handling - occurs due to unsatisfactory storage of products or their transportation;
- work place arrangement - due to the inconvenience of the workplace, cramped living space of the premise, increased noise and temperature, insufficient illumination.

Human errors can be divided into three levels:

- 1-level, in which it is possible to eliminate errors;
- 2- level – it is possible of consequence prevention;
- 3- level is elimination of repetitive occurrence of errors.

In general, the high quality of the personnel activity is determined by two characteristics: speed and reliability.

The performance criterion is the time of solving the problem - the time interval from the moment of response to the information received until the end of the control actions, mathematically determined by the formula (1.1) [12]:

$$t + b * h = a + h/v \tag{1.1}$$

where *a* - latent reaction time, usually $a = 0{,}2 \dots 0{,}6\ sec$;

b - processing time of a unit of information, $b = 0{,}15 \dots 0{,}35\ sec$;

h - amount of processed information;

v - the average speed of information processing, $v = 2 \dots 4\ ms$.

The reliability of personnel determines its ability to perform the functions assigned to it in a full range under the certain working conditions.

Reliability is determined by characteristics:

- readiness;
- faultlessness;
- recoverability;
- promptness;
- accuracy.

The availability factor K is determined by the formula (1.2):

$$K = 1 - T_b/T \tag{1.2}$$

where T_b - time during which the staff cannot receive information;

T – total working time.

However, complex manufacturing processes are characterized by additional characteristics:

– faultlessness is assessed by the probability of error-free operation at the level of an individual operation or in a full range;

– recoverability of personnel is the likelihood of correcting an error made by them;

– the promptness of personnel actions is characterized by the likelihood of completing the task within a given time;

– accuracy - the degree of deviation of a quantitative parameter of the system measured by a person from its true, specified or nominal value.

Quantitatively, the accuracy is estimated by the error:

$$A = A_I - A_F \tag{1.3}$$

where A_I – true, nominal value of the parameter;

A_F - Actual, measured or controlled by a person value of parameter.

As a quantitative parameter, accuracy depends on the following characteristics:

- quantitative values of information;
- complexity of the task;
- condition and speed of performance of work;
- functional and psychological states of personnel;
- professional qualifications;
- tediousness and the other factors.

Stimulating the psychological safety of activities in production is also an additional characteristic that affects the mental and emotional state of personnel, since subsequent behavior depends on their consequences and is built on the principle of reward and punishment.

Stimulation is assessed by the coefficient of performing discipline, determined by the formula (1.4):

$$K_{ID} = M_V - M_P \tag{1.4}$$

where M_V – the number of completed activities;

M_P - planned activities.

Encouragement for behavior stimulates staff to increase responsibility, punishment - to decrease and gives a negative result:

- hiding shortcomings in work and in psychology;
- unwillingness to improve the process;
- a sense of punishment among all colleagues.

Stimulating the activities of structural divisions and functional services to ensure a high level of occupational health and safety

To assess the observance of labor protection norms, rules and instructions by employees, the coefficient of the level of compliance with labor protection rules and norms is used, determined by the formula (1.5):

$$K_{SP} = P_C - P_O \tag{1.5}$$

where P_C – part of the staff who comply with the rules;

P_O – Headcount.

In general, for a quantitative assessment of the state of safety of technical and technological equipment in production, a safety factor is used, and mathematically defined as

$$K_B = T_B / T_O \tag{1.6}$$

where T_B – the amount of equipment that meets safety rules and regulations;

T_O – Total number of technical and technological equipment.

An integral assessment of the psychological safety of labor in a production environment can be estimated by the generalized coefficient of the level of occupational health and safety:

$$K_{ING} = (K_{ID} + K_{SP} + K_B)/3 \tag{1.7}$$

The value of the incentive depends on this integral assessment, and most importantly, productivity and occupational health and safety can significantly increase with a decrease in the number of injuries and accidents.

1.2 The human factor in civil aviation

At present, aviation in the Republic of Kazakhstan, based on the field of application, can be divided into two groups: state and civil (experimental).

Civil aviation [13] is an air transport industry that provides a successful solution to a wide range of tasks:

1) transportation of passengers, mail and cargo;

2) protecting agricultural plants from pests;

3) aerial photography of the area;

4) mineral exploration;

5) protection of forest tracts (including extinguishing fires);

6) For medical and sanitary purposes.

Civil aviation is a national strategic pillar of the economy of any state, since it has a fleet of aircraft (airplanes, helicopters and unmanned systems), a network of airlines, airports, airfields with a system of engineering structure, radio and meteorological stations, factories and repair and technical bases, scientific research and educational institutions [13].

Each civil aviation enterprise has its own production cycle, which includes a set of specific technological operations: from aircraft control to service and maintenance.

As on any production plants, accidents, occurrences and disasters are possible in civil aviation, but the consequences from them are of a special nature, since the consequences are accompanied by large human toll and damage of property.

In accordance with the Law of the Republic of Kazakhstan dated 15.07.2010 No339-IV ЗРК "On the use of the airspace of the Republic of Kazakhstan and aviation activities", Chapter 12. Aviation accidents and incidents and their investigation (hereinafter the quotation): *"Article 92":*

– an aviation accident in state aviation is recognized as an event related to the flight operation of an aircraft, which led to the fatality (physical damage resulting in death) of people on board the aircraft and (or) the loss of this aircraft;

– an aviation accident in civil (experimental) aviation is recognized as an aviation event associated with the use of a civil aircraft with the intention to fly, which, in the case of a manned aircraft, occurs from the moment when a person stepped on board with the intent to fly until the moment when all persons on board for the purpose of the flight have left the aircraft; or, in the case of an unmanned aircraft, occurs from the moment the aircraft is ready to initiate movement for the purpose of flight until it stops at the end of the flight and the main power plant is turned off, and during which:

1) any person suffers a physical damage resulting in death (including cases where in connection with the inflicted bodily injury, death occurred within thirty calendar days from the date of the accident) as a result of being on the aircraft, except in cases where the bodily injury is due to natural causes, self-inflicted or inflicted by others, or when injuries are caused by a stowaway hiding outside areas normally accessible to passengers and crew;

2) the aircraft is damaged or its structure is destroyed, as a result of which the strength of the structure is impaired, the technical or flight characteristics of the aircraft deteriorate, major repairs or replacement of the damaged element are required, except in the following cases:

- failure or engine damage when only one engine is out of order and its hoods or auxiliary units are damaged;

- damages only for aircraft propellers, main rotor blades, tail rotor blades, non-power elements of the airframe, fairings, wingtips, antennas, sensors, blades, pneumatics, brake gears, windshields, wheels or, when the landing gear, landing gear panels are slightly damaged, or when the skin has small dents or holes, including minor damage caused by hail or bird strikes (including holes in the radar dome);

- damage to other elements that do not violate the overall strength of the structure;

- damage to the elements of the main and tail rotor, the hub of the main or tail rotor, transmission, damage to the fan unit or gearbox, if these cases did not lead to damage or destruction of the power elements of the fuselage (beams), damage to the skin of the fuselage (beams) without damaging the power elements;

3) aircraft lost or it is in a place where access is absolutely impossible.

– an aviation incident in state aviation is recognized as an event related to the flight operation of an aircraft that could create or created a threat to the integrity of the aircraft and (or) the lives of people on board, but did not end in an aircraft accident;

– an aviation incident in civil (experimental) aviation is recognized as an event related to the use of a civil aircraft, which occurs from the moment when a person boarded with the intent to fly until the moment when all persons on board for the purpose of flying left the aircraft, and due to deviations from the normal functioning of the aircraft, crew, control and flight support services, the impact of the external environment, which may have an impact on flight safety, but not ending in an aviation accident".

Conventionally, the causes of accidents and incidents can be divided into the following groups:

– technical, for example, the destruction of individual aircraft structures, engine failure, malfunction of control systems, power supply, communication, piloting, lack of fuel, interruptions in the life support of the crew and passengers, etc;

– caused by the influence of the human factor - errors in the organization of the operation of air transport, in the management of aircraft and flight personnel, errors of air traffic controllers, engineering and technical personnel and other employees;

– hostilities, including exercises and terrorism;

– severe weather conditions;

– the mistake of the military, in particular the air defense troops.

Namely, after any plane crash or incident, a special group is created and an investigation is carried out in order to immediately answer two questions: "How and why does a group of well-intentioned, highly motivated and competent professionals in all respects commit exactly the combination of mistakes and safety violations necessary for an accident to occur?" and "Could something like this happen again?" Organizational and managerial factors also often contribute to normal, healthy, experienced, skilled, motivated and well-equipped personnel committing human errors.

Although information on the results of investigations is closed, but according to information from the media in the Republic of Kazakhstan, statistical information for the period 1991-2013 in aviation is as follows [14]:

– 33 aviation accidents occurred, including 20 with aircraft of civil aviation (including the Ministry of Emergency Situations), 2 accidents with agricultural aircraft, and 4 are with training and private aircraft;

– 169 people died and 178 were injured;

– 4 aviation accidents occurred due to failures in AT (16.7%), 20 aviation accidents were caused by the human factor (HF), i.e. about 83.3%.

The aviation industry is effectively introducing high technologies of microelectronics and artificial intelligence, which makes it possible to reduce the number of accidents and incidents that arise for technical reasons.

However, as follows from the world practice [14], where it is shown that three out of four aircraft accidents and incidents occur as a result of operational errors by human factor, committed by the apparently healthy and properly certified people. In the pursuit of new technologies, people, who interact with each other and use these technologies and who are peculiar to make mistakes, are often forgotten.

Disruptions in human performance are cited as the cause of most aviation accidents [15].

The term “human factor” must be clearly defined, since these words, when used in everyday life, usually cover all aspects of the human activities [15, p. 10].

A person is the most flexible, adaptable and important element of the aviation system, but at the same time he is the most vulnerable to negative influences on his activities.

The science of the human factor is the science of humans, the environments in which they work and live, and their interactions with machines, equipment and procedures. It is also important to understand that it is also involved in the relationship of people with each other and within teams.

The term “human error” hides underlying causal factors that need to be brought into the foreground in order to prevent aviation accidents [16].

The price, both in relation to human life and from a material point of view, paid for less-than-optimal human activity, has now increased so much that an amateurish or intuitive approach to the human factor is no longer acceptable.

Based on the analysis of a number of scientific sources [15, p. 11] and the legislation of ICAO and EASA, it is possible to distinguish three strategies for approaching the aspects of the impact of HF in aviation, reflected in the document Doc 9758, based on the Euro control document “Human Factors Module–A Business Case or Human Factors Investment” [15, p. 11-13]:

1. “No actions” approach: no initiatives are taken to prevent HF-related problems; they are solved only after they have arisen.

2. The “retroactive” approach solving HF-related problems is postponed to the last stages of the system development process.

3. “Proactive” approach: the problems of the human factor are solved before they occur.

In most cases, human error is cited as a factor in causing or contributing to an aviation accident. Often human errors are committed by qualified employees, although it is obvious that they did not plan any incident. Errors are not some kinds of abnormal behavior; they are naturally occurring spin-off effects of almost all human efforts. An error must be perceived as a normal component of any system in which humans and technology interact.

Errors can be caused by:

– improper maintenance of equipment;

– insufficient professional training (theory, internship, practice);

– imperfect regulatory legal acts (rules, guidelines, instructions, technologies, etc.).

In this case, human errors can be further subdivided into two groups:

- the first is based on norms that are wrong or inappropriate for a given task;

- the second is based on knowledge, i.e. the correct method has not been chosen for tasks for which there are no predefined norms.

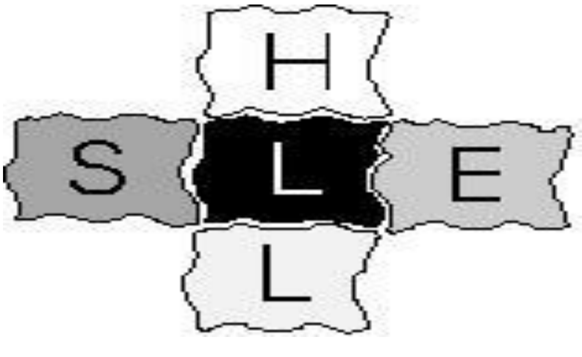
Errors and mistakes are in random character.

Violations are intentional [17], i.e., a person has an intention to commit actions that deviate from the rules with undesirable consequences. As well as errors, violations are divided into two groups:

- routine - to get the job done with minimal effort or to satisfy aggressive instincts;

- necessary - noncompliance with forms simply to get the job done in the absence of adequate tools, equipment or procedures.

The influence of the HF is very clearly reflected in the SHELL model shown in figure 1.10. The model includes the following segments: Software - software installations, Hardware - object, Environment - environmental conditions, and Liveware - subject.



S - installations, H – object, E - environment, L – subject

Figure 1.1 – The model of SHELL

In the research, the following interpretation of the segments is proposed: the subject is a person, hardware is a machine, software installations are procedures, symbol systems, etc., and environment is the conditions in which a system consisting of elements L, H, S should function.

The block diagram of the model does not cover interactions between building blocks that are not related to human factors and is only intended to facilitate understanding of the role of human factors [18].

In this model, coinciding or non-coinciding segment boundaries - interfaces are as informative as the characteristics of the blocks themselves. The discrepancy between the boundaries can be a source of error.

The subject, in this case the person, is in the center of the model. He is considered the most critical and most flexible building block of the system. People are characterized by significant differences in their performance features and many limitations, most of which can be currently broadly foreseen. The boundaries of this block are jagged, so in order to avoid stressful situations and the final destruction of the system, they must be precisely mated with the boundaries of other segments. Understanding the characteristics of this central segment is very important to achieve this coupling. Examples of such important characteristics are the following:

1. Physical size and shape, as it is vital to consider body size and kinematics of movement, which can vary based on factors such as age, ethnicity and gender when designing the workplace and equipment.

2. Physiological needs because human needs are related to the intake of food, water and oxygen.

3. Characteristics of information perception because a person has various senses that allow him to react to events and perform the required tasks, but they are subject to degradation, scientific information about which can be gleaned from the field of psychology and physiology.

4. Information processing since the functions performed by a person are limited, while poor design of the device or warning system is very often the result of the fact that the design did not take into account the capabilities and limitations of the person in relation to information processing, i.e., factors such as nervous tension, motivation, short-term and long-term memory.

5. Characteristics of the reaction to input information since the received and processed information forms decisions and (or) the reaction signal is transmitted to the muscles in the form of physical control movements at the beginning of communication, and for which knowledge of biomechanics, physiology and psychology is required about acceptable control efforts and an acceptable direction of movement.

6. The range of permissible environmental conditions, such as temperature, vibration, pressure, humidity, noise, time of day, lighting and overload components, can adversely affect the performance and well-being of a person, information on which is contained in medicine, psychology, physiology and biology.

The subject represents the centerpiece of the SHELL model of the human factor. The rest of the elements must adapt to it and fit into this centerpiece. Let’s consider the combinations of segments of the model:

1. The interaction “Subject – object”, in relation to the influence of HF in the system “man – machine”, is considered as, for example:

– designing chairs that match the characteristics of the human body; displays corresponding to the user's ability to assimilate information;

– design of controls with the correct choice of direction of movement, marking and placement.

Due to the natural ability of a person to adapt to system defects, which masks their impact, the user may not be aware of the defects in the interconnection of the L - H elements, even if they ultimately lead to a catastrophe. However, defects continue to exist and can pose a potential hazard.

2. The relationship “Subject - software installations” deals with the interaction between humans and non-physical components of the system, such as rules, guidelines, checklists of operations, symbolics and computer software. Problems with such a relationship may be less visible than with the relationship “subject-object”, and therefore more difficult to detect and resolve (for example, an incorrect interpretation of a checklist of operations or symbolic designation).

3. The relationship “Subject – environment” is recognized in aviation as one of the first, because now new problems have arisen: the concentration of ozone and high levels of radiation during flights at high altitudes, as well as problems associated with disruption of biological rhythms and sleep due to fast intercontinental flights. As the causes of many accidents are related to inadequate perception of the situation and loss of orientation, when considering the relationship “subject-environment”, it is necessary to pay attention to the errors of perception associated, for example, with environmental conditions (for example, the effects of optical illusion during approach and landing). The aviation system operates within the framework of broad organizational, managerial, political and economic constraints. These elements of the environment interact with a person through devices for his interface with the environmental conditions. And, although the adjustment of the influence of these factors are usually beyond specialists’ compass on the human factor, they should be considered and evaluated by responsible executives, who have such opportunities.

4. The relationship “Subject – subject” is considered as a type of interaction between people. Flight crew training and proficiency testing has traditionally been done on an individual basis. It was believed that if each member of the crew had a professional training, then the entire crew was professionally fit and efficiently copes with the duties. However, this is not always the case, and over the years more and more attention is focused on the violations of teamwork in the crew. Flight crews act as groups, so the relationships within the group affect the behavior and activities of its members. Associated with this type of interaction there are concepts such as leadership, the interaction of crew members, its well-coordinated work and interpersonal relationships. ICAO Handbook No. 2 on HF describes the currently accepted in the aviation industry approach to resolving problems associated with this

type of interaction: CRM is a crew resource management and LOFT is a line-oriented flight training. The relationship between administration and personnel is also considered to be the type discussed here, because the corporate climate and the degree of exploitation of people in a company can significantly affect their work. In addition, the Handbook No. 2 demonstrates the important role of administration in accident prevention.

“Human factor” as a term requires a clear definition because when it is used in everyday life, it often covers all aspects of human activity. The human being is the most flexible, adaptable and important element of the aviation system, but also the most vulnerable from the point of view of the possibility of a negative impact on his activities. For many years, every three out of four accidents have occurred as a result of a malfunction in human performance. These failures are usually classified as “human error”.

The term “human error” does not play a positive role from the point of view of prevention of aviation accidents, as it most often can only determine where a failure occurred in the system, but not establish why it occurred. An error connected with human activity in the system can be predetermined at the design stage of the system or provoked by insufficient personnel training, poorly developed procedures, imperfect concept and format of the current checklists or manuals. In addition, the definition of the term “human error” does not take into account some hidden factors that must be carefully analyzed in order to prevent accidents.

1.3 Human errors in aircraft maintenance

Today, human errors, rather than technical failures, represent the greatest potential threat in aviation safety. The commercial airline Boeing, after analyzing 220 documented aircraft accidents, found that the three most common causes of accidents are:

– non-compliance of established procedures by flight crews (70 out of 220);

– errors in maintenance and inspection (34 out of 220);

– structural defects (33 of 220).

Currently, aviation is conducting scientific research on the construction of safety systems that would prevent the appearance of not only physical defects on the aircraft, but also human-centered errors. In the scientific literature, there is a natural increase in the importance of accounting for and preventing human error for aviation technology. For example, in the 1960s, when this problem first began to attract serious attention, the “contribution” of human errors to the set of reasons that cause aviation accidents was estimated at about 20%. In 90s, this indicator increased four-fold, reaching 80%. There are many reasons for this dramatic growth, but only three of them are related to aviation technology:

– over the past thirty years, the reliability of mechanical and electronic components has increased markedly, but the level of intelligence of the operating personnel has remained the same;

– air transport has become more automated and complex, for example, modern aircraft of the Boeing-747-400 and Airbus-A340 types have up to three redundant flight control systems, which reduce the workload on the flight crew, but increase the requirements for engineering and technical specialists in the field of mechanics, electronics and computing, which is not ensured by the correct interaction between the elements (L - H) and (L - S) of the SHELL model;

– the increased complexity of aviation technology creates the potential for accidents due to the redistribution of errors from one category of service personnel to another.

As follows from, one of the reasons for well-known aviation accidents was human error during maintenance. For example, the crash of DC-10 airplane of the American Airlines in Chicago in 1979 resulted from a disruption in engine replacement technology. The pylon and engine were dismantled and reassembled rather than individually, causing it to fly off the wing.

In a careful analysis of the major accidents that have occurred worldwide, the following list of the leading causes of accidents in percentage terms is presented in table 1.1.

Table 1.1 – List of the main causes of aviation accidents

The reason of accident	Percent
Pilot Violation of Standard Operating Procedure	33
Insufficient cross control from the co-pilot	26
Structural defects	13
Disadvantages of maintenance	12
Lack of guidance on approach	10
Ignoring by the aircraft commander of the messages of the crew members	10
Error / failure of the air traffic control service	9

Incorrect component installation, inattentive inspection and the quality control are the most frequently recurring maintenance errors. Let's look at some examples.

In May 1983 the plane “Lockheed L-1011” of the “Eastern Airlines”, carrying out flight 855, took off from Miami International Airport in Nassau in the Bahamas. Shortly after takeoff, the warning lamp of drop in pressure in engine No. 2 burst into flame. As a precaution, the crew turned off the engine and the pilot decided to return to Miami. Shortly thereafter, the indicators of both remaining engines showed zero oil pressure, and they failed. Attempts were made to start all three engines. At a distance of 22 miles from Miami, after descending to an altitude of 4000 feet, the crew was able to start engine # 2 and land on one engine running, which was fuming heavily. It was found that all three main sensor - detector chips were installed without gasket rings [19].

In June 1990, BAC 1-11 (“British Airways”, Flight 5390) departed from Birmingham International Airport to Malaga, Spain with 81 passengers, four flight attendants and two flight crew members. The takeoff was performed by the co-pilot, and after the transition to a steady climb, the pilot-in-command took control of the aircraft in accordance with the airline's rules. At this point, both pilots released the shoulder safety belt, and the captain also released the slip safety harness. While climbing 17,300 feet, a sharp sonic boom was heard and a thick fog enveloped the fuselage; that is a sign of rapid depressurization. The windshield in the cockpit flew out, and the commander was pulled into the windshield opening, where he was stuck. The door to the cockpit abruptly opened inward and hit the control panel and control of radio-technical and navigation equipment. The co-pilot immediately took control of the aircraft again and began an emergency descent to level 110. The flight attendants tried to pull the commander back into the cockpit, but the suction stream did not allow them to do this. They held him in this position by his knees until the plane landed. As a result of the investigation, it was established that the cause of the flight accident was the fact that during the replacement the windshield was fixed with the wrong bolts.

In September 1991, the plane "Embraer 120" of the "Continental Express" Airlines, on Flight 2574, departed from Laredo International Airport, Texas, to Houston International Airport. The plane suddenly collapsed in flight and run into an accident, killing all 13 people on board. During the investigation, it was determined that the incident occurred due to the fact that the fastening screws on the upper surface of the left side of the front edge of the horizontal stabilizer were unscrewed and not replaced, as a result of which the leading-edge de-icing assembly was secured to the stabilizer with only the lower fastening screws.

Due to the specific nature of human error in the maintenance environment, it manifests itself in a different form from that in which it occurs in other work environments, for example, in the flight deck or in the air traffic controllers' room. If the wrong button is pressed or the handle is pulled out of the wrong lever, or the wrong command is sent, the pilot or air traffic controller will see the consequences of their mistake before the aircraft finishes its flight. If an aircraft accident or incident occurs, the pilot is always "on stage" while it occurs. If the aviation accident is connected with the work of air traffic controller, who monitors the air traffic, then ATC is almost always "on stage" or following the event in real time. Although this important feature seems quite natural for errors of the flight crew or ATC, it is not always typical for the errors committed in the maintenance of the aircraft [20].

In contrast to the "real-time" nature of errors of an air traffic management and flight deck, a maintenance error is very often not apparent during its occurrence. In some cases, the aircraft maintenance technician will never know about the error, because its identification can occur after a few days, months or several years. In the event of an engine disk failure on a "Su-21 City" Airlines of DC-10 in 1989, the alleged error in the aircraft inspection was made seventeen months prior to the accident.

This usually happens when the system malfunctions is a human error made during maintenance, we often only know about the malfunction of the aircraft, to which it led. But we very rarely know why it happened. In the field of aircraft maintenance, there are no analogues to the cockpit conversation recorder, flight data recorder or ATC tape recorder, that is, there are no devices that record in detail the process of performing maintenance operations.

Flight accidents with aircraft BAC 1-11 and “Embraer 120”, which took place for reasons connected with errors of maintenance and inspections are exceptions in the sense that they occurred soon after the active faults. This allowed the investigators to focus their efforts on the site and on the actions of individuals, as well as on the activities of the organization. The classical factor of “remoteness in time and space” not only did not hinder the investigation of these cases, but also did not slow down their implementation. This identified organizational errors, individual errors and methods of organization of work, contributing to make mistakes, which made it possible to focus on the root causes of practice leading to aviation accidents [20].

Statistics show that organizational or systematic errors in aircraft maintenance organizations are not limited to one organization or one region. From the results of the analysis of three flight accidents carried out here, it is clear that the behavior of organizations and their individual employees before the events described was the same. For example:

– maintenance staff and inspectors violated established methods and procedures (active failure);

– the persons responsible for ensuring compliance with the established procedures and methods did not check not only "single violations", but also, symptomatically, incorrect actions performed over a long period of time (active and latent failures);

– upper executive management, responsible for maintenance did not take the necessary steps to unconditionally follow the procedures prescribed in their respective organizations (latent failures);

– maintenance operations were performed by persons not designated for these duties, who, with the best of intentions, began work on their own initiative (active failure, facilitated by the two previously discussed latent failures);

– the lack of complete and / or properly transmitted information is obvious, which increases the chain of errors leading to aviation accidents (latent failure).

It follows from the material in this subsection, aviation personnel, making decisions, including responsible executives, corporate or regulatory authorities are responsible for establishing objectives and management of all available resources to achieve two clearly defined goals for air transport:

– accidents prevention;

– timely and economically rational transportation of passengers and goods.

At the same time, new scientific approaches and technologies should be created to reduce the number of human errors, i.e., to reduce the impact of HF on flight safety.

1.4 Analysis of flight accidents and incidents

The statistical analysis of accidents and incidents in aviation is open information. Currently, there are several organizations, whose main activity is to collect, analyze and disseminate information about the incidents, but the information service “Aviation Safety Network”, which has its own website, is more open.

After conducting analytical research on the database [21] of registered aircraft accidents and air crashes that resulted in human casualties, the following conclusion can be drawn:

The dynamics are as follows. From 1996 to 2004, the number of air crashes decreased from 52 accidents (1104 fatalities) to 28 (431 victims). It increased further, but since 2005 there has been a steady decline. In particular, in 2009, by August 12, 33 accidents were registered, during the same time period in 2008 - 38 such crashes, also in 2007 - 30 and in 2006 - 35 until August 12.

According to the FORINSURER information service, the number of fatal accidents from the beginning of 2000 to 2020 was 106 cases and 9,962 deaths [14].

The statistics of the largest plane crashes in the world for 2000-2021 show that the main cause of tragedies in the air is the human factor (crew or dispatcher’s error).

The following information provided by the US Department of Transportation [22] is quite interesting: air transport is less dangerous than road transport, because the risk of dying while flying on an airliner is 1 to 52.6 million, and when driving a car is 1 to 7.6 million.

As follows from the materials [23] published by Boeing company, the distribution of flight safety is as follows: at the time of landing is 45% of air crashes, in flight it is 6%, while preparing for take-off and landing of passengers is 5%.

The Interstate Aviation Committee published a report [24] on the state of flight safety, where it is noted that in 2015 the relative indicator of aviation accidents and high-incidence in civil aviation is the worst since 2011. Moreover, most of the accidents occurred on the territory of Russia. For example, in the Russian Federation in 2015 there were 41 aviation accidents (AA), in which 60 people died, and in Kazakhstan there are 4 aviation accidents with seven fatalities. At the same time, there were no disasters in the field of passenger transportation on heavy aircraft in 2015.

However, in 2015, there was one accident with 224 dead, which occurred with the A-321 aircraft over the Sinai Peninsula, as a result of an act of unlawful interference [25].

As follows from the analysis of the above sources [26], in general, the problem of HF in aviation arose for many reasons, including due to the discrepancy between the capabilities of aviation personnel and professional requirements and competencies. At the same time, as the functional systems of aircraft became more complex and the requirements for air transportation increased, the psychological stress on aviation personnel - pilots, air traffic controllers, engineering and technical personnel and flight attendants, increased. Their fatigue and overloading are the causes of aviation accidents.

Recently, the pilots’ behavior, which is the result of a negative emotional state, has become the cause of aircraft crashes. The crash in the Alps of the aircraft “Airbus-A320” Germanwings [27] in March 2015 with 250 people on board was for this reason.

Similar situations periodically arise in civil aviation. In particular, due to such a state of pilots, there were aircraft crashes in 1997 in Indonesia of “Silk Air” airlines with 104 people on board and in October 1999 of an Egyptian Boeing-767 airliner that crashed in the Atlantic Ocean with 217 people on board.

A similar situation occurred during the crash in March 2014 with a Malaysian “Boeing-777-200” plane with 239 people on board, when the co-pilot of the plane was previously treated for depression, and it is highly likely that he was in an ill-fated flight in a negative emotional state.

In 2014, after the disappearance of the Malaysian Boeing-777-200 aircraft, one of the articles [14] emphasized that a person’s negative emotional state is often associated with his anxiety state, which does not arise instantly, but is formed as a result of a sufficiently long-time interval when repeated intake of irritants, assessed as negative.

Six years ago, there was an aircraft crash of the airliner A320-211 of the German company called “Germanwings” [27], following the route Barcelona - Dusseldorf. As a result of the intentions of co-pilot Andreas Lubitz to commit suicide due to psychological problems, the plane crashed into the mountains of the Southern Alps and 150 passengers died.

The role of flight attendants in the safety system is increasing, since human lives depend on their well-organized and coordinated actions during emergency situations. Let's consider some examples. At the Sheremetyevo airport on June 18, 2019, when the plane caught fire, the flight attendant took the wrong action by opening the back door, creating an emergency, which led to the rapid spread of fire and smoke in the cabin. As a result of the aircraft crash 41 people died [25]. At the same time, there are many stories when flight attendants saved many lives in plane crashes [28].

According to IAC statistics [29], about 50% of all aviation accidents in the world are caused by the aircraft crew team, 22% is due to equipment failure, 12% is due to weather conditions, 9% for terrorist threats, 7% is to errors of ground personnel: air traffic controllers and aircraft technicians.

As follows from the research results given in [30], in critical situations, including an aircraft accident, emotionally shocking influences, accompanied by physical and physiological actions, appear in a person:

1) the readiness to respond is increased and at the same time the range of possible options for action is reduced to a minimum;

2) stiffness and angularity in movement;

3) the processes of response to stimuli are accelerated with a decrease in the level of reliability of the developed solutions;

4) the level of selection to negative stimuli ticks upward, increasing sensitivity to them, and, as a result of psycho-emotional selection of incoming information in an anxiety state, current information with a negative assessment, adequate to the person's state, accumulates in the person's memory;

5) the anxiety state can be fixed during the next several hours, free from external stimuli, as well as during sleep, due to feedbacks from the default nervous network;

6) selective and retroactive enhancement of the memory, after 6 hours or more, about the event that occurred in conjunction with the shock

As you can see, a person can be a weak link in the aviation system, and solving the problem of reducing the impact of HF on aviation accidents and incidents by neutralizing or timely removal of an aviation specialist due to unpredictable behavior is an urgent task.

1.5 The impact of the human factor on flight safety

Safety in aviation is the main task of all personnel, for pilots, air traffic controllers, engineering and technical staff and flight attendants, specialists of aviation security at the airport, etc. In particular, the flight outcome depends on the actions of the aircraft crew in emergency situations during the approach phase. Aviation accidents can go from the accident stage to a catastrophic one due to the psychological stress of the aircraft crew and the air traffic controller, who in such situations can make mistakes or make the wrong decision.

A special case in flight during the approach phase can be caused by HF, in particular due to a violation of the aircraft control technology or a violation by the technical service, and a technical factor, for example, due to lack of pressure in the hydraulic system, malfunction of the cleaning switch - landing gear, failure of blocking the chassis control system, etc.

The Human factor plays a very important role in making decisions when a special event occurs in flight for the following reasons:

– determines the correctness and timeliness of decision-making by the controller and the pilot, aviation specialists;

– insufficient research, i.e., there are still many questions, and therefore, the governing documents serve as guidelines;

– analyzing aviation accidents and incidents, ways of dealing with their occurrences are formed;

– in aviation there are many non-standard situations that lead to stressful situations.

Currently, the new technologies (methods, algorithms and tools) are being developed and implemented to reduce the impact of the human factor. For example, in the work [31], algorithms, that reduce the time for thinking about the situation, for thinking about the next step and the sequence of actions, are proposed. But they are not effective enough, because in each case, an accident or incident may develop unpredictably, and each person from the aviation personnel may operate wrong.

The real goal of ongoing research is to develop intelligent technologies (methods, algorithms and tools) for processing and analyzing personnel conditions to reduce the number of accidents and minimize the impact of HF on the aviation security system and flight safety.

Many extraneous factors can affect human behavior during an aircraft accident and incident, that is why the behavior of aviation personnel should be automated and regulated as much as possible, and for this it is necessary to introduce the methodologies for analyzing factors affecting the behavior of aviation personnel, determining the reasons for the occurrence of special cases and forming expert opinions and recommendations for reducing the influence of HF [32].

Since it is known [33] that most aviation accidents are caused by HF of the aviation personnel and are the result of suboptimal human actions, therefore, any improvements in this area can significantly contribute to improving the level of flight safety.

Support for the safe functioning of the aviation industry is a primary task in the formation of scientific and technical tasks for aviation universities and research centers. The main scientific research on flight safety support at the expense of HF is developing in the following areas:

– development of technologies for eliminating the causes associated with the activities of the aircraft crew team (ACT), whose inadequate decision-making accounts for 90% of the causes of aviation accidents and incidents in the world, i.e., with the right actions, the ACT performs the effective flight management;

– creating a methodological framework for preparing the Air Traffic Controller to issue competent recommendations and instructions to the ACT, since a timely and correct prompt can prevent the situation on board from developing to a catastrophic one;

– simulation of situations and special conditions preceding the accidents and incidents to prepare the aircraft crew team and air traffic controllers for correct behavior;

– research for determination of the interaction of factors affecting flight safety.

Let's consider some aspects of the listed scientific directions.

In the case of special flight conditions [34], the traffic controller is required to take additional measures to ensure the safety of air traffic. At the same time, special conditions cannot be predicted, and they do not interfere with the flight, and the aircraft crew and the air traffic controller must make adequate decisions. Currently, hardware and software tools allow you to simulate special conditions, so the aircraft crew team and air traffic controllers are trained on simulators to work out the right solutions.

For the air traffic controller, in turn, it is important to control the situation, to give the full necessary information to the crew of the aircraft and to issue timely recommendations regarding the flight, because in such a situation there is a small-time limit for decision-making and a tense psychophysiological state of the air traffic controller, which is characterized by a high level of incompleteness and uncertainty of information. In such situations, the task of quantifying the possible options for completing the flight is relevant, which allows the aviation operator to choose a strategy of actions with a minimum level of the potential damage. Searching for the effective solution in such conditions requires processing a significant amount of additional information about the aircraft and the external ATC zone [35].

Often, aviation accidents go from an emergency stage to a catastrophic one due to the psychological stress of the aircraft crew team, which, under these conditions, makes mistakes or makes the wrong decision. Therefore, it is very important that in these cases the psycho-emotional state of the aviation personnel was assessed with the removal of untrained specialists, and the most optimal and timely recommendation for the completion of the flight was issued.

Human error is considered to be the cause or one of the main factors in most aviation accidents. At the same time, at first glance, very competent personnel make “critical” mistakes, which are explained by the impossibility of predicting accidents.

Therefore, two reasons for errors can be distinguished:

7) deviation in the behavior of aviation personnel;

8) a natural product of the virtuality of all human efforts.

Errors should be accepted as a normal component of any aviation system, including those in which technicians and humans interact. In this situation, it is important to predict the occurrence of errors in a timely manner.

The effectiveness of the operation of the management systems on flight safety, especially when planning them, is determined by the factors that affect the flight safety, the analysis of their interaction and influence among themselves, and, most importantly, the formation of a methodology for the optimization of processes.

The author of the dissertation, having systematized the existing approaches, proposed a new structure of the security system at aircraft enterprises, the functioning of which can be represented in the form of an algorithm of procedures, which is shown in figure 1.2.

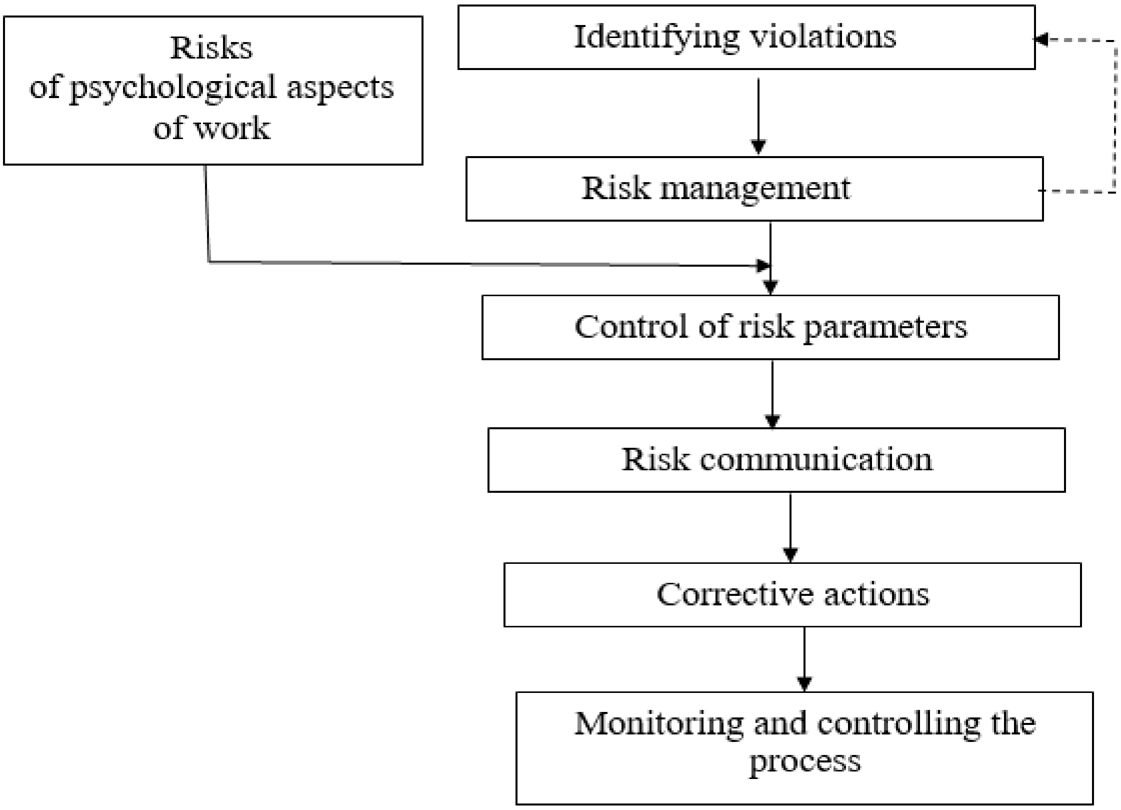

Risks of psychological aspects of work
Identifying violations
Risk management
Control of risk parameters
Risk communication
Corrective actions
Monitoring and controlling the process

Figure 1.2 – Algorithm of procedures "Flight safety systems"

Step 1. Identifying violations in the aviation safety and security manual and other regulatory documents, using various methods and tools in the components, in accordance with the identification and prioritization of:

- aviation violations and incidents of the notification system;
- investigation technology and notification;
- system deviation analysis;
- feedback on pre-flight training;
- analysis of flight data and parameters;
- security system inspection and security monitoring audit;
- control over the sequence of operations;
- state investigation of accidents and serious accidents;
- alerting service and information exchange system.

This procedure is quite complex, as it requires the evaluation and analysis of all available information.

The assessment is based on determining the size of violations in the work, its consistency, uniqueness and features. In this case, you will need to use computer technologies with specialized software to compile a database for storing, filtering and searching information.

Step 2. Risk management – their assessment based on the detection of defects in the security system with decisions on the prediction and elimination of violations, i.e., reducing the impact of the associated risk. The decision should be made individually and taking into account all local conditions. It is important to take into account that the decision does not lead to new violations.

At this stage, the following controlled and positive safety results can be obtained:

9) elimination of the violation completely or, at least so that the associated risk was reduced as far as possible and serious;

10) a set of measures is being developed to eliminate the violation satisfactorily;

11) exclusion of all conditions for the occurrence of problems.

If the results are unsatisfactory, the whole process is repeated again.

Step 3. Control of risk parameters is a management of airlines and enterprises, taking into account the compliance of production goals (such as departure on time, profit) and safety goals. The working space of aviation is filled with unsafe conditions that cannot be avoided, however, operations must continue [36].

Income or loss is an indicator of success in pursuing your productive goals, and safety is a prerequisite for a sustainable aviation business. For companies, safety problems only after an accident or loss will lead to losses/reduced profits, so a company can operate for years, having unsafe conditions without adverse consequences.

In the absence of an effective control system, direct and indirect, quantitative and qualitative parameters that characterize the degree of risk are the key to successful operation. An example is the situation with the assessment of psychological safety at work, when it is possible to solve the issues of financial stability of airlines due to strict control of the parameters of psychological safety of aviation personnel that arise in the course of work, for example, job satisfaction, interest, indifference, somnolence, apathy, etc. Indeed, quality organizations understand that the cost of correcting unsafe conditions is an investment in further long-term profits. Accidents are expensive.

Step 4. Risk communication is the investigation of the impact of possible and existing risks on the processes and economic sustainability of the aircraft enterprises.

Any company operates in the information space; therefore, it must strictly apply modern expert opinions, methods and information processing tools that allow adapting the results of risk analysis for all levels of aviation personnel responsible for safety and decision-making, as well as informing persons who may become victims of aviation violations and incidents.

However, when several risks appear, it is necessary to determine the relationship between the factors of their occurrence and the consequences, which in turn will be subdivided into the following groups:

12) direct - obvious, associated with physical injury, repairs, and property damage, while compensation and replacement for damage to aircraft equipment is covered by insurance payments or savings from special funds to cover risk costs [30, p. 2615];

13) indirect - includes all costs not covered by insurance, are the result of an accident to a greater extent and are accompanied by the following measures:

a) the loss of business or irreparable damage to the organization's reputation;
b) significant losses in the use of equipment with a significant loss of income;
c) loss of staff productivity;
d) investigation and cleaning;
e) payment of insurance and reduction of funds;
f) legal acts and claims for compensation for damages;
g) fines, legal liability and closure of unsafe transactions;

14) the industrial and social impact - along with monetary impact, aviation accidents and incidents can lead to a significant decline in the reputation of the industry and the market of air services in general.

Step 5. Corrective actions is a set of effective measures to eliminate risks; prevent the recurrence of direct or indirect inconsistencies in the production process.

Corrective action should be initiated during audits of the forecast and occurrence of the problem. At the same time, elimination of deficiencies is not a corrective action, but is aimed at correcting errors.

When developing a corrective action plan, it is necessary to take into account additional signs of the emergence of risks: complaints and threats from consumers, expert recommendations, letters of thanks, etc.

Step 6. Monitoring the security process is a part of the management system for ensuring flight safety and aviation security using mechanisms for the prevention and detection of potentially dangerous aviation accidents and incidents, actions and conditions. In aviation science, there are many techniques and technical solutions for optimal and effective monitoring of safety processes [36, p. 59]. Behavioral safety audit (hereinafter BSA) refers to one of the types of such mechanisms and is an interactive process of professional engagement and interaction between the auditor and the employee, whose behavior is being monitored. Information obtained during the conduct of a behavioral safety audit is subject to registration and analysis. As a result, a set of corrective actions is formed: short-term and long-term, solved in working order or with the involvement of higher management, indicating the timing and employees responsible for the correction.

The proposed system is quite universal, but it has a drawback associated with the lack of the possibility of considering the psychological occupational safety. In particular, if the mental and emotional states of aviation personnel are recognized in a timely manner, which are not appropriate for the conditions of high-quality and effective work, then when implementing corrective actions in the form of suspension from production work or aircraft control, it is possible to reduce the impact of HF on aviation accidents and incidents.

1.6 Biometric identification methods

Currently, the term “biometrics” is being introduced into all areas of our life. For the mankind, the definitions for “Biometric identification”, “biometric scanner (BS)”, “biometric passport” is already clear. If earlier they belonged only to the field of biology and mathematical statistics, now they are associated with automatic or automated methods of recognizing a person’s personality by his biological characteristics or manifestations.

The modern biometric system [37] consists of a BS, a physical device for measuring biometric characteristics and processors with algorithms for comparing the measured characteristic with a previously registered standard - a biometric template.

Biometric systems operate in two modes:

1. Verification - comparing one to one to verify the veracity by entering a name, password or pin-code, presenting an identity card or an electronic card, etc.

2. Identification - comparison of one with many to make a decision about whether the user belongs to the number of known individuals based on the measurement of biometric characteristics.

In practice, identification and verification is carried out using three biometric recognition methods: fingerprint, face image (2D photo and 3D photo), and iris [38].

The effectiveness of each method is determined by the following characteristics [38]:

- measurability of biometric characteristics;
- average recognition time;
- environmental resistance;
- counterfeit resistance;
- recognition accuracy.

The above biometric technologies have certain advantages and disadvantages indicated in a number of sources. However, one of the most effective is the biometric characteristic of a person's voice for the following reasons:

- speech signal is individual;

– easily measurable, for example, by the frequency spectrum;

– low cost of devices used for identification, for example, microphones;

– application possibilities for solving a wide range of tasks, especially in systems for differentiating access to physical objects and information resources, in telecommunication channels for management, information protection, forensics and anti-terrorist activities.

Currently, there are scientific approaches to recognizing the biometric characteristics of the voice, and the determination of these characteristics by the speaker's voice, such as gender, age, nationality, dialect, emotional coloring of speech, are also important in the field of criminal investigation technique and anti-terrorist actions [39].

Despite the wide applicability and the advantages listed above, the methods used for identifying a person by voice data have a number of serious disadvantages:

– during identification, all kinds of hardware distortion and interference occur;

– external acoustic noises are inevitably superimposed on the voice signal, which can significantly distort individual informative characteristics.

Voice identification includes a complex of technical, algorithmic and mathematical methods covering all stages, from voice recording to the classification of voice data.

The considered advantages and disadvantages lead to the conclusion that the further development of voice identification systems is promising, relevant and urgently requires the development of new intelligent approaches aimed at processing large arrays of experimental speech signals, their effective analysis and reliable classification.

The author of the dissertation proposes to use speech signals to determine the emotional states of aviation personnel to reduce the impact of the human factor on aviation accidents and incidents. This scientific approach testifies to the relevance of research on the creation of new mathematical methods for processing, analyzing and classifying voice data, ensuring the reliability and veracity of personal identification.

Conclusions on the first section

Based on the conducted analytical studies to assess the impact of the human factor on flight safety, the following scientific results were obtained. It has been established that at enterprises it is important to pay special attention to the psychological safety of labor, since not only accidents with the loss of large material values, but also catastrophes associated with injuries and deaths occur from the erroneous actions of personnel. There are certain signs of behavior that make it possible to assess the psychological state, which are reflected, among other things, on the speech signal.

As in any manufacturing plant, cases of accidents, crashes and disasters are possible in civil aviation, but the consequences from them are of a specific nature, since the consequences are accompanied by large human losses and the damage of property.

The causes of aviation accidents and incidents are divided into two groups: technical and due to the influence of the human factor. At the same time, as follows from world practice, three out of four aviation accidents and incidents occur as a result of the impact of HF for reasons caused by the presence of errors in the organization of the operation of air transport, in the management of aircraft and flight personnel, in the activities of air traffic controllers, engineering and technical personnel and other staff, etc.

Based on the statistical analysis of accidents and incidents in aviation, presented on the information resources of foreign organizations, data on quantitative characteristics of the impact of human factor on flight safety, depending on the professional activities of aviation personnel, have been established.

Based on the systematization of existing scientific approaches to reduce the impact of human factor, a new structure of the safety system at the aviation enterprises, based on the assessment and analysis of the risks in ensuring the flight safety, is proposed. Its feature is the possibility to take into account the risks of psychological safety at work for the recognition of psycho-emotional states of aviation personnel on-time, which are not appropriate for the conditions of high-quality and efficient work. Immediate suspension from manufacturing activities or aircraft control is a factor in reducing the impact of HF on aviation accidents and incidents.

It has been established that the main solution to the problem of reducing the number of accidents and incidents is to create new scientific approaches and intelligent technologies.

To reduce the impact of HF on aviation accidents and incidents, it is proposed to determine the psycho-emotional states of aviation personnel based on the recognition of speech signals, since this characteristic is individual, easily measurable, and the hardware and software implementation has a low cost and applicability for a wide range of tasks.

2 DEVELOPMENT OF AN INTEGRATED APPROACH TO INTELLECTUAL RECOGNITION OF PES BY SPEECH SIGNAL IN FLIGHT SAFETY TASKS

2.1 An approach for recognition of PEC using intelligent information analysis methods in aviation sphere

Over the past decade, automatic speech recognition systems have been actively used to solve a wide range of problems in the field of building modern human-machine interfaces, as well as in the field of security support, where an operational assessment of the situation is required based on incoming voice data. In many ways, this became possible due to the increase in the number and greater availability of high-performance computing systems, and the active development of info communication technologies.

However, the current market conditions and socio-economic trends in society require the solution of new problems. Such spheres of human activity as medicine, marketing, security support, monitoring the state of personnel in hazardous industries or in transport stimulate researchers to search for new effective tools for automatic recognition of the psycho-emotional state of a person by his voice. Automatic recognition of PES is also necessary to raise the implemented interface systems of human-computer communication to a higher level. In addition, when solving this problem, it becomes possible to automatically determine the level of stress and fatigue, recognize depressive states, prevent tediousness, etc.

A significant positive effect from the introduction of automatic recognition systems of PES based on voice can be expected in those industries where communication is mainly supported by speech, without visual contact, and where it is extremely important to reduce possible sources of danger to people and property through a continuous process of identifying and controlling risk factors. These requirements primarily relate to the field of air transportation and are formulated in the manuals on flight safety management [40]. For these reasons, automatic detection of a critical change in PES in the dialogue “Pilot - Air traffic controller”, “Pilot – Crew”, etc. can play a decisive role in the process of identifying risk factors in the event of emergency situations.

The possibility of using automatic means of PES recognition in air transport is also justified by the specifics of building a dialogue. This industry uses a special aviation English language.

As you know, the emergence of a special professional language in aviation is dictated by the requirement to enhance flight safety. According to [41], the language factor is associated with aviation accidents and incidents in the following cases:

– the crew or air traffic controller does not use standard radiotelephony phraseology when performing routine procedures;

– pilots do not speak English at a sufficient level to explain the problem on board;

– the crew or air traffic controller switches from English to their native language while communicating in the same airspace.

The need to prevent the outlined situations made aviation English the only language in the airspace, determined its conciseness, intelligibility, limited words and phraseological units used, as well as low emotional color in the process of radiotelephone communication.

These circumstances can have a positive effect on the effectiveness of automatic recognition systems of PES based on voice for a given applied problem, since in this case it will be necessary to analyze only English speech with a low emotional content, which is rather limited in terms of a set of words, over an overwhelming time interval. At the same time, it can be expected that fragments of emotionally colored speech will be more efficiently distinguished by the applied automatic algorithms.

It should be noted that, despite the importance and relevance of the indicated problem, at the moment there is no general theory that reveals the relationship between the speaker's PES and the characteristics of the voice signal. This circumstance largely determines the approach being formed in the development of methods for automatic classification of PES according to the characteristics of the voice signal: to solve specific applied problems, it is necessary to develop new algorithms based on modern advances in the area of digital signal processing (DSP) and info communication technologies, as well as deep optimization of existing solutions.

An important factor determining the difficulty of achieving the goal of automatic recognition of PES is the vagueness and ambiguousness of the existing formulations of the very concept of emotion, as well as theoretical models of their classification. In this regard, when implementing the automatic recognition system of PES, which implies application in the aviation field, one should single out a set of archetypal psycho-emotional states, which include joy, fear, anger, sadness, disgust, surprise and neutral state (calmness) [42]. In this case, the purpose of automatic classification will be to determine the probability of attributing the speaker's emotional state to each of the seven listed classes. In this case, for the implemented multiclass classifier, it is possible to determine the threshold of the probability function, according to which the choice of the dominant emotion in the speaker's speech is performed.

Taking into account the proven high efficiency of using intelligent analysis methods for automatic recognition of human speech [43], it is also advisable to use well-proven analysis methods based on technologies of computer assisted learning (CAL) and deep neural networks for the problem of detection PES.

The main advantages of using CAL methods are the ability to analyze large amounts of information to find hidden patterns in the data. This allows you to compare some of its implicit characteristics to perform classification with the object of study. Effective work with large arrays of heterogeneous data, using intelligent methods, allows using various informative features that characterize the object under study, and highlight among them those that are really important from the perspective of information gain in the process of analysis. In addition, intelligent CAL algorithms are capable of self-learning, optimizing their parameters for a specific problem being solved.

The use of intelligent methods of analysis makes it possible to effectively automate the tasks associated with the processing of large flows of information, to develop decision support tools for human personnel in the aviation industry, reducing the risk of errors and reduced attention. The use of tools of the theory of artificial intelligence opens up opportunities for solving difficult problems from the perspective of automation, for which a certain limit has been reached at the moment in terms of the quality of their functioning.

Based on the foregoing, this work solves the problem of developing an effective method for automatic recognition of PES of the speaker using intelligent analysis tools, which can be used in air transport in the acoustic analysis of negotiations between crews and ground services.

2.2 An approach to automatic recognition of PEC using intelligent information analysis methods

A block diagram explaining the proposed process of developing an intelligent method for automatic recognition of emotions from the speaker's speech is shown in figure 2.1.

In accordance with Figure 2.1, to solve the PES recognition problem, it is necessary to develop a mathematical model that will be able to perform a multi-class classification according to seven types of emotional states with adequate accuracy. The model receives as an input the features of the classified object, which are extracted as a result of preprocessing. At the output of the model, the classification result is a vector of values of the probabilities of assigning the object under study to one of the seven classes of PES, *Y*.

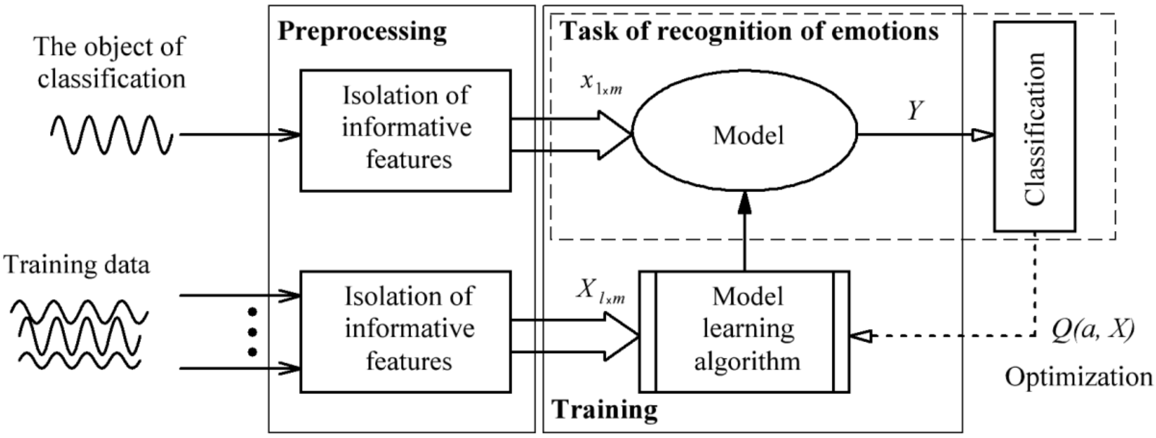


Figure 2.1 – The structure of the automatic classification process of PES by intelligent methods of information analysis

To synthesize the required model, algorithmic methods from the theory of computer assisted learning are used. In the process of tuition, the training data is fed to the input of the model in the form of a training sample from audio recordings of speech signals with various manifestations of PES, from which informative features are extracted (preprocessing). For the training sample, for each file it is known in advance what type of emotion it corresponds to.

As a result, the model building process is an intelligent CAL method known as supervised learning or labeled data learning. The essence of this method can be formulated as follows. For existing training sample $X=(x_i, y_i)_{i=1}^{l}$ it is necessary to find algorithm $a \in \mathrm{A}$, for which the minimum error functionality will be achieved *Q* (*a*, *X*):

$$Q(a, X) \to \min_{a \in \mathrm{A}}. \tag{2.1}$$

Thus, depending on the input object, the model forms the probability of its assignment to one of the classes of PES. The *Y* model's response is compared with the known correct response. The comparison result is expressed in the form of some accepted error functional *Q* (*a*, *X*), as shown in figure 2.1. The self-learning process of the algorithm consists in striving to reduce the value of the error functional by sequentially changing the value of the parameters of model. The model learning algorithm is terminated when the global minimum of the error functional or one of its local minima, satisfying the conditions imposed on the quality of the classification, is reached.

The existence of local minima greatly complicates the learning process of the model. Also, the correct choice of the minimized error functional has a great influence on the quality of classification. For the model training algorithm, it is necessary to select the correct hyper parameters that determine the efficiency of its work. With the correct tuning of hyper parameters, the phenomenon of overfitting can be avoided, providing the model with a high generalizing ability on new data. [44].

Thus, to obtain a classifier model, it is necessary, first of all, to have a training dataset in the form of records of human speech with different emotional coloring.

2.3 Formation of the training dataset

Due to the specifics of the problem being solved, during the formation of the emotional corpus (a set of voice recordings with different emotional coloration), difficulties inevitably arise in obtaining a spontaneous emotional speech of the speaker not only in the field of air transportation, but also in any other spheres of human activity. At the same time, the process of accumulation of a large number of marked sound recordings with an emotional connotation, which is rarely manifested in a person's daily activities, is greatly hampered and stretched in time. In addition, the receipt of sound recordings with certain types of emotional states encounters certain moral barriers on its way.

A generally applicable solution to these problems is the use of model bases formed with the participation of professional actors as an emotional corpus of sound recordings. It can be expected that the use of the automatic classification system of PES, built using model bases, will lead to a decrease in efficiency when switching to spontaneous speech.

However, there is definitely a clear similarity in the manifestation of PES in conversation among arbitrary speakers, as evidenced by research in evolutionary biology [45]. Therefore, corpuses from the recordings of professional actors can be effectively used to create and initially evaluate systems for voice analysis of emotional state. The use of representative databases of records that have proven their reliability among other researchers will help to avoid obvious difficulties when working with spontaneous speech and to reveal the relative efficiency of the developed algorithms at the initial stage of testing [46].

An analysis of the existing emotional corpuses available today showed that the following bases of audio recordings will satisfy the objectives of this investigation:

– the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [47];

– surrey Audio-Visual Expressed Emotion (SAVEE) [48];

– Toronto Emotional Speech Set (TESS) [49].

The RAVDESS corpus is an audiovisual database of the emotional speech and songs. From the entire corpus, audio recordings of the emotional speech, obtained from 24 professional actors (12 women and 12 men), who voiced two lexically selected utterances with a neutral North American accent, were selected. Speech includes joy, fear, anger, sadness, disgust, surprise, calmness, and neutrality.

Each speech pattern is spoken at two levels of emotional intensity: normal and strong. Sampling frequency of audio files is 48 kHz, in *.wav format.

The SAVEE database consists of entries from 4 male actors for 7 psycho-emotional states: joy, fear, anger, sadness, disgust, surprise and neutrality. A total of 480 sayings were in British English. The sentences were selected from the standard TIMIT corpus [50] and are phonetically balanced for each emotion. Audio sampling rate is 44.1 kHz, in *.wav format.

In the TESS record base, information is presented as a set of 200 target words spoken in the carrier phrase “Say the word ...” by two actresses (aged 26 and 64) in the seven above-mentioned emotional states. The total number of records is 2800. Both actresses speak English. Audio sampling rate is 24.414 kHz, in * .wav format.

In figure 2.2, the statistical characteristics of the duration of audio recordings from the selected databases are presented in the form of boxplot diagrams [51].

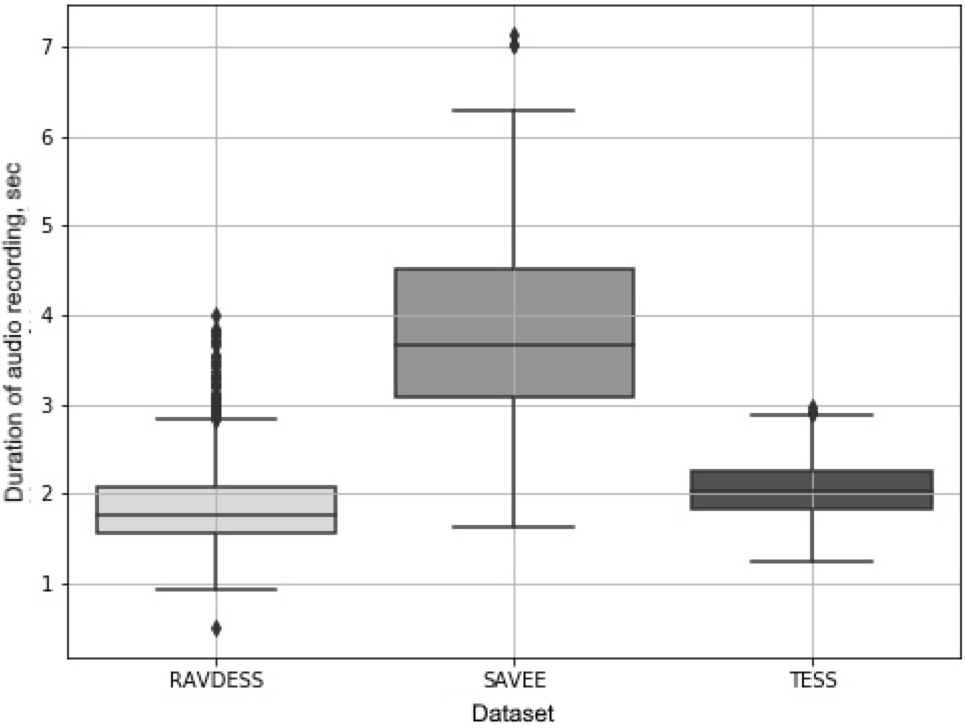


Figure 2.2 – Statistics on the duration of audio recordings from selected databases

As follows from Figure 2.2, the overwhelming majority of audio recordings have duration in the range of 1-7 seconds. In this regard, files with duration of less than 1 second and more than 7 seconds were removed from the training sample, since they will differ significantly from other samples.

Table 2.1 presents data on the total number of audio recordings included in the final training dataset, as well as their distribution by PES grades.

Table 2.1 – Distribution of audio recordings by PES grades in the training dataset

Emotion type	RAVDESS	SAVEE	TESS	Total
Angry	192	60	400	652
Disgusted	192	60	400	652
Fearful	192	60	400	652
Happy	192	60	400	652
Neutral	96	120	400	616
Sad	191	60	400	651
Surprised	192	60	400	652
Altogether	1247	480	2800	4527

As follows from the description of the prepared training sample, the dataset contains recordings of seven types of psycho-emotional state, but audio recordings differ significantly in the form of expression of emotions, the composition of speakers and pronunciation options. This structure of the corpus of emotions will facilitate a more varied representation of samples for training the model of classifier.

Audio recordings from these databases were downloaded from the Internet to the hard drive of a personal computer for further processing. All received audio files were indexed, resulting in a catalog of available samples in the form of a file in CSV format. This file consists of 4527 records, each of which contains information about the location of the audio file on the computer disk, the name of the data set to which the file belongs, the speaker's unique alphanumeric cipher, gender, and the total length of the recording in seconds.

Having a sufficient number of samples of speech signal is fundamental to building the model of classifier. However, for a productive search for a solution to the posed problem of automatic PES classification, it is necessary to form a clear understanding of the physical nature of the speech signal and to determine the processes and phenomena that generate it.

2.4 Equivalent model of the speech production process

To be able to use intelligent methods of information analysis in the problem of automatic recognition of PES, human speech should be represented by a set of characteristics that can act as informative features for the model of classifier. For this reason, in a technical system, it is advantageous to interpret speech as a signal, the physical carrier of which is an acoustic vibration. That is, speech is a sequence of sounds separated by pauses of various lengths.

In accordance with this, when using modern DSP methods in speech analysis tasks, it is necessary to have an idea of the processes of speech production in discrete time. For this purpose, the speech signal is represented as a response of a non-stationary linear system to the effect of noise or a quasi-periodic sequence of pulses [52], as shown in figure 2.3.

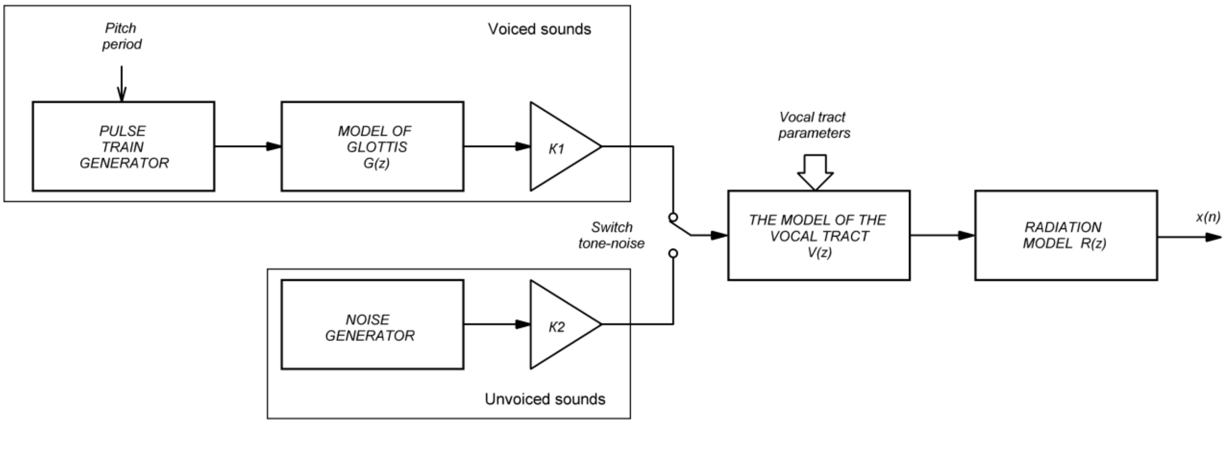


Figure 2.3 – Block diagram of the discrete model of speech production

In accordance with Figure 2.3, the vocal tract can be represented by an equivalent discrete-time model with a transfer function of the form:

$$V(z)=\frac{G}{1+\sum_{k=1}^{n}a_k z^{-k}}, \tag{2.2}$$

where G, a_k – these are coefficients that change over time and are determined by the parameters of the vocal tract: the dependence of its cross-sectional area on the distance along the longitudinal axis.

From the position of DSP, the filter coefficients (2.2) are responsible for the position of the maxima on the amplitude-frequency characteristic, which are called formant frequencies. Formant frequencies (formants) are the resonant frequencies of the vocal tract. They have a direct impact on the formation of individual speech sounds.

In turn, the sounds used in speech production are called phonemes. For each phoneme, the envelope of the spectrum of the model (2.2) acquires a certain shape, depending on the position of the formant frequencies. In the process of pronouncing speech, phonemes change and formant transitions appear.

In the English language there are 42 phonemes, which are subdivided into vowels, diphthongs, semi-vowels and consonants [52, p. 45]. However, most speech sounds can be conditionally divided into those formed with the participation of the vocal cords - voiced, and those formed without the use of the cords - unvoiced.

Vibration of the vocal cords creates an intermittent movement of air passage from the lungs, which can be considered periodical. The corresponding repetition period of the air flow pulses is called the pitch period. As can be seen from the Figure 2.3, in the speech production model, for generating voiced sounds, the pulse train generator activates a sequence of single pulses with the pitch frequency F_0. The pulse shape is determined by the transfer function G (z) of the linear system, the impulse response of which corresponds to the shape of the oscillation in the glottis. Block K1 determines the intensity of the voice signal using the appropriate gain.

In turn, the process of forming unvoiced sounds consists in using a noise generator (figure 2.3), the power of which is regulated by the gain K2.

The structural diagram of Figure 2.3 also takes into account the nature of the change in sound pressure near the lips in the form of a radiation model. This effect can be represented as a first approximation in the form of a differentiator of the form [52, p. 102]:

$$R(z)=G'(1-z^{-1}) \tag{2.3}$$

The gain ratio G' determines the intensity of the voice excitation.

As a result, the general transfer function of the discrete model of speech production can be represented as:

$$H(z) = V(z)G(z)R(z). \tag{2.4}$$

It should be noted that the presented model has many limitations associated with the ability to describe all the phonemes in the language. However, based on the structure of figure 2.3, a number of important conclusions can be drawn.

Firstly, for application of the technical means of DSP in tasks of research of speech, it is necessary to apply short-term signal analysis, since the model parameters will be constant only at certain time intervals.

Secondly, the structure and parameters of the model in figure 2.3 suggest that useful information about the speech signal will be predominantly located in the frequency domain. That is, the study of the spectral composition of the speech signal will reveal significant informative signs. The reason for this is that the model of speech production in figure 2.3 is a linear system that is excited periodically or randomly. Therefore, one should expect that the spectrum of the output signal will reflect the properties of both the vocal tract and the excitation itself.

Thirdly, the form of the speech signal will mainly have the form of quasi-periodic oscillations and noise. Moreover, the spectral composition of quasi-periodic oscillations will be determined by the frequency of the fundamental tone F_0 and formant frequencies. The noise spectrum is distributed over the entire frequency range, and the distribution function is not determinative.

For example, figure 2.4 shows the form of a speech signal when a man pronounces the word "two", recorded with a sampling frequency f_S = 44100 Hz. Unvoiced sound |*t*| appears as a noise signal with a time of 2.6 seconds. From about 2.7 seconds on the graph, you can observe a quasi-periodic process that refers to the voiced sound |*u*:|.

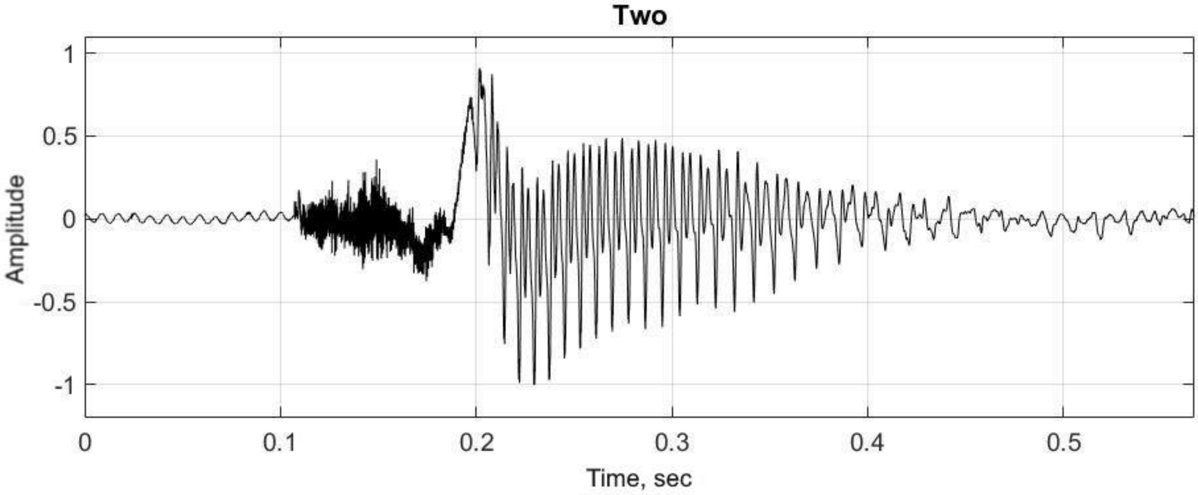


Figure 2.4 – The form of the speech signal when pronouncing the word “two”

Figure 2.5 shows the spectral composition of the speech signal when pronouncing the word “two”. Phoneme spectra are superimposed on each other, which makes it difficult to determine the frequency of the main tone F_0 and first formants.

Thus, having information about the characteristics and basic features of the speech signal, it is possible to develop the structure of the preparatory digital processing of data, called preprocessing.

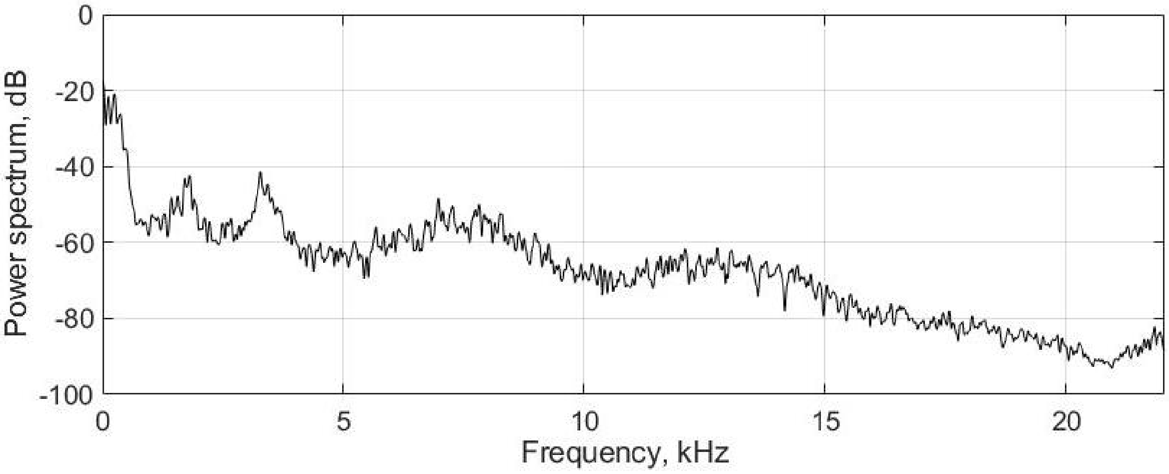


Figure 2.5 – Spectral composition of the speech signal when pronouncing the word “two”

2.5 Preprocessing in the task of automatic recognition of PES based on a speech signal

The main task of the preprocessing stage in automatic recognition of PES is to select informative features of the objects under study. Informative features are used as input parameters for the classification model. However, before extracting informative features, a number of important transformations must be performed on the data of speech signals. These transformations should ensure the extraction of the speech signal from the sound recording with the least noise component and the removal of long pauses in the analyzed utterances.

Figure 2.6 shows the generalized structure of preprocessing in the task of automatic recognition of PES.

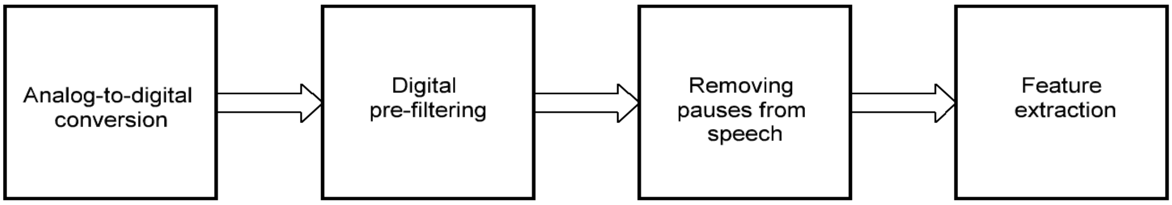


Figure 2.6 – The structure of the process of information transformation at the preprocessing stage

Depending on the nature of the speech audio signals used for automatic recognition of PES, some of the preprocessing steps shown in Figure 2.6 may be skipped. For a clear understanding of the need for certain manipulations with data, we will analyze the purpose of the procedures performed at each of the stages listed.

2.5.1 Analog-to-digital conversion

As follows from Figure 2.6 at the first step of data processing, the analog-to-digital conversion (ADC) procedure is performed the analog signal of the audio range is converted into digital format.

In turn, the stage of analog-to-digital conversion consists of the following components.

1. Low-pass filtering of an analog input signal using an anti-aliasing filter.

This procedure is performed to prevent the initiation of effect of aliasing occurring after sampling the signal. Anti-aliasing filter should suppress all frequencies above half the sample rate $f_S/2$ [53].

2. Sampling of an analog signal with a given frequency f_S.

The sampling rate f_S are selected based on the condition formulated in the Kotelnikov theorem. That is, the sampling frequency of the signal must be at least twice as high as its upper frequency component [54].

Human hearing organs perceive sound vibrations in the frequency range from 20 Hz to 20 kHz. Moreover, the frequency of the fundamental tone for male and female voices lies in the range of 70 - 450 Hz [55]. In accordance with the described model of speech production (figure 2.3), the speech signal will not be limited in frequency band, but its spectrum will decay rapidly for high frequencies, as can be seen from Figure 2.5. For voiced sounds, the highest frequency below which the spectrum maxima are less than 40 dB is about 4 kHz. For unvoiced sounds, the spectral components remain high for a frequency of 8 kHz [52, p. 161]. At the same time, some authors limit the frequency range of the speech signal within 70 - 7000 Hz [55, p. 7].

Regardless, for legible transmission of human speech, it is sufficient to accept f_S = 8 kHz, that is used in telephony. However, for accurate reproduction of the entire variety of speech sounds, a sampling rate f_s = 20 kHz is required. At the same time, since it is rather difficult to perform an anti-aliasing analog filter with a steep slope in the frequency response, the sampling frequency is chosen slightly higher than the required value, namely f_S = 22050 Hz.

Thus, in practice, the following sampling rates are used for digital audio recording [56]:

- 22050 Hz, 44100 Hz – Audio CD;
- 48000 Hz – DVD, DAT.

3. Quantization and coding of signal amplitude values.

To determine the required number of digits, with uniform quantization, you can use the following technique. As you know, for an analog-to-digital converter, the value of the signal-to-noise ratio can be approximately calculated using the following formula (2.5) [57]:

$$SNR = 6.02n + 1.76\partial Б, \tag{2.5}$$

where *n* – number of digits of ADC.

Assuming that the signal-to-noise ratio in sound recording should be about 45 dB, from expression (2.5) we obtain the value $n \approx 8$ bits.

However, modern common uncompressed digital audio formats such as wav can be in higher bit depths - 16-bit, 24-bit, 32 bits.

After the quantization process, each quantized signal sample is encoded in a binary code of the corresponding length.

4. Anti-jamming coding

Anti-jamming coding uses redundant codes to prevent errors that occur when reading files from a medium (for example, a hard disk of a personal computer).

Thus, to unify the process of obtaining digital recordings of speech signals in the systems of automatic recognition of PES, it is advantageous to use the sampling frequency of the common Audio CD format with a value f_S = 22050 Hz and quantization bit depth of 16 bits. On an electronic medium, data must be stored without compression in order not to lose the informative components of the signals.

2.5.2 Pre-filtration

In conditions of high noisy audio recordings at the preprocessing stage, additional preliminary digital signal filtering can be applied (figure 2.6). This procedure is especially important when the speech signal is received in the presence of strong concomitant additive noise from operating equipment, as well as background noise from AC voltage sources.

Since, as noted earlier, the frequency range of the human voice is not very informative below 70 Hz, a high-frequency filter with a cutoff frequency f_C = 70Hz and a sufficiently narrow transition region will be able to suppress low-frequency noise and quasi-harmonic processes represented by the components of the voltage pickup at the industrial frequency.

As a high-pass filter, the use of a Butterworth filter [58] with a transfer function of the form is justified:

$$H(z) = \frac{\left(1 + z^{-1}\right)^N}{\sum_{n=0}^{N} a_n z^{-n}} \tag{2.6}$$

The filter order in (2.6) is adopted to provide a steep slope N = 6 of the amplitude-frequency characteristics (AFC) in the transition area. The use of a Butterworth filter also provides a flat amplitude-frequency characteristic across the passband.

In addition, in a wide range of sound intensity, human hearing organs do not respond to changes in the phase relationships between individual harmonic components of the signal spectrum during monophonic reproduction. Therefore, in the problem of automatic recognition of PES, there is no need to limit or normalize the values of phase-frequency distortion [59].

As a result, the use of a digital filter with infinite impulse response becomes an effective solution.

Figure 2.7 shows the amplitude-frequency characteristic of the 6th order high-pass Butterworth filter. As follows from Figure 2.7, the synthesized filter has a high slope of the AFC after the cutoff frequency f_C = 70 Hz.

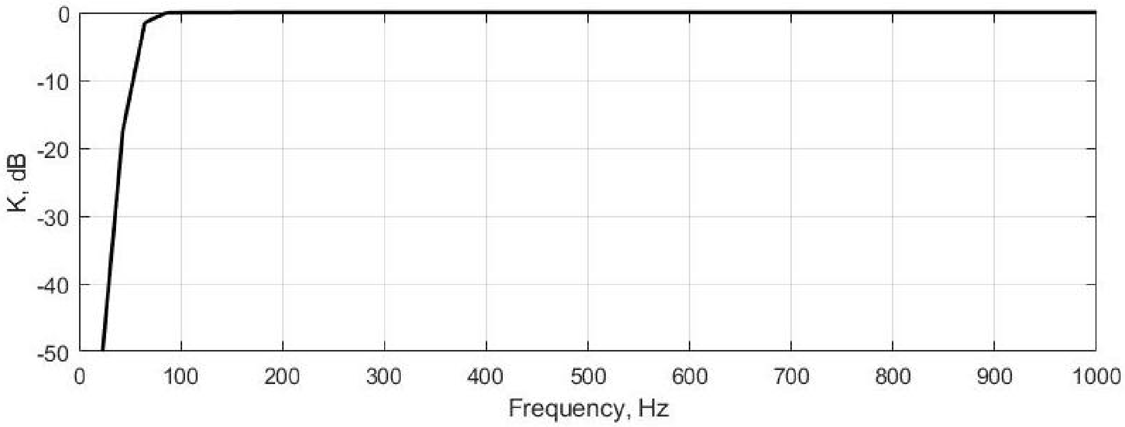


Figure 2.7 – AFC of a high-frequency Butterworth filter of the 6th order

Likewise, a low-pass filter can be synthesized to limit the frequency range of the speech signal from above. It should be borne in mind that after sampling the signal spectrum will already be limited by the frequency value f_S/2.

A special situation is presented by cases when the spectra of the noise and the useful signal mutually overlap or the frequency composition of the accompanying interference is not known in advance. Such cases are typical for situations when the transmission of a speech signal occurs against the background of noise from operating equipment, for example, in the cockpit of an aircraft.

It is impossible to get rid of such noise using the previously described method, but the obvious solution is adaptive filtering [60]. Figure 2.8 shows a diagram explaining the implementation mechanism of the adaptive filtering method.

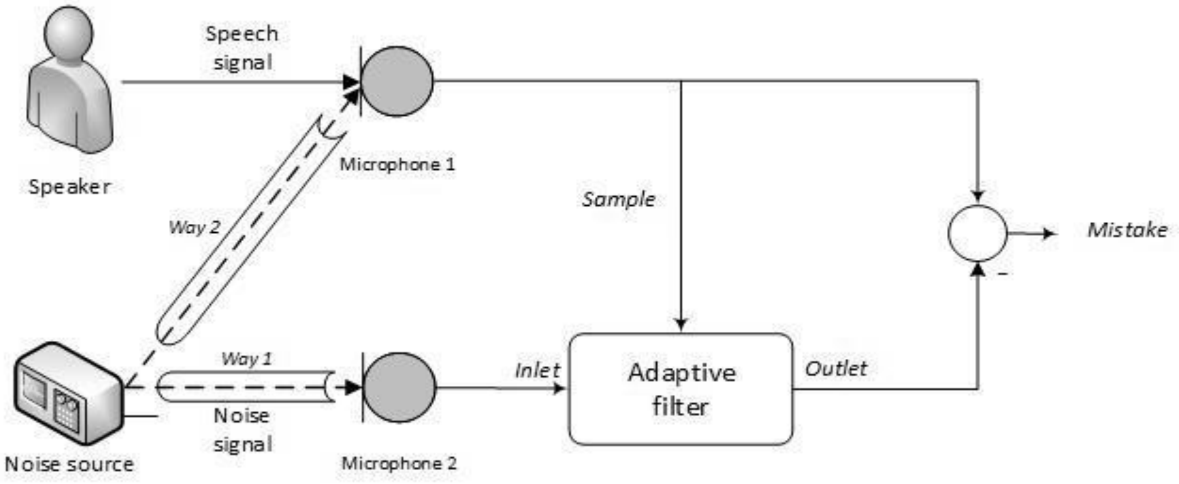


Figure 2.8 – The principle of implementation of adaptive filtration systems

As can be seen from the figure 2.8, in addition to the noisy speech signal, you should also receive a sample of the noise signal using a second microphone located in the immediate vicinity of the noise source (at a distance from the speaker).

The noise signal arriving at microphone 1 in figure 2.8 differs from the signal at microphone 2, since the paths and physical environment in which the sound wave propagates will be different. In this case, both noise signals on microphone 1 and microphone 2 will correlate with each other, since they have a common nature. In contrast to it, both noise and speech signals will be uncorrelated.

The adaptive filtering algorithm adjusts the weights of the non-recursive digital filter so that the noise signal from microphone 2 is as close as possible to the reference signal from microphone 1. Since only the noise component of the mixture of the speech signal and interference is correlated with the input signal of the filter, an estimate of the noise included in the sample from microphone 1 will be formed at the output of the filter. The difference between the reference signal and the resulting noise estimate will be considered as a conversion error, which in turn is a noise-free speech signal.

Among the common algorithms for adaptive filtering are *RLS* – Recursive Least Square; *LMS* – Least Mean Square; Kalman filter estimator. These implementations differ in the quality of filtration, speed, and resource consumption, and are used depending on the conditions of the applied filtration problem to be solved [61].

Thus, on the basis of the foregoing, it can be concluded that modern digital filtering methods are capable of providing high-quality removal of accompanying noise from the speech signal even in conditions of strong noise components.

2.5.3 Removing pauses from speech

In the spontaneous speech signal received for the task of automatic recognition of PES, there will always be a large number of pauses of various lengths. Long pauses in the speaker’s speech will not contain useful information about the emotional color of the statement. Moreover, when extracting informative features from a signal, the presence of pauses in a sound recording with sharply differing temporal, spectral, statistical and other characteristics can lead to an incorrect interpretation of the selected speech parameters. Having information about the moment of the beginning and end of the phrase will significantly reduce the number of arithmetic operations on the samples of the signal under study. Therefore, at the preprocessing stage (figure 2.6), it is necessary to remove long pauses from speech recordings.

As it is known [62], in human speech, each individual sentence is divided into semantic groups called speech measures. A speech measure consists of one or more words and is separated from the others by the logical pauses of different lengths.

Depending on the duration, logical pauses in speech can be divided into four groups:

1. Short pauses or backlash pauses, which are necessary for breathing in, as well as focusing on an important word.

2. Pauses separating speech measures.

3. Longer pauses between sentences that are related in meaning.

4. The longest pauses between the statements that are not related in meaning.

Along with other linguistic factors, pauses of the first and second types will play an important role in the formation of the emotional coloring of speech. However, the pauses of the third and fourth types can already be more than 1.5 seconds long, and when processing a speech signal, they must be removed to reduce the size of the processed sample and correctly interpret the received informative features.

Thus, when removing pauses from speech, it is necessary to preserve the integrity of the sequence of speech measures in phrases and sentences, but at the same time to remove long empty sections on the sound recording.

For modern digital communication systems, the task of automatically removing pauses from speech is currently solved and is reduced to the use of speech activity detection devices, known in the English literature as VAD systems (VAD stands for Voice Activity Detector). When pauses are detected, the telecommunications system equipped with VAD stops the audio signal transmission and only broadcasts information about the general description of the background noise present [63].

Telecommunication systems that use VAD to compress pause sections have many advantages, including increased communication channel capacity, energy efficiency, reduced packet loss, etc.

An example of an effective solution for pause detection is the VAD encoder used in the GSM (global standard for digital mobile cellular communications) standard. Figure 2.9 shows a generalized block diagram of such a VAD encoder.

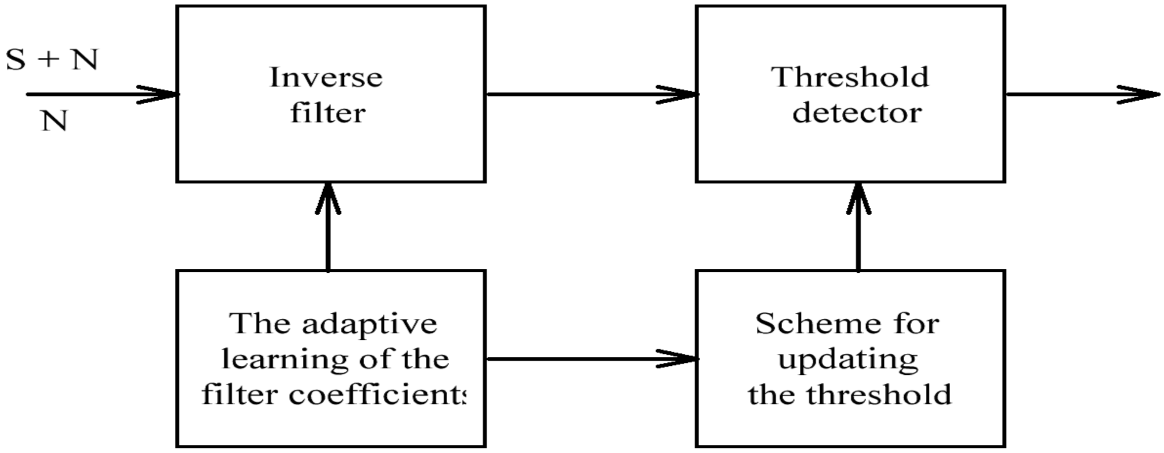

S + N
N
Inverse filter
Threshold detector
The adaptive learning of the filter coefficients
Scheme for updating the threshold

Figure 2.9 – Generalized structure of the VAD encoder used in the GSM communication standard

The GSM standard adopts the VAD scheme with frequency domain processing. The principle of operation of such a VAD is based on the difference between the spectral characteristics of human speech and noise that exists during pauses. In this case, the background noise can be considered stationary with a sufficient degree of reliability over a long-time interval, i.e., its spectrum is subject to significantly slower changes than the spectrum of transmitted speech.

Based on this, the VAD algorithm determines the spectral deviations of the input stimulus (S + N in figure 2.9) from the background noise spectrum (N in figure 2.9). This is achieved with an inverse filter. When a mixture of speech and noise (S + N) is present at the input, the filter suppresses the spectral components of the noise, thereby reducing its intensity. After that, the energy of the signal and noise mixture at the output of the inverse filter is compared with the threshold, which is calculated when only noise is present at the input.

Since the calculated threshold is above the energy level of the noise signal, the excess of the threshold level is taken as the presence at the input of the S + N implementation.

For correct operation of the encoder, the coefficients of the inverse filter and the threshold value adaptively change over time, depending on the current value of the noise level [64].

In addition to the described variant of VAD implementation, many other algorithms are used in practice (for example, G.729B / G.723.1A, AMR, IS-127/133, etc. [65]), differing in the principle of operation, speech detection accuracy depending on the level of interference, complexity of implementation, speed, etc. Also, methods for detecting pauses in the time domain, using a threshold procedure for the magnitude of the signal energy, as well as the number of signal transitions through zero, are known [52, p. 123].

In such a situation, the problem of removing pauses from the analyzed speech with automatic recognition of PES can be successfully solved on the basis of existing algorithms.

2.5.4 Feature extraction

As follows from the model in figure 2.3, an important feature of the speech signal is its statistical quasi-stationarity for short periods of time. For this reason, when working with speech signals, it is beneficial to perform short-term analysis by examining individual sections of the sound recording, called frames. Then, within the frame, the speech signal will not undergo significant changes. The size of the frame is determined by the size of the window that moves along the signal under investigation. When forming frames, overlap between adjacent areas can be organized, as shown in the figure 2.10.

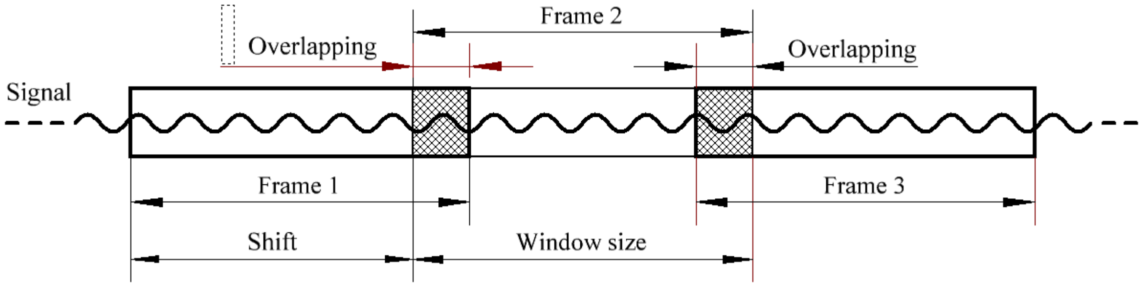


Figure 2.10 – The process of forming frames for the investigated speech signal

The need to overlap adjacent frames may arise when sound distortions appear at the edges of the processed signal fragments.

At the same time, splitting the speech signal into frames is effective from the perspective of reducing the computational complexity, since with the sampling frequency used in audio recording (22.05 kHz and higher), the number of unique signal values for performing arithmetic operations becomes too large.

In practice, the window size is chosen in the range of 20-40 m per s, since in this interval the properties of the voice path can be considered unchanged.

Thus, in the process of identifying informative features, frames of a speech signal can be considered as a sequence of separate sounds with different properties. Then the result of processing each frame is the calculation of a set of features for the sound contained in it. As a result, during short-term analysis, at each time interval, certain signal features are calculated, which can act as some time-dependent informative characteristic.

Feature extraction methods based on short-term analysis can generally be described by an expression of the form

$$Q_m = \sum_{n=-\infty}^{+\infty} F\left[x(n)\right] w(n-m) \tag{2.7}$$

where F [•] – is a linear or nonlinear transformation;

$x(n)$ – discrete speech signal;

$w(m)$ – window function of length M;

Q_m – sequence weighted value $F[x(n)]$.

The most common window functions for analyzing speech signals can be considered a rectangular window and a Hamming window [66], which is described by the expression:

$$w(m) = 0,54 - 0,46\cos\left(\frac{2\pi n}{M-1}\right), \tag{2.8}$$

where M – number of samples of the window function.

In accordance with the adopted discrete model of speech production (Figure 2.3), it is determined that in the study of speech signals, their spectral representation will be of great informative value. In such a situation, to obtain information about the dependence of the spectral characteristics of a speech signal on time, it is necessary to use a short-term (the windowed) Fourier transformation (FT).

The FT on an infinite interval does not reflect the local properties of the signal, and, therefore, does not have the ability to analyze the frequency characteristics of the signal at individual points in time. Short Time Fourier Transform (STFT) allows you to obtain a spectral representation that reflects the change in a speech signal over time.

The result of executing a windowed-discrete Fourier transform (WDFT) will be a matrix, the number of rows of which will be equal to the number of components of the discrete FT, and the number of columns is the number of frames. The frame number is determined by the expression:

$$k = \left\lfloor \frac{N - L}{M - L} \right\rfloor, \tag{2.9}$$

where N – number of samples in a signal $x(n)$;

M – window counts $w(m)$;

L – number of shift counts (figure 2.10).

Thus, as a result of performing WDFT over the speech signal, we will have a matrix of the form

$$\mathbf{X}_{m\times k} = [X_1(\omega)\ X_2(\omega)\ \ldots\ X_k(\omega)] \tag{2.10}$$

Where each m component is calculated by the formula (2.11):

$$X_m(\omega) = \sum_{n=-\infty}^{+\infty} x(n) w\big(n - m(M - L)\big) e^{-j\omega n}. \tag{2.11}$$

Figure 2.11 shows the short-term power spectrum for the signal in figure 2.4 at f_S= 22050 Hz, obtained as a result of the WDFT for 1024 samples using the Hamming window (2.8) with the size M = 1024 samples and the shift value L = 512. The frequency axis is shown on a logarithmic scale.

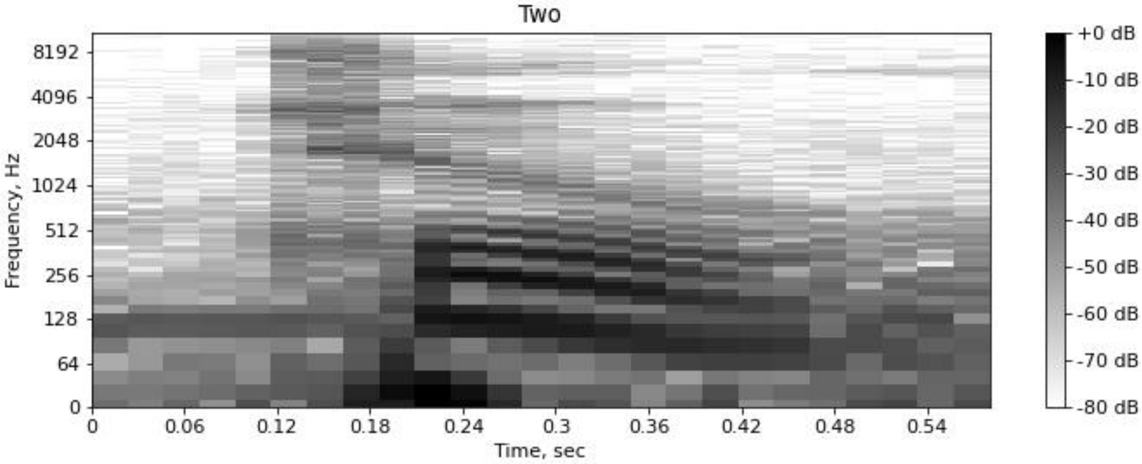


Figure 2.11 – Short-term power spectrum of a speech signal

Figure 2.11 shows that the spectrum of unvoiced sound (0.12-0.18 s) is distributed over the entire frequency range. The voiced sound (0.2-0.5 s) has pronounced maxima at the pitch and formant frequencies.

However, for the study of speech signals and recognition of the emotional state, it is ineffective to use information about the spectral composition of the signal obtained with the help of WDFT. It is necessary to take into account the subjective nature of the perception of sounds by the human hearing organs. In addition, it is probably necessary to monitor the intonation of speech and its change, manifested in the change in the frequency of the base tone [55, p. 6].

For these reasons, at the next stage of research, it is required to determine a method for representing the characteristic properties of a time-varying speech signal, which is able to take into account the peculiarities of the psychophysical perception of sounds by a person.

2.6 Selection of informative features in accordance with the objectives of the study

2.6.1 Preparation of audio recordings

Since specially prepared databases of speech signal records are used to build a classifier model, at the preprocessing stage (figure 2.6) it is only necessary to perform operations to remove pauses and extract features. As the analysis of audio files showed, the speech samples of the speakers contained in them were recorded without strong additive noise.

However, since the signal samples contained in the prepared training dataset have different sampling rates, it is necessary to apply the resampling operation to the frequency f_S = 22050 Hz.

Due to the fact that on certain audio recordings (in particular, for recordings from the RAVDESS and TESS databases), it is necessary to lower the sampling frequency not by an integer number of times, but by the value of p / q, the resampling process is performed according to the following algorithm.

1. Between the samples of the original signal with a sampling frequency f_S, (p – 1) of zero samples are placed.

2. To obtain one sample, the required oversampled signal with a sampling frequency $f_S{\cdot}p/q$, the signal obtained in the previous step is fed to the low-pass filter in portions of q samples. During the filtering process, arithmetic operations are not performed on zero samples.

For oversampling, the value q = 22050 Hz, and the value p corresponds to the sampling frequencies of the original signals (see subsection 2.3). The need for a low-pass filter (LPF) is explained by the occurrence of the same effects as when sampling an analog signal - in the process of oversampling, spurious frequencies appear. The cutoff frequency of the low-pass filter is taken to be $f_S/2{\cdot}\min(1, p/q)$ [54, p. 629].

The low level of background noise in the signals under investigation makes it possible to use a simplified procedure for removing pauses from the speech of speakers. In particular, a simple threshold procedure is applied, where on each segment the selected parameter of the speech signal is compared with a predetermined threshold, on the basis of which a decision is made whether the current segment belongs to either a pause in speech or an utterance.

The energy of the speech signal $x^2(n)$ or its absolute value $|x(n)|$can be used as a signal parameter for which the threshold procedure will be applied. When using energy, the ratio between the signal samples $x(n)$ will be violated, since the squaring operation is required.

Then, to exclude from the speech signals the moments of time in which there is no speaker's speech (pauses), it is necessary to perform a comparison with a given threshold *thr* of the signal $\hat{x}(n)$. In turn, the signal $\hat{x}(n)$ is a sampled sequence of absolute values of the analyzed audio signal $\hat{x}(n) = h(k) \bullet |x(n)|$, filtered with a moving average filter, where $h(k)$ – the impulse response of a filter, $x(n)$ – analyzed speech signal. In this case, the transfer function of the filter used has the form:

$$H_{MA}(z) = \frac{1}{W} \sum_{k=1}^{W} z^{k-\frac{W}{2}}. \qquad (2.12)$$

Expression (2.12) defines a moving average filter, the size of the moving window is chosen equal to W=0,1f_S, where f_S – the sampling frequency of the signal.

As a result of applying the threshold procedure at the transformation output, we obtain the desired signal *s*, the counts of which will correspond to the counts of $\hat{x}(n)$, taken under the condition $\hat{x}(n)$>*thr*.

Figure 2.12 illustrates the described process of removing pauses from the audio signal in the speaker's speech.

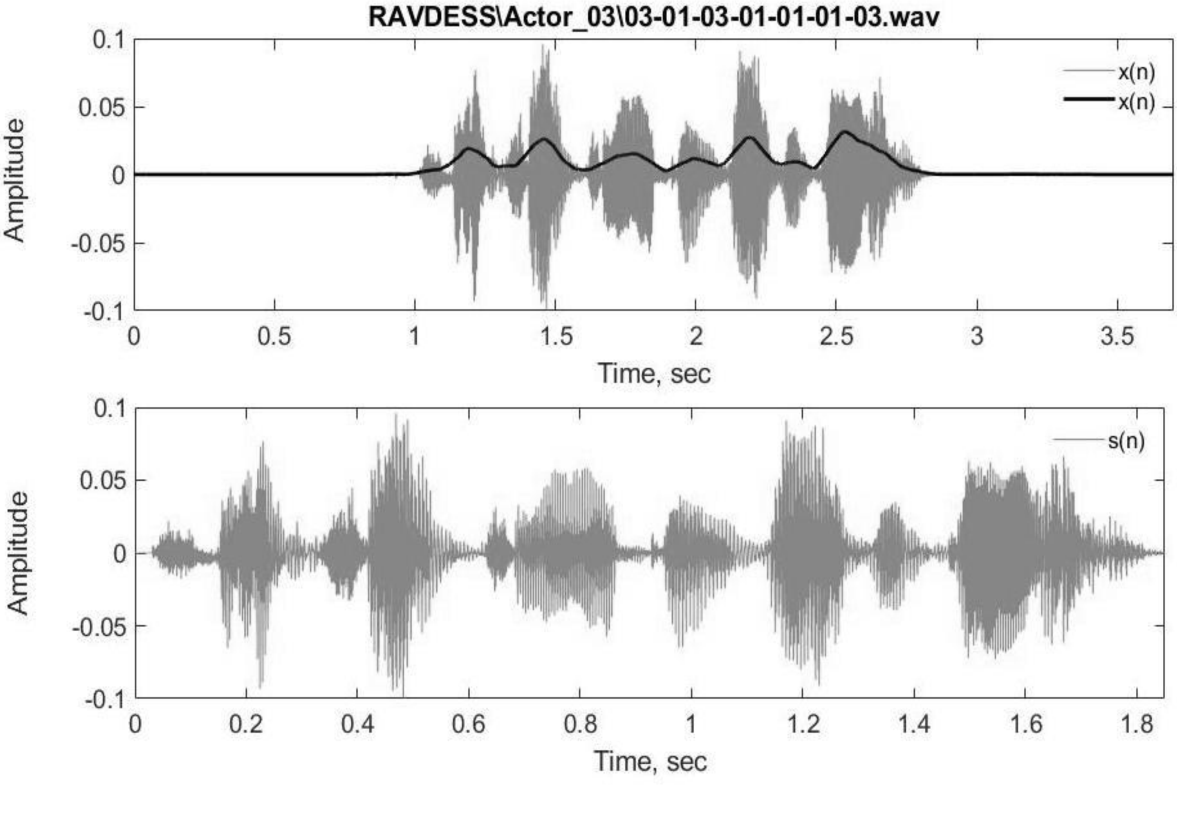


Figure 2.12 – The process of extracting fragments of speech from the audio signal

A record from the RAVDESS corpus was used as a speech signal in figure 2.12. As you can see from figure 2.12, in the process of removing pauses, the fragments of the audio recording before and after the start of the utterance are discarded. This allows you to reduce the size of the processed data, as well as analyze only moments of speech. The described procedure was applied to all available samples from the generated training dataset.

Next, the search for characteristic parameters of the speech signal is carried out, on the basis of which it will be possible to create a system for automatic detection of the emotional state of a person.

2.6.2 Mel-spectrogram of a speech signal

From the standpoint of the theory of signals and systems, the perception of acoustic waves by the human hearing organs can be considered as the equivalent of passing an audio signal through a bank of band-pass digital filters. Moreover, the bandwidth of these filters raises with increasing frequency in accordance with the expression [67]:

$$\Delta F_C = 25 + 75\left(1 + 1.4\left(\frac{f_C}{100}\right)^2\right)^{0.69} \qquad (2.13)$$

where ΔF_C – filter bandwidth;

f_C – center frequency. Frequency ranges of such filters will mutually overlap [68].

Such a representation of the system for the perception of acoustic vibrations by the human hearing organs is very important for understanding the informative value of the frequency response of a speech signal. It is known, however, that the frequency of an audio signal is related to a subjective characteristic known as pitch.

In accordance with the above, in order to take into account the psychophysical characteristics of the perception of sounds by a person, a special nonlinear frequency scale is introduced, called a mel-scale. Mel-scale reflects the relationship between the pitch of a tone and base frequency. This scale was obtained on the basis of statistical processing of a large number of data on the subjective perception of sound frequencies by humans and can be approximated by an equation of the form:

$$B(f) = 1125 \ln(1 + f / 700). \qquad (2.14)$$

Figure 2.13 shows a graph of the dependence of the pitch in mels on the frequency of the main tone. As can be seen from Figure 2.13, a pitch frequency of 1000 Hz corresponds to a pitch of 1000 mels. Below 1000 Hz, the ratio between pitch of a tone and frequency is almost proportional. However, for higher frequencies, the dependence is nonlinear.

Thus, by comparing the values of the frequency to the magnitude of the pitch, it is possible to construct the spectral characteristic of the sound signal, where a mel-scale will be used instead of the frequency axis. Such an estimate of the power spectrum of a sound signal will be called a mel-spectrogram.

Mel-spectrogram turns out to be a more informative characteristic of a speech signal, since when it is constructed; the psychophysical characteristics of the perception of sounds by a person are taken into account. We can say that such a presentation of the signal is more focused on the lower frequency range and with an increase in frequency, the information content decreases.

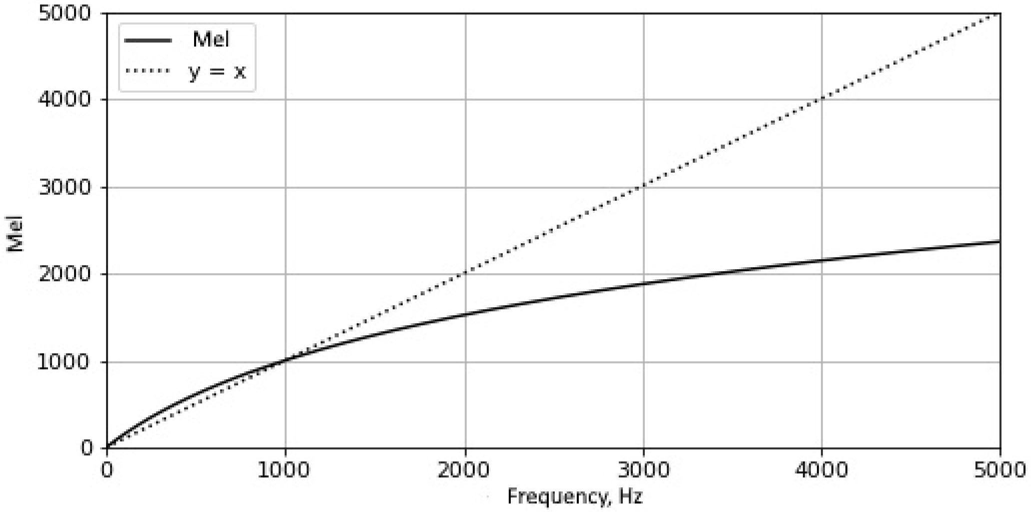


Figure 2.13 – Graph of the dependence of the pitch in mels on the frequency of the fundamental tone in Hertz

To construct a mel-spectrogram, a bank of M special triangular filters of the following form is applied to the original spectrogram:

$$H_m(k)=\begin{cases}0,\, k< f(m-1);\\ \dfrac{k-f(m-1)}{f(m)-f(m-1)},\ f(m-1)\le k\le f(m);\\ \dfrac{f(m+1)-k}{f(m+1)-f(m)},\ f(m)\le k\le f(m+1);\\ 0,\, k> f(m+1).\end{cases} \qquad (2.15)$$

where $m=\overline{1,M}$, and expression (2.15) satisfies the condition $\sum_{m=0}^{M-1} H_m(k)=1$.

For the filter bank (2.15), evenly spaced points are determined within a limited frequency range from f_{min} доf_{max}:

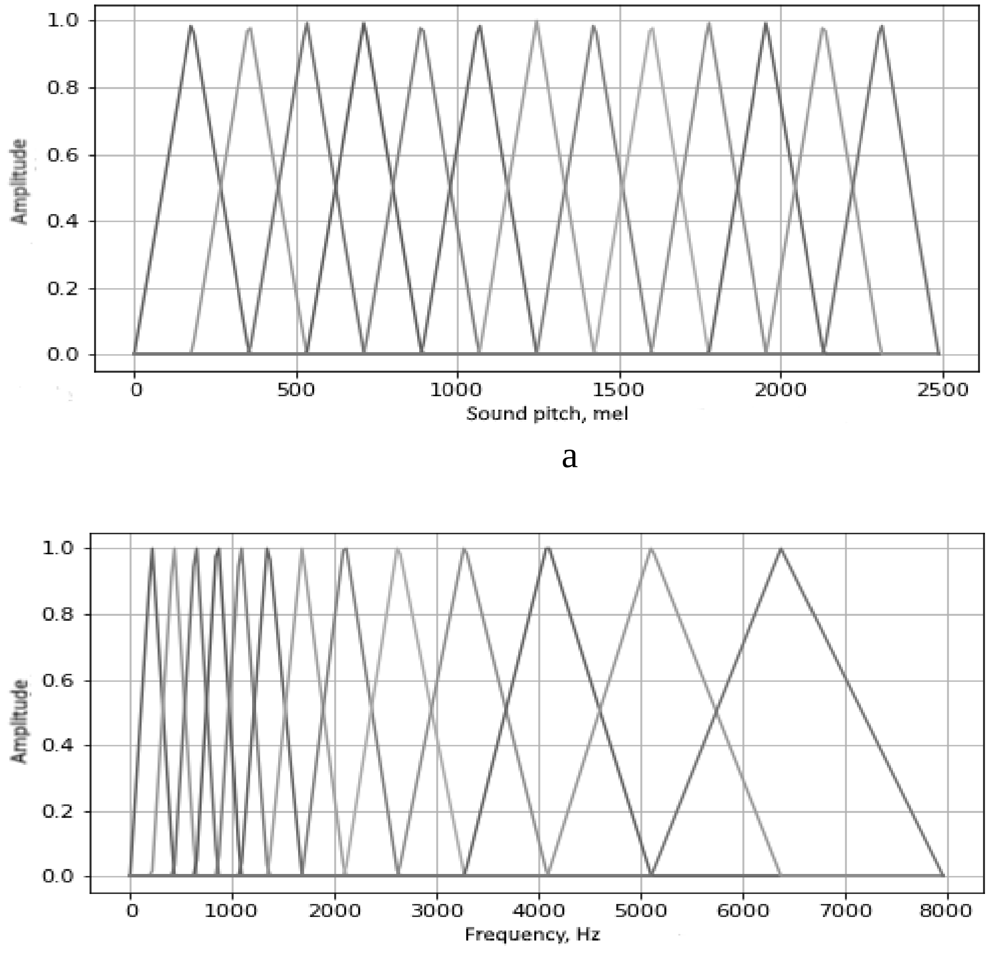
$$f(m)=\left(\frac{N}{f_S}\right)B^{-1}\left(B(f_{\max})+m\frac{B(f_{\max})-B(f_{\min})}{M+1}\right), \qquad (2.16)$$

where B^{-1} – this is the inverse transformation (3.14)

$$B^{-1}(b) = 700(e^{b/1125} - 1), \tag{2.17}$$

and $N = 2^j, j \in \mathrm{N}$ – number of samples of the analyzed signal frame.

Figure 2.14 shows the shape and arrangement of filters for *M* = 13, plotted on a frequency scale and a scale in mels.



a

b

a – arrangement of filters on mel-frequency scale; b – arrangement of filters on a frequency scale

Figure 2.14 – Bank of triangular filters for constructing a mel-spectrogram

When constructing the power spectrum, then the energy values of the spectrum components at the output of each filter are calculated (2.15):

$$P_i(m) = \sum_{k=0}^{N-1} \left(S_i(k)\right)^2 H_m(k), \quad 0 \le m < M. \tag{2.18}$$

In the expression (2.18) $S_i(k)$ – fast DFT of speech signal *s*, obtained by short analysis (2.7):

$$S_i(k) = \sum_{n=0}^{N-1} s_i(n)w(n)e^{-j2\pi kn/N}, \qquad (2.19)$$

where $w(n)$ – Hamming window function (2.8) used to reduce DFT leakage over a finite interval;

k – DFT index in the frequency domain.

Figure 2.15 shows the result of constructing a mel-spectrogram (*melspec*) when performing a short-term analysis for the audio signal from Figure 2.4. Mel-scale is shown on a logarithmic scale.

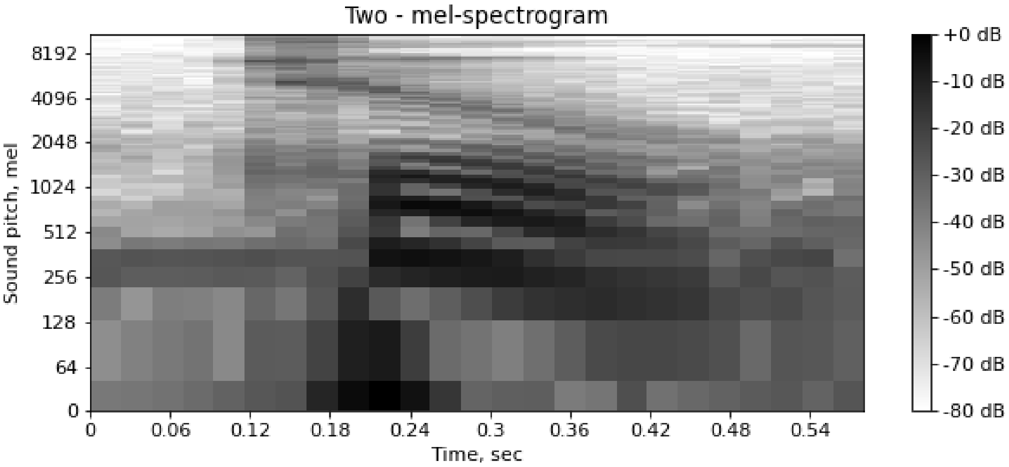


Figure 2.15 – Mel-spectrogram (*melspec)* of a speech signal

Thus, in the problem of automatic recognition of PES based on a speech signal, the mel-spectrogram is a more informative characteristic than the calculation of the power spectrum in the frequency range, since the mel-scale allows one to assess the subjective psychophysical perception of the pitch of the sound by the human hearing organs.

The use of short-term analysis in calculating the mel-spectrogram makes it possible to evaluate the change in pitch over time, and, therefore, to track the change in the speaker's intonation.

2.6.3 Mel-frequency cepstral coefficients

In addition to the analysis of mel-spectrograms, in the practice of automatic speaker-independent speech recognition, a significant effect is achieved when using Linear Prediction Coefficients (*LPC*) and Linear Prediction Cepstral Coefficients (*LPCC*) as informative features. These features are used in speech recognition systems based on the use of hidden Markov models [69].

However, at present, in systems for processing speech signals, the Mel Frequency Cepstral Coefficients (*MFCC*) are more widespread [70]. Unlike *LPC* and

LPCC, the information content of *MFCC* is based on the use of patterns of sound perception by human hearing organs based on the mel-frequency scale. Because of this, *MFCC*s are less sensitive to the features of the speaker's vocal tract and, unlike a mel-spectrogram, significantly reduce the space of individual characteristics of the speech signal. In this regard, *MFCC* can also be used in the problem of classification of emotions by voice.

The structure of the extraction process of the speech signal of mel-frequency cepstral *MFCC* coefficients presented in figure 2.16.

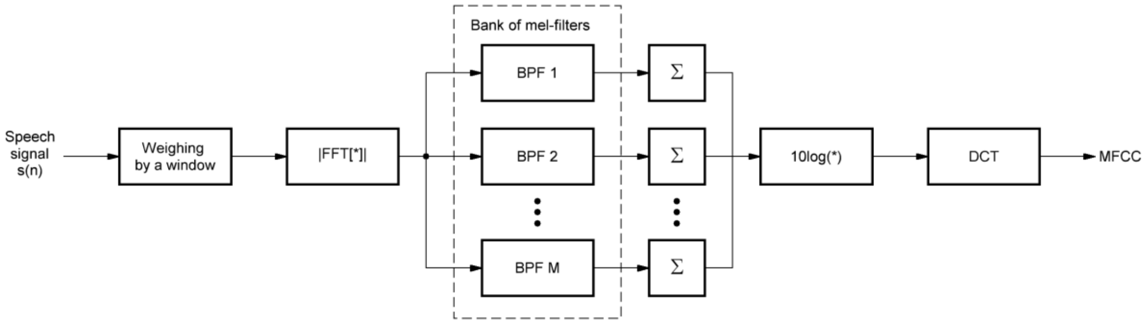


Figure 2.16 – Scheme for calculating the mel-frequency cepstral coefficients for the input speech signal

The process of calculating the *MFCC* at the initial stage is similar to the process of calculating the mel-spectrogram. As follows from Figure 2.16, the input signal frame *s (n)* is weighted by the window (2.8). Next, the FFT operation is performed, indicated in the figure as | FFT [*] |, since only real DFT coefficients are used for further calculations. With the help of the synthesized filters (2.15), the spectral components are selected, after which the obtained components are summed up.

At the final stages, in contrast to the calculation of the mel-spectrogram, when determining the *MFCC*, the logarithmic value of the energy of the spectrum components at the output of each filter is calculated (2.15):

$$P_i(m) = 10\log\left(\sum_{k=0}^{N-1}\left(S_i(k)\right)^2 H_m(k)\right), \quad 0 \le m < M. \tag{2.20}$$

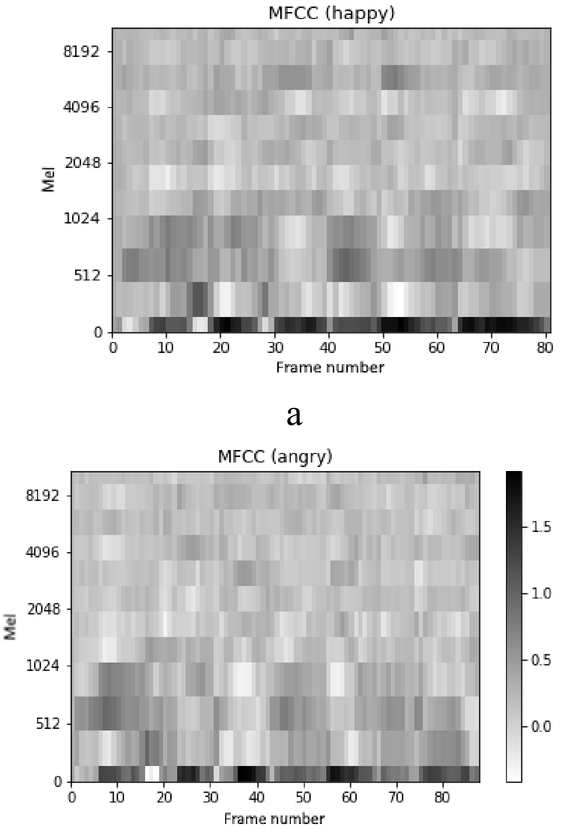
At the last step of the *MFCC* calculation (Figure 2.15), a discrete cosine transform (DCT) is performed for *P* (*m*):

$$c_i(l) = \sum_{m=0}^{M-1} P_i(m)\cos\left(\left(\pi l(m+\frac{1}{2}\right)/M\right), \quad 0 \le l < M. \tag{2.21}$$

DCT (2.20) is necessary when calculating *MFCC,* since the DFT of the impulse responses of the synthesized filters (2.15) mutually intersect, and the energies at the output of the filters are significantly correlated. DCT allows you to eliminate the emerging correlations.

After receiving c_i (l), coefficient c_i (0) is discarded, since it does not carry information about the speaker's speech and sets a constant offset.

Figure 2.17 graphically shows the standardized score of the calculated *MFCCs* c_i (1) – c_i (13) for processed audio recordings of the speech of the same speaker, but with a different emotional color. Audio recordings are obtained at a sampling rate f_S = 22050 Hz, frame length - 512 samples (about 23 m/s).



a

b

a – happy; b – angry

Figure 2.17 – Calculated values of *MFCC* c_i (1) – c_i (13)

As follows from the presented algorithm for calculating *MFCC*, the vector of these features describes the envelope of the power spectrum of the frame of the speech signal. However, the dynamics of changes in these parameters of the signal in time will also be of certain interest. Therefore, to determine the trajectory of the *MFCC* change, the calculation of their differential parameters, the so-called deltas (*delta*), can be performed as [71]:

$$d_l = \frac{\sum_{n=1}^{N} n(c_{l+n} + c_{l-n})}{2\sum_{n=1}^{N} n^2}, \qquad (2.22)$$

where N= 2.

Despite its advantages in the analysis of speech signals, the mel-spectrogram and mel-frequency cepstral coefficients do not allow to clearly distinguish information about the frequency of the fundamental tone for the analyzed fragment. It is assumed that the measurement of the frequency of the main tone in the speaker's speech can in a certain way characterize its intonation, and, consequently, the emotional color.

In this regard, below there is information on the use of special aggregated coefficients containing information about the frequency of the main tone.

2.6.4 Sound Pitch Classes

Speaking from the perspective of the subjective perception of acoustic vibrations by a person, in order to assess the frequency of the fundamental tone, the speech signal should be analyzed for the selection of a sound of the dominant pitch.

To calculate the gradations of the height of speech signals, the practice of assessing the equivalence of sounds in music can be taken as a basis. In particular, pitch classes can be applied. The meaning of splitting signals into these classes is the following observation: people perceive two sounds as similar (having the same chromaticity [72]) if the interval between their frequencies is equal to an integer number of octaves. From a formal mathematical point of view, sounds will belong to the same class if the ratio of their frequencies is equal to a whole positive or negative power of two.

Thus, the pitch can be characterized in terms of pitch classes, or in other words, its chromaticity, without the need to define a specific frequency. To do this, when creating a uniform scale, 12 classes or chromaticity values are considered, and their designation uses the corresponding symbols adopted in Western musical notation: {C, C♯, D, D♯, E, F, F♯, G, G♯, A, A♯, B} [73].

The main advantage of using the characteristics of pitch classes is that they can be used to aggregate all information about the pitch into one coefficient for a separate section of sound recording, thereby significantly reducing the dimension of the feature space.

When performing short-term analysis, the resulting representation of the chromaticity of a sound over time is called a chromogram or chromaticity spectrogram (*chroma*). Figure 2.18 shows, as an example, the calculated chromogram for a sample of a speaker's speech with a pronounced emotional coloring (angry).

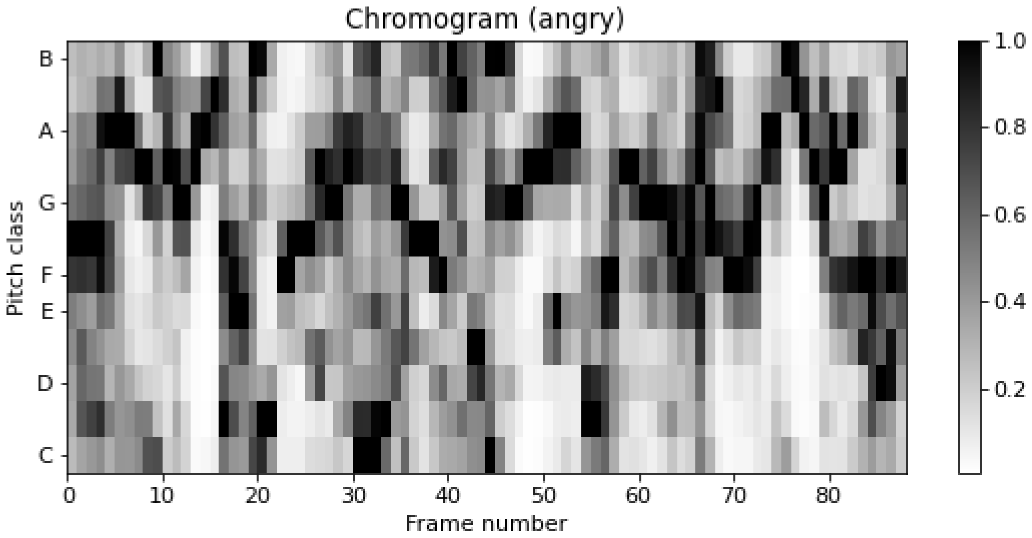


Figure 2.18 – Chromogram (*chroma)* of a speech signal

Calculation and construction of a chromogram and other informative features of a speech signal was carried out in Python 3.7 using the Librosa 0.8.0 library for audio data analysis [74].

In Figure 2.18, on the ordinate axis, the average normalized energy value of each of the 12 semitones over all octaves is presented. On the abscissa axis, the number of the analyzed frame is plotted for short-term analysis.

Calculation of the chromogram for the analyzed speech signals can provide their comparison with each other by the aggregated coefficients of the fundamental tone without the influence of the formant frequencies.

In addition to the described informative features, various characteristics of the spectral function, such as spectral centroid, spectral decay, spectral flux, rms energy, etc. are also used in the practice of analyzing audio signals. However, these features are more widely used in the tasks of automatic analysis of musical works, since they reflect the low-level characteristics of a sound object.

2.7 The general principle of constructing a mathematical model of the PES classifier based on a speech signal

Based on the analysis of speech formation process the known features of the psychophysical perception of sounds by a person, the vagueness and ambiguity of the existing formulations of the concept of emotion, as well as the ambiguity and complexity of identifying significant informative signs, it can be argued that the phenomena that generate data about the emotional state of a person by a speech signal are complex multifactorial process. In this connection, any mathematical model synthesized for the problem of the classification of psycho-emotional state by a speech signal will contain a certain amount of uncertainty, which does not allow making unambiguous conclusions as a result of classification.

Then, in the learning process when creating a mathematical model of the classifier, in accordance with figure 2.1, it is necessary to apply a probabilistic approach, i.e., it is advantageous to consider the CAL process from the standpoint of its probabilistic interpretation.

The probabilistic models of ML are based on Bayes' theorem, presented in the following form:

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)}, \tag{2.23}$$

where X – values of attributes of objects of classification;

Y – a plurality of target variables (object classes);

$P(Y|X)$ – posterior probability;

$P(X|Y)$ – likelihood function;

$P(Y)$ – prior probability;

$P(X)$ – the probability of observing data or features of objects obtained at the preprocessing stage (figure 2.6).

Prior probability $P(Y)$ represents the likelihood of each emotional class before observing the data X. The probability of observing data $P\ (X)$ does not depend on Y and can be eliminated by performing normalization, since the following expression is valid for it:

$$P(X) = \sum_{Y} P(X \mid Y)P(Y). \tag{2.24}$$

When performing the classification, the mathematical model must implement the decision rule, according to which the class with the maximum posterior probability must be selected. That is, in accordance with (2.22), we have the posterior maximum rule (MAP):

$$\begin{gathered} y_{MAP} = \arg\max_{Y} P(Y \mid X) = \arg\max_{Y} \frac{P(X \mid Y)P(Y)}{P(X)} = \\ = \arg\max_{Y} P(X \mid Y)P(Y). \end{gathered} \tag{2.25}$$

In accordance with (2.24), the construction of a mathematical model of the classifier will consist in solving the problem of optimizing the posterior distribution:

$$\begin{gathered} y_{MAP} = \arg\max_{Y} P(Y) \prod_{x \in X} p(x \mid Y) = \\ = \arg\max_{Y} \left(\log P(Y) + \sum_{x \in X} \log P(x \mid Y) \right), \end{gathered} \tag{2.26}$$

where $P(X\,|\,Y)=\prod_{x\in X}P(x\,|\,Y)$ – marginal likelihood function, according to which it is assumed that the probabilities of various features *x* within a class are independent.

If we make an assumption about the uniformity of the prior distribution *P (Y)*, then we can get the decision rule of maximum likelihood (ML):

$$y_{ML}=\arg\max_{Y}\sum_{x\in X}\log P(x\,|\,Y). \qquad (2.27)$$

In expressions (2.25), (2.26), the transition to the sum of logarithms is done to simplify the optimization process, since the logarithm is a monotonic function and argmax will not change.

Thus, the desired model of the PES classifier based on the speech signal will be obtained by solving the CAL problem, which in turn consists in finding and maximizing the distribution *P (Y | X).* In this case, it is necessary to find out which parameters best correspond to the available data, as well as the existing a priori representations. In practice, this problem is realized in the course of optimizing the logarithm of the likelihood of the model and regularizers [75].

Conclusions on the second section

Based on the studies carried out within the framework of the tasks of this dissertation work, it can be argued that today there is no general theory that reveals the relationship between the speaker's emotions and the characteristics of the voice signal. In this regard, automatic classification of PES by speech signal requires the development of new intelligent algorithms based on modern advances in digital signal processing and information and communication technologies, as well as deep optimization of existing solutions.

The task of automatic classification of a person’s emotional state based on his speech is distinguished by a number of difficulties among which the main ones are the following: the existing ambiguity in the formulation of the concept of emotions, the complex structure of the speech signal and the processes that generate it, the peculiarities of the psychophysical perception of sounds by a person, and, consequently, uncertainty in the choice of characteristics of speech signal.

In such a situation the use of intelligent methods of CAL theory can provide a solution to the problem of automatic classification of PES according to the speaker’s speech, since they allow revealing hidden patterns in the data, including in the presence of some uncertainty.

In turn, the problem of automatic classification by PES by CAL methods requires the formation of a representative set of training data. In accordance with this, the work formed a corpus of recordings of emotionally colored speech for seven grades in English, characterized by a variety of speakers of both sexes, a pronounced set of phrases, and the degree of emotional coloring.

As a result of the analysis of a discrete model of speech production, a preprocessing structure is proposed for the selection of informative features. The necessity of performing special DSP procedures for preliminary filtration and removal of pauses was established. The expediency of using short-term analysis of speech signals for PES classification is shown. On the basis of this, signs of objects for training the mathematical model of the classifier, which may contain information about the emotional color of speech, are proposed.

On the basis of a probabilistic approach to constructing a classifier model, the general principle of its training, which satisfies both CAL algorithms and deep learning methods, has been determined.

3 DEVELOPMENT AND EXPERIMENTAL STUDY OF INTELLIGENT METHODS OF DATA ANALYSIS FOR AUTOMATIC RECOGNITION OF PSYCHOEMOTIONAL STATE BY SPEECH SIGNAL TO INCREASE FLIGHT SAFETY

3.1 Development of the architecture of a deep convolutional neural network for automatic recognition of PES by a speech signal

3.1.1 Benefits of using convolutional neural networks

As presented in the previous section of the work, it is advisable to obtain informative features of a speech signal for recognizing a PES of a speaker using short-term analysis. Then, for a separate sample of a speech signal, each investigated feature will be a two-dimensional data structure (figures 2.15, 2.17, 2.18) in the form of an array of size $m\times n\times 1$, where m – the number of coefficients used when calculating the analyzed feature, and n – a number of signal frame. At the same time, it is reasonable to assume that not only the value of a particular coefficient, but also its location in space has informative value. In particular, the second dimension of the array carries information about the time scale, so the order along it will have a certain value, and the first dimension is responsible for the relative position of the frequency components. Consequently, the informative features selected for the classification of emotions have their own internal structure, the format of which is known to us. Array of calculated coefficients $X_{m\times n}$ can be considered as a monochrome image (having one channel), where a certain value of an informative feature acts as each pixel.

As practice shows [76], the most effective type of classifier when working with multidimensional data arrays at the moment are artificial convolutional neural networks (CNN).

When using fully connected neural networks, the additional knowledge about the internal data structure is not used in any way, and, therefore, part of the information is lost. In contrast, CNN is a neural network model, the format of which is focused on processing data with a mesh structure. This is achieved by using the convolution operation on at least one layer of the neural network.

As applied to the convolutional layer of a neural network, the operation of two-dimensional convolution can be represented in the form [75, p. 186]:

$$y^l_{i,j} = \sum_{-d\le a,b\le d} W_{a,b} x^l_{i+a,j+b}, \tag{3.1}$$

where $x^l_{i,j}$ – convolution entry;

W – convolution kernel with size weights matrix $(2d + 1)\times(2d + 1)$;

$y^l_{i,j}$ – the result of convolution at the l-th level (feature map).

The use of the convolution operation in the implementation of an artificial neural network makes it possible to radically improve the CAL algorithm. In particular, convolutional layers are characterized by the following features:

manifestation of sparse interaction, separation of parameters, and equivariant representations.

Due to the sparse interaction in the CNN, the neurons of the lower levels can interact with only a small part of the input neurons. That is, there is a local selection of features without reference to the specific position of the analyzed area. The network gains the ability to efficiently interpret complex relationships between many variables by using convolutional kernels, each of which implements only sparse interactions. When dividing parameters in a CNN, neurons with associated weights, when a weight value applied to one input appears, occurs elsewhere. As a result, this significantly reduces the memory requirements and computational complexity, in contrast to fully connected layers. CNN learns faster on large amounts of data with less time consuming. In addition, the convolution operation is equivalent to a relatively parallel transfer, which makes the CNN much less sensitive to parallel transfer or input shift.

However, in most cases, one convolutional layer is not enough to reveal relationships between data that are significantly separated in space. Therefore, it is necessary to use several consecutive convolutional layers with a nonlinear activation function and additional downsampling.

Thus, CNNs are significantly superior to other intelligent ML algorithms in the tasks of analyzing images or other data, where the spatial arrangement of features is important. At the same time, the use of deep learning technologies, when more than one convolutional layer is used in the network, allows for better analysis by identifying complex interactions in the data structure.

3.1.2 The architecture of a deep convolutional neural network for automatic recognition of PES by a speech signal

The search for the optimal configuration of the neural network was carried out by dividing the prepared training dataset (see subsection 2.3) into three subsamples for training, validation, and testing. To do this, using a random stratified partition [77], all data were divided in the following proportion: 70% is for network training, 20% is for testing, and 10% is for final verification. Isolation of the test subsample made it possible to compare different variants of network architectures with each other, as well as to make the selection of optimal hyper parameters.

As a result, within the framework of the problem being solved, the architecture of a deep convolutional neural network (DCNN), shown in figure 3.1, was developed (the dimension of the data is shown to the right of the diagram). The notation “div” in figure 3.1 is used to describe the mathematical operation of division without regard to remainder.

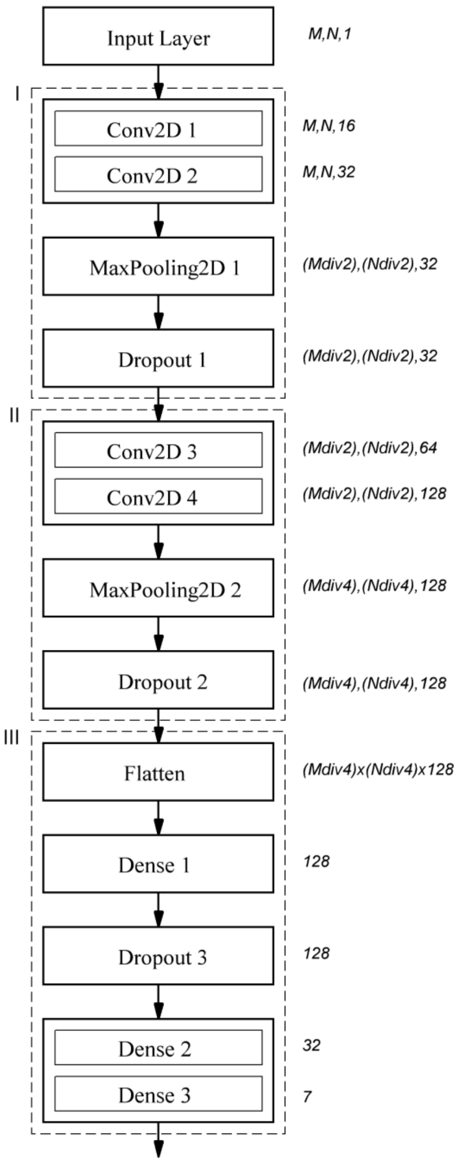


Figure 3.1 – Deep CNN Architecture

In accordance with figure 3.1, as a result of the conducted investigation, the network architecture consists of three main segments - two convolutional and one fully connected (the network segments in figure 3.1 are highlighted with a dotted line and denoted by Roman numerals). The first segment uses two consecutive convolutional layers Conv2D 1 and Conv2D 2 with 16 and 32 filters, respectively.

For all convolutional layers, the kernel size is 3 × 3 elements, and it is shifted by one element for each dimension. The choice of the kernel size is dictated by the requirement for the presence of a center in the filters used and the ability to express the ratio of top-bottom, right-left. At the output of convolutional layers, the size of the array does not change due to the padding of the input tensor with zeros outside the dimension.

The number of filters for the convolutional layers of the second segment Conv2D3 and Conv2D 4 is 64 and 128, respectively. Also, for each convolutional layer, an activation function is used in the form of nonlinearity ReLU (rectified linear units) [78]:

$$h(x) = \max(0, x). \tag{3.2}$$

The ReLU (3.2) function will return the value x if $x> 0$, and zero in all other cases. The calculation of this function is less expensive in terms of the computer time used. In addition, ReLU introduces the required nonlinearity, and also allows for sparse activation of neurons, thereby facilitating the network and reducing the resource intensity of computations in the learning process.

The network layers Max Pooling 2D 1 and Max Pooling 2D 2 in the network structure in figure 3.1 are responsible for the downsampling operation. In this case, for each local group of neurons of 2 × 2 size, the operation of finding the maximum is performed:

$$x_{i,j}^{l+1} = \max_{-2 \le a \le 2, -2 \le b \le 2} z_{i+a, j+b}^{l}, \tag{3.3}$$

where $z_{i,j}^{l} = h(y_{i,j}^{l})$ – values at the output of the activation function (3.2).

For regularization in the learning process, Dropout layers are used (figure 3.1). On Dropout 1 and Dropout 2, every fourth block of input data is discarded, on Dropout 3 - every second. Subsequent coats Flatten, Dense 1 and Dense 2 are compacted. That is, a fully connected neural network is already being implemented in this segment.

Dense output layer 3 is a softmax classifier with 7 output neurons for C = 7 types of classified emotions:

$$\sigma(x)_i = \frac{e^{x_i}}{\sum_{c=1}^{C} e^{x_k}}. \tag{3.4}$$

In accordance with expression (3.4), the developed DCNN as a result of classification will return a vector of seven numbers, representing the probabilities of seven types of PES for the analyzed sample of the speech signal.

Thus, the developed DCNN, using sequential convolutions, selects various feature maps, which can be considered as information transmission channels. Due to the MaxRooling 2D layers (see figure 3.1); the dimensionality of the analyzed data array will gradually decrease. As a result, the lower convolutional layers will be able to interpret not only individual information within the convolution kernel, but work with filtered data from the entire input array of informative features. Fully connected layers with a sequentially decreasing number of neurons combine local features identified by convolutional layers and form a classification result for the required number of classes.

3.2 Deep convolutional neural network training

DCNN training was carried out on selected informative features. The learning algorithm implements the backpropagation method. In this case, for the convolutional layers at each iteration of the network training, the error function Q is optimized by expressing its gradients from the obtained weights:

$$\frac{\partial Q}{\partial w_{a,b}^{l}} = \sum_{i}\sum_{j}\frac{\partial Q}{\partial y_{i,j}^{l}}\frac{\partial y_{i,j}^{l}}{\partial w_{a,b}^{l}}, \tag{3.5}$$

where i, j – dimensional data coordinates;

w – weight coefficients. As a result, to calculate the gradients of the error function, we have the following expression [75, p. 196]:

$$\frac{\partial Q}{\partial x_{i,j}^{l}} = \sum_{i}\sum_{j}\frac{\partial Q}{\partial y_{i-a,j-b}^{l}}w_{a,b}. \tag{3.6}$$

The optimization problem is solved by implementing stochastic gradient descent over mini-batches using the adaptive SGD algorithm [79] for the differential parameters of the mel-frequency cepstral coefficients (2.21) and using the Adam algorithm [80] for other types of informative features.

Thus, in the process of training the neural network, the search for the optimal values of the weights w^*, for which the network error takes the minimum value

$$w^* = \arg\min_{w} Q(w). \tag{3.7}$$

To perform a multiclass classification of PES, categorical cross-entropy is used as an error function:

$$loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} I\left[y_i \in C_c\right]\log p\left[y_i \in C_c\right], \tag{3.8}$$

where N – sample size;

$p\left[y_i \in C_c\right]$ – probability of belonging of the i-th sample to the class c;

$I\left[y_i \in C_c\right]$– indicator function that takes the value one if the sample actually belongs to the class c, and zero - in the opposite case.

The classifier model based on the developed DCNN was studied for the following types of informative features: mel-spectrograms (2.18), mel-frequency cepstral coefficients (2.20), differential parameters of mel-frequency cepstral coefficients (2.21) and pitch classes (chromogram is in figure 2.18).

To train the neural network, data obtained as a result of calculating the corresponding informative features for a randomly taken time interval within the analyzed sound was fed to its input. The above explains the algorithm shown in figure 3.2.

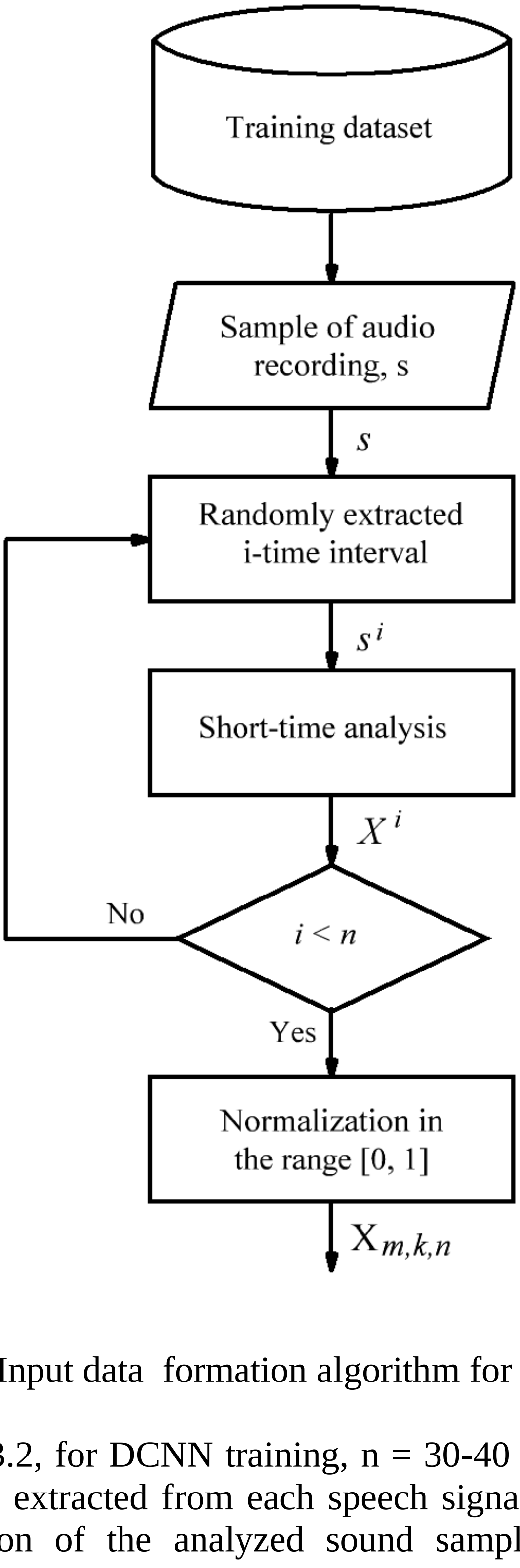

Training dataset
Sample of audio recording, s
s
Randomly extracted i-time interval
s^i
Short-time analysis
X^i
No
$i < n$
Yes
Normalization in the range [0, 1]
$\mathrm{X}_{m,k,n}$

Figure 3.2 – Input data formation algorithm for DCNN training

Based on figure 3.2, for DCNN training, n = 30-40 of two-dimensional arrays of input data X^i will be extracted from each speech signal sample *s*. The number *n* depends on the duration of the analyzed sound sample. The duration of each randomly extracted i-time interval *s1* is taken equal to 400 ms.

As a result, for each speech signal sample, we have an array of informative features $\mathrm{X}_{m,k,n}$, where *m* – number of the calculated characteristic value, *k* – frame number, *n* – number of randomly selected time interval. The obtained data is subjected to a linear operation of normalization in the range [0, 1]:

$$\widehat{X}^{i} = \frac{X^{i} - \min(\mathrm{X})}{\max(\mathrm{X}) - \min(\mathrm{X})}. \tag{3.9}$$

When performing short-term analysis, the frame duration is assumed to be *M* = 25 ms, the overlap interval is *L* = 10 ms. Taking this into account, the number of frames for each feature will be *k* = 27.

Table 3.1 shows the dimension of the data and the number of trained parameters for each layer of the developed DCNN when training on the selected informative features. The batch size in all cases was taken equal to 32.

In the case of using mel-spectrograms (*melspec* in table 3.1), the number of triangular filters (2.15) was taken equal to *M* = 128. To calculate the mel-frequency cepstral coefficients (*MFCC* in Table 3.1), the value *M* = 40 (3.20), and the coefficient c_i (0) is discarded. The calculation of the differential parameters of the chalk-frequency cepstral coefficients (*delta* in Table 3.1) is carried out through the obtained *MFCC* values (2.21) while maintaining the data dimension. When calculating the pitch classes (*chroma* in table 3.1), 12 corresponding coefficients are calculated.

Table 3.1 – Dimension of data and the number of trained parameters for each layer of the developed DCNN

Layer type	*Melspec*		*MFCC*		*Delta*		*Chroma*	
	output size	number of learning parameters	output size	number of learning parameters	output size	number of learning parameters	output size	number of learning parameters
1	2	3	4	5	6	7	8	9
Input data	(128, 27, 1)	0	(39, 27, 1)	0	(39, 27, 1)	0	(12, 16, 1)	0
Conv2D 1	(32, 128, 27, 16)	160	(32, 39, 27, 16)	160	(32, 39, 27, 16)	160	(32, 12, 16, 16)	160
Conv2D 2	(32, 128, 27, 32)	4640	(32, 39, 27, 32)	4640	(32, 39, 27, 32)	4640	(32, 12, 16, 32)	4640
Max Pooling 2D 1	(32, 64, 13, 32)	0	(32, 19, 13, 32)	0	(32, 19, 13, 32)	0	(32, 6, 8, 32)	0
Dropout1	(32, 64, 13, 32)	0	(32, 19, 13, 32)	0	(32, 19, 13, 32)	0	(32, 6, 8, 32)	0
Conv2D 3	(32, 64, 13, 64)	18496	(32, 19, 13, 64)	18496	(32, 19, 13, 64)	18496	(32, 6, 8, 64)	18496
Conv2D 4	(32, 64, 13, 128)	73856	(32, 19, 13, 128)	73856	(32, 19, 13, 128)	73856	(32, 6, 8, 128)	73856

Table continuation 3.1

1	2	3	4	5	6	7	8	9
MaxPooling2D 2	(32, 32, 6, 128)	0	(32, 9, 6, 128)	0	(32, 9, 6, 128)	0	(32, 3, 4, 128)	0
Dropout 2	(32, 32, 6, 128)	0	(32, 9, 6, 128)	0	(32, 9, 6, 128)	0	(32, 3, 4, 128)	0
Flatten	(32, 24576)	0	(32, 6912)	0	(32, 6912)	0	(32, 1536)	0
Dense 1	(32, 128)	3145856	(32, 128)	884864	(32, 128)	884864	(32, 128)	196736
Dropout 3	(32, 128)	0	(32, 128)	0	(32, 128)	0	(32, 128)	0
Dense 2	(32, 32)	4128	(32, 32)	4128	(32, 32)	4128	(32, 32)	4128
Dense 3	(32, 7)	231	(32, 7)	231	(32, 7)	231	(32, 7)	231
Total		3247367		986375		986375		298247

The synthesis and training of the DCNN model was carried out in the Python 3.7 programming language using the open neural network library Keras 2.3.1 [81, 82], working on the TensorFlow 2.1.0 framework [83].

The DCNN learning process is presented graphically in figure 3.3. The multiclass share of correct answers was used as a quality metric on the training subsample - *accuracy*:

$$acc = \frac{\sum_{c=1}^{C} \frac{tp_c + tn_c}{tp_c + tn_c + fp_c + fn_c}}{C}, \tag{3.10}$$

where $C = 7$ – the number of classified emotions;

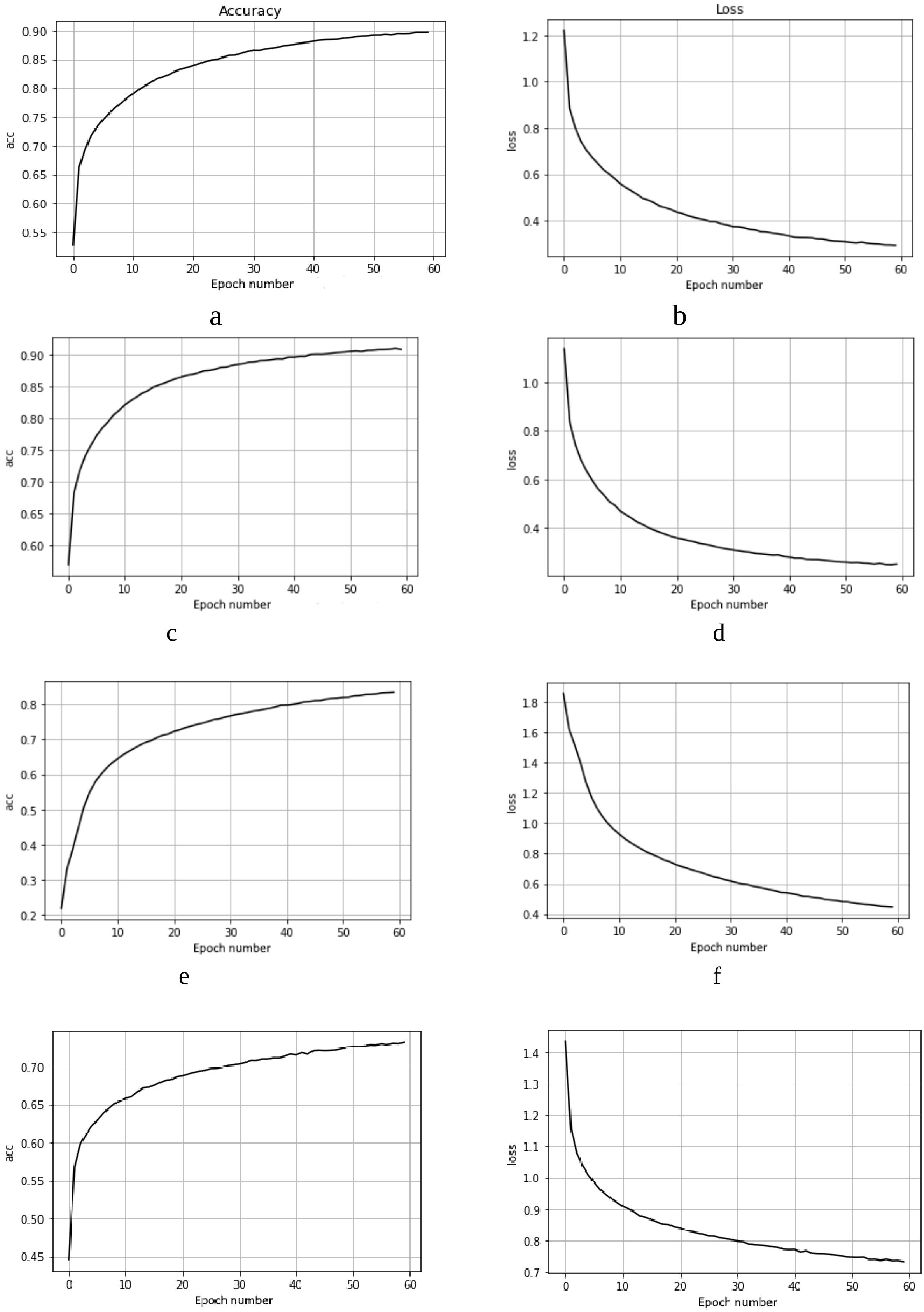
tp_c (*true positive*) – these are correctly classified specimens as belonging to the class c;

tn_c (*true negative*) – correctly classified samples as not belonging to the class c;

fp_c (*false positive*) – samples do not belong to the class c, but are incorrectly classified as belonging to it*;*

fn_c (*false negative*) – samples belonging to the class c, but are wrongly classified as not belonging to it.

The neural network has been trained for 60 epochs. As follows from the graphs in figure 3.3, in the learning process, there is a sequential increase in the *acc* value and a corresponding decrease in the error function (3.8). It can also be seen from the graphs in figure 3.3 that the type of the selected informative feature affects the quality of DCNN training.



a b c d e f g h

a, b – *MFCC*; c, d – *melspec*; e, f – *delta*; g, h – *chroma*

Figure 3.3 – DCNN training process: change in the proportion of correct answers *acc* and cross-entropy *loss* on the trained subset

3.3 Evaluation of the efficiency of classifier models based on the developed DCNN

The process of classifying speech signal samples in order to identify the emotional state of the speaker is carried out according to the method, the algorithm of which is shown in figure 3.4.

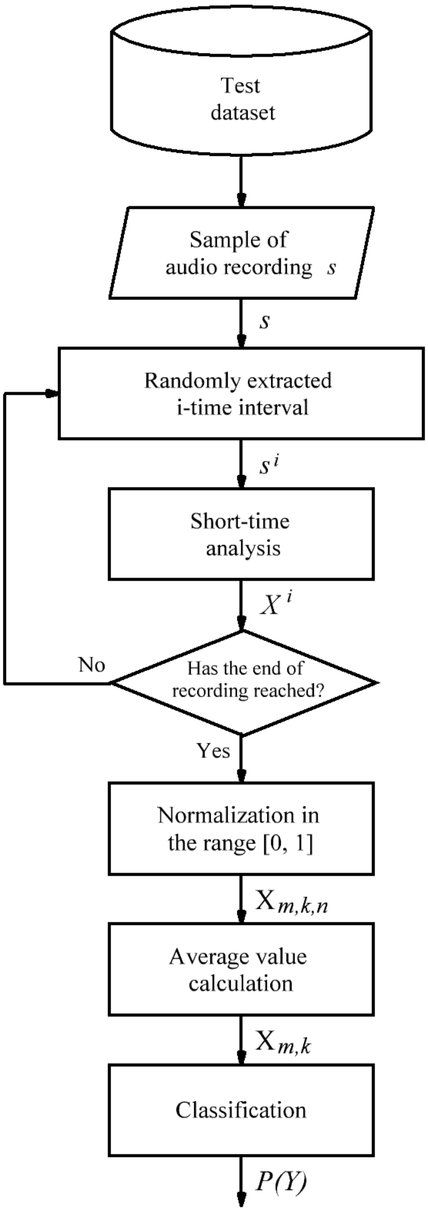


Figure 3.4 – Methodology for performing the classification of speech samples using a trained DCNN

In accordance with Figure 3.4, the classified speech signal sample s is divided into successive sections s^i of the same duration of 400 ms. Using a short-term analysis with a frame duration of M = 25 ms and an overlap of L = 10 ms, the required informative features are calculated. The resulting data are normalized and performed averaging over all n, obtained by two-dimensional arrays. As a result, the required informative feature $X_{m,k}$ is fed to the input of the DCNN model, which, in turn, forms the result of the classification *P (Y)* at the output, in terms of a vector of probabilities that the sample *s* belongs to each of the seven classes. Then the corresponding class of emotion will be defined as

$$c = \sum_{i=1}^{C} I\left[y_i = \arg\max(\mathrm{Y})\right] \cdot i. \tag{3.11}$$

Thus, the efficiency of DCNN models trained on various types of informative features can be estimated by the method of figure 3.4 on a delayed test subsample.

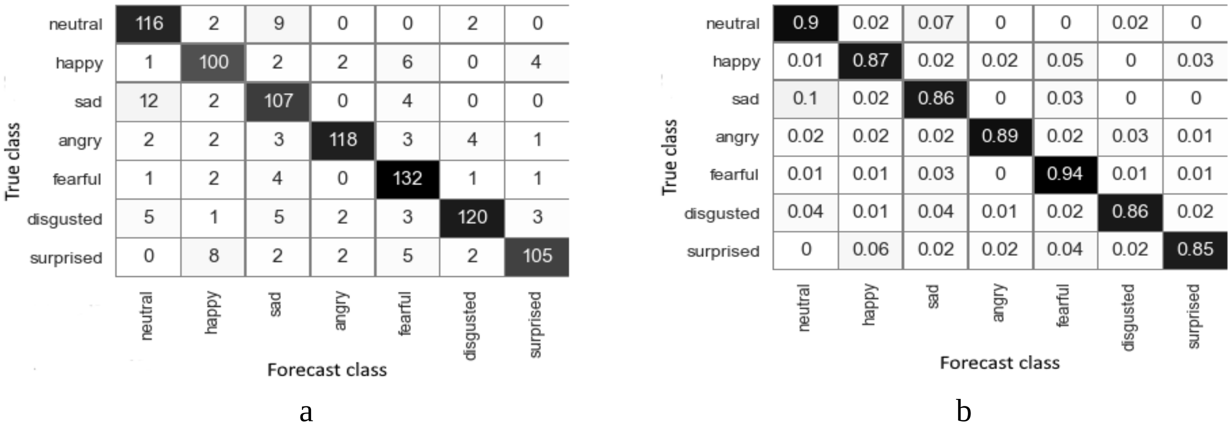
It is convenient to represent the classification results in the form of an error matrix $M = \left\{n_{i,j}\right\}_{i,j=1}^{C}$, which shows the number of objects belonging to the class c_i, but classified by the classifier to c_j. Figure 3.5 illustrates the process of forming the error matrix.

		Forecast class			
		1	2	…	C
True class	1	n_{11}	n_{12}	...	n_{1C}
	2	n_{21}	n_{22}	...	n_{2C}
	…	...	...	...	...
	C	n_C 1	n_C 2	...	n_C C

Figure 3.5 – Structure of the error matrix

In figure 3.5 the diagonal elements of the matrix n_{11}, n_{22}, …, n_{CC} correspond to a successful classification, and off-diagonal ones correspond to an erroneous.

In accordance with this, Figure 3.6 shows the calculated error matrix for the test subsample, when used as a DCNN classifier, trained on informative features in the form of mel-frequency cepstral coefficients (*MFCC*).



True class	neutral	happy	sad	angry	fearful	disgusted	surprised
neutral	116	2	9	0	0	2	0
happy	1	100	2	2	6	0	4
sad	12	2	107	0	4	0	0
angry	2	2	3	118	3	4	1
fearful	1	2	4	0	132	1	1
disgusted	5	1	5	2	3	120	3
surprised	0	8	2	2	5	2	105

Forecast class

a

True class	neutral	happy	sad	angry	fearful	disgusted	surprised
neutral	0.9	0.02	0.07	0	0	0.02	0
happy	0.01	0.87	0.02	0.02	0.05	0	0.03
sad	0.1	0.02	0.86	0	0.03	0	0
angry	0.02	0.02	0.02	0.89	0.02	0.03	0.01
fearful	0.01	0.01	0.03	0	0.94	0.01	0.01
disgusted	0.04	0.01	0.04	0.01	0.02	0.86	0.02
surprised	0	0.06	0.02	0.02	0.04	0.02	0.85

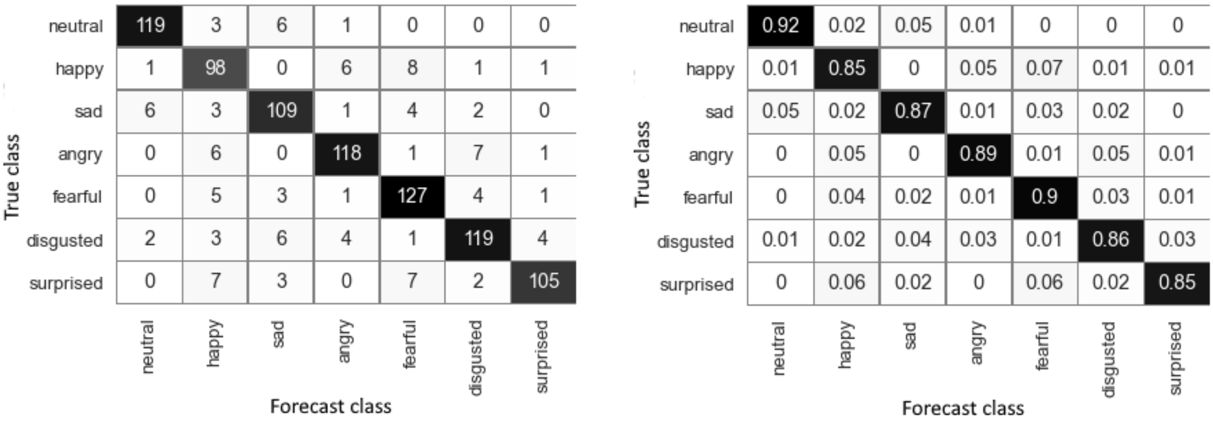
Forecast class

b

a – error matrix; b – the matrix of errors recalculated in fractions

Figure 3.6 – Matrix of errors for *MFCC*

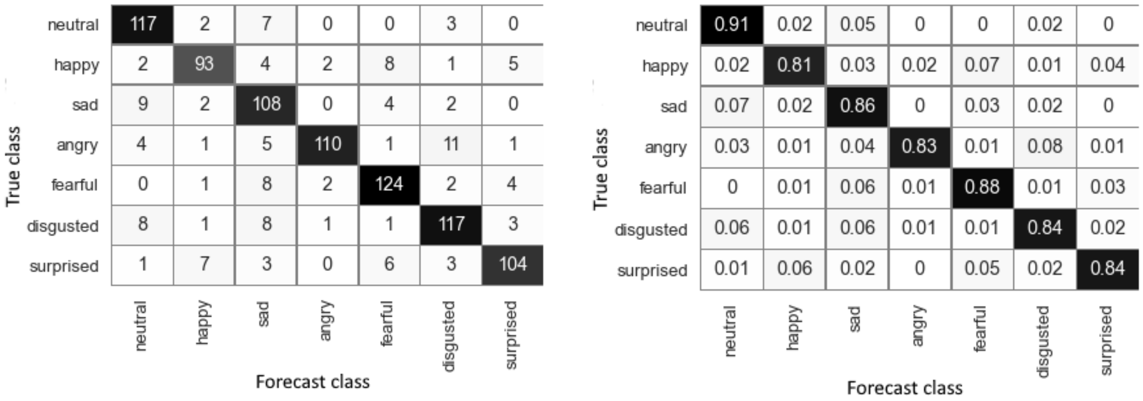
Figures 3.7, 3.8, and 3.9 show similar error matrices for mel-spectrograms (*melspec*), differential parameters of mel-frequency cepstral coefficients (*delta*) and pitch classes (*chroma*), respectively.



a b

a – error matrix; b – the matrix of errors recalculated in fractions

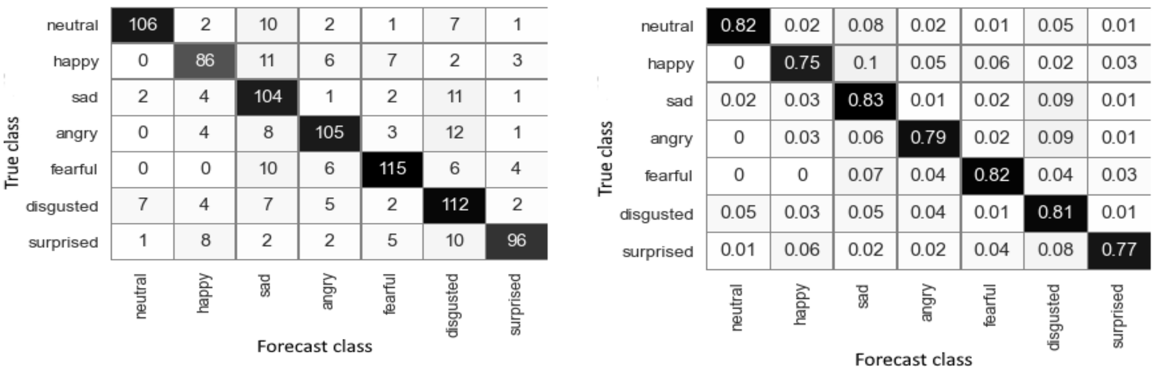
Figure 3.7 – Matrix of errors for *melspec*



a b

a - error matrix; b - the matrix of errors recalculated in fractions

Figure 3.8 – Error matrix for *delta*



	neutral	happy	sad	angry	fearful	disgusted	surprised
neutral	106	2	10	2	1	7	1
happy	0	86	11	6	7	2	3
sad	2	4	104	1	2	11	1
angry	0	4	8	105	3	12	1
fearful	0	0	10	6	115	6	4
disgusted	7	4	7	5	2	112	2
surprised	1	8	2	2	5	10	96

a

	neutral	happy	sad	angry	fearful	disgusted	surprised
neutral	0.82	0.02	0.08	0.02	0.01	0.05	0.01
happy	0	0.75	0.1	0.05	0.06	0.02	0.03
sad	0.02	0.03	0.83	0.01	0.02	0.09	0.01
angry	0	0.03	0.06	0.79	0.02	0.09	0.01
fearful	0	0	0.07	0.04	0.82	0.04	0.03
disgusted	0.05	0.03	0.05	0.04	0.01	0.81	0.01
surprised	0.01	0.06	0.02	0.02	0.04	0.08	0.77

b

a – error matrix; b – the matrix of errors recalculated in fractions

Figure 3.9 – Error matrix for *chroma*

For correct comparison of the classification results presented in the form of error matrices in figures 3.6, 3.7, 3.8, 3.9, we use the metrics of *precision, recall, F-measure* [84], as well as the proportion of correct answers *accuracy* (3.10).

In the case of a binary classification, accuracy is understood as a metric that represents the predictive value of positive results, which is calculated from the expression:

$$pre = \frac{tp}{tp + fp}. \tag{3.12}$$

From (3.12) it follows that accuracy is the proportion of truly positive samples from the total number of predicted positive samples.

The completeness metric can be viewed as the percentage of successful responses, which is calculated as follows:

$$rec = \frac{tp}{tp + fn}.. \tag{3.13}$$

That is, completeness is the proportion of truly positive samples out of the total number of truly positive samples.

To combine the metrics of accuracy and completeness into a general criterion of classification quality, their harmonic mean is used in the form of an *F*-of measure *F*1:

$$F1 = 2\frac{pre \cdot rec}{pre + rec}. \tag{3.14}$$

In the case of multi-class prediction in the specified metrics, the class of interest is taken as a positive class, and all others are considered as negative.

The use of these types of metrics to evaluate the models of classifiers will also take into account the existence of some imbalance between the classes of objects.

In accordance with the above, Table 3.2 presents the calculated quality metrics for the developed DCNN, this was trained on four different types of features.

Table 3.2 – Quality metrics of the DCNN classifier model

Metric type	*melspec*	*MFCC*	*delta*	*chroma*	The number of samples in the test subsample
Multi-class *acc*	0,8775	0,8808	0,8532	0,7991	906
Average for classes *pre*	0,8785	0,8825	0,8575	0,8088	906
Weighted average *pre*	0,8796	0,8836	0,8579	0,8086	906
Average for classes *rec*	0,8768	0,8798	0,8524	0,7981	906
Weighted average *rec*	0,8775	0,8808	0,8532	0,7991	906
Average for classes *F*1	0,8770	0,8802	0,8532	0,8008	906
Weighted average *F*1	0,8779	0,8812	0,8539	0,8012	906

In table 3.2, the class-average value of the metric is calculated by taking the arithmetic mean of this metric for each of the seven classes of emotional state. The weighted average takes into account the presence of some class imbalance (see table 3.1) by multiplying the metric of each class by the appropriate weighting factor when calculating the arithmetic mean.

3.4 Construction of the final model of the PES classifier for the recognition system of PES based on the speech signal of the aviation personnel

The data in table 3.2 show that for all classification quality metrics, the best result is achieved using the DCNN model trained on features in the form of mel-frequency cepstral coefficients (*MFCC*). At the same time, comparable results are obtained when training a neural network on the basis of mel-spectrograms (*melspec*). Less effective was the use of the chalk-frequency cepstral coefficients (*delta*) and pitch classes (*chroma*) as informative features. The corresponding error matrices in Figures 3.8 and 3.9 for any class are inferior in terms of the number of correctly classified samples to the characteristics *MFCC* and *melspec*.

In turn, a detailed analysis of the error matrices in figures 3.6 and 3.7 shows that for the "neutral" and "sad" classes, the classification quality when using the *melspec* features is superior to the model trained on the *MFCC*, although the latter shows higher results on average across classes. Then it is possible to improve the quality of the classification if we use two DCNN models trained on the *MFCC* and *melspec* signs in the process of predicting the class of the speaker's emotional state. That is, the decision about the belonging of the analyzed sample to a specific class will be made on the basis of forecasts from two models.

The diagram in figure 3.10 illustrates the process of using two DCNNs in the classification process. At the output of each neural network, a vector of probabilities (*P*(*Y*)$_1$, *P*(*Y*)$_2$ is formed in figure 3.10) belonging of sample *s* to one of the *C* = 7 classes.

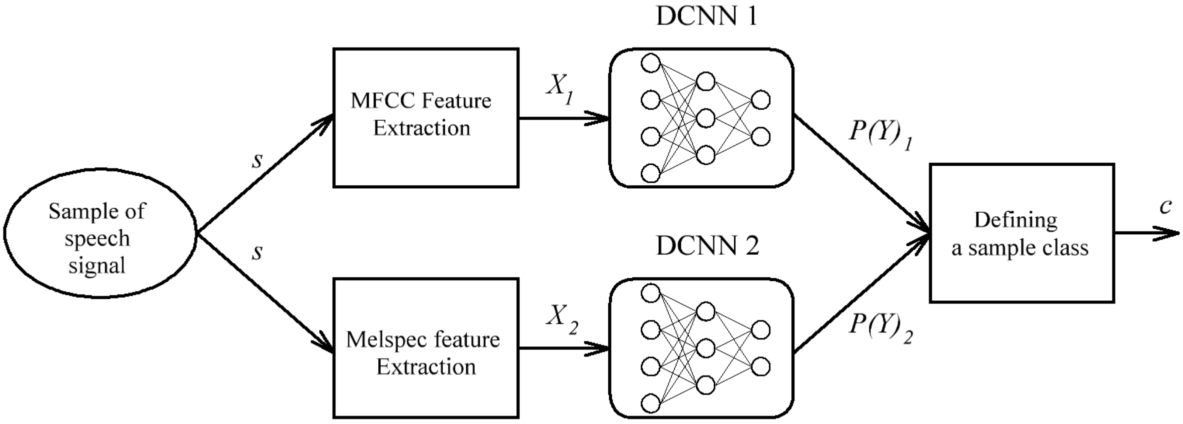


Figure 3.10 – Using two DCNNs for classification

The strategy for the final determination of the class according to the data *P (Y) 1*, *P (Y) 2* can be different.

1. You can calculate the arithmetic mean of the obtained probabilities:

$$c = \sum_{i=1}^{C} I\left[y_i = \arg\max(\frac{P(Y)_1 + P(Y)_2}{2}) \right] \cdot i. \tag{3.15}$$

2. Take as a positive class the one with the highest probability among both neural network predictions:

$$c = \sum_{i=1}^{C} I\left[y_i = \arg\max(P(Y)_1, P(Y)_2) \right] \cdot i. \tag{3.16}$$

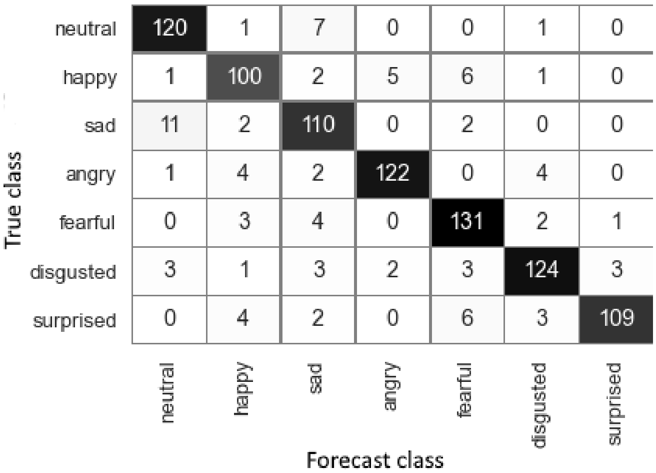
Table 3.3 presents, for comparison, the classification quality metrics using the two described strategies.

Table 3.3 – Metrics of classification quality using two DCNN models

Metric type	Strategy №1 (3.15)	Strategy №2 (3.16)
Multi-class *acc*	0,9007	0,8962
Average for classes *pre*	0,9017	0,8966
Weighted average *pre*	0,9023	0,8974
Average for classes *rec*	0,8996	0,8950
Weighted average *rec*	0,9007	0,8962
Average for classes *F*1	0,9001	0,8952
Weighted average *F*1	0,9009	0,8962

Thus, the data in table 3.3 show that the use of two DCNNs trained on the *MFCC* and *melspec* features for the classification of PES contributes to a tangible improvement in the quality of the classification (in comparison with the data in table 3.2). At the same time, for the formation of the final answer, the use of strategy No. 1 (3.15) when combining the probability vectors of the current sample belonging to each of the seven classes is more preferable from the point of view of the obtained quality metrics of classification.

Figure 3.11 shows the error matrix for the final classifier of a person's emotional state based on a speech signal. This classifier was obtained using two DCNN models trained on the *MFCC* and *melspec* features. In this case, the probability of belonging of the desired sample to each of the classes is calculated as the arithmetic mean for the obtained probabilities from each neural network (3.15).



True class
Forecast class

	neutral	happy	sad	angry	fearful	disgusted	surprised
neutral	120	1	7	0	0	1	0
happy	1	100	2	5	6	1	0
sad	11	2	110	0	2	0	0
angry	1	4	2	122	0	4	0
fearful	0	3	4	0	131	2	1
disgusted	3	1	3	2	3	124	3
surprised	0	4	2	0	6	3	109

Figure 3.11 – Matrix of errors of the final model of the classifier

Table 3.4 shows the quality metrics of classification calculated by the classes of emotional states for the resulting final model of the classifier.

Table 3.4 – Quality metrics of classification by class of samples

Emotion type	Accuracy, *pre*	Completeness, *rec*	Measure *F*1	The number of samples in the test subsample
Neutral	0,8824	0,9302	0,9057	129
Happy	0,8696	0,8696	0,8696	115
Sad	0,8462	0,8800	0,8627	125
Angry	0,9457	0,9173	0,9313	133
Fearful	0,8851	0,9291	0,9066	141
Disgusted	0,9185	0,8921	0,9051	139
Surprised	0,9646	0,8790	0,9198	124

3.5 Analysis of the effectiveness of the developed method for classifying the emotional state by the speech signal

3.5.1 Comparison of the developed classification method with other machine learning algorithms

For a comparative assessment of the proposed classification method based on the use of two DCNNs, let us consider the performance of other CAL methods that have proven themselves well in practice. For this, the following algorithms were investigated in the work:

- fully connected neural network [75, p. 117];
- logistic regression [44, p. 234];
- random forest [85];
- gradient boosting [86].

For these models, a vector of informative features *MFCC*, *melspec*, *delta*, *chroma* was used as input data. However, the values of the feature coefficients were averaged over the number of frames in the audio sample. As a result, the vector of features of objects becomes one-dimensional and contains 218 elements in its composition (39 coefficients of *MFCC*, 128 *melspec*, 39 *delta*, 12 *chroma*).

In the process of searching for the optimal parameters of these models, the following configurations of the CAL algorithms were found.

The structure of a fully connected neural network consists of four fully connected layers: an input layer with 218 neurons in accordance with the dimension of the input vector, 2 hidden by 512 neurons in each, and an output layer with seven neurons according to the number of classes. The output layer is a softmax classifier (3.4). Regularization layers are located between fully connected layers: in the first, every fourth block of input data is discarded, in the second and third, every second block. The nonlinearity ReLU (3.2) is used as the activation function of neurons. As an optimization method Adam algorithm is used. The categorical cross-entropy (3.8) acts as an error function. The proportion of correct answers *acc* is taken as a quality metric when training a network on a training subset (3.10).

In the process of selecting the hyperparameters of algorithms for logistic regression, the degree of regularization was taken to be L_2 = 10. For the random forest model, the number of trees is set equal to *n_estimators* = 300. When implementing the gradient boosting algorithm, the number of trees was chosen equal to *n_estimators* = 200.

For fully connected neural network and logistic regression models, a standardized estimate is applied to the training data:

$$z_x = \frac{x - \mu_x}{\sigma_x}, \tag{3.17}$$

where x – the element of vector of informative features;

μ_x – the average of this element over all objects in the subsample;

σ_x – standard deviation of a given element for all objects in the subsample.

The described models were implemented and trained in Python 3.7 using the machine learning libraries Scikit-learn 0.23.1 [87] and Keras 2.3.1.

Table 3.5 shows the results of comparison of the proposed classifier model, based on DCNN, with other types of considered CAL models.

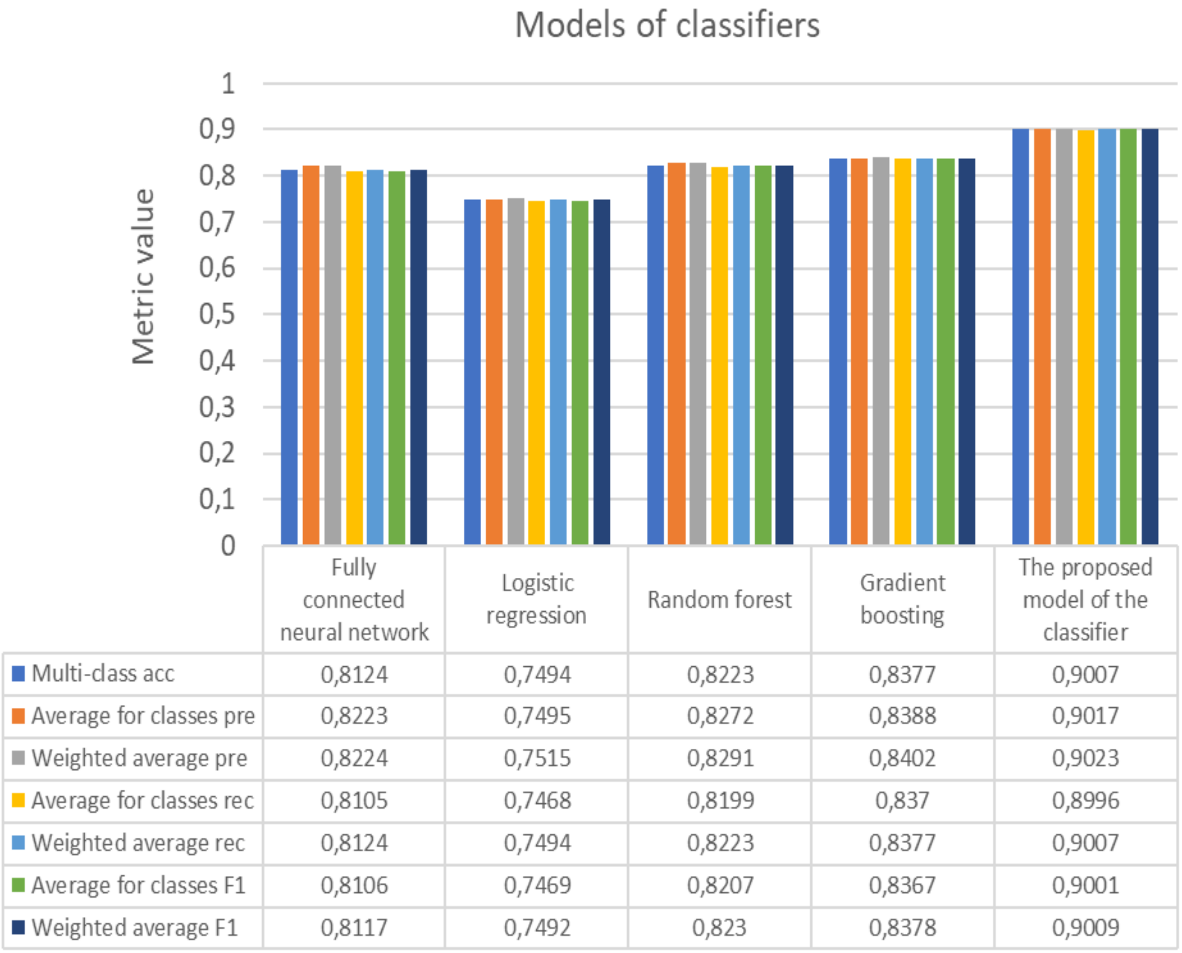
Table 3.5 – Results of comparison of the proposed method of the classification of PES by speech signal with other types of CAL models

Metric type	Fully connected neural network	Logistic regression	Random forest	Gradient boosting	The proposed model of the classifier	The number of samples in the test subsample
Multi-class *acc*	0,8124	0,7494	0,8223	0,8377	0,9007	906
Average for classes *pre*	0,8223	0,7495	0,8272	0,8388	0,9017	906
Weighted average *pre*	0,8224	0,7515	0,8291	0,8402	0,9023	906
Average for classes *rec*	0,8105	0,7468	0,8199	0,8370	0,8996	906
Weighted average *rec*	0,8124	0,7494	0,8223	0,8377	0,9007	906
Average for classes *F1*	0,8106	0,7469	0,8207	0,8367	0,9001	906
Weighted average *F1*	0,8117	0,7492	0,8230	0,8378	0,9009	906

Figure 3.12 presents the data from table 3.5 in the form of histograms.

As follows from the data obtained from the results of comparing the classifier models (table 3.5, figure 3.12), the proposed method for predicting the class of the speaker's emotional state by voice is superior to the rest of the considered CAL algorithms in all accepted types of metrics. The use of two DCNNs in the classifier model makes it possible to achieve 90% of accuracy on the test sample. The small scatter of parameters by types of metrics for the proposed classifier indicates the adequate operation of the model on seven accepted types of PES of a person. The developed classification method outperforms such effective CAL algorithms as gradient boosting and random forest in terms of performance.

The obtained results indicate a correctly chosen approach in the design and selection of informative features. The proposed method for detecting PES from a speech signal avoids the need to recognize said phrases in the analysis process, which greatly simplifies the classification procedure. The model uses only acoustic data for prediction. At the output of the model, the probabilities of the sample belonging to each of the seven classes are generated. On the basis of this, it is possible to build a fuzzy logic of the operation of automatic systems for monitoring the state of a person.



	Fully connected neural network	Logistic regression	Random forest	Gradient boosting	The proposed model of the classifier
■ Multi-class acc	0,8124	0,7494	0,8223	0,8377	0,9007
■ Average for classes pre	0,8223	0,7495	0,8272	0,8388	0,9017
■ Weighted average pre	0,8224	0,7515	0,8291	0,8402	0,9023
■ Average for classes rec	0,8105	0,7468	0,8199	0,837	0,8996
■ Weighted average rec	0,8124	0,7494	0,8223	0,8377	0,9007
■ Average for classes F1	0,8106	0,7469	0,8207	0,8367	0,9001
■ Weighted average F1	0,8117	0,7492	0,823	0,8378	0,9009

Figure 3.12 – Results of comparison of the proposed model of the classifier of the emotional state by the speech signal with other algorithms of the CAL

3.5.2 Comparison of the proposed classification model with other studies in the field

To compare the proposed method of the recognition of PES by speech signal with the research results obtained by other authors, the available sources of information in this area were analyzed. First of all, it was found that a large number of works are devoted to solving the problem of recognition of PES based on complex information about acoustic and linguistic speech data. This approach requires the existence of a special effective language model [88], which in turn significantly complicates the classification process. Moreover, it can be expected that when using aviation English with specific phraseology of radio communication, the existing language models will be ineffective. In this regard, in order to compare the research results, the classification quality metrics obtained only from the acoustic data of the speech signal were analyzed.

In addition, significant difficulties in comparing research results arise due to the use by the authors of different databases in different languages and with a different number of types of allocated PES.

In accordance with this, table 3.6 presents the results of the analysis performed comparing the quality of the classification obtained in this work, and in studies with the closest characteristics of the data used and the requirements for the results.

Table 3.6 – The comparative analysis of the quality of the classification of the emotional state by the speech signal, obtained in this work and the similar studies

A source	Classification methods	Database	Quality metrics in %
B. Schulleretal	Gaussian Mix Model (GMM), k-Means, Support Vector Machine (SVM), Multilayer Perceptron (MLP)	B. Schuller et al.	Share of correct answers *acc* = 74,2%
C. Lee, S. Narayanan	Evaluation of the emotional weight	Private database	Classification error *err* = 1 – *acc*. *err* = 17,85% – 25,45% for men; *err* = 12,04% – 24,25% for women.
H. Goetal	Wavelet analysis, linear discriminant analysis	Private database	Accuracy *pre* = 57% – 93,3% for men; *pre* = 68% – 93,3% for women.
M.M.H. El Ayadi et al	Gaussian mixture of vector autoregressive model (GMVAR)	F. Burkhardt at al	*acc* = 76%
B. Schuller et al	Hidden Markov Model (HMM)	Private database	*acc* = 86,8%
Javier Getal	MLP, decision trees	O. Martin et al	*acc* = 96,97%
This work	DCNN	RAVDESS; SAVEE; TESS	*acc* = 90,07%, *pre* = 90.17%
Note – Compiled from sources [89-97]			

The data in Table 3.6 show that the method of proposed classification outperforms most of the known models of detection the PES from a speech signal. Moreover, in [96, p. 24], the share of correct answers is 96.97%, which is 6.9% higher than the results of this study. However, it should be noted that the classification by the base [97, p. 1 - 8] in the work [96, p. 20-27] was produced only for 6 types of PES without determining the neutral state. Also, in the work [96, p. 21], 264 samples of audio signals extracted from video recordings were used for the study, with one utterance for each emotion. For these reasons, it can be assumed that there is insufficient generalizing ability of those proposed in the study [96, p. 20-27] of classification algorithms.

In turn, the developed classifier based on DCNN was trained immediately on data from three different emotional corpuses (table 3.6), which significantly increases the generalizing ability of the final model.

Thus, based on the comparative analysis of the developed model of the speaker-independent classifier of the emotional state of a person based on his speech signal with other intelligent algorithms of CAL and proposed methods in the works of other researchers, it can be argued that the use of two DCNNs trained on the signs of mel-frequency cepstral coefficients and mel-spectrograms, is an effective solution. Moreover, the proposed classification method makes it possible to obtain a high quality of automatic detection of PES only from the acoustic data of the speech signal.

Conclusions on the third section

Modern technologies of data mining make it possible to achieve high quality results in the tasks of automatically extracting useful information from various kinds of features of the objects under study. The use of deep learning technologies in the form of artificial convolutional neural networks opens up new possibilities for analyzing data of a two-dimensional structure. In particular, informative features of a speech signal have such a dimension when performing its short-term analysis to obtain mel-spectrograms, mel-frequency cepstral coefficients, and differential parameters of chalk-frequency cepstral coefficients and pitch classes.

The proposed DCNN architecture and the algorithm for its training on the selected informative features allow one to obtain high results in the classification of the emotional state of a person for seven classes of objects only on the basis of the acoustic data of the studied samples. The classifier model based on DCNN of the proposed architecture allows obtaining the best results of classification when training it on informative features in the form of mel-frequency cepstral coefficients. In this case, the result of the classification is considered as independent of the speaker, since data from three different emotional corpuses are used to train the neural network.

To improve the parameters of classification of PES, a method is proposed that combines the classification results from two DCNNs trained on different types of informative features: mel-spectrograms and mel-frequency cepstral coefficients. As a result, the result of classification of PES is formed in the form of the average value of the probabilities of belonging of the studied sample to each of the seven classes of PES predicted by each neural network. With this approach to solving the problem of classification of PES, it is possible to achieve a multiclass fraction of correct answers equal to 0.9007 on a deferred test subsample.

During the analysis of the results obtained, it was found that the calculated indicators of the classification quality according to the proposed method are superior to the results for other effective CAL algorithms, such as a random forest, a fully connected neural network, gradient boosting, etc. An analysis of sources based on similar studies also shows that when using only acoustic information of a speech signal to recognize seven types of PES, the proposed method surpasses the existing models in terms of quality metrics.

4 METHODOLOGY FOR REDUCING THE IMPACT OF THE HUMAN FACTOR ON FLIGHT SAFETY BASED ON INTELLECTUAL ANALYSIS OF THE SPEECH SIGNAL

4.1 Proficiency in aviation and general English as a human factor, affecting flight safety

To achieve the goal of the dissertation, we will conduct analytical studies to determine the impact of HF, based on insufficient knowledge of aviation and general English on flight safety, having considered the following issues:

– the role of language in the world practice of aviation accidents and incidents;

– to systematize the scientific and practical foundations of the formation of phraseology of radio communication;

– for in-depth study of the impact of HF of engineering and technical personnel and flight attendants, who have a key role in ensuring flight safety, systematize the scientific and practical foundations for the formation of standard phrases;

– to explore aspects of ensuring effective communication in aviation;

– to determine the conditions for the formation of stable phraseological units for aviation personnel, for whom English is a non-native language.

In 2004, the International Civil Aviation Organization (ICAO) put forward a number of advisory language proficiency requirements for aviation personnel, namely for pilots and controllers working on international airlines. This provision was intended to avoid the language problems by openly demonstrating the basic level of English proficiency (operational level 4): aviation and general (plain). But the main goal of this innovation is to reduce the impact of HF on insufficient language proficiency on flight safety, i.e. avoid language problems during “air-to-ground” communications and emphasize the critical importance of English as a prerequisite for doing this job reliably.

ICAO is an entity of the United Nations Organization and has no regulatory oversight but only an advisory function [98]. Despite the fact that English was recognized as the language of international communication, until 2008 (for some countries until 2011) there were no regulatory requirements for training and testing. Moreover, ICAO did not mandate the use of English internationally for ATC communications, but only recommended communication in the language of “commonly used ground station”, which in many cases resulted in an inability to understand the message by part of each crew in the airspace.

Over the years, ICAO has documented and investigated a large number of incidents and accidents both in flight and on the ground, ranging from severe accidents to minor incidents. In these cases, miscommunication was known to be the key, fatal factor. According to Feldman, misunderstanding was a concomitant factor in 70% of cases. The literature [99] regularly mentions the fact that ICAO reported 7 major, significant accidents from 1976 to 2001, which were a direct result of misunderstandings and resulted in ICAO’s decision to strengthen measures related to the introduction of language training.

However, from the inception of the ICAO to the present day, there have been many more accidents and incidents. LOT Airlines: the plane was lost over London on June 4, 2007, the pilots of the Boeing-737-500, SP-LKA of the Polish airlines “LOT”, en route from London to Warsaw, noticed that the information on the two main navigation devices disappeared due to an error in entering their navigation location. This situation forced them to enter instrumental meteorological conditions (IMC), which means they were forced to use onboard instruments because they did not have any landmarks. The co-pilot controlled the aircraft using backup instruments, while the co-pilot tried to solve the problem. In this situation, the pilots were forced to ask the control tower for directions to the airport, which meant that they had to rely entirely on the instructions of the air traffic controllers, which, of course, were given in English. The aircraft flew around the surrounding airspace for almost half an hour because the two pilots could not understand the instructions of ATC.

According to a report prepared by the British AAIB, a series of conversations took place between the pilots and the controller during the flight, in which it was obvious that the commander making the radio calls could not understand some of the instructions [100].

Fortunately, they managed to land safely and no one was hurt, but the risk they were exposed to during this unfortunate incident was great. First, they could cause a mid-air collision with another aircraft that was flying at the same altitude, but managed to maintain separation thanks to revised tower instructions, the report said. Second, a runway incursion can be caused by a misunderstanding of the correct alignment with the runway.

Other fatal accidents:

– 1976 is a mid-air collision near Zagreb, caused by a sudden switch to Serbo-Croatian at a critical moment by one of the pilots aboard the Inex Adria aircraft and 176 deaths;

– 1977 is one of the worst plane crashes in aviation history occurred in Tenerife due to the misuse of a single grammatical element; Dutch-speaking pilot lacked knowledge of English, which led to 583 fatalities [101, p. 22.1-22.4];

– 1990 is the Colombian crew aboard the Avianca 52 aircraft approaching New York failed to declare a "fuel emergency" and crashed because they were not granted priority landing; 73 people died [100].

Unfortunately, there are (too) many other accidents to be mentioned. There is a consensus in the literature that failures are the result of many factors, and language proficiency is just one of many attendant problems [100]. Nevertheless, it remains the most important component of the “air-ground station” exchanges.

4.1.1 Scientific and practical bases for the formation of the phraseology of AEL

The purpose of phraseology is to provide a clear, concise, unambiguous language for conveying messages of a day-to-day nature [102].

The scope of application of aviation English is quite wide, since it includes all types of distribution (spoken, radio communication, etc.) of the languages used by aviation personnel as pilots and controllers. However, ICAO began to pay special attention to expanding the scope of AEL for engineering and technical personnel, air and ground service specialists, commercial personnel, etc.

Standard radiotelephone phraseology is a semi-artificial sublanguage used by pilots and controllers to conduct standard, general, and conventional radiotelephony communications (RTC).

During normal flights, pilots and air traffic controllers adhere to the RTC phraseology, which ICAO defines as “a template code that includes certain words that, in the context of aviation operations, have precise and exclusive operational significance”. According to Phillips [103], the main features of the RTC phraseology are a number of documents.

There are differences between plain English and radio communication phraseology (RTC) at all linguistic levels. Let’s consider the representation of the various levels in terms described in the Aitchison wheel diagram [104] shown in Figure 4.1, where the “inner circles” represent the core disciplines such as phonology, vocabulary, syntax, and semantics. It follows from this diagram that communication in an aviation context is not just a “coherent two-level lexicography” (phraseology and general English), but the interaction of three language levels.

The phonology of the AEL includes the specific pronunciation of numbers (table 4.1) and letters (table 4.2).

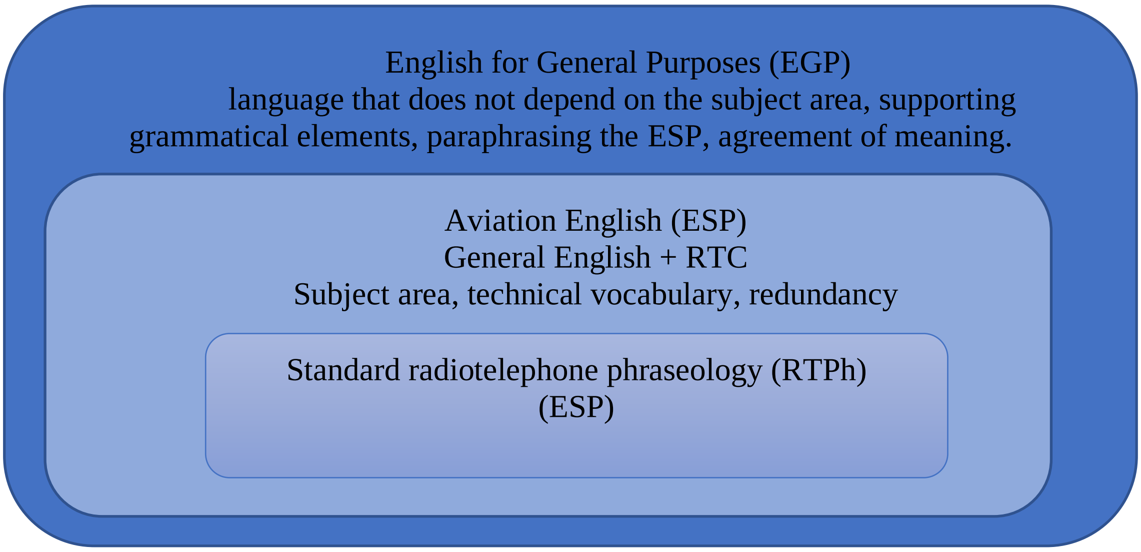

English for General Purposes (EGP)
language that does not depend on the subject area, supporting grammatical elements, paraphrasing the ESP, agreement of meaning.
Aviation English (ESP)
General English + RTC
Subject area, technical vocabulary, redundancy
Standard radiotelephone phraseology (RTPh)
(ESP)

Figure 4.1 – Interaction of three language levels

Table 4.1 – Pronunciation of numerals in aviation

1 – One	WUN	6 – six	SIX
2 – two	TOO	7 – seven	SEVEN
3 – three	TREE	8 – eight	AIT
4 – four	FOW-ER	9 – nine	NINER
5 – five	FIFE	0 – Zero	ZEE-RO

Table 4.2 – Phonetic alphabet

A - Alfa	N – November
B – Bravo	O – Oscar
C – Charlie	P – Papa
D – Delta	Q – Quebec
E – Echo	R – Romeo
F – Foxtrot	S – Sierra
G – Golf	T- Tango
H – Hotel	V – Victor
I – India	U – Uniform
J – Juliet	W – Whiskey
K – Kilo	X – Xray
L – Lima	Y – Yankee
M – Mike	Z – Zulu

Pilots and controllers pronounce words differently to avoid misunderstandings. Moreover, they also expand the words to make them clearer and more specific to other pilots.

An example of radiotelephone communication is:

1. *Tower, this is YANKEE ZULU ONE NINER NINER TANGO (YZ199T), ready for takeoff.*

2. *Roger, YANKEE ZULU ONE NINER NINER TANGO. Keep on the runway ONE NINER RIGHT (19 Right).*

The vocabulary of AEL is based on phraseology, which uses “certain expressions to convey meaning” that are significantly different from those used in plain English: for example, confirm (yes; that's right), negative (no), roger (confirm), request (I would like to, can I ..?), repeat (could you repeat, please?).

The syntax of AEL is very simplified, which leads to a reduction in vocabulary (about 400 words), very short sentences with the removal of official words and prepositions (climb 150; cf. climb to [flight level] 150), aids and the topic of the pronoun (soon you will lose radar contact; cf. you will soon lose radar contact) followed by the production of nominal sentences, usually imperative or passive [105].

4.1.2 Scientific and practical foundations of the formation of the PEL

Simple, in some literature referred to as a common language - English, on the other hand, is used in all those situations when the phraseology of radio communication is not enough, because it does not provide ready-made forms for communication. In particular, in the case of an unexpected turn of events, when pilots and controllers must use natural, accessible language.

ICAO gives a clear definition of a simple language, considering it the best tool for interacting with people in unforeseen situations: “Linguistic research now clearly shows that there is no more suitable form of speech for human communication than natural language. Human language is characterized in part by its ability to create new meanings and use words in new contexts. This creative function of language is especially useful for accommodating the complex and unpredictable nature of human interaction, including in the context of aviation communication. There is simply no more suitable form of speech for human communication than natural language."

Examples of possible emergencies in which plain language needs to be used are technical problems on board, ATC equipment failure, warnings, etc.

ICAO very clearly describes the use of a simple language: "*First of all, you should always use standard phraseology of ICAO"*.

ICAO standard phraseology should be used in all situations for which it is provided. Only in cases where standard phraseology cannot serve the intended transmission, the plain language should be used.

However, ICAO uses terminology very clearly and refers to the language to be used when phraseology is not enough as a simple language. In fact, ICAO defines it as follows:

"Plain language in aviation RTC means the spontaneous, creative and uncoded use of a given natural language, albeit limited to features and topics (aviation-related and not), as required for aerial radiotelephony communications, as well as specific safety critical requirements for legibility, straightforwardness, relevance, unambiguousness and conciseness".

As follows from the above, it is possible to define the boundaries of AEL and PEL: characteristics of PEL include a wider range of vocabulary outside the field of aviation (medicine, information technology, military organizations, etc.) and the ability to construct more complex sentences grammatically compared to AEL

The ICAO Guidelines (2010) emphasize the importance of "free, clear, concise and unambiguous language". In light of what has already been said about plain language, it is extremely important to emphasize the importance of language proficiency among learners, test takers and, most importantly, users.

The importance of a high level of English becomes obvious, especially during urgent or emergency incidents, when phraseology does not cover the corresponding needs, and a low level of language proficiency can become another obstacle to successfully doing your job.

According to Emery [106], plain language should always focus on possible non-standard, real-life situations in aviation operations. An example of communication for which phraseology is not a standard construction, and the ability to obtain results at an adequate level of language is required, is the following excerpt from the communication between the air traffic controller and the pilot (ICAO, 2010):

Air traffic controller: Will you let me know your intentions for the main landing gear?

Pilot: WT, "will comply": We'll try to turn off the landing gear again, and if it stays on and I can't let go of the front landing gear, we'll land with all three up.

Air traffic controller: Roger. So, if you wish, you can go around and visually inspect the landing gear.

Pilot: Ok. Roger

Air traffic controller: WT, Do you see the field?

Pilot: WT, Affirm.

Air traffic controller: Roger. You ... you fly over the field and make a low pass over runway 29 to check the landing gear.

This passage clearly shows that phraseology is not sufficient to adequately express the intentions of the air traffic controller and pilot. The effective transition from standard phraseology to simple language is what ICAO calls “code switching” and is a critical component of ICAO's language proficiency requirements.

Figure 4.1 clearly illustrates the hierarchical relationship described above: phraseology as a subgroup of Aviation English, which in turn is a subgroup of the PEL. A good level of PEL will provide an opportunity to skillfully study aviation English, like English for special purposes (ESP) and, as a result, standard radiotelephone phraseology.

Uplinger [107] argues that mastering specialized terms is easier when other aspects of the language, such as sentence structure and word formation, are first mastered. However, “English in international aviation should not be regarded as English for General Purposes” [108], in other words, it should not be regarded as General English. In conclusion, it should be noted that the high level of English proficiency, combined with a good command of phraseology is essential for the ensuring safety during flight operations.

4.1.3 Aspects of Effective Communication in Aviation

According to the Aviation Safety Foundation [109] “Until data link communication becomes widespread, air traffic control (ATC) will primarily rely on voice communications”. However, there are various linguistic factors involved, and achieving effective radio communications involves many of them. Let’s consider aspects that can contribute to effective communication in aviation.

Prosody - effective radio communication in aviation English is not only a matter of using phraseology correctly or having a high level of English proficiency in terms of vocabulary or grammar construction. In fact, it has been demonstrated that the prosody (intonation, stress and rhythm) may have a positive impact on the comprehension of messaging. According to McMillan [110], speed and lack of pauses during message delivery were the main causes of read back errors, and in one of his studies he states that “the high speed at which air traffic controllers deliver instructions is probably the most common complaint of misunderstandings, received from the pilots”. Moreover, another study by Neville [111] confirms that time, silence and intonation are necessary conditions for the verbal communication, in particular, they play an important role in the cabin, as they help to distinguish “tolerances” from “test questions” (for example, cleared before take-off versus cleared for take-off?). This communication complication is further exacerbated in the context of ESL, and one of the objectives of this investigation is to understand whether the misunderstanding of speech could be the result of misuse of prosodic functions that negatively affect the English language of the non-native speakers.

A direct consequence of the globalization of aviation is the formation of multinational and multicultural crews with different linguistic and cultural origins. Communication in English is much more difficult for the flight crew, whose native language is non-English, especially when they are working under pressure, for example, in case of emergency [112]. According to the Aviation Safety Foundation, studies of crew resource management (CRM) have shown that “language differences in the cockpit are a major barrier for safety”. The problem is connected not only with ELS pilots, who must demonstrate a high level of English proficiency in accordance with ICAO rules, but also with native English speakers who may not understand certain types of exchanges due to the many regional accents and dialects. The methodology used to teach aviation English should include a set of listening comprehension tasks that will help students [113] improve their ability to understand many of the foreign accents that can be found in the cockpit and air traffic control room, as well as their ability to use strategies in explanations to overcome possible problems associated with the inability to understand them. This is another aspect analyzed in the survey, proposed to pilots, and interviews conducted with aviation English instructors at some flight training organizations in Kazakhstan.

There are a number of high-quality data-driven articles on the topic of cognitive load that affects language proficiency. To highlight one of them, Farris and others [114] conducted several experiments to measure the level of ESL proficiency for pilots operating at different levels of workload, given the focus of this thesis of a non-psycholinguistic nature. The results of these experiments show that message repetition (reading back) is less accurate and the first language accent is broader in candidates with low ESL proficiency. They also demonstrate that short messages are more likely to be understood than longer messages. Generally speaking, an increase in cognitive load during radiotelephone communication increases language misunderstandings leading to error.

Communication problems can occur due to poor radio transmission, resulting in deterioration of the speech signal. Not only that, but the cockpit is a very noisy environment that makes everyone, ESL pilots in particular, have difficulty in understanding radio communications. This is why, according to Mell [115], communication should be “smooth and easy” and resolved as quickly as possible.

Communication between air and ground operators is generally very predictable because the air traffic controller usually receives the details of the flight plan in advance. Thus, each message along the route appears in a predetermined sequence [116]. However, even in these limited circumstances, misunderstandings do occur as a result of a number of factors, which may include overlapping calls, disrupted radio signal, etc. An important acknowledgment-correction process by which safe and redundant communication can be guaranteed is the “pilot-to-controller” communication loop.

4.1.4 Aviation as Lingua franca

Unlike many other varieties of English for Specific Purposes, Aviation English is a legally prescribed language that is strictly regulated at the international level. It defines and requires the language to be used by aviation personnel.

Estival and Farris [117, 118] have put forward a clear distinction between aviation English and English as a common language (hereinafter - EFL). They argue that aviation English can be viewed as an interlanguage, that is, a stable variety of

English used as the working language around the world, but cannot be considered ELF. The opposite view is suggested by Barbara Seidlhofer [119] who points out that aviation English can be viewed as a language without native speakers as it must also be learned by native English speakers. It is in this context that the original meaning of the Lingua franca would fit perfectly, as it is considered a language used by native speakers who do not share a common language. Moreover, aviation English has such a limited domain that it can only be used in the aviation field.

Since a large number of speakers in this community are not all native speakers, it is more appropriate to think of English as an interlanguage [120] although there is a different view of the Lingua franca that includes native speakers.

In other words, aviation English can be defined as a language for a specific purpose, used by both native speakers (hereinafter referred to as NS) and non-native speakers, who use English in a work context, where English is a second language for many users. However, communicative success should also be understood as the interaction between the listener and the interlocutor, regardless of their nationality [121]. It is from this point of view that while ELF often refers to the common language used by the NNS of English, native speakers are responsible for modulating their speech in order to be understood by an international audience. [122].

The role of both NS and NNS is discussed here in terms of language requirements. Both of them must guarantee clear intelligibility during flight, which is the ultimate goal. In addition, NS status does not necessarily imply a high level of proficiency in aviation English.

As previously described, clear, unambiguous and fluent communication in aviation English is of paramount importance to aviation safety, and all native speakers must take responsibility for achieving this goal as a long-term goal, regardless of their native language.

Kim and Elder [123] believe that the solution lies in the training of the flight crew and air traffic controllers. It is through training that pilots and controllers must learn certain strategies, such as speech simplification and speech rephrasing, when they may be perceived as solutions to misunderstandings. In addition, training should contribute to improving the use of aviation phraseology as required by ICAO. Cookson [101] argues that native speakers should simplify their language by avoiding idiomatic expressions and correcting interactions by adjusting accents when this may undermine the understanding of international interlocutors.

In conclusion, given the status of aviation English as the Lingua franca, it is important to carefully consider the choice of the teaching methodology to improve those linguistic aspects that facilitate the effective communication on an international level.

4.1.5 Language Proficiency Requirements (LPR)

Technical investigations into the most serious accidents and incidents in which language proficiency was considered an unsafe activity in the system led to the need for certification of English proficiency among pilots in the international community in order to cope with unusual situations. As a result, National Aviation Authorities have integrated LPRs into their regulatory frameworks, and course developers have prepared the language training materials that can help pilots and controllers reach ICAO Level 4. In addition, ICAO agreed that this competence should not be limited to the use of English in standard radiotelephony phraseology for air-ground communications, but should be extended to the use of ordinary English. On March 5, 2003, the ICAO Council adopted Amendment 164 to amended ICAO Document 9835, which introduces new language requirements for both flight crew and air traffic controllers. In particular, paragraph 1.2.9.4 of the aforementioned Annex states that the deadline for compliance with ICAO rules by all Member States was March 5, 2008, although some Member States complied with ICAO requirements only in 2011. Commercial pilots, private pilots, helicopter pilots and air traffic controllers must demonstrate the ability to speak and understand the language used in aeronautical communications in accordance with the Language Proficiency Requirements (LPR) and holistic descriptions.

Despite the fact that it was a lot of criticism of the reality of LPR according to ICAO [124], they should be considered the standard requirements that must be used in command of the language in the aviation industry worldwide. According to Emery, ICAO Level 4 is also a mandatory requirement for pilots to enter ab-initio, and he argues that it is important for airlines and flight training providers to provide language training prior to commencing flight practice in accordance with ICAO standards.

It is widely recognized by linguistic experts that standard phraseology cannot fully describe all possible circumstances arising in non-standard situations, and therefore it is necessary to constantly improve language competence [125].

In this case, we cannot disagree with the fact that the English language is of paramount importance for aviation specialists in the performance of their direct functional responsibilities, which in turn guarantees safety at all stages of the flight.

As a result of these investigations, it was found that a high percentage of incidents and accidents really depended on the difficulties faced by pilots and air traffic controllers when decoding messages in plain English, for example, to explain some unexpected circumstances that were not included in phraseology. According to Uplinger "ATC terminology is narrowly specialized and does not occur frequently in the common language, so proficiency in ATC terminology does not provide a functional proficiency in English on its own terms. Mastering the terminology of ATC requires a sufficiently high level of functional training".

The importance of expanding skills and competencies in the English language used in this specific context is that proficiency in phraseology combined with a good command of English will result in less stressful and more effective management of unexpected, exceptional flight situations or its effective use in different contexts, for example, work-related [126].

Literature on the subject "English for Aviation" is divided into two distinct periods of time, and the introductory section represented ICAO declaration in 2003 that all pilots of its Member States should demonstrate a working level of English (level 4 of ICAO language proficiency scale) until March 5, 2008, a deadline that was later extended to 2011 as many Member States found it difficult to comply with this schedule. “Qualification standards, developed by the International Civil Aviation Organization (ICAO), encourage pilots on international routes, air traffic controllers and aviation station operators to speak and understand English at an ‘operational’ level 4 [98]. As shown in paragraph 1.7.1, the need for simple language skills can arise quickly on board when an unforeseen situation arises.

4.2 Methodology for reducing the impact of the human factor on flight safety

Based on the research results presented in the previous sections, the author proposed a new methodology for reducing the impact of the human factor on flight safety based on recognition of PES by the voice signal of aviation personnel, which includes the following elements:

- rules for the formation of stable phraseological units;
- integral method and algorithm for recognition of PES by speech signal;
- intellectual recognition system of PES.

Let’s consider the rules for the formation of stable phraseological units, which can later be used for recognition and assessment of PES of aviation personnel for intelligent speech signal processing.

Rule 1.

The optimal duration of an utterance for the formation of a stable phraseological unit, and accordingly a speech signal, is a range with a time interval t = [1, 5/5], sec. It is with such a time interval that a high-quality training of the model takes place on the data.

Rule 2.

For recognition of PES, phraseological units must be strictly formed in plain and aviation English, since it was this speech that was used in training the model.

Rule 3.

The maximum duration of pauses in an utterance during the formation of a stable phraseological unit, and, accordingly, a speech signal, should be no more than t=1, 5, sec. Although the pauses in the algorithm in the processing of the speech signal are removed, prolonged silence can make it difficult to intelligently analyze the phrase.

Rule 4.

In an utterance, voiced sounds are the most informative and effective for recognition of PES and for training a model; therefore, to form stable phraseological units, words with a smaller number of voiced sounds N_{NV} should be selected.

Rule 5.

Phraseological units should be from frequently used and correspond to the professional field for the qualitative accumulation of samples in the database, and, therefore, for effective training.

Rule 6.

When using frequently encountered phraseological units, training must be carried out taking into account the analysis of changes in intonation or comparison of samples with each other.

Rule 7.

For high-quality and effective recognition of PES, you should select phraseological units in simple and aviation English, where non-native speakers make the least mistakes in pronunciation, since the efficiency of solving the problem of detection the speech signal increases [127].

In accordance with the proposed rules and on the basis of the recommendations of experts - specialists of the aviation industry, the author of the dissertation proposed stable phraseological units corresponding to four groups of professions of aviation personnel.

Some examples are given below:

– for flight personnel:

Request level change; Traffic is not in sight; Astana Line 331; Affirm, to keep heading 035, Astana Line 331; Roger, contact Almaty Tower on 123,6; Request low pass; RW vacated; Request push back clearance; Traffic in-sight; Traffic 12 o'clock in sight;

– for air traffic controllers:

You are not cleared, Astana Line 331; Stop immediately, I say again, stop immediately; Report Runway vacated; Astana Line 331 climb and maintain FL 300; Verify; Abort take-off; At own discretion; Negative; Read back; Approved; Acknowledge;

– for engineering and technical staff:

A tongue-and-groove joint; We need a solid rivet; A pressure gauge is out of order; Chocks away; Disconnect;

– for flight attendants:

Do you need medical attention; Are you injured; Remain seated; It's a matter of safety; Now, we are boarding all passengers in rows 1 through 3

Taking into account the high efficiency of the use of intelligent technologies for analyzing speech signals based on machine learning methods and deep neural networks, proved in Chapter 2 of the dissertation, and taking into account the rules for the formation of phraseological units, the author proposed an integral recognition technique - PES assessment of aviation personnel [128, 129].

The technique is implemented in two modes of operation:

– training - for the formation of standards;

– monitoring - recognition, or assessment, PES.

The methodology includes four separate speech-based PES assessment methods for four classes of aviation personnel: pilots, controllers, engineering and technical staff and flight attendants.

In general, each mode is implemented step by step as follows.

Step 1. Recording phrases.

In the training mode, control phrases are recorded, which subsequently serve as standards. This operation is carried out if there are no speech signals in the database with the corresponding information signs.

In the monitoring mode, the studied phrases (statements) of aviation personnel are recorded into a computer system, a specialized device or a mobile device (tablet, smartphone, cell phone) with PES recognition software:

– for the flight personnel - before the flight, during a medical examination, by pronouncing control phrases in the phraseology of radio exchange;

– for the dispatching staff - before leaving the shift;

– for engineering and technical personnel - before receiving an assignment for servicing aviation equipment;

– for flight attendants - during briefing.

Step 2. Formatting audio recording.

For both modes, preliminary preparation of audio recording for the formation of a speech signal with placement in the database.

Step 3. Preprocessing.

For both modes, at this stage, an operation is performed to remove pauses and extract features. Functionally, preprocessing includes a number of sequential operations, the dependence of which can be represented by the structural diagram shown in figure 2.6. Also at this stage, additional transformations are also carried out over the speech signals, which ensure the extraction of noise components from the sound recording of the speech signal.

Since a speech signal is analog, an important parameter in digital processing is its sampling with a frequency f_S; the value is selected based on the condition formulated in the Kotelnikov theorem. That is, the sampling frequency of the signal must be at least twice its upper frequency component.

Human hearing organs perceive sound vibrations in the frequency range from 20 Hz to 20 kHz. Moreover, the frequency of the main tone for male and female voices lies in the range of 70-450 Hz. In accordance with the described speech production model, the speech signal will not be limited in frequency band, but its spectrum will decay rapidly for high frequencies. For voiced sounds, the highest frequency below which the spectrum maxima are less than 40 dB is about 4 kHz. For unvoiced sounds, the spectral components remain high for a frequency of 8 kHz. At the same time, some authors limit the frequency range of the speech signal within 70 - 7000 Hz.

Regardless of this, for intelligible transmission of human speech, it is quite sufficient to accept f_S = 8 kHz, which is used in telephony. However, for accurate reproduction of the entire variety of speech sounds, a sampling frequency of f_S = 20 kHz is required. At the same time, since it is quite difficult to perform an anti-aliasing analog filter with a steep slope in the frequency response, the sampling frequency is chosen slightly higher than the required value, namely f_S = 22050 Hz.

Thus, in practice, the following sampling rates are used for digital audio recording:

– 22050 Hz, 44100 Hz – Audio CD;

– 48000 Hz – DVD, DAT.

Step 4. Allocation of informative features.

For both modes, a recording from the RAVDESS corpus was used as a speech signal. In the process of removing pauses, fragments of audio recordings before and after the start of the utterance are discarded, which makes it possible to reduce the size of the processed data, and also to analyze only moments of speech. Since the main task of this stage is to extract informative features from the speech signal in accordance with (2.7), which are obtained using short-term analysis. Then, for a separate sample of a speech signal, each test feature will be a two-dimensional data structure (figures 2.15, 2.17, 2.18) in the form of an array of size $m \times n \times 1$, where m – the number of coefficients used when calculating the analyzed feature, and n – the number of signal frame.

At the same time, it is reasonable to assume that not only the value of a specific coefficient, but also its location in space has an informative value. In particular, the second dimension of the array carries information about the time scale, so the order along it will have a certain value, and the first dimension is responsible for the relative position of the frequency components. Consequently, the informative features selected for the classification of PES have their own internal structure, the format of which is known to us. The array of calculated coefficients $X_{m \times n}$ can be considered, as a monochrome image (having one channel), where a certain value of an informative feature acts as each pixel. The sample speech signal s breaks down into successive sections s^i with the duration of 400 ms. Using a short-term analysis with the frame duration of M = 25 ms and an overlap of L = 10 ms, the required informative features are calculated. The resulting data are normalized and averaged over all n, obtained by two-dimensional arrays.

Step 5. The classification of voice samples using trained DCNN.

At this stage, in the training mode, information signs are entered into the database, and the corresponding expert solution P is assigned, which is formed on the basis of expert opinions of specialists, medical workers and psychologists of the aviation industry.

In the monitoring mode, as a result of digital processing, the required informative features $X_{m,k}$ are fed to the input of the DCNN model, which, in turn, at the output forms the result of the classification $P\ (Y)$ in the form of a vector of probabilities that the sample s belongs to each of the seven classes. Then the corresponding class of PES will be defined as

$$c = \sum_{i=1}^{C} I\left[y_i = \arg\max(\mathrm{Y}) \right] \cdot i. \qquad 4.1$$

Step 6. Decision making based on recognition of PES.

После распознавания - оценки PES принимается решение P, которое впоследствии должно снизить влияние HF на безопасность полетов, в частности:

– for the flight personnel:

If “PES” = {angry}, then P = {remove from flight}.
If “PES” = {disgusted}, then P = {send to rest}.
If “PES” = {fearful}, then P = {psychological counseling}.
If “PES” = {happy}, then P = {flight clearance}.
If “PES” = {neutral}, then P = {flight clearance}.
If “PES” = {sad}, then P = {psychological counseling}.
If “PES” = {surprised}, then P = {supervised flight}.

– for the dispatching staff:
If “PES” = {angry}, then P = {remove from air traffic control}.
If “PES” = {disgusted}, then P = {send to rest}.
If “PES” = {fearful}, then P = {psychological counseling}.;
If “PES” = {happy}, then P = {admission to service}.
If “PES” = {neutral}, then P = {admission to service}.
If “PES” = {sad}, then P = {psychological counseling}.
If “PES” = {surprised}, then P = {supervised flight}.

– for engineering and technical personnel:
If “PES” = {angry}, then P = {suspend from work}.
If “PES” = {disgusted}, then P = {send to rest}.
If “PES” = {fearful}, then P = {psychological counseling}.
If “PES” = {happy}, then P = {permit to work}.
If “PES” = {neutral}, then P = {permit to work}.
If “PES” = {sad}, then P = {psychological counseling}.
If “PES” = {surprised}, then P = {supervised flight}.

– for flight attendants:
If “PES” = {angry}, then P = {suspend from work}.
If “PES” = {disgusted}, then P = {send to rest}.
If “PES” = {fearful}, then P = {psychological counseling}.
If “PES” = {happy}, then P = {permit to work}.
If “PES” = {neutral}, then P = {permit to work}.
If “PES” = {sad}, then P = {psychological counseling}.
If “PES” = {surprised}, then P = {supervised flight}.

Step 7. Go to step 1. Repetitive recognition of PES.

A feature of the proposed technique is the formation of special recommendations with frequent repetitions of some signs.

Step 8. For flight-dispatching staff and flight attendants:

If “PES” = {angry}, {disgusted}, {sad} > 5 times, then P = {re-passage of MFEC}.

Step 9. For flight-dispatching staff and engineering and technical personnel and flight attendants:

If “PES” = {sad}, {fearful}, {surprised} > 5 times, then P = {send to rest};

If “PES” = {happy}, {neutral} > 10 times, then P = {person - balanced, responsible and worthy of promotion}.

The block diagram of the algorithm for the implementation of the integrated technique of recognition of PES by the speech signal is shown in figure 4.2.

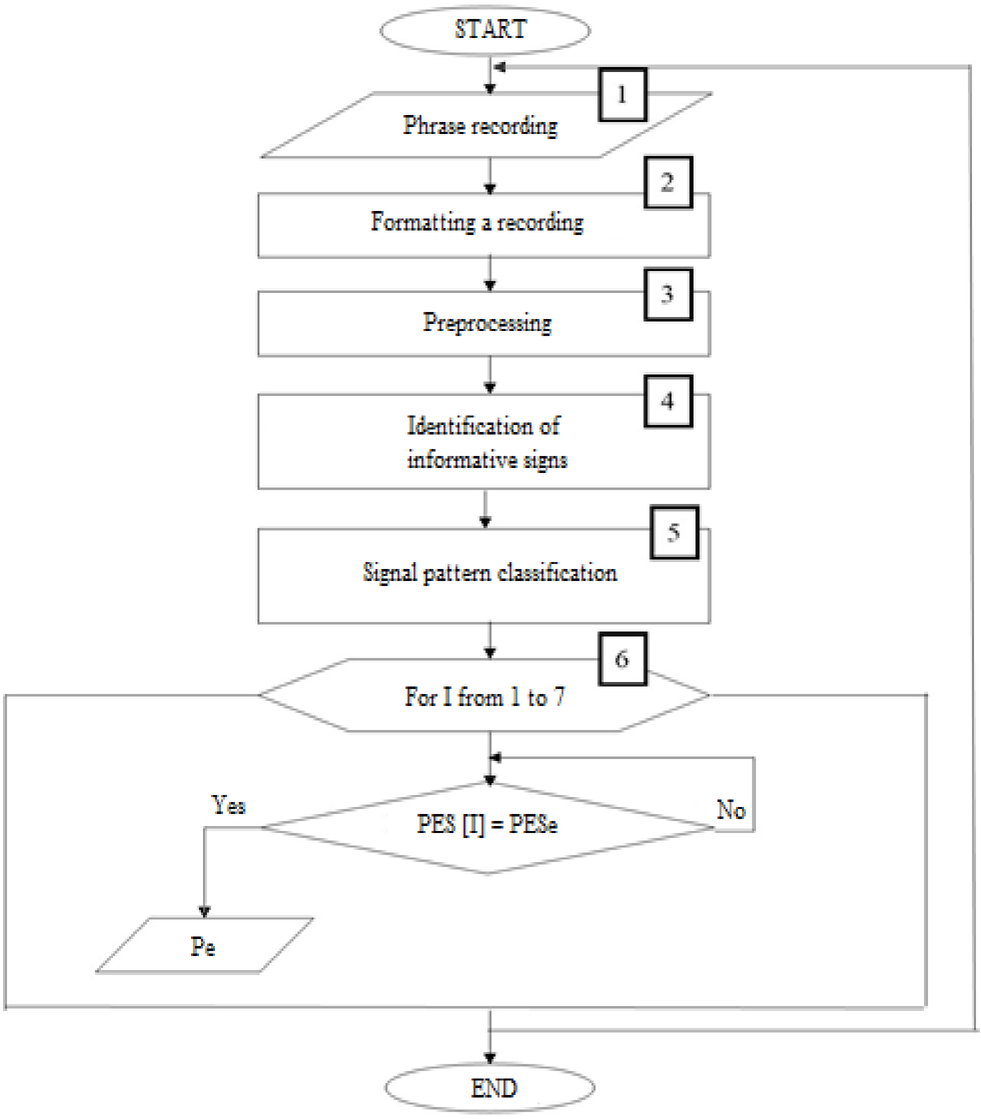


Figure 4.2 - The block diagram of the algorithm for the implementation of the integrated technique of recognition of PES by the speech signal

The author of the dissertation, having studied artificial intelligence technologies and based on the developed methodology and algorithm, proposed a new automated intelligent recognition system of PES based on the speech signal of aviation personnel, the structural diagram of which is shown in figure 4.2, and Figure 4.3 shows a block diagram of the algorithm for continuous monitoring.

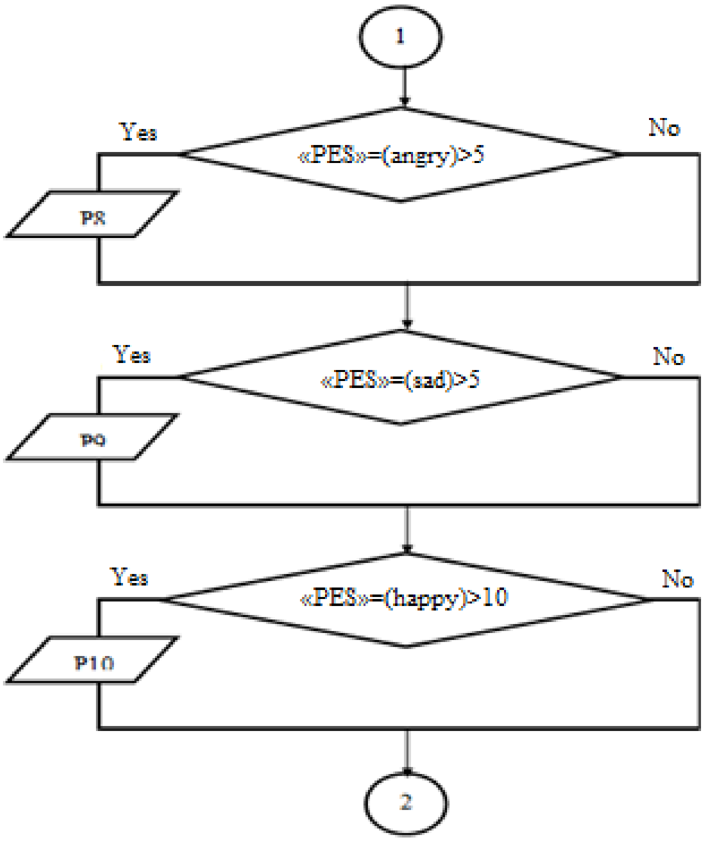


Figure 4.3 – Continuous monitoring block diagram algorithm

In figure 4.4, there are a user-flight safety specialist, ship commander, doctor on MFEC, shift supervisor, senior flight attendant, i.e., those who value the PES assessment of aviation personnel.

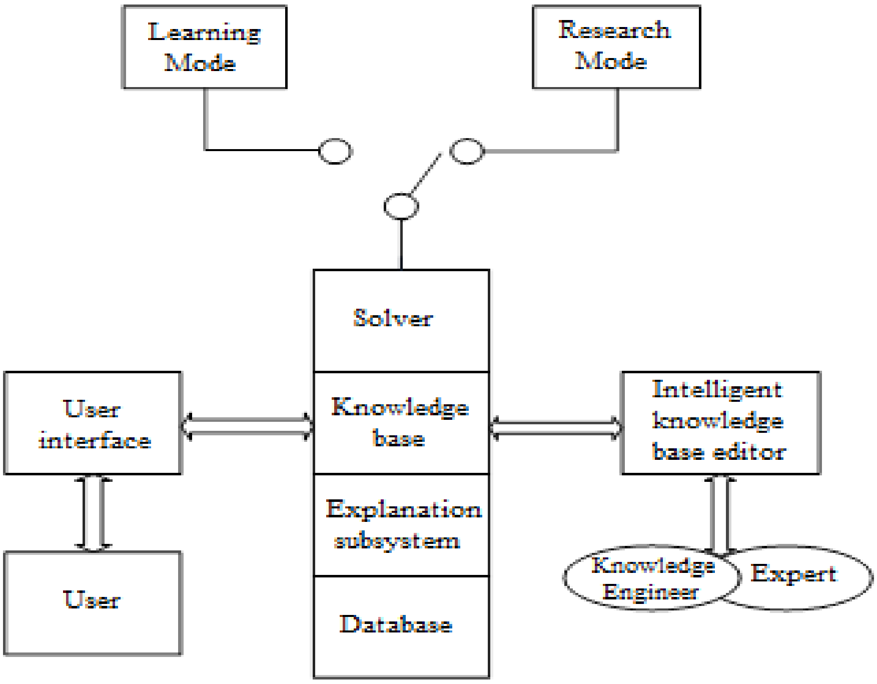


Figure 4.4 – Block diagram of the intellectual system of recognition of PES on the voice signal of aviation personnel

Knowledge engineer is an artificial intelligence specialist (analyst, interpreter engineer) who acts as an intermediate buffer between the expert and the knowledge base.

Expert - a specialist for the formation of a knowledge base in conjunction with the subject area: PES during control with entry into the database. Psychologists, medical professionals, specialists from the aviation industry with extensive experience, etc. can act as an expert.

The interface of a user is a program that implements a user dialogue with intelligent systems and devices at the stage of information input and output of results.

The knowledge base is the core of an intelligent system based on the voice signal of the aviation personnel. This is a knowledge body of the subject area, recorded on a machine medium in the form of software that implements an integral technique and an algorithm for PES recognition from a speech signal.

The database contains the values of information signs and expert opinions of standards.

Knowledge has the following properties:

– internal interpretability is along with information in the knowledge base, information structures, that allow not only storing information about PES, but also issuing descriptions of enterprises, results of medical examinations and other information necessary for profiling, should be presented;

– connectedness is compatible software based on artificial intelligence;

– activity is knowledge which involves the purposeful use of information for the ability to manage information processes.

Data are the quantitative and qualitative characteristics of information features of speech signals.

Solver is a program that simulates an expert's reasoning based on knowledge. Explanation subsystem is a program that allows the user to provide information in the form of recommendations.

The intelligent knowledge base editor is a program that provides a knowledge engineer with the opportunity to modify knowledge bases in an interactive mode and create new ones. It includes a system of nested menus, information signs and forms of speech signals with recommendations, as well as service tools that facilitate work with the base.

The user interface is necessary for the transmission of information signs and recommendations in a form convenient for it, as well as for manipulating knowledge.

Additional achievements in the implementation of the intelligent system of recognition of PES based on the speech signal of aviation personnel are the following capabilities:

– noise-immune recognition of complex speech signals;

– scalability of databases;

– the ability to issue quantitative characteristics of signals (root-mean square values, statistical characteristics, dynamic changes in PES over time, amplitude values, frequencies) and qualitative (linguistic - PES names, recommendations for further actions, etc.);

– dispatching control of the state of PES of aviation personnel with the transfer of information, in this case, recommendations via wireless communication channels of IEEE standard 802.11 [130].

Conclusions on the fourth section

The analytical studies carried out in this section have shown that HF, due to the insufficient level of knowledge of aviation and plain English, significantly affects flight safety.

Over the years, ICAO has documented and investigated a large number of incidents and accidents both in flight and on the ground, ranging from severe accidents to minor incidents.

Systematization of the existing rules and regulations of radiotelephony phraseology for pilots and dispatchers, as well as standard phrases for engineering and technical personnel and flight attendants made it possible to form a scientific and theoretical basis for the formation of phraseological units, which can subsequently be used for the recognition (assessment) of PES according to seven archetypal classes.

A new methodology for reducing the impact of HF on flight safety based on PES recognition by the voice signal of aviation personnel is proposed, which includes the following elements:

– rules for the formation of stable phraseological units of aviation English for pilots and air traffic controllers, as well as simple English for engineering and technical personnel and flight attendants;

– an integral method and algorithm for recognition of PES from a speech signal with the issuance of recommendations to experts to reduce the effect of HF on FS;

– an intelligent system of recognition of PES based on the speech signal of aviation personnel, which allows, among other things, to provide noise-immune recognition of complex speech signals, the scalability of databases, the ability to issue various quantitative and qualitative characteristics (linguistic - PES names, recommendations for further actions, etc.), dispatching etc.

CONCLUSION

In the dissertation research, a theoretical generalization is given and a solution to the urgent scientific problem of reducing the impact of the HF on the safety of the aviation transport system is proposed. The essence of the proposed solution lies in the assessment of PES of aviation personnel based on intelligent processing of the speech signals of stable phraseological units of aviation and plain English.

The results of the presented dissertation research are the following main scientific **findings** and practical solutions:

1. The results of the conducted analytical study to assess the impact of HF on flight safety create an important methodological effect that makes it possible to identify the existing problems in ensuring the psychological safety of work.
2. There are certain signs of behavior that make it possible to assess the psycho-emotional state, which are reflected, among other things, on the speech signal.
3. A professional group of aviation personnel, whose wrong actions are HF influencing accidents and incidents: pilots, air traffic controllers, engineers and flight attendants; has been identified.
4. Based on the statistical analysis of accidents and incidents in aviation, presented on the information resources of foreign organizations, quantitative characteristics of the impact of HF on flight safety have been established depending on the professional activities of aviation personnel.
5. A new effective approach to solving the problem of reducing the number of accidents and incidents by determining the PES of aviation personnel based on speech recognition is proposed, since this characteristic is individual, easily measurable, and the hardware and software implementation has a low cost and is applicable for a wide range of tasks.
6. It has been established that the problem of automatic classification of PES of aviation personnel by their speech is distinguished by a number of difficulties, among which the main ones are: the existing ambiguity in the formulation of the concept of emotion, the complex structure of the speech signal and the processes that generate it, the peculiarities of the psychophysical perception of sounds by a person, and, consequently, uncertainty in the choice of the characteristics of the speech signal.
7. Methods of the theory of CAL are effective intelligent technologies for the automatic classification of PES based on the speaker's speech, since they allow revealing hidden patterns in the data, including in the presence of some uncertainty.
8. The task of automatic classification of PES by CAL methods requires the formation of a representative set of training data, for which it is necessary to form a corpus of sound recordings of emotionally colored speech for seven classes in aviation and plain English, characterized by a variety of speakers of both sexes, a pronounced set of phrases, and the degree of emotional coloring.
9. A new discrete model of speech production is proposed for the selection of informative features in the preprocessing structure. The necessity of performing

special DSP procedures for preliminary filtration and removal of pauses has been established.

10. The expediency of using short-term analysis of speech signals for the classification of PES has been established, which made it possible to propose signs of objects for training the mathematical model of the classifier, containing information about the emotional color of the speech.

11. On the basis of a probabilistic approach to constructing a classifier model, the general principle of its training is determined, which satisfies both CAL algorithms and deep learning methods.

12. It is proposed to use deep learning technology in the form of artificial convolutional neural networks for analyzing data of a two-dimensional structure, since informative features of a speech signal have such a dimension when performing its short-term analysis to obtain mel-spectrograms, mel-frequency cepstral coefficients, differential parameters of mel-frequency cepstral coefficients and pitch classes.

13. The architecture of DCNN and an algorithm for its training on the selected informative features have been determined, which makes it possible to obtain high results of the PES classification of aviation personnel for seven classes of objects only on the basis of the acoustic data of the samples under study. It was found that the DCNN-based classifier model allows obtaining the best classification results when it is trained on informative features in the form of mel-frequency cepstral coefficients.

14. To improve the PES classification parameters, a method is proposed that combines the classification results from two DCNNs trained on different types of informative features as mel-spectrograms and mel-frequency cepstral coefficients. As a result, the result is formed in the form of the average value of the probabilities of belonging of the studied sample to each of the seven PES classes predicted by each neural network, which provides a multiclass share of correct answers equal to 0.9007 on the deferred test subsample.

15. During the analysis of the results obtained, it was found that the calculated indicators of the classification quality according to the proposed method are superior to the results for other effective CAL algorithms, such as a random forest, a fully connected neural network, gradient boosting, etc.

16. An analysis of sources based on similar studies shows that when using only acoustic information of a speech signal to recognize seven types of PES, the proposed method surpasses the existing models in quality metrics.

17. Analytical studies carried out in this work have shown that HF, due to the insufficient level of knowledge of aviation and simple English, significantly affects flight safety.

18. The existing rules and regulations of the phraseology of radio communication for pilots and air traffic controllers, as well as standard phrases for engineering and technical personnel and flight attendants, have been systematized, which made it possible to propose a scientific and theoretical basis for the formation

of phraseological units, which can subsequently be used for recognition of PES according to seven archetypal classes.

19. A new methodology for reducing the impact of HF on flight safety based on recognition of PES by the speech signal of aviation personnel is proposed, which includes the following elements: rules for the formation of stable phraseological units, an integral technique and algorithm, an intelligent system of recognition of PES by a speech signal with the issuance of recommendations to experts to reduce the influence of HF on FS.

20. An intelligent system of recognition of PES has been developed based on the speech signal of aviation personnel, which allows, among other things, to provide noise-immune recognition of complex speech signals, the scalability of databases, the ability to issue various quantitative and qualitative characteristics (linguistic - PES names, recommendations for further actions, etc.), dispatching, etc.

LIST OF SOURCES USED

1 Human factors training manual. – Ed. 1st // https://www.globalairtraining.com/resources/DOC-9683.pdf. 20.04.2019.

2 Gippenreiter Yu.B, Petukhova V.V. Psychology of thinking: chrestomathy on General Psychology. Ed. – M.: Publishing house, 1982. – 356 p.

3 Resolution of the Government of the Republic of Kazakhstan. On the state program "Digital Kazakhstan": appr. on December 12, 2017, №827 // https://adilet.zan.kz/rus/docs/P1700000827. 20.04.2019.

4 The Global operations support plan. June 2007 / International Civil Aviation Organization // https://www.icao.int/secretariat/AirNavigation/Documents/GASP/GASP_ru.pdf. 19.07.2018.

5 Bodrov V.A. Psychology of professional suitability: a study guide for universities. – M.: PER SE. 2001. – 511 p.

6 Klebelsberg D. Transport psychology. – M.: Transport, 1989. – 366 p.

7 Кружалов А.И. Влияние человеческого фактора на безопасность полетов при техническом обслуживании авиатехники // Научный вестник Московского государственного технического университета гражданской авиации. – 2006. – №100. – С. 55-59.

8 Anayatova R.K. The Dirty Dozen: The need and Possibility of minimizing the HF influences on FS // Mater. of the 5st internat. scient.-pract. conf. «Integration of the Scientific Community to the Global Challenges of Our Time». – Tokyo, 2020. – P. 381-387.

9 Theory of ERM. Special mental conditions // https://bgdstud.ru/bezopasnost-zhiznedeyatelnosti/638-osobye-psixicheskie-sostoyaniya. 5.05.2019.

10 Anayatova R.K., Karipbaev S.Zh. Actual problems of aviation security in the Republic of Kazakhstan // Mater. of the 3th internat. scient. and pract. youth conf. "Creative potential of youth in solving aerospace problems." – Baku, 2018. – P. 244-246.

11 Hallinan J. Why are we wrong? Thinking traps in action. – M.: Mann, Ivanov and Ferber, 2014. – 240 p.

12 Dismukes R.K. Human error in aviation. – NY.: NASA Ames Research Center, 2016. – 604 p.

13 Law of the Republic of Kazakhstan. On the use of the airspace of the Republic of Kazakhstan and aviation activities: appr. July 15, 2010, №339-IV // https://online.zakon.kz/Document/?doc_id=30789893#pos=1;20. 19.03.2019.

14 Statistics of the largest aircraft crashes in the world // https://forinsurer.com/public/17/01/10/3824. 19.03.2019.

15 Doc. 9806: Basic Principles of Human Factors Consideration in the guide on conducting safety inspections / International Civil Aviation Organization. – Montréal; Quebec, 2010. – 224 p.

16 Titz S.N. Human factor. – Samara, 2012. – 64 p.

17 Martinussen M., Hunter D.R. Aviation Psychology and Human Factors. – Boca Raton CRC Press, 2017. – 364 p.

18 Koshekov K.T., Anayatova R.K. Ways to reduce the negative impact of the human factor on flight safety // Bulletin of PSU. – 2020. – Vol. 2. – P. 286-299.

19 Salas E., Maurino D. Human Factors in Aviation. – Ed. 2nd. – NY.: Academic Press, 2010. – 732 p.

20 Order of the acting Minister of Transport and Communications of the Republic of Kazakhstan. On approval of the Instruction on the organization and maintenance of air traffic: appr. May 16, 2011, №279 // https://adilet.zan.kz/rus/docs. 25.02.2019.

21 Anayatova R., Janpeisova Zh., Essenaliyeva M. et al. Problems of International air cargo transportation in Kazakhstan: New strategies and its implementation // Bulletin of the Donetsk Academy of Automobile Transport. – 2018. – №1. – P. 49-56.

22 Prabhu P.V., Drury C.G. Information Requirements of Aircraft Inspection: Framework and Analysis // The International Journal of Man Machine Studies. – 1995. – Vol. 45(6). – P. 679-695.

23 Vidulich M.A. et al. Advances in Aviation Psychology. – Boston: Routledge, 2016. – 286 p.

24 Anodina T.G., Kuranov V.P., Fedorov Yu.M. The concept of a promising air traffic management system // Proceedings of the State Research Institute of Civil Aviation - 1989. – Vol. 286. – P. 3-5.

25 Internet portal. Aviation accidents // http://uaecis.com. 06.07.2018.

26 Anayatova R.K. The role of Human Factor and its impact on Flight Safety in Civil Aviation // International scientific Journal Science and Life of Kazakhstan. – 2020. – Vol. 10/2(142). – P. 444-448.

27 Report on the collision of aircraft over Germany, published on May 19, 2004 // ttps://www.bfu-web.de/DE/Unfallmeldung/unfallmeldung_node. html;jsessionid=8CC76A1398EEAB7A49EC8C1F061FE06F.live11291. 10.11.2019.

28 Shumilov I.S., Aviation accidents. Causes of occurrence and possibilities of prevention. Moscow, 2006.

29 Interstate Aviation Committee // www.mak-iac.org. 25/09.2019.

30 Imasheva G.M., Dolzhenko N.A., Anayatova R.K. et al. Commercial use of Aircraft based on Safety Risk // Journal of Advanced Research in Law and Economics. – 2018. – Vol. 9, №8. – P. 2615-2622.

31 Koshekov K.T., Anayatova R.K. The mechanism of flight safety management in the Republic of Kazakhstan based on the factor of human behavior // Bulletin of KazATK. – 2020. – Vol. 2(113). – P. 344-353.

32 Koshekov A.K., Kobenko V.Yu., Kislov A.P. et al. A method of pattern recognition based on identification measurements // Bulletin of PSU. Energy series. - Pavlodar. – 2019. – №4. – P. 448-460.

33 Anayatova R.K., Turdyakynova M. The role of English in ensuring flight safety in civil aviation // Mater. of the internat. scient.-theor. conf. of students and young scientists "Rukhani zhangyru - the choice of the President, supported by the society" and the World Cosmonautics Day. – Almaty, 2018. – P. 166-168.

34 Doc. 9758-AN/966: Human Factors For ATM Systems / ICAO. – Montreal, 2000. – 127 p.

35 Zubkov B.V., Prozorov S.E. Flight safety: textbook. – Ulyanovsk, 2013. – 451 p.

36 Zadorozhny A.I., Shcheglov I.N. Ways of development of automated air traffic control systems based on the use of the principles of artificial intelligence // Problems of air traffic management. Flight safety: col. of scientific papers. – M.: State Research Institute "Air Navigation", 2000. – P. 59-70.

37 Мельник В.Н. Анализ ошибочных действий авиадиспетчеров в психологическом аспекте // Новое слово в науке: перспективы развития: матер. 11-й междунар. науч.-практ. конф. – Чебоксары: ЦНС «Интерактив плюс», 2017. – С. 118-127.

38 Biometrics from A to Z, a complete guide to biometric identification and authentication // https://securityrussia.com/blog/biometriya.html. 4.05.2019.

39 Малини П.В. Технология голосовой идентификации личности на основе проекционных методов анализа многомерных данных: дис. … канд. техн. наук: 05.13.19. – Барнаул, 2015. – 139 с.

40 Doc. 9859. AN/474: Safety Management Manual. – Ed. 2nd. / International Civil Aviation Organization. – Montréal; Quebec, 2013. – 254 p.

41 Doc. 9835. AN/453: Guidelines for implementation of the requirements of ICAO on Language Proficiency. – Ed. 2nd. – Montreal: ICAO, 2010. – 254 p.

42 Ekman P. Universals and Cultural Differences in Facial Expressions of Emotions // Nebraska Symposium on Motivation. – Lincon: University of Nebraska. 1972. – P. 207-283.

43 Kamath U., Liu J., Whitaker J. Deep Learning for NLP and Speech Recognition. – N-Y.: Springer, 2019. – 649 p.

44 Flach P. Machine learning. Science and art of building algorithms that extract knowledge from data / transl. from english A.A. Slinkina. – M.: DMK Press, 2015. – 400 p.

45 Cornelius R.R. The science of emotion: Research and tradition in the psychology of emotions. – New Jersey: Prentice-Hall, 1996. – 260 p.

46 Davydov A.G., Kiselev V.V., Kochetkov D.S. Classification of the speaker's emotional state by voice: problems and solutions // Proceed. of the internat. conf. "Dialogue 2011". – M.: RGTU, 2011. – P. 178-185.

47 Livingstone S.R., Russo F.A.The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English // PLoS ONE. – 2018. – Vol. 13(5). – P. e0196391.

48 Haq S., Jackson P.J.B. Speaker-Dependent Audio-Visual Emotion Recognition // http://personal.ee.surrey.ac.uk/Personal/P.Jackson/. 5.04.2019.

49 Pichora-Fuller M. et al. Toronto emotional speech set (TESS) // https://doi.org/10.5683/SP2/E8H2MF. 5.04.2020.

50 Garofolo J.S. et al. TIMIT Acoustic-Phonetic Continuous Speech Corpus. – Philadelphia: Linguistic Data Consortium, 1993. – 95 p.

51 Тьюки Д.В. Анализ результатов наблюдений: разведочный анализ / пер. с англ. – М.: Мир, 1981. – 693 с.

52 Rabiner L.R., Schafer R. B. Digital processing of speech signals: Translated from English. – M.: Radio and communication, 1981. – 496 p.

53 Goldenberg L.M. et al. Digital signal processing: handbook. – M.: Radio and communication, 1985. – 312 p.

54 Sergienko A.B. Digital Signal Processing: A Textbook for Universities. – Ed. 2nd. – SPb.: Peter, 2006. – 751 p.

55 Akhmad Kh.M. et al. Introduction to digital processing of speech signals: textbook. – Vladimir: Publishing house of Vladimir State University, 2007. – 192 p.

56 Nikamin V.A. Digital sound recording. Technologies and standards. – SPb: Science and Engineering, 2002. – 256 p.

57 Design of devices for digital and mixed signal processing / ed. W. Kester. – M.: Technosphere, 2010. – 328 p.

58 Оппенгейм А.В., Шафер Р.В. Цифровая обработка сигналов / пер. с англ. – М.: Связь, 1979. – 416 с.

59 Self D. Design of audio power amplifiers. – Ed. 3rd. – M.: DMK Press, 2009. – 536 p.

60 Widrow B., Stearns S.D. Adaptive Signal Processing. – M.: Radio and Communication, 1989. – 440 p.

61 Sergienko A.B. Adaptive filtering algorithms: implementation features in MATLAB. EXPonentaPro // Mathematics in applications. – 2003. – №1. – P. 18-28.

62 Ivanova L.V. The logic of speech in performing arts: study guide. – Oryol: Oryol State Institute of Arts and Culture, 2013. – 80 p.

63 Shelukhin O.I. et al. Digital processing and speech transmission. – M.: Radio and communication, 2000. – 456 p.

64 Freeman D. et al. A Voice Activity Detector for the Pan-European Digital Cellular Mobile Telephone Service // IEE Colloquium "Digitized Speech Communication via Mobile Radio". – London, 1988. – P. 6-5.

65 Volchenkov V.A., Vityazev V.V. Methods and algorithms for detecting speech activity // Digital Processing of Originals. – 2013. – №1. – P. 54-60.

66 Sergienko A.B. Digital signal processing: study guide. – Ed. 3rd. – SPb.: BHV-Petersburg, 2011. – 768 p.

67 Zwicker E., Fastl H. Psychoacoustics. – Ed. 2nd. – Berlin: Springer-Verlag, 1990. – 463 p.

68 Rabiner L.R., Schafer R.W. Introduction to Digital Speech Processing: Foundations and Trends in Signal Processing. – Boston, 2007. – 213 p.

69 Rabiner L.R., Juang B.H. Hidden Markov Models for Speech Recognition // Technometrics. – 1991. – Vol. 33, №3. – P. 251-272.

70 Davis S.B., Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences // IEEE Transactions on ASSP. – 1980. – Vol. 28. – P. 357-366.

71 Huang X., Acero A., Hon H. Spoken Language Processing: A guide to theory, algorithm, and system development. – New Jersey: Prentice Hall, 2001. – 965 p.

72 Shepard R.N. Circularity in judgments of relative pitch // Journal of the Acoustical Society of America. – 1964. – Vol. 36(212). – P. 2346-2353.

73 Lindley M., Turner-Smith R. Mathematical models of musical scales: A new approach. – Bonn, 1993. – 308 p.

74 Librosa // https://librosa.org/doc/latest/index.html. 21.09.2020.

75 Nikolenko S., Kadurin A., Arkhangelskaya E. Deep learning. – SPb.: Peter, 2018. – 480 p.

76 Гудфеллоу Я., Бенджио И. и др. Глубокое обучение / пер. с англ. – Изд. 2-е. – М.: ДМК Пресс, 2018. – 653 с.

77 Raska S. Python and Machine Learning / transl. engl. – M.: DMK Press, 2017. – 418 p.

78 LeCun Y., Bengio Y., Hinton G. Deep learning // Nature. – 2015. – Vol. 521. – P. 436-444.

79 Sutskever I., Martens J. et al. On the importance of initialization and momentum in deep learning // Proceed. of the 30th internat. conf. on Machine Learning. – Atlanta, 2013. – P. 1139-1147.

80 Kingma D., Ba J. Adam: A Method for Stochastic Optimization // http: //arxiv.org/abs/1412.6980. 16.02.2019.

81 Julie A., Pal S. Keras Library - A Deep Learning Tool. – M.: DMK Press, 2017. – 294 p.

82 Keras // https://keras.io/. 21.09.2020.

83 Tensor Flow // https://www.tensorflow.org/. 21.09.2020.

84 Davis J., Goadric M. The relationship between Precision-Recall and ROC curves // ICML '06 Proceedings of the 23rd international conference on Machine learning. – Pittsburgh, PA, 2006 – P. 233-240.

85 Leo B. Random Forests // Machine Learning. – 2001. – Vol. 45(1). – P. 5-32.

86 Friedman J.H. Stochastic Gradient Boosting // Computational Statistics and Data Analysis. – 1999. – Vol. 38. – P. 367-378.

87 Scikit-learn // https://scikit-learn.org/stable/. 21.09.2020.

88 Ayadi M.E. et al. Survey on speech emotion recognition: Features, classification schemes, and databases // Pattern Recognition. – 2011. – Vol. 44. – P. 572-587.

89 Schuller B., Rigoll G., Lang M. Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture // Proceedings of the ICASSP. – 2004. – Vol. 1. – P. 577-580.

90 Schuller B. Towards intuitive speech interaction by the integration of emotional aspects // Procced. internat. conf. on Systems, Man and Cybernetics (IEEE. 2002). – Yasmine Hammamet, Tunisia, 2002. – P. 6.

91 Lee C., Narayanan S. Toward detecting emotions in spoken dialogs // IEEE Trans. Speech Audio Process. – 2005. – Vol. 13(2). – P. 293-303.

92 Go H., Kwak K., Lee D., Chun M. Emotion recognition from the facial image and speech signal // Proceedings of the IEEE SICE. – 2003. – Vol. 3. – P. 2890-2895.

93 Kamel M.S., Karray F. et al. Speech emotion recognition using Gaussian mixture vector autoregressive models // ICASSP. – 2007. – Vol. 4. – P. 957-960.

94 Burkhardt F., Paeschke A., Rolfes M. et al. A database of German emotional speech // Procced. 9th European conf. on Speech Communication and Technology Proceedings of the Interspeech 2005. – Lissabon, 2005. – P. 1517-1520.

95 Schuller B., Rigoll G., Lang M., Hidden Markov model-based speech emotion recognition // International Conference on Multimedia and Expo (ICME). – 2003. – Vol. 1. – P. 401-404.

96 Javier G., Sundgren D. et al. Speech emotion recognition in emotional feedback for Human-Robot Interaction // International Journal of Advanced Research in Artificial Intelligence. – 2015. – Vol. 4, №2. – P. 20-27.

97 Martin O., Kotsia I., Macq B. et al. Pitas The eNTERFACE’05 Audio-Visual Emotion Database // Data Engineering Workshops: proceed. 22nd internat. conf. – Atlanta, 2006. – P. 8.

98 Guidelines for Aviation English Training Programs: cir 323 AN/185 / International Civil Aviation Organization. – Montréal, 2010. – 80 p.

99 Feldman, J.M. (1998). Speaking with one voice // Air Transport World. – 1998. – Vol. 35(11). – P. 42-51.

100 Incident report EW/C2007/0602 // http://www.securiteaerienne.com /ill/files/Boeing-737-500-SP-LKA-06-08-LOT.pdf. 6.08.2019.

101 Cookson S. Zagreb and Tenerife, Airline accidents involving linguistic factors // Australian review of applied linguistics. – 2009. – Vol. 32(3). – P. 22.1-22.14.

102 Zhambylkyzy M., Kotiyeva L.M., Anayatova R.K. et al. Lexical-phraseological features of phrasal verbs and difficulties in their study // X Linguae, European Scientific Language Journal. – 2018. – Vol. 11, №2. – P. 292-302.

103 Philips D. Linguistic Security in the Syntactic Structures of Air Traffic Control English // English World-Wide. – 1991. – Vol. 12(1). – P. 103-124.

104 Aitchison J. Aitchison’s Linguistics: A practical introduction to contemporary linguistics. – London: Hodder Headlines, 2010. – 318 p.

105 Zhunussova D., Anayatova R., Kashkinbayeva K. Effective ways of teaching aviation English through CLIL approach // International scientific journal Science and Life of Kazakhstan, Almaty. – 2019. – Vol. 5/2. – P. 222-226.

106 Emery H. Aviation English for the Next Generation // In book: Changing Perspectives of Aviation English Training. – Warsawa, 2016. – P. 8-34.

107 Uplinger S. English-language Training for Air Traffic Controllers Must Go beyond Basic ATC Vocabulary // Flight Safety Foundation Airport Operations. – 1997. – Vol. 23(5). – P. 1-6.

108 Friginal E., Matthews E., Roberts J. English in Global Aviation: Context, Research, and Pedagogy. – London: Bloomsbury Academic, 2020. – 320 p.

109 Werfelman L. et al. AeroSafety World. – Alexandria VA, 2008. – 63 p.

110 McMillan D et al. "…Say again?..." Miscommunications in air traffic control: unpubl. mas. thes. … – Brisbane: Queensland University of Technology, 1998. – 60 p.

111 Nevile M. Being out of order: Overlapping talk as evidence of trouble in airline pilot's work // In book: Advances in Discourse Studies. – Abingdon: Routledge, 2008. – 262 p.

112 Anayatova R.K., Dzhanpeisova Zh.M. Formation of intercultural communicative competencies of students in teaching aviation English // Implementation of Benchmark test in CAA and its role in flight safety: col. mater. 1st internat. pre-service teachers conf. – Shymkent, 2018. – P. 123-127.

113 Anayatova R.K., Dzhanpeisova Zh.M. Formation of intercultural communicative competencies of students in teaching aviation English // Bulletin of KazNPU named after Abay. – 2018. – Vol. 2(58). – P. 197-201.

114 Farris C. et al. Air Traffic Communication in a Second Language: Implications of Cognitive Factors for Training and Assessment // Tesol Quarterly. – 2008. – Vol. 42(3). – P. 397-410.

115 Mell J. Emergency Calls – Messages out of the blue. Toulouse: Ecole Nationale de l'Aviation Civile // http://www.icao.it/anb/sg/pricesg/ backgrounds/OotB.htm. 15.08/2019.

116 Anayatova R.K., Koshekov K.T., Savostin A.A. et al. Automatic emotion recognition by speech signal in aviation security tasks // Bulletin of the Academy of Civil Aviation. – 2020. – Vol. 3(18). – P. 10-18.

117 Estival D., Molesworth B. A study of ELS Pilots' Radio Communication in the General Aviation Environment // Australian Review of Applied Linguistics. – 2011. – Vol. 32(3). – P. 24.1-24.16.

118 Farris C. The effects of message length, L2 proficiency and cognitive workload on performance accuracy and speech production in a simulated pilot navigation task: master degree in applied linguistics. – Monreal, Quebec: Concordia University, 2007. – 168 p.

119 Seidlhofer B. Research Perspective on Teaching English as a Lingua Franca // Annual Review of Applied Linguistics. – 2004. – Vol. 24. – P. 209-239.

120 Anayatova R.K., Turdyakinova M. The role of English in ensuring flight safety in civil aviation // Mater. of the internat. scient.-theoret. conf. of students and young scientists "Rukhani zhangyru - the choice of the President, supported by the society" and the World Cosmonautics Day. – Almaty, 2018. – P. 166-168.

121 Anayatova R.K. Flight Safety is adversely affected by human factor // Mater. of the 7th internat. scient. and pract. conf. "Science and Education in the Modern World: Challenges of the 21st Century". – Nur-Sultan, 2020. – P. 139-142.

122 Bullock N. Wider consideration in teaching speaking of English in the context of aeronautical communications // Journal of the IATEFK ESP SIG. – 2015. – Issue 45. – P. 4-11.

123 Kim H., Elder C. Understanding aviation English as a lingua franca: Perceptions of Korean aviation personnel // Australian Review of Applied Linguistics. – 2009. – Vol. 32(3). – P. 23.1-23.17.

124 Friginal E., Matthews E., Roberts J. English in Global Aviation. Context, Research, and Pedagogy. – London: Bloomsbury Academic, 2020. – 320 p.

125 Anayatova R.K. Possible ways to solve the problem of reducing the accident operation of the aviation transport // Bulletin of the Academy of Civil Aviation. – 2019. – Vol. 2(13). – P. 21-26.

126 Anayatova R.K., Yessenamanova K.M. Influence of multicultural environment on personal development // Bulletin of KazNPU named after Abai. Psychology series. – 2017. – Vol. 3(52). – P. 101-104.

127 Anayatova R.K., Koshekov K.T., Savostin A.A. Automatic emotion recognition by speech signal // Bulletin of the Academy of Civil Aviation. – 2020. – Vol. 3(18). – P. 31-38.

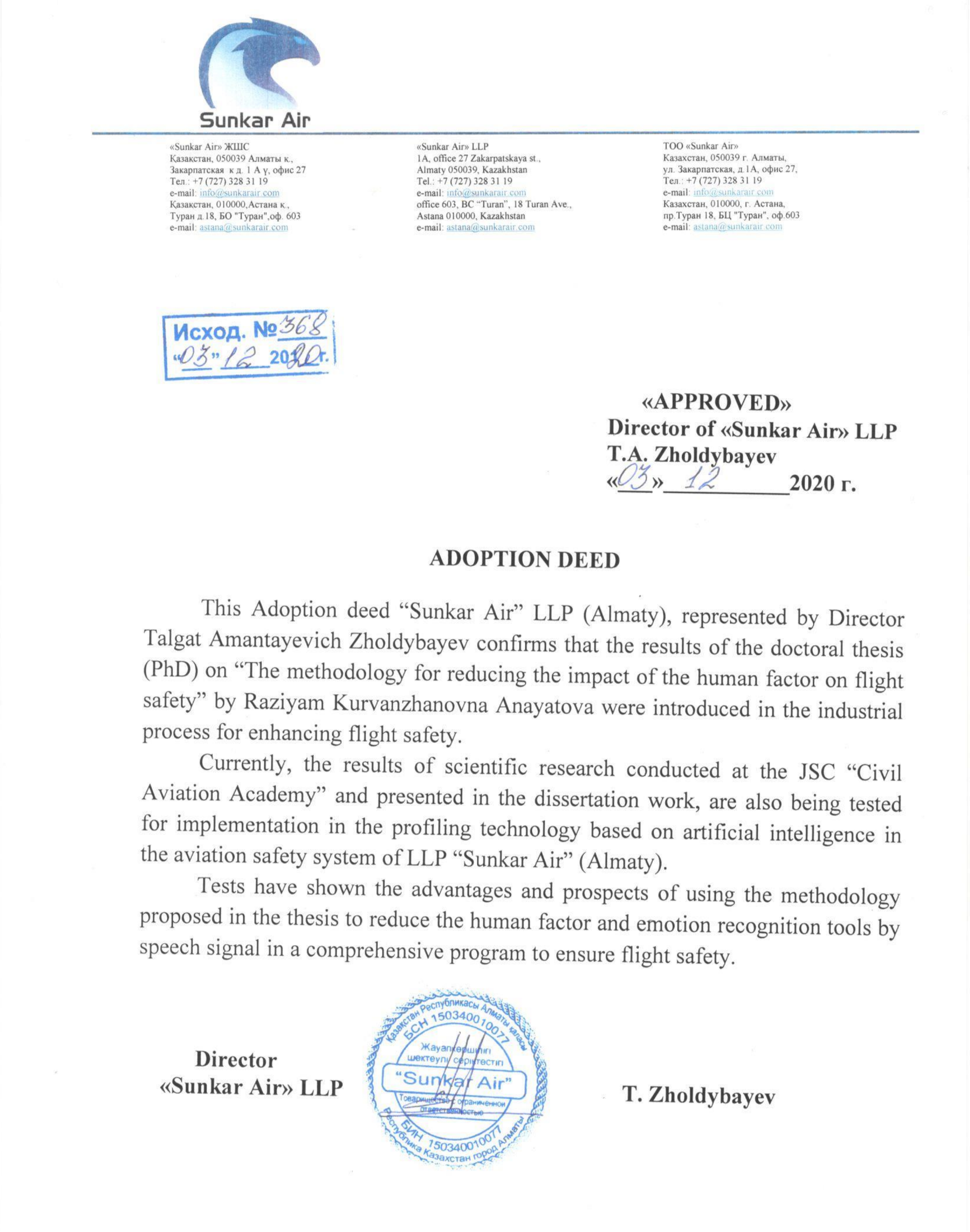
128 Anayatova R.K., Kobenko V.Yu., Koshekov K.T. Sound model of distributions of random signals // Bulletin of the Academy of Civil Aviation. – 2019. – Vol. 4(15). – P. 76-81.

129 Koshekov K.T., Kobenko V.Yu., Anayatova R.K. et al. Method of automatic classification of the speaker's emotional state by voice // Mater. of the 14th internat. scient. and techn. conf. «Dynamics of Systems, Mechanisms and Machines». – Omsk, 2020. – P. 51-59.

130 Koshekov K.T. et al 2021. Automatic classification method of the speaker’s emotional state by voice // Journal of Physics. – 2021. – Vol. 1791. – P. 1-9.

APPENDIX A

Act of implementation



Sunkar Air

«Sunkar Air» ЖШС
Қазақстан, 050039 Алматы қ.,
Закарпатская к д. 1 А ү, офис 27
Тел.: +7 (727) 328 31 19
e-mail: info@sunkarair.com
Қазақстан, 010000,Астана қ.,
Туран д.18, БО "Туран",оф. 603
e-mail: astana@sunkarair.com

«Sunkar Air» LLP
1A, office 27 Zakarpatskaya st.,
Almaty 050039, Kazakhstan
Tel.: +7 (727) 328 31 19
e-mail: info@sunkarair.com
office 603, BC "Turan", 18 Turan Ave.,
Astana 010000, Kazakhstan
e-mail: astana@sunkarair.com

ТОО «Sunkar Air»
Казахстан, 050039 г. Алматы,
ул. Закарпатская, д.1А, офис 27,
Тел.: +7 (727) 328 31 19
e-mail: info@sunkarair.com
Казахстан, 010000, г. Астана,
пр.Туран 18, БЦ "Туран", оф.603
e-mail: astana@sunkarair.com

Исход. №368
«03» 12 2020г.

«APPROVED»
Director of «Sunkar Air» LLP
T.A. Zholdybayev
«03» 12 2020 г.

ADOPTION DEED

This Adoption deed "Sunkar Air" LLP (Almaty), represented by Director Talgat Amantayevich Zholdybayev confirms that the results of the doctoral thesis (PhD) on "The methodology for reducing the impact of the human factor on flight safety" by Raziyam Kurvanzhanovna Anayatova were introduced in the industrial process for enhancing flight safety.

Currently, the results of scientific research conducted at the JSC "Civil Aviation Academy" and presented in the dissertation work, are also being tested for implementation in the profiling technology based on artificial intelligence in the aviation safety system of LLP "Sunkar Air" (Almaty).

Tests have shown the advantages and prospects of using the methodology proposed in the thesis to reduce the human factor and emotion recognition tools by speech signal in a comprehensive program to ensure flight safety.

Director
«Sunkar Air» LLP

Қазақстан Республикасы Алматы қаласы
БСН 150340010077
Жауапкершілігі шектеулі серіктестігі
"Sunkar Air"
Товарищество с ограниченной ответственностью
БИН 150340010077
Республика Казахстан город Алматы

T. Zholdybayev

"CIVIL AVIATION ACADEMY" JSC

APPROVED
Vice-rector for academic
affairs and collaboration
JSC "Civil Aviation Academy"
______________K.B. Zhakupov
«26» января 20____г.

Қазақстан Республикасы Алматы қ.
Республика Казахстан г. Алматы
Азаматтық Авиация Академиясы
Академия Гражданской Авиации
БСН/БИН: 010340003480
Акционерлік қоғамы
Акционерное общество

ADOPTION DEED

Full commission

Chairman: Arken S. Shanlayakov, Candidate of Technical Sciences, Director of Academic Affair Department of Civil Aviation Academy (hereinafter - the Academy)

Committee members:

1. Yerbol A. Ospanov, Head of the Flight Operation of Aircraft Department;
2. Inna N. Ryabchenko, Senior teacher of Aviation English Department

have drawn up the present Deed that the results of scientific research stated in the thesis work of Raziyam K. Anayatova "The methodology for reducing the impact of the human factor on flight safety" have been tested and implemented in the educational process of the Academy, in particular:

1. The rules of phraseology formation were introduced in the curricula of the following disciplines "Phraseology of radio communication in English", "Aviation and technical English", "Professional English" on the specialties 6B07112 "Flight operation of air transport (pilot)" and 6B07108 "Technical operation of aircraft".

2. Methodology of applying intelligent technologies in civil aviation to the curriculum of the discipline "Theory and Practice of Flight / Safety Management / Aviation Safety" for the Educational Programme 6B07103 "Air Traffic Services".

According to the results of the research, it was found that the results of the dissertation of Raziyam K. Anayatova are reliable and able to have a positive effect on the learning process and outcomes.

Chairman of Commission **A. Shanlayakov**

Members of Commission

Ye. Ospanov

I. Ryabchenko