

ISSN 2078-9181

DOI 10.15622/sp.60

РОССИЙСКАЯ АКАДЕМИЯ НАУК
Отделение нанотехнологий и информационных технологий

САНКТ-ПЕТЕРБУРГСКИЙ
ИНСТИТУТ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ РАН

ТРУДЫ СПИИРАН

proceedings.spiiras.nw.ru



ВЫПУСК 5(60)



Санкт-Петербург
2018

18+

SPIIRAS PROCEEDINGS

Issue № 5(60), 2018

Scientific, educational, and interdisciplinary journal primarily specialized
in computer science, automation, and applied mathematics

Trudy SPIIRAN ♦ Founded in 2002 ♦ Труды СПИИРАН

Founder and Publisher

St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences

Editor-in-Chief

R. M. Yusupov, Prof., Dr. Sci., Corr. Member of RAS, St. Petersburg, Russia

Editorial Board Members

A. A. Ashimov ,	Prof., Dr. Sci., Academician of the National Academy of Sciences of the Republic of Kazakhstan, Almaty, Kazakhstan
N. P. Veselkin ,	Prof., Dr. Sci., Academician of RAS, St. Petersburg, Russia
O. Yu. Gusikhin ,	Ph. D., Dearborn, USA
V. Delic ,	Prof., Dr. Sci., Novi Sad, Serbia
A. Dolgui ,	Prof., Dr. Habil., St. Etienne, France
M. Zelezny ,	Assoc. Prof., Ph.D., Plzen, Czech Republic
I. A. Kalyaev ,	Prof., Dr. Sci., Academician of RAS, Taganrog, Russia
A. A. Karpov ,	Assoc. Prof., Dr. Sci., St. Petersburg, Russia
D. A. Ivanov ,	Prof., Dr. Habil., Berlin, Germany
K. P. Markov ,	Assoc. Prof., Ph.D., Aizu, Japan
Yu. A. Merkuriev ,	Prof., Dr. Habil., Academician of the Latvian Academy of Sciences, Riga, Latvia
R. V. Meshcheryakov ,	Prof., Dr. Sci., Tomsk, Russia
N. A. Moldovian ,	Prof., Dr. Sci., St. Petersburg, Russia
V. E. Pavlovskiy ,	Prof., Dr. Sci., Moscow, Russia
A. A. Petrovsky ,	Prof., Dr. Sci., Minsk, Belarus
V. A. Putilov ,	Prof., Dr. Sci., Apatity, Russia
V. K. Pshikhopov ,	Prof., Dr. Sci., Taganrog, Russia
A. L. Ronzhin	(Deputy Editor-in-Chief), Prof., Dr. Sci., St. Petersburg, Russia
A. I. Rudskoi ,	Prof., Dr. Sci., Academician of RAS, St. Petersburg, Russia
H. Samani ,	Assoc. Prof., Ph.D., New Taipei City, Taiwan, Province of China
V. Sgurev ,	Prof., Dr. Sci., Academician of the Bulgarian academy of sciences, Sofia, Bulgaria
V. Skormin ,	Prof., Ph.D., Binghamton, USA
A. V. Smirnov ,	Prof., Dr. Sci., St. Petersburg, Russia
B. Ya. Sovetov ,	Prof., Dr. Sci., Academician of RAE, St. Petersburg, Russia
V. A. Soyfer ,	Prof., Dr. Sci., Academician of RAS, Samara, Russia
B. V. Sokolov ,	Prof., Dr. Sci., St. Petersburg, Russia
L. V. Utkin ,	Prof., Dr. Sci., St. Petersburg, Russia
A. L. Fradkov ,	Prof., Dr. Sci., St. Petersburg, Russia
H. Kaya ,	Assoc. Prof., Ph.D., Tekirdag, Turkey
L. B. Sheremetov ,	Assoc. Prof., Dr. Sci., Mexico, Mexico

Editor: A. I. Motienko

Editor: E. P. Miroshnikova

Technical editor: M. S. Avstriyskaya

Translator: N. V. Kashina

Editorial Board's address

14-th line VO, 39, SPIIRAS, St. Petersburg, 199178, Russia,
e-mail: publ@ias.spb.su, web: <http://www.proceedings.spiiras.nw.ru/>

The journal is indexed in Scopus

© St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences, 2018

ТРУДЫ СПИИРАН

Выпуск № 5(60), 2018

Научный, научно-образовательный, междисциплинарный журнал с базовой специализацией в области информатики, автоматизации и прикладной математики
Журнал основан в 2002 году

Учредитель и издатель

Федеральное государственное бюджетное учреждение науки
Санкт-Петербургский институт информатики и автоматизации Российской академии наук
(СПИИРАН)

Главный редактор

Р. М. Юсупов, чл.-корр. РАН, д-р техн. наук, проф., С-Петербург, РФ

Редакционная коллегия

- А. А. Ашимов**, академик национальной академии наук Республики Казахстан д-р техн. наук, проф., Алматы, Казахстан
Н. П. Веселкин, академик РАН, д-р мед. наук, проф., С.-Петербург, РФ
О. Ю. Гусихин, Ph.D., Диаборн, США
В. Делич, д-р техн. наук, проф., Нови-Сад, Сербия
А. Б. Долгий, Dr. Habil., проф., Сент-Этьен, Франция
М. Железны, Ph.D., доцент, Пльзень, Чешская республика
Д. А. Иванов, д-р экон. наук, проф., Берлин, Германия
И. А. Каляев, академик РАН, д-р техн. наук, профессор, Таганрог, РФ
А. А. Карпов, д-р техн. наук, доцент, С.-Петербург, РФ
К. П. Марков, Ph.D., доцент, Аизу, Япония
Ю. А. Меркурьев, академик Латвийской академии наук, Dr. Habil., проф., Рига, Латвия
Р. В. Мещеряков, д-р техн. наук, профессор, Томск, РФ
Н. А. Молдовян, д-р техн. наук, проф., С.-Петербург, РФ
В. Е. Павловский, д-р физ.-мат. наук, профессор, Москва, РФ
А. А. Петровский, д-р техн. наук, проф., Минск, Беларусь
В. А. Путилов, д-р техн. наук, проф., Апатиты, РФ
В. Х. Пшихопов, д-р техн. наук, профессор, Таганрог, РФ
А. Л. Ронжин (зам. главного редактора), д-р техн. наук, проф., С.-Петербург, РФ
А. И. Рудской, академик РАН, д-р техн. наук, проф., С.-Петербург, РФ
Х. Самани, Ph.D., доцент, Синьбэй, Тайвань, КНР
В. Сгурев, академик Болгарской академии наук, д-р техн. наук, проф., София, Болгария
В. А. Скормин, Ph.D., проф., Бингемптон, США
А. В. Смирнов, д-р техн. наук, проф., С.-Петербург, РФ
Б. Я. Советов, академик РАО, д-р техн. наук, проф., С.-Петербург, РФ
В. А. Соифер, академик РАН, д-р техн. наук, проф., Самара, РФ
Б. В. Соколов, д-р техн. наук, проф., С.-Петербург, РФ
Л. В. Уткин, д-р техн. наук, проф., С.-Петербург, РФ
А. Л. Фрадков, д-р техн. наук, проф., С.-Петербург, РФ
Х. Кайя, Ph.D., доцент, Текирдаг, Турция
Л. Б. Шереметов, д-р техн. наук, Мехико, Мексика

Редактор: А. И. Мотиенко

Литературный редактор: Е. П. Мирошникова

Технический редактор: М. С. Австрийская

Переводчик: Н. В. Кашина

Адрес редакции

199178, Санкт-Петербург, 14-я линия, д. 39,
e-mail: publ@iias.spb.su, сайт: <http://www.proceedings.spiiras.nw.ru/>

Журнал индексируется в международной базе данных Scopus

Журнал входит в «Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени доктора и кандидата наук»

© Федеральное государственное бюджетное учреждение науки

Санкт-Петербургский институт информатики и автоматизации Российской академии наук, 2018
Разрешается воспроизведение в прессе, а также сообщение в эфир или по кабелю опубликованных в составе печатного периодического издания–журнала «Труды СПИИРАН» статей по текущим экономическим, политическим, социальным и религиозным вопросам с обязательным указанием имени автора статьи и печатного периодического издания–журнала «Труды СПИИРАН»

CONTENTS

Robotics, Automation and Control Systems

M. Negrete, J. Savage, L.A. Contreras-Toledo A MOTION-PLANNING SYSTEM FOR A DOMESTIC SERVICE ROBOT	5
V.Kh. Pshikhopov, M.Yu. Medvedev GROUP CONTROL OF AUTONOMOUS ROBOTS MOTION IN UNCERTAIN ENVIRONMENT VIA UNSTABLE MODES	39
E.Ya. Rapoport, Yu.E. Pleshivtseva METHOD OF MULTIOBJECTIVE OPTIMIZATION OF CONTROLLED SYSTEMS WITH DISTRIBUTED PARAMETERS	64
D.A. Ivanov, M.A. Ivanova, B.S. Sokolov ANALYSIS OF TRANSFORMATION TRENDS IN ENTERPRISE MANAGEMENT PRINCIPLES IN THE ERA OF INDUSTRY 4.0 TECHNOLOGY	97

Artificial Intelligence, Knowledge and Data Engineering

A.V. Kozachok, S.A. Kopylov, R.V. Meshcheryakov, O.O. Evsutin, L.M. Tuan AN APPROACH TO A ROBUST WATERMARK EXTRACTION FROM IMAGES CONTAINING TEXT	128
R.A. Neydorf, A.G. Aghajanyan DUAL OPTIMIZATION OF MONOCHROME IMAGES TONE APPROXIMATION USING PARALLEL EVOLUTIONARILY GENETIC SEARCH	156
R.D. Yoshinov, O.P. Iliev THE STRUCTURAL WAY FOR BINDING A LEARNING MATERIAL WITH PERSONAL PREFERENCES OF LEARNERS	189
S. Suzić, T.V. Delić, S. Ostrogonac, S. Đurić, D.J. Pekar STYLE-CODE METHOD FOR MULTI-STYLE PARAMETRIC TEXT-TO-SPEECH SYNTHESIS	216
A.A. Ashimov, A.S. Geida, I.V. Lysenko, R.M. Yusupov SYSTEM FUNCTIONING EFFICIENCY AND OTHER SYSTEM OPERATIONAL PROPERTIES: RESEARCH PROBLEMS, EVALUATION METHOD	241

СОДЕРЖАНИЕ

Робототехника, автоматизация и системы управления

М. Негрете, Х. Саваж, Л.Э. Контрерас-Толедо СИСТЕМА ПЛАНИРОВАНИЯ ДВИЖЕНИЯ БЫТОВОГО СЕРВИСНОГО РОБОТА НА ОСНОВЕ АЛГОРИТМОВ ПРОСТРАНСТВЕННОГО ПРЕДСТАВЛЕНИЯ И АКТИВНОЙ НАВИГАЦИИ	5
В.Х. Пшихолов, М.Ю. Медведев ГРУППОВОЕ УПРАВЛЕНИЕ ДВИЖЕНИЕМ МОБИЛЬНЫХ РОБОТОВ В НЕОПРЕДЕЛЕННОЙ СРЕДЕ С ИСПОЛЬЗОВАНИЕМ НЕУСТОЙЧИВЫХ РЕЖИМОВ	39
Э.Я. Рапопорт, Ю.Э. Плешивцева МЕТОД МНОГОКРИТЕРИАЛЬНОЙ ОПТИМИЗАЦИИ УПРАВЛЯЕМЫХ СИСТЕМ С РАСПРЕДЕЛЕННЫМИ ПАРАМЕТРАМИ	64
Д.А. Иванов, М.А. Иванова, Б.В. Соколов АНАЛИЗ ТЕНДЕНЦИЙ ИЗМЕНЕНИЯ ПРИНЦИПОВ УПРАВЛЕНИЯ ПРЕДПРИЯТИЯМИ В УСЛОВИЯХ РАЗВИТИЯ ТЕХНОЛОГИЙ ИНДУСТРИИ 4.0	97
<i>Искусственный интеллект, инженерия данных и знаний</i>	
А.В. Козачок, С.А. Копылов, Р.В. Мещеряков, О.О. Евсютин, Л.М. Туан ПОДХОД К ИЗВЛЕЧЕНИЮ РОБАСТНОГО ВОДЯНОГО ЗНАКА ИЗ ИЗОБРАЖЕНИЙ, СОДЕРЖАЩИХ ТЕКСТ	128
Р.А. Нейдорф, А.Г. Агаджанян ДУАЛЬНАЯ ОПТИМИЗАЦИЯ ТОНОВОЙ АППРОКСИМАЦИИ МОНОХРОМНЫХ ИЗОБРАЖЕНИЙ ПАРАЛЛЕЛЬНЫМ ЭВОЛЮЦИОННО-ГЕНЕТИЧЕСКИМ ПОИСКОМ	156
Р.Д. Йошинов, О.П. Илиев СТРУКТУРНЫЙ СПОСОБ И МОДЕЛЬ АВТОМАТИЧЕСКОЙ ГЕНЕРАЦИИ ПЕРСОНАЛИЗИРОВАННЫХ УЧЕБНЫХ МАТЕРИАЛОВ	189
С. Сузич, Т.В. Делич, С. Острогонац, С. Джурич, Д.И. Пекар МЕТОД СТИЛЕВЫХ КОДОВ ДЛЯ МНОГОСТИЛЕВОГО ПАРАМЕТРИЧЕСКОГО СИНТЕЗА РЕЧИ ПО ТЕКСТУ	216
А.А. Ашимов, А.С. Гейда, И.В. Лысенко, Р.М. Юсупов ЭФФЕКТИВНОСТЬ ФУНКЦИОНИРОВАНИЯ И ДРУГИЕ ОПЕРАЦИОННЫЕ СВОЙСТВА СИСТЕМ: ЗАДАЧИ И МЕТОД ОЦЕНИВАНИЯ	241

M. NEGRETE , J. SAVAGE , L. CONTRERAS -TOLEDO
**A MOTION-PLANNING SYSTEM FOR A DOMESTIC SERVICE
ROBOT**

Negrete M., Savage J., Contreras L. **A Motion-Planning System for a Domestic Service Robot.**

Abstract. Service robots are intended to help humans in non-industrial environments such as houses or offices. To accomplish their goal, service robots must have several skills such as object recognition and manipulation, face detection and recognition, speech recognition and synthesis, task planning and, one of the most important, navigation in dynamic environments. This paper describes a fully implemented motion-planning system which comprehends from motion and path planning algorithms to spatial representation and behavior-based active navigation. The proposed system is implemented in Justina, a domestic service robot whose design is based on the ViRBot, an architecture to operate virtual and real robots that encompasses several layers of abstraction, from low-level control to symbolic planning. We evaluated our proposal both in simulated and real environments and compared it to classical implementations. For the tests, we used maps obtained from real environments (the Biorobotics Laboratory and the Robocup@Home arena) and maps generated from obstacles with random positions and shapes. Several parameters were used for comparison: the total traveled distance, the number of collisions, the number of reached goal points and the average execution speed. Our proposal performed significantly better both in real and simulated tests. Finally, we show our results in the context of the RoboCup@Home competition, where the system was successfully tested.

Keywords: Autonomous navigation, behavior-based robotics, domestic service robots, path planning.

1. Introduction. According to the International Organization for Standardization a service robot is a robot that performs useful tasks for humans or equipment excluding industrial automation applications [13]. This norm also states that a robot, in general, requires a certain degree of autonomy which, in this context, is considered as the ability to perform intended tasks based on current state and sensing, without human intervention. The degree of autonomy in a service robot ranges from partial autonomy, which includes human-robot interaction, to full autonomy (without human intervention). Although by definition service robots do not need to be fully autonomous, the more degree of autonomy, the more useful the robot will be. For example, in space applications, due to the delay caused by the enormous distances, teleoperation is not a feasible solution, thus a fully autonomous robot is needed. In domestic applications, if a robot is going to clean a room, it should do it without human help. Thus, a high degree of autonomy is desired in a domestic service robot.

Domestic service robots need several skills: motion planning in dynamic environments, object recognition and manipulation, human-robot interaction, which can include human detection, gesture recognition and speech under-

standing and synthesis, task planning, real-time awareness, among others. The problem of motion planning refers to the robot's ability to localize itself, map an environment and plan secure paths from one point to another [6].

A motion planning system consists of many components that are interconnected and whose performance depend on other modules' performance. For example, following a path depends on the path planner and the low level control, and both depend on the localizer, but also the localizer depends on the control module because if robot moves abruptly or shows oscillations, localization will be more likely to fail. In principle, one can test each module in an isolated manner and the reliability of the whole system can be argued based on the reliability of each component. This is possible when only a feedforward interaction is present, nevertheless when feedback is involved, the whole system performance must be tested since it can differ from each module's performance.

As stated in [1], robotic competitions are useful to test and compare different algorithms and systems by providing an objective performance evaluation under controlled and replicable conditions. Robocup@Home [30] and RoCKIn@Home [1] are examples of competitions for domestic service robots to evaluate specific and general robot's performance. Contrary to experiments, where specific hypothesis are tested, a competition usually evaluates general abilities of robot systems. Testing a robot system in a competition also helps to find solutions to problems that only arise in fully integrated systems.

There is a large amount of work reported in the scientific literature about motion planning for autonomous robots, nevertheless, it is still an open and challenging problem when talking about robots navigating in real environments [5]. This can be seen, for example, in the Robocup@Home 2018 rulebook where the tasks of three tests are focused, among others, in safe navigation and obstacle avoidance. Autonomous navigation is developed to the point that there are open libraries to implement navigation systems, but these libraries commonly implement basic skills and a lot of work can be done, for example, for improving path planning in hard-to-navigate environments such as messy rooms. Robocup@Home rulebook also contemplates tests where the robot has to evade hard-to-see obstacles such as an apple (small object), glass or Lego brick. The fact that very recent competitions still pose challenges for testing autonomous navigation in domestic service robots shows that navigation is still an open problem.

This paper is organized as follows: In Section 2 we present the related work and highlight the novelties of our approach. Section 3 is dedicated to explain ViRBot, the architecture our domestic service robot is based on. In Sections 4 to 8 we explain in detail every module of the proposed motion planning. In Section 9 we describe the results of applying the proposed system in the con-

text of the tasks required by the international competitions Robocup@Home and we compare the performance with one of the most used packages for navigation. We state our conclusions and sketch the future work in Section 10.

2. Related Work.

2.1. Complete motion planning systems. Most of the literature in this topic reports isolated problems and experiments, such as mapping algorithms, path planning algorithms, localization or obstacle avoidance but there is only a few works reporting fully implemented and tested motion planning systems for domestic service robots. Examples of these works are the robots Markovito [4], LISA [26] and Cosero [27].

The work of [4] propose a navigation system based on a three-level software architecture: functional, execution and decision. This motion planning system uses dynamic programming for path planning, particle filters and geometric landmarks (corners and lines) for localization, occupancy grids for environment representation and include semantic information for a better motion planning. This motion planning system is based on a well defined architecture both, contrary the work presented in this paper, they do not use biology models of behavior.

The work of [27] is focused on the design and building of a cognitive service robot. To building a map, their motion planning system uses simultaneous localization and mapping based on Rao-Blackwellized particle filters and the Adaptive Monte Carlo Localization for estimating the robot's position. Authors of this work use 3D surfel grids from which they extract 2D navigation maps by exploring the traversability of surfels. When using a 2D sensors, they extract line segments from 2D scans and align them with the surfels in the map. Their software is also designed based on an architecture with four layers: task, subtask, action and perception, and sensori-motor.

The work of [26] developed a ROS-based autonomous service robot called LISA. Its motion planning system also uses particle filters for map building and localization. A* algorithm on the occupancy grid is used for path planning and the calculated path is post processed to get smooth trajectories. The smooth path is followed using waypoints along the planned path. Authors mention the close integration with the graphical user interface as the main benefit of their motion planning system. Contrary to this work, [26] do not mention the use of reactive behaviors to deal with unexpected obstacles.

2.2. Localization and Mapping. In order to fulfill their tasks, service robots need a representation of the environment (a map either geometric, topological or semantic), and its current configuration within such environment. The problem of building a map while estimating robot's position at the same time is called SLAM (Simultaneous Localization and Mapping) and is commonly

solved using Kalman [9] or particle filters [11]. More recent work has tackled this problem using cameras and data compression techniques such as [8] and [7]. To solve only the localization problem, one of the most used techniques in the development of service robots is the Adaptive Monte Carlo Localization [28] which is also used in this work.

Roadmaps are a class of topological maps that are useful for mobile robots' navigation in structured environments. There are several techniques for roadmap construction when a geometrical representation of the objects in the environment is available, for example, Voronoi diagrams [17], visibility maps [19] or probabilistic roadmap methods [15]. In this work we describe how we build roadmaps using data from RGB-D cameras and vector quantization techniques. A similar work is presented in [24]. Vector quantization has a very high computational cost and processing can be very slow. In this proposal computation time is significantly reduced by implementing such algorithms in parallel.

2.3. Obstacle Avoidance. Artificial Potential Fields [16] are a technique for path planning and obstacle avoidance. In this work, potential fields are used as a behavior and they are part of the behavior-based methods of the ViRBot architecture. Costmaps are also a technique commonly used for obstacle avoidance and path planning, for example, the work of [20] uses a scheme of layered costmaps to plan paths in dynamic environments. Nevertheless a costmap requires a map while artificial potential fields can be purely reactive. As it will be discussed later, costmaps may become ineffective when robot is navigating in narrow spaces.

2.4. The ROS package *nav2d*. Robot Operating System (ROS) [23] is an open source middleware which provides the functionality commonly needed in the development of software for autonomous mobile robots such as message passing and package management. ROS also has several packages with algorithms to implement the most common skills needed in an autonomous robot such as perception of objects and people, knowledge representation and navigation. *Nav2d* is an open source package for 2D navigation (<http://wiki.ros.org/nav2d>) whose main features, according to its web page, are a purely reactive obstacle avoidance, a simple path planner and a graph-based SLAM that allows multiple robots to cooperatively build a map. The last update was done in 2017 (according to the web page). This package was used to compare the performance of the proposed system.

3. ViRBot: an architecture for the operation of service robots. To develop domestic service robots a minimum hardware is required and also a design architecture must be followed in order to integrate the huge amount of software that is required to operate this kind of autonomous robots. To perform

the intended tasks, domestic service robots must have from low-level controls, to make actuators move according to a set point, to high-level task planning, to understand, plan and execute commands given by humans. Designing a domestic service robot requires much more than merging and coordinating top trending technologies in machine learning, computer vision, control, navigation, real time mapping, artificial intelligence, and human-computer interaction. A service robot interacts with an ever-changing environment in which the “optimal” conditions of a research laboratory are almost never met. For these reasons, a design architecture is mandatory for succeeding in the development of autonomous robots.

ViRBot is an architecture to design, organize, integrate and test software for autonomous service robots [25]. In the ViRBot architecture the operation of a service robot is divided in four general layers: Inputs, planning, knowledge management and execution, having each of them several subsystems, see Figure 1. Each subsystem has a specific function that contributes to the final operation of the robot. This architecture has similar features to the ones presented in the INTERRAP agent architecture [21].

Input layer. This layer encompasses all proprioceptive and exteroceptive robot’s sensors. Every sensor has also a simulation mode so that, when a virtual robot is used, algorithms can be tested using simulated sensor data. This layer also includes data coming from the human-robot interaction subsystem, which can be recognized voice or gesture commands. Digital signal processing techniques are applied to data coming from all sensors to obtain a symbolic representation of the environment and, with this representation, a series of beliefs are generated. A set of programmed tasks are also included in this layer since they establish a set of goals used to plan actions.

Planning layer. In cooperation with the Knowledge Management layer, beliefs are validated in the Planning layer and thus, a situation recognition is created. Given such recognized situation, a set of goals are activated and then, the action planner finds a sequence of physical operations to achieve the generated goals. The motion planner module is in charge of planning the movements of the actuators (mobile base and manipulators) to perform the sequence of physical operations generated by the action planner. If during the execution of a plan, something not considered in it happens, the exception recognizer module will try to solve the problem by setting new goals and re-planning the overall sequence of operations.

Knowledge Management layer. This layer builds and manages the robot’s knowledge both declarative and procedural. The cartographer module is in charge of three tasks: to keep a representation of the environment using geometric maps and roadmaps, to build such representation using SLAM and

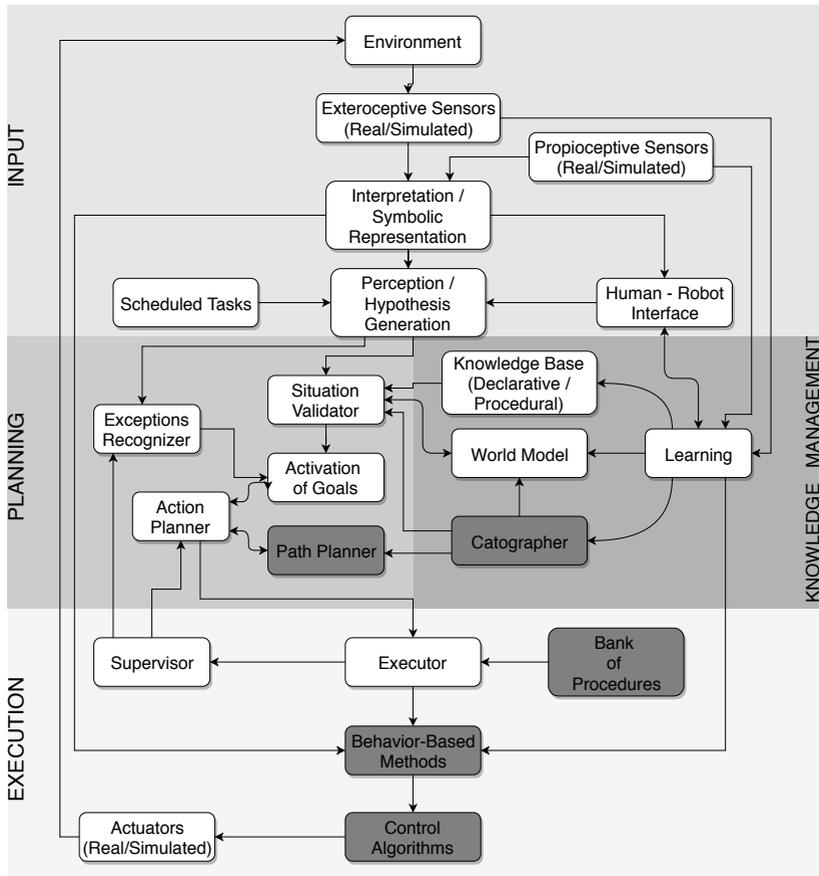


Fig. 1. The ViRBot architecture: several layers and subsystems are integrated to operate autonomous service robots. Highlighted modules are those involved in this paper

clustering algorithms, and to localize the robot using an Extended Kalman Filter and the Adaptive Monte-Carlo Localization. This layer also includes a knowledge base which keeps both procedural and declarative knowledge. The first one encodes the knowledge of an expert using a rule-based system written in CLIPS, a logic language developed by NASA. The declarative knowledge encompasses patterns of objects and faces, predefined locations and regions in the environment and positions of objects and persons. Learning algorithms are used to train patterns and to generate new behaviors.

Execution layer. The first module of this layer is the executor. It takes a set of hardwired procedures to transform plans generated by the action planner into simpler sequences of actions. The executor informs the result of the execution of each hardwired procedure to the supervisor, which checks that all actions and movements are performed according to the plans. The bank of hardwired procedures is a set of augmented finite state machines used to partially solve specific problems such as taking and object or ask and memorize a name. Once the executor generates simple sequence of actions, a set of behaviors are used to cope with problems not foreseen by the action planner, like the avoidance of unknown obstacles. Behaviors generate signals taken as desired values by the control algorithms and finally, the outputs of such low-level controls are used to move the actuators.

The ViRBot and the proposed system. The ViRBot architecture is an abstraction of the tasks and interactions between of the software used to operate a service robot. A program running on the robot can perform more than one task of the ViRBot and also a subsystem of the ViRBot can be implemented in more than one programs. The proposed system is part of the path planner, cartographer, human-robot interface, bank of procedures, behaviors and control algorithms. Figure 1 shows, highlighted in green, those subsystems involved in this proposal. In the following sections we describe each module of the proposed system and their corresponding parts in the ViRBot architecture.

4. Control Algorithms. In Sections 5 to 8 it is explained how the motion planning system plans paths, represents the environment and uses behaviors to avoid obstacles, nevertheless, all those algorithms will require control laws to guaranty that all planned motions will be performed correctly. For that purpose, we designed a non-linear control law.

Consider an omnidirectional mobile base like the one shown in Figure 2, where configuration is determined by three variables $[x_r, y_r, \theta_r]$. If only the kinematic part is considered and assuming there is no slip, the model of the

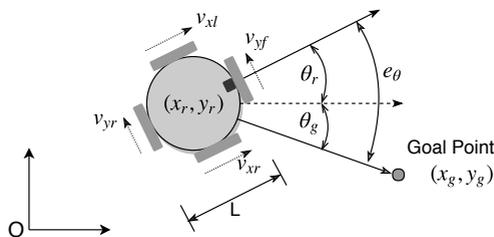


Fig. 2. Omnidirectional mobile base, the current and the goal position

robot is given by

$$\dot{x}_r = v_x \cos \theta_r - v_y \sin \theta_r, \quad (1)$$

$$\dot{y}_r = v_x \sin \theta_r + v_y \cos \theta_r, \quad (2)$$

$$\dot{\theta}_r = \omega, \quad (3)$$

where v_x , v_y and ω are the frontal, lateral and angular speeds, respectively, taken as input signals.

Note that the dynamic part of the robot is not modeled, i.e., the configuration $[x_r, y_r, \theta_r]$ is also considered as the state vector. Also, it is assumed that v_x , v_y and ω can be set arbitrarily. In the real system this is not the case, the real input signal is the voltage set in the motor terminals, nevertheless, we can assume that electrical and mechanical responses of the four motors are fast enough and thus, we can set any desired motor speed and it will be reached in a very small time.

The objective of the control laws is to determine signals v_x , v_y and ω such that it can be ensured that the robot will reach the goal position (x_g, y_g) even in the presence of uncertainties (such as non modeled dynamics) and perturbations. Although the mobile base is omnidirectional, in this work all movements are done taking it as a differential mobile base. Omnidirectionality is used in our service robot in other tasks where fine lateral movements are needed (e.g., when the robot needs to move laterally for grasping and object) but that is out of the scope of this work.

Consider the scheme shown in Figure 2. Let θ_g be the desired angle, which corresponds to the angle of the position error vector $[x_g - x_r, y_g - y_r]^T$, calculated as

$$\theta_g = \text{atan2}(y_g - y_r, x_g - x_r),$$

and then, angle error can be calculated as

$$e_{\theta} = \theta_g - \theta_r = \text{atan2}(y_g - y_r, x_g - x_r) - \theta_r. \quad (4)$$

It is important to note that e_{θ} , same as any other angular measure, must be always in the interval $(-\pi, \pi]$. If the difference given by (4) results in an angle greater than π or less than $-\pi$, it must be corrected so that $e_{\theta} \in (-\pi, \pi]$ and the correct performance of the control laws can be guaranteed.

To model an omnidirectional base as a differential one, it is enough to set $v_y = 0$. Assuming we have a mobile robot whose model is given by (1)-(3), then the control law

$$v_x = v_{max} e^{-\frac{e_{\theta}^2}{\alpha}}, \quad (5)$$

$$\omega = \omega_{max} \left(\frac{2}{1 + e^{-\frac{e_{\theta}}{\beta}}} - 1 \right), \quad (6)$$

ensures the robot will reach the goal position $(x_g, y_g)^T$. Given v_x , v_y and ω , the four motor speeds (left v_{xl} , right v_{xr} , front v_{yf} and rear v_{yr}) can be calculated as

$$\begin{aligned} v_{xl} &= v_x - \frac{L}{2}\omega, & v_{yf} &= v_y + \frac{L}{2}\omega, \\ v_{xr} &= v_x + \frac{L}{2}\omega, & v_{yr} &= v_y - \frac{L}{2}\omega, \end{aligned}$$

where L is the robot diameter (see Figure 2).

The control law (5)-(6) has four parameters: v_{max} and ω_{max} are the maximum linear and angular speeds, respectively, that the robot can reach during its movement. These parameters, in principle, can be set arbitrarily, nevertheless, in a real implementation, they are bounded by the actuators capabilities.

For a better understanding of constants α and β , consider the Figure 3. It can be seen that constant α determines how fast linear speed v decreases when error angle increases. A small α will make linear speed to decrease too fast, that is, robot will not move forward until it is pointing directly to the goal point. In other words, it will first turn until angle error is very small and then it will move towards the goal point.

Constant β determines how fast increases angular speed ω when error angle increases. In general, a small β will make the robot to keep always “point-

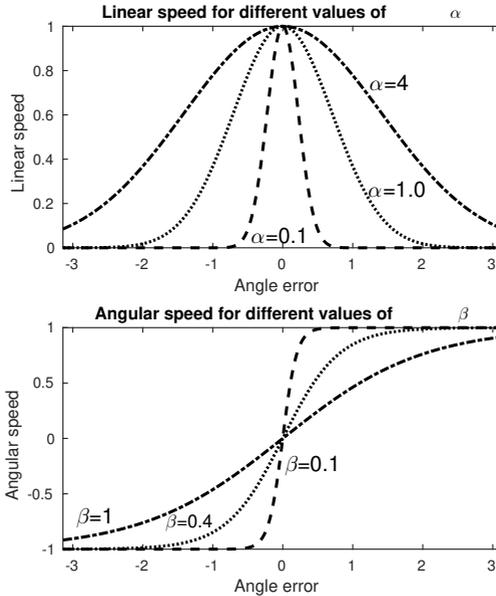


Fig. 3. Linear and angular speeds for different values of α y β

ing” to the goal point, i.e., it will help to a better path tracking, nevertheless, if β is too small, it can generate undesired oscillations.

5. Cartographer. Cartographer, in the ViRBot architecture, is part of the Knowledge Management layer. It does not make any movement plan but builds and keeps a representation of the environment. Using this representation and the information coming from sensors, it also estimates the robot configuration.

5.1. Environment Representation. Occupancy grids are a type of geometric maps in which the space is discretized with a given resolution. Each resulting cell is assigned to a number $p \in [0, 1]$ which indicates the occupation level. In the simplest interpretation, $p = 1$ means the cell is occupied and $p = 0$, the cell is free space. Occupancy grids can also be used with a probabilistic approach where p indicates the certainty that a given cell is occupied. In this work we use occupancy grids for representing the environment due to the ease to build them and the availability of standard formats to represent them such as the *OccupancyGrid* message in the platform ROS.

5.2. Localization. Adaptive Monte Carlo Localization is a probabilistic method used to determine robot’s position based on particle filters to track the 2D position against a known map. It is implemented in the open source ROS

package *amcl*. The theoretical base is explained in [28] and the implementation details are documented in the package web page. In this work, we used the *amcl* package to localize our domestic service robot.

5.3. Roadmap Construction using Clustering Algorithms.

Roadmaps are useful for mobile robots' navigation in structured environments. If such roadmaps are constructed with a fast enough sampling time and based on information extracted from robot's sensors, they can be used for obstacle avoidance. If a geometric representation is not available, it can be obtained by vector quantization (VQ) methods and, based on this representation, it is possible to build a roadmap. Roughly explained, the process for building such roadmaps is as follows: we process the point cloud acquired from a Kinect device to separate free and occupied space. Then, both the free and occupied space are clustered. Centroids and sizes of the occupied space clusters are used as a geometrical representation of the objects in the environment. Centroids of the free space are used as nodes of the roadmap.

Separation of Free and Occupied Space. RGB-D cameras provide information through an RGB image and a point cloud that represents the spatial position of each pixel of the captured image. This research uses only the spatial information which comes as a set R of triplets of the form $r_j = (x_{screen_j}, y_{screen_j}, d_j)$, where $(x_{screen_j}, y_{screen_j})$ is the pixel location in the image and d_j , the distance to the object on the line of sight. The point cloud S expressed in cartesian coordinates $s_j = (x_j, y_j, z_j)$ of the objects w.r.t. the RGB-D camera plane, can be obtained with the transformation

$$s_j = Mr_j, \quad (7)$$

with M , the matrix of intrinsic parameters of the RGB-D camera, which can be obtained by several methods (e.g., the work of [14]).

Since the camera is mounted on a mechatronic head with pan and tilt movements, it is convenient to apply a homogeneous transformation to express point with respect to the mobile base. After transformation, a point p_j is classified as free space if its z component is less than a constant K_h and, as occupied space, otherwise.

Vector Quantization. Vector Quantization [18] is commonly used for data compression in telecommunications and digital signal processing. In the field of robotics, it is also used to compress data and get a smaller but significant set of data. In this work, we use VQ to cluster the free and occupied space and, based on this clusters, to construct a roadmap.

Given a point cloud P , i.e., a set of n_v vectors $p_j = (x_j, y_j, z_j); j = 1, \dots, n_v$ that represent the position in space, a set of centroids that represent

these vectors is found. A collection of centroids is called a codebook and it is designed from a long training sequence that is representative of all vectors p_j to be encoded. The codebook is created with the Linde-Buzo-Gray (LGB) algorithm [18] as explained in Algorithm 1.

Algorithm 1. Vector Quantization by the Linde-Buzo-Gray algorithm

Require:

Point cloud $P = \{p_j | p_j = (x_j, y_j, z_j), j \in [1, n_v]\}$

L_d the desired codebook size

Ensure: Codebook $D = \{C_{L_m}\}$

$m \leftarrow 1$ //Current iteration

$L_m \leftarrow 1$ //Current number of centroids

//First centroid is the average of all points

$$C_1 \leftarrow \frac{1}{n_v} \sum_{j=1}^{n_v} p_j$$

$D_m \leftarrow \{C_1\}$ //Initial codebook

while $L_m < L_d$ **do**

for $\forall C_i \in D_m$ **do**

$\psi \leftarrow$ disturbance of small magnitude

 Obtain two new centroids by adding $\pm\psi$

end for

// D_{m+1} will now contain L_{m+1} new centroids

$$L_{m+1} \leftarrow 2L_m$$

//Set of clusters for each centroid

$$R \leftarrow \{R_k\}, k \in [1, L_{m+1}]$$

$$\bar{d}_t \leftarrow \infty$$
 //Average change in centroid values

while $\bar{d}_t > \varepsilon$ **do**

for $\forall p_j \in P$ **do**

 Assign p_j to the nearest cluster R_k whose corresponding centroid is

$C_k \in D_{m+1}$.

end for

for $\forall R_k \in R$ **do**

 Recompute centroid C_k by averaging all vectors $p_j \in R_k$

end for

 Calculate \bar{d}_t as the average distance between the current centroid C_k and its value in the previous iteration

end while

end while

The desired codebook size L_d (number of regions in the environment) is chosen with a tradeoff between computation time limitations for real time operation and the desired precision. In this work, we used $|\psi| = 0.01$, $\varepsilon = 0.03$ and $L_d = 64$. The LGB algorithm is applied to cluster both the free and occupied space and a total of 128 centroids are calculated, 64 for the free space and 64 for the occupied space. Figure 4, left and center, shows the resulting clusters.

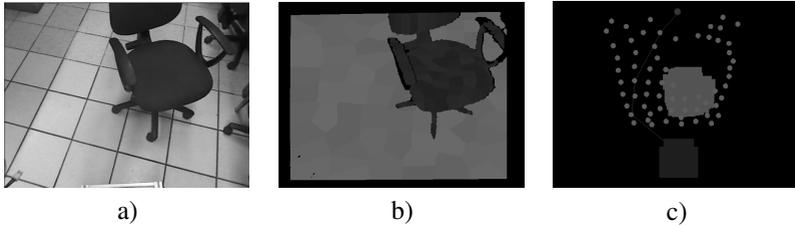


Fig. 4. a) original image captured by the Kinect; b) free space clusters are colored in light gray and occupied space clusters, in dark gray. The black regions are those points with no depth information; c) resulting environment representation. Dots represent the nodes (free space centroids) used to build the roadmap and calculate a path. Light gray rectangles represent obstacles from which paths are calculated by A* algorithm to reach a goal point

Roadmap Construction After the free and occupied spaces are separated and clustered, the roadmap is built following the Algorithm 2. In this algorithm $e(v, v')$ represents the edge between nodes v and v' and $Vis(v, v', p)$ is a function that determines if it is possible to reach the node v from node v' without crashing with the obstacle whose centroid is p . *NV* means Not Visible and function Vis return this value when node v is not visible from v' .

Algorithm 2. Roadmap construction from clusterized space

Require:

$P = \{p_i\}$ //Centroids of the quantized occupied space

$C = \{c_j\}$ //Centroids of the quantized free space

Ensure:

Roadmap $G(V, E)$ with

E : The edges of the roadmap

V : The nodes of the roadmap

$E \leftarrow \emptyset$

$V \leftarrow C$

$i \leftarrow 0$

```

for  $\forall v \in V$  do
  for  $\forall v' \in V$  do
    if  $e(v, v') \notin V \wedge \text{Vis}(v, v', p) \neq NV \forall p \in P$  then
       $E \leftarrow E \cup \{e(v, v')\}$ 
    end if
  end for
end for

```

Figure 4 shows the original image, the clusterized free and occupied space and the resulting roadmap. In general, clustering algorithms have a very high computational cost, which results in latencies that could make not possible to use them online as a method for obstacle avoidance. Nevertheless, in this work vector quantization was implemented in parallel allowing to get much shorter latencies. Details of this implementation are given in [22].

6. Path Planner. In the ViRBot architecture, the Path Planner is part of the planning layer. It takes information coming from the Action Planner and the environment representation kept by the Cartographer to calculate safe paths.

If environment is represented with an occupancy grid, the problem of planning a path can be solved by applying a search algorithm in graphs. In this case, each cell represents a node in the graph and it is considered to be connected only with the neighbor cells belonging to the free space. To determine the neighbor nodes, either four or eight-connectivity can be used.

A* is a search algorithm which explores the path with the lowest expected cost. For a given node n , the expected cost $f(n)$ is calculated as

$$f(n) = g(n) + h(n),$$

where $g(n)$ is the cost of the path from the start node until node n and $h(n)$ is an heuristic function which determines a cost that would be expected from node n to the goal node. The expected cost $h(n)$ must actually subestimate the real value, i.e., it must be hold that $h(n) \leq g(n) \quad \forall n \in \text{Graph}$.

The A* algorithm finds a path that is optimal with respect to the cost function $g(n)$. This function should be designed according to what is needed in the robot navigation: traveling the shortest distance, the fastest path or the most energy-efficient trajectory, for example. In this proposal, the risk of collision is taken as part of the cost function such that the robot will always navigate through the safest paths. To achieve this, we define the cost function as

$$g(n) = d(n) + r(n), \tag{8}$$

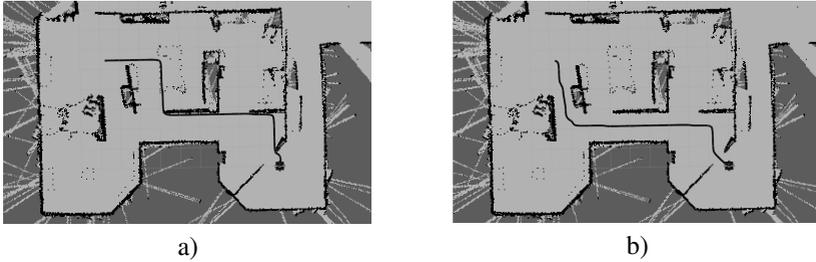


Fig. 5. Comparison of paths using: a) only distance as cost function; b) distance plus nearness to obstacles

where $d(n)$ is accumulated distance from the starting node to the current node n and $r(n)$ is a value inversely proportional to the distance from node n to the nearest obstacle. Figure 5 shows a comparison with and without the component $r(n)$ in the cost function. As can be seen in the upper Figure, if only distance is taken as a cost function, the path is calculated with positions next to the walls, since this is the shortest distance. Lower Figure shows the result considering the nearness to obstacles as a cost function and, as it can be seen, it is much safer for navigation.

A similar path can be obtained by augmenting the size of all obstacles in the map, nevertheless, with this approach the planner may be unable to calculate paths through small spaces. Costmaps are also a similar approach but, as it will be shown in the results section, they also have the problem of being unable to calculate paths through narrow spaces.

The heuristic function we use is the Manhattan distance from the current node to the goal node. After the path is calculated, it is post-processed to obtain a smooth path such that the control signals can also be smooth and avoid damages to the motors.

7. Behavior-Based Methods. Deliberative and reactive are two paradigms in robotics each one with advantages and disadvantages. Deliberative paradigm assumes an environment representation and this gives robots high prediction capabilities and thus, more complex tasks can be solved. Since the robot actuates based on a knowledge representation, the response is slow and in general it is difficult to handle fast changes in the environment. In the opposite side, the reactive paradigm does not depend on a knowledge representation and its response is faster. Reactive paradigm is better to handle high dynamic environments but their prediction capabilities are poor.

Behavior-based robotics states that intelligence in robots can be achieved as an emergent property of a large enough set of behaviors interacting with

each other [3]. A behavior is a pair of stimulus-response and, since there is no symbolic planning, response is fast and computational cost is low. Since each behavior represents a “direct connection” of sensors with actuators, it is possible to have several behaviors trying to control the robot’s actuator. A solution for this problem is the use of an arbiter which can make a selection of only one behavior, based on a hierarchy or the amplitude of the response, or it can produce a response by weighting all behaviors responses.

Hybridizing paradigms allows to take the advantages of each one and that is the case of the ViRBot architecture. Environment representation and planning algorithms allows the robot to perform more complex movements while behavior-based methods allows it to deal with changes in the environment such as obstacles not considered in the map. In our work, once a path is planned, it is executed with a behavior-based approach. There are three behaviors running concurrently: *go-to-goal-point* behavior, *avoid-obstacle* behavior and *collision-risk* behavior. To assemble all behaviors, we use a priority-based arbiter where the first behavior has the lowest priority and the last one, the highest priority.

7.1. The Go-To-Goal-Point Behavior. This behavior takes as input signals the desired positions calculated by the path planner (see Section 6) and the robot’s position estimated by the localization algorithms (see Section 5). The output of this behavior are the tire speeds calculated according to the control laws explained in Section 4. This behavior is considered to be activated any time a new path is calculated.

7.2. The Avoid-Obstacles Behavior. We use the potential fields method as a behavior to avoid unexpected obstacles. The potential fields method proposes to design a potential function $U(q) : R^n \rightarrow R$, where q is the robot position, such that it has a minimum in the goal point and local maxima in every obstacle’s position. If U is taken as a potential energy function, then its gradient $\nabla U(q)$ is a force vector whose direction points to the direction of maximum change. If the robot moves always in the opposite direction to the gradient (a method called gradient descent), then it will stop its movement in a local minimum. Thus, if the potential function is correctly designed (with a minimum in the goal point and local maxima in every obstacle) and the robot moves following the gradient descent, then it will reach the goal point evading obstacles at the same time [16].

There are several methods to design the potential function U and one of the most used is the attractive and repulsive field method. Since the movement will be guided by the gradient ∇U and not the function U , it is easier to design directly ∇U , i.e., to design the attraction and rejection forces. In this work we use the forces proposed by [2] given by

$$F_a = \zeta \frac{(q - q_g)}{\|q - q_g\|}, \quad \zeta > 0, \quad (9)$$

$$F_r = \begin{cases} \eta \left(\sqrt{\frac{1}{d} - \frac{1}{d_0}} \right) \frac{q_{oi} - q}{d} & , \text{ if } d < d_0, \\ 0 & , \text{ otherwise,} \end{cases} \quad (10)$$

where q is the current robot position, q_g is the goal position, q_{oi} is the position of obstacle i and $d = \|q_{oi} - q\|$. These forces have three design parameters $\zeta > 0$, $\eta > 0$ and $d_0 > 0$. Constant ζ defines how big is the attraction force and constant η , how big is the rejection force. d_0 is called the influence distance. If the distance between an object and the robot is greater than d_0 , then that object will not affect the robot's movement.

The resulting force is calculated as the sum of the attractive force and the mean of all repulsive forces:

$$F = F_a + \frac{1}{n} \sum_{i=1}^n F_r, \quad (11)$$

Once calculated the gradient, we move the robot using the Algorithm 3 (gradient descent). Constant $\alpha > 0$ must be small enough to avoid oscillations but also considering that a too small α will imply a high computational cost.

Algorithm 3. Gradient descent to move robot through a potential field

Require: Initial position q_s , goal position q_g , obstacle positions q_{oi}

Ensure: A sequence of points $\{q_0, q_1, q_2, \dots\}$

$q_0 \leftarrow q_s$

$i \leftarrow 0$

while $\|\nabla U(q_i)\| > 0$ **do**

$q_{i+1} \leftarrow q_i - \alpha \nabla U(q_i)$

$i \leftarrow i + 1$

end while

Finally, to move the robot towards each point q_i we use the control laws given by (5)-(6). Designing the potential function U using attractive and repulsive fields has the problem of local minima, i.e., the robot can stop in a point (local minimum) because gradient ∇U at that point is zero, but it is not the goal point. Nevertheless this problem can be overwhelmed if behaviors are assembled correctly, as it will be explained in the following subsections.

It is worth to note that, although Algorithm 3 assumes all obstacle positions are known, it is not necessary to have an environment representation

with all obstacle positions. Since the robot has a laser sensor, every reading can be regarded as an obstacle and thus potential fields can be calculated directly from robot's sensors.

7.3. The Collision-Risk Behavior. The goal of this behavior is to stop the robot if a risk of collision is detected. Since the laser sensor makes only 2D scans, it is not possible to detect obstacles that do not cross the plane of readings, that's why we use the RGB-D camera mounted on the robot's head. The criterium for determining if there is a risk of collision is the detection of an obstacle in front of the robot but only if the robot is going to move forward. To determine the presence of an obstacle, we count the number of pixels in the RGB-D image that have a Z-value greater than zero and are within a rectangular area in front of the robot; if the counting exceeds a threshold K_{obs} , then there is an obstacle in front of the robot. Consider the Figure 6. The yellow square in front of the robot is the detection region and is defined by the bounding box $[(x_{min}, y_{min}), (x_{max}, y_{max})]$. Collision risk detection is performed according to Algorithm 4.

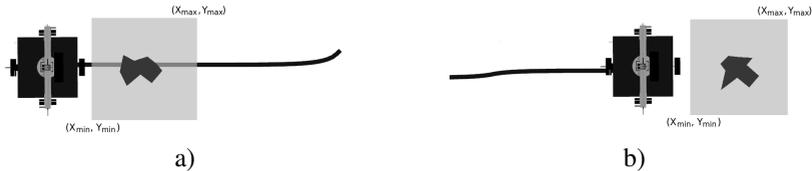


Fig. 6. a) risk of collision is detected since the obstacle is in the robot's path; b) obstacle is not in the path of the robot, thus, no risk of collision is detected

Algorithm 4. Detection of collision risk

Require:

Point cloud $P = \{p_j | p_j = (x_j, y_j, z_j), j \in [1, n]\}$

Control signals v and ω

Minimum number of points $K_{obstacle}$

Minimum linear speed v_{fwd} to consider the robot is moving forward

Ensure: Collision risk flag

$i \leftarrow 0$

Collision risk \leftarrow False

for $\forall p_j \in P$ **do**

if $x_j > x_{min} \wedge x_j < x_{max} \wedge y_j > y_{min} \wedge y_j < y_{max}$ **then**

$i \leftarrow i + 1$

end if

end for

```

if  $v > v_{fwd} \wedge i > K_{obstacle}$  then
//The number of points is greater than a threshold and the robot is moving
forward
    Collision risk  $\leftarrow$  True
end if

```

The output of this behavior is always zero since it is intended to stop the robot in case of a risk of collision, thus, this behavior is considered activated any time the conditions described in Algorithm 4 are hold.

7.4. The Arbiter. Figure 7 shows a block diagram of the behavior based movement. Arbiter takes the activated behavior with the highest priority. The *go-to-goal-point* behavior activates any time a new path is planned. *Avoid-obstacles* behavior activates when the calculated repulsive force is greater than zero and the *collision-risk* behavior, when a collision risk is detected. With this scheme, robot is able to avoid unexpected obstacles and also the problem of local minima in potential fields is overwhelmed since, every time an obstacle is in front of the robot (possible causing local minima), the robot is stopped and a new path is calculated using a map built as described in Section 5.3.

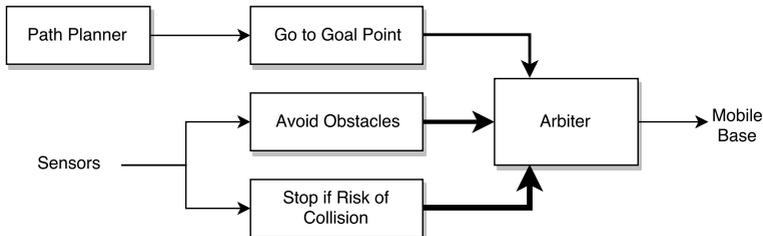


Fig. 7. The priority-based arbiter. Arrow thickness indicate the behavior priority

8. Bank of Procedures. In the ViRBot architecture, the bank of procedures is a set of state machines used to perform simple tasks such as taking an object (once position has been determined using computer vision algorithms), asking and memorizing a name and following a path, among others.

The control law (5)-(6) has the disadvantage of depending only on the angle error, i.e., robot does not decelerates as it approaches to the goal point. This will cause the robot to have strong oscillations when it is in a small region around the goal point. There are two ways to address this problem. The easiest one is to execute the control law only if the distance to the goal point $d = \sqrt{(x_g - x_r)^2 + (y_g - y_r)^2}$ is greater than a tolerance ε :

$$\begin{bmatrix} v_l \\ v_r \end{bmatrix} = \begin{cases} \begin{bmatrix} v + \frac{L}{2}\omega \\ v - \frac{D}{2}\omega \end{bmatrix}, & \text{if } d > \varepsilon, \\ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, & \text{otherwise.} \end{cases}$$

In this case the robot will stop abruptly when the error distance is less than ε , which is not desirable since the robot should decelerate slowly as it approaches to the goal point. It is also desirable that robot increases its linear speed slowly at the beginning of the movement, which is not yet considered, because according to (5)-(6), the robot can have a linear speed of v_{max} since the beginning of the movement.

In the other hand, control law (5)-(6) is designed to reach a goal point, but what we have to follow is not a point but a path composed by a sequence of points. Acceleration and deceleration can be achieved by parameterizing with respect to time the full path. This is commonly done by adjusting a fifth degree polynomial, or a higher degree if more variations of speed and acceleration are required. Nevertheless, expressing positions x and y as a function of time is easy when only straight-line movements are required, but in the case of a service robot moving in a domestic environment, paths can be too complex and parameterization with respect to time can become unfeasible.

Augmented Finite State Machine. An easy way to make the robot to accelerate and decelerate, that is, to get a speed profile, is through an Augmented Finite State Machine (AFSM) that establishes the value of v_{max} in control law (5)-(6). Let v_{sm} be the new maximum speed, such that

$$v = v_{sm} e^{-\frac{e_\theta}{\alpha}}, \quad (12)$$

$$\omega = \omega_{max} \left(\frac{2}{1 + e^{-\frac{e_\theta}{\beta}}} - 1 \right). \quad (13)$$

Control given by (12)-(13) is designed to reach a goal point (x_g, y_g) , nevertheless, what we want is to track a path given by N points (x_i^p, y_i^p) $i \in [0, N - 1]$. To follow this path it is enough to set as a goal point, the point (x_i^p, y_i^p) which is “in front of” the robot, that is, a point in the path whose position is a small distance away from the current position. In this work, we consider “small distance” a distance of 0.3 [m].

Figure 8 shows the AFSM used to determine v_{sm} and the goal point. In general, v_{sm} is increased slowly until it reaches the final value of v_{max} and it

starts to decrease when robot is “near” the last point of the path, that is, when distance r is less than a constant r_d . When distance to current goal point is less than 0.3 [m], then it is changed by the next point of the path. Values of r_d and Δv are determined depending on how fast the robot should accelerate and decelerate.

9. Results.

9.1. The Service Robot Justina. Justina is a domestic service robot built at the Biorobotics Laboratory of the National Autonomous University of Mexico and developed under the ViRBot architecture (see Section 3). This robot and its predecessors have been participating in the Robocup@Home league [30] since 2006 performing several tasks such as cleaning a table, serving drinks and several other tasks that humans ask for. To sense the environment, Justina has several sensors: two laser range finders, an RGB-D camera, a directional microphone, an array of omnidirectional microphones and encoders in each motor. The actuators of Justina consist of an omnidirectional mobile base, a 2-DOF head (pan and tilt movements) where the RGB-D camera and the directional microphone are mounted, two 7-DOF anthropomorphic manipulators and a 1-DOF torso to modify Justina’s height. Figure 9 shows robot Justina and the location of its sensors and actuators.

As mentioned in Section 2, the *nav2d* package was used as a baseline to compare the performance of our proposal. Experiments were conducted in real and simulated environments as follows.

9.2. Simulation Results. To test the performance of our proposal and compare it with the *nav2d* package, we generated 10 artificial maps consisting, each one, of polygons with random shapes and positions. Figure 10 shows examples of the artificial maps and the random points. We also used 2 real maps: The Biorobotics Laboratory and the @Home arena of the Robocup 2018 (shown in Figure 13). The *nav2d* package was used with all constants and parameters in their default values, except for the maximum linear speed, which was set to 0.7 to be the same than the value used with the control laws (5)-(6), and the *map inflation radius*, whose value of 0.2 was selected as the greatest value that allowed *nav2d* to calculate a path through the doors of the environment.

The low-level control was implemented using as constants $\alpha = 0.6$, $\beta = 0.09$, $\omega_{max} = 1.0$ and $V_{max} = 0.7$. Path planning was implemented according to the algorithms described in Section 6. For RoC detection we used $x_{min} = 0.3$, $y_{min} = -0.25$, $x_{max} = 0.9$, $y_{max} = 0.25$, $K_{obstacle} = 30$, $K_{\Pi} = 0.05$ and $v_{fwd} = 0.1$. Potential fields used the constants $\zeta = 1.0$, $\eta = 5.0$ and $d_0 = 0.8$.

The baseline (*nav2d* package) and the proposal were tested with 100 random goal points for each map. Comparison was made base on three param-

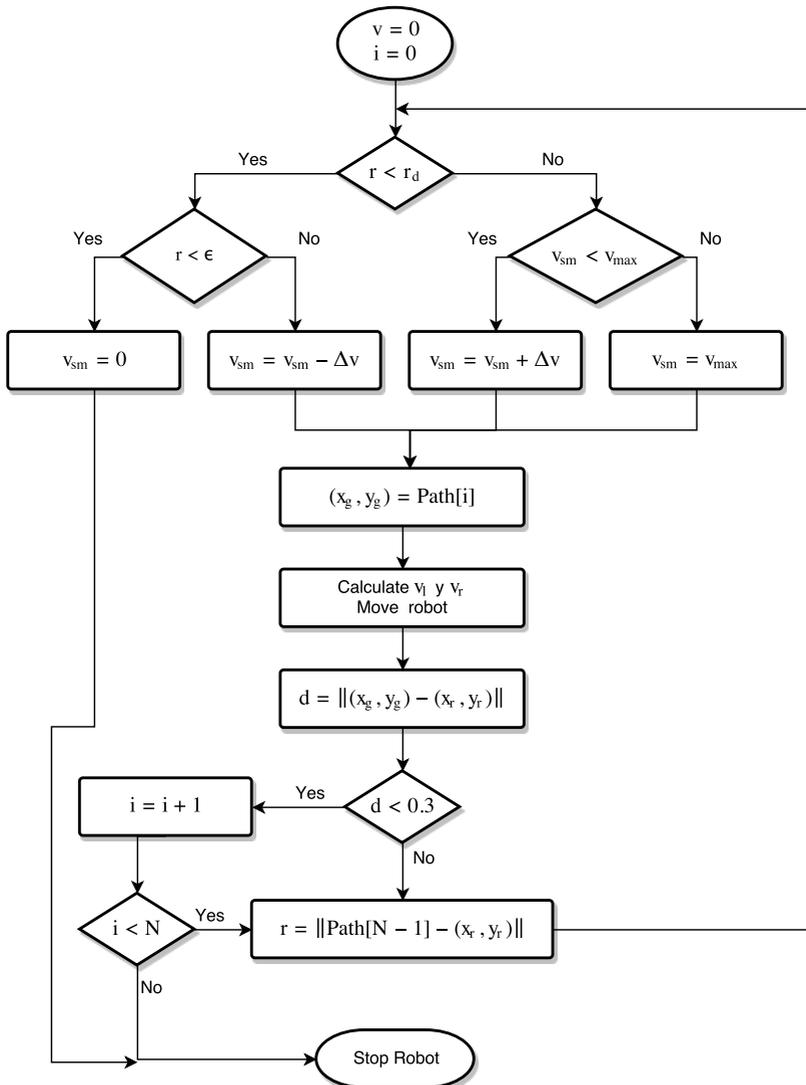


Fig. 8. Augmented Finite State Machine used to generate a speed profile

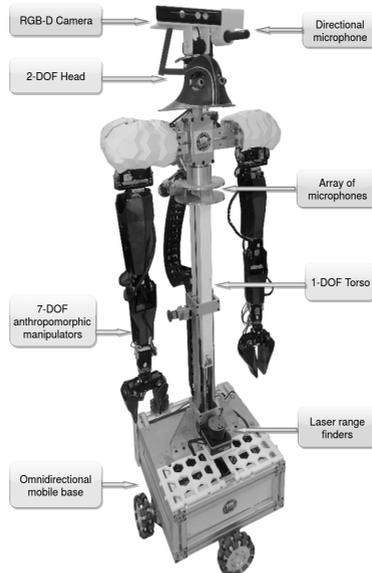


Fig. 9. The domestic service robot Justina

eters: **TDEDR**. Traveled Distance to Euclidean Distance Ratio. Calculated as the total traveled distance divided by the euclidean distance from the start to the goal point. We consider that the longer the path traveled, the less efficient the moving planning system. **AMpS**. Average Meters per Second. Calculated as the total traveled distance divided by the total time spent to reach the goal. This measure can give an insight of how fast the robot moves towards the goal. **NoC**. The total number of collisions (NoC) with any part of the map. Since these tests were simulated, we counted the times the robot touched any part of the map.

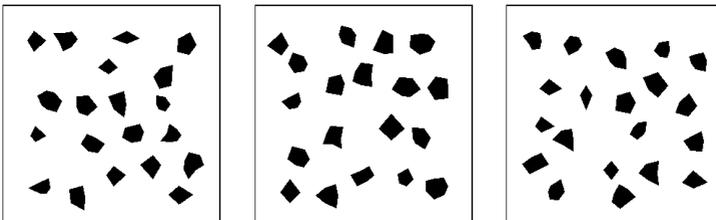


Fig. 10. Examples of the worlds used to test path planning and the random goal points

Table 1 shows the mean and standard deviation of the TDEDR for the 100 random goal points for each world. In order to determine if there is a significant difference between both navigation systems, a two-tailed t-Student test was performed. Last columns show the statistic and the p-value. In a similar way, tables 2 and 3 show the means, standard deviations and t-Student tests for AMpS and NoC, respectively.

Table 1. Ratio of the Traveled Distance to the Euclidean Distance (TDEDR) for our proposal and the baseline (the *Nav2d* package) for simulated scenes. One-hundred random paths were evaluated per map

World	Proposal		Nav2d		Significance	
	Mean	SD	Mean	SD	t	p-value
Biorobotics L.	1.29	0.39	1.18	0.20	2.45	0.01534
@Home arena	1.56	0.98	1.19	0.18	3.73	0.00030
Random 1	1.31	0.35	1.12	0.20	4.41	1.88E-5
Random 2	1.26	0.27	1.13	0.22	3.76	0.00022
Random 3	1.30	0.24	1.10	0.13	7.55	3.4E-12
Random 4	1.31	0.26	1.11	0.15	6.62	5.5E-10
Random 5	1.27	0.22	1.13	0.21	4.43	1.54E-5
Random 6	1.24	0.21	1.11	0.17	4.73	4.22E-6
Random 7	1.24	0.18	1.10	0.13	6.19	3.83E-9
Random 8	1.22	0.27	1.15	0.22	1.96	0.05032
Random 9	1.22	0.27	1.13	0.26	2.44	0.01540
Random 10	1.26	0.22	1.08	0.14	7.01	5.6E-11

Table 1 shows that TDEDR was greater for the proposal, i.e., distance traveled using the *nav2d* package was significantly less than the distance using the proposed navigation system in 11 out of 12 tested worlds (using 95% of confidence). This is due to the cost function described in Section 6. Since the nearness to obstacles is taken as part of the cost function, paths are in general longer because robot will tend to be away from all obstacles. Figure 11 shows a comparison of the paths calculated by Justina and by the *nav2d* package. As mentioned in Section 6, safer paths can be obtained by the *nav2d* package by incrementing the obstacle inflation radius, but doing so, disables the package to calculate paths through small doors. Despite the longer paths, Justina reached the goal points in shorter times, as it can be seen in table 2, which shows that the Average Meters per Second (AMpS) was significantly greater for the proposed system than the *Nav2d* package.

For both navigation systems, we counted the NoC (any touch to some part of the map was considered a collision). Table 3 shows the statistics of NoC for the 100 goals for every map. As it can be seen, NoC with Justina's system was significantly lower for all maps, as a result of the safer paths calculated by the path planner. Figure 11 shows a comparison of the paths calculated by the

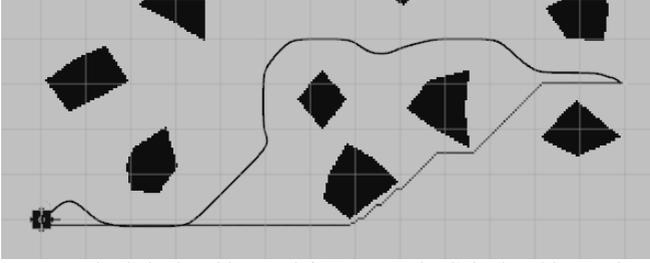


Fig. 11. Lower path: Calculated by *nav2d*. Upper path: Calculated by Justina. Justina's path is safer

Table 2. Average Meters per Second (AMpS) for our proposal and the baseline (the *Nav2d* package) for simulated scenes. One-hundred random paths were evaluated per map

World	Proposal		Nav2d		Significance	
	Mean	SD	Mean	SD	t	p-value
Biorobotics L.	0.43	0.107	0.20	0.042	20.08	2.2E-16
@Home arena	0.43	0.109	0.19	0.048	19.62	2.2E-16
Random 1	0.41	0.110	0.20	0.046	17.98	2.2E-16
Random 2	0.43	0.099	0.20	0.040	21.61	2.2E-16
Random 3	0.46	0.091	0.20	0.033	22.81	2.2E-16
Random 4	0.43	0.092	0.21	0.047	21.73	2.2E-16
Random 5	0.42	0.086	0.20	0.037	23.42	2.2E-16
Random 6	0.43	0.090	0.21	0.036	23.14	2.2E-16
Random 7	0.44	0.091	0.20	0.034	24.64	2.2E-16
Random 8	0.42	0.113	0.20	0.035	18.37	2.2E-16
Random 9	0.42	0.111	0.20	0.046	18.32	2.2E-16
Random 10	0.43	0.089	0.21	0.036	23.09	2.2E-16

baseline and the proposal. Justina's system gets safer paths which results in less number of collisions.

9.3. Results with the real robot. Similar to the simulated results, we tested our proposal and the *nav2d* package with 20 points but only in the Biorobotics Laboratory. For security reasons, once we detected the robot is going to crash or even touch some obstacle, we stopped the robot, thus, contrary to the simulated results, we don't have a measure of the number of collisions, instead, we have a Number of Reached Goal Points (**NoRGP**). For all goal points, we put some obstacles in front of the robot to check the performance for obstacle avoidance. All parameters and constants were the same than those used in the simulated results.

Of the 20 goal points, Justina reached 18 goals without crashing or touching the unexpected obstacles using the proposed system. When using the

Table 3. Number of Collisions (NoC) for our proposal and the baseline (the *Nav2d* package) for simulated scenes. One-hundred random paths were evaluated per map

World	Proposal		Nav2d		Significance	
	Mean	SD	Mean	SD	t	p-value
Biorobotics L.	0.22	0.48	1.96	1.76	-9.51	3.7E-16
@Home arena	0.43	0.97	2.70	1.82	-11.00	<2.2E-16
Random 1	0.10	0.30	1.88	1.66	-10.55	<2.2E-16
Random 2	0.08	0.31	2.19	1.61	-12.90	<2.2E-16
Random 3	0.13	0.34	1.91	1.62	-10.74	<2.2E-16
Random 4	0.08	0.27	2.27	1.68	-12.86	<2.2E-16
Random 5	0.05	0.22	1.85	1.59	-11.20	<2.2E-16
Random 6	0.07	0.25	2.08	1.64	-12.12	<2.2E-16
Random 7	0.15	0.48	2.40	1.73	-12.50	<2.2E-16
Random 8	0.14	0.38	2.43	2.10	-10.71	<2.2E-16
Random 9	0.15	0.39	1.99	1.50	-11.82	<2.2E-16
Random 10	0.07	0.26	1.93	1.68	-10.92	<2.2E-16

Table 4. Statistics of the Traveled Distance to Euclidean Distance Ratio (TDED R) and the Average Meters per Second (AMpS) for the proposal and the baseline in the real environment. Last row shows the Number of Reached Goal Points (NoRGP)

Param	Proposal		Nav2d		Significance	
	Mean	SD	Mean	SD	t	p-value
TDED R	1.06	0.337	1.36	0.365	-2.6376	0.01253
AETPM	0.19	0.052	0.17	0.054	1.4153	0.1661
NoRGP	18		14			

nav2d package, it only reached 14 goal points. Figure 12 shows the performance of the behavior-based approach for obstacle avoidance. The upper Figure shows the robot at the beginning of the movement. At this point, *go-to-goal-point* and *avoid-obstacles* behaviors are activated. The center Figure shows the moment when a collision risk is detected and the robot is stopped. Although a person was in front of the robot, what could cause a local minimum, robot was able to reach the goal due to the activation of the *collision-risk* behavior and the calculation of a new path, as shown in the lower Figure.

Table 4 shows the statistic data for the TDED R and AMpS for both navigation systems. As it can be seen, in the real world with unexpected obstacles the mean distance traveled using the *nav2d* package was significantly longer than the distance with the proposed system. This means that the proposed system avoids objects more efficiently since it travels a shorter distance. The AMpS had no significant difference for both systems, which is actually a good indicator: the proposed system traveled shorter paths which, at the same average speed, means shorter travel times.

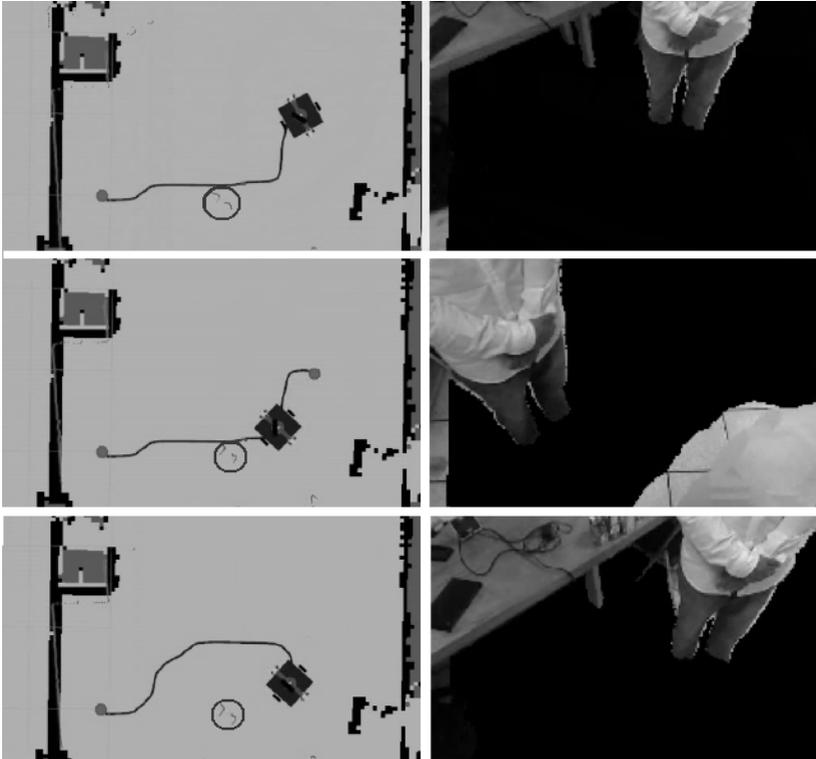


Fig. 12. Behavior-based obstacle avoidance. Top: justina starts its movement. Center: a risk of collision is detected. Bottom: a new path is calculated

9.4. Service Robots Competitions. To promote and foster development of autonomous robots, there are international competitions, such as Robocup [10] and RoCKIn [1]. Both competitions have the advantage of providing a standard problem where a wide range of technologies and approaches can be integrated, examined and evaluated. As stated in [12], competitions are useful when it is necessary to evaluate the general performance of a whole system and not only isolated hypothesis. For these reasons, robot Justina and its predecessors have been participating in the Robocup@Home since 2006. The motion-planning system described in this paper was tested in the last Robocup@Home edition held in Montreal, Canada, where Justina achieved the second place. In the @Home league, navigation was necessary in 6 of 7 tests [29]. In all of them, Justina successfully performed all the required motions. Figure 13 shows the map of the @Home arena where the proposal was tested.

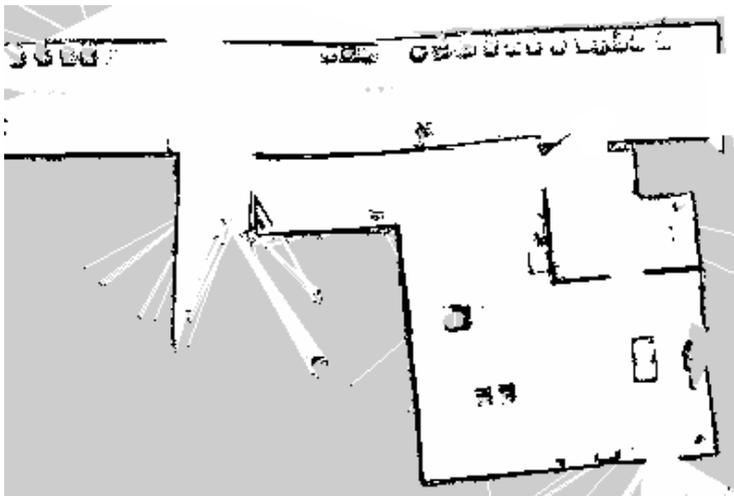


Fig. 13. The @Home arena of the Robocup 2018. Even though a clear drift is observed in the map during the slam process, our system performs well due to the active navigation module

10. Conclusions. We built a motion-planning system following the guidelines proposed by the ViRbot architecture. The motion planning was implemented in the service robot Justina and was tested both in real and simulated worlds. To evaluate the performance, we compared it with the nav2d package taking the Traveled Distance to Euclidean Distance Ratio (TDED), Average Meters per Second (AMpS), Number of Collisions (NoC) and Number

of Reached Goal Points (NoRGP) as evaluation parameters. Distances traveled using the proposed system were significantly longer, in the simulated test, than the distances traveled using the *Nav2d* package, according to the TDED. This is due to the cost function used in the A* algorithm which results in longer but safer paths. Safety of paths was verified by the significantly lower Number of Collisions (NoC) when using our motion-planning system, in the simulation tests, and the greater Number of Reached Goal Points (NoRGP), in the real experiments. Contrary to the simulation results, in the experiments with the real robot, distances using our proposal were significantly shorter. This is due to the presence of unexpected obstacles, i.e., the proposed system avoided obstacles more efficiently.

As stated in the introduction of this work, testing whole systems is important when there is a feedback between the involved subsystems and thus, the reliability of each module does not guarantee the reliability of the whole system. In this work, we implemented several algorithms already described in the literature but we integrated them following the ViRBot architecture and we tested the performance of the whole system in the context of the Robocup@Home competition, where navigation was required in 6 of the 7 tests. Justina was able to successfully perform all the required motions.

As future work, we plan to make a more extensive comparison with other service robot architectures in similar conditions by promoting open source access and research collaboration.

References

1. Amigoni F. et al. Competitions for benchmarking: task and functionality scoring complete performance assessment. *IEEE Robotics&Automation Magazine*. 2015. vol. 22(3). pp. 53–61.
2. Arambula F., Padilla M.A. Autonomus robot navigation using adaptive potential fields. *Mathematical and Computer Modelling*. 2011. Vol. 40. pp. 1141–1156.
3. Arkin R.C. Behavior-based robotics. MIT Press. 1998. 491 p.
4. Avilés-Arriaga H.H. et al. Markovito: A Flexible and General Service Robot. Springer. 2009. pp. 401–423.
5. Banino A. et al. Vector-based navigation using grid-like representations in artificial agents. *Nature*. 2018. vol. 557. no. 7705. pp. 429.
6. Choset H.M. et al. Principles of Robot Motion. MIT Press. 2005. 626 p.
7. Contreras L., Mayol-Cuevas W. Trajectory-driven point cloud compression techniques for visual slam. IEEE/RSJ International Conference on Robots and Systems (IROS). 2015. pp. 133–140.
8. Contreras L., Mayol-Cuevas W. O-poco: Online point cloud compression mapping for visual odometry and slam. IEEE/RSJ International Conference on Robots and Systems (IROS). 2017. pp. 4509–4514.
9. Dissanayake M.W.M.G. et al. A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions on robotics and automation*. 2001. vol. 17(3). pp. 229–241.

10. Ferrein A., Steinbauer G. 20 years of robocup. *KI-Künstliche Intelligenz*. 2016. vol. 30(3-4). pp. 225–232.
11. Grisetti G., Stachniss C., Burgard W. Improved techniques for grid mapping with rao-blackwellized particle filters. *IEEE transactions on Robotics*. 2007. vol. 23(1). pp. 34–46.
12. Iocchi L. et al. RoboCup@ Home: Analysis and results of evolving competitions for domestic and service robots. *Artificial Intelligence*. 2015. vol. 229. pp. 258–281.
13. Robots and robotic devices - vocabulary. ISO 8373-201. International Organization for Standardization. Geneva. Switzerland. 2012.
14. Jin B., Lei H., Geng W. Accurate intrinsic calibration of depth camera with cuboids. European Conference on Computer Vision. 2014. pp. 788–803.
15. Kavraki L., Svestka P., Latombe J-C., Overmars M.H. Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE Transactions on Robotics and Automation*. 1996. vol. 12(4). pp. 566–580.
16. Khatib O. Real-time obstacle avoidance for manipulators and mobile robots. *The international journal of robotics research*. 1986. vol. 5(1). pp. 90–98.
17. Latombe J.C. Robot Motion Plannin. Kluwer Academic. 1991.
18. Linde Y., Buzo A., Gray R.M. An algorithm for vector quantizer design. *IEEE Transactions on Communications*. 1980. vol. 28(1). pp. 84–95.
19. Lozano-Pérez T., Wesley M.A. An algorithm for planning collision-free paths among polyhedral obstacles. *Communications of the ACM*. 1979. vol. 22(10). pp. 560–570.
20. Lu D.V., Hershberger D., Smart, W.D. Layered costmaps for context-sensitive navigation. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2014 . 2014. pp. 709–715.
21. Müller J.P. The design of intelligent agents: a layered approach. Springer Science & Business Media. 1996. vol. 1177. 233 p.
22. Negrete M., Savage J., Cruz J., Marquez J. Parallel Implementation of Roadmap Construction for Mobile Robots using RGB-D Cameras. Open German-Russian Workshop on Pattern Recognition and Image Understanding (OGRW). 2014. 184 p.
23. Quigley M., Gerkey B., Smart W.D. Programming Robots with ROS: a practical introduction to the Robot Operating System. "O'Reilly Media, Inc.". 2015. 448 p.
24. Savage J. et al. Construction of roadmaps for mobile robots' navigation using rgb-d cameras. *Intelligent Autonomous Systems*. Springer. 2016. vol. 13. pp. 217–229.
25. Savage J. et al. VIRbot: a system for the operation of mobile robots. RoboCup 2007: Robot Soccer World Cup XI. Springer. 2008. pp. 512–519.
26. Seib V., Memmesheimer R., Paulus D. Aros-based system for an autonomous service robot. Robot Operating System (ROS). Springer. 2016. pp. 215–252.
27. Stückler J., Schwarz M., Behnke S. Mobile manipulation, tool use, and intuitive interaction for cognitive service robot cosero. *Frontiers in Robotics and AI*. 2016. vol. 3. 58 p.
28. Thrun S., Burgard W., Fox D. Probabilistic robotics. MIT Press. 2005. 672 p.
29. Van Beek L. et al. RoboCup@Home 2018: Rules and Regulations. 2018. 120 p. Available at: http://www.robocupathome.org/rules/2018_rulebook.pdf. 2018 (accessed: 29.08.2018)
30. Wachsmuth S., Holz D., Rudinac M., Ruiz-del-Solar J. RoboCup@Home-benchmarking domestic service robots. Association for the Advancement of Artificial Intelligence (AAAI). 2015. pp. 4328–4329.

Negrete Marco — Master in Engineering, Ph.D. student of Laboratory of Biorobotics of Faculty of Engineering, National Autonomous University of Mexico (UNAM). Research interests: domestic service robots, control and automation, autonomous navigation, computer vision, cognitive and behavioral sciences. The number of publications — 12. mnegretev@gmail.com; Circuito Exterior S/N, Ciudad Universitaria, 04510, Mexico City, Mexico; office phone: +52(55)56223041.

Savage Jesús — Ph.D., professor, full professor of the Electrical Engineering Department, National Autonomous University of Mexico (UNAM). Research interests: autonomous mobile robots, digital signal processing, computer architectures. The number of publications —100. robotssavage@gmail.com; Circuito Exterior S/N, Ciudad Universitaria, 04510, Mexico City, Mexico; office phone: +52(55)56223041.

Contreras-Toledo Luis Angel — Ph.D., postdoctoral researcher of Advance Intelligence and Robotics Research Center, Tamagawa University. Research interests: computer vision, visual SLAM, robotics. The number of publications — 20. tenshi.unam@gmail.com; 6-1-1, Tamagawagakuen, Machida, Tokyo, 194-8610, Japan; office phone: +5215510323399.

Acknowledgements. This research is supported by UNAM-DGAPA-PAPIIT (grant IG110818).

М. НЕГРЕТЕ , Х. САВАЖ , Л.Э. КОНТРЕРАС -ТОЛЕДО
**СИСТЕМА ПЛАНИРОВАНИЯ ДВИЖЕНИЯ БЫТОВОГО
СЕРВИСНОГО РОБОТА НА ОСНОВЕ АЛГОРИТМОВ
ПРОСТРАНСТВЕННОГО ПРЕДСТАВЛЕНИЯ И АКТИВНОЙ
НАВИГАЦИИ**

Негрете М., Саваж Х., Контрерас-Толедо Л.Э. Система планирования движения бытового сервисного робота на основе алгоритмов пространственного представления и активной навигации.

Аннотация. Главным предназначением сервисных роботов является помощь людям в непромышленных средах, таких как дома или офисы. Для достижения своей цели сервисные роботы должны обладать несколькими навыками, например распознавание и манипулирование объектом, обнаружение и распознавание лиц, распознавание и синтез речи, планирование задач и одним из самых важных навыков — навигация в динамических средах. В статье описывается полностью внедренная система планирования движения, которая учитывает все: начиная от алгоритмов движения и планирования пути до пространственного представления и активной навигации на основе поведения. Предлагаемая система реализована в бытовом сервисном роботе под названием «Юстина», конструкция которого основана на робототехнической архитектуре под названием «ViRBot», использующейся для контроля действий виртуальных и реальных роботов, которая охватывает несколько уровней абстракции от низкоуровневого управления до символического планирования. Мы оценили наш проект как в симулированной, так и в реальной среде и сравнили его с классическими реализациями. Для тестов мы использовали карты, полученные из реальных сред (Лаборатория биороботов и Robocup@Home arena), и карты, созданные из препятствий со случайными положениями и формами. Для сравнения использовались несколько параметров: общее пройденное расстояние, количество столкновений, количество достигнутых целей и средняя исполнительная скорость. Наш проект значительно улучшился как в реальных, так и в симуляционных тестах. Представлены экспериментальные результаты успешно протестированной системы в контексте конкурса RoboCup@Home.

Ключевые слова: Автономная навигация, поведенческая робототехника, бытовые сервисные роботы, планирование маршрута.

Негрете Марко — аспирант лаборатории биоробототехники, инженерного факультета, Национальный автономный университет Мексики. Область научных интересов: сервисный робот, управление и автоматизация, автономная навигации, компьютерное зрение, когнитивистика, бихевиоризм. Число научных публикаций — 12. mnegretev@gmail.com; Сиркуито Экстериор, Съезд Университария, Мехико, 04510, Мексика; р.т.: +52(55)56223041.

Саваж Хесус — Ph.D., профессор, профессор электротехнического отдела, Национальный автономный университет Мексики. Область научных интересов: автономные мобильные роботы, обработка цифрового сигнала, компьютерная архитектура. Число научных публикаций — 100. robotssavage@gmail.com; Сиркуито Экстериор, Съезд Университария, Мехико, 04510, Мексика; р.т.: +52(55)56223041.

Контрерас-Толедо Луис Энджел — Ph.D., научный сотрудник передового исследовательского центра искусственного интеллекта и робототехники, Университет Тамагава. Область научных интересов: компьютерное зрение, визуальный метод одновременной локализации и построения карты, робототехника. Число научных публикаций

— 20. tenshi.unam@gmail.com; Тамагава Гакуэн, 6-1-1, Матида, Токио, 194-8610, Япония; р.т.: +5215510323399.

Поддержка исследований. Работа выполнена при финансовой поддержке UNAM-DGAPA-PAPIIT (грант IG110818).

Литература

1. *Amigoni F. et al.* Competitions for benchmarking: task and functionality scoring complete performance assessment // *IEEE Robotics&Automation Magazine*. 2015. vol. 22(3). pp. 53–61.
2. *Arambula F., Padilla M.A.* Autonomus robot navigation using adaptive potential fields // *Mathematical and Computer Modelling*. 2011. Vol. 40. pp. 1141–1156.
3. *Arkin R.C.* Behavior-based robotics // MIT Press. 1998. 491 p.
4. *Avilés-Arriaga H.H. et al.* Markovito: A Flexible and General Service Robot // *Springer*. 2009. pp. 401–423.
5. *Banino A. et al.* Vector-based navigation using grid-like representations in artificial agents // *Nature*. 2018. vol. 557. no. 7705. pp. 429.
6. *Choset H.M. et al.* Principles of Robot Motion // MIT Press. 2005. 626 p.
7. *Contreras L., Mayol-Cuevas W.* Trajectory-driven point cloud compression techniques for visual slam // *IEEE/RSJ International Conference on Robots and Systems (IROS)*. 2015. pp. 133–140.
8. *Contreras L., Mayol-Cuevas W.* O-poco: Online point cloud compression mapping for visual odometry and slam // *IEEE/RSJ International Conference on Robots and Systems (IROS)*. 2017. pp. 4509–4514.
9. *Dissanayake M.W.M.G. et al.* A solution to the simultaneous localization and map building (SLAM) problem // *IEEE Transactions on robotics and automation*. 2001. vol. 17(3). pp. 229–241.
10. *Ferrein A., Steinbauer G.* 20 years of robocup // *KI-Künstliche Intelligenz*. 2016. vol. 30(3-4). pp. 225–232.
11. *Grisetti G., Stachniss C., Burgard W.* Improved techniques for grid mapping with raoblackwellized particle filters // *IEEE transactions on Robotics*. 2007. vol. 23(1). pp. 34–46.
12. *Iocchi L. et al.* RoboCup@ Home: Analysis and results of evolving competitions for domestic and service robots // *Artificial Intelligence*. 2015. vol. 229. pp. 258–281.
13. *Robots and robotic devices - vocabulary. ISO 8373-201* // *International Organization for Standardization*. Geneva. Switzerland. 2012.
14. *Jin B., Lei H., Geng W.* Accurate intrinsic calibration of depth camera with cuboids // *European Conference on Computer Vision*. 2014. pp. 788–803.
15. *Kavraki L., Svestka P., Latombe J-C., Overmars M.H.* Probabilistic roadmaps for path planning in high-dimensional configuration spaces // *IEEE Transactions on Robotics and Automation*. 1996. vol. 12(4). pp. 566–580.
16. *Khatib O.* Real-time obstacle avoidance for manipulators and mobile robots // *The international journal of robotics researc.* 1986. vol. 5(1). pp. 90–98.
17. *Latombe J.C.* *Robot Motion Plannin* // *Kluwer Academic*. 1991.
18. *Linde Y., Buzo A., Gray R.M.* An algorithm for vector quantizer design // *IEEE Transactions on Communications*. 1980. vol. 28(1). pp. 84–95.
19. *Lozano-Pérez T., Wesley M.A.* An algorithm for planning collision-free paths among polyhedral obstacles // *Communications of the ACM*. 1979. vol. 22(10). pp. 560–570.

20. *Lu D.V., Hershberger D., Smart, W.D.* Layered costmaps for context-sensitive navigation // IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2014. 2014. pp. 709–715.
21. *Müller J.P.* The design of intelligent agents: a layered approach // Springer Science & Business Media. 1996. vol. 1177. 233 p.
22. *Negrete M., Savage J., Cruz J., Marquez J.* Parallel Implementation of Roadmap Construction for Mobile Robots using RGB-D Cameras // Open German-Russian Workshop on Pattern Recognition and Image Understanding (OGRW). 2014. 184 p.
23. *Quigley M., Gerkey B., Smart W.D.* Programming Robots with ROS: a practical introduction to the Robot Operating System // "O'Reilly Media, Inc.". 2015. 448 p.
24. *Savage J. et al.* Construction of roadmaps for mobile robots' navigation using rgb-d cameras // Intelligent Autonomous Systems. Springer. 2016. vol. 13. pp. 217–229.
25. *Savage J. et al.* VIRbot: a system for the operation of mobile robots // RoboCup 2007: Robot Soccer World Cup XI. Springer. 2008. pp. 512–519.
26. *Seib V., Memmesheimer R., Paulus D.* Aros-based system for an autonomous service robot // Robot Operating System (ROS). Springer. 2016. pp. 215–252.
27. *Stückler J., Schwarz M., Behnke S.* Mobile manipulation, tool use, and intuitive interaction for cognitive service robot cosero // Frontiers in Robotics and AI. 2016. vol. 3. 58 p.
28. *Thrun S., Burgard W., Fox D.* Probabilistic robotics // MIT Press. 2005. 672 p.
29. *Van Beek L. et al.* RoboCup@Home 2018: Rules and Regulations. 2018. 120 p. Available at: http://www.robocupathome.org/rules/2018_rulebook.pdf. 2018 (дата обращения: 29.08.2018)
30. *Wachsmuth S., Holz D., Rudinac M., Ruiz-del-Solar J.* RoboCup@Home-benchmarking domestic service robots // Association for the Advancement of Artificial Intelligence (AAAI). 2015. pp. 4328–4329.

В.Х. ПШИХОПОВ, М.Ю. МЕДВЕДЕВ
**ГРУППОВОЕ УПРАВЛЕНИЕ ДВИЖЕНИЕМ МОБИЛЬНЫХ
РОБОТОВ В НЕОПРЕДЕЛЕННОЙ СРЕДЕ С
ИСПОЛЬЗОВАНИЕМ НЕУСТОЙЧИВЫХ РЕЖИМОВ**

Пшихопов В.Х., Медведев М.Ю. Групповое управление движением мобильных роботов в неопределенной среде с использованием неустойчивых режимов.

Аннотация. Рассматриваются алгоритмы управления, обеспечивающие согласованное перемещение группы роботов в неопределенной трехмерной среде с препятствиями. Неопределенность среды заключается в наличии априори неизвестных препятствий, часть которых может быть нестационарными. Мобильные роботы группы должны автоматически распределиться в заданной прямоугольной области на плоскости и двигаться в направлении, перпендикулярном указанной области, по возможности сохраняя заданное взаимное расположение. В данной статье предлагаются новые алгоритмы автоматического распределения роботов на плоскости, не предполагающие предварительного назначения места каждого робота. Эта задача решается с применением триангуляции Делоне и дальнейшей оптимизации положения робота. Для коррекции движения отдельного робота и всей группы при сближении с препятствием предложены алгоритмы, базирующиеся на неустойчивых режимах, позволяющих трансформировать препятствия в репеллеры. Рассмотрено два варианта алгоритмов обхода препятствий. В первом варианте используются только неустойчивые режимы, а во втором варианте — гибридный алгоритм, включающий интеллектуальный анализ текущей ситуации и неустойчивый режим движения. Предложенные алгоритмы могут реализовываться децентрализованно. В статье анализируются два варианта алгоритмов группового управления, а также выполняется численное моделирование группы из 5 гексакоптеров в неопределенной среде с неподвижными и подвижными препятствиями. Также приведены экспериментальные данные, подтверждающие работоспособность предлагаемых алгоритмов на примере полета двух гексакоптеров в среде с неподвижным препятствием. Разработанные алгоритмы могут применяться в системах управления мобильными роботами при их групповом движении в неопределенных 3D средах.

Ключевые слова: групповое управление, неопределенная среда, неустойчивые режимы, управление движением.

1. Введение. В неопределенных средах с препятствиями система управления мобильного робота, как правило, строится по иерархической структуре, включающей уровни управления исполнительными механизмами, управления движением и планирования движения [1]. В настоящее время большое внимание исследователи уделяют алгоритмам планирования движения мобильных роботов и их групп [2-5]. При планировании движения требуется на основе данных системы технического зрения доопределять среду и выбирать оптимальный маршрут. Решение таких задач делает актуальным использование интеллектуальных подходов и технологий [6-9].

В связи со сложностью проблем планирования и группового управления движением мобильных роботов единого подхода, дающего

эффективное решение в общем случае, не существует. Поэтому в различных условиях применяются разные подходы. Общая проблематика планирования движения мобильных роботов освещена в работах [2, 3]. В [10-14] представлены основные методы группового управления, используемые в мобильной робототехнике.

Анализ работ в области группового управления движением роботов показывает, что в данном направлении исследования нацелены на решение проблем группового планирования траекторий, формирования и поддержания строя.

Среди подходов, используемых для решения указанных задач, можно выделить метод искусственных потенциальных полей, обучающиеся системы на базе искусственных нейронных сетей, методы мультиагентного управления, методы эволюционных вычислений, логического вывода и методы, базирующиеся на использовании природных аналогий.

Метод потенциальных полей [23] является одним из самых распространенных, однако он подразумевает картографирование области функционирования и имеет проблему попадания в локальные минимумы. Кроме того, метод потенциальных полей требует применения дополнительных приемов, позволяющих учитывать переменные скорости и ускорения объектов среды.

Интеллектуальные обучающиеся системы демонстрируют высокий потенциал в задачах управления движением. Последние достижения в этой области связаны с нейронными сетями глубокого обучения [26], которые успешно решают задачи классификации в системах технического зрения и управления. Вместе с тем следует отметить, что нейронные сети глубокого обучения требуют наличия больших объемов обучающих выборок и длительного обучения для достижения необходимой надежности. В связи с недостаточно высокой надежностью в системах управления нейронные сети, как правило, используются в сочетании с традиционными регуляторами.

Мультиагентные технологии [27] также часто применяются для построения систем группового управления мобильными роботами, так как они позволяют организовывать взаимодействие объектов в условиях возможных конфликтов, конкуренции, противодействия. Мультиагентные технологии в том или ином виде применяются при решении задач управления в неопределенных и конфликтных средах.

Эволюционные алгоритмы, хорошо зарекомендовавшие себя при решении задач многомерной оптимизации, в последнее время приобрели актуальность в решении задач группового управления мобильными роботами. Это связано с появлением систем группового управления, включающих в себя сотни и тысячи мобильных роботов.

Так, Министерство обороны США провело в 2015 году «бой» двух групп квадрокоптеров численностью 20 штук в каждой. В ходе эксперимента проведена отработка различных стратегий группового управления. В октябре 2016 года с трех истребителей F/A-18 осуществлен сброс 103 миниатюрных беспилотников Perdix, которые объединились в стаю, реализовали несколько конфигураций и произвели имитацию атаки на объект. В начале 2017 года в Китае состоялась демонстрация группы из тысячи квадрокоптеров Ghostdrone 2.0, которыми управлял один инженер. В больших группах роботов оптимизация движения с помощью эволюционных алгоритмов может демонстрировать высокую эффективность. Вместе с тем следует отметить, что эволюционные алгоритмы наиболее эффективны при наличии карты среды [1].

Для управления группами мобильных роботов большой размерности могут также применяться методы, использующие природные аналогии — стайные принципы управления, муравьиные и пчелиные алгоритмы, метод неустойчивых режимов. Применение таких подходов позволяет не задавать положение каждого робота группы при достижении цели.

Многочисленные результаты проведенных исследований указывают на эффективность гибридизации различных алгоритмов и методов в задачах планирования и управления движением роботов [1, 15-19]. В частности, в работе [1] показано, что гибридизация методов при решении задач планирования пути приводит к увеличению интегрального критерия качества до 50 %. В этой связи в данной статье разрабатываются гибридные алгоритмы управления движением в трехмерной среде в условиях неопределенности. Предлагается система планирования и управления движением, в которой присутствуют как реактивная составляющая [2], так и интеллектуальная. Эти две составляющие дополняют друг друга и позволяют функционировать в среде, априорная карта которой отсутствует.

Структура системы планирования и управления мобильного робота в неопределенной динамической среде представлена на рисунке 1 [1].

Блок планирования движения на основе навигационных данных (y, x) и данных о положении препятствий U_{Π} формирует желаемую траекторию движения ψ_{tr} . Блок управления движением формирует вектор управления u исполнительными механизмами, которые воздействуют на мобильного робота.

Рассмотрим следующую задачу. Имеется трехмерная прямоугольная область, ограниченная координатами (L_1, L_2, L_3) (см. рисунок 2), содержащая точечные препятствия Π_i , расположение которых заранее неизвестно. Препятствия Π_i могут перемещаться с кусочно-постоянными скоростями.

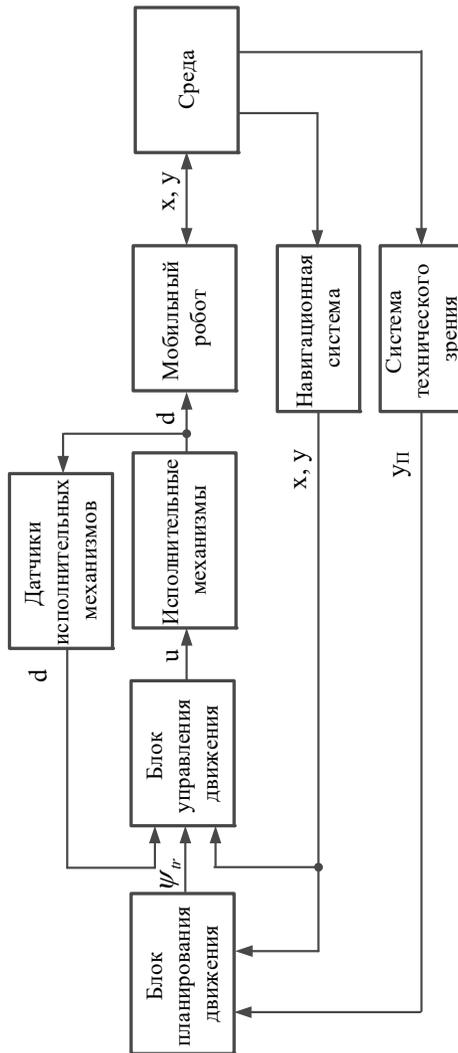


Рис. 1. Структура системы планирования и управления мобильным роботом

Математическая модель мобильного робота представлена уравнениями кинематики и динамики твердого тела [20]:

$$\dot{y}_i = R(y_i)x_i; \quad (1)$$

$$\dot{x}_i = M_i^{-1}(F_{ui} + F_{di}), \quad (2)$$

где y_i — вектор, определяющий положение i -го робота в неподвижных декартовых координатах; x_i — вектор скоростей в связанных с мобильным роботом координатах; $R_i(y_i)$ — матрица, описывающая кинематику робота; M_i — матрица инерционных параметров; F_{ui} — вектор управления; F_{di} — вектор прочих сил и моментов, действующих на мобильного робота; $i = \overline{1, N}$; N — число роботов в группе.

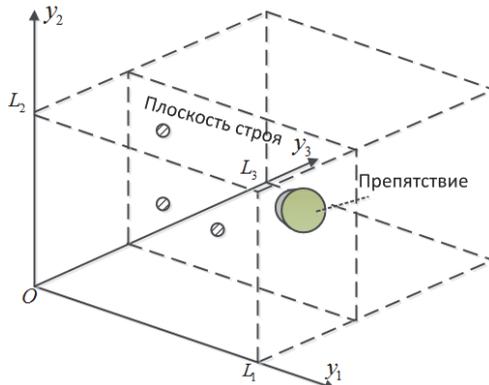


Рис. 2. Задача обследования области группой роботов

Мобильные роботы оснащены навигационной системой, которая определяет их положения в неподвижной системе координат. Роботы группы также оснащены системой технического зрения, которая определяет текущее положение препятствий и других объектов, и системой связи.

Группа из N мобильных роботов должна распределиться в плоскости строя (см. рисунок 2) таким образом, чтобы оптимизировать функционал, зависящий от расстояний между роботами, границами области и препятствиями $\Omega(r_{ij})$. При этом оптимизируемыми параметрами являются координаты y_{1i} и y_{2i} каждого робота группы. Кроме того, группа мобильных роботов должна перемещаться вдоль оси Oy_3 с

заданной постоянной скоростью $\dot{y}_{3i} = V_{II}$, $V_{II} = const$, одновременно удовлетворяя равенствам $y_{3i} = y_{3j}$, $i, j = \overline{1, N}$.

Для решения поставленной задачи в данной статье рассматриваются следующие вопросы:

- предлагаются и исследуются алгоритмы автоматического распределения мобильных роботов в плоскости строя;
- предлагаются и исследуются два децентрализованных алгоритма планирования движения группы мобильных роботов в среде с нестационарными препятствиями.

2. Распределение роботов в плоскости строя. Рассматривается совокупность роботов и препятствий, обнаруженных системами технического зрения. Назовем такую совокупность обобщенной группой подвижных объектов.

Решение задачи распределения роботов в плоскости строя можно разделить на три основных этапа.

На первом этапе выделяются объекты (роботы или препятствия), находящиеся на минимальном расстоянии к границам прямоугольной области. В качестве указанных границ могут использоваться либо вершины прямоугольников, либо их ребра. Например, на рисунке 3 представлен вариант формирования дополнительных препятствий на основе координат объектов, находящихся наиболее близко к границам прямоугольной области. Эти объекты отмечены на рисунке 3 крестиками. При этом точки пересечения перпендикуляров от ближайших объектов до границ прямоугольной области также считаются объектами обобщенной группы (препятствиями). Таким образом, обобщенная группа включает в себя мобильные роботы, препятствия внутри прямоугольной области и дополнительные препятствия на границах области.

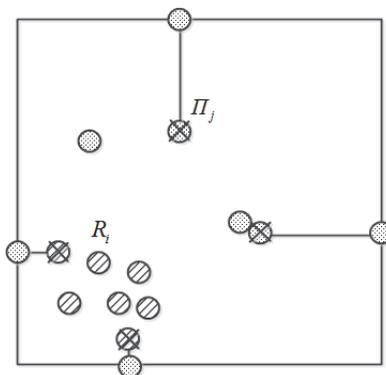


Рис. 3. Формирование препятствий на границах области

При использовании в качестве границ области функционирования вершин прямоугольника, выделение ближайших объектов не производится. В обобщенную группу подвижных объектов включаются четыре препятствия с координатами указанных вершин.

На втором этапе осуществляется кластеризация объектов обобщенной группы, включающей в себя роботов и препятствия. Для осуществления кластеризации используется механизм триангуляции. В частности, может быть использована триангуляция Делоне, обладающая известными преимуществами [21]. При триангуляции учитываются все объекты — и роботы, и препятствия. Результатом триангуляции для каждого робота R_i , $i = \overline{1, N}$ является множество Ω_i объектов, с которыми i -й робот связан ребрами. Таким образом, в результате кластеризации формируется N подгрупп, включающих в себя роботов и препятствия. Данная операция позволяет выделить понятие «соседние объекты» на плоскости. Возможно использование и других способов кластеризации. Например, в [1] применено разделение на секторы пространства, окружающего мобильного робота. В этом случае для i -го робота «соседним объектом» в j -м секторе является объект, входящий в этот сектор и находящийся на минимальном расстоянии от него. Однако такой способ требует обоснованного выбора количества секторов, на которые разделяется окружающее робота пространство.

На рисунке 4 приведена графическая иллюстрация результатов триангуляции Делоне для группы из 5 роботов. Триангуляция осуществлена в среде Matlab. На рисунке 4 в качестве границ области использованы вершины прямоугольника, ограничивающего область функционирования группы роботов.

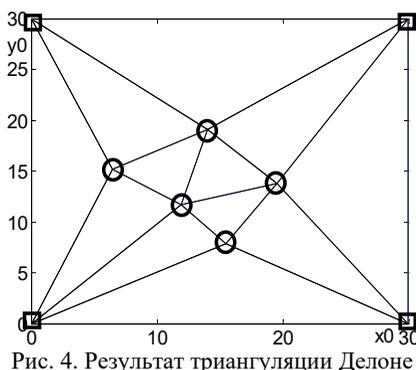


Рис. 4. Результат триангуляции Делоне

В результате кластеризации с использованием триангуляции Делоне формируется массив связей каждого робота с другими робота-

ми и препятствиями. Далее выполняется обработка и сортировка полученного массива связей, в ходе которой удаляются повторяющиеся элементы и все роботы сортируются в порядке возрастания их номеров. Результатом является матрица соседних объектов каждого мобильного робота группы. В ней по одному разу перечислены номера соседних объектов для данного робота в порядке возрастания. Данная матрица описывает i -й кластер группы роботов.

На основе указанной матрицы формируется функция, являющаяся суммой расстояний от i -го робота до всех соседних объектов:

$$S_i = \sum_{j=1}^{n_i} r_{ij} = \sum_{j=1}^{n_i} \left[(y_{1i} - y_{1j})^2 + (y_{2i} - y_{2j})^2 \right], \quad (3)$$

где n_i — число соседних объектов для i -го робота, полученных в результате триангуляции Делоне; r_{ij} — расстояние от i -го робота до j -го объекта кластера.

На третьем этапе для нахождения требуемого положения i -го робота ставится задача оптимизировать полученную функцию (3), используя в качестве независимых переменных координаты y_{1i} , y_{2i} . Так как функция (3) не имеет особенностей, то необходимые условия экстремума имеют вид:

$$\frac{\partial S_i}{\partial y_{1i}} = 2 \sum_{j=1}^{n_i} (y_{1i} - y_{1j}) = 0, \quad \frac{\partial S_i}{\partial y_{2i}} = 2 \sum_{j=1}^{n_i} (y_{2i} - y_{2j}) = 0. \quad (4)$$

Из уравнений (4) находятся оптимальные значения координат i -го робота в плоскости Oy_1y_2 :

$$y_{1i}^* = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{1j}, \quad y_{2i}^* = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{2j}. \quad (5)$$

Выражения (5) определяют желаемые значениям координат i -го робота в плоскости Oy_1y_2 (см. рисунок 2) в ситуации, когда отсутствуют дополнительные препятствия, кроме границ области.

Выражения (5) дополняются желаемыми значениями углов ориентации i -го робота, которые зависят от конструктивных особенностей и размещения управляющих элементов [20].

Синтез алгоритмов управления нижнего уровня, обеспечивающих отработку желаемых значений координат и углов ориентации подвижного объекта, осуществляется на основе метода позиционно-

траекторного управления [1, 20], с помощью которого решается обратная задача динамики по определению сил и моментов, требуемых для реализации заданной траектории. Траектория мобильного робота задается в виде вектора, содержащего функции, определяющие требования к линейным и угловым координатам мобильного робота [1]:

$$\Psi_{tr}^i = A_{1i}Y_i y_i + A_{2i}y_i + A_{3i} = 0, \quad (6)$$

где Ψ_{tr}^i — вектор функций базовых координат робота, формируемых на основе желаемых траекторий движения; A_{ij} — матрицы постоянных коэффициентов, определяющие конкретные траектории. Размерность вектора Ψ_{tr}^i равна $\mu \times 1$. Матрицы A_{1i} , A_{2i} имеют размерность $\mu \times 6$, матрица A_{3i} имеет размерность $\mu \times 1$.

Матрица Y_i является диагональной, определяемой следующим выражением:

$$Y_i = \text{diag}([y_{1i} \ y_{2i} \ y_{3i} \ y_{4i} \ y_{5i} \ y_{6i}]).$$

Выражение (6) позволяет задавать движение i -го робота в точку, вдоль прямой линии или вдоль кривой второго порядка.

Выражение (6) дополняется требованиями к скорости движения группы роботов, которые в общем случае задаются в виде:

$$\dot{y}_{3i} = V_{II} - \lambda_V \frac{1}{n_i^*} \sum_{j=1}^{n_i^*} (y_{3i} - y_{3j}), \quad (7)$$

где разность положений $y_{3i} - y_{3j}$ вычисляется только для мобильных роботов группы без учета препятствий, λ_V — постоянный коэффициент настройки; n_i^* — число роботов, входящих в кластер, полученный в результате триангуляции Делоне; V_{II} — заданная постоянная скорость движения плоскости строя.

В соответствии с выражением (7), если текущее положение i -го робота меньше среднего значения положений всех остальных объектов группы, то задание по скорости увеличивается. Если текущее положение i -го робота больше среднего значения положений всех остальных объектов группы, то задание по скорости уменьшается. При значении $\lambda_V = 0$ роботы группы обрабатывают уставку по скорости V_{II} независимо друг от друга.

Требования (7) могут быть сформулированы в виде:

$$\Psi_{ck}^i = A_{4i}x_i + A_{5i} + \lambda_V \frac{1}{n_i} \sum_{j=1}^{n_i^*} (y_{3i} - y_{3j}), \quad (8)$$

где матрица A_{4i} имеет размерность $v \times 6$, а матрица A_{5i} имеет размерность $v \times 1$.

3. Гибридный алгоритм реконфигурации строя при движении. Обход препятствий группой роботов осуществляется с использованием неустойчивых по расстоянию до препятствий режимов движения, впервые предложенных в [22] и развитых в [1]. Данный метод имеет сходство с методом потенциальных полей. Основное отличие метода неустойчивых режимов от метода искусственных потенциальных полей [23–25] заключается в следующем. Как известно, в методе потенциальных полей притягивающие (отталкивающие) силы являются функциями координат y :

$$F_{attr} = F_{attr}(y), \quad F_{rep} = F_{rep}(y),$$

где $F_{attr}(y)$ — результирующая притягивающих сил; $F_{rep}(y)$ — результирующая отталкивающих сил.

В предложенном в работах [1, 22] методе неустойчивых режимов притягивающие и отталкивающие силы являются функциями координат, скоростей и ускорений подвижного объекта и бифуркационного параметра β . Они формируются как решение дифференциального уравнения вида:

$$F_{ar}(y, \dot{y}, \ddot{y}, \beta) = 0,$$

где $F_{ar}(y, \dot{y}, \ddot{y}, \beta)$ — функция, которая в зависимости от значения параметра β является притягивающей или отталкивающей силой.

Идея, предложенная в [1, 22], состоит в применении теоремы Ляпунова о неустойчивости для линейного дифференциального уравнения замкнутой системы управления относительно желаемой траектории $\psi(y)$:

$$\ddot{\psi} + T_1(\beta)\dot{\psi} + T_2(\beta)\psi = 0, \quad (9)$$

где $T_1(\beta)$, $T_2(\beta)$ — матрицы коэффициентов.

Если расстояния от мобильного робота до препятствий больше допустимого значения r_d , то есть:

$$r_{ij} > r_d, j = \overline{1, n_i}, \quad (10)$$

то бифуркационный параметр β принимается равным нулю, а матрицы $T_1(\beta), T_2(\beta)$ положительно определенные. В этом случае решение $\psi(y)=0$ уравнения (9) является асимптотически устойчивым.

Если одно из неравенств (10) нарушается, то бифуркационный параметр β отличен от нуля, а матрицы $T_1(\beta), T_2(\beta)$ становятся отрицательно определенными. В этом случае положение равновесия $\psi(y)=0$ является неустойчивым.

В работах [1, 22] бифуркационный параметр вычисляется в соответствии с выражением:

$$\beta_i = \sum_{j=1}^{n_i} \left[|r_{ij} - r_d| - (r_{ij} - r_d) \right]. \quad (11)$$

В данной статье бифуркационный параметр вычисляется в соответствии с (11), однако предлагается комбинированный алгоритм, реализующийся структурой, представленной на рисунке 5.

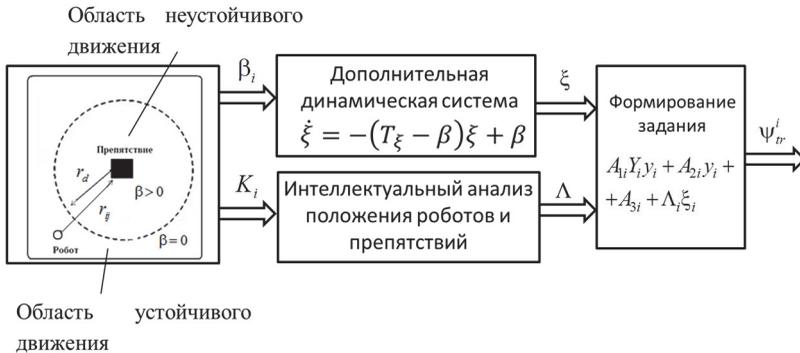


Рис. 5. Формирование траектории с учетом препятствия

Предлагаемый алгоритм состоит в последовательном выполнении следующих шагов.

Шаг 1. На основе данных, поступающих от системы технического зрения i -го робота, в соответствии с выражением (11) вычисляется бифуркационный параметр β_i и набор векторов K_i , связывающих

робота с ближайшими до него точками на обнаруженных препятствиях. Бифуркационный параметр β_i поступает на вход дифференциального уравнения вида:

$$\dot{\xi}_i = -(T_{\xi_i} - I\beta_i)\xi_i + \beta_i I, \quad (12)$$

где ξ_i — вектор дополнительных переменных; T_{ξ_i} — положительно определенная матрица.

Из (12) следует, что при попадании i -го робота в область препятствия параметр β_i начинает расти. При выполнении условия $T_{\xi_i} < \beta_i$ решение уравнения (12) становится неустойчивым, поэтому вектор ξ_i экспоненциально нарастает.

Шаг 2. Формируется матрица Λ_i алгоритмом, задающим направление обхода препятствия. Данный алгоритм анализирует расположение соседних объектов в кластере и ранжирует координаты роботов и препятствий по оси OY_1 и оси OY_2 . Для i -го робота рассматриваются расстояния до $i-1$ -го и $i+1$ -го объектов. Если расстояние между i -м и $i-1$ -м объектами меньше, чем расстояние между i -м и $i+1$ -м объектами, то принимается решение обходить препятствие так, чтобы соответствующая координата увеличивалась. Пусть, например, алгоритм верхнего уровня определил, что требуется осуществить облет препятствия в сторону увеличения координаты y_{1i} и уменьшения координаты y_{2i} . Тогда $\Lambda_i(1,1) = 1$, $\Lambda_i(2,2) = -1$.

Шаг 3. Вычисляется траекторная ошибка системы управления робота на основе выражения:

$$\psi_{tr}^i = A_{1i}Y_{1i} + A_{2i}Y_{2i} + A_{3i} + \Lambda_i\xi_i. \quad (13)$$

Шаг 4. Траекторная ошибка (13) передается в систему управления движением мобильного робота.

Отметим, что задающий сигнал (13) обрабатывается системой управления i -го подвижного объекта в устойчивом режиме. Иными словами, неустойчивый режим реализуется на уровне планирования.

4. Негибридный алгоритм реконфигурации строя при движении. Рассмотрим аналитический алгоритм, не требующий эвристического анализа при определении направления обхода препятствий, а базирующийся только на неустойчивых режимах. Данный алгоритм состоит из последовательного выполнения шагов (1)-(4), описанных в пункте 3, однако он отличается способом расчета бифуркационных параметров β_i и дополнительных переменных ξ_i .

Для каждой линейной координаты, требования к которой определяются вектором (6), вводится два дополнительных бифуркационных параметра, которые вычисляются в соответствии с выражениями:

$$\beta_i^{xa} = |r_i^{xa} - r_d| - (r_i^{xa} - r_d), \quad \beta_i^{xb} = |r_i^{xb} - r_d| - (r_i^{xb} - r_d); \quad (14)$$

$$\beta_i^{ya} = |r_i^{ya} - r_d| - (r_i^{ya} - r_d), \quad \beta_i^{yb} = |r_i^{yb} - r_d| - (r_i^{yb} - r_d); \quad (15)$$

$$\beta_i^{za} = |r_i^{za} - r_d| - (r_i^{za} - r_d), \quad \beta_i^{zb} = |r_i^{zb} - r_d| - (r_i^{zb} - r_d), \quad (16)$$

где $r_i^{xa}, r_i^{xb}, r_i^{ya}, r_i^{yb}, r_i^{za}, r_i^{zb}$ — показания датчиков расстояния до препятствий по соответствующим координатам.

На основании бифуркационных параметров (14), (15), (16) вычисляются векторы дополнительных переменных как решения следующих дифференциальных уравнений:

$$\dot{\xi}_i^{xa} = -\left(T_{\xi_i} - I\beta_i^{xa}\right)\xi_i^{xa} + \beta_i^{xa}I, \quad \dot{\xi}_i^{xb} = -\left(T_{\xi_i} - I\beta_i^{xb}\right)\xi_i^{xb} + \beta_i^{xb}I; \quad (17)$$

$$\dot{\xi}_i^{ya} = -\left(T_{\xi_i} - I\beta_i^{ya}\right)\xi_i^{ya} + \beta_i^{ya}I, \quad \dot{\xi}_i^{yb} = -\left(T_{\xi_i} - I\beta_i^{yb}\right)\xi_i^{yb} + \beta_i^{yb}I; \quad (18)$$

$$\dot{\xi}_i^{Vza} = -\left(T_{\xi_i} - I\beta_i^{za}\right)\xi_i^{Vza} + \beta_i^{za}I, \quad \dot{\xi}_i^{Vzb} = -\left(T_{\xi_i} - I\beta_i^{zb}\right)\xi_i^{Vzb} + \beta_i^{zb}I. \quad (19)$$

Пусть требования (6), (8) сформированы в виде:

$$\Psi_{tr}^i = A_{2i}y_i + A_{3i} + \left[\xi_i^{xb} - \xi_i^{xa} \quad \xi_i^{yb} - \xi_i^{ya} \quad 0_{m-2} \right]^T, \quad (20)$$

$$\Psi_{ck}^i = A_{4i}x_i + A_{5i} + \left[0 \quad \xi_i^{Vzb} - \xi_i^{Vza} \quad 0_{v-3} \right]^T. \quad (21)$$

При этом требования к скоростям дополнительно ограничиваются в соответствии с характеристиками подвижного объекта.

Для нахождения вектора управлений F_{ui} используется метод обратных задач динамики, в соответствии с которым требуется, чтобы переменные (20), (21) удовлетворяли уравнениям:

$$\ddot{\Psi}_{tr}^i + T_2\dot{\Psi}_{tr}^i + T_1\Psi_{tr}^i = 0; \quad (22)$$

$$\dot{\Psi}_{ck}^i + T_3\Psi_{ck}^i = 0, \quad (23)$$

находится вектор F_{ui} .

Анализ устойчивости замкнутой системы проводится с использованием следующей функции Ляпунова:

$$V_i = \frac{1}{2} \psi_{ck}^{iT} \psi_{ck}^i + \psi_{tr}^{iT} T_1 \psi_{tr}^i + \frac{1}{2} \dot{\psi}_{tr}^{iT} \dot{\psi}_{tr}^i + \frac{1}{2} (\dot{\psi}_{tr}^i + T_2 \psi_{tr}^i)^T (\dot{\psi}_{tr}^i + T_2 \psi_{tr}^i). \quad (24)$$

Производная от выражения (24), вычисленная в силу уравнений замкнутой системы (22), (23), равна:

$$\begin{aligned} \dot{V}_i &= -\psi_{ck}^{iT} T_3 \psi_{ck}^i + 2\psi_{tr}^{iT} T_1 \dot{\psi}_{tr}^i + \dot{\psi}_{tr}^{iT} (-T_1 \psi_{tr}^i - T_2 \dot{\psi}_{tr}^i) + \\ &+ (\dot{\psi}_{tr}^i + T_2 \psi_{tr}^i)^T (-T_1 \psi_{tr}^i - T_2 \dot{\psi}_{tr}^i + T_2 \dot{\psi}_{tr}^i) = \\ &= -\psi_{ck}^{iT} T_3 \psi_{ck}^i - \psi_{tr}^{iT} T_1 T_2 \psi_{tr}^i - \dot{\psi}_{tr}^{iT} T_2 \dot{\psi}_{tr}^i. \end{aligned} \quad (25)$$

Из выражений (24), (25) следует, что:

$$\psi_{ck}^i \rightarrow 0, \quad \psi_{tr}^i \rightarrow 0.$$

Учитывая (20), (21), получаем в установившемся режиме:

$$\begin{aligned} A_{2i} y_i &= -A_{3i} - \left[\xi_i^{xb} - \xi_i^{xa} \quad \xi_i^{yb} - \xi_i^{ya} \quad 0_{\mu-2} \right]^T, \\ A_{4i} x_i &= -A_{5i} - \left[00 \xi_i^{Vzb} - \xi_i^{Vza} \quad 0_{v-3} \right]^T. \end{aligned}$$

Тогда в области, ограниченной неравенствами:

$$r_i^{xa} > r_d, \quad r_i^{xb} > r_d, \quad r_i^{ya} > r_d, \quad r_i^{yb} > r_d, \quad r_i^{za} > r_d, \quad r_i^{zb} > r_d, \quad (26)$$

установившийся режим определяется выражениями:

$$\begin{aligned} \tilde{y}_i &= -A_{2i}^{-1} A_{3i}, \\ \tilde{x}_i &= -A_{4i}^{-1} A_{5i}. \end{aligned}$$

Вне области, определенной неравенствами (26), установившийся режим достигается при условиях:

$$\xi_i^{xb} = \xi_i^{xa} = \xi_i^x, \quad \xi_i^{yb} = \xi_i^{ya} = \xi_i^y, \quad \xi_i^{Vzb} = \xi_i^{Vza} = \xi_i^{Vz}. \quad (27)$$

Дифференцируя равенства (27), с учетом выражений (17)-(19), получаем:

$$\beta_i^{xa} = \beta_i^{xb}, \quad \beta_i^{ya} = \beta_i^{yb}, \quad \beta_i^{za} = \beta_i^{zb}. \quad (28)$$

Таким образом, в замкнутой системе управления установившийся режим по координатам и скоростям робота группы достигается как при устойчивых решениях дифференциальных уравнений (17)–(19), так и при неустойчивых решениях указанных уравнений. При одинаковых параметрах T_{ξ_j} уравнений (17)–(19) условия (28) приводят к тому, что роботы распределяются равномерно в заданной области. Задавая различные параметры T_{ξ_j} можно изменять установившиеся значения положений роботов группы.

5. Результаты численных исследований и эксперимента. В качестве примера рассмотрим систему управления группой роботов, состоящих из пяти однотипных гексакоптеров. Масса гексакоптера составляет 1.08 кг, его диаметр 0,55 м; масса полезной нагрузки до 2.7 кг. При стандартном аккумуляторе гексакоптер позволяет летать до 20 минут с максимальной скоростью до 4 м/с. Количество гексакоптеров выбрано из соображений достаточного их количества для демонстрации алгоритмов, с учетом имеющихся вычислительных мощностей.

Модель гексакоптера (1), (2) описывается следующими матрицами и параметрами (индекс i для упрощения выражений опущен):

$$A_{\omega} = \begin{bmatrix} 0 & \frac{\cos \gamma}{\cos \theta} & -\frac{\sin \gamma}{\cos \theta} \\ 0 & \sin \gamma & \cos \gamma \\ 1 & -\operatorname{tg} \theta \cos \gamma & \operatorname{tg} \theta \sin \gamma \end{bmatrix};$$

$$A = \begin{bmatrix} \cos \psi \cos \vartheta & -\cos \psi \sin \vartheta \cos \gamma + \sin \psi \cos \gamma + & \sin \psi \cos \gamma + \\ & \sin \psi \sin \gamma & \cos \psi \sin \vartheta \sin \gamma \\ \sin \vartheta & \cos \vartheta \cos \gamma & -\cos \vartheta \sin \gamma \\ -\sin \psi \cos \vartheta & \cos \psi \sin \gamma + & \cos \psi \cos \gamma - \\ & \sin \psi \sin \vartheta \cos \gamma & \sin \psi \sin \vartheta \sin \gamma \end{bmatrix};$$

$$m = 3.2 \text{ kg}, J_x = J_z = 0.032 \text{ kg} \cdot \text{m}^2, J_y = 0.062 \text{ kg} \cdot \text{m}^2;$$

$$M = \operatorname{diag} \left(\left[m, m, m, J_x, J_y, J_z \right] \right);$$

$$F_d = \begin{bmatrix} m(-g \sin \vartheta - \omega_y V_z + \omega_z V_y) \\ m(-g \cos \vartheta \cos \gamma - \omega_z V_x + \omega_x V_z) \\ m(g \cos \vartheta \sin \gamma - \omega_x V_y + \omega_y V_x) \\ -(J_z - J_y) \omega_y \omega_z \\ -(J_x - J_z) \omega_x \omega_z \\ -(J_y - J_x) \omega_y \omega_x \end{bmatrix};$$

$$F_{ui} = \begin{bmatrix} 0000 \\ 1000 \\ 0000 \\ 0100 \\ 0010 \\ 0001 \end{bmatrix} \begin{bmatrix} P_y \\ N_x \\ N_y \\ N_z \end{bmatrix},$$

где P_y, N_x, N_y, N_z — тяга и моменты, создаваемые винтами.

На рисунках 6–8 представлены результаты моделирования при движении группы с неподвижным препятствием, расположенным в точке $y_p=[16; 14; 15; 0; 0; 0]^T$. На рисунке 6 представлены траектории группы подвижных объектов в трехмерном пространстве со сферическим препятствием, а на рисунке 7 их проекции в плоскости Ox_0y_0 .

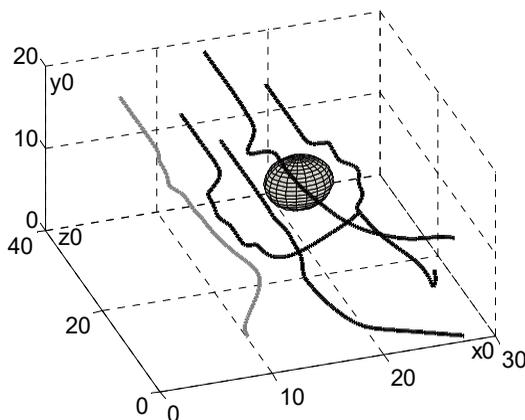


Рис. 6. Траектории движения группы в области с неподвижным препятствием

Моделирование проводилось при следующих параметрах: размер области $La=30$ м; безопасное расстояние 3 м, $T_{\xi_i} = 3$.

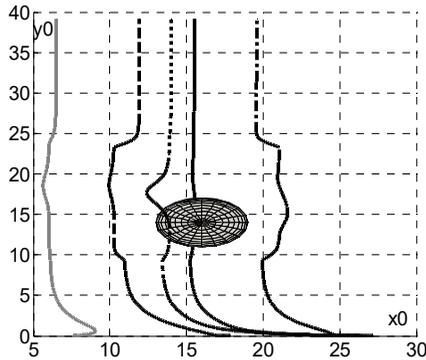


Рис. 7. Проекция траекторий движения группы в области с неподвижным препятствием

На рисунке 8 представлены переменные ξ_i , обеспечивающие неустойчивый режим движения, и распределение группы подвижных объектов на плоскости Oy_1y_2 в конечный момент времени $t = 30$ с.

В начальный период времени группа, включающая пять гексакоптеров, и четыре препятствия с координатами $(0, 0)$, $(La, 0)$, (La, La) и $(0, La)$, с помощью триангуляции Делоне осуществляет кластеризацию на подгруппы. Система планирования движения каждого гексакоптера группы определяет свое место на плоскости Oy_1y_2 , одновременно двигаясь с заданной скоростью V_{II} вдоль оси Oy_3 .

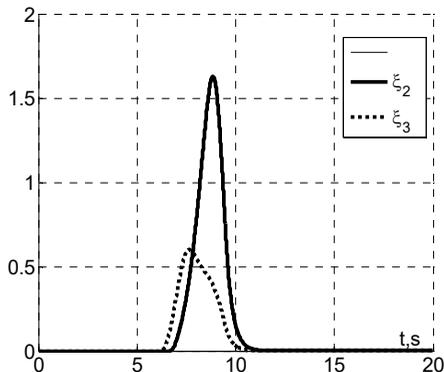


Рис. 8. Дополнительные переменные ξ_i группы подвижных объектов

Приблизительно в момент времени $t = 5$ с группа обнаруживает препятствие, в результате чего оно учитывается при распределении на плоскости Oy_1y_2 , и группа начинает перестроение. При перестроении второй подвижный объект приближается к препятствию на расстояние около 3,3 метра, поэтому у него включается неустойчивый режим, корректирующий уставки по положению. В этой связи группа продолжает перестроение. После облета препятствия, приблизительно в момент времени $t = 11$ с, группа выходит из неустойчивого режима, и в момент времени $t = 12$ с препятствие уже не учитывается при формировании строя. Группа гексакоптеров снова перестраивается и продолжает прямолинейное движение.

Результаты моделирования группы гексакоптеров в симуляторе при движении в среде с несколькими подвижными препятствиями представлены на рисунках 9 и 10.

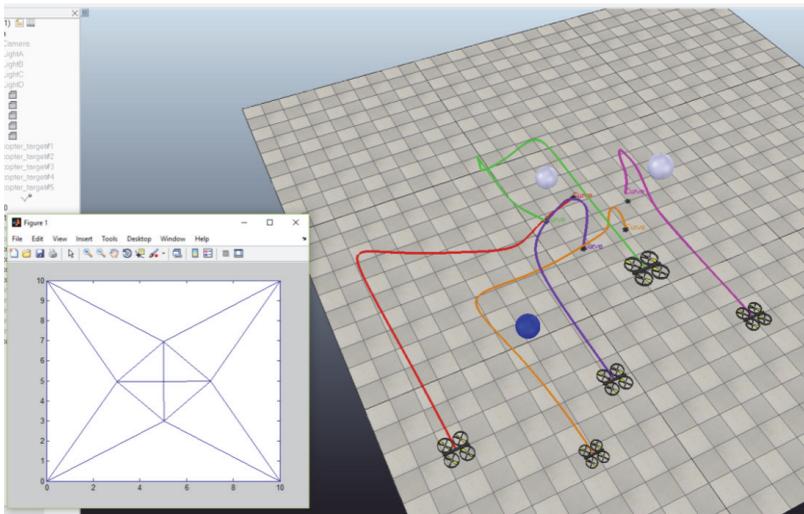


Рис. 9. Полет группы квадрокоптеров в среде с тремя нестационарными препятствиями, вид в плоскости OY_1Y_2

Как видно из рисунка 9, группа успешно осуществляет облет подвижных препятствий, которые отмечены тремя сферами. При этом группа после облета препятствий восстанавливает структуру, сформированную до обнаружения препятствий.

Количественная оценка системы группового управления по результатам численного моделирования осуществляется по минимальному расстоянию до препятствия и СКО отработки скорости при движении группы. Численные оценки указанных параметров представлены в таблице 1.

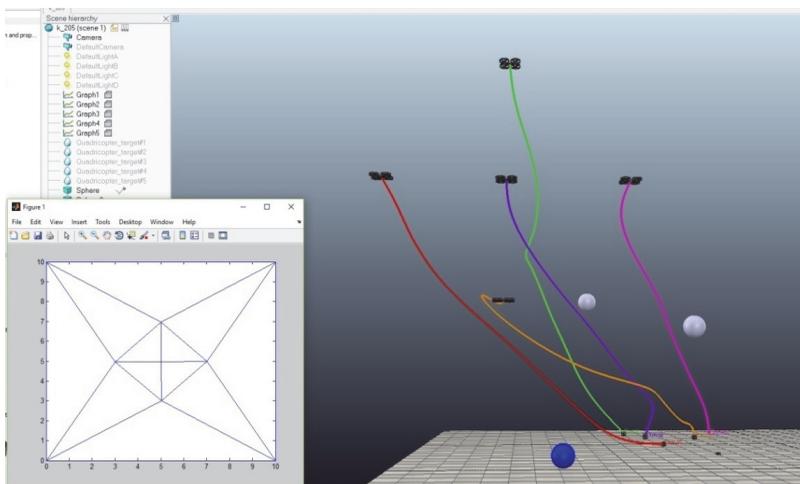


Рис. 10. Полет группы гексакоптеров в среде с тремя нестационарными препятствиями, вид в плоскости OY_1Y_2

Оценка требуемых вычислительных ресурсов осуществлялась путем сравнительного моделирования предлагаемых алгоритмов с методами, базирующимися на виртуальных полях и на использовании нейронных сетей. Результаты представлены в таблице 2. Моделирование проводилось в системе Matlab на ЭВМ с процессором Icore 7. С точки зрения вычислительных затрат, предлагаемые алгоритмы показали высокую эффективность.

Таблица 1. Численные оценки качества системы управления

Параметр	R1	R2	R3	R4	R5
Минимальное расстояние до препятствия, м	8,4	3,3	10,0	5,5	4,5
СКО по скорости, м/с	0.07	0.12	0.15	0.18	0.21

Таблица 2. Время выполнения алгоритмов управления

Метод	Виртуальные поля	Нейронные сети	Неустойчивые режимы
Время выполнения, мс	0,4	1,2	0,4

Экспериментальные исследования проводились с использованием гексакоптера, оснащенного полетным контроллером Pixhawk 2.4.8, спутниковой навигационной системой GPS RadioLink M8N SE100 с компасом, оптическим модулем PX4FLOW, бортовым вычислителем с процессором Intel Atom® x5-Z8350, стереокамерой ELP-960P2CAM-V90 с углом обзора 90 градусов. Заявленное СКО навигационной системы 2,5 м по положе-

нию, 0,1 м/с по скорости. В ходе экспериментов использовано два гексакоптера на полигоне с шириной и высотой 30 м. В целях безопасности неподвижное препятствие симулировалось программно. Результаты экспериментального полета представлены на рисунке 11, на котором построены графики, полученные в результате моделирования (серая линия) и записанные с навигационных систем гексакоптеров (черная линия).

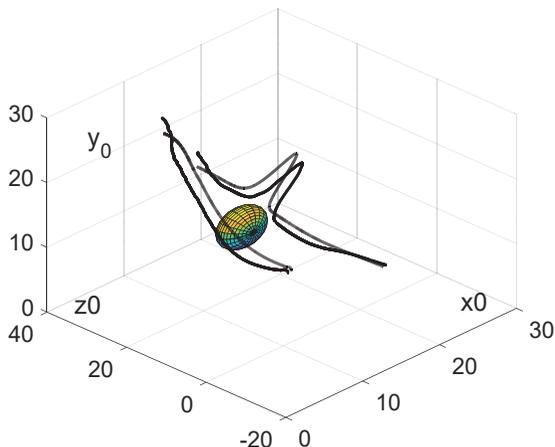


Рис. 11. Экспериментальные и моделируемые траектории

По результатам 8 полетов СКО экспериментальных траекторий от моделируемых составило около 2,0 м. Минимальное расстояние до препятствия составило 4,0 м.

6. Заключение. В статье предложены алгоритмы децентрализованного планирования движения группы мобильных роботов в среде с препятствиями. Разработаны алгоритмы распределения роботов на плоскости, не требующие предварительного задания их координат, основанные на триангуляции Делоне. Мобильные роботы вычисляют свое расположение таким образом, чтобы оптимизировать сумму расстояний между объектами в кластере. Скорость каждого мобильного робота формируется на основе заданной скорости движения группы и текущего положения объектов в кластере. Для обхода препятствий предложены децентрализованные алгоритмы планирования движения, включающие в себя три уровня. На верхнем уровне производится учет обнаруженного препятствия при кластеризации группы на основе триангуляции Делоне. На среднем уровне осуществляется анализ текущей обстановки в кластере и определяется наиболее безопасное направление обхода препятствия. На нижнем уровне при попадании мобильного робота в область препятствия формируются неустойчивые компоненты, корректирующие сфор-

мированные ранее уставки. Указанные алгоритмы планирования движения могут применяться как совместно, так и по отдельности. При этом алгоритмы верхнего уровня функционируют на основе информации о координатах мобильных роботов, границах области функционирования и координатах препятствий. Алгоритмы среднего и нижнего уровня требуют информацию только о расстоянии между соседними объектами в кластере и не требуют картографирования. В результате численных и натуральных экспериментов получены следующие усредненные оценки: минимальное расстояние до препятствия — 3,3 м (моделирование) и 4,0 м (эксперимент); СКО экспериментальных траекторий от моделируемых составило около 2,0 м. Учитывая, что погрешности навигационной системы составляют 2,5 м, можно сделать вывод о хорошем совпадении экспериментальных данных с результатами численного моделирования.

Литература

1. *Pshikhopov V.Kh. et al.* Path Planning for Vehicles Operating in Uncertain 2D Environments // Butterworth-Heinemann: Elsevier. 2017. 312 p.
2. *Yang L. et al.* Survey of Robot 3D Path Planning Algorithms // Journal of Control Science and Engineering. 2016. vol. 2016. 22 p.
3. *LaValle S.* Planning Algorithms // Cambridge University Press. 2006. 842 p.
4. *Каляев И.А., Гайдук А.Р., Капустян С.Г.* Модели и алгоритмы коллективного управления в группах роботов // М.: ФИЗМАТЛИТ. 2009. 280 с.
5. *Макаров И.М. и др.* Модели и алгоритмы планирования действий и распределения заданий в мультиагентных робототехнических системах // Мехатроника, автоматизация, управление. 2012. № 5. С. 44–50.
6. *Mac T.T., Copot C., Tran D.T., De Keyser R.* Heuristic approaches in robot path planning: a survey // Robotics and Autonomous Systems. 2016. vol. 86. pp. 13–28.
7. *Jha A.K.* Intelligent Control and Path Planning of Multiple Mobile Robots Using Hybrid Ai Techniques // PhD thesis. 2016. 213 p.
8. *Лохин В.М. и др.* Механизмы интеллектуальных обратных связей, обработки знаний и самообучения в системах управления автономными роботами и мультиагентными робототехническими группировками // Мехатроника, автоматизация, управление. 2015. Т. 16. № 8. С. 545–555.
9. *Каляев И.А., Капустян С.Г.* Проблемы группового управления роботами // Мехатроника, автоматизация, управление. 2009. № 6. С. 33–40.
10. *Duan H., Qiao P.* Pigeon-inspired optimization: A new swarm intelligence optimizer for air robot path planning // International Journal of Intelligent Computing and Cybernetics. 2014. vol. 7(1). pp. 24–37.
11. *Schulman J. et al.* Motion planning with sequential convex optimization and convex collision checking // International Journal of Robotics Research. 2014. vol. 33(9). pp. 1251–1270.
12. *Montiel O., Orozco-Rosas U., Sepúlveda R.* Path planning for mobile robots using Bacterial Potential Field for avoiding static and dynamic obstacles // Expert Systems with Applications. 2015. vol. 42(12). pp. 5177–5191.
13. *Архипкин А.В. и др.* Задачи группового управления роботами в робототехническом комплексе пожаротушения // Труды СПИИРАН. 2016. Вып. 2(45). С. 116–129.
14. *Пицхонков В.Х., Медведев М.Ю.* Децентрализованное управление группой однородных подвижных объектов в двумерной среде с препятствиями // Мехатроника, автоматизация, управление. 2016. Том 17. № 5. С. 346–353.

15. *Liu Z. et al.* Distributed sampled-data control of nonholonomic multi-robot systems with proximity networks // *Automatica*. 2017. vol. 77. pp. 170–179.
16. *Chetty R.M.K., Singaperumal M., Nagarajan T., Tetsunari I.* Coordination control of wheeled mobile robots – A hybrid approach // *International Journal of Computer Applications in Technology*. 2011. vol. 41(3-4). pp. 195–204.
17. *Williams R.K., Sukhatme G.S.* Constrained interaction and coordination in proximity-limited multiagent systems // *IEEE Transactions on Robotics*. 2013. vol. 29(4). pp. 930–944.
18. *Макаров И.М. и др.* Смешанные стратегии группового управления в многоагентных робототехнических системах // *Известия ЮФУ. Технические науки*. 2012. № 3(128). С. 8–13.
19. *Белоглазов Д.А. и др.* Интеллектуальные технологии планирования перемещений подвижных объектов в трехмерных недетерминированных средах / под ред. В.Х. Пшихопова // М.: Наука. 2017. 232 с.
20. *Pshikhovop V. et al.* Mathematical model of robot on base of airship // *Proceedings of the IEEE Conference on Decision and Control*. 2013. pp. 959–964.
21. *Mulzer W., Rote G.* Minimum-weight triangulation is NP-hard // *Proc. of 22nd Annual Symposium on Computational Geometry*. 2006. pp. 1–10.
22. *Pshikhovop V.Kh., Ali A.S.* Hybrid motion control of a mobile robot in dynamic environments // *Proceedings of IEEE International Conference on Mechatronics*. 2011. pp. 540–545.
23. *Khatib O.* Real-Time Obstacles Avoidance for Manipulators and Mobile Robots // *Int. Journal of Robotics Research*. 1986. vol. 5(1). pp. 90–98.
24. *Bounini F., Gingras D., Pollart H., Gruyer D.* Modified Artificial Potential Field Method for Online Path Planning Applications // *IEEE Intelligent Vehicles Symposium (IV)*. 2017. pp. 1–7.
25. *Платонов А.К., Карпов И.И., Кирильченко А.А.* Метод потенциалов в задаче прокладки трассы // М.: Препринт Института прикладной математики АН СССР. 1974. 27 с.
26. *LeCun Ya., Yoshua B., Geoffrey H.* Deep learning // *Nature*. 2015. vol. 521. pp. 436–444.
27. *Городецкий В.И., Скобелев П.О.* Многоагентные технологии для промышленных приложений: реальность и перспектива // *Труды СПИИРАН*. 2017. Вып. 6(55). С. 11–45.

Пшихопов Вячеслав Хасанович — д-р техн. наук, профессор, директор научно-исследовательского института робототехники и процессов управления, Южный федеральный университет (ЮФУ). Область научных интересов: управление подвижными объектами в условиях неопределенности, методы оптимального управления роботами, анализ и синтез систем группового управления роботами, интеллектуальное управление и планирование в робототехнических системах. Число научных публикаций — 250. pshichop@rambler.ru, <https://sfedu.ru/person/vhpshichop>; ул. Шевченко, 2, Таганрог, 347922; р.т.: +7(8634)371-694, Факс: +7(8634)681-894.

Медведев Михаил Юрьевич — д-р техн. наук, доцент, профессор кафедры электротехники и мехатроники, Южный федеральный университет (ЮФУ). Область научных интересов: адаптивное и робастное управление подвижными объектами, оценивание, методы анализа и синтеза систем управления. Число научных публикаций — 210. medvmihal@gmail.com, <https://sfedu.ru/person/medvmihal>; ул. Шевченко, 2, Таганрог, 347922; р.т.: +7(8634)371-694, Факс: +7(8634)681-894.

Поддержка исследований. Работа выполнена при поддержке РНФ (грант 16-19-00001). Раздел 4 выполнен при поддержке РФФИ (грант 16-08-00013).

V.KH. PSHIKHOPOV, M.YU. MEDVEDEV
**GROUP CONTROL OF AUTONOMOUS ROBOTS MOTION IN
UNCERTAIN ENVIRONMENT VIA UNSTABLE MODES**

Pshikhopov V.Kh., Medvedev M.Yu. Group Control of Autonomous Robots Motion in Uncertain Environment via Unstable Modes.

Abstract. The paper is devoted to the methods and algorithms of group control, providing a consistent movement of a moving objects group in a partially uncertain three-dimensional environment with obstacles. The uncertainty of the environment is a priori unknown moving obstacles. Mobile robots of the group should create a formation in a given rectangular area in the plane. The robots formation must move in the given direction. The problem solution is carried out by using algorithms based on unstable modes. The unstable modes allow to transform obstacles into repellers. The proposed algorithms can be implemented decentralized. Two algorithms options for group control are analyzed in the paper. Also a numerical simulation of a hexacopters group in an uncertain environment with obstacles is performed. The developed algorithms are used in the control system of mobile robots in their group locomotion in an uncertain 3-d environment. A group of mobile robots should be self-organized into a structure that does not require a preliminary assignment of the place of each robot. The group of robots can adjust the formation at occurrence of obstacles.

Keywords: group control, uncertain environment, unstable mode, motion control.

Pshikhopov Viacheslav Khasanovich — Ph.D., Dr. Sci., professor, head of research and development institute of robotics and control systems, Southern Federal University (SFedU). Research interests: control of mobile robots in uncertain environment, optimal control of robots, analysis and synthesis of collective robots control, intelligent control and planning in the robotics. The number of publications — 250. pshichop@rambler.ru, <https://sfedu.ru/person/vhpshichop>; 2, Shevchenko str., Taganrog, 347922, Russia; office phone: +7(8634)371-694, Fax: +7(8634)681-894.

Medvedev Mikhail Yurievich — Ph.D., Dr. Sci., associate professor, professor of electrical sciences and mechatronics department, Southern Federal University (SFedU). Research interests: adaptive and robust control of mobile robots, estimation, analysis and synthesis attitudes of control systems. The number of publications — 210. medvmihal@gmail.com, <https://sfedu.ru/person/medvmihal>; 2, Shevchenko str., Taganrog, 347922, Russia; office phone: +7(8634)371-694, Fax: +7(8634)681-894.

Acknowledgements. This research is supported by Russian Science Foundation (grant 16-19-00001). Section 4 is supported by Russian Foundation for Basic Research (grant 16-08-00013).

References

1. Pshikhopov V.Kh. et al. Path Planning for Vehicles Operating in Uncertain 2D Environments. Butterworth-Heinemann: Elsevier. 2017. 312 p.
2. Yang L. et al. Survey of Robot 3D Path Planning Algorithms. *Journal of Control Science and Engineering*. 2016. vol. 2016. 22 p.
3. LaValle S. Planning Algorithms. Cambridge University Press. 2006. 842 p.
4. Kaliaev I.A., Gaiduk A.R., Kapustyan S.G. *Modeli i algoritmy kolektivnogo upravleniya v gruppah robotov* [Models and algorithms of collective control in groups of robots]. M.: FIZMATLIT. 2009. 280 p. (In Russ.).

5. Makarov I.M. et al. [Action Planning and Tasks Distribution Models and Algorithms for Multiagent Robotic Systems]. *Mehatronika, avtomatizacija, upravlenie – Mechatronics, Automation, Control*. 2012. vol. 5. pp. 44–50. (In Russ.).
6. Mac T.T., Copot C., Tran D.T., De Keyser R. Heuristic approaches in robot path planning: a survey. *Robotics and Autonomous Systems*. 2016. vol. 86. pp. 13–28.
7. Jha A.K. Intelligent Control and Path Planning of Multiple Mobile Robots Using Hybrid Ai Techniques. PhD thesis. 2016. 213 p.
8. Lokhin V.M. et al. [Intelligent Feedback, Knowledge Processing and Self Learning in the Control Systems of the Autonomous Robots and Multi-Agent Robotic Groups]. *Mehatronika, avtomatizacija, upravlenie – Mechatronics, Automation, Control*. 2015. Issue 16. vol. 8. pp. 545–555. (In Russ.).
9. Kaliev I.A., Kapustjan S.G. [Problems of Group Control by Robots]. *Mehatronika, avtomatizacija, upravlenie – Mechatronics, Automation, Control*. 2009. vol. 6. pp. 33–40. (In Russ.).
10. Duan H., Qiao P. Pigeon-inspired optimization: A new swarm intelligence optimizer for air robot path planning. *International Journal of Intelligent Computing and Cybernetics*. 2014. vol. 7(1). pp. 24–37.
11. Schulman J. et al. Motion planning with sequential convex optimization and convex collision checking. *International Journal of Robotics Research*. 2014. vol. 33(9). pp. 1251–1270.
12. Montiel O., Orozco-Rosas U., Sepúlveda R. Path planning for mobile robots using Bacterial Potential Field for avoiding static and dynamic obstacles. *Expert Systems with Applications*. 2015. vol. 42(12). pp. 5177–5191.
13. Petrov V.F. et al. [Problems of Group Control of Robots in the Robotic Complex of Fire Extinguishing]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2016. vol. 2(45). pp. 116–129. (In Russ.).
14. Pshikhopov V.Kh., Medvedev M.Yu. [Decentralized Control of a Group of Similar Objects Moving in a Two-Dimensional Environment with Obstacles]. *Mehatronika, avtomatizacija, upravlenie – Mechatronics, Automation, Control*. 2016. Issue 17. vol. 5. pp. 346–353. (In Russ.).
15. Liu Z. et al. Distributed sampled-data control of nonholonomic multi-robot systems with proximity networks. *Automatica*. 2017. vol. 77. pp. 170–179.
16. Chetty R.M.K., Singaperumal M., Nagarajan T., Tetsunari I. Coordination control of wheeled mobile robots – A hybrid approach. *International Journal of Computer Applications in Technology*. 2011. vol. 41(3-4). pp. 195–204.
17. Williams R.K., Sukhatme G.S. Constrained interaction and coordination in proximity-limited multiagent systems. *IEEE Transactions on Robotics*. 2013. vol. 29(4). pp. 930–944.
18. Makarov I.M. et al. [The mixed strategy of group control in multi-robotics systems]. *Izvestija JuFU. Tehnicheskie nauki – Izvestiya SFEDU. Engineering Sciences*. 2012. vol. 3(128). pp. 8–13. (In Russ.).
19. Beloglazov D.A. et al. *Intellektual'nye tehnologii planirovanija peremeshhenij podvizhnyh ob#ektov v trehmernyh nedeterminirovannyh sredah. Pod red. V.H. Pshihopova* [Intelligent attitudes of mobile objects locomotion planning in 3-D uncertain environment. Edited by pod red. V.H. Pshihopov]. Moscow. Nauka. 2017. 232 p. (In Russ.).
20. Pshikhopov V. et al. Mathematical model of robot on base of airship. Proceedings of the IEEE Conference on Decision and Control. 2013. pp. 959–964.
21. Mulzer W., Rote G. Minimum-weight triangulation is NP-hard. Proc. of 22nd Annual Symposium on Computational Geometry. 2006. pp. 1–10.

22. Pshikhopov V.Kh., Ali A.S. Hybrid motion control of a mobile robot in dynamic environments. Proceedings of IEEE International Conference on Mechatronics. 2011. pp. 540–545.
23. Khatib O. Real-Time Obstacles Avoidance for Manipulators and Mobile Robots. *Int. Journal of Robotics Research*. 1986. vol. 5(1). pp. 90–98.
24. Bounini F., Gingras D., Pollart H., Gruyer D. Modified Artificial Potential Field Method for Online Path Planning Applications. IEEE Intelligent Vehicles Symposium (IV). 2017. pp. 1–7.
25. Platonov A.K., Karpov I.I., Kiril'chenko A.A. *Metod potencialov v zadache prokladki trassy* [Potential field attitude in the problem of path finding]. Moscow. Institute of Applied Mathematics. Academy of Science USSR. 1974. 27 p. (In Russ.).
26. LeCun Ya., Yoshua B., Geoffrey H. Deep learning. *Nature*. 2015. vol. 521. pp. 436–444.
27. Gorodetsky V.I., Skobelev P.O. [Industrial Applications of Multi-agent Technology: Reality and Perspectives]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2017. vol. 6(55). pp. 11–45. (In Russ.).

Э.Я. РАПОПОРТ, Ю.Э. ПЛЕШИВЦЕВА
**МЕТОД МНОГОКРИТЕРИАЛЬНОЙ ОПТИМИЗАЦИИ
УПРАВЛЯЕМЫХ СИСТЕМ С РАСПРЕДЕЛЕННЫМИ
ПАРАМЕТРАМИ**

Рапопорт Э.Я., Плешивцева Ю.Э. Метод многокритериальной оптимизации управляемых систем с распределенными параметрами.

Аннотация. Предлагается конструктивный метод многокритериальной оптимизации процессов управления детерминированными и не полностью определенными управляемыми системами с распределенными параметрами, описываемыми линейными многомерными уравнениями в частных производных параболического типа с внутренними и граничными управляющими воздействиями в условиях заданной точности приближения конечного состояния объекта к требуемому пространственному распределению управляемой величины. Предлагаемый в статье подход базируется на однокритериальной версии в виде минимаксной свертки нормализуемых критериев качества и последующем переходе к эквивалентной форме типовой вариационной задачи с ограничениями. Анализ проводится применительно к детерминированной модели объекта, описываемой бесконечной системой дифференциальных уравнений относительно временных мод разложения управляемой величины в ряд по собственным функциям начально-краевой задачи. Дальнейшие процедуры предварительной параметризации управляющих воздействий на основе аналитических условий оптимальности и редукции к задачам полубесконечного программирования позволяют найти искомые экстремали с использованием их чебышёвских свойств и фундаментальных закономерностей предметной области в типичных для приложений условиях оценки в равномерной метрике точности приближения конечного состояния объекта к требуемому. Полученные результаты распространяются на задачи программного управления по принципу гарантированного результата ансамблями траекторий движения объекта в условиях интервальной неопределенности параметрических характеристик распределенной системы и множественных внешних возмущений. Приводится представляющий самостоятельный интерес пример решения предлагаемым методом задачи многокритериальной оптимизации передовой технологии индукционного нагрева металлических полуфабрикатов перед последующей обработкой давлением. В качестве компонент векторного критерия оптимизации рассматриваются расход энергии на нагрев, потери металла на термохимические взаимодействия и точность достижения заданной конечной температуры нагрева.

Ключевые слова: многокритериальная оптимизация, системы с распределенными параметрами, управляемые системы, полубесконечная оптимизация, равномерные приближения, альтернативный метод, интервальные неопределенности, принцип гарантированного результата.

1. Введение. Характерные требования к оценке эффективности функционирования сложных технических систем по целому ряду конфликтующих друг с другом показателей качества приводят к актуальной задаче многокритериальной оптимизации (ЗМО) искомых управляющих воздействий (УВ) в подобных условиях неопределенности целей управления [1-5]. При этом отбор

приемлемых вариантов производится на парето-эффективном множестве решений (ПЭМР), не улучшаемых ни по одному из критериев без ухудшения показателей по какому-либо из остальных. Для определения ПЭМР в настоящее время используются два основных подхода, базирующихся на предварительной параметризации искомым УВ [1-9].

Во-первых, в целях непосредственного построения ПЭМР эффективным оказывается применение различных модификаций генетических алгоритмов, отличающихся друг от друга способами агрегирования популяций для частных критериев эффективности в сочетании с современными методами численного моделирования оптимизируемых процессов [5, 6]. Этот подход успешно реализован на практике, в частности в сложных ЗМО интегрированного проектирования объектов технологической теплофизики с распределенными параметрами [6-10].

Во-вторых, используется традиционная технология различных вариантов однокритериальной свертки взвешиваемых компонентов векторной целевой функции с последующим построением ПЭМР, которому принадлежат все решения соответствующих однокритериальных задач при варьировании весовых коэффициентов в пределах типового симплекса [1-4].

ЗМО в реальных условиях неполной определенности исходных данных могут быть решены на основе указанных подходов различными методами. В том числе путем формирования критериев оптимальности в терминах функций принадлежности в рамках теории нечетких множеств [11] или в форме математического ожидания при заданной функции плотности распределения вероятности неопределенной величины [12]; путем получения недостающей информации методами регрессионного анализа [2]; решением ЗМО в условиях ограниченной неопределенности по принципу гарантированного результата [4].

После построения ПЭМР возникает самостоятельная проблема выбора в пределах множества Парето единственной альтернативы при наличии значительного числа возможных вариантов. В частности, при переходе к относительным оценкам предварительно нормализуемых равнозначных составляющих векторного критерия качества с последующим использованием их минимаксной (или максиминной) свертки с единичными весовыми коэффициентами решение соответствующей однокритериальной задачи заведомо является парето-эффективным и отвечает искомому альтернативному варианту решения ЗМО по принципу гарантированного результата [2, 3].

Искомые УВ при этом непосредственно задаются в параметризованной форме в ЗМО проектных решений или статических режимов работы объекта, которые в результате сводятся к типовым задачам математического программирования (ЗМП) с учетом заданных ограничений [2-4].

При многокритериальной оптимизации динамических процессов предварительная процедура параметризации УВ требует применения известных аналитических условий оптимальности и представляет собой отдельную достаточно сложную задачу [13, 14]. Трудности перехода к ЗМП в динамических ЗМО в условиях необходимости использования прямых решений системы дифференциальных уравнений модели объекта с параметризованными УВ резко возрастают с повышением порядка этих уравнений и приобретают принципиальный характер применительно к бесконечномерным объектам с распределенными параметрами (ОРП). В этом случае ЗМО сводятся вместо типовых задач к специальным задачам полубесконечной оптимизации с бесконечным числом ограничений на конечное состояние ОРП в характерной ситуации его оценки в равномерной метрике на множестве определения пространственных аргументов управляемых величин [8, 13, 14].

В настоящей работе предлагаются постановки и конструктивный способ решения отличающихся указанными затруднениями малоисследованных многокритериальных задач оптимального управления ОРП, в значительной степени опирающийся на разработанный ранее авторами альтернативный метод параметрической оптимизации управляемых процессов [8, 13, 14].

2. Задача многокритериального управления детерминированными моделями объекта с распределенными параметрами.

Пусть управляемая функция состояния $Q(X, t)$ объекта с распределенными параметрами описывается в зависимости от времени $t \in (0, t^*)$ и пространственных координат $X \in V$; $X = (x_i)$, $i = \overline{1, m}$;

$1 \leq m \leq 3$ в пределах заданной односвязной области V m -мерного евклидова пространства E^m с кусочно-гладкой границей S линейным неоднородным уравнением в частных производных параболического типа с постоянными во времени коэффициентами:

$$\frac{\partial Q}{\partial t} = \sum_{i=1}^m a_i(X) \frac{\partial^2 Q}{\partial x_i^2} + \sum_{i=1}^m b_i(X) \frac{\partial Q}{\partial x_i} + c(X)Q + g_V(X)u_V(t) + g_{V1}(X)\eta_1(t), \quad (1)$$

с начальными и граничными условиями:

$$\begin{aligned} Q(X, 0) &= f(X), \quad X \in V \cup S; \\ \frac{\partial Q}{\partial N} + \beta Q &= g_S(X)u_S(t) + g_{S1}(X)\eta_2(t), \quad X \in S; \end{aligned} \quad (2)$$

и сосредоточенными внутренними $u_V(t)$ или (и) граничными $u_S(t)$ управляющими воздействиями при изначально фиксируемых функциях $g_V(X)$, $g_S(X)$ их пространственного распределения. Здесь N – вектор внешней нормали к S ; $\beta = const \geq 0$; t^* — конечный момент процесса управления; дифференциальный оператор в правой части (1) самосопряженный; функция $f(X) \in L_2(V)$ и коэффициенты $a_i(X), b_i(X), c(X)$ являются известными достаточно гладкими функциями своих аргументов, причем не все $a_i(X)$ в (1) одновременно равны нулю; $\eta_1(t), \eta_2(t)$ — аддитивные по отношению к $u_V(t)$ и $u_S(t)$ кусочно-непрерывные возмущения с заданным характером $g_{V1}(X)$ и $g_{S1}(X)$ их зависимости от X .

Управляющие воздействия $u(t) \in \{u_V(t), u_S(t)\}$ должны быть подчинены ограничениям:

$$u_{\min} \leq u(t) \leq u_{\max}, \quad t \in [0, t^*], \quad (3)$$

с известными предельно допустимыми значениями u_{\min} и u_{\max} .

Рассмотрим далее ЗМО детерминированной модели ОРП (1)-(3) при $\eta_1(t) = \eta_2(t) = 0$ в (1), (2).

Применение к уравнениям (1), (2) метода конечных интегральных преобразований приводит к описанию модели ОРП бесконечной системой обыкновенных дифференциальных уравнений относительно временных мод $\bar{Q}_n(t)$ — коэффициентов разложения $Q(X, t)$ в сходящийся в среднем ряд по ортогональной с весом $r(X)$ системе собственных функций $\phi_n(X)$ [15]:

$$\frac{d\bar{Q}_n}{dt} = -\mu_n^2 \bar{Q}_n + k_{Vn} u_V(t) + k_{Sn} u_S(t), \quad n = 1, 2, \dots; \quad \bar{Q}_n(0) = \bar{f}_n, \quad (4)$$

$$Q(X, t) = \sum_{n=1}^{\infty} \bar{Q}_n(t) \phi_n(X). \quad (5)$$

Здесь μ_n^2 — собственные числа, и

$$\begin{aligned} \bar{f}_n &= \int_V f(Y)\phi_n(Y)r(Y)dY; \bar{k}_{Vn} = \int_V g_V(Y)\phi_n(Y)r(Y)dY; \\ \bar{k}_{Sn} &= \int_S g_S(Y)\phi_n(Y)r(Y)dS(Y). \end{aligned} \quad (6)$$

Рассмотрим $q > 1$ различных частных задач оптимального управления объектом (4)-(5) с интегральными критериями качества:

$$I_p(u) = \int_0^{t^*} f_{0p}(\bar{Q}_p, u)dt \rightarrow \min_{u \in \Omega_p}, p = \overline{1, q}, \bar{Q}_p = (\bar{Q}_{pn}(t)), \quad (7)$$

с заданными подынтегральными достаточно гладкими функциями f_{0p} своих аргументов (задачи на максимум $I_p, p \in \{\overline{1, q}\}$ приводятся к минимизации $-I_p$). Задачи (7) формулируются в условиях ограничений на управляющие воздействия, отвечающих требованиям обеспечения допустимой точности ε_{0p} приближения $Q_p(X, t^*)$ к заданному пространственному распределению $Q^{**}(X) = Q_c = const$, которая оценивается в равномерной метрике применительно к представляющим наибольший интерес ситуациям в прикладных задачах и определяет целевое множество достижимых конечных состояний ОРП [13, 14]:

$$\Omega_p = \{u : \Phi_p(u) \leq \varepsilon_{0p}, u \in [u_{\min}, u_{\max}]\}; \quad (8)$$

$$\Phi_p(u) = \max_{X \in V} |Q_p(X, t^*) - Q_c| = \max_{X \in V} \left| \sum_{n=1}^{\infty} \bar{Q}_{pn}(t^*)\phi_n(X) - Q_c \right|. \quad (9)$$

Здесь в (7)-(9) \bar{Q}_p — вектор модальных переменных $\bar{Q}_{pn}(t)$ управляемой величины $Q_p(X, t^*)$ и $\varepsilon_{0p} \geq \varepsilon_{p \inf} \geq 0$, где $\varepsilon_{p \inf}$ — точная нижняя граница достижимых значений ε_{0p} [13].

При необходимости оценки качества процесса управления одновременно по всем q показателям эффективности I_p в (7) может быть сформулирована на основании (7)-(9) многокритериальная задача [1-5]:

$$I_{\Sigma} \rightarrow \min; I_{\Sigma} = (I_p(u)), p = \overline{1, q}, \quad (10)$$

$$\Omega_{\Sigma} = \{u : \Phi_{\Sigma}(u) \leq \varepsilon_0, u \in [u_{\min}, u_{\max}]\}, 0 \leq \varepsilon_{\inf} \leq \varepsilon_0 \leq \max_p \varepsilon_{0p};$$

$$\Phi_{\Sigma}(u) = \max_{X \in V} |Q(X, t^*) - Q_c| = \max_{X \in V} \left| \sum_{n=1}^{\infty} \overline{Q}_n(t^*) \phi_n(X) - Q_c \right| \quad (11)$$

минимизации q критериев I_p согласно (10) с учетом ограничения (11) на априори фиксируемую по условиям выполнения неравенств $\varepsilon_0 \leq \max_p \varepsilon_{0p}$ величину ε_0 погрешности равномерного приближения $Q(X, t^*)$ к Q_c при управлении объектом (4). Здесь исключается тривиальный случай совпадения решений u_p^* задач (7)-(9) по каждому из критериев I_p для всех $p = \overline{1, q}$.

Общая теория управления не содержит методов решения ЗМО управляемых бесконечномерных объектов вида (4), (10), (11) с ограничением на оцениваемую в равномерной метрике точность приближения конечного состояния управляемой величины к требуемому. В настоящем разделе статьи предлагается новый конструктивный способ решения такой задачи.

2.1. Эквивалентная форма детерминированной однокритериальной задачи оптимального управления. Не влияющий на результат решения задачи (10)-(11) переход от I_p к относительным равнозначным оценкам [2, 3]:

$$\lambda_p(u) = \frac{I_p(u) - I_p^*}{I_p^0 - I_p^*}; 0 \leq \lambda_p(u) \leq 1; p = \overline{1, q}, \quad (12)$$

приводит (в отличие от типовой формы с весовыми коэффициентами) к минимаксной свертке ЗМО (10)-(11) следующего вида:

$$\lambda(u) = \max_{p \in \{\overline{1, q}\}} \lambda_p(u) \rightarrow \min, \quad u \in \Omega_{\Sigma} \quad (13)$$

точка оптимума которой:

$$u^{**} = \arg \min_{u \in \Omega_{\Sigma}} \left[\max_{p \in \{\overline{1, q}\}} \lambda_p(u) \right], \quad (14)$$

априори принадлежит множеству Парето исходной ЗМО и может рассматриваться в качестве искомого оптимального управления в

исходной задаче [2, 3]. В выражении (12) I_p^* — минимальная величина I_p , достигаемая при управлении u_p^* в условиях (8):

$$I_p^* = \min_{u \in \Omega_p} I_p(u); u_p^* = \arg \min_{u \in \Omega_p} I_p(u); \lambda_p(u_p^*) = 0, \quad (15)$$

и I_p^0 — «наихудшее» значение $I_p(u)$ на множестве Ω_p .

Минимаксная задача (13) эквивалентна обычной однокритериальной задаче с интегральным функционалом качества:

$$I = \frac{1}{t^*} \int_0^{t^*} \lambda^0 dt = \lambda^0 \rightarrow \min; \lambda^0 = const, \quad (16)$$

ограничениями:

$$\lambda_p(u) \leq \lambda^0, p = \overline{1, q}, \quad (17)$$

и минимизируемым параметром λ^0 , вводимым согласно уравнению:

$$d\lambda^0/dt = 0. \quad (18)$$

В итоге исходная ЗМО (10)-(11) сводится без каких-либо погрешностей в рамках рассматриваемых моделей к отысканию решения u^{**} задачи (16)-(18) с одним критерием оптимальности I в условиях заданных ограничений (11) и (17).

Введем в рассмотрение описываемые уравнениями

$$dz_p/dt = f_{0p}; z_p(t^*) = I_p(u); p = \overline{1, q} \quad (19)$$

новые переменные $z_p(u); p = \overline{1, q}$, в терминах которых ограничения (17) принимают следующий вид:

$$\frac{z_p(t^*) - I_p^*}{I_p^0 - I_p^*} \leq \lambda^0, p = \overline{1, q}. \quad (20)$$

Дополнив модель объекта (4) уравнениями (18), (19), приходим теперь к следующей эквивалентной форме задачи (16)-(18).

Необходимо определить стесняемое условиями (3) управляющее воздействие $u^{**}(t)$, которое переводит объект

управления (4), (18), (19) в требуемое согласно (11), (20) конечное состояние при минимальном значении критерия оптимальности (16).

На задачу (11), (16), (20) оптимального управления бесконечномерным объектом (4), (19) распространяется принцип максимума Понтрягина.

Базовое условие:

$$H(\bar{Q}^{**}(t), u^{**}(t), \bar{\Psi}^{**}(t)) = \max_{u \in [u_{\min}, u_{\max}]} H(\bar{Q}^{**}(t), u(t), \bar{\Psi}^{**}(t)), t \in (0, t^*) \quad (21)$$

достижения максимума функции Понтрягина [16]:

$$\begin{aligned} H(\bar{Q}(t), u(t), \bar{\Psi}(t)) = & -\lambda^0 + \sum_{n=1}^{\infty} \Psi_n(t) (-\mu_n^2 \bar{Q}_n + k_{vn} u_v(t) + k_{sn} u_s(t)) + \\ & + \sum_{p=1}^q \tilde{\Psi}_p f_{0p}, \end{aligned} \quad (22)$$

в оптимальном процессе $(\bar{Q}^{**}(t), u^{**}(t), \bar{\Psi}^{**}(t))$, где вектор сопряженных переменных $\bar{\Psi}(t) = (\tilde{\Psi}_p, p = 1, q; \Psi_n(t), n = 1, 2, \dots)$ описывается уравнениями:

$$\frac{d\tilde{\Psi}_p}{dt} = -\frac{\partial H}{\partial z_p} = 0, p = 1, q; \frac{d\Psi_n}{dt} = -\frac{\partial H}{\partial \bar{Q}_n}, n = 1, 2, \dots \quad (23)$$

и, как правило, позволяет найти $u^{**}(t)$ в форме явной функции $u^{**}(\bar{Q}^{**}(t), \bar{\Psi}^{**}(t))$ переменных $\bar{Q}(t) = (\bar{Q}_n)$ и $\bar{\Psi}(t)$, вид которой не зависит от ограничений (11) и (20) на конечное состояние объекта управления.

2.2. Редукция к задаче полубесконечной оптимизации.

Найденные согласно (21)-(23) зависимости $u^{**}(\bar{Q}^{**}(t), \bar{\Psi}^{**}(t))$ во многих случаях непосредственно устанавливают структуру искомого оптимального алгоритма $u^{**}(t)$ (и аналогичным образом $u_p^*(t)$ в задаче (15)) в форме их параметрического описания с точностью до вектора $\Delta^{(N)} = (\Delta_i^{(N)}), i = 1, N$, упорядоченной последовательности конечного числа N параметров $\Delta_1^{(N)}, \Delta_2^{(N)}, \dots, \Delta_N^{(N)}$, непосредственно характеризующих поведение $u^{**}(t)$, $u_p^*(t)$ в области их определения [13, 14].

Интегрирование уравнений объекта (4) с параметризованным управлением $u(\Delta^{(N)}, t)$ приводит в таких случаях к представлению критериев оптимальности I_p в (7), λ_p в (12) и конечных состояний $Q_p(X, t^*)$, $Q(X, t^*)$ в (9), (11) в виде явных достаточно гладких зависимостей $I_p(\Delta^{(N)})$, $\lambda_p(\Delta^{(N)})$, $Q_p(X, \Delta^{(N)})$ и $Q(X, \Delta^{(N)})$ от $\Delta^{(N)}$, если считать известными значения I_p^* и I_p^0 в (12).

Размерность N вектора $\Delta^{(N)}$ устанавливается по известным правилам в зависимости от значения ε_0 в (11), либо априори фиксируется возможностями технической реализации $\Delta^{(N)}$ — параметризуемых управляющих воздействий [13].

В результате производится точная редукция задачи (16)-(20) оптимального управления объектом (4), (5), (18), (19) к задаче минимизации целевой функции:

$$I = \lambda^0(\Delta^{(N)}) \rightarrow \min_{\Delta^{(N)} \in G_N}, \quad (24)$$

конечного числа N переменных $\Delta_i^{(N)}, i = \overline{1, N}$ на допустимом множестве $G_N \ni \Delta^{(N)}$ их изменения с бесконечным числом ограничений:

$$\Phi_\Sigma(\Delta^{(N)}) = \max_{X \in V} |Q(X, \Delta^{(N)}) - Q_c| \leq \varepsilon_0, \quad (25)$$

диктуемых требованиями (11) для всех $X \in V$ и дополняемых в соответствии с (17), (20) неравенствами:

$$\lambda_p(\Delta^{(N)}) = \frac{I_p(\Delta^{(N)}) - I_p^*}{I_p^0 - I_p^*} \leq \lambda^0, p = \overline{1, q}. \quad (26)$$

(задача полубесконечной оптимизации (ЗПО)) [13].

ЗПО (24)-(26) эквивалентна минимаксной задаче полубесконечной оптимизации без дополнительных условий (26) с критерием оптимальности:

$$I_1 = \max_{p \in \{\overline{1, q}\}} \lambda_p(\Delta^{(N)}) \rightarrow \min_{\Delta^{(N)} \in G_N}, \quad (27)$$

и континуумом ограничений (25).

Рассматривая вместо (27) критерий оптимальности:

$$I_{2p_1} = \lambda_{p_1}(\Delta^{(N)}) \rightarrow \min_{\Delta^{(N)} \in G_N}, \quad (28)$$

последовательно для всех $p_1 = \overline{1, q}$, получим ряд ЗПО (25), (28) рассмотренного в [13] типового вида, на решениях которых $\Delta^{(N)} = \Delta^{[p_1]}$, определяемых по схеме альтернансного метода [13], вычисляются $\lambda_p(\Delta^{[p_1]}) \forall p = \overline{1, q}$. За решение $\overline{\Delta}^{(N)}$ минимаксной задачи (25), (27) принимается $\Delta^{[p_1^*]}$ при таких $p_1 = p_1^*$, для которых выполняются неравенства $\lambda_p(\Delta^{[p_1^*]}) \leq \lambda_{p_1^*}(\Delta^{[p_1^*]})$ для всех $p = \overline{1, q}$.

Исходная ЗПО (24)-(26) значительно упрощается как раз в наиболее характерной для приложений ситуации в (25) с минимально достижимой величиной ε_0 в классе $\Delta^{(N)}$ -параметризуемых управлений:

$$\varepsilon_0 = \min \varepsilon(\Delta^{(N)}) = \min_{\Delta^{(N)} \in G_N} \left[\max_{X \in V} |Q(X, \Delta^{(N)}) - Q_c| \right]. \quad (29)$$

В таком случае задача (24)-(26), (29) сводится к минимаксной задаче без ограничений

$$\Phi_{\Sigma}(\Delta^{(N)}) \rightarrow \min_{\Delta^{(N)} \in G_N}, \quad (30)$$

на решениях которой автоматически удовлетворяются условия (26) при определении в качестве λ_0 максимального из значений λ_p .

Решения $\overline{\Delta}^{(N)}$ задач (25), (28) при требуемом значении ε_0 в (25) обладают альтернансными свойствами [13], согласно которым в условиях некоторых малостеснительных допущений выполняются равенства:

$$\left| Q(X_j^0, \overline{\Delta}^{(N)}) - Q_c \right| = \varepsilon_0, j = \overline{1, R}, R = \begin{cases} N, \varepsilon_0 > \min \varepsilon(\Delta^{(N)}); \\ N+1, \varepsilon_0 = \min \varepsilon(\Delta^{(N)}); \end{cases} \quad (31)$$

в некоторых точках $X_j^0 \in V, j = \overline{1, R}$. Число R этих точек оказывается равным числу всех искомым параметров оптимального процесса, включая все компоненты $\overline{\Delta}_i^{(N)}, i = \overline{1, N}$ вектора $\overline{\Delta}^{(N)}$ при заданной величине $\varepsilon_0 > \min \varepsilon(\Delta^{(N)})$ и наряду с ними априори неизвестную величину минимакса $\min \varepsilon(\Delta^{(N)})$ в условиях (29).

При наличии дополнительной информации из предметной области о конфигурации зависимостей $Q(X, \bar{\Delta}^{(N)})$ от $X \in V$, позволяющей идентифицировать значения X_j^0 и знаки разностей $Q(X_j^0, \bar{\Delta}^{(N)}) - Q_c$, равенства (31) вместе с условиями:

$$\frac{\partial}{\partial X} Q(X_j^0, \bar{\Delta}^{(N)}) = 0, v = \overline{1, R_1}, \quad (32)$$

существования экстремума функции $Q(X, \bar{\Delta}^{(N)})$ в точках $X_{jv}^0 \in \text{int } V, v = \overline{1, R_1}, R_1 \leq R$, трансформируются к замкнутой системе $R + R_1$ уравнений. Решение этой системы (как правило, единственное) относительно всех искомых величин $\bar{\Delta}_i^{(N)}, i = \overline{1, N}$; $\min \varepsilon(\Delta^{(N)})$ в случае (29) и промежуточных неизвестных $X_{jv}^0, v = \overline{1, R_1}$ находится известными численными методами.

По найденному указанным способом вектору $\bar{\Delta}^{(N)}$ и известным зависимостям $\lambda_p(\Delta^{(N)})$ вычисляется минимальное значение λ_{\min}^0 целевой функции (24):

$$\lambda_{\min}^0 = \min_{\Delta^{(N)} \in G_N} \lambda^0(\Delta^{(N)}) = \lambda^0(\bar{\Delta}^{(N)}) = \max_{p \in \{\overline{1, q}\}} \lambda_p(\bar{\Delta}^{(N)}), \quad (33)$$

обеспечивая выполнение неравенств (26) и завершая тем самым решение задачи (24)-(26) в условиях $\varepsilon_0 \geq \min \varepsilon(\Delta^{(N)})$.

Для решения задач (25), (28) и определения λ_{\min}^0 в (33) необходимо предварительно найти I_p^* и I_p^0 для всех $p = \overline{1, q}$ в (12).

Частные однокритериальные задачи оптимизации (7)-(9) после параметризации искомых управлений $u_p^*(t)$ в (15) с использованием стандартной процедуры принципа максимума вида (21)-(23) приводятся подобно (24), (25) к ЗПО:

$$I_p([\Delta^{(N)}]_p) \rightarrow \min_{[\Delta^{(N)}]_p}; \Phi_p([\Delta^{(N)}]_p) = \max_{X \in V} |Q(X, [\Delta^{(N)}]_p) - Q_c| \leq \varepsilon_{0p}, \quad (34)$$

которую необходимо рассматривать, как это следует из выражения (22) для функции Понтрягина, на элементах $[\Delta^{(N)}]_p$

подмножества $\{\Delta^{(N)}\}_p \subset \{\Delta^{(N)} \in G_N\}$ значений $\Delta^{(N)}$ в p -ой задаче (34).

При этом в условиях (29) достаточно рассмотреть задачи (7)-(9), принимая аналогичным образом $\varepsilon_{0p} = \min \varepsilon([\Delta^{(N)}]_p)$, где

$$\min \varepsilon([\Delta^{(N)}]_p) = \min_{[\Delta^{(N)}]_p} \left[\max_{X \in V} |Q(X, [\Delta^{(N)}]_p) - Q_c| \right], \quad (35)$$

и, следовательно, заведомо выполняется условие $\varepsilon_0 \leq \max_p \varepsilon_{0p}$ в (10), (11).

На решениях $\Delta^{[p]}$ задач (34), определяемых по описанной схеме альтернансного метода, находятся величины:

$$I_p^* = I_p(\Delta^{[p]}), p = \overline{1, q}. \quad (36)$$

Максимально возможные значения I_p^0 критерия I_p формально задаются зависимостями:

$$I_p^0 = I_p(\tilde{\Delta}^{[p]}); \tilde{\Delta}^{[p]} = \arg \max_{[\Delta^{(N)}]_p} \left(I_p([\Delta^{(N)}]_p : \Phi_p([\Delta^{(N)}]_p) \leq \varepsilon_{0p} \right), \quad (37)$$

однако величины $I_p^0, p = \overline{1, q}$ могут быть установлены проще, например, по следующему алгоритму:

$$I_k^0 = \max_{\substack{p \in \{\overline{1, q}\} \\ p \neq k}} I_k(\Delta^{[p]}), k = \overline{1, q} \quad (38)$$

или исходя из физических соображений в каждой конкретной задаче.

В итоге решение по описанной схеме альтернансного метода задач полубесконечной оптимизации (34) и (24)-(26) с вычислением λ_{\min}^0 , I_p^* и I_p^0 согласно (33), (36) и (37) (или (38)) исчерпывает решение исходной ЗМО (13).

3. Задача многокритериальной оптимизации в условиях интервальной неопределенности параметрических характеристик объекта управления. Самостоятельный интерес представляет ЗМО ОРП в характерных условиях интервальной неопределенности начального состояния $\bar{Q}^0 = (\bar{Q}_n(0))$ и неизменного во времени вектора w параметрических характеристик объекта управления (1)-(4), с

точностью до которых определены коэффициенты a_i, b_i, c и β в начально-краевой задаче (1), (2), где по-прежнему, как и в разделе 2, положим $\eta_1(t) = \eta_2(t) = 0$.

Все сведения о \bar{Q}^0 и w исчерпываются в такой ситуации информацией об их принадлежности заданным компактным множествам B и W :

$$\bar{Q}^0 \in B, w \in W. \quad (39)$$

Каждой фиксированной паре значений неопределенных факторов $y = (\bar{Q}^0, w) \in A = B \times W$ при любом допустимом согласно (3) управлении $u(t)$ соответствует определяемая решением системы уравнений (4) траектория процесса $\bar{Q}(t, y) = (\bar{Q}_n(t, y))$ и управляемая функция состояния:

$$Q(X, y, t) = \sum_{n=1}^{\infty} \bar{Q}_n(t, y) \phi_n(X), \quad (40)$$

согласно (5). Объединение этих траекторий по всем допустимым величинам y при одном и том же управляющем воздействии образует ансамбль:

$$\bar{Q}^{\Sigma}(t, y) = \cup \left\{ \bar{Q}(t, \bar{Q}^0, w) : \bar{Q}^0 \in B, w \in W, t \in [0, t^*] \right\}, \quad (41)$$

вместо единственной траектории в его составе при фиксированных значениях y в детерминированном случае.

Конструктивные способы решения подобных ЗМО ОРП неизвестны. В настоящем разделе на многокритериальные задачи оптимизации поведения распределенных объектов в условиях неопределенности (39) распространяется предлагаемый метод решения детерминированных ЗМО.

Стратегия управления ансамблем траекторий (41) по принципу гарантированного результата приводит теперь вместо (7)-(9) к частным задачам минимаксной оптимизации:

$$\begin{aligned} \tilde{I}_p(u) = \max_{y \in A} I_p(u, y) = \max_{y \in Y} \int_0^{t^*} f_{0p}(\bar{Q}_p(y), u) dt \rightarrow \min_{u \in \Omega_p}; p = \bar{1}, q; \\ \tilde{\Omega}_p = \left\{ u : \tilde{\Phi}_p(u) \leq \varepsilon_{0p}; u \in [u_{\min}, u_{\max}] \right\}; \end{aligned} \quad (42)$$

$$\begin{aligned}
 \tilde{\Phi}_p(u) &= \max_{y \in A} \left[\max_{X \in V} \left| Q_p(X, y, t^*) - Q_c \right| \right] = \\
 &= \max_{y \in A} \left[\max_{X \in V} \left| \sum_{n=1}^{\infty} \bar{Q}_{pn}(t^*, y) \phi_n(X) - Q_c \right| \right].
 \end{aligned} \tag{43}$$

Здесь учтены зависимости подынтегральных функций f_{0p} в (7) от u и условия необходимости достижения заданной в (8) величины ε_{0p} для всех допустимых реализаций u на множестве A .

На этом основании формулируется следующая многокритериальная задача оптимизации вместо (10), (11):

$$\tilde{I}_\Sigma \rightarrow \min; \tilde{I}_\Sigma = (\tilde{I}_p(u)), p = \overline{1, q}; \tag{44}$$

$$\tilde{\Omega}_\Sigma = \{u : \tilde{\Phi}_\Sigma(u) \leq \varepsilon_0, u \in [u_{\min}, u_{\max}]\}; \varepsilon_0 \leq \max_p \varepsilon_{0p};$$

$$\begin{aligned}
 \tilde{\Phi}_\Sigma(u) &= \max_{y \in A} \left[\max_{X \in V} \left| Q(X, y, t^*) - Q_c \right| \right] = \\
 &= \max_{y \in A} \left[\max_{X \in V} \left| \sum_{n=1}^{\infty} \bar{Q}_n(t^*, y) \phi_n(X) - Q_c \right| \right].
 \end{aligned} \tag{45}$$

В качестве искомого оптимального управления в ЗМО (44), (45) опять можно рассматривать экстремали $\tilde{u}^{**}(t)$ однокритериальной вариационной версии, переход к которой осуществляется по схеме (12)-(18):

$$\begin{aligned}
 \tilde{I} &= \frac{1}{t^*} \int_0^{t^*} \tilde{\lambda}_0 dt = \tilde{\lambda}_0 \rightarrow \min_{u \in \Omega_\Sigma}; \frac{d\tilde{\lambda}_0}{dt} = 0; \tilde{\lambda}_p(u) = \frac{\tilde{I}_p(u) - \tilde{I}_p^*}{\tilde{I}_p^0 - \tilde{I}_p^*} \leq \tilde{\lambda}_0; p = \overline{1, q}; \\
 \tilde{I}_p^* &= \min_{u \in \Omega_p} \tilde{I}_p(u); \tilde{u}_p^*(t) = \arg \min_{u \in \Omega_p} \tilde{I}_p(u); \tilde{I}_p^0 = \max_{u \in \Omega_p} \tilde{I}_p(u).
 \end{aligned} \tag{46}$$

Эквивалентный характер решений задач управления ансамблем (41) и «изолированной» траекторией в его составе приводит к аналогичной детерминированной задаче структуре алгоритмов $\tilde{u}^{**}(t)$ и $\tilde{u}_p^*(t)$, получаемых с использованием принципа максимума подобно (21)-(23) в параметризованной форме с точностью до конечномерного вектора $\Delta^{(N)}$ параметров, к определению которого и

сводится дальнейшая проблема. В итоге, аналогично (24)-(26), с учетом выражений для $\tilde{I}_p(u)$ и $\tilde{\Phi}_\Sigma(u)$ в (42), (45) производится точная редукция задачи (46) к задаче полубесконечной оптимизации:

$$\tilde{I} = \tilde{\lambda}_0(\Delta^{(N)}) \rightarrow \min_{\Delta^{(N)} \in G_N} \quad (47)$$

$$\tilde{\Phi}_\Sigma(\Delta^{(N)}) = \max_{z \in Z} |Q(z, \Delta^{(N)}) - Q_c| \leq \varepsilon_0; z = (X, y); Z = V \times A \quad (48)$$

$$\tilde{\lambda}_p(\Delta^{(N)}) = \max_{y \in A} \frac{I_p(\Delta^{(N)}, y) - \tilde{I}_p^*}{\tilde{I}_p^0 - \tilde{I}_p^*} \leq \tilde{\lambda}_0; p = \overline{1, q}. \quad (49)$$

Эта задача отличается от (24)-(26), во-первых, ограничением на функцию максимума в (48), которое рассматривается на расширенном по сравнению с (25) множестве элементов, включающем кроме пространственных аргументов $X \in V$ все допустимые реализации неопределенных факторов $y \in A$, и, во-вторых, определением в (49) значения $\tilde{\lambda}_p(\Delta^{(N)})$ по максимальной величине $I_p(\Delta^{(N)}, y)$ на множестве $A \ni y$.

Задача (47)-(49) сводится по схеме (25), (27) к ЗПО:

$$\tilde{I}_1 = \max_{p \in \{\overline{1, q}\}} \tilde{\lambda}_p(\Delta^{(N)}) \rightarrow \min_{\Delta^{(N)} \in G_N}; \quad (50)$$

$$\tilde{\Phi}_\Sigma(\Delta^{(N)}) \leq \varepsilon_0 \quad (51)$$

без ограничений (49). Такая задача, в свою очередь, редуцируется к q задачам типовой формы вида (28), (25) с критериями оптимальности:

$$\tilde{I}_{2p_1} = \tilde{\lambda}_{p_1}(\Delta^{(N)}) \rightarrow \min_{\Delta^{(N)} \in G_N}, p_1 = \overline{1, q} \quad (52)$$

и ограничениями (51), либо в частном случае:

$$\varepsilon_0 = \min \tilde{\varepsilon}(\Delta^{(N)}) = \min_{\Delta^{(N)} \in G_N} \left[\max_{z \in Z} |Q(z, \Delta^{(N)}) - Q_c| \right], \quad (53)$$

аналогичном (29), исходная ЗПО (47)-(49) сводится к минимаксной задаче без ограничений:

$$\tilde{\Phi}_\Sigma(\Delta^{(N)}) \rightarrow \min_{\Delta^{(N)} \in G_N}. \quad (54)$$

На задачу (51), (52) распространяются при малостеснительных для приложений допущениях, подобно детерминированному случаю, технология альтернативного метода отыскания ее решения $\Delta^{(N)}$, базирующаяся на его альтернативных свойствах. Эти свойства усложняются по сравнению с (31), (32) заменой точек $X_j^0 \in V$ на элементы $z_j^0 = (X_j^0, y_j^0) \in Z$, в которых достигается максимум в (48), равный ε_0 , и дополнительной системой соотношений для значений $\max_{y \in A} \lambda_{p_l}(\Delta^{(N)}, y)$, достигаемых в точках \tilde{y}_i^0 [13]:

$$\begin{aligned} & \left| Q(z_j^0, \bar{\Delta}^{(N)}) - Q_c \right| = \varepsilon_0, j = \overline{1, R_X}; \\ & \frac{\partial}{\partial X} Q(z_{j\nu}^0, \bar{\Delta}^{(N)}) = 0, \nu = \overline{1, R_{X1}}; R_{X1} \leq R_X, z_{j\nu}^0 \in \{z_j^0\}, X_{j\nu}^0 \in \text{int } V; \\ & \lambda_{p_l}(\bar{\Delta}^{(N)}, \tilde{y}_i^0) = \tilde{\lambda}_{p_l}(\bar{\Delta}^{(N)}); i = \overline{1, R_Y}; \tilde{y}_i^0 \in A; \\ & \frac{\partial \lambda_{p_l}(\bar{\Delta}^{(N)}, \tilde{y}_{ik}^0)}{\partial y} = 0, k = \overline{1, R_{y1}}; R_{y1} \leq R_Y, \tilde{y}_{ik}^0 \in \text{int } A, \tilde{y}_{ik}^0 \in \{\tilde{y}_i^0\}. \end{aligned} \quad (55)$$

Здесь

$$\begin{aligned} & R_X + R_Y = N + 1, \text{ если } \min \tilde{\varepsilon}(\Delta^{(N)}) < \varepsilon_0 < \bar{\varepsilon}; \\ & R_X = N + 1, \text{ если } \varepsilon_0 = \min \tilde{\varepsilon}(\Delta^{(N)}); \\ & R_Y = N + 1, \text{ если } \varepsilon_0 \geq \bar{\varepsilon}; \\ & \bar{\varepsilon} = \tilde{\Phi}_\Sigma(\Delta_*^{(N)}); \Delta_*^{(N)} = \arg \min_{\Delta^{(N)} \in G_N} \tilde{\lambda}_{p_l}(\Delta^{(N)}), \end{aligned} \quad (56)$$

точки \tilde{y}_i^0 отличаются от y_j^0 ; значения $\tilde{\Phi}_\Sigma(\Delta^{(N)})$ в (51) могут достигаться в разных точках X_j^0 для разных значений неопределенных факторов y_j^0 и по-прежнему система равенств (55), (56) оказывается замкнутой относительно всех неизвестных, к числу которых теперь относятся $\bar{\Delta}^{(N)}$, $\tilde{\lambda}_{p_l}(\bar{\Delta}^{(N)})$, $X_{j\nu}^0$, \tilde{y}_{ik}^0 и $\min \tilde{\varepsilon}(\Delta^{(N)})$, если $\varepsilon_0 = \min \tilde{\varepsilon}(\Delta^{(N)})$.

В частном случае минимаксной задачи (54) соотношения (55)-(58) упрощаются и сводятся к подобной (31), (32)

системе первых двух равенств в (55), где R_X и R_{X1} определяются аналогично R и R_1 в (31), (32).

Аналогичным образом усложняются частные задачи оптимизации по сравнению с (34). Специфическая проблема выявления характера распределения $Q(z, \bar{\Delta}^{(N)})$ на множестве Z и $I_p(\bar{\Delta}^{(N)}, y)$ на множестве A в (48), (49) в целях редукции системы равенств вида (31), (32) к расчетным системам уравнений часто может быть решена на основе свойств результирующих состояний ОРП для детерминированных моделей объекта и базовых закономерностей предметной области исследуемых процессов [8, 13, 14].

4. Задача многокритериальной оптимизации в условиях воздействия множественных возмущений. Рассмотрим далее ЗМО объектом (1)-(3) с априори фиксируемыми значениями \bar{Q}^0 и w его параметрических характеристик в (39) в условиях наличия возмущений $\eta(t) \in \{\eta_1(t), \eta_2(t)\}$ в (1), (2), рассматриваемых в качестве не полностью определенных факторов $y_1 = (\eta_1, \eta_2)$, вся априорная информация о которых исчерпывается заданными границами диапазона их возможного изменения:

$$\eta(t) \in W_1 = \{\eta(t) : \eta_{\min} \leq \eta(t) \leq \eta_{\max}, t \in [0, t^*]\}. \quad (57)$$

Предполагается, что ограничения (57) согласованы с (3) и сохраняют управляемость объекта по требованиям, аналогичным (42), (43), где следует заменить y на y_1 .

Условия неопределенности (57) порождают ансамбль траекторий движения объекта (1), (2) вида (41):

$$\bar{Q}^\Sigma(t, y_1) = \cup \{\bar{Q}(t, \eta) : \eta \in W_1, t \in [0, t^*]\}, \quad (58)$$

описываемый вместо (4) системой уравнений:

$$\begin{aligned} \frac{d\bar{Q}_n}{dt} &= -\mu_n^2 \bar{Q}_n + k_{Vn} u_V(t) + k_{Sn} u_S(t) + k_{V1n} \eta_1(t) + k_{S1n} \eta_2(t), \\ k_{V1n} &= \int_V g_{V1}(Y) \phi_n(Y) r(Y) dY; \quad k_{S1n} = \int_S g_{S1}(Y) \phi_n(Y) r(Y) dS(Y). \end{aligned} \quad (59)$$

Принцип гарантированного результата формально приводит к подобным (42), (43) частным задачам оптимизации;

аналогичной (44), (45) ЗМО и однокритериальной версии (46), где лишь следует заменить u на η и A на W_1 . Однако эти задачи существенно усложняются по сравнению с (47)-(49) необходимостью поиска функций максимума на множестве допустимых согласно (57) возмущений $\eta_1(t), \eta_2(t)$ с заранее неизвестным характером их изменения во времени. В этой ситуации предлагается рассматривать такие задачи в игровой постановке в качестве дифференциальной игры с антагонистическими управлениями $u(t)$ и $\eta(t)$ объектом (59) [16, 17] с последующим решением по схеме детерминированных ЗМО.

В типичном случае представления подынтегральных функций вида (42) в форме аддитивной зависимости от $u(t)$ и $\eta(t)$ искомые оптимальные управления $u^*(t), \eta^*(t)$ в частных задачах управления и $u^{**}(t), \eta^{**}(t)$ в однокритериальной версии ЗМО могут быть найдены из условий существования седловой точки игры:

$$\begin{aligned} I_p(u^*(t), \eta^*(t)) &= \min_{u \in \tilde{\Omega}_p} \max_{\eta \in W_1} I_p(u(t), \eta(t)) = \\ &= \max_{\eta \in W_1} \min_{u \in \tilde{\Omega}_p} I_p(u(t), \eta(t)); p = \overline{1, q}; \end{aligned} \quad (60)$$

$$I(u^{**}(t), \eta^{**}(t)) = \min_{u \in \tilde{\Omega}_\Sigma} \max_{\eta \in W_1} I(u(t), \eta(t)) = \max_{\eta \in W_1} \min_{u \in \tilde{\Omega}_\Sigma} I(u(t), \eta(t)), \quad (61)$$

где обозначения $\tilde{\Omega}_p$ и $\tilde{\Omega}_\Sigma$ соответствуют принятым в (42)-(46) с заменой u на η .

Специальная форма принципа максимума для минимаксных задач (60), (61) [16] с функцией Понтрягина вида (22), определяемой для модели объекта (59) вместо (4), опять позволяет подобно детерминированному случаю найти зависимости $u^{**}(\overline{Q}^{**}(t), \overline{\Psi}^{**}(t)), \eta^{**}(\overline{Q}^{**}(t), \overline{\Psi}^{**}(t))$ (и аналогично для u^* и η^*), с помощью которых устанавливается структура искомых алгоритмов и их параметрическое представление с точностью до векторов параметров $\Delta_u^{(N_u)}$ и $\Delta_\eta^{(N_\eta)}$. В итоге опять может быть осуществлена процедура редукции к задаче полубесконечной оптимизации вида (24)-(26):

$$I = \lambda^0 \left(\Delta_u^{(N_u)}, \Delta_\eta^{(N_\eta)} \right) \rightarrow \min_{\Delta_u^{(N_u)}, \Delta_\eta^{(N_\eta)}} ;$$

$$\Phi_\Sigma \left(\Delta_u^{(N_u)}, \Delta_\eta^{(N_\eta)} \right) = \max_{X \in V} \left| Q \left(X, \Delta_u^{(N_u)}, \Delta_\eta^{(N_\eta)} \right) - Q_c \right| \leq \varepsilon_0 ; \quad (62)$$

$$\lambda_p \left(\Delta_u^{(N_u)}, \Delta_\eta^{(N_\eta)} \right) = \frac{I_p \left(\Delta_u^{(N_u)}, \Delta_\eta^{(N_\eta)} \right) - I_p^* (u^*, \eta^*)}{I_p^0 - I_p^* (u^*, \eta^*)} \leq \lambda^0, \quad p = \overline{1, q},$$

которая отличается от (24)-(26) большим числом $N_u + N_\eta$ искомым неизвестных.

Дальнейшая редукция ЗПО (62) к аналогичной детерминированной ситуации форме (25), (28) (или (30) в частном случае, подобном (29)), и последующее ее решение может быть выполнено по схеме альтернативного метода на базе альтернативных свойств (31). Аналогичным образом по схеме, подобной (34)-(38), могут быть решены частные задачи оптимизации и найдены величины $I_p^* (u^*, \eta^*)$, I_p^0 в (62).

Исходная задача многокритериальной оптимизации в условиях действия множественных возмущений кардинально упрощается в характерном частном случае ее линейности по управлению $u(t)$. Применительно к этому случаю $u^{**}(t)$ является кусочно-постоянной (релейной) функцией времени:

$$u^{**}(t) = \frac{u_{\max} + u_{\min}}{2} \pm (-1)^{j+1} \frac{u_{\max} - u_{\min}}{2}, \quad (63)$$

$$\sum_{i=0}^{j-1} \Delta_{iu}^{(N)} < t < \sum_{i=0}^j \Delta_{iu}^{(N)}, \quad j = \overline{1, N}, \Delta_{0u}^{(N)} = 0,$$

априори определяемой с точностью до вектора $\Delta_u^{(N)}$ моментов ее переключения $\Delta_{iu}^{(N)}, i = \overline{1, N}$, и до знака u^{**} на первом интервале постоянства. В такой ситуации антагонистический алгоритм $\eta^{**}(t)$

заведомо формируется в подобном (63) виде с обратным знаком перед множителем $(-1)^{j+1}$:

$$\eta^{**}(t) = \frac{\eta_{\max} + \eta_{\min}}{2} \mp (-1)^{j+1} \frac{\eta_{\max} - \eta_{\min}}{2},$$

$$\sum_{i=0}^{j-1} \Delta_{iu}^{(N)} < t < \sum_{i=0}^j \Delta_{iu}^{(N)}, j = \overline{1, N}, \Delta_{0u}^{(N)} = 0. \quad (64)$$

Здесь в (63), (64) $N = N_u = N_\eta$.

5. Многокритериальная задача управления объектом технологической теплофизики. В качестве примера, представляющего самостоятельный интерес, рассмотрим многокритериальную задачу оптимизации процесса индукционного нагрева металлических полуфабрикатов перед обработкой давлением [6-9, 14, 18-22].

Пусть температурное поле $Q(x, t)$ тела цилиндрической формы в процессе индукционного нагрева описывается в зависимости от радиальной координаты x и времени t неоднородным одномерным уравнением теплопроводности в относительных единицах вида (1)-(2) [14]:

$$\frac{\partial Q(x, t)}{\partial t} = \frac{\partial^2 Q(x, t)}{\partial x^2} + \frac{1}{x} \frac{\partial Q(x, t)}{\partial x} + g_V(x) u_V(t), 0 < t < t^*, 0 < x < 1,$$

$$Q(x, 0) = Q^0 = const, 0 \leq x \leq 1; \frac{\partial Q(0, t)}{\partial x} = 0; \frac{\partial Q(1, t)}{\partial x} + \beta Q(1, t) = \beta Q^0.$$

Здесь заданный коэффициент $\beta = const > 0$ определяет уровень тепловых потерь в окружающую среду с температурой Q^0 на внешней поверхности цилиндра $x = 1$; $g_V(x)$ — известная функция радиального распределения электромагнитных источников тепла [14]; $u_V(t)$ — суммарная удельная мощность внутреннего тепловыделения, рассматриваемая в качестве управляющего воздействия, стесненного ограничением (3):

$$0 \leq u_V(t) \leq 1, 0 < t < t^*,$$

и в условиях интервальной неопределенности (39) вся информация о значениях Q^0 и β исчерпывается сведениями о диапазонах их максимально возможного изменения:

$$Q_{\min}^0 \leq Q^0 \leq Q_{\max}^0; \beta_{\min} \leq \beta \leq \beta_{\max}.$$

Пусть качество процесса индукционного нагрева с априори фиксируемой длительностью t^* описывается тремя рассматриваемыми в качестве критериев оптимальности типичными технико-экономическими показателями вида (7)-(9), характеризующими расход энергии на нагрев ($I_1(u_V)$), потери металла в окалину ($I_2(u_V)$) и точность приближения к заданному конечному температурному состоянию ($I_3(u_V)$), оцениваемую в равномерной метрике [8, 14]:

$$I_1(u_V) = \int_0^{t^*} u_V dt \rightarrow \min_{u_V \in \Omega_1};$$

$$I_2(u_V) = \int_0^{t^*} f_{02}(Q(1,t)) dt \rightarrow \min_{u_V \in \Omega_2};$$

$$f_{02}(Q(1,t)) = \begin{cases} 0, Q(1,t) \leq Q_h, \\ (Q(1,t) - Q_h)^{r+1}, Q(1,t) \geq Q_h; \end{cases}$$

$$I_3(u_V) = \int_0^{t^*} \frac{\tilde{\varepsilon}}{t^*} dt = \tilde{\varepsilon} \rightarrow \min_{u_V \in \Omega_3}; \frac{d\tilde{\varepsilon}}{dt} = 0.$$

Здесь Q_h — известная температурная граница процесса окисления; $Q(1,t)$ — температура поверхности нагреваемой заготовки; r — заданное число; требуемая конечная температура $Q_c > Q_h$ и I_3 представляется в известной форме интегрального функционала с дополнительными фазовыми ограничениями, где $\tilde{\Omega}_3$ совпадает с $\tilde{\Omega}_2$ в (44), (45).

Соответствующая многокритериальная задача управления с учетом интервальной неопределенности значений Q^0 и β приводится к виду (44)-(45) при

$$y = (Q^0, \beta); A = \{y : Q_{\min}^0 \leq Q^0 \leq Q_{\max}^0; \beta_{\min} \leq \beta \leq \beta_{\max}\},$$

где температурное поле $Q(x, y, t^*), x \in V = [0, 1]$, представляется разложением (5) в ряд по собственным функциям Бесселя нулевого порядка $\phi_n(x) = J_0(\mu_n x)$ [14]:

$$Q(x, y, t^*) = \sum_{n=1}^{\infty} \frac{2\mu_n^2 J_0(\mu_n x)}{(\mu_n^2 + \beta^2) J_0^2(\mu_n)} \bar{Q}_n(t^*),$$

и $\bar{Q}_n(t^*)$ определяется решениями уравнений (4) при $u_S(t) = 0$ в зависимости от значения Q^0 .

Последующий переход к однокритериальной версии (46) и применение по схеме (21)-(23) стандартной процедуры принципа максимума Понтрягина для любой допустимой пары значений (Q^0, β) приводит к искомой зависимости $u^{**}(\bar{Q}^{**}, \bar{\Psi}^{**})$ для оптимального управления в форме релейной функции времени:

$$u_V^{**}(t) = \frac{1}{2} \left[1 + \text{sign} \left(\tilde{\Psi}_1^{**} + \sum_{n=1}^{\infty} k_{Vn} \Psi_n^{**}(t) \right) \right],$$

на всех интервалах изменения оптимальной программы, где не выполняется тождественно равенство:

$$\tilde{\Psi}_1^{**} + \sum_{n=1}^{\infty} k_{Vn} \Psi_n^{**}(t) \equiv 0,$$

и в форме особого управления $u_0(t)$ на тех участках в ее составе, где последнее соотношение имеет место.

Можно показать, что искомая оптимальная программа действительно содержит особые участки, на которых минимизируются потери металла в окалину, существенно усложняя проблему однозначного выбора структуры $u_V^{**}(t)$ из множества допустимых вариантов [14].

Ограничимся далее типичной наиболее просто реализуемой ситуацией с программой оптимального управления [14]:

$$u_V^{**}(t) = \begin{cases} 1, & 0 \leq t \leq t_1; \\ u_0(t), & t_1 < t < t_2; \\ 0, & t_2 < t \leq t^*, \end{cases} \quad (65)$$

которая при заданной величине $t^* > t_2$ содержит единственный участок длительностью $t_2 - t_1$ с особым управлением (рисунок 1а),

определяемым в первом приближении в зависимости от t_1 известной функцией $u_0(t, t_1)$ экспоненциального вида [14].

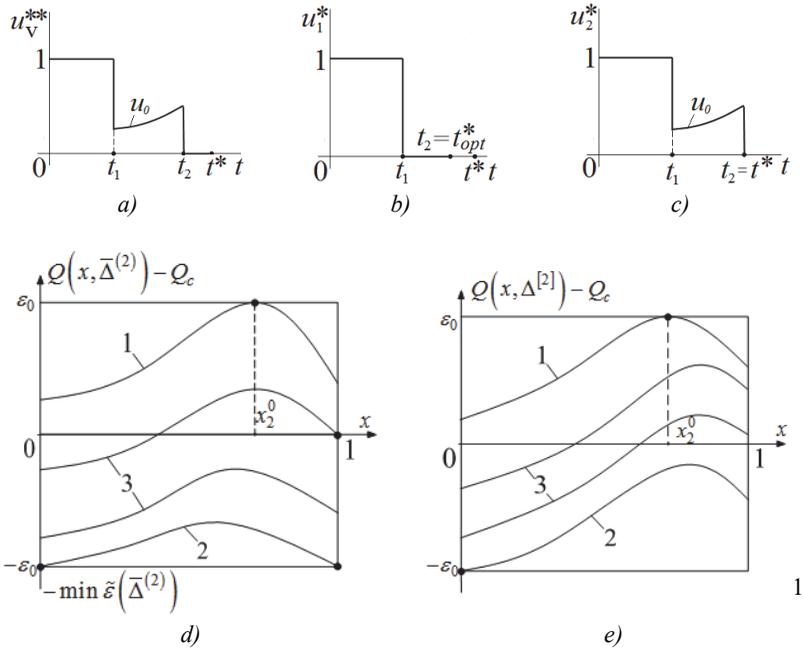


Рис. 1. Базовые характеристики оптимального процесса: а-с) структура управляющих воздействий; d, e) формы кривых пространственного распределения конечных температурных состояний в процессе многокритериальной оптимизации

В итоге при фиксированной величине t^* оптимальное управление $u^{**}(t)$ однозначно характеризуется его параметрическим представлением с точностью до вектора $\Delta^{(N)} = \Delta^{(2)} = (\Delta_1^{(2)}, \Delta_2^{(2)})$ параметров $\Delta_1^{(2)} = t_1$; $\Delta_2^{(2)} = t_2$ — начального и конечного моментов особого участка. Последующее интегрирование модели объекта с параметризованным управлением $u_V(\Delta^{(2)}, t)$ обеспечивает при $N=2$ редукцию к ЗПО вида (47)-(49) с явными выражениями для $Q(z, \Delta^{(2)})$, $\tilde{I}_p(\Delta^{(2)}, y)$, $\tilde{\lambda}_p(\Delta^{(2)})$, \tilde{I}_p^* , \tilde{I}_p^0 , где $z = (x, Q^0, \beta)$; $Z = [0, 1] \times A$. Применительно к наиболее характерной для

приложений ситуации (53), ЗПО (47)-(49) сводится к минимаксной задаче вида (54):

$$\tilde{\Phi}_{\Sigma}(\Delta^{(2)}) = \max_{z=(Q^0, \beta) \in A} \left[\max_{x \in [0,1]} |Q(X, Q^0, \beta, \Delta^{(2)}) - Q_c| \right] \rightarrow \min_{\Delta^{(2)}}. \quad (66)$$

Закономерности нестационарных процессов теплопроводности позволяют установить показанную на рисунке 1d форму кривой $Q(X, Q^0, \beta, \bar{\Delta}^{(2)})$ пространственного распределения конечного температурного состояния на решении $\bar{\Delta}^{(2)}$ задачи (66) [13, 14]. Согласно рисунку 1d система первых двух альтернансных равенств в (55) при $R_X = N + 1 = 3$ выполняется в трех точках $z_1^0 = (x_1^0 = 0, Q_1^0 = Q_{\min}^0, \beta_1 = \beta_{\max})$, $z_2^0 = (x_2^0 \in (0,1), Q_2^0 = Q_{\max}^0, \beta_2 = \beta_{\min})$, $z_3^0 = (x_3^0 = 1, Q_3^0 = Q_{\min}^0, \beta_3 = \beta_{\max})$ и редуцируется к расчетной системе четырех уравнений альтернансного метода:

$$\begin{aligned} & (1- Q^0 = Q_{\max}^0; \beta = \beta_{\min}; 2- Q^0 = Q_{\min}^0; \beta = \beta_{\max}; \\ & 3- Q_{\min}^0 < Q^0 < Q_{\max}^0; \beta_{\min} < \beta < \beta_{\max}). \end{aligned}$$

$$\begin{aligned} & Q(0, Q_{\min}^0, \beta_{\max}, \bar{\Delta}^{(2)}) - Q_c = -\min \tilde{\varepsilon}(\Delta^{(2)}); \\ & Q(x_2^0, Q_{\max}^0, \beta_{\min}, \bar{\Delta}^{(2)}) - Q_c = \min \tilde{\varepsilon}(\Delta^{(2)}); \\ & Q(1, Q_{\min}^0, \beta_{\max}, \bar{\Delta}^{(2)}) - Q_c = -\min \tilde{\varepsilon}(\Delta^{(2)}); \\ & \frac{\partial Q(x_2^0, Q_{\max}^0, \beta_{\min}, \bar{\Delta}^{(2)})}{\partial x} = 0. \end{aligned} \quad (67)$$

Эта система разрешается известными численными методами относительно всех четырех неизвестных параметров оптимального процесса $\bar{\Delta}_1^{(2)} = t_1$, $\bar{\Delta}_2^{(2)} = t_2$, $\min \tilde{\varepsilon}(\Delta^{(2)})$ и x_2^0 . Последующее вычисление по найденному указанным способом вектору $\bar{\Delta}^{(2)}$ значений $\tilde{\lambda}_p(\bar{\Delta}^{(2)})$ в (49) при известных \tilde{I}_p^* , \tilde{I}_p^0 и подобно (33) величины $\tilde{\lambda}_{\min}^0 = \max_{p=1,2,3} \tilde{\lambda}_p(\bar{\Delta}^{(2)})$ обеспечивает выполнение неравенств (49), завершая решение исходной ЗПО (47)-(49).

Для определения \tilde{I}_p^* необходимо предварительно решить частные задачи оптимизации вида (34):

$$\tilde{I}_p([\Delta^{(2)}]_p) \rightarrow \max_{Q^0, \beta} I_p([\Delta^{(2)}]_p, Q^0, \beta) \rightarrow \min_{[\Delta^{(2)}]_p}; p = 1, 2, 3;$$

$$\tilde{\Phi}_p([\Delta^{(2)}]_p) = \max_{Q^0, \beta} \left[\max_{x \in [0,1]} |Q_p(x, Q^0, \beta, [\Delta^{(2)}]_p) - Q_c| \right] \leq \varepsilon_{0p},$$

на элементах $[\Delta^{(2)}]_p$ подмножества $\{[\Delta^{(2)}]_p\} \subset \{\Delta^{(2)}\} \in G_N$ значений $\Delta^{(2)} = (t_1, t_2)$. Здесь \tilde{I}_p и $\tilde{\Phi}_p$ определяются согласно (42), (43); в рассматриваемых условиях (53) следует аналогичным образом принять:

$$\varepsilon_{0p} = \min \tilde{\varepsilon}([\Delta^{(2)}]_p) = \min_{[\Delta^{(2)}]_p} \left[\max_{x, Q^0, \beta} |Q_p(x, Q^0, \beta, [\Delta^{(2)}]_p) - Q_c| \right],$$

и в таком случае частные задачи оптимизации опять сводятся подобно (66) к минимизации $\tilde{\Phi}_p([\Delta^{(2)}]_p)$:

$$\tilde{\Phi}_p([\Delta^{(2)}]_p) = \max_{Q^0, \beta} \left[\max_{x \in [0,1]} |Q_p(x, Q^0, \beta, [\Delta^{(2)}]_p) - Q_c| \right] \rightarrow \min_{[\Delta^{(2)}]_p}. \quad (68)$$

Применительно к критерию I_1 минимальный расход энергии \tilde{I}_1^* достигается при нефиксируемом заранее моменте t_{opt}^* окончания процесса управления с релейной формой не содержащего особого участка алгоритма (65) (рисунок 1b), параметрически представляемого в этих условиях на подмножестве $\{[\Delta^{(2)}]_{p=1}\}$ векторов $\Delta^{(2)} = (t_1, t_2 < t^*)$, где $\{[\Delta^{(2)}]_{p=1}\} = \{\Delta^{(2)} = (t_1, t_2) : t_2 = t_{opt}^*, u_0 = 0\}$.

Здесь t_{opt}^* совпадает с минимальным временем t_{min}^* достижения требуемой величины $\varepsilon_{01} = \min \tilde{\varepsilon}([\Delta^{(2)}]_{p=1})$, и тогда всегда выполняется условие $t_{opt}^* < t^*$ для фиксируемого значения t^* [14].

В таком случае задержка начала процесса управления на время $t^* - t_{opt}^*$ обеспечивает поиск \tilde{I}_1^* на подмножестве $\{[\Delta^{(2)}]_{p=1}\}$. В итоге задача сводится к виду (68) при $p = 1$.

Ее решение $\Delta^{[1]} = (t_{1opt}, t_{opt}^*)$ обладает подобными (67) альтернансными свойствами, которые при сохранении характера зависимости $Q_p(x, Q^0, \beta, \Delta^{[1]})$ от x, Q^0, β , показанного на рисунке 1,d [14], приводят к аналогичной (67) системе уравнений. Здесь требуется только заменить $\bar{\Delta}^{(2)}$ решением $\Delta^{[1]}$, $\min \tilde{\varepsilon}(\Delta^{(2)})$ — значением $\min \tilde{\varepsilon}([\Delta^{(2)}]_{p=1})$ и представить $Q_1(x, Q^0, \beta, \Delta^{[1]})$ в виде, отвечающем алгоритму управления $u_1^*(t)$ релейной формы.

Последующее вычисление $\Delta^{[1]}$ позволяет найти $\tilde{I}_1^* = t_{1opt}$.

Минимизация I_2 обеспечивается в условиях максимальной длительности особого участка при $t_2 = t^*$ в (65) (рисунок 1c) [14], и, следовательно, в данном случае будем иметь:

$$\{[\Delta^{(2)}]_{p=2}\} = \{\Delta^{(2)} = (t_1, t_2) : t_2 = t^*\},$$

при фиксируемой заведомо величине t^* . В итоге, получаем однопараметрическое представление $[\Delta^{(2)}]_{p=2} = t_1$.

Таким образом, вторая частная задача оптимизации сводится к виду (68) при $p = 2$ и альтернансные свойства ее решения $\Delta^{[2]} = t_{1opt}$ согласно (55), (56) выполняются при условиях $N=1$, $R_X = 2$. Можно показать аналогично [13, 14], что конфигурация распределения $Q_2(x, Q^0, \beta, \Delta^{[2]})$ по x в зависимости от значений Q^0 и β принимает вид, показанный на рисунке 1e с двумя точками альтернанса $z_1^0 = (x_1^0 = 0, Q_1^0 = Q_{min}^0, \beta_1 = \beta_{max})$; $z_2^0 = (x_2^0 \in (0, 1), Q_2^0 = Q_{max}^0, \beta_2 = \beta_{min})$. Соответствующая система теперь уже трех уравнений для определения $\Delta^{[2]} = t_{1opt}$, $\min \tilde{\varepsilon}([\Delta^{(2)}]_{p=2})$, и x_2^0 выглядит следующим образом:

$$\begin{aligned} Q_2(0, Q_{min}^0, \beta_{max}, \Delta^{[2]}) - Q_c &= -\min \tilde{\varepsilon}([\Delta^{(2)}]_{p=2}); \\ Q_2(x_2^0, Q_{max}^0, \beta_{min}, \Delta^{[2]}) - Q_c &= \min \tilde{\varepsilon}([\Delta^{(2)}]_{p=2}); \\ \frac{\partial Q_2(x_2^0, Q_{max}^0, \beta_{min}, \Delta^{[2]})}{\partial x} &= 0. \end{aligned}$$

По найденным корням этой системы вычисляется $\tilde{I}_2^* = I_2(t_{1opt}, t^*)$ для заданной величины t^* .

Применительно к критерию I_3 подмножество $\{[\Delta^{(2)}]_{p=3}\}$ совпадает со всем допустимым множеством значений $\Delta^{(2)} \in G_2$, и задача (68) не отличается от (66), то есть $\Delta^{[3]} = \bar{\Delta}^{(2)}$ и $\tilde{I}_3^* = \min \tilde{\varepsilon}(\Delta^{(2)})$.

Значения $\tilde{I}_p^0, p = 1, 2, 3$ в (49) могут быть определены по алгоритму, подобному (38).

На рисунке 2 приведены некоторые результаты решения рассматриваемой задачи при $Q_{\min}^0 = -0,15$; $Q_{\max}^0 = 0$; $\beta_{\min} = 0,2$; $\beta_{\max} = 0,5$; $Q_h = 0,375$; $Q_c = 0,5$; $r = 3,5$; $t^* = 0,43$.

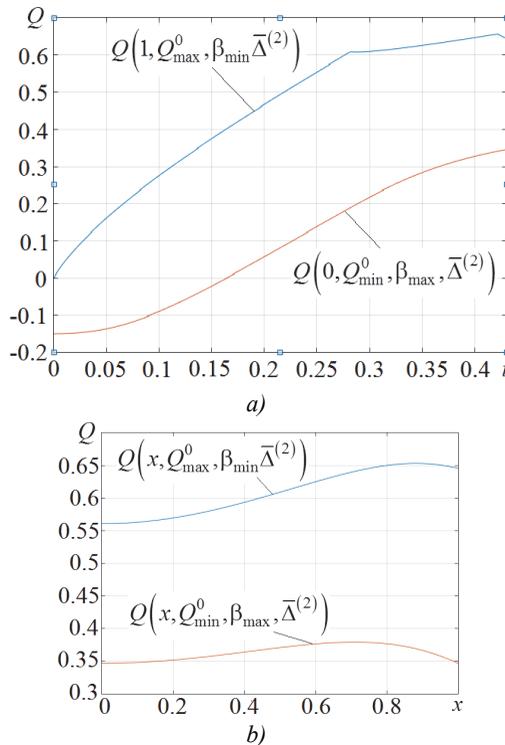


Рис. 2. Температурное поле (а) и результирующие радиальные распределения температур (б) в оптимальном процессе нагрева

Применительно к этому случаю находятся значения $\lambda_p(\bar{\Delta}^{(2)})$ и λ_{\min}^0 , определяемые для вычисляемых по указанной схеме значений

$$I_p^*, I_p^0 \quad \text{при} \quad p = 1, 2, 3: \quad \tilde{\lambda}_1(\bar{\Delta}^{(2)}) = 0.081845; \quad \tilde{\lambda}_2(\bar{\Delta}^{(2)}) = 0.09654;$$

$$\tilde{\lambda}_3(\bar{\Delta}^{(2)}) = 0; \quad \tilde{\lambda}_{\min}^0 = \tilde{\lambda}^0(\bar{\Delta}^{(2)}) = \max_{p=1,2,3} \tilde{\lambda}_p(\bar{\Delta}^{(2)}) = \tilde{\lambda}_2(\bar{\Delta}^{(2)}) = 0.09654.$$

6. Заключение. Предлагаемые способы поиска парето-эффективных решений достаточно широкого круга задач многокритериальной оптимизации систем с распределенными параметрами параболического типа сводятся к последовательно осуществляемым процедурам минимаксной свертки нормализуемых критериев качества; перехода к эквивалентной версии в форме типовой вариационной задачи с ограничениями; точной (в рамках исследуемых моделей) редукции к задаче полубесконечного программирования после параметризации искомым управляющих воздействий и ее алгоритмически точному решению альтернативным методом, который распространяет на рассматриваемые задачи известные результаты теории нелинейных чебышёвских приближений с существенным использованием закономерностей предметной области. Подобная схема реализована в наиболее характерных для приложений условиях оценки в равномерной метрике требований к конечному состоянию распределенной системы.

Разработанная применительно к детерминированным моделям объекта технология может быть распространена на задачи управления по принципу гарантированного результата системами с распределенными параметрами с интервальными неопределенностями параметрических характеристик объекта и (или) внешних возмущающих воздействий. По существу, такие задачи сводятся к детерминированному варианту для наиболее неблагоприятной из возможных реализации неопределенных факторов.

Приведенный пример иллюстрирует конструктивный характер предлагаемой технологии в условиях наличия дополнительной информации о базовых свойствах исследуемых процессов в каждой отдельно взятой задаче многокритериального управления.

Литература

1. *Гермейер Ю.Б.* Введение в теорию исследования операций // М.: Наука. 1971. 383 с.
2. *Мацунин К.Ю., Мацунин Ю.К.* Моделирование технических систем в условиях неопределенности и принятие оптимального решения // Изв. РАН. Теория и системы управления. 2013. № 4. С. 19–35.
3. *Мацунин К.Ю., Мацунин Ю.К.* Векторная оптимизация с равнозначными и приоритетными критериями // Изв. РАН. Теория и системы управления. 2017. № 6. С. 80–99.

4. *Новикова Н.М., Поспелова И.И., Зенюков А.И.* Метод сверток в многокритериальных задачах с неопределенностью // Изв. РАН. Теория и системы управления. 2017. № 5. С. 27–45.
5. *Konak A., Coit D.W., Smith A.E.* Multi-objective optimization using genetic algorithms: a tutorial // Reliability Engineering and Systems Safety. 2006. vol. 9. pp. 992–1007.
6. *Di Barba P.* Basic principles of optimal design of electromagnetic devices and multi-objective optimization // ASM Handbook, Induction Heating and Heat Treatment. 2014. vol. 4C. pp. 359–365.
7. *Di Barba P. et al.* Induction heating of a graphite disk: a benchmark for multi-physics platform for multi-objective design // Proceedings of XVIII International UIE-Congress «Electrotechnologies for Material Processing». 2017. pp. 520–525.
8. *Pleshivtseva Yu. et al.* Design concepts of induction mass heating technology based on multiple-criteria optimization // Intern. Journal for Computation and Mathematics in Electrical and Electronic Engineering. 2017. vol. 36. no. 2. pp. 386–400.
9. *Sieni E., Forzan M., Di Barba P.* Migration NSGA: Method to Improve a Non-Elitist Searching of Pareto front with Application in Magnetics // Inverse Problems in Science and Engineering. 2016. vol. 24. no. 1. pp. 543–566.
10. *Klonk S., Bay F.* Numerical Analysis of Computational Models for Induction Heat Treatment of Complex Geometrical Parts // International Journal of Microstructure and materials Properties. 2016. vol. 11. no. 1-2. pp. 48–70.
11. *Дилигенский Н.В., Дымова Л.Г., Севастьянов П.В.* Нечеткое моделирование и многокритериальная оптимизация производственных систем в условиях неопределенности: технология, экономика, экология // М.: Машиностроение-1. 2004. 336 с.
12. *Волин Ю.М., Островский Г.М.* Многокритериальная оптимизация технологических процессов в условиях неопределенности // Автоматика и телемеханика. 2007. № 3. С. 165–180.
13. *Рапопорт Э.Я.* Альтернативный метод в прикладных задачах оптимизации // М.: Наука. 2000. 336 с.
14. *Рапопорт Э.Я., Плишевицева Ю.Э.* Оптимальное управление температурными режимами индукционного нагрева // М.: Наука. 2012. 309 с.
15. *Рапопорт Э.Я.* Структурное моделирование объектов и систем с распределенными параметрами // М.: Высшая школа. 2003. 299 с.
16. *Пантелеев А.В., Бортакоский А.С.* Теория управления в примерах и задачах // М.: Высшая школа. 2003. 583 с.
17. *Красовский Н.Н., Субботин А.И.* Позиционные дифференциальные игры // М.: Наука. 1974. 456 с.
18. *Di Barba P., Forzan M., Sieni E.* Multiobjective design optimization of an induction heating device: A benchmark problem // International Journal of Applied Electromagnetics and Mechanics. 2015. vol. 47. no. 4. pp. 1003–1013.
19. *Di Barba P., Dughiero F., Forzan M., Sieni E.* Improved solution to a multi-objective benchmark problem of inverse induction heating // International Journal of Applied Electromagnetics and Mechanics. 2015. vol. 49. no. 2. pp. 279–288.
20. *Bay F., Alves J., Barlier J.* Modelling and optimisation of electromagnetically coupled solid manufacturing processes // Proceedings of XVIII International UIE-Congress «Electrotechnologies for Material Processing». 2017. pp. 436–441.
21. *Rudnev V., Loveless D., Cook R.* Handbook of Induction Heating: 2nd edition // CRC Press. 2017. 780 p.
22. *Brown D., Rudnev V., Dickson P.* Induction Heating of Billets, Rods and Bars // ASM Handbook. 2014. vol. 4. pp. 330–345.

Рапопорт Эдгар Яковлевич — д-р техн. наук, профессор, профессор кафедры автоматизации и управления в технических системах института автоматизации и информационных технологий, Самарский государственный технический университет (СамГТУ). Область научных интересов: теория управления системами с распределенными параметрами, оптимизация объектов технологической теплофизики. Число научных публикаций — 400. edgar.rapoport@mail.ru; ул. Молодогвардейская, 244, Самара, 443100; р.т.: +7(846)337-0700.

Плешивцева Юлия Эдгаровна — д-р техн. наук, профессор, профессор кафедры управления и системного анализа теплоэнергетических и социотехнических комплексов теплоэнергетического факультета, Самарский государственный технический университет (СамГТУ). Область научных интересов: теория управления системами с распределенными параметрами, оптимизация объектов технологической теплофизики. Число научных публикаций — 165. yulia_pl@mail.ru; ул. Молодогвардейская, 244, Самара, 443100; р.т.: +7(846)332-4234.

Поддержка исследований. Работа выполнена при финансовой поддержке РФФИ (проекты №18-08-00048 и №16-08-00945).

E.YA. RAPOPORT, YU.E. PLESHIVTSEVA
**METHOD OF MULTIOBJECTIVE OPTIMIZATION OF
CONTROLLED SYSTEMS WITH DISTRIBUTED PARAMETERS**

Rapoport E.Ya., Pleshivtseva Yu.E. Method of Multiobjective Optimization of Controlled Systems with Distributed Parameters.

Abstracts. The constructive method for multicriteria optimization of control processes of deterministic and not fully defined controlled systems with distributed parameters, described by linear multidimensional parabolic partial differential equations with internal and boundary control actions, is proposed. The optimization problem is considered when an uniform approximation accuracy of object's final state to the required spatial distribution of controlled function is given. The suggested approach is based on the one-criterion option in the form of minimax convolution of the normalized quality criteria and the subsequent transition to the equivalent form of a typical variational problem with constraints. It is applied to the deterministic model of an object described by an infinite system of differential equations with respect to time-dependent modes of the controlled quantity expansion in a series of eigenfunctions of the initial-boundary value problem. Further procedures for the preliminary parametrization of control actions, based on analytical optimum conditions and reduction to semi-infinite programming problems, allow one to find the desired extremals using their Chebyshev properties and fundamental laws of the domain in typical application conditions of estimating the accuracy of approaching the object's final state to the required one in a uniform metric. The obtained results are extended to the tasks of program control on the principle of guaranteed result by ensembles of object trajectories under conditions of interval uncertainty of the parametric characteristics of the distributed system and multiple external disturbances. A demonstrated example of a multicriteria optimization of an innovation technology of metal induction heating prior a hot forming is of special interest. The typical optimization criteria such as energy consumption, metal loss due to scale formation, and heating accuracy are considered as components of vector optimization criterion.

Key words: multiobjective optimization, system with distributed parameters, semi-infinite optimization, controlled systems, uniform approximation, alternance method, interval of uncertainty, principle of guaranteed result.

Rapoport Edgar Yakovlevich — Ph.D., Dr. Sci., professor, professor of automatics and control in technical systems department of Institute of automatics and information technologies, Samara State Technical University. Research interests: theory of control of systems with distributed parameters, optimization of objects of technological thermal physics. The number of publications — 400. edgar.rapoport@mail.ru; 244, Molodogvardeyskaya Str., Samara, 443100, Russia; office phone: +7(846)337-0700.

Pleshivtseva Yuliya Edgarovna — Ph.D., Dr. Sci., professor, professor of control and system analysis department of Faculty of heat-and-power engineering, Samara State Technical University. Research interests: theory of control of systems with distributed parameters, optimization of objects of technological thermal physics. The number of publications — 165. yulia_pl@mail.ru; 244, Molodogvardeyskaya Str., Samara, 443100, Russia; office phone: +7(846)332-4234.

Acknowledgements. This research is supported by RFBR (grants no. 18-08-00048 and no. 16-08-00945).

References

1. Germejer Yu.B. *Vvedenie v teoriju issledovanija operacij* [Introduction in theory research operations]. M.: Nauka. 1971. 383 p. (In Russ.).
2. Mashunin K.Yu., Mashunin Yu.K. [Simulating engineering systems under uncertainty and optimal decision making]. *Izv. RAN. Teorija i sistemy upravleniya – Journal of Computer and Systems Sciences International*. 2013. vol. 4. pp. 19–35. (In Russ.).
3. Mashunin K.Yu., Mashunin Yu.K. [Vector optimization with equal and priority criteria]. *Izv. RAN. Teorija i sistemy upravleniya – Journal of Computer and Systems Sciences International*. 2017. vol. 6. pp. 80–99. (In Russ.).
4. Novikova N.M., Pospelova I.I., Zenyukov A.I. [Method of convolution in multicriteria problems with uncertainty]. *Izv. RAN. Teorija i sistemy upravleniya – Journal of Computer and Systems Sciences International*. 2017. vol. 5. pp. 27–45. (In Russ.).
5. Konak A., Coit D.W., Smith A.E. Multi-objective optimization using genetic algorithms: a tutorial. *Reliability Engineering and Systems Safety*. 2006. vol. 9. pp. 992–1007.
6. Di Barba P. Basic principles of optimal design of electromagnetic devices and multi-objective optimization. ASM Handbook, Induction Heating and Heat Treatment. 2014. vol. 4C. pp. 359–365.
7. Di Barba P. et al. Induction heating of a graphite disk: a benchmark for multi-physics platform for multi-objective design. Proceedings of XVIII International UIE-Congress “Electrotechnologies for Material Processing”. 2017. pp. 520–525.
8. Pleshivtseva Yu. et al. Design concepts of induction mass heating technology based on multiple-criteria optimization. *Intern. Journal for Computation and Mathematics in Electrical and Electronic Engineering*. 2017. vol. 36. no. 2. pp. 386–400.
9. Sieni E., Forzan M., Di Barba P. Migration NSGA: Method to Improve a Non-Elitist Searching of Pareto front with Application in Magnetics. *Inverse Problems in Science and Engineering*. 2016. vol. 24. no. 1. pp. 543–566.
10. Klonk S., Bay F. Numerical Analysis of Computational Models for Induction Heat Treatment of Complex Geometrical Parts. *International Journal of Microstructure and materials Properties*. 2016. vol. 11. no. 1–2. pp. 48–70.
11. Diligenskij N.V., Dymova L.G., Sevast'janov P.V. *Nechetkoe modelirovanie i mnogokriterial'naja optimizacija proizvodstvennyh sistem v uslovijah neopredelennosti: tehnologija, jekonomika, jekologija*. [Fuzzy modeling and multi-criteria optimization of production systems under uncertainty: technology, economy, ecology]. M.: Mashinostroenie-1. 2004. 336 p. (In Russ.).
12. Volin Yu.M., Ostrovskij G.M. [Multicriteria optimization of technological processes under uncertainty conditions]. *Avtomatika i telemekhanika – Automation and Remote Control*. 2007. vol. 3. pp. 165–180. (In Russ.).
13. Rapoport E.Ya. *Al'ternansnyj metod v prikladnyh zadachah optimizacii* [Alternative Method for Solving Applied Optimal Control Problems]. M.: Nauka. 2000. 336 p. (In Russ.).
14. Rapoport E.Ya., Pleshivtseva Yu.E. *Optimal'noe upravlenie temperaturnymi rezhimami indukcionnogo nagreva* [Optimal Control of Induction Heating Processes]. M.: Nauka. 2012. 309 p. (In Russ.).
15. Rapoport E.Ya. *Strukturnoe modelirovanie objektov i system upravlenija s raspredelennymi parametrami* [Structural parametric synthesis of automatic control systems with distributed parameters]. M.: Vysshaya Shkola. 2003. 299 p. (In Russ.).
16. Panteleev A.V., Bortakovskij A.S. *Teorija upravlenija v primerah i zadachah* [Control Theory in Examples and Problems]. M.: Vysshaja shkola. 2003. 583 p. (In Russ.).

17. Krasovskij N.N., Subbotin A.I. *Pozicionnye differencial'nye igry* [Positional Differential Games]. M.: Nauka. 1974. 456 p. (In Russ.).
18. Di Barba P., Forzan M., Sieni E. Multiobjective design optimization of an induction heating device: A benchmark problem. *International Journal of Applied Electromagnetics and Mechanics*. 2015. vol. 47. no. 4. pp. 1003–1013.
19. Di Barba P., Dughiero F., Forzan M., Sieni E. Improved solution to a multi-objective benchmark problem of inverse induction heating. *International Journal of Applied Electromagnetics and Mechanics*. 2015. vol. 49. no. 2. pp. 279–288.
20. Bay F., Alves J., Barlier J. Modelling and optimisation of electromagnetically coupled solid manufacturing processes. Proceedings of XVIII International UIE-Congress “Electrotechnologies for Material Processing”. 2017. pp. 436–441.
21. Rudnev V., Loveless D., Cook R. Handbook of Induction Heating: 2nd edition. CRC Press. 2017. 780 p.
22. Brown D., Rudnev V., Dickson P. Induction Heating of Billets, Rods and Bars. ASM Handbook. 2014. vol. 4. pp. 330–345.

Д.А. ИВАНОВ, М.А. ИВАНОВА, Б.В. СОКОЛОВ
**АНАЛИЗ ТЕНДЕНЦИЙ ИЗМЕНЕНИЯ ПРИНЦИПОВ
УПРАВЛЕНИЯ ПРЕДПРИЯТИЯМИ В УСЛОВИЯХ РАЗВИТИЯ
ТЕХНОЛОГИЙ ИНДУСТРИИ 4.0**

Иванов Д.А., Иванова М.А., Соколов Б.В. Анализ тенденций изменения принципов управления предприятиями в условиях развития технологий Индустрии 4.0.

Аннотация. Объектами данного исследования являются существующие и перспективные информационные технологии, формирующие основу концепций «Индустрия 4.0» и киберфизических систем (КФС), и соответствующие предприятия, на которых они внедряются. В статье в качестве примеров таких предприятий рассматриваются транспортно-логистические и промышленные предприятия. Цель исследований состоит в анализе влияния интеллектуальных информационных технологий (ИИТ), разрабатываемых в рамках проектов «Индустрия 4.0» и соответствующих киберфизических систем, на изменения принципов и способов управления указанными предприятиями в настоящее время и в ближайшем будущем. В статье на конкретных практических примерах проводится иллюстрация происходящих изменений, вызванных внедрением ИИТ. Анализ литературы и практических примеров Индустрии 4.0 показал, что наряду с развитием информационных и инженерных технологий Индустрии 4.0 в экономической части управления транспортно-логистическими и промышленными предприятиями отсутствует четкое понимание и описание бизнес-моделей, в рамках которых эти технологии могут применяться. Проанализированные примеры из практики позволяют сделать вывод, что технологии Индустрии 4.0 и киберфизических систем используются контекстно в виде лишь ограниченного набора той или иной функциональности. Основная же идея Индустрии 4.0 и КФС — создание самоорганизующихся и самоадаптирующихся динамических сетевых структур поставок на протяжении всего жизненного цикла изделий для реализации максимально гибкого индивидуального производства с затратами массового поточного производства — на данный момент реализована лишь в ряде исследовательских проектов. В статье предложена оригинальная интерпретация сервис-ориентированного подхода к описанию и решению задач оценивания и выбора эффективных стратегий использования ИИТ на существующих и перспективных предприятиях. Данная интерпретация базируется на разрабатываемой авторами статьи теории управления структурной динамикой сложных объектов.

Ключевые слова: киберфизические системы и «Индустрия 4.0», интеллектуальные информационные технологии, цепи поставок, промышленные предприятия, эффективность использования информационных технологий

1. Введение. В настоящее время мы являемся свидетелями наступления нового этапа научно-технической революции, характеризующейся интенсивным и повсеместным созданием и внедрением интернет-сервисов и новых информационных технологий. В условиях, когда ресурсы экстенсивного роста экономики за счет наращивания производства новых товаров и услуг замедляются, ключевым фактором дальнейшего развития становится рост эффективности производственно-сбытовых процессов, важная роль в которых отводится как транспортно-логистическим, так и собственно промышленным пред-

приятиям (ТЛП и ПП)). Данные процессы и системы претерпевают постоянные изменения, что приводит к появлению новых концепций организации и управления промышленными и транспортными предприятиями, адекватных уровню развития рыночных отношений, информационных технологий (ИТ) и инженерных систем. Современная степень развития ТЛП и ПП требует того, что они должны проектироваться как распределенные, унифицированные и многофункциональные системы с рекурсивным, самопроизводящимся адаптивно замкнутым характером коммутационных отношений. В адаптивно замкнутых ТЛП и ПП формируются (синтезируются) такие механизмы их избирательной чувствительности, которые обеспечивают ослабление нежелательных воздействий внешней среды за счет контуров отрицательной обратной связи и усиление полезных воздействий (сигналов) по соответствующим каналам положительной обратной связи. В ТЛП и ПП должны быть реализованы технологии управления, обеспечивающие гибкое сочетание принципов иерархического и сетевого управления при динамическом структурно-функциональном синтезе ее облика. Кроме того, создание программно-математического и информационного обеспечения (ИО) и соответствующих подсистем должно базироваться на концепции их самоподобного рекурсивного построения. Самоподобие является одним из необходимых условий жизнеспособности ТЛП и ПП. Предварительные исследования показали, что применительно ТЛП и ПП самоподобие проявляется, прежде всего, через концепцию развивающейся ситуации, которая на формально-математическом уровне описания реализуется с помощью задания состояний, структурных состояний, макросостояний, многоструктурных макросостояний, классов эквивалентных структурных состояний [1-6].

Наиболее распространенной формой ТЛП и ПП являются цепи поставок (ЦП) [5-6]. Цепь поставок (процессное понимание) — это совокупность потоков и соответствующих им кооперационных и координационных процессов между различными участниками цепи создания стоимости для удовлетворения требований потребителей в товарах и услугах.

Цепь поставок (объектное понимание) — это совокупность организаций (предприятий-изготовителей, складов, дистрибуторов, 3PL и 4PL провайдеров, экспедиторов, оптовой и розничной торговли), взаимодействующих в материальных, финансовых и информационных потоках, а также потоках услуг от источников исходного сырья до конечного потребителя.

Развитие и видоизменение принципов управления цепями поставок имеет объективные основы. Предпосылками происходящих изменений в подходах к организации и управлению цепями поставок являются, прежде всего, появление новых форм конкурентной борьбы, основанных

на стратегическом взаимодействии (Collaboration) предприятий, а также все большее проникновение информационных технологий в бизнес.

Мощным стимулом синтеза новых поколений цепей поставок (ЦП) стали интенсивно развиваемые информационные технологии и соответствующие информационные системы (ИС) [7-9]. К числу указанных технологий можно, в первую очередь, отнести: технологию совмещенного проектирования; технологию удовлетворения ограничений; технологии радиочастотной идентификации и мобильные информационные технологии; технологию системного моделирования и интеллектуального проактивного управления; технологию создания объектно-ориентированных и интеллектуальных баз данных; технологию интеллектуальных геоинформационных систем; технологию проектирования и применения многоагентных и гибридных систем. Все перечисленные технологии получают широкую практическую реализацию в рамках таких глобальных бизнес-проектов, как Интернет вещей и Промышленный Интернет, который ассоциируется с программой Индустрия 4.0 [10-14]. Наряду с указанными технологиями необходимо учитывать также и перспективы их дальнейшего развития, которые базируются на тенденциях в ИТ сфере, связанных с переходом от классических вычислений к альтернативным способам организации вычислительного процесса; с использованием технологии активных объектов; ориентацией на приоритет моделей, а не алгоритмов; реализацией естественного параллелизма вычислений; проактивности и самоорганизации вычислений [10, 15].

При этом технологии проактивного мониторинга и управления ТЛП и ПП в отличие от традиционно используемого на практике реактивного управления, ориентированного на оперативное реагирование и последующее недопущение инцидентов, предполагает предотвращение возникновения данных инцидентов за счет создания в системах мониторинга и управления принципиально новых прогнозирующих и упреждающих возможностей при формировании и реализации управляющих воздействий, базирующихся на концепции системного (комплексного) моделирования. Данная концепция, в свою очередь, базируется на полимодельном описании соответствующей предметной области, а также применении комбинированных методов, алгоритмов и методик для решения задач многокритериального оценивания, анализа и выбора наиболее предпочтительных альтернативных управленческих решений [15, 21-23].

В связи со сказанным весьма актуальными становятся вопросы анализа влияния существующих и перспективных интеллектуальных информационных технологий (ИИТ), положенных в основу проектов Индустрия 4.0 и киберфизических систем, на организацию функционирования транспортно-логистических и промышленных предприятий.

2. Индустрия 4.0: четвертая индустриальная революция.

2.1. Экономический и управленческий смысл концепции

Индустрия 4.0. Из курса современной истории человечества известно о трех индустриальных революциях: первая индустриальная революция произошла в конце 18 века, когда были изобретены и усовершенствованы паровые двигатели, заменившие ручной труд машинным и многократно увеличившие темпы производства и транспортировки продукции; 1863–1947 гг. называют эпохой второй индустриальной революции, в основе которой — изобретение Генри Фордом конвейерной линии, массовая электрификация производства, вызвавшие резкий подъем промышленного производства и снизившие стоимость готовой продукции, например, автомобиля, сделав ее доступной для массового покупателя.

Третья индустриальная революция началась в 1960 гг. Обычно ее называют цифровой революцией, потому что это время послужило началом использования полупроводников и ЭВМ. Наше время — это переходный период. Четвертая революция еще не наступила, но инновации третьей уже достигли своего промышленного предела.

На основании данных, которые были получены от квалифицированных экспертов, Клаус Шваб предполагает, что временем начала четвертой промышленной революции можно считать 2025 год [24]. Именно в этот период множество экспертов предсказывают масштабные изменения во всех отраслях нашей жизни. Одно из главных отличий этой революции от остальных состоит в том, что изменения будут происходить стремительными темпами. Пройдет всего 10 лет, и мир кардинально изменится.

Четвертую индустриальную революцию или Индустрию 4.0 можно охарактеризовать следующими особенностями [3, 6, 8, 17-20, 24]: комплексная интеллектуализация и автоматизация производственных и транспортных процессов; радиочастотная идентификация деталей (RFID), технологического оборудования и роботов, которая осуществляется с помощью идентифицирующей и управляющей электроники; способность быстро и гибко перенастраивать технологическое оборудование; детали и оборудование самостоятельно «договариваются» между собой с помощью смарт-сенсоров, когда и на каком оборудовании должен производиться продукт, чтобы скорость и эффективность производства были максимальными. Возникает новое поколение smart factories — «умных фабрик». Основное их отличие от классических фабрик, работающих на принципах линейных производственных процессов, состоит в том, что все производственные циклы гибко конфигурируются в соответствии с запросами покупателя и загрузкой оборудования. Это повышает гибкость, эффективную загрузку и про-

изводительность «умных фабрик». Производство индивидуализированной продукции становится рентабельным и возможным в отраслях массового серийного производства, таких как автомобилестроение, машиностроение, авиастроение и другие.

Генри Форд, основатель одного из крупнейших автомобильных концернов и идеолог конвейерного производства, говорил, что покупатель имеет право сам выбрать цвет своего автомобиля, если этот цвет будет черным. Современная автоиндустрия давно отошла от этих принципов. На передний план выходят индивидуализация продукции и удовлетворение запросов покупателя. Индустрия 4.0, основой которой являются высокоавтоматизированные и компьютеризированные сетевые фабрики, фокусируется главным образом на удовлетворении запросов покупателя. Ученые расходятся во мнениях о том, что представляет собой Индустрия 4.0, и следует ли говорить об эволюционном или революционном пути развития индустриализации. Скептики считают, что сама терминология «Индустрия 4.0» придумана политиками и робототехниками в маркетинговых целях. Однако нельзя отрицать тот факт, что научные достижения в области микроэлектроники, сенсорики, робототехники, программного обеспечения и радиоэлектроники и их успешное применение на практике создали новое поколение «умных фабрик».

Аналоги программы Индустрия 4.0 существуют и в других странах: Smart Factory в Нидерландах, Usine du Futur во Франции, High Value Manufacturing Catapult в Великобритании, Fabbrica del Futuro в Италии, «Сделано в Китае-2025» и тому подобное.

В США тоже задумываются о будущем индустриального производства. В 2014 году компании General Electric, AT&T, Cisco, IBM и Intel создали Консорциум промышленного интернета (Industrial Internet Consortium), который сегодня включает уже 170 членов.

Таким образом, Индустрию 4.0 (или по-другому — Производственный Интернет) можно считать ключевым, но более узким понятием, чем Интернет вещей (Internet of Things), поскольку последний объединяет не только оборудование и производственные циклы внутри производственной сферы, но и ИТ вне ее, как, например, смартфоны, умные часы (smart watches), бытовые приборы, «умные дома» (smart homes), автомобили [2, 8]. Такой взгляд на Индустрию 4.0 особенно популярен в Германии, поскольку эта страна является не только лидером мировой индустрии, но и ведущим производителем оборудования и программного обеспечения для автоматизированных производственных систем. Перед страной открываются неограниченные возможности стать главным поставщиком электротехники для «умных фабрик».

2.2. Технологические основы Индустрии 4.0: смарт-сенсоры и искусственный интеллект. Рассмотрим ключевые компоненты, лежащие в основе «умных фабрик». Среди них важнейшим компонентом являются киберфизические системы (Cyber-Physical Systems, далее в тексте — КФС). Термин «киберфизические системы» предложила в 2006 году Хелен Джилл, в то время директор по встроенным и гибридным системам в Национальном научном фонде США, желая подчеркнуть отличительную особенность организованного ею семинара NSF CPS Workshop. Ей удалось уловить общую тенденцию пересмотра роли встроенных систем. Согласно современному взгляду КФС — это системы, состоящие из различных материальных объектов, искусственных подсистем и управляющих контроллеров, которые представляют собой единое целое. В КФС обеспечивается тесная связь и координация между вычислительными и физическими ресурсами. Область действия КФС распространяется на робототехнику, транспорт, энергетику, управление промышленными процессами и крупными инфраструктурами [3, 19].

От существующих мехатронных систем КФС отличаются наличием интеллекта (знаний и процедур их формирования, пополнения, манипулирования) и способностью взаимодействовать со своим окружением; планировать и адаптировать свое собственное поведение согласно окружающим условиям; учиться новым моделям и линиям поведения и быть самооптимизирующимися [13-15, 20]. В перспективе КФС будут также обладать такими свойствами как самосознание и проактивность, самоконфигурирование, самосовершенствование, самолечение, самосохранение [13-14, 20].

Ключевым в КФС является модель, используемая в системе управления, — от того, как она соотносится с реальностью, зависит работоспособность КФС. Разработка унифицированных языков описания разнородных объектов для построения моделей их функционирования, а также методов оценки качества этих моделей являются чрезвычайно актуальными научными задачами. Наряду с этим большое значение приобретает развитие теории оценивания качества моделей и полимодельных комплексов, с тем чтобы повысить обоснованность автоматического выбора моделей своего поведения самими КФС [15].

На рисунке 1 в качестве примера приведена обобщенная структура системы управления жизненным циклом перспективной сложной организационно-технической системы, в состав которой включены как существующие, так и перспективные КФС, которые в сочетании с соответствующими организационными структурами уже образуют социокриберфизические системы (СКФС).

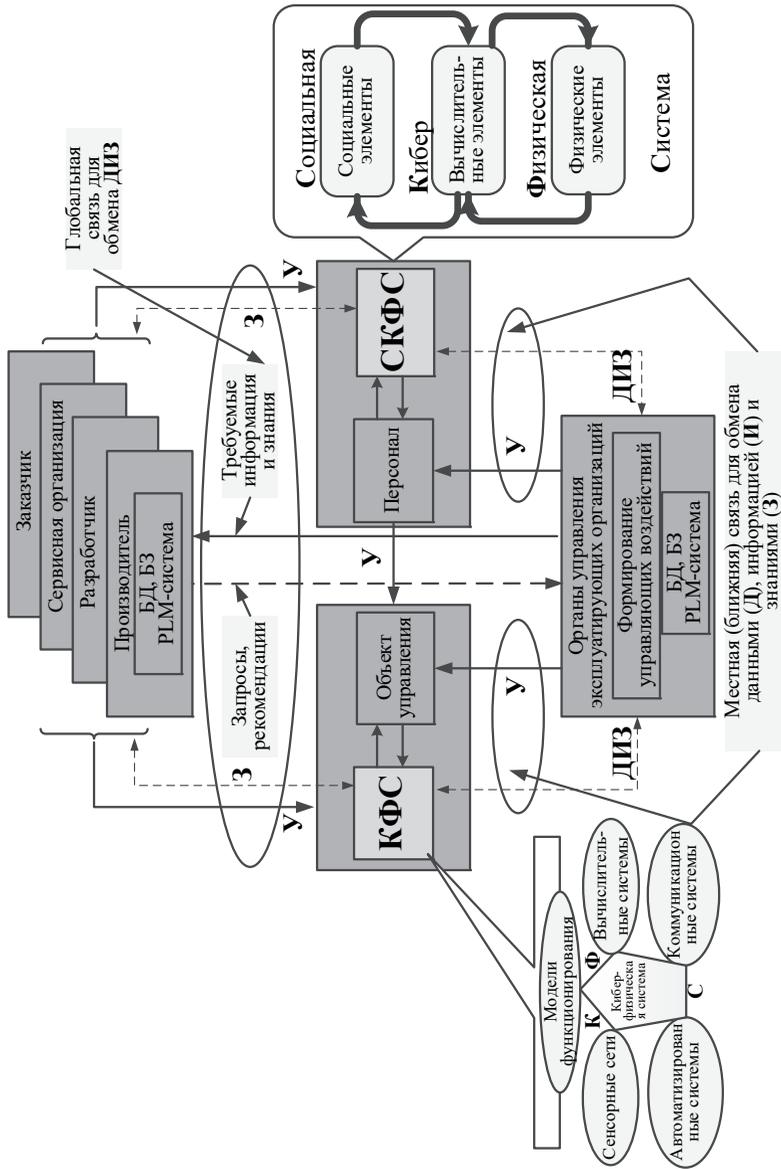


Рис. 1. Обобщенная структура системы управления жизненным циклом сложной организационно-технической системы

На рисунке 1 приняты следующие условные сокращения: З — знания, У — управление, ДИЗ — данные, информация, знания, БД, БЗ — базы данных, базы знаний. Данный рисунок иллюстрирует, как основные компоненты и технологии Индустрии 4.0 взаимодействуют между собой на различных этапах создания и использования перспективных производственных процессов и систем (в том числе и применительно к основным объектам данного обзора — ПЛП, ПЛС, ЦП и соответствующим ИС).

Необходимо подчеркнуть, что в настоящее время в результате повсеместного внедрения на практике как КФС, так и СКФС происходит эволюционный переход от существующих систем управления жизненным циклом (СУ ЖЦ) (зарубежный аналог — Product Lifecycle Management (PLM) системы) к более совершенной технологии управления ЖЦ, которая получила название CL2M (Closed Loop Lifecycle Management — управление ЖЦ с обратной связью).

Новое поколение PLM систем (СУ ЖЦ СТО) должно быть сопряжено с распределенными базами знаний. В связи с этим в дальнейшем предусматривается разработка (адаптация), внедрение и использование следующих новейших информационных технологий и элементов новой СУ ЖЦ: встраиваемых в изделие интеллектуальных информационных сенсоров и приборов (например, радиочастотных меток — RFID tags, приемников GPS, GSM сигналов), позволяющих осуществлять глобальное и локальное позиционирование изделий, получать и предварительно обрабатывать данные и информацию о состоянии изделия, а также сведения о необходимом объеме его обслуживания; встраиваемых в изделие интеллектуальных информационных сенсоров для контроля технического состояния изделий (smart tags), обеспечивающих непрерывное накопление и обработку данных, информации и знаний о режимах функционирования изделия в течение его ЖЦ, а также сведений о степени удовлетворенности заказчика готовым изделием и создания на этой основе лучших, ориентированных на заказчика и устойчиво развивающихся изделий, процессов и услуг; мобильных телекоммуникационных систем и средств для беспроводного считывания и передачи данных со встроенных в изделие интеллектуальных информационных сенсоров; программного обеспечения для беспроводного Интернет соединения мобильных средств считывания информации с PLM системами в любом месте мира; программного обеспечения, которое позволит с учетом разграничения доступа распределять данные, информацию и знания между субъектами управления в любое время и в любом месте. На базе перечисленных возможностей и технологий использования КФС (как нового поколения встроенных систем) наряду с принципиальным изменением технологии управления ЖЦ изделий происходят такие же масштабные изменения в организации использования произ-

водственного оборудования, о состоянии которого дистанционно получают данные и информацию не только люди, его эксплуатирующие, но и производители данного оборудования. Все это позволяет своевременно и обоснованно проводить регламентные работы, предсказывать аварии, проводить плановые предупредительные ремонты, заранее подготавливать замену тех или иных деталей.

В РФ данные системы уже сейчас получили самое широкое распространение в ракетно-космической отрасли.

На многих современных фабриках в РФ и за рубежом уже существует компьютеризованное управление станками и роботами с помощью распределенной сети компьютеров. На данный момент эта технология распространяется на некоторых наиболее технически оснащенных заводах в автомобилестроении и производстве микрочипов.

Так, например, для устранения простоев на «умных автомобильных фабриках» существуют управляющие компьютеры, которые на основе динамического ситуационного решения задачи теории расписаний осуществляют коррекцию производственного плана и перенаправляют кузов к соответствующему роботу.

Подобные задачи сложны как с модельно-алгоритмической, так и технической точки зрения и их успешное решение возможно только в случае, если соответствующие детали, роботы, станки и оборудование обладают способностью формировать и реализовывать в проактивном режиме соответствующие управляющие воздействия.

Проведенный анализ показывает, что в современных условиях для обеспечения требуемой степени автономности, качества и оперативности управления промышленными предприятиями (ПП) и соответствующими производственными процессами (ПрП) необходимо решить следующие основные научно-технические проблемы: во-первых, обеспечить модельно-алгоритмическое описание процессов смысловой интерпретации всех возможных штатных и нештатных состояний их функционирования, и во-вторых, на этой основе решить весь перечень задач комплексной автоматизации и интеллектуализации процессов мониторинга и управления ПП и ПрП в различных условиях обстановки. Однако, к сожалению, в подавляющем большинстве случаев на практике мониторинг, прогнозирование и управление их состояниями в указанной выше трактовке автоматизирован в лучшем случае лишь частично. Как правило, в современных системах мониторинга и управления (СМУ) ПП и соответствующими ПрП операторам представляется смысловая информация только о состояниях их элементов, а не объектов контроля в целом. Указанные обстоятельства приводят к тому, что интегральную оценку их состояния так же, как и формирование необходимых управ-

ляющих воздействий операторы выполняют в основном вручную, базирясь на тех или иных эвристических правилах.

Эти задачи в проекте Индустрия 4.0 планируется возложить, прежде всего, на smart-sensors — «умные сенсоры» и соответствующие управляющие подсистемы. «Умный сенсор» (в отличие от классических встроенных систем) способен измерять многочисленные функционально связанные физические параметры. Эти параметры могут характеризовать в каждый конкретный момент времени состояние заданного объекта, обрабатывать и анализировать полученные данные и информацию и осуществлять их преобразование в цифровой вид для последующей передачи управляющему компьютеру. На управляющем компьютере осуществляется интеграция данных, информации и знаний, поступающих от распределенной системы умных сенсоров. КФС системы вместе с компьютерами позволяют реализовать на практике технологии повсеместных вычислений и коммуникаций при решении задач анализа и синтеза бизнес-процессов предприятий в динамике, а также задач оценивания таких параметров, как качество, время, расход ресурсов и тому подобное.

К настоящему времени современная наука создала богатый методологический и методический аппарат, который наряду с КФС может быть положен в основу рассматриваемых технологий Интернета вещей и Промышленного Интернета, позволяющий успешно преодолевать трудности, связанные с воздействием факторов сложности в современном мире. В основу этого аппарата положена междисциплинарная отрасль системных научных знаний. Возникновение системной отрасли научных знаний является велением времени, так как на данном этапе развития науки (этапе интеграции научных знаний) на передний план в развитии научных знаний выступает методология, требующая сочетания (единства) процессов анализа и синтеза при изучении свойств сложных объектов и процессов (к числу которых относятся и рассматриваемые в статье ПрП) как целостных образований, состоящих из взаимосвязанных частей и обладающих качественно новыми свойствами по сравнению со свойствами этих частей. При этом в настоящее время речь должна идти не о взаимном поглощении, а о взаимном дополнении, концептуальном и идейном взаимообогащении, гармоничном и согласованном развитии базовых междисциплинарных направлений, в числе которых, прежде всего, выделяются такие научные направления, как информатика, общая теория систем, кибернетика (в современных условиях — неокибернетика) [23-29].

Говоря о последнем научном направлении (неокибернетика), получившем существенное развитие в нашей стране и за рубежом, следует указать, что основные научные направления неокибернетики

связаны с исследованием проблематики управления сложностью и организации деятельности социокберфизических систем (СКФС) в указанных условиях (problem of complexity control and management). Следует отметить, что рассматриваемые в данной статье цепи поставок тоже относятся к классу сложных СКФС.

Организация управления СКФС требует (в отличие от управления КФС) проведения междисциплинарных исследований с привлечением специалистов разных специальностей: экономистов, биологов, физиков, математиков, специалистов в области компьютерных технологий. При этом само управление в его классическом понимании (то есть реализация функций целеполагания, планирования, оперативного управления, мониторинга, контроля, координации и т.п.) в таких системах без учета человеческого фактора невозможно, так как поведение (функционалирование) данных систем слабо предсказуемо, и поэтому в конечном итоге сводится к организации целенаправленного формирования возбуждающих и тормозящих стимулов (управляющих воздействий) в многочисленных контурах прямых и обратных связей, обладающих большой динамической инерционностью и требующих проведения многовариантного упреждающего (проактивного) прогнозирования сценариев функционирования СКФС [15, 23].

Важная роль в функционировании СКФС отводится процессам их самоорганизации, которую можно трактовать как их способность управлять разнообразием своих состояний и состояний внешней среды. Борьба с разнообразием внешней среды возможна на основе динамической адаптации структур системы к изменяющимся условиям функционирования. Следует признать, что концепция централизованного управления с отрицательной обратной связью, лежащая в основе классической кибернетики Н. Винера, в настоящее время не всегда адекватна современным процессам управления [15, 24-26]. В частности, она не позволяет описывать поведение сложных распределенных самоорганизующихся систем, к числу которых относятся СКФС. В связи с этим в книге [28] систематизированы методологические подходы современной кибернетики, применимые к моделированию распределенных самоорганизующихся систем. Эти подходы в основном базируются на известных концепциях, рассмотренных выше, но в дополнение к ним предоставляют конструктивные механизмы описания процессов саморегуляции и эволюции, применимые на практике. Примеры, рассмотренные в книге, не исчерпываются биологическими системами, речь идет прежде всего о системах естественного происхождения. При этом процессы взаимодействия СКФС с внешней средой, с одной стороны, обеспечивают поддержание динамического равновесия системы, а с другой стороны, наиболее важны с точки зрения внешне-

го наблюдателя. Для описания процессов взаимодействия Н. Нуўтуніемі в [28] использует понятие абстрактной энергии (информационной энергии), применяя термин «эмергия» (как результат объединения таких понятий как энергии и эмерджентность). Можно говорить о потоке «эмергии» из системы во внешнюю среду и об обратном потоке. Теперь, опираясь на сформулированные принципы моделирования самоорганизующихся систем, можно построить общую модельную схему самоорганизации, обобщающую схему классического централизованного управления с обратной связью (рисунок 2) [26].

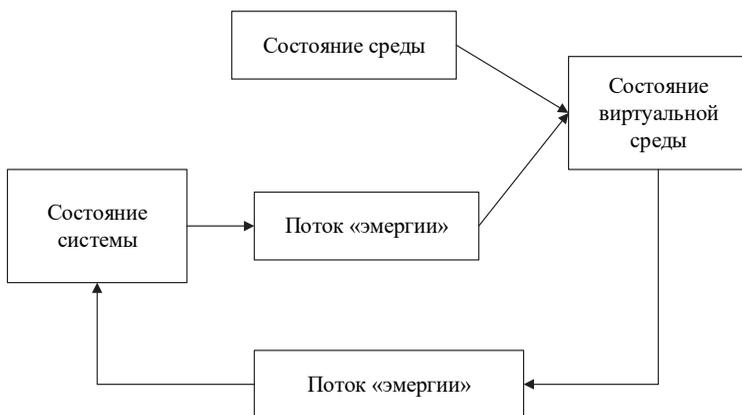


Рис. 2. Самоорганизация с обратной связью, замыкающейся через виртуальную среду

На приведенной схеме показано, что система непосредственно взаимодействует с некоторой абстрактной виртуальной средой. Заметим, что целенаправленная система может использовать модель внешней среды для формирования выходного потока «эмергии». При реализации схемы стрелки заменяются функциональными зависимостями, которые в общем случае могут быть определены неявно, например, алгоритмически, в виде систем алгебраических или обыкновенных дифференциальных уравнений.

Эту модельную схему можно использовать также для исследования целенаправленных воздействий на самоорганизующуюся систему, то есть для управления указанной системой.

Для этого совокупность характеристик виртуальной внешней среды следует разделить на группы регулируемых и нерегулируемых характеристик. Далее управляющие воздействия на систему заключаются в непосредственном изменении регулируемых характеристик. Более того, управлению могут подвергаться и структуры. В работах [15, 29]

представлены основы управления структурной динамикой сложных технических объектов, которые могут быть положены в основу еще более интересного научного направления, связанного с управляемой самоорганизацией. При этом, говоря о данном варианте реализации процессов самоорганизации, следует подчеркнуть, что управление в рассматриваемом случае трактуется не в классическом (кибернетическом) смысле, а более широко. Предлагается управлять не самим объектом управления (в нашем случае цепями поставок и соответствующими информационными системами, обеспечивающими их эффективное функционирование), а разнообразием потенциальных состояний, в которых он может находиться. Другими словами, в отличие от классической самоорганизации в рамках предлагаемого подхода на этапе применения целенаправленно и динамически формируются (выявляются) дополнительные свойства и характеристики объекта управления. К их числу могут относиться требуемая степень функциональной избыточности в виде дополнительных параметров, структур, режимов функционирования, ситуационное изменение границ между объектом управления и внешней средой и т.п.). Это позволяет существенно повысить эффективность процессов реализации традиционной самоорганизации. Авторами данной статьи предложено несколько вариантов реализации специального модельно-алгоритмического и программного обеспечения проактивного управления структурной динамикой КФС (в том числе и цепями поставок), в рамках которых на конструктивном уровне были продемонстрированы возможности предлагаемого подхода к проведению управляемой самоорганизации применительно к различным видам технических систем. На сайте [30] можно подробно ознакомиться с проектами и публикациями авторов, в которых использовались предложенные подходы.

В результате таких воздействий опосредованно, а именно через изменение состояния систем, могут изменяться значения непосредственно нерегулируемых характеристик. Как правило, цель управления системой формулируется именно в терминах изменения указанных нерегулируемых характеристик виртуальной среды. Регулируемые характеристики (входные воздействия на систему) можно в свою очередь разделить на два класса. Значения характеристик первого класса непосредственно воздействуют на многие элементы системы. Для экономических систем примером таких характеристик могут служить величины налогов, пошлин, акцизов и так далее. Входные воздействия второго класса влияют на состояние одного или нескольких элементов системы. Последствия таких воздействий с течением времени могут сказываться на состоянии других элементов за счет взаимодействия с уже изменившимися элементами. Например, банкротство крупного предприятия оказывает влияние на его поставщиков и клиентов. У по-

ставщиков и клиентов, в свою очередь, могут существовать свои поставщики и клиенты, условия деятельности которых также изменятся.

Процесс изменения состояния элементов самоорганизующейся системы имеет прямые аналогии с процессами диффузии в жидкостях и газах, поэтому для их описания могут быть использованы известные модели.

Следует отметить, что развивая концепции классической кибернетики второго порядка Фоерстера, Н. Нуйтүниemi ввел следующие конструктивные определения кибернетических систем высоких порядков исходя из их способности к самоорганизации и эволюции [28]: *кибернетическая система первого порядка* находит равновесие в условиях возмущающих воздействий внешней среды; *кибернетическая система второго порядка* настраивает внутренние структуры для лучшего соответствия наблюдаемым воздействиям внешней среды в интересах увеличения жёсткости системы (по критерию минимума наблюдаемых изменений в системе); *кибернетические системы более высоких порядков* перестраивают внешние структуры для создания более комфортной внешней среды. Внешние структуры перестраиваются по критерию максимального потока «эмергии» (максимального переноса «эмергии» через внешнюю среду). Для экономических систем этот критерий может принимать форму максимизации оборота.

К настоящему времени на практике приходится сталкиваться с многочисленными примерами реально существующих кибернетических систем различных порядков, получен ряд интересных теоретических и практических результатов при исследовании проблем управления структурной динамикой ЦП, которые можно повсеместно использовать в условиях расширяющейся технологической базы, предоставляемой программой Индустрия 4.0 [3, 4, 19].

Проведем краткий анализ того, в какой степени перечисленные научно-методические подходы реализуются на практике уже сейчас в рамках программы Индустрия 4.0.

3. Анализ практических примеров Индустрии 4.0. В представленных далее примерах проиллюстрированы те принципиально новые возможности по организации процессов управления, а также оптимизации цепочек создания стоимости, которые появляются при внедрении в существующие производственно-логистические системы новых информационных технологий, разрабатываемых в рамках Индустрии 4.0

3.1. Пример компании по производству микрочипов. Компания Globalfoundries — известный в мире производитель микрочипов. На заводе в Дрездене уже производятся кремниевые шайбы (так называемые «вафли»). Транспортировку «вафлей» осуществляют FOUPS —

«Front Opening Unified Pods», каждый из которых перевозит до 25 «вафлей». Они передвигаются по рельсам через все сборочные цеха, а специальные роботы осуществляют погрузку и разгрузку «вафлей» на всех этапах сборки, считывают с них электронную информацию, на основе которой самостоятельно определяют следующий пункт назначения на сборочной линии для каждой «вафли».

При этом ИИТ внедрены не только в используемых роботах, но и в самих FOUPS: на их поверхность нанесены электронные модули, передающие центральному компьютеру информацию о местонахождении каждого из FOUPs на сборочной линии. Центральный компьютер может дать им команду «припарковаться» на конкретном сборочном узле и ждать, чтобы избежать «пробок» на особенно загруженных участках линии или, например, пропустить вперед более срочный заказ от покупателя [2-4].

3.2. Пример из автоиндустрии. Ведущие производители автомобилей, такие как концерн «Volkswagen», активно применяют принципы и технологии проекта Индустрия 4.0 на практике. Основные производственные процессы, такие как сварка, лакировка, монтаж колес и стекол давно роботизированы. Компания успешно внедряет ИИТ при сборке деталей: каждый автомобильный кузов в момент производства получает свой RFID-чип, на котором сохранена информация о том, каким должен быть цвет автомобиля, какую оснастку он будет иметь, когда должен быть изготовлен и так далее. «Volkswagen» применяет цифровые и Интернет-технологии не только на собственных заводах, но и в цепях поставок. Например, специальный Big Data Monitor, созданный для завода Audi в немецком городе Карлсруэ позволяет в режиме реального времени проводить мониторинг логистических потоков 12 тысяч поставщиков компании. И компания, и поставщики получают актуальную информацию о заказах, могут точно прогнозировать спрос и сроки поставок, сокращая простои оборудования и складские запасы [2, 7].

3.3. Пример компании, организующей цепи поставок. SupplyOn — компания, занимающаяся с 2000 года созданием единой информационной платформы в области управления цепями поставок для таких компаний, как Bosch, Continental, Siemens, Airbus и другие и их поставщиков. Основными клиентами компании являются предприятия авиа- и автомобилестроения, электроники, промышленные предприятия. С момента своего основания в 2000 году SupplyOn стала использовать информационные технологии для взаимодействия компаний с поставщиками и транспортными компаниями (например, Software-as-a-service (SaaS) задолго до того, как SaaS и Cloud-computing стали популярными).

лярными в мире. Сегодня единая информационная платформа Supply On объединяет свыше 12 тысяч компаний в 70 странах мира.

Задачей SupplyOn является перевод цепи поставок в цифровую форму. Наличие прозрачной и высокотехнологичной единой информационной платформы — залог успешного внедрения Индустрии 4.0 в цепь поставок. Supply On разработала цифровое решение для стандартизированных процессов цепей поставок железнодорожных компаний — приложение RailSupply, позволяющее отображать в электронном виде процессы заказов, поставок и учета крупных ж/д перевозчиков и их компаний-клиентов [3, 8].

3.4. Adidas: 3D-принтеры и первый размер партии. Компания Адидас представила в 2015 году свою «фабрику будущего» в Германии (г. Ансбах) — Speed Factory, на которой планируется производство спортивных товаров в максимально короткие сроки по принципу «размер партии = 1». Ранее почти вся продукция компании производилась в Китае, сроки поставки потребителю составляли в среднем 3 месяца. На Speed Factory активно применяются роботы, способные изготавливать высоко функциональную продукцию в максимально сжатые сроки и по индивидуальным заказам. На фабрике в немецком Ансбахе работает 6 современных единиц оборудования: вязальная машина производит из полимерных нитей высококачественный материал для кроссовок и передает его на лазерную кройную машину, одновременно с этим на 3D-принтере изготавливается подошва. Последний этап производства — «сварка» подошвы и верха обуви. Использование современных технологий при производстве обуви позволило практически исключить использование клея, сделав готовую продукцию более экологичной. Цель Speed Factory — гибкая и быстрая реакция на запросы потребителя и локализация вблизи основных покупателей, позволяющая сократить время нахождения в пути, снизить транспортные расходы и повысить независимость от поставщиков. Для производства одной пары обуви требуется приблизительно 5 часов.

Автоматизация производства на Speed Factory позволит создавать запасы там, где локализованы основные потребители. Открытие «фабрики будущего» в декабре 2015 года явилось важнейшей инновацией для компании Адидас и позволило за счет создания автоматизированного, децентрализованного и гибкого производства изготавливать индивидуализированные спортивные товары высокого качества, находясь вблизи основного рынка в Европе, снижая транспортные затраты и повышая экологичность готовой продукции [9, 16].

3.5. Amazon Robotics: Индустрия 4.0 в дистрибуционной логистике. Дистрибуционные центры сегодня сталкиваются с необходи-

мостью управлять возрастающими потоками товаров, увеличивающимися объемами информации, рисками и ожиданиями других участников цепи поставок. Внедрение Индустрии 4.0 в дистрибуционные центры может повысить гибкость при выполнении операций, снизить операционные затраты, автоматизировать процессы и, как следствие, обеспечить экономический рост.

Технологии, применяемые в Индустрии 4.0, позволяют создавать полностью роботизированные «умные» дистрибуционные центры, объединяя интеллектуальные возможности существующих и перспективных производственных роботов. Специфика дистрибуционных центров заключается в том, что в отличие от фабрик со сборочными линиями и цехами решения для дистрибуционных центров в сфере интеллектуального проактивного мониторинга и управления должны быть более гибкими и высокотехнологичными. Это вызвано колоссальным ассортиментом товаров, поступающим в крупные дистрибуционные центры, постоянно меняющимся спросом и ожиданиями других участников цепи поставок. Транспортировка товаров в крупных дистрибуционных центрах осуществляется автоматизированными роботами-погрузчиками (AGV), передвигающихся по складу с использованием локальных навигационных устройств, обеспечивающих их точное позиционирование. Современные сенсорные технологии позволяют роботам самостоятельно выбирать оптимальный маршрут, корректировать задачи и взаимодействовать с другими роботами. Рассмотрим пример американской компании Amazon Robotics, крупнейшего производителя роботизированных систем.

Amazon Robotics разработала роботов для передвижения стеллажей независимо от размера продуктов и форм их размещения в стеллажах. Роботы управляются централизованно на основе беспроводных технологий. Передвижения основаны на чтении QR-кодов на полу складского комплекса. Роботы являются обучаемыми и самоадаптируемыми элементами системы управления складом, обладая к тому же способностью к коммуникации друг с другом. По состоянию на вторую половину 2015, Amazon использует более 30000 роботов в 13 дистрибуционных центрах, что позволило компании существенно ускорить внутрискладские операции и сократить число ошибочных управленческих решений [4, 17, 18].

4. Анализ основных направлений влияния Индустрии 4.0 на изменения принципов управления цепями поставок. Проведенный анализ результатов внедрения технологий Индустрии 4.0 показывает, что предприятия различных форм собственности, предполагающие внедрение указанных технологий, должны быть хорошо подготовлены к

указанным процессам как в организационно-техническом, так и финансово-экономическом плане. При этом только комплексная автоматизация, информатизация и интеллектуализация основных технологий функционирования ТЛП и ПП будет обеспечивать гарантированный рост оборота, гибкости производства, производительности и в целом эффективности компаний, входящих в состав указанных цепей поставок.

Однако внедрение и использование технологий Индустрии 4.0 и киберфизических систем будет проходить в условиях влияния различного рода неопределенных факторов (возмущающих воздействий). Это в конечном итоге будет приводить к ситуациям принятия управленческих решений в ТЛП и ПП в условиях риска и неопределенности. Указанные факторы условно можно разделить на три большие класса [5]: индифферентные неопределенные факторы (возмущающие воздействия 1 типа); целенаправленные неопределенные факторы (возмущающие воздействия 2 типа); факторы, связанные с неопределенностью целей субъектов, участников ТЛП и ПП (возмущающие воздействия 3 типа).

Индифферентные неопределенные факторы, или по-другому неопределенность воздействий природы, характеризует ограниченность наших знаний об изучаемых и используемых объектах. В этих случаях, как правило, неизвестные факторы представляют собой обычные объекты изучения теории вероятности, математической статистики, в последнее время теории нечетких множеств. При этом предполагается, что вероятностно-статистические характеристики этих факторов известны или потенциально могут быть получены от экспертов в виде соответствующих функций принадлежности. Следует отметить, что во многих случаях невозможно математическое описание степени влияния различных факторов на процесс достижения целей, стоящих перед участниками ТЛП и ПП, либо же это описание будет сделано с недостаточной степенью точности и достоверности. Это связано с тем, что процессы в реальных ТЛП и ПП, как правило, носят неповторяющийся, нестационарный характер, а также с отсутствием или неполнотой необходимой ретроспективной статистической информации.

Неопределенность, вызванная деятельностью конкурентов (либо каких-то других участников рыночных отношений) обусловлена активностью агентов (предприятий) и центра (органа координации ТЛП и ПП) и их предпочтениями (интересами). Анализ неопределенности действий конкурентов базируется на мультиагентных системах, игровых моделях, методе нечеткой логики. В данных подходах реализуются принципы управления конфликтами, а также учета так называемых «мягких» трудно формализуемых факторов, таких как доверие, репутация и так далее.

Неопределенность целей, или иными словами — многокритериальность, связана с невозможностью однозначного формулирования цели и условия соответствующих задач планирования и управления ТЛП и ПП. Для решения многокритериальных задач большой размерности с необходимой оперативностью в настоящее время широко используются нейронные сети, генетические алгоритмы, АСО-алгоритмы, мультиагентные системы.

В целом для обоснованной реализации ранее перечисленных интеллектуальных информационных технологий, составляющих базу программы Индустрия 4.0. в существующих и перспективных ТЛП и ПП представляется целесообразным проводить соответствующее комплексное моделирование и анализ различных сценариев внедрения и использования данных технологий.

Кратко остановимся еще на одном весьма важном вопросе, имеющем непосредственное отношение к проблеме проактивного мониторинга и управления ТЛП и ПП, а именно на проблеме многокритериального количественного оценивания и анализа вкладов ИИТ в основную деятельность предприятий (фирм), из которых формируются указанные предприятия.

При этом возьмем современные цепи поставок (ЦП) в качестве примера, объединяющего подходы, используемые при оценивании и анализе вкладов ИИТ как в деятельность ТЛП, так и ПП. Содержание многочисленных публикаций по данной теме показывает, что, к сожалению, на данный момент времени нет корректных методик, позволяющих установить однозначную связь между повышением производительности ИИТ и информационных систем (ИС) и увеличением доходности и прибыльности бизнеса в соответствующей организации (в нашем случае в ЦП). Так применительно к оцениванию эффективности управления ЦП необходимо понимать, что цели, которые ставятся перед управлением цепями поставок (например, уровень сервиса или прибыль) связаны с так называемой потенциальной эффективностью. Реальная же эффективность реализуется через устойчивость ЦП. Поэтому, наряду с экономической эффективностью ЦП, большое значение при комплексном оценивании и выборе эффективных стратегий управления ЦП в динамически изменяющейся обстановке является обеспечение их безопасности и устойчивости. Как показывает практика, наблюдается значительное снижение экономических эффектов управления ЦП вследствие повреждений грузов, коллапсов транспортных систем, нарушений в финансовых потоках, а также недостаточной координации в ЦП (колебания спроса, несоответствия объемов производства и закупок и т.д.). Современной тенденцией понимания эффек-

тивности ЦП является такое их проектирование управление, которые бы характеризовались высоким уровнем экономической эффективности и необходимым уровнем устойчивости. В основу соответствующих расчетов показателей эффективности управления ЦП положены общеизвестные SCOR-модели, а также методики их использования, которые представлены в работах [1, 5, 9], а также на сайтах Supply Chain Council (www.supply-chain.org и supply-chain.ru). Однако в условиях широкомасштабного цифрового производства и цифровых экономик, данные модели должны быть дополнены моделями, с помощью которых учитывались эффекты, создаваемые ИИТ в рамках их повсеместного использования в ЦП.

Анализ показывает, что в разных предметных областях существуют отдельные частные методики оценивания как на макро, так и на микроуровне эффективности внедрения ИИТ, в том числе и технологий управления жизненным циклом (ЖЦ) (PLM технологий) промышленных предприятий (ПП) и транспортно-логистических компаний (ТЛК) [1, 5]. Данные методики в своем большинстве базируются на методологии функционально-стоимостного анализа (ФСА). Говоря о роли ФСА при решении проблем управления ЖЦ изделий с целью повышения их качества и использования соответствующих PLM систем для информационной поддержки принятия решений, необходимо отметить, что в настоящее время бизнес и государство готовы оплачивать ровно столько информационных ресурсов, сколько им необходимо для организации управленческой деятельности.

При этом они исходят из таких классических показателей эффективности, используемых в настоящее время на рынке компьютерных услуг, как показатель возврата инвестиций (return on investment, ROI), показатель качества обслуживания (quality of service, QoS), показатель совокупной стоимости владения (CCV) (total cost of ownership (TCO)). Лишние информационные ресурсы и избыточные статьи ИТ бюджета — это замороженные инвестиции и ресурсы (более того, потерянные ресурсы, учитывая быстрое моральное старение аппаратно-программных средств и оборудования). Недостаточные ресурсы — это упущенная выгода. Особенностью показателя ССВ является возможность оценки совокупных затрат на создание (приобретение) и использование готовых изделий и ИТ с последующим анализом эффективности сценариев их использования. Методика ССВ позволяет выявить избыточные статьи расходов и дает возможность оценки возврата вложенных в ИИТ и ИС соответствующих организаций, входящих в ЦП. В общем виде показатель ССВ не позволяет оценивать прямой экономический эффект от внедрения ИИТ и ИС, поэтому для ре-

шения данной задачи чаще всего применяется показатель возврата инвестиций (return on investment, ROI). Результатом расчета методики является коэффициент возврата инвестиций в ИС предприятия. С помощью данного показателя можно оценить экономический эффект от вложения средств в повышения качества создаваемых в ЦП новых сервисов на основе внедрения ИИТ и ИС. Главной трудностью данной методики является определение прямых выгод, получаемых компанией от внедрения новых ИИТ и ИС в различных частях ЦП. При анализе указанных выгод выделяются направления бизнеса и цели, стоящие при внедрении ИИТ и ИС, а затем рассчитывается выгода, получаемая соответствующими предприятиями, входящими в ЦП, от их достижения. Это могут быть как повышение качества сервиса, так и возможность разработки и производства новых товаров, дающие данной организации конкурентные преимущества на рынке. В методике ROI важную роль играет показатель TCO, так как он отражает фактические расходы на ИИТ и ИС [32]. Отметим, что показатель ROI рассматривается на протяжении довольно длительного срока и является оценкой инвестиционных проектов. Для того чтобы связать все ранее перечисленные технологии и показатели, с помощью которых оцениваются выгоды и затраты при том или ином варианте создания и развития изделия, с информационными ресурсами, обеспечивающими их реализацию, был предложен сервис-ориентированный подход [15, 31, 32]. С помощью понятия сервиса можно на конструктивном уровне описать взаимодействия ИИТ и ИС системы с бизнес-процессами. По сути, любая информационная система предоставляет набор сервисов для пользователей. Каждый из этих сервисов должен выполнять поставленные перед ним задачи, имеет ряд функции и использует определенное количество ресурсов. Для описания этих сервисов в работе [32] было предложено использовать типовую модель информационных служб ITSM (ITSM). Данная модель представляет собой мощный инструмент для контроля и управления бизнес-процессами компании. Также с ее помощью достигается четкое представление об информационных службах (ИТ сервисах) компании, их структуре и компонентах, а также их затратные составляющие. Другими словами, в рамках данной модели выделяется взаимосвязь информационных служб с финансовыми ресурсами компании. Причем информационные службы являются связующим звеном между ИИТ, ИС и бизнес-процессами, происходящими в ЦП. В этом случае при решении задачи расчета ССВ с помощью модели ITSM информационные службы принимаются в качестве объектов затрат, и для каждой из них решается задача расчета ССВ. Для этого выделяются ресурсы каждой информационной служ-

бы. Затем рассчитывается стоимость этих ресурсов. ССВ рассматривается как обобщенный показатель для набора информационных служб, предоставляющих необходимые сервисы пользователям ИИТ и ИС. В рамках такой модели легко выявляются наиболее проблемные участки существующих и перспективных ИС и, как следствие, удастся с точки зрения экономической эффективности оптимизировать и сокращать расходы на ИС ЦП в целом. Важной проблемой при определении совокупной стоимости сервисов является тот факт, что разные сервисы могут использовать один и тот же ресурс. Для решения этой задачи применяется метод функционально-стоимостного анализа. В общем виде показатель ССВ (ТСО), как уже указывалось ранее, не показывает прямой экономический эффект от внедрения PLM системы, поэтому на практике для решения данной задачи применяется показатель ROI:

$$ROI(t) = \frac{Ef(t)}{TCO(t)}, \quad (1)$$

где $Ef(t)$ — экономическая эффективность внедряемой PLM системы.

С помощью данного показателя можно количественно оценить экономический эффект от вложения средств в ИИТ и ИС. Это может быть как повышение качества сервиса, создаваемых в ЦП, так и возможность разработки и производства новых товаров, дающих организациям, входящих в ЦП конкурентные преимущества на рынке. В работах [5, 32] представлены конкретные пути расчета показателя ROI, базирующиеся на комбинированном использовании данных, формируемых в реальных системах бухгалтерского, финансового и управленческого учета и генерируемых с использованием аналитико-имитационных моделей. В работе [33] проведен анализ подходов и выработка предложений по решению задачи оценивания эффективности информационных систем и технологий. В данной статье проводится дальнейшее развитие содержания [33] в части учета факторов, описывающих эволюционное развитие как самих ПП (на примере ЦП), так и информационных систем, обеспечивающих их функционирование.

Еще одной из важнейших проблем модернизации ЦП на основе повсеместного использования ИИТ и соответствующих ИС, разрабатываемых в рамках программы Индустрия 4.0, является проблема управления указанной модернизацией, в ходе которой надо комплексно решить *пять основных вопросов*: сформировать (синтезировать) конечный облик соответствующих модернизируемых ИС, определить сроки, к которым должна завершиться указанная модернизация, определить технологии модернизации ИС (в рамках которой на каком-то

интервале времени будут одновременно функционировать элементы и подсистемы «старой» и «новой» ИС, чтобы обеспечить устойчивость функционирования в целом ЦП), определить программу перехода от «старой» и «новой» ИС, и наконец сформировать программы перепланирования либо коррекции исходной программы модернизации ИС, если появятся непредусмотренные возмущающие воздействия.

В настоящее время существует большое разнообразие подходов к решению перечисленных пяти задач. Основной их недостаток состоит в разобленном решении. При этом на практике очень трудно согласовать решения, полученные с использованием различных классов моделей, методов и алгоритмов. В рамках разрабатываемой авторами статьи теории проактивного управления цепями поставок удалось все перечисленные задачи формально описать и исследовать с единых методологических и методических позиций. В этом случае задача многокритериального выбора эффективных стратегий создания, применения и модернизации как самих ЦП, так и соответствующих ИИТ и ИС, используемых в них может быть формально представлена в следующем виде: необходимо разработать принципы, подходы, модели, методы, алгоритмы, позволяющие находить такие $\langle U^t, S_{\delta}^{*t_f} \rangle$, при которых выполняются следующие условия:

$$\begin{aligned}
 J_{\theta} \left(X_{\chi}^t, \Gamma_{\chi}^t, Z_{\chi}^t, F_{\langle \chi, \chi' \rangle}^t, \Pi_{(\tilde{\delta}, \tilde{\delta})}^t, t \in (t_0, t_f] \right) &\rightarrow \underset{\langle U^t, S_{\delta}^{*t_f} \rangle \in \Delta_g}{extr}, \\
 \Delta_g = \left\{ \langle U^t, S_{\delta}^{*t_f} \rangle \mid R_{\beta} \left(X_{\chi}^t, \Gamma_{\chi}^t, Z_{\chi}^t, F_{\langle \chi, \chi' \rangle}^t, \Pi_{(\tilde{\delta}, \tilde{\delta})}^t \right) \leq \tilde{R}_g; \right. & \quad (2) \\
 \left. U^t = \Pi_{\langle \delta_1, \delta_2 \rangle}^t \circ \Pi_{\langle \delta_2, \delta_3 \rangle}^t \circ \Pi_{\langle \delta, \delta \rangle}^t; \chi \in \mathbf{B} \right\}, &
 \end{aligned}$$

где χ — индекс, характеризующий различные типы структур ЦП и соответствующих модернизируемых ИС, $\chi \in \{\text{Топ, Фун, Тех, ПМО, ИО, Ор}\}$ — множество индексов, соответствующих топологической, функциональной и технической структурам, структурам программно-математического и информационного обеспечения (ПМО, ИО), организационной структуре, $t \in T$ — множество моментов времени; $X_{\chi}^t = \{x_{\chi l}^t, l \in L_{\chi}\}$ — множество элементов, входящих в состав структуры динамического альтернативного системного графа (ДАСГ) G_{χ}^t (множество вершин ДАСГ), помощью которого задается управляемая структурная динамика как ЦП, так и соответствующих ИС в момент времени t ; $\Gamma_{\chi}^t = \{\gamma_{\langle \chi, l, l' \rangle}^t, l, l' \in L_{\chi}\}$ — множество дуг ДАСГ типа G_{χ}^t , отражаю-

щих взаимосвязи между его элементами в момент времени t ; $Z_{\chi}^t = \{z_{\langle \chi, l, l' \rangle}^t, l, l' \in L_{\chi}\}$ — множество значений параметров, количественно характеризующих взаимосвязь соответствующих элементов ДАСГ. $F_{\langle \chi, \chi' \rangle}^t$ — отображения различных структур ЦП и ИС друг на друга в момент времени t , $\Pi_{\langle \tilde{\delta}, \tilde{\delta} \rangle}^t$ — операция композиции многоструктурных макросостояний с номерами $\tilde{\delta}, \tilde{\delta}$ в момент времени t ; U^t — управляющие (программные и в реальном масштабе времени) воздействия, позволяющие синтезировать как структуры ЦП и соответствующих модернизируемых ИС, так и процессы функционирования предприятий, входящих ЦП; J_{θ} — стоимостные, временные, ресурсные показатели, характеризующие качество функционирования виртуальных предприятий, $q \in Q = \{1, \dots, l\}$ — множество номеров показателей; Δ_g — множество динамических альтернатив (множество структур и параметров ЦП, ИС и предприятий ЦП, множество программ их функционирования); \mathbf{B} — множество номеров пространственно-временных, технических и технологических ограничений, определяющих процессы реализации ЦП для различных сценариев возмущающих воздействий; \tilde{R}_g — заданные величины; $T = (t_0, t_f]$ — интервал времени, на котором синтезируются и модернизируются как ЦП, ИС, так и облики соответствующих предприятий, входящих в ЦП [5, 6, 15, 29, 30]. Особенностью модели вида (2) является то, что с ее помощью можно конструктивно описать и решить традиционные задачи управления ЖЦ ЦП и ИС (в рамках технологий PLM), но и задачи проактивного управления ЖЦ ЦП и ИС уже в рамках принципиально новой концепции управления ЖЦ указанных объектов, базирующейся на технологии CL2M.

Представленную формулой (2) проблему многокритериального выбора эффективных стратегий (программ) модернизации ЦП на основе технологий, разрабатываемых в проекте Индустрия 4.0. предлагается решить в рамках следующей многоэтапной процедуры, в соответствии с которой *на первой фазе* должно осуществляться формирование (генерирование) допустимых вариантов многоструктурных макросостояний ЦП (в том числе ИС) или, другими словами, должен проводиться структурно-функциональный синтез нового облика ЦП, соответствующего складывающейся (прогнозируемой) обстановке. На данном этапе на основе многомерного ортогонального проектирования на множество достижимости, сформированного в результате комплексного моделирования структурной динамики ЦП, множества, с помощью которого задаются требования, предъявляемые к частным показателям

эффективности ЦП (новому облику ЦП), происходит формирование множества неокончательных решений (множества недоминируемых, эффективных альтернатив, множества Парето). Окончательный выбор эффективных системотехнических решений должен осуществляться из указанного множества. *На второй фазе* проводится выбор и реализация конкретного варианта многоструктурного макросостояния ЦП с одновременным синтезом (построением) адаптивных планов (программ) управления переходом ЦП из текущего в требуемое (выбранное) макросостояние. При этом рассматриваемые планы должны обеспечивать такое эволюционное развитие ЦП, при котором наряду с реализацией программ перехода из соответствующих макросостояний предусматривается одновременно и реализация программ устойчивого управления ЦП в промежуточных макросостояниях.

5. Заключение. В статье рассмотрены ИИТ, формирующие основу концепций Индустрии 4.0 и киберфизических систем для выявления основных направлений влияния указанных концепции и технологий на изменения принципов управления цепями поставок в ближайшем будущем. Вместе с тем анализ литературы и практических примеров Индустрии 4.0 показал, что наряду с развитием информационных и инженерных технологий Индустрии 4.0 в экономической части управления ТЛП и ПП отсутствует четкое понимание и описание бизнес-моделей, в рамках которых эти технологии могут применяться. Проанализированные примеры из практики позволяют сделать вывод, что технологии Индустрии 4.0 и киберфизических систем используются контекстно в виде лишь ограниченного набора той или иной функциональности (например, RFID, сенсоры или 3D принтеры). Основная же идея Индустрии 4.0 и КФС — создание самоорганизующихся и самоадаптирующихся динамических сетевых структур поставок на протяжении всего жизненного цикла изделий для реализации максимально гибкого индивидуального производства с затратами массового поточного производства — на данный момент реализована лишь в ряде исследовательских проектов.

К причинам подобного развития следует отнести преобладание инженерных разработок в области технологии Индустрии 4.0 и КФС и недостаточное участие экономистов и управленцев-практиков в данных проектах. Для менеджеров цепей поставок пока остается непонятным влияние инвестиций в технологии Индустрии 4.0 и киберфизических систем на повышение доходности бизнеса. Вместе с тем потенциал технологий Индустрии 4.0 и киберфизических систем наряду с уже существующими практическими реализациями не оставляет сомнений в том, что принцип «технология определяет организацию» и на этот раз подтвердится на практике. В связи со ска-

занным в статье ее авторами предложена оригинальная интерпретация сервис-ориентированного подхода применительно к задачам оценивания и выбора эффективных стратегий внедрения и использования ИИТ и ИТ в существующих и перспективных цепях поставок, базирующаяся на разрабатываемой авторами прикладной теории проактивного управления структурной динамикой ЦП.

В рамках предложенного подхода к формализации и решению одной из центральных проблем промышленного Интернета вещей, связанной с ситуационным структурно-функциональным синтезом как ЦП, так и соответствующих ИИТ и ИС удастся с единых методологических и методических позиций одновременно решить следующие пять основных задач комплексного планирования функционирования и модернизации рассматриваемого класса СКФС. Во-первых, удастся сформировать (синтезировать) конечный облик соответствующих модернизируемых СКФС. Во-вторых появляется возможность определить сроки, к которым должна завершиться указанная модернизация. В-третьих, можно определить технологии модернизации СКФС (в рамках которой на каком-то интервале времени будут одновременно функционировать элементы и подсистемы «старой» и «новой» СКФС, чтобы обеспечить устойчивость функционирования в целом СКФС). В-четвертых, появляется возможность определить программу перехода от «старой» и «новой» СКФС. В-пятых, удастся сформировать программы перепланирования либо коррекции исходной программы модернизации СКФС, если появятся непредусмотренные возмущающие воздействия.

При этом в рамках предложенного формального описания (см. формулу (2) и конкретные варианты ее реализации [30]) удастся описать как процессы совместного функционирования бизнес-систем (БС), связанных с функционированием ЦП, так и обеспечивающих их работу ИС, а также процессы модернизации ИС с учетом различных схем их финансирования и сценариев реализации возмущающих воздействий со стороны внешней среды. К настоящему времени, базируясь на сервис-ориентированном подходе к оцениванию эффективности ЦП, ИТ и ИС удастся, с одной стороны, зафиксировать в стоимостной форме объем услуг, оказываемых ИС соответствующей БС (ЦП), а с другой стороны, каждую услугу (сервис) можно непосредственно связать с конкретными информационными ресурсами, которые необходимы для ее реализации. Указанное обстоятельство очень важно при обосновании распределения денежных средств, выделяемых на приобретение и эксплуатацию аппаратно-программных средств современных ИС в рамках создаваемых структур производственного Интернета вещей.

Литература

1. *Orunfleh S., Tarafdar M.* Supply chain management practices — IT utilisation alignment: Impact on supply chain performance and firm performance // *International Journal of Business Information Systems*. 2015. vol. 18(4). pp. 364–389.
2. *Kumar A., Shankar R., Choudhary A., Thakur L.S.* A big data MapReduce framework for fault diagnosis in cloud-based manufacturing // *International Journal of Production Research*. 2016. vol. 54(23). pp. 7060–7073.
3. *Nayak A., Reyes Levalle R., Lee S., Nof S.Y.* Resource sharing in cyber-physical systems: modelling framework and case studies // *International Journal of Production Research*. 2016. vol. 54(23). pp. 6969–6983.
4. *Theorin A. et al.* An event-driven manufacturing information system architecture for Industry 4.0 // *International Journal of Production Research*. 2016. pp. 1297–1311.
5. *Ivanov D., Sokolov B.* Adaptive Supply Chain Management // London: Springer. 2010. 245 p.
6. *Ivanov D. et al.* A dynamic model and an algorithm for shortterm supply chain scheduling in the smart factory Industry 4.0 // *International Journal of Production Research*. 2016. vol. 54(2). pp. 386–402.
7. *Weyer S., Schmitt M., Ohmer M., Gorecky D.* Towards Industry 4.0 — Standardization as the crucial challenge for highly modular, multi-vendor production systems // *Proceedings of IFAC-PapersOnLine*. 2015. vol. 48. Issue 3. pp. 579–584.
8. *Zhong R.Y., Xu C., Chen C., Huang G.Q.* Big Data Analytics for Physical Internet-based intelligent manufacturing shop floors // *International Journal of Production Research*. 2015. pp. 2610–2621.
9. *Oesterreich T.D., Teuteberg F.* Understanding the implications of digitisation and automation in the context of Industry 4.0: A triangulation approach and elements of a research agenda for the construction industry // *Computers in Industry*. 2016. vol. 83. pp. 121–139.
10. *Велихов Е.П., Бетелин В.Б., Кушницренко А.Г.* Промышленность, инновации, образование и наука в России // М.: Наука. 2009. 141 с.
11. *Вонг Р., Перинг Т., Тенненхау Д.* Адаптивные и проактивные компьютерные системы // *Открытые системы*. 2003. № 7. С. 4–9.
12. *Добрынин А.П. и др.* Цифровая экономика — различные пути к эффективному применению технологий (BIM, PLM, CAD, IOT, Smart City, BIG DATA и другие) // *International Journal of Open Information Technologies*. 2016. Т. 4. № 1. С. 4–11.
13. *Черняк Л.* Платформа для облаков // *Открытые системы. СУБД*. 2014. № 8. С. 42–46.
14. *Черняк Л.* Гибридные облака: новый виток виртуализации // *Открытые системы. СУБД*. 2014. № 9. С. 11–15.
15. *Охтилев М.Ю., Соколов Б.В., Юсупов Р.М.* Интеллектуальные технологии мониторинга и управления структурной динамикой сложных технических объектов // М.: Наука. 2006. 410 с.
16. *Erol S. et al.* Tangible Industry 4.0: A Scenario-Based Approach to Learning for the Future of Production // *Procedia CIRP*. 2016. vol. 54. pp. 13–18.
17. *Hwang G., Lee J., Park J., Chang T.-W.* Developing performance measurement system for Internet of Things and smart factory environment // *International Journal of Production Research*. 2016. pp. 2590–2602.
18. *Kolberg D., Zühlke D.* Lean Automation enabled by Industry 4.0 Technologies // *Proceedings of IFAC-PapersOnLine*. 2015. vol. 48. Issue 3. pp. 1870–1875.
19. *Zhuge H.* Semantic linking through spaces for cyber-physical-socio intelligence: A methodology // *Artificial Intelligence*. 2015. vol. 175. Issues 5–6. pp. 988–1019.

20. *Намиот Д.Е.* Умные города 2016 // International Journal of Open Information Technologies. 2016. Т. 4. № 1. С. 1–3.
21. *Кульба К.С.* Автоматизация технической подготовки и управления производством на основе PLM — системы // М.: Машиностроение. 2008. 256 с.
22. *Соломенцев Ю.М., Волкова Г.Д.* Когнитивные технологии в конструкторско-технологической информатике // Вестник МГТУ «Станкин». 2008. № 4. С. 132–135.
23. *Хищенко В.Е.* Самоорганизация: элементы теории и социальные приложения // М.: КомКнига. 2005. 224 с.
24. *Шваб К.* Четвертая промышленная революция // М.: Эксмо. 2018. 278 с.
25. *Бир С.* Мозг фирмы // М.: УРСС. 2005. 416 с.
26. *Верзилин Д.Н., Соколов Б.В., Юсупов Р.М.* Неокибернетика — вчера, сегодня, завтра // Материалы конференции «Информационные технологии в управлении» (ИТУ-2014). СПб.: ОАО «Концерн «ЦНИИ «Электроприбор». 2014. С. 192–201.
27. *Крылов С.М.* Неокибернетика: Алгоритмы, математика эволюции и технологии будущего // М.: Издательство ЛКИ. 2008. 288 с.
28. *Hübner H.* Neocybernetics in Biological Systems. Helsinki University of Technology, Control Engineering Laboratory. Report 151. 2006. 273 p.
29. *Юсупов Р.М., Соколов Б.В.* Проблемы развития кибернетики и информатики на современном этапе // Сб. «Кибернетика и информатика». 2006. С. 6–21.
30. Официальный сайт лаборатории информационных технологий в системном анализе и моделировании (ЛИТСАМ). URL: <http://litsam.ru> (дата обращения: 01.08.2016).
31. *Васильев С.Н.* От классических задач регулирования к интеллектуальному управлению // Теория и системы управления. 2001. № 1. С. 5–22; № 2. С. 5–21.
32. *Скрипкин К.Г.* Экономическая эффективность информационных систем // М.: ДМК Пресс. 2002. 256 с.
33. *Юсупов Р.М., Мусаев А.А.* Особенности оценивания эффективности информационных систем и технологий // Труды СПИИРАН. 2017. № 2(51). С. 5–34.

Иванов Дмитрий Александрович — д-р экон. наук, профессор, заведующий кафедрой управления международными цепями поставок, Берлинская школа экономики и права. Область научных интересов: структурная динамика цепей поставок, риски в цепях поставок, динамическое планирование расписаний. Число научных публикаций — 300. dmitri.ivanov@mail.ru; ул. Баденштрассе, 50-51, 10825, Берлин, Германия; р.т.: +49(0)3030877-1155.

Иванова Марина Александровна — к-т экон. наук, научный сотрудник отдела бизнеса и экономики, Берлинская школа экономики и права. Область научных интересов: индустрия 4.0, планирование и оперативное управление в цепях поставок. Число научных публикаций — 40. marina.ivanova@hwr-berlin.de; ул. Баденштрассе, 50-51, 10825, Берлин, Германия; р.т.: +49(0)3030877-1155.

Поддержка исследований. Результаты исследований, представленные в 1-3 разделах статьи, проводились при частичной финансовой поддержке грантов РФФИ (№№ 16-07-00779, 16-08-00510, 16-08-01277, 17-08-00797, 17-06-00108, 17-01-00139, 17-20-01214, 17-29-07073- оф-м, 18-07-01272, 18-08-01505), госзадания Министерства образования и науки РФ №2.3135.2017/4.6, результаты исследований, представленные в 4 и 5 разделах статьи, проводились в рамках финансирования по бюджетной теме №№0073–2018–0003.

D.A. IVANOV, M.A. IVANOVA, B.V. SOKOLOV
**ANALYSIS OF TRANSFORMATION TRENDS IN ENTERPRISE
MANAGEMENT PRINCIPLES IN THE ERA OF INDUSTRY 4.0
TECHNOLOGY**

Ivanov D.A., Ivanova M.A., Sokolov B.V. Analysis of Transformation Trends in Enterprise Management Principles in the Era of Industry 4.0 Technology.

Abstract. This study is devoted to the existing and future information technology that forms a base for concepts of Industry 4.0 and cyber-physical systems using examples of firms that apply this technology. Examples include manufacturing and transportation companies. The objective of this paper is to analyse the impact of intelligent information technologies in Industry 4.0 and cyber-physical systems on the enterprise management principles in near future. The transformation trends are illustrated on case-study examples. The literature and case-study analysis shows that enterprise management in manufacturing and transportation lacks understanding of Industry 4.0 and cyber-physical systems in terms of application to concrete business models. The reviews cases lead to the conclusion that the Industry 4.0 and cyber-physical technology applications are rather contextual and engineering-oriented. The implementation of the main principle of the Industry 4.0 and cyber-physical technology, i.e., the self-organisation and self-adaption of dynamic networked structures along the entire product lifecycle to ensure the flexible individual manufacturing at the mass production costs, has been yet applied very restrictively and mostly in research projects. The paper presents an original interpretation of the service-oriented approach to modeling and solution of the information technology assessment and selection problems to facilitate a broader application in firms. This approach is based on the structural dynamics control theory developed by the authors.

Keywords: Cyber-Physical system, Industry 4.0, intelligent information technology, supply chain, manufacturing, effectiveness of information technology utilization.

Ivanov Dmitry Alexandrovich — Dr. Habil., professor, head of international supply chain management department, Berlin School of Economics and Law (BSEL). Research interests: structural dynamics of supply chains, risks in supply chains, dynamic scheduling of schedules. The number of publications — 300. dmitri.ivanov@mail.ru; 50, Badensche Str., 10825, Berlin, Germany; office phone: +49(0)3030877-1155.

Ivanova Marina Alexandrovna — Ph.D., research associate of department of business and economics, Berlin School of Economics and Law (BSEL). Research interests: industry 4.0 systems, supply chain planning and scheduling. The number of publications — 40. marina.ivanova@hwr-berlin.de; 50, Badensche Str., 10825, Berlin, Germany; office phone: +49(0)3030877-1155.

Sokolov Boris Vladimirovich — Ph.D., Dr. Sci., professor, Honored scientist of Russian Federation, head of information technologies in systems analysis and modeling laboratory, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: development of research fundamentals for the control theory by structural dynamics of complex organizational-technical systems. The number of publications — 560. sokol@iias.spb.su; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7-812-328-3311.

Acknowledgements. The research described in 1-3 paper sections is partially supported by the Russian Foundation for Basic Research (grants 16-07-00779, 16-08-00510, 16-08-01277, 17-08-00797, 17-06-00108, 17-01-00139, 17-20-01214, 17-29-07073-ofi-i, 18-07-01272, 18-08-01505)), state order of the Ministry of Education and Science of the Russian Federation №2.3135.2017/4.6, the research described in 4 paper section is partially supported by the state research №№0073–2018–0003.

References

1. Qrunfleh S., Tarafdar M. Supply chain management practices — IT utilisation alignment: Impact on supply chain performance and firm performance. *International Journal of Business Information Systems*. 2015. vol. 18(4). pp. 364–389.
2. Kumar A., Shankar R., Choudhary A., Thakur L.S. A big data MapReduce framework for fault diagnosis in cloud-based manufacturing. *International Journal of Production Research*. 2016. vol. 54(23). pp. 7060–7073.
3. Nayak A., Reyes Levalle R., Lee S., Nof S.Y. Resource sharing in cyber-physical systems: modelling framework and case studies. *International Journal of Production Research*. 2016. vol. 54(23). pp. 6969–6983.
4. Theorin A. et al. An event-driven manufacturing information system architecture for Industry 4.0. *International Journal of Production Research*. 2016. pp. 1297–1311.
5. Ivanov D., Sokolov B. Adaptive Supply Chain Management. London: Springer. 2010. 245 p.
6. Ivanov D. et al. A dynamic model and an algorithm for shortterm supply chain scheduling in the smart factory Industry 4.0. *International Journal of Production Research*. 2016. vol. 54(2). pp. 386–402.
7. Weyer S., Schmitt M., Ohmer M., Gorecky D. Towards Industry 4.0 — Standardization as the crucial challenge for highly modular, multi-vendor production systems. *Proceedings of IFAC-PapersOnLine*. 2015. vol. 48. Issue 3. pp. 579–584.
8. Zhong R.Y., Xu C., Chen C., Huang G.Q. Big Data Analytics for Physical Internet-based intelligent manufacturing shop floors. *International Journal of Production Research*. 2015. pp. 2610–2621.
9. Oesterreich T.D., Teuteberg F. Understanding the implications of digitisation and automation in the context of Industry 4.0: A triangulation approach and elements of a research agenda for the construction industry. *Computers in Industry*. 2016. vol. 83. pp. 121–139.
10. Velihov E.P., Betelin V.B., Kushnirenko A.G. *Promyshlennost', innovacii, obrazovanie i nauka v Rossi* [The industry, innovations, science and education in Russia]. M.: Nauka. 2009. 141 p. (In Russ.).
11. Vong R., Pering T., Tennenhou D. [Adaptive and pro-active computer systems]. *Otkrytye sistemy – Open systems*. 2003. vol. 7. pp. 4–9. (In Russ.).
12. Dobrynin A.P. et al. [Digital economy — various ways to effective use of technologies (BIM, PLM, CAD, IOT, Smart City, BIG DATA et al.)]. *International Journal of Open Information Technologies*. 2016. vol. 4. no. 1. pp. 4–11. (In Russ.).
13. Chernjak L. [Platform for clouds: fabric computing]. *Otkrytye sistemy. SUBD – Open systems*. 2014. vol. 8. pp. 42–46.
14. Chernjak L. [Hybrid clouds: a new round of virtualization]. *Otkrytye sistemy. SUBD – Open systems*. 2014. vol. 9. pp. 11–15.
15. Ohtilev M.Yu., Sokolov B.V., Yusupov R.M. *Intellektual'nye tehnologii monitoringa i upravleniya strukturnoj dinamikoj slozhnykh tehnikeskikh ob'ektov* [Intellectual technologies of monitoring and management of complex technical objects structural dynamics] M.: Nauka. 2006. 410 p. (In Russ.).
16. Erol S. et al. Tangible Industry 4.0: A Scenario-Based Approach to Learning for the Future of Production. *Procedia CIRP*. 2016. vol. 54. pp. 13–18.

17. Hwang G., Lee J., Park J., Chang T.-W. Developing performance measurement system for Internet of Things and smart factory environment. *International Journal of Production Research*. 2016. pp. 2590–2602.
18. Kolberg D., Detlef Zühlke. Lean Automation enabled by Industry 4.0 Technologies. Proceedings of IFAC-PapersOnLine. 2015. vol. 48. Issue 3. pp. 1870–1875.
19. Zhuge H. Semantic linking through spaces for cyber-physical-socio intelligence: A methodology. *Artificial Intelligence*. 2015. vol. 175. Issues 5–6. pp. 988–1019.
20. Namiot D.E. [Smart cities 2016]. *International Journal of Open Information Technologies*. 2016. vol. 4. no. 1. pp. 1–3. (In Russ.).
21. Kul'ga K.S. *Avtomatizacija tehničkoj podgotovki i upravlenija proizvodstvom na osnove PLM — sistemy* [Automation of technical training and production management on the basis of PLM — systems]. M.: Mashinostroenie. 2008. 256 p. (In Russ.).
22. Solomencev Ju.M., Volkova G.D. [Cognitive technologies in design-technology informatics]. *Vestnik MGTU «Stankin» – Vestnik MSTU «Stankin»*. 2008. vol. 4. pp. 32–135. (In Russ.).
23. Hicenko V.E. *Samoorganizacija: jelementy teorii i social'nye prilozhenija* [Self-organization: elements of the theory and social applications]. M.: KomKniga. 2005. 224 p. (In Russ.).
24. Schwab K. *Chevertaya promyshlennaya revolyuciya*. [Fourth industrial revolution]. M.: Eksmo. 2018. 278 p. (In Russ.).
25. Bir S. *Mozg firmy* [Firm's brain]. M.:URS. 2005. (In Russ.).
26. Verzilin D.N., Sokolov B.V., Yusupov R.M. [Neocybernetics — yesterday, nowadays, tomorrow]. *Materialy konferencii «Informacionnye tehnologii v upravlenii» (ITU-2014)* [Information Technologies in Control: Collected papers]. SPb.: JSC «Concern CRI «Elektropribor». 2014. pp. 192–201. (In Russ.).
27. Krylov S.M. *Neokibernetika. Algorytmy, matematika evoljutsii I tehnologii budushego* [Neocybernetics. Algorithms, mathematics of evolution and technologies of the future]. M.: Izdatelstvo LKI. 2008. 288 p. (In Russ.).
28. Hyötyniemi H. Neocybernetics in Biological Systems. Helsinki University of Technology, Control Engineering Laboratory. Report 151. 2006. 273 p.
29. Yusupov R.M., Sokolov B.V. [Cybernetics development problems and computer science at the present stage]. *Problemy razvitiya kibernetiki i informatiki na sovremennom etape* [Cybernetics and informatics: Collected papers]. 2006. pp. 6–21. (In Russ.).
30. Oficial'nyj sajt laboratorii informacionnyh tekhnologij v sistemnom analize i modelirovanii (LITSAM) [Official web site of information technologies in systems analysis and modeling laboratory]. Available at: <http://litsam.ru> (accessed: 01.08.2016).
31. Vasil'ev S.N. [From classical problems of regulation to intellectual management] *Teorija i sistemy upravlenija*. 2001. vol. 1. pp. 5–22; vol. 2. pp. 5–21. (In Russ.).
32. Skripkin K.G. *Jekonomicheskaja jeffektivnost' informacionnyh system* [Economic efficiency of information systems]. M.: DMK Press. 2002. 256 p. (In Russ.).
33. Yusupov R.M., Musaev A.A. [Efficiency of Information Systems and Technologies: Features of Estimation]. *Trudy SPIIRAN – SPIIRAS Proceeding*. 2017. vol. 2(51). pp. 5–34. (In Russ.).

А.В. Козачок, С.А. Копылов, Р.В. Мещеряков, О.О. Евсютин,
Л.М. Туан
**ПОДХОД К ИЗВЛЕЧЕНИЮ РОБАСТНОГО ВОДЯНОГО ЗНАКА
ИЗ ИЗОБРАЖЕНИЙ, СОДЕРЖАЩИХ ТЕКСТ**

Козачок А.В., Копылов С.А., Мещеряков Р.В., Евсютин О.О., Туан Л.М. Подход к извлечению робастного водяного знака из изображений, содержащих текст.

Аннотация. Представлен подход к извлечению робастного водяного знака из изображений, содержащих текст. Извлечение данных основано на разработанном подходе к внедрению в текстовые данные робастного водяного знака, отличающегося устойчивостью к преобразованию текстовых данных в формат изображения. Проведен сравнительный анализ существующих подходов к стеганографическому внедрению информации в текстовые данные, определены их достоинства и недостатки. Обоснован выбор группы методов стеганографического внедрения информации, основанных на форматировании текста. В качестве алгоритма встраивания выбран подход на основе изменения величины межстрочных интервалов. Приведены блок-схема и описание разработанного алгоритма встраивания информации в текстовые данные. Осуществлена экспериментальная оценка емкости встраивания и границ перцептивной невидимости встроенных данных. На основе существующих ограничений разработан подход к извлечению встроенной информации из изображений, содержащих робастный водяной знак. Базовым в процедуре извлечения встроенной информации выбрано преобразование Радона, позволяющее извлекать значения величин межстрочных интервалов. Для выделения значений битов встроенной информации использован подход на основе разделения смеси нормальных распределений, поскольку полученное распределение является бимодальным. Экспериментально установлены границы применимости, а также дана оценка робастности разработанного подхода встраивания к осуществлению различных преобразований. Определены следующие параметры робастности разработанного подхода к осуществлению преобразований: поворот изображения, содержащего встроенные данные на любой угол, масштабирование изображения с множителем масштабирования, не превышающим 1,5; преобразование в любой формат растрового изображения; применение медианного фильтра к изображению с пределом ядра свертки не более 9 пикселей, гауссовского фильтра размытия — с пределом показателя размытия не более 8 пикселей и усредненного фильтра с пределом ядра свертки не более 5 пикселей.

Ключевые слова: защита информации, текстовая стеганография, цифровой водяной знак, преобразование Радона, разделение смеси распределений.

1. Введение. Стремительное развитие «Интернета вещей» привело к увеличению устройств, имеющих доступ к глобальной сети [1, 2]. С ростом объемов данных и степени разнородности устройств в значительной степени возросло и количество инцидентов информационной безопасности, связанных с нарушением авторских прав владельцев текстовой информации и подменой их содержимого. По данным аналитической компании Solar Security во втором полугодии 2017 года наблюдался стабильный рост инцидентов, связанный с утечками конфиденциальных

данных и компрометацией учетных записей [3]. Одним из типов данных, подвергшихся компрометации являются текстовые данные, содержащие конфиденциальную информацию. Повышению защищенности текстовых данных, представленных в электронном виде, посвящено множество работ [4-8]. В то же время задача по защите текстовых данных, представленных в напечатанном виде, остается наиболее актуальным направлением исследований. Данная задача может быть решена посредством скрытого внедрения в текстовые данные идентифицирующей информации, которая может характеризовать владельца данных, сами данные или содержать другую метаинформацию [9].

В настоящее время для встраивания информации в текстовые данные используется технология внедрения цифровых водяных знаков (ЦВЗ). ЦВЗ могут быть использованы для защиты авторских прав, обнаружения подмены данных, проверки подлинности или вставки дополнительной информации. ЦВЗ представляет собой видимый или невидимый знак (информационную последовательность), встраиваемый в исходные данные. В аспекте работы классификация ЦВЗ рассматривается по двум направлениям (рисунок 1) [10].



Рис. 1. Классификация ЦВЗ

Примером внедрения видимого ЦВЗ в текстовые данные является встраивание электронной подписи. Недостаток использования видимых ЦВЗ заключается в возможности удаления или модификации встроенных данных, что, в свою очередь, не позволяет правильно идентифицировать владельца либо определить подмену области исходных данных. Невидимые ЦВЗ лишены данного недостатка, однако в процессе их извлечения необходимы дополнительные этапы, направленные на определение наличия водяного знака и восстановление данных, которые они содержат.

По критерию надежности ЦВЗ могут быть представлены следующими группами:

- робастный водяной знак (РВЗ) — обеспечивает устойчивость к внесению искажений и осуществлению различных преобразований;
- хрупкий водяной знак — разрушается при внесении искажений и осуществлении различных преобразований;
- полухрупкий водяной знак — разрушается отдельным типом искажений (преобразований), однако обеспечивает устойчивость к другим типам искажений (преобразований) [11].

В случае применения преобразований и внесения искажений в хрупкие и полухрупкие водяные знаки невозможно осуществить извлечение встроенной информации ввиду разрушения водяного знака. Указанная особенность не позволяет использовать данные ЦВЗ в процессе защиты авторских прав и проверки подлинности текстовых данных. В свою очередь РВЗ способен обеспечить свойство инвариантности встроенных данных в случае применения преобразований и внесения искажений.

Невидимые робастные водяные знаки характеризуются следующими параметрами [12-14]:

- емкость встраивания (полезная нагрузка) — количество информации, которая может быть встроена (внедрена) в контейнер. В качестве контейнера могут выступать типы данных, пригодные для встраивания информации;
- невидимость (перцепционная прозрачность) — качественная характеристика, отражающая степень искажения контейнера встраиваемыми данными. Данная характеристика основана на перцептивном восприятии человека;
- необнаруживаемость (сложность обнаружения) — качественная характеристика, отражающая степень искажения статистических характеристик контейнера, не связанных с перцептивным восприятием человека;
- робастность — способность встроенных данных сохранять свойство инвариантности после осуществления различных преобразований над контейнером, подмены или удаления встроенных данных;
- извлекаемость — способность правильного извлечения встроенных данных из контейнера.

Для обеспечения невидимости и необнаруживаемости РВЗ в процессе встраивания в текстовые данные широкое распространение получили стеганографические методы, позволяющие осуществлять скрытое встраивание дополнительной информации в цифровые объекты за счет внесения некоторых изменений в элементы данных, составляющие цифровой объект [15, 16]. Для определения наиболее подходящего алгоритма

стеганографического встраивания информации в текстовые данные необходимо провести сравнительный анализ существующих методов текстовой стеганографии.

2. Обоснование выбора метода встраивания информации в текстовые данные. Методы текстовой стеганографии принято разделять на следующие группы [17, 18]:

- методы, основанные на создании последовательности с учетом статистических характеристик исходного текста;
- лингвистические методы;
- методы, основанные на форматировании текста.

Методы текстовой стеганографии, основанные на создании последовательности с учетом статистических характеристик исходного текста, могут быть разделены на методы, генерирующие случайную последовательность символов, и методы, основанные на учете статистических характеристик исходного текста. В первом случае генерируется случайная последовательность символов, которая встраивается в текст. Недостаток данного метода заключается в отсутствии скрытости внедряемых данных по отношению к семантическому и визуальному анализу текста. Второй метод лишен данного недостатка. Он получил название текстовая мимикрия. При использовании текстовой мимикрии генерируется осмысленный текст на основе синтаксиса, описанного в контекстно-свободной грамматике, и встраивается информация за счет выбора из нее определенных фраз и слов [19].

Методы данной группы позволяют сохранить внедренную информацию как внутри электронного документа, так и в случае вывода его на печать. Данная группа методов характеризуется большой емкостью встраивания, наличием стойкости к операциям повторного набора текста и обработке текста средствами оптического распознавания символов. Недостатком данной группы методов является возможность обнаружения встроенных данных в процессе семантического анализа текста.

Лингвистические методы текстовой стеганографии для встраивания информации используют языковые свойства и структуру текста. Лингвистические методы сочетают в себе синтаксические, пунктуационные добавления (добавление знаков пунктуации в строго определенные места в тексте) и семантические методы (замена определенных слов в тексте на синонимы, использование кодовых таблиц, словарей сокращений и аббревиатур) [20-23] встраивания информации в исходный текст. К недостаткам данного метода относится небольшая емкость встраивания данных и высокая сложность восстановления.

Методы данной группы применяются для встраивания информации во все виды текстов. Данная группа методов обеспечивает стойкость к операциям повторного набора текста и обработки текста средствами оптического распознавания символов. В свою очередь данная группа, как и предыдущая, характеризуется невозможностью применения к официальным текстовым документам, подготовленным для печати и публикации.

К методам текстовой стеганографии, основанным на форматировании текста, относятся методы, осуществляющие форматирование текста и текстового пространства. Под форматированием текста и текстового пространства подразумевается изменение структуры текста, а также формы и представления шрифтов. К данной группе относятся следующие подходы:

- основанные на использовании пространства интервалов и отступов (сдвиг линии текста, сдвиг слова текста, сдвиг символа внутри слова, сдвиг абзацного отступа);
- основанные на использовании открытого пространства (внесение дополнительных пробелов);
- основанные на кодировании признаков (замена символов).

В алгоритмах встраивания, реализующих изменение положения линии текста (рисунок 2), кодирование информации осуществляется посредством вертикального сдвига линии текста на определенное расстояние [24, 25].

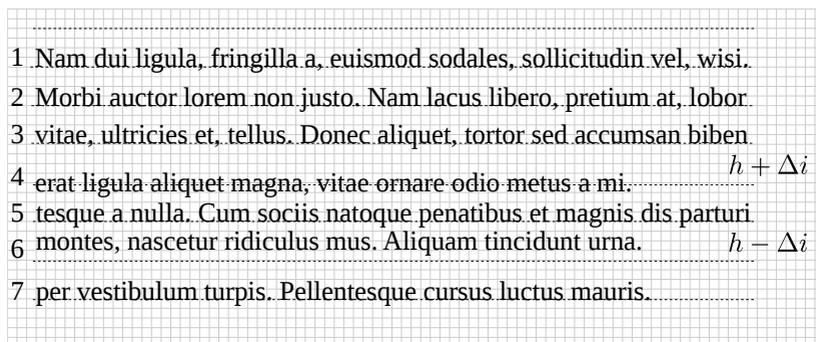


Рис. 2. Сдвиг линии текста

Алгоритмы встраивания, основанные на изменении положения слов [26], реализуют кодирование информации посредством горизонтального сдвига слов (влево или вправо на определенное расстояние) либо увеличения (уменьшения) интервала как между словами, так и между символами внутри слова.

Подход к встраиванию, основанный на сдвиге символа внутри слова, реализуется посредством кодирования информации за счет модуляции локального индекса символов текста [27] (рисунок 3). В отличие от рассмотренных ранее алгоритмов встраивания положение символа кодируется двумя битами.

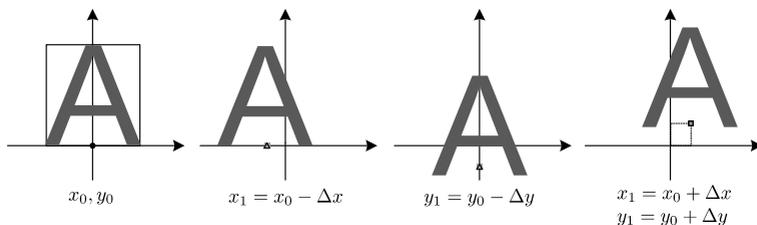


Рис. 3. Изменение положения символа

В алгоритмах встраивания, основанных на изменении значения отступа и параметров текста, кодирование информации осуществляется посредством изменения величины отступа (в том числе абзацного), табуляции, значения полей, колонтитулов и выравнивания текста.

Особенность подходов встраивания, основанных на использовании пространства интервалов и отступов, заключается в применении небольших значений величин отступа и сдвига для обеспечения перцептивной невидимости встраиваемой информации.

Вторая группа подходов встраивания, основанных на использовании открытого пространства, реализует схему внесения дополнительных пробелов в исходный текст. В качестве свободного (пустого) пространства может выступать: пустое пространство в конце каждой строки, пробелы между словами в предложении и пробелы, следующие за пунктуационными знаками [28].

Алгоритмы встраивания, основанные на внесении в текст дополнительных пробелов [12], осуществляют кодирование информации посредством сопоставления единичному дополнительному пробелу значения «1», а двойному пробелу — «0».

Алгоритмы встраивания, основанные на внедрении дополнительных пробелов в конец каждой строки, кодируют один бит встраиваемой информации двумя дополнительными пробелами в конце строки, два бита информации — четырьмя пробелами и так далее в зависимости от достижимой емкости встраивания [26].

В подходах к встраиванию, основанных на кодировании признаков, встраивание информации в исходный текст осуществляется посредством

замены одного и более символов текста. Для этого исходный текст анализируется и производится выборка символов, пригодных для встраивания. Такими символами могут быть символы, содержащие диакритические знаки, кодирование информации в которых производится путем их смещения (удаления), или символы, содержащие выносные линии, кодирование информации в которых осуществляется посредством изменения их длины и ширины [29].

Помимо изменения формы и размера символов алгоритмы встраивания, основанные на кодировании признаков, могут использовать особенности стандартов кодирования. Алгоритм встраивания, основанный на стандарте кодирования Unicode [30], осуществляет встраивание информации посредством замены одного из 13 символов английского алфавита, имеющих аналогичное представление в других алфавитах, на один из трех символов-аналогов в формате Unicode.

Кодирование признаков в алгоритмах встраивания может быть основано на сходстве некоторых типов шрифтов и реализовано посредством замены символов одного типа шрифта на другой. Для реализации алгоритма встраивания информации необходимо определить набор шрифтов схожих с исходным, составить таблицу замены и осуществить внедрение информации.

Методы текстовой стеганографии, основанные на форматировании текста, могут применяться в случаях, когда необходимо сохранение внедренной информации при печати электронного документа. Данная группа методов обеспечивает высокий уровень невидимости встраиваемых данных по отношению к визуальному и семантическому анализу текста. Недосток, присущий всем подходам данной группы, заключается в отсутствии стойкости к операциям повторного набора текста и обработки текста средствами оптического распознавания символов. Кроме того, рассмотренные подходы характеризуются небольшой емкостью встраивания данных.

Указанные достоинства и недостатки методов текстовой стеганографии позволяют сделать вывод о применимости методов, основанных на форматировании текста, для внедрения РВЗ в текстовые данные, подготовленные для печати. Особенность текстовых данных, подготовленных к печати, заключается в невозможности изменения семантической и синтаксической структуры текста. Данная особенность накладывает ограничение на использование лингвистических методов и методов, основанных на создании последовательностей, в процессе внедрения РВЗ в текстовые данные.

Использование в процессе внедрения РВЗ подходов встраивания, основанных на внесении дополнительных пробелов и кодировании признаков в текстовые данные, подготовленные для печати, накладывает дополнительные требования на этапы обнаружения и извлечения встроенных данных. В процессе извлечения встроенных данных необходимо точно определить количество дополнительно встроенных пробелов или расположение измененных символов внутри текста, что является трудновыполнимой задачей для существующих систем оптического распознавания символов. Для решения данной задачи наиболее подходящими являются следующие подходы к встраиванию: подход, основанный на изменении величины межстрочного интервала, и подход, основанный на изменении величины межсимвольного интервала.

Подход к внедрению информации, основанный на изменении межсимвольного интервала, применим при наборе текста с использованием моноширинных шрифтов, что не соответствует существующим правилам подготовки официальных документов и документов, подготовленных к печати или опубликованию. Подход к внедрению данных, основанный на изменении межстрочного интервала (положении линий текста) лишен указанного недостатка и может быть использован в процессе стеганографического встраивания РВЗ в текстовые данные, подготовленные для печати.

Таким образом, указанные ограничения позволяют выбрать в качестве подхода к встраиванию информации в текстовые данные разработанный алгоритм стеганографического внедрения РВЗ, основанный на изменении величины межстрочного интервала.

3. Подход к встраиванию информации в текстовые данные.

Вариант встраивания информации на основе изменения межстрочного интервала (рисунок 2) позволяет кодировать 1 бит информации с помощью изменения межстрочного интервала текста на величину Δ . В разработанном алгоритме встраивание информации реализуется следующим образом:

- увеличение величины межстрочного интервала на установленное значение Δ между соседними строками текста интерпретируется как значение «1»;
- отсутствие изменений в величине межстрочного интервала между соседними строками текста интерпретируется как значение «0».

Блок-схема алгоритма стеганографического внедрения информации в текстовые данные представлена в листинге 1.

Исходными данными для работы алгоритма является текстовый документ TD_0 (документ-контейнер), содержащий текстовые данные.

Data: Текстовый документ TD_0 , встраиваемая информация I

Result: Подписанный текстовый документ TD_s

```

1  $Len \leftarrow \text{GetLength}(I)$ 
2  $N \leftarrow \text{CountLines}(TD_0)$ 
3 if  $N > (Len + 1)$  then
4   for  $i \leftarrow 0$  to  $(N - 2)$  do
5      $j \leftarrow i \bmod Len$ 
6     if  $I_j = 1$  then
7        $TD_0 \leftarrow \text{Embed}(TD_0, i)$ 
8    $TD_s \leftarrow TD_0$ 
9 return  $TD_s$ 

```

Листинг 1. Алгоритм внедрения информации в текстовый документ

Допускается использование следующих форматов текстовых документов:

- RTF (Rich Text Format) — спецификация [31];
- PDF (Portable Document Format) — стандарт [32, 33].

Первым этапом работы алгоритма стеганографического внедрения информации (шаги 1-2) является определение размера встраиваемой информации и достижимой емкости встраивания функцией *GetLength* и *CountLines* соответственно.

На втором этапе (шаг 3) определяется возможность встраивания информации в исходные текстовые данные. На этапе кодирования встраиваемой информации (шаги 4-7) кодовая последовательность встраивается в исходный текстовый документ следующим образом:

- внедрение информации осуществляется с первой (i) до $(N - 2)$ строки, где N определяет достижимую емкость встраивания;
- в i строке считывается символ кодовой комбинации I ;
- если считанный символ соответствует «1», то межстрочный интервал между i и $i + 1$ строками увеличивается на величину Δ .

По окончании работы алгоритма стеганографического внедрения информации формируется подписанный текстовый документ TD_s , содержащий встраиваемую информацию I .

Для определения возможности реализации указанного подхода встраивания, достижимой емкости встраивания и установления границ перцептивной невидимости встроенной информации в текстовые данные была проведена их экспериментальная оценка разработанного алгоритма.

4. Оценка емкости встраивания и перцептивной невидимости подхода на основе изменения межстрочного интервала. Проведенная группа экспериментов позволила определить величину предельно до-

стижимой емкости встраивания, установить зависимость достижимой емкости встраивания от размера и гарнитуры используемого шрифта и изменения величины межстрочного интервала. Кроме того, проведена оценка перцептивной невидимости встроенных данных и определены пороговые значения изменения величины межстрочного интервала.

Исходными данными являлся текст, созданный в системе компьютерной верстки \LaTeX , размером не менее 10 страниц со следующими размерами полей:

- верхнее, нижнее — 20 мм;
- левое — 30 мм;
- правое — 10 мм.

Предельно достижимая емкость встраивания алгоритма характеризуется количеством межстрочных интервалов текста на одной странице. Для установления зависимости между емкостью встраивания, параметрами шрифта и величиной межстрочного интервала были выбраны следующие варьируемые параметры:

- межстрочный интервал: одинарный, полуторный, двойной.
- кегль шрифта: 10 пт, 12 пт и 14 пт.
- гарнитура используемого шрифта:

1. Шрифт с засечками (serif) — Computer Modern Roman (аналог Times New Roman).

2. Шрифт без засечек (sans serif) — Computer Modern Sans Serif (аналог Arial).

3. Моноширинный шрифт (monospace) — Computer Modern Typewriter (аналог Courier New).

Полученные значения предельно достижимой емкости встраивания представлены в таблице 1.

Таблица 1. Зависимость емкости встраивания от изменения межстрочного интервала и кегля шрифта

Кегль шрифта	Величина межстрочного интервала	Предельно достижимая емкость встраивания
10	1	60
10	1,5	40
10	2	30
12	1	49
12	1,5	33
12	2	24
14	1	42
14	1,5	28
14	2	21

Анализ полученных результатов позволяет сделать вывод о зависимости емкости встраивания от величины межстрочного интервала и размера кегля шрифта, а также об отсутствии зависимости от гарнитуры используемого шрифта. Максимально достижимое значение емкости встраивания составляет 60 бит (при использовании шрифта с кеглем 10 пт и величине межстрочного интервала 1), минимальное — 21 бит (при использовании шрифта с кеглем 14 пт и величине межстрочного интервала 2).

В ходе проведения второй группы экспериментов была произведена оценка перцептивной невидимости встраиваемых данных. Кодирование информации реализовано следующим образом:

- величина межстрочного интервала после каждой нечетной строки текста не изменяется и интерпретируется как значение «0»;
- величина межстрочного интервала после каждой четной строки текста увеличивается на установленное значение и интерпретируется как значение «1».

В данном эксперименте использовался один абзац текста с гарнитурой Computer Modern Roman и кеглем 14 пт. Для определения порога перцептивной невидимости использовались следующие варьируемые параметры:

1. Шрифт Computer Modern Roman 14 пт (4,94 мм) с межстрочным интервалом, равным 1 (4,94 мм):

- увеличение межстрочного интервала до 1,25 (6,17 мм) с шагом 0,05 (0,245 мм);
- уменьшение межстрочного интервала до 0,75 (3,71 мм) с шагом 0,05 (0,245 мм).

2. Шрифт Computer Modern Roman 14 пт (4,94 мм) с межстрочным интервалом, равным 1,5 (7,41 мм):

- увеличение межстрочного интервала до 1,25 (8,64 мм) с шагом 0,05 (0,245 мм);
- уменьшение межстрочного интервала до 0,75 (6,17 мм) с шагом 0,05 (0,245 мм).

В результате визуального анализа полученных данных экспертным путем был установлен порог перцептивной невидимости встроенных данных:

- 0,85-1,15 (4,20-5,68 мм) — для одинарного межстрочного интервала;
- 1,35-1,65 (6,67-8,15 мм) — для полуторного межстрочного интервала.

Полученные результаты емкости встраивания и перцептивной невидимости позволяют использовать разработанный подход к внедрению РВЗ в текстовые данные для решения задачи по защите текстовых данных, подготовленных к печати.

Процесс извлечения встроенных данных накладывает дополнительные требования по точности обнаружения встроенного РВЗ и определения величин межстрочных интервалов. Кроме того, извлечение встроенного РВЗ происходит из изображения, содержащего текст, что требует введения в процесс извлечения этапа предварительной обработки изображения.

5. Подход к извлечению информации из изображений со встроенным РВЗ. Для реализации указанных требований и обеспечения достоверности извлечения встроенных данных был разработан подход к извлечению информации из изображений со встроенным РВЗ, состоящий из следующих этапов:

- предварительная обработка изображения;
- обнаружение линий текста;
- определение массива значений межстрочных интервалов;
- декодирование встроенной информации.

Отличительной особенностью разработанного подхода является применение нормального преобразования Радона для обнаружения линий текста.

5.1. Преобразование Радона. Преобразование Радона $\check{g}(p, \tau)$ непрерывной двумерной функции $f(x, y)$ определяется посредством суммирования (интегрирования) значений f вдоль наклонных линий [34]. Положение линии определяется параметрами наклона p и смещения линии τ :

$$\check{g}(p, \tau) = \int_{-\infty}^{\infty} f(x, p \cdot x + \tau) dx. \quad (1)$$

К свойствам преобразования Радона относятся:

- обратимость — возможность восстановить исходную функцию по ее преобразованию;
- линейность — преобразование Радона взвешенной суммы функций равно взвешенной сумме отдельно преобразованных функций;
- смещение функции — наклон линии не может быть изменен посредством применения преобразования;
- масштабируемость.

Выражение (1) представляет собой одну из нескольких форм записи преобразования. В общем виде преобразование Радона описывается следующим образом:

$$\check{g}(\xi_0, \xi_1, \xi_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(\xi_0 - \xi_1 \cdot x - \xi_2 \cdot y) dx dy. \quad (2)$$

Особенность прямого преобразования Радона заключается в преобразовании каждой линии изображения в определенным образом расположенные пики, соответствующие параметрам этих линий. Обратное преобразование Радона позволяет восстановить из полученных пиков исходные линии. Из свойства обратимости преобразования Радона следует, что преобразование Радона, примененное к точке, трансформирует ее в линию, а примененное к линии — в точку.

Если базовая функция $f(x, y)$ не имеет предпочтительной ориентации, то уравнение прямой может быть представлено в нормальном виде:

$$\rho = x \cdot \cos \theta + y \cdot \sin \theta. \quad (3)$$

Подставляя полученные значения $(\xi_0, \xi_1, \xi_2) = (\rho, \cos \theta, \sin \theta)$ в выражение (2), получим нормальное преобразование Радона:

$$\check{g}(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(\rho - x \cdot \cos \theta - y \cdot \sin \theta) dx dy. \quad (4)$$

Нормальное преобразование Радона имеет простой геометрический смысл — это интеграл от функции $f(x, y)$ вдоль прямой, описанной выражением (3) и повернутой на угол θ , перпендикулярной вектору $(\cos \theta, \sin \theta)$, проходящей на расстоянии ρ от начала координат [35].

В случае применения нормального преобразования Радона $R_\theta(x')$ к изображению, описанному функцией $f(x, y)$, результат преобразования представляет собой набор линейных интегралов от функции $f(x, y)$, полученный при вращении оси абсцисс (Рисунок 4). Другими словами, для представления изображения через нормальное преобразование Радона необходимо взять несколько параллельно-лучевых отображений изображения, вычисленных путем поворота исходной линии (оси абсцисс) вокруг центра изображения на различные углы [36].

Рассмотренное преобразование применяется в разработанном алгоритме извлечения встроенных данных из изображений, содержащих текст со встроенным РВЗ, на этапе обнаружения линий текста. Также

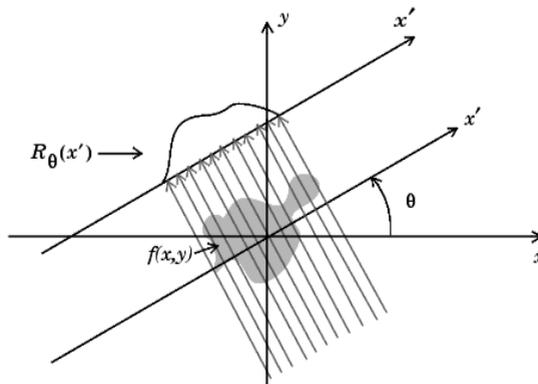


Рис. 4. Графическая интерпретация преобразования Радона

данное преобразование позволяет определить угол поворота изображения, содержащего текст, и учесть его при извлечении линий текста.

5.2. Алгоритм извлечения встроенной информации из изображений. Алгоритм извлечения информации, встроенной в текстовые данные из изображений, содержащих робастный водяной знак, представлен в листинге 2.

В качестве исходных данных выступает изображение, содержащее встроенную информацию в текстовые данные посредством алгоритма стеганографического внедрения, основанного на изменении межстрочного интервала (листинг 1).

На первом этапе осуществляется предварительная обработка исходного изображения Im_t (шаги 1-2). Функция *ConvertToGray* преобразует цветное пространство изображения Im_t к полутоновому изображению (в градациях серого) Im_{grey} . Компонента яркости полутонового изображения Im_{grey} подвергается среднеарифметической фильтрации (функция *Filtration*). В результате обработки исходного изображения формируется нормализованное изображение в градациях серого Im_{filt} .

В процессе обнаружения линий текста (шаг 3) нормализованное изображение Im_{filt} преобразуется в синограмму *sinogram* посредством нормального преобразования Радона. Под синограммой понимается двумерное распределение одномерных проекций слоя объекта как функции угла проекции, где угол проекции расположен по оси ординат, а линей-

Data: Изображение, содержащее текст Im_t
Result: Встроенная информация I_e

```

1  $Im_{grey} \leftarrow \text{ConvertToGray}(Im_t)$ 
2  $Im_{filt} \leftarrow \text{Filtration}(Im_{grey})$ 
3  $sinogram \leftarrow \text{RadonTransform}(Im_{filt})$ 
4 for  $l \leftarrow 0$  to 180 do
5    $R[l] \leftarrow \text{RmsCalculation}(sinogram, l)$ 
6  $rot \leftarrow \text{Argmax}(R)$ 
7  $row \leftarrow sinogram[rot]$ 
8  $M \leftarrow \text{FindPicks}(row)$ 
9  $min\_mode, max\_mode \leftarrow \text{FindModes}(M)$ 
10  $D \leftarrow \text{CorrectErrors}(M)$ 
11 if  $\text{Std}(D) > 0, 8$  then
12    $B \leftarrow \text{GaussianMixtureModel}(D, min\_mode, max\_mode)$ 
13 else
14    $b_i \leftarrow 0, i = \overline{1, |D|}$ 
15    $B \leftarrow \{b_i\}$ 
16  $I_e \leftarrow B$ 
17 return  $I_e$ 
    
```

Листинг 2. Извлечение встроенной информации из изображения

ная проекция расположена по оси абсцисс [37]. Данное преобразование осуществляется функцией *RadonTransform*.

Из сформированной синограммы посредством функции *RmsCalculation* для каждого ряда (угла поворота) вычисляется среднеквадратичное значение ($R[l]$) согласно выражения (шаги 4-5):

$$Rms = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}},$$

где x_1, x_2, \dots, x_n — значения синограммы по углу l .

Затем из синограммы посредством функции *Argmax* выбирается ряд row , обладающий наиболее выровненным по линии чередованием белых полос и черных точек. Данный ряд определяет положение текста и угол поворота rot исходного изображения.

Этап вычисления величин межстрочных интервалов текста изображения состоит из обнаружения линий текста и вычисления величин межстрочных значений функцией *FindPicks*. На первом этапе происходит обнаружение линий текста. Под линией текста понимается виртуальная линия, по которой выровнен или объединен текст [36]. Для извлечения линий текста в row осуществляется поиск индексов смены знака, которые соответствуют линиям исходного текста. Расстояние между индексами смены знака определяет величину межстрочного интервала между двумя

строками исходного текста. Массив величин межстрочных интервалов M вычисляется как разность между двумя индексами смены знака.

На этапе коррекции ошибок (шаг 10) полученные значения M проверяются на наличие ошибок первого и второго рода. Под ошибками первого рода понимается извлечение межстрочных интервалов, не присутствующих в исходном тексте. Ошибки второго рода характеризуются пропуском межстрочных интервалов в исходном тексте. Для обнаружения ошибок функцией *FindModes* вычисляется верхняя и нижняя моды min_mode, max_mode массива M . Если элемент массива больше, чем значение $1,7 \cdot max_mode$, принимается решение о наличии ошибки второго рода. Для исправления обнаруженной ошибки указанный элемент массива заменяется на два элемента, каждый из которых имеет значение max_mode . Ошибкам первого рода соответствуют величины межстрочных интервалов массива M меньше, чем $0,6 \cdot min_mode$. Исправление ошибок первого рода осуществляется путем удаления данного элемента из массива M .

На шаге 11-15 осуществляется преобразование скорректированного массива D в двоичный вид. Полученный массив D можно интерпретировать как бимодальное распределение (значений нуля и единицы). Таким образом, задача преобразования полученной последовательности данных в двоичный вид может быть решена посредством применения методов математической статистики, направленных на разделение смеси распределений. В качестве метода разделения смеси используется модель разделения смеси нормальных распределений [38].

5.3. Разделение смеси распределений. Основная задача разделения заключается в восстановлении компонентов $f_{\omega}(X; \theta(\omega))$ и смешивающей функции $P(\omega)$ по заданной левой части $f(X)$ соотношения:

$$f(X) = \int_{-\infty}^{\omega} f_{\omega}(X; \theta(\omega)) dP(\omega) \quad (5)$$

и называется задачей идентификации компонентов смеси [39].

Выражение (5) представляет собой функцию плотности (полигон вероятностей) распределения называемой смесью (или P -смесью) распределения семейства F .

Данное разделение может быть реализовано в случае, если $P = \{P(\omega)\}$ представляет собой семейство смешивающих функций распределения и имеется двухпараметрическое семейство p -мерных плотностей (полигонов вероятностей) распределения:

$$F = \{f_{\omega}(X; \theta(\omega))\}, \quad (6)$$

где одномерный (целочисленный или непрерывный) параметр ω в качестве нижнего индекса функции f определяет специфику общего вида каждого компонента — распределения смеси, а в качестве аргумента при многомерном параметре θ определяет зависимость значений хотя бы части компонентов этого параметра от того, в каком именно составляющем распределении f_{ω} он присутствует.

Тогда задача разделения смеси распределений заключается в выборке классифицируемых наблюдений X_1, X_2, \dots, X_n , извлеченной из генеральной совокупности, являющейся смесью, описываемой выражением 5, генеральных совокупностей типа, приведенных в выражении 6.

Если в качестве функции плотности вероятности выступает плотность вероятности нормального закона распределения:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

то данный метод разделения смесей распределений называется гауссовской моделью разделения смеси распределений.

Преобразование массива D в двоичный массив B зависит от величины среднеквадратического отклонения значений элементов массива D . Функция *Std* осуществляет вычисление среднеквадратического отклонения значений элементов массива D . Если полученное значение среднеквадратического отклонения превышает 0,8, то бинаризация осуществляется с использованием функции *GaussianMixtureModel* (шаг 12), реализующей модель разделения смеси нормальных распределений. Интерпретация массива D как дискретного бимодального распределения позволяет разделить его на два класса. Класс, характеризующийся меньшим значением моды элементов, соответствует «0», а класс, характеризующийся большим значением моды, соответствует «1» в двоичном массиве B . При значении $Std(D) < 0,8$ массив D считается унимодальным и принимается решение об отсутствии встроенной информации. В результате чего на шагах 14-15 формируется двоичный массив B , состоящий из «0» элементов. Полученный двоичный массив B представляет собой встроенные данные I_e .

По окончании работы алгоритма извлечения информации, встроенной в изображение, содержащее робастный водяной знак, делается вывод

о подлинности полученных данных и соответствии их подписанному владельцу.

Для разработанного алгоритма извлечения информации, встроенной в текстовые данные из изображения, была произведена оценка следующих основных свойств:

- результативность;
- элементарность;
- корректность;
- вычислительная сложность.

Результативность (отсутствие аварийного останова) алгоритма заключается в том, что исполнение алгоритма завершится за конечное число шагов.

Данный алгоритм является элементарным, так как содержит блоки, выполняющие простые операции.

Доказательство корректности алгоритма определяется следующим образом: для заданных параметров алгоритм всегда позволяет определить информацию, представленную РВЗ, внедренную в изображение.

Временную сложность алгоритма определим для наиболее вычислительно сложных процедур:

- преобразования Радона, имеющего сложность $T_1(180 \cdot x \cdot y)$, где x, y — размер изображения в пикселях;
- разделения смеси нормальных распределений $T_2(n \cdot k \cdot p \cdot i)$, где n — число элементов в массиве, k — число модальностей ($k = 2$), p — размерность пространства ($p = 1$), i — число итераций алгоритма.

Таким образом, общую временную сложность алгоритма можно представить как $T(180 \cdot x \cdot y + 2 \cdot n \cdot i)$.

Сложность алгоритма по памяти равна $O(x \cdot y)$. Такая оценка сложности является линейной.

Для определения границ извлекаемости встроенных данных и робастности разработанного подхода к стеганографическому внедрению информации в текстовые данные РВЗ была проведена экспериментальная оценка.

6. Экспериментальная оценка применимости и робастности разработанного алгоритма извлечения встроенной информации из изображений. В ходе проведения эксперимента кодирование информации производилось путем увеличения межстрочного интервала на значения, находящиеся в пределах перцептивной невидимости (от 0,01 до 0,1). Внедрение информации осуществлялось в текстовые данные, набранные шрифтом Computer Modern Roman с размером кегля 14 пт.

РВЗ представлен последовательностью «0101», циклически встроенной в исходный текст.

Результат извлечения РВЗ из изображения, содержащего встроенные данные, представлен на рисунке 5 (Рисунок 5а — исходное изображение, содержащее РВЗ; Рисунок 5б — сформированная синопграмма; Рисунок 5в — результат преобразования Радона, где пики соответствуют линиям текста на изображении).

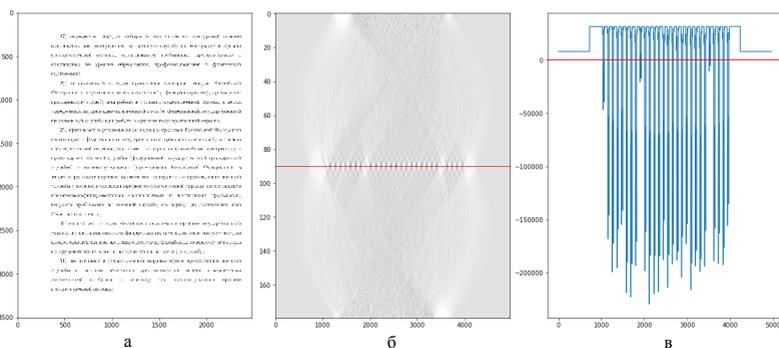


Рис. 5. Пример извлечения линий текста из изображения, содержащего текст

Полученные значения количества строк и соответствующий им массив величин межстрочных интервалов представлены в таблице 2.

Таблица 2. Массив величин межстрочных интервалов

Число строк в исходном документе	Вычисленное число строк	Массив извлеченных величин межстрочных интервалов
28	30	[83, 80, 83, 80, 82, 80, 83, 10, 4, 2, 80, 83, 80, 82, 80, 83, 80, 82, 80, 83, 80, 83, 188, 80, 83, 80, 83, 79, 83]
28	31	[85, 80, 86, 80, 85, 80, 86, 80, 12, 8, 86, 79, 86, 80, 86, 79, 86, 80, 86, 79, 86, 80, 86, 79, 86, 24, 80, 86, 80, 85]
28	27	[80, 88, 80, 89, 80, 88, 80, 89, 79, 89, 80, 88, 194, 80, 88, 80, 89, 79, 89, 80, 88, 80, 89, 80, 88, 80]

Результаты процесса исправления ошибок межстрочных интервалов приведены в таблице 3.

Результаты бинаризации извлеченных данных из изображений со встроенным РВЗ представлены в таблице 4.

Точность извлечения встроенных данных из изображений, содержащих РВЗ, представлена в таблице 5.

волит повысить точность извлечения встроенных данных из изображений, содержащих РВЗ.

Помимо оценки границ применимости разработанного алгоритма была произведена оценка устойчивости изображений со встроенным РВЗ к различным преобразованиям над изображениями.

Оценка стойкости разработанного алгоритма извлечения информации из изображений со встроенным РВЗ состоит в возможности достоверного извлечения информации из изображений после применения к ним следующих преобразований:

- поворот изображения;
- масштабирование;
- преобразование формата (JPEG, PNG, BMP, PDF);
- медианная фильтрация;
- гауссовская фильтрация;
- усредненная фильтрация.

Полученные результаты экспериментальной оценки стойкости разработанного алгоритма извлечения информации из изображения со встроенным РВЗ представлены в таблице 6. Пример результата извлечения данных из изображения, содержащего РВЗ, после применения гауссовского фильтра размытия с пределом показателя размытия 8 пикселей представлен на рисунке 6.

Таблица 6. Стойкость разработанного алгоритма к преобразованиям применяемым к изображениям

Тип преобразования	Устойчивость разработанного алгоритма к преобразованию
Поворот	поворот на любой угол
Масштабирование	множитель масштабирования до 1,5
Преобразование формата	JPEG, PNG, GIF, BMP, PDF
Медианная фильтрация	с пределом ядра свертки 9 пикселей
Гауссовская фильтрация	с пределом показателя размытия 8 пикселей
Усредненная фильтрация	с пределом ядра свертки 5 пикселей

7. Заключение. Разработанный подход к внедрению робастного водяного знака в текстовые данные позволяет обеспечить защиту авторских прав владельцев данных посредством внедрения в качестве РВЗ информации, идентифицирующей владельца данных, а также другой метаинформации. Кроме того, разработанный подход к внедрению робастного водяного знака обеспечивает инвариантность к рассмотренным преобразованиям и позволяет достоверно извлекать встроенные данные. Повышение точности извлечения, а также снижение количества ошибок встраивания являются направлением дальнейших исследований.

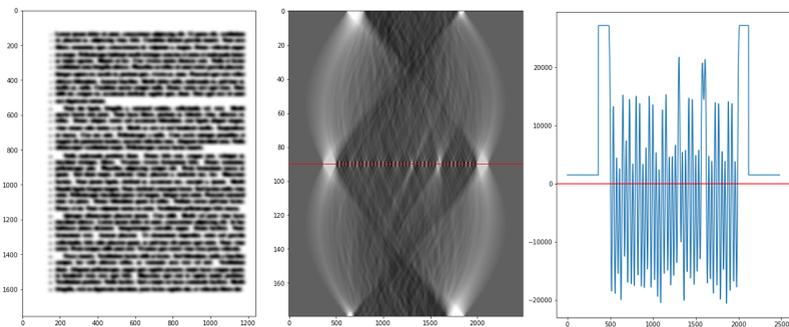


Рис. 6. Пример извлечения линий текста из изображения, содержащего текст

Литература

1. *Weber R.H.* Internet of Things – New security and privacy challenges // *Computer Law & Security Review*. 2010. vol. 26. no. 1. pp. 23–30.
2. *Chaudhuri A.* Internet of things data protection and privacy in the era of the General Data Protection Regulation // *Journal of Data Protection & Privacy*. 2016. vol. 1. no. 1. pp. 64–75.
3. JSOC Security flash report второе полугодие 2017 года. URL: solarsecurity.ru/upload/iblock/54f/flash_report_Q2_2017_041.pdf (дата обращения: 02.05.2018).
4. *Ramapriya B.* An Improved Approach of Text Steganography in Application with Rotational Symmetry // *International Journal of Innovative Research in Computer and Communication Engineering*. 2017. vol. 5. no. 7. pp. 12939–12947.
5. *Elmahi M.Y., Kosti Su., Sayed M.H.* Text Steganography Using Compression and Random Number Generators // *International Journal of Computer Applications Technology and Research*. 2017. vol. 6. no. 6. pp. 259–263.
6. *Kingslin S., Kavitha N.* Evaluative approach towards text steganographic techniques // *Indian Journal of Science and Technology*. 2015. vol. 8. no. 29. pp. 1–8.
7. *Odeh A., Elleithy K., Faezipour M., Abdelfattah E.* Highly efficient novel text steganography algorithm // *Systems, Applications and Technology Conference (LISAT)*. 2015. pp. 1–7.
8. *Rathore A.S., Rawat S.S.* A Secure Image and Text Steganography Technique // *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*. 2015. vol. 3. no. 5. pp. 506–509.
9. *Резанова З.И., Романов А.С., Мещеряков Р.В.* О выборе признаков текста, релевантных в авторведческой экспертной деятельности // *Вестник Томского государственного университета. Филология*. 2013. Т. 26. №. 6. С. 38–52.
10. *Грибунин В.Г., Оков И.Н., Турицев И.В.* Цифровая стеганография // М.: СОЛОН-Пресс. 2017. 262 с.
11. *Козачок А.В., Копьлов С.А., Мещеряков Р.В., Евсютин О.О.* Анализ текущего состояния научных исследований в области робастного хэширования изображений // *Компьютерная оптика*. 2017. Т. 41. №. 5. С. 743–755.
12. *Salomon D.* Data privacy and security: encryption and information hiding // *Springer Science & Business Media*. 2003. pp. 469.
13. *Woo C.-S.* Digital image watermarking methods for copyright protection and authentication // Ph.D. Thesis. Queensland University of Technology. 2007. 223 p.

14. *Phadikar A.* Robust Watermarking Techniques for Color Images // 2009.
15. *Евстигин О.О., Шелупанов А.А., Меццержков Р.В., Бондаренко Д.О.* Алгоритм встраивания информации в сжатые цифровые изображения на основе операции замены с применением оптимизации // Компьютерная оптика. 2017. Т. 41. №. 3. С. 412–421.
16. *Ватаманюк И.В., Ронжин А.Л.* Применение методов оценивания размытости цифровых изображений в задаче аудиовизуального мониторинга // Информационно-управляющие системы. 2014. Т. 71. №. 4. С. 16–23.
17. *Agarwal M.* Text steganographic approaches: a comparison // 2013. arXiv preprint arXiv:1302.2718. 16 p.
18. *Pamulaparty L., Rao N.M.* Text Steganography: Review // International Journal of Computer Science and Information Technology & Security (IJCSITS). 2016. vol. 6. no. 4. pp. 80–83.
19. *Ярмолик С.В., Листонад Ю.Н.* Стеганографические методы защиты информации // Информатизация образования. 2005. №. 5. С. 64–74.
20. *Shirali-Shahreza M., Shirali-Shahreza M.H.* Text steganography in SMS // Proceedings of IEEE International Conference on Convergence Information Technology. 2007. pp. 2260–2265.
21. *Бондарчук С.С., Давыдова Е.М., Костюченко Е.Ю.* Встраивание цифровых знаков для обеспечения защиты информации // Доклады Томского государственного университета систем управления и радиоэлектроники. 2011. Т. 24. №. 2. С. 228–235.
22. *Govada S.R., Kumar B.S., Devarakonda M., Stephen M.J.* Text steganography with multi level shielding // IJCSI International Journal of Computer Science Issues. 2012. vol. 9. no. 5. pp. 401–405.
23. *Nagarhalli T.P., Bakal J.W., Jain N.A* Survey of Hindi Text Steganography // International Journal of Scientific & Engineering Research. 2016. vol. 7. no. 3. pp. 55–61.
24. *Low S.H., Maxemchuk N.F., Brassil J.T., O’Gorman L.* Document marking and identification using both line and word shifting // Proceedings of IEEE INFOCOM’95 on Fourteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Bringing Information to People. 1995. vol. 2. pp. 853–860.
25. *Saraswathi V., Kingslin M.S.* Different Approaches to Text Steganography: A Comparison // International Journal of Emerging Research in Management & Technology. 2014. vol. 9359. no. 11. pp. 124–127.
26. *Por L.Y., Delina B.* Information hiding: A new approach in text steganography // Proceedings of WSEAS International Conference on Mathematics and Computers in Science and Engineering. 2008. vol. 7. pp. 689–695.
27. *Villán R. et al.* Tamper-proofing of electronic and printed text documents via robust hashing and data-hiding // Security, Steganography, and Watermarking of Multimedia Contents. 2007. pp. 6505–6517.
28. *Kumar K.A., Pabboju S., Desai N.M.S.* Advance text steganography algorithms: an overview // International Journal of Research and Applications. 2014. vol. 1. no. 1. pp. 31–35.
29. *Bhaya W., Rahma A.M., Al-nasrawi D.* Text steganography based on font type in MS-Word documents // Journal of Computer Science. 2013. vol. 9. no. 7. pp. 898–904.
30. *Rahma A.M.S., Bhaya W.S., Al-Nasrawi D.A.* Text steganography based on unicode of characters in multilingual // International Journal of Engineering Research and Applications. 2013. vol. 3. no. 4. pp. 1153–1165.
31. Word 2007: Rich Text Format (RTF) Specification, version 1.9.1. URL: www.microsoft.com/en-us/download/details.aspx?id=10725 (дата обращения: 02.05.2018).

32. 171/SC 2: ISO 32000–1: 2008 Document Management-Portable Document Format-Part 1 // ISO. TC. 2008. 747 p.
33. 171/SC 2: ISO 32000–2: 2017 Document Management-Portable Document Format-Part 2 // ISO. TC. 2017. 371 p.
34. *Toft P.A., Sørensen J.A.* The Radon transform-theory and implementation // Ph.D. Thesis. Technical University of Denmark. 1996. 321 p.
35. *Волков В.А.* Ряды Фурье. Интегральные преобразования Фурье и Радона: учебно-методическое пособие // Издательство Уральского университета. 2014. 32 с.
36. *Bahaghighat M.K., Mohammadi J.* Novel approach for baseline detection and Text line segmentation // International Journal of Computer Applications. 2012. vol. 51. no. 2. pp. 9–16.
37. ГОСТ Р МЭК/ТО 61948–2–2008 Оборудование для радионуклидной диагностики. Эксплуатационные испытания. Часть 2. Сцинтилляционные камеры и однофотонные компьютерные томографы // М.: Госстандарт России. 2009. 8 с.
38. *Reynolds D.* Gaussian mixture models // Encyclopedia of biometrics. 2015. pp. 827–832.
39. *Айвазян С.А., Буххутабер В.М., Енюков Е.С.* Прикладная статистика: Классификация и снижение размерности // М.: Финансы и статистика. 1989. 607 с.

Козачок Александр Васильевич — к-т техн. наук, сотрудник, Академия Федеральной службы охраны Российской Федерации. Область научных интересов: информационная безопасность, защита от несанкционированного доступа, математическая криптография, теоретические проблемы информатики. Число научных публикаций — 105. alex.totrin@gmail.com; Приборостроительная, 35, Орел, 302034; р.т.: +7(486) 254-99-33.

Копылов Сергей Александрович — сотрудник, Академия Федеральной службы охраны Российской Федерации. Область научных интересов: информационная безопасность, распознавание образов, обработка изображений. Число научных публикаций — 5. gremlin.kor@mail.ru; Приборостроительная, 35, Орел, 302034; р.т.: +7(4862)54-99-33.

Мещеряков Роман Валерьевич – д-р техн. наук, доцент, проректор по научной работе и инновациям, Томский государственный университет систем управления и радиоэлектроники (ТУСУР). Область научных интересов: системный анализ, информационная безопасность, вопросы обработки информации в интеллектуальных системах, информационно-безопасные системы, идентификация диктора, обработка речи, машинное обучение, программно-аппаратные средства защиты информации. Число научных публикаций — 400. mgrv@ieec.org; пр. Ленина, 40, Томск, 634050; р.т.: +7(3822)900111, Факс: +7 (3822) 900-111.

Евсютин Олег Олегович — к-т техн. наук, доцент кафедры безопасности информационных систем, Томский государственный университет систем управления и радиоэлектроники (ТУСУР). Область научных интересов: информационная безопасность, обработка цифровых изображений, цифровая стеганография, приложения клеточных автоматов. Число научных публикаций — 50. eoo@keva.tusur.ru; пр. Ленина, 40, Томск, 634050; р.т.: +7(3822)701529, Факс: +7(3822)513262.

Туан Лай Минь — к-т техн. наук, доцент, Академия криптографической техники СШПР Социалистической республики Вьетнам. Область научных интересов: информационная безопасность, защита от несанкционированного доступа. Число научных публикаций — 10. lmtuan.1989@gmail.com; Нгуен Чи Тан, 105, Донг Да, Ханой, 10, Вьетнам; р.т.: +438357975.

Поддержка исследований. Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации в рамках проектной части государственного задания ТУСУР на 2017-2019 гг. (проект № 2.3583.2017/ПЧ)

A.V. KOZACHOK, S.A. KOPYLOV, R.V. MESHCHERYAKOV,
O.O. EVSUTIN, L.M. TUAN

AN APPROACH TO A ROBUST WATERMARK EXTRACTION FROM IMAGES CONTAINING TEXT

Kozachok A.V., Kopylov S.A., Meshcheryakov R.V., Evsutin O.O., Tuan L.M. An Approach to a Robust Watermark Extraction from Images Containing Text.

Abstract. This paper presents an approach to a robust watermark extraction from images containing text. Data extraction based on developed approach to robust watermark embedding into text data, characterizing by conversion invariance of text data into an image format. The comparative analysis of existing approaches of steganographic data embedding into text data is carried out, their advantages and disadvantages are determined. The choice of groups to steganographic data embedding methods based on text formatting is justified. As an embedding algorithm is determined approach based on interline space shifting. The block diagram and the description of the developed algorithm of data embedding into text data are given. An experimental estimation of the embedding capacity and perceptual invisibility of the developed data embedding approach was carried out. An approach to extract embedded information from images containing a robust watermark, based on the existing limitations, has been developed. The Radon transform is chosen as the basic extraction procedure of embedded information, allowing to extract values of the interline spacing. An approach based on Gaussian mixture model separating to isolate the values of the bits was chosen. The limits of the retrieval of embedded data have been experimentally established, and the robustness of the developed embedding approach to the implementation of various transformations has been estimated. The following parameters of robustness developed approach are defined: rotation of an image containing embedded data at any angle; scaling an image with a scaling factor not exceeding 1.5; conversion to any bitmap format; the application of a median filter to an image with a convolution core limit of not more than 9, a Gaussian blur filter with a blurring limit not exceeding 8 and an average filter with a convolution kernel limit of not more than 5.

Keywords: information security, text steganography, digital watermarking, Radon transform, separation of mixture distribution.

Kozachok Alexander Vasilievich — Ph.D., researcher, The Academy of Federal Security Guard Service of the Russian Federation. Research interests: information security, unauthorized access protection, mathematical cryptography, theoretical problems of computer science. The number of publications — 105. alex.totrin@gmail.com; 35, Priborostroitel'naya Street, Orel, 302034, Russia; office phone: +7(486) 254-99-33.

Kopylov Sergey Alexandrovich — researcher, The Academy of Federal Security Guard Service of the Russian Federation. Research interests: information security, protection from unauthorized access, image processing. The number of publications — 5. gremlin.kop@mail.ru; 35, Priborostroitel'naya Street, Orel, 302034, Russia; office phone: +7(4862)54-99-33.

Meshcheryakov Roman Valeryevich — Ph.D., Dr. Sci., professor, vice-rector for research and innovation, Tomsk State University of Control Systems and Radioelectronics (TUSUR). Research interests: speech analysis, speech recognition, medical technology, information security, speaker recognition, speech processing, machine learning, hardware-software data protection solutions. The number of publications — 400. mrv@ieee.org; 40, Lenin-avenue Tomsk, 634050, Russia; office phone: +7(3822)900111, Fax: +7 (3822) 900-111.

Evsutin Oleg Olegovich — Ph.D., associate professor of information system security department, Tomsk State University of Control Systems and Radioelectronics (TUSUR). Research interests: information security, digital images processing, applications of cellular automata theory. The number of publications — 50. eoo@keva.tusur.ru; 40, Lenin-avenue Tomsk, 634050, Russia; office phone: +7(3822)701529, Fax: +7(3822)513262.

Tuan Lai Minh — Ph.D., associate professor, Cryptographic technology Academy. Research interests: information security, unauthorized access protection. The number of publications — 10. lmtuan.1989@gmail.com; 105, Nguen Chi Than, Dong Da, Hanoi, 10, Vietnam; office phone: +438357975.

Acknowledgements. This work is supported by the Ministry of Education and Science of the Russian Federation within the framework of the project part of the state task of TUSUR for 2017-2019 (Project № 2.3583.2017/Pch)

References

1. Weber R.H. Internet of Things – New security and privacy challenges. *Computer Law & Security Review*. 2010. vol. 26. no. 1. pp. 23–30.
2. Chaudhuri A. Internet of things data protection and privacy in the era of the General Data Protection Regulation. *Journal of Data Protection & Privacy*. 2016. vol. 1. no.1. pp. 64–75.
3. JSOC Security flash report 2017 second half year. Available at: solarsecurity.ru/upload/iblock/54f/flash_report_Q2_2017_041.pdf (accessed: 02.05.2018) (In Russ.).
4. Ramapriya B. An Improved Approach of Text Steganography in Application with Rotational Symmetry. *International Journal of Innovative Research in Computer and Communication Engineering*. 2017. vol. 5. no. 7. pp. 12939–12947.
5. Elmahi M.Y., Kosti Su., Sayed M.H. Text Steganography Using Compression and Random Number Generators. *International Journal of Computer Applications Technology and Research*. 2017. vol. 6. no. 6. pp. 259–263.
6. Kingslin S., Kavitha N. Evaluative approach towards text steganographic techniques. *Indian Journal of Science and Technology*. 2015. vol. 8. no. 29. pp. 1–8.
7. Odeh A., Elleithy K., Faezipour M., Abdelfattah E. Highly efficient novel text steganography algorithmss. Systems, Applications and Technology Conference (LISAT). 2015. pp. 1–7.
8. Rathore A.S., Rawat S.S. A Secure Image and Text Steganography Technique. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*. 2015. vol. 3. no. 5. pp. 506–509.
9. Rezanova Z.I., Romanov A.S., Meshcheryakov R.V. [About choice of signs of the text relevant in the autor expert's activity]. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal* 2013. vol. 26. no. 6. pp. 38–52. (In Russ.).
10. Gribunin V.G., Okov I.N., Turincev I.V. *Cifrovaya steganografiya*. [Digital steganography]. M.: SOLON-Press., 2017. 262 p. (In Russ.).
11. Kozachok A.V., Kopylov S.A., Meshcheryakov R.V., Evsutin O.O. [Robust image hashing survey]. *Kompjuternaja Optika – Computer Optics*. 2017. vol. 41. no. 5. pp. 743–755. (In Russ.).
12. Salomon D. Data privacy and security: encryption and information hiding. *Springer Science & Business Media*. 2003. 469 p.
13. Woo C.-S. Digital image watermarking methods for copyright protection and authentication. Ph.D. Thesis. Queensland University of Technology. 2007. 223 p.
14. Phadikar A. Robust Watermarking Techniques for Color Images. 2009.

15. Evsutin O.O., Shelupanov A.A., Meshcheryakov R.V., Bondarenko D.O. [An algorithm for information embedding into compressed digital images based on replacement procedures with use of optimization]. *Kompjuternaja Optika – Computer Optics*. 2017. vol. 41. no. 3. pp. 412–421. (In Russ.).
16. Vatomanyuk I.V., Ronzhin A.L. [Application of digital images blur estimation methods for audiovisual monitoring]. *Informatsionno-upravliaiushchie sistemy – Information and Control Systems*. 2014. vol. 71. no. 4. pp. 16–23. (In Russ.).
17. Agarwal M. Text steganographic approaches: a comparison. 2013. arXiv preprint arXiv:1302.2718. 16 p.
18. Pamulaparty L., Rao N.M. Text Steganography: Review. *International Journal of Computer Science and Information Technology & Security (IJCSITS)*. 2016. vol. 6. no. 4. pp. 80–83.
19. Jarmolik S.V., Listopad J.N. [Steganographic methods of information protection]. *Informacionnoe obrazovanie – Information education*. 2005. vol. 5. pp. 64–74. (In Russ.).
20. Shirali-Shahreza M., Shirali-Shahreza M.H. Text steganography in SMS. Proceedings of IEEE International Conference on Convergence Information Technology. 2007. pp. 2260–2265.
21. Bondarchuk S.S., Davydova E.M., Kostjuchenko E.J. [Integration of digital characters for information security]. *Doklady Tomskogo Gosudarstvennogo Universiteta Sistem Upravlenija i Radioelektroniki – Proceedings of Tomsk State University of Control Systems and Radioelectronics*. 2011. vol. 24. no. 2. pp. 228–235. (In Russ.).
22. Govada S.R., Kumar B.S., Devarakonda M., Stephen M.J. Text steganography with multi level shielding. *IJCSI International Journal of Computer Science Issues*. 2012. vol. 9. no. 5. pp. 401–405.
23. Nagarhalli T.P., Bakal J.W., Jain N. A Survey of Hindi Text Steganography. *International Journal of Scientific & Engineering Research*. 2016. vol. 7. no. 3. pp. 55–61.
24. Low S.H., Maxemchuk N.F., Brassil JT., O’Gorman L. Document marking and identification using both line and word shifting. Proceedings of IEEE INFOCOM’95 on Fourteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Bringing Information to People. 1995. vol. 2. pp. 853–860.
25. Saraswathi V., Kingslin M.S. Different Approaches to Text Steganography: A Comparison. *International Journal of Emerging Research in Management & Technology*. 2014. vol. 9359. no. 11. pp. 124–127.
26. Por L.Y., Delina B. Information hiding: A new approach in text steganography. Proceedings of WSEAS International Conference on Mathematics and Computers in Science and Engineering. 2008. vol. 7. pp. 689–695.
27. Villán R. et al. Tamper-proofing of electronic and printed text documents via robust hashing and data-hiding. *Security, Steganography, and Watermarking of Multimedia Contents*. 2007. pp. 6505–6517.
28. Kumar K.A., Pabboju S., Desai N.M.S. Advance text steganography algorithms: an overview. *International Journal of Research and Applications*. 2014. vol. 1. no. 1. pp. 31–35.
29. Bhaya W., Rahma A.M., Al-nasrawi D. Text steganography based on font type in MS-Word documents. *Journal of Computer Science*. 2013. vol. 9. no. 7. pp. 898–904.
30. Rahma A.M.S., Bhaya W.S., Al-Nasrawi D.A. Text steganography based on unicode of characters in multilingual. *International Journal of Engineering Research and Applications*. 2013. vol. 3. no. 4. pp. 1153–1165.
31. Word 2007: Rich Text Format (RTF) Specification, version 1.9.1. Available at: www.microsoft.com/en-us/download/details.aspx?id=10725 (accessed: 02.05.2018).
32. 171/SC 2: ISO 32000-1: 2008 Document Management-Portable Document Format-Part 1. ISO. TC. 2008. 747 p.

33. 171/SC 2: ISO 32000–2: 2017 Document Management-Portable Document Format-Part 2. ISO. TC. 2017. 371 p.
34. Toft P.A., Sørensen J.A. The Radon transform-theory and implementation. Ph.D. Thesis. Technical University of Denmark. 1996. 321 p.
35. Volkov V.A. *Rjady Fur'e. Integral'noe preobrazovanie Fur'e i Radona: uchebno-metodicheskoe posobie*. [Fourier series. Integral Fourier and Radon transform: a teaching aid]. Ekaterinburg: Ural University Publ., 2014. 32 p. (In. Russ.)
36. Bahaghighat M.K., Mohammadi J. Novel approach for baseline detection and Text line segmentation. *International Journal of Computer Applications*. 2012. vol. 51. no. 2. pp. 9–16.
37. GOST R IEC/TR 61948-2–2008 [Nuclear medicine instrumentation. Routine tests. Part 2. Scintillation cameras and single photon emission computed tomography imaging] M.: Gosstandart Rossii. 2009. 8 p. (In. Russ.)
38. Reynolds D. Gaussian mixture models. *Encyclopedia of biometrics*. 2015. pp. 827–832.
39. Ajvazjan S.A., Buhshtarev V.M., Enjukov E.S. *Prikladnaja statistika. Klassifikacija i snizhenie razmernosti*. [Applied statistics. Classification and reduction of dimension]. M.: Finance and statistics. 1989. 607 p. (In. Russ.)

Р.А. НЕЙДОРФ, А.Г. АГАДЖАНЯН
**ДУАЛЬНАЯ ОПТИМИЗАЦИЯ ТОНОВОЙ АППРОКСИМАЦИИ
МОНОХРОМНЫХ ИЗОБРАЖЕНИЙ ПАРАЛЛЕЛЬНЫМ
ЭВОЛЮЦИОННО-ГЕНЕТИЧЕСКИМ ПОИСКОМ**

Нейдорф Р.А., Агаджанян А.Г. Дуальная оптимизация тоновой аппроксимации монохромных изображений параллельным эволюционно-генетическим поиском.

Аннотация. Рассматривается оптимизация процедуры тоновой аппроксимации полутоновых (например, в палитре серого цвета) изображений. Процедура тоновой аппроксимации подразумевает сокращение в палитре аппроксимированного изображения количества используемых тонов по сравнению с количеством тонов в палитре исходного изображения. Оптимизация этой процедуры заключается в минимизации потери качества передачи графической информации, которая оценивается суммарным или усредненным по изображению отклонением тонов координатно-идентичных пикселей аппроксимированного изображения от тонов исходного. В качестве инструмента оптимизации предлагается гибридный алгоритм, который совмещает эвристический и детерминированный алгоритмы поиска наилучшей по критерию минимизации ошибки аппроксимации структуры аппроксимирующей палитры. Эвристический алгоритм реализован на основе эволюционно-генетической парадигмы. Его задачей является поиск области тоновых структур аппроксимирующей палитры, максимально близких к оптимальной. Цель детерминированного алгоритма направленного перебора — найти ближайший к полученному предыдущим поиском результату экстремум критерия качества аппроксимации. Эвристический алгоритм, как более быстродействующий, нацелен на оперативное сокращение области поиска, а детерминированный, как более затратный, — на нахождение хотя бы локального экстремума (а, возможно, и глобального) по максимально сокращенному предыдущим алгоритмом пути. Совместная работа этих алгоритмов позволяет обеспечить процессу тоновой аппроксимации эффект оптимизации, названный в статье дуальной. Под этим термином подразумевается получение результата, при котором достигается экстремум критерия качества аппроксимации при минимизации времени его достижения. Описываемое в статье исследование посвящено повышению результативности гибридного алгоритма на эвристическом этапе, в качестве которого используется модифицированный эволюционно-генетический алгоритм. Рассматриваются перспективы разработки и оценки эффективности внедрения модели параллельного использования алгоритмов с различными параметрами настройки. Обсуждаются первичные эксперименты, а их результаты сравниваются с известным алгоритмом решения поставленной задачи.

Ключевые слова: тоновая аппроксимация, гибридизация, параллельная модель, оптимизация, адаптивная схема, эволюционно-генетический алгоритм, аппроксимирующая палитра.

1. Введение. Тоновая аппроксимация представляет собой понижение размерности палитры, воспроизводящей цифровое изображение. Данная процедура предполагает замену оригинальных тонов изображения, исходной палитры (ИП), чаще всего стандартной, тонами аппроксимирующей палитры (АП). Поскольку ИП имеет большое количество тонов для воспроизведения изображения, а тоновая размерность АП значительно меньше результатом тоновой аппроксимации оказывается упрощение

образов, передаваемых изображением. Если при этом его основные визуальные признаки сохраняются, то возникает возможность выполнять обработку изображения с меньшими затратами. Это востребовано во многих задачах (сжатие цифровых файлов [1], распознавание образов [2, 3], в технологиях защиты изображений [3] и т.д.). Несмотря на то, что проблема тоновой аппроксимации возникла еще в начале 80-х годов, в современных реалиях эта проблема все еще актуальна [3-12].

В настоящей работе процедура тоновой аппроксимации рассматривается применительно к монохромным мультитоновым изображениям (ММИ), состоящим из 256 градаций одного цвета, которые имеют широкое распространение в технических задачах, и их обработка является также актуальной проблемой, несмотря на серьезное развитие цифровых технологий [2, 6]. Это обусловлено широким использованием ММИ в алгоритмах распознавания образов, в частности препятствий типа помех безопасному движению, что актуально в связи с распространением беспилотных транспортных средств [2, 6-8]. Таким образом, тоновая аппроксимация ММИ востребована в сжатии информации [9], распознавании образов [10-12], и может быть использована в ряде других технических отраслей [14, 15]. При этом очевидно, что для успешного развития этого научно-технического направления важно выявить базовые закономерности тоновой аппроксимации ММИ.

Качество процедуры тоновой аппроксимации определяется степенью средней близости тонов АП и ИП по всему изображению. Это, в первую очередь, определяется структурой АП, то есть тонами, входящими в эту палитру. Поэтому границы покрытия тонов ИП одним тоном АП должны быть сформированы из условия их наименьшего среднего взаимного отклонения в аппроксимированном ММИ.

Наиболее популярным и используемым на практике алгоритмом подбора АП для тоновой аппроксимации применительно к цветным растровым изображениям является алгоритм «медианного сечения», предложенный Полом Хэкбертом [16]. Его широкое применение обусловлено простотой и высокой скоростью обработки. Построение варианта алгоритма медианного сечения применительно к ММИ приводит к равномерному распределению тонов АП по шкале ИП. Это означает разделение стандартной палитры ММИ на приблизительно равные участки, количество которых идентично размеру АП, а средний тон каждого участка ИП входит в АП.

Однако исследования показали, что АП с равномерно распределенными тонами из ИП лишь в редких случаях дает

удовлетворительное качество аппроксимации. Поисковая оптимизация с целью улучшить качество аппроксимации варьированием тонов приводят к существенно неравномерной АП [17-22], что обуславливает целесообразность исследования способов усовершенствования алгоритма понижающей тоновой аппроксимации. Однако поисковая оптимизация, во-первых, не гарантирует получения оптимального результата аппроксимации, а, во-вторых, может обеспечить приближение к нему лишь за счет существенного увеличения временного ресурса работы алгоритма.

Целью представленного исследования являлась разработка алгоритма оптимальной или предельно близкой к оптимальной по качеству тоновой аппроксимации монохромных изображений, обладающий при этом свойством субоптимальности по быстродействию.

2. Сущность и инструменты аппроксимации изображений.

Ниже рассматривается пример тоновой аппроксимации произвольного ММИ производственного помещения, выполненного в стандартной ИП при разрешении 400 на 261 пиксель. Решается задача преобразования его в аппроксимированный ММИ, выполненный в 8 тоновой АП.

2.1. Влияние структуры АП на результат аппроксимации.

Изображение на рисунке 1 аппроксимировалось двумя алгоритмами преобразования:

- с использованием АП, полученного с помощью алгоритма медианного сечения;
- с использованием АП с доказанной экстремальностью качества аппроксимации [20-22].



Рис. 1. Оригинальное ММИ с отмеченными фрагментами I, II и III

Результаты аппроксимации оригинального ММИ двумя алгоритмами анализировались для трех фрагментов I, II и III снимка, выделенных рамками на рисунке 1. Они рассматриваются далее в увеличенном формате.

На увеличенном участке №I на рисунке 2 видно, что после тоновой аппроксимации алгоритмом медианного сечения, кронштейн и колесная база, отчетливо различимые на оригинальном ММИ (рисунок 2а), отображаются на аппроксимированном ММИ, чья палитра получена с помощью алгоритма медианного сечения (рисунок 2б), с сильными погрешностями. В качестве причины такого эффекта можно выдвинуть предположение о недостаточности выбранной размерности АП для более корректного отображения данного объекта. Однако дальнейшие исследования показывают, что это не так.

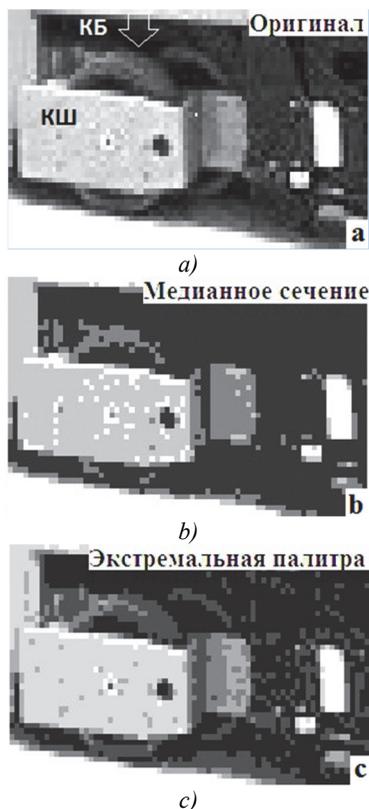


Рис. 2. Увеличенный участок №I: а) оригинального изображения; б, с) аппроксимированного различными алгоритмами (КШ — кронштейн, КБ — колесная база).

Равномерное распределение тонов АП способно обеспечить эффективную аппроксимацию лишь в том случае, когда частотно-яркостная диаграмма оригинального ММИ достаточно равномерна, что, конечно, возможно, но на практике встречается крайне редко. На рисунке 3 представлены такие диаграммы трех различных ММИ, на которых видно, что ни одно из них не имеет равномерной диаграммы.

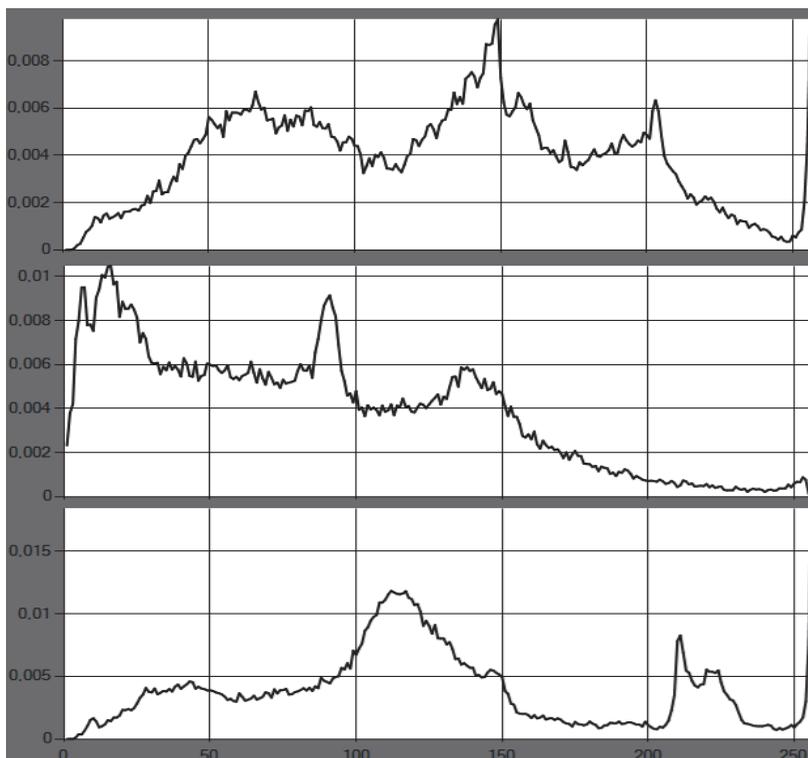


Рис. 3. По оси абсцисс отложены значения тонов ИП, а по оси ординат — количество пикселей этого тона в ММИ

2.2. Гибридный алгоритм оптимизации структуры АП.

Таким образом, одной из составляющих задачи тоновой аппроксимации оригинального ММИ является построение оптимальной или близкой к ней АП. Высокая размерность объекта, связанная как с количеством пикселей в оригинальном ММИ, так и с показательной зависимостью количества комбинаций m тонов АП из n тонов ИП, обуславливает, во-первых, поисково-оптимизационный

характер этой задачи, а во-вторых, предпочтительное применение эвристических алгоритмов для ее решения. Поэтому авторами проведены исследования по применению эволюционно-генетического алгоритма (ЭГА) в качестве инструмента решения этой задачи. Эффективность такого подхода с точки зрения возможности получения высокого качества аппроксимации продемонстрирована в работах [17-19].

Тем не менее использование ЭГА в чистом виде недостаточно эффективно. С одной стороны, эвристическая и в значительной степени вероятностная сущность ЭГА не позволяет гарантировать нахождения не только глобального, но и достаточно близкого к нему локального экстремума. С другой стороны, многовариантность объекта оптимизации приводит к низкому быстродействию ЭГА. Для решения первой проблемы был разработан детерминированный алгоритм проверки экстремальности произвольной АП [20]. Суть данного метода заключается в формировании ближайшей окрестности исследуемой АП [20], который можно рассматривать как вектор (см. раздел 2), и полным перебором в поиске лучшей точки (АП) согласно критерию качества (см. раздел 2). Соответственно, отсутствие более лучшей точки, нежели исследуемая АП, говорит о статусе локального экстремума. Алгоритм проверки решения на экстремальность [20] показал, что ЭГА не всегда выдает «экстремальный» выходной результат.

Таким образом, целесообразно использовать данный алгоритм для итеративного поиска экстремальной АП. Если очередная найденная АП оказывается не экстремальной, то из исследованной окрестности выбирается наилучший вариант и также исследуется на экстремальность. Объединение этого детерминированного алгоритма с ЭГА позволило построить гибридный алгоритм субоптимизации структуры АП [21, 22]. В нем ЭГА выполняет задачу сужения области поиска субоптимальной АП, а алгоритм поиска ближайшего экстремума обеспечивает субоптимальность решения. Поскольку алгоритм поиска экстремума, входящий в гибридный, основан на полном переборе, общее быстродействие поисковой оптимизации гибридным алгоритмом зависит от близости к субоптимальному решению найденного ЭГА. Таким образом, повышение эффективности ЭГА влечет за собой сокращение количества итераций алгоритма поиска ближайшего экстремума, что уменьшает общее время работы гибридного алгоритма.

В результате нужно отметить, что разработанная гибридная модель [21, 22] обнаруживает свойство дуальности оптимизации в том

смысле, что гибридный алгоритм обеспечивает как субоптимизацию быстродействия, так и гарантированную экстремальность точности аппроксимации [21, 22]. Это позволяет при экспериментальной настройке этого алгоритма в качестве критерия субоптимизации использовать лишь общее вычислительное время.

2.3. Качественное сравнение результатов аппроксимации ММИ алгоритмом медианного сечения и гибридным алгоритмом субоптимизации. Результат обработки ММИ разработанным гибридным алгоритмом субоптимизации АП представлен на рисунке 2с с надписью: «Экстремальная палитра». Визуальный анализ показывает, что полученная экстремальная 8-тоновая АП позволяет значительно качественнее, чем после алгоритма медианного сечения, отразить как колесную базу, так и кронштейн механизма (см. рисунок 2b и 2с). Таким образом, гибридный алгоритм субоптимизации АП обеспечивает улучшение качества тоновой аппроксимации, выполняя более эффективную тоновую аппроксимацию ММИ, что может быть использовано в соответствующих областях. Тот же эксперимент показал, что в большинстве случаев гибридный алгоритм демонстрирует повышенное быстродействие нахождения экстремальной палитры.

В подтверждение универсальности (многоцелевого характера) гибридный алгоритм субоптимизации АП исследованы результаты аппроксимации фрагментов II и III.

На увеличенном фрагменте II наиболее интересна корректность отображение границы кузова погрузочного аппарата (рисунок 4а), что может быть потенциально использовано для повышения точности алгоритма распознавания образов, в том числе препятствий для автономно движущихся объектов.

На представленном примере видно, что корректный подбор АП позволяет точнее отразить границы исследуемого объекта (рисунок 4b и 4с), но следует отметить, что граница кузова несколько расплывчата и на самом оригинальном ММИ (рисунок 4а), а это, несомненно, отражается и на результате разработанного алгоритма.

Последний фрагмент (рисунок 5) демонстрирует эффективность разработанного алгоритма как с точки зрения повышения качества тоновой аппроксимации и применения с целью эффективного сжатия объема файла, так и с точки зрения применения с целью эффективного сжатия объема файла, так и с точки зрения применимости в техническом зрении. Металлическая труба, составляющая каркас кабины, четко различима на оригинальном ММИ (рисунок 5а), практически полностью размывается, после обработки алгоритмом медианного сечения (рисунок 5b).



Рис. 4. Увеличенный участок №II: а) оригинального изображения; б, с) аппроксимированного различными алгоритмами

Результат тоновой аппроксимации разработанным гибридным алгоритмом субоптимизации АП позволил корректно отразить как металлическую трубу, так и отдельные мелкие элементы исследуемого участка (рисунок 5с).

Необходимо отметить, что фрагменты I и III были подвергнуты значительному повышению контрастности сторонним редактором, чтобы продемонстрировать их в печатной версии статьи.

Положительные результаты кратко описанных визуальных исследований послужили основанием для дополнительных структурных и количественных уточняющих исследований, результаты которых использованы при написании данной статьи.

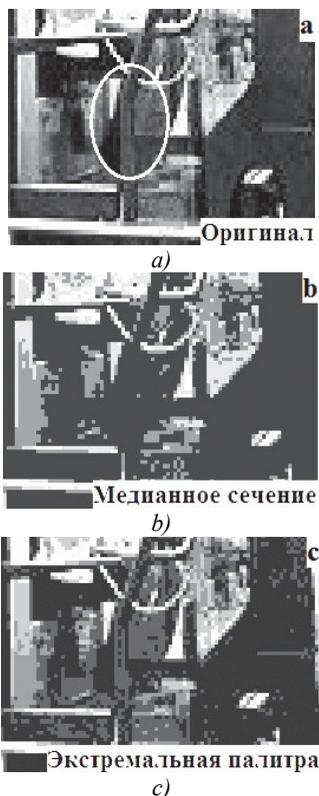


Рис. 5. Увеличенный участок №III: а) оригинального изображения; б, в) аппроксимированного различными алгоритмами

2.4. Возможные пути решения поставленной задачи.

Исследование разработанного авторами ранее гибридного алгоритма субоптимальной аппроксимации с использованием ЭГА и алгоритма нахождения ближайшего экстремального АП дает основание предполагать, что, повышение поисковой точности ЭГА за счет адаптивного подбора его наилучших вероятностных настроек, упростит задачу алгоритма поиска экстремальной АП, что сократит время его работы. Это должно обеспечить одновременно и высокоточную и быстродействующую аппроксимацию ММИ.

Для объективной оценки результатов исследований используются математические модели описания и аппроксимации ММИ, сущность которых раскрывается далее, а также описывается разработанный вариант предметной модификации используемого для субоптимизации ЭГА.

3. Алгоритмы оптимальной аппроксимации ММИ. Ранее было отмечено, что процедура тоновой аппроксимации заключается в сокращении размерности яркостной палитры изображения. Рассмотрим математическую базу данной процедуры.

3.1. Математическая модель аппроксимации ММИ. Одним из основных свойств, влияющих на качество воспроизведения любого изображения (в частности ММИ), является палитра, которая может быть представлена как упорядоченное множество тонов:

$$P^S = \{p^S \mid s \in [1, S]\} = (p^1, \dots, p^S, \dots, p^S), \quad (1)$$

где S — ее размерность, то есть количество различных тонов для зрительной передачи изображения.

Процедура тоновой аппроксимации осуществляется посредством замены пикселей ИП в оригинальном ММИ на пиксели АП. Структура ИП имеет вид (1) с исходными пикселями $p_{s^0}^0$ и представлена записью:

$$P^o = \{p_{s^o}^o \mid s^o \in [1, S^o]\} = (p_1^o, \dots, p_{s^o}^o, \dots, p_{s^o}^o). \quad (2)$$

Структура АП с размерностью $S^a < S^o$ и аппроксимирующими пикселями $p_{s^a}^a$ также имеет вид (1):

$$P^a = \{p_{s^a}^a \mid s^a \in [1, S^a]\} = (p_1^a, \dots, p_{s^a}^a, \dots, p_{s^a}^a). \quad (3)$$

Данная операция может реализовываться различными алгоритмами, на что было указано в предыдущих разделах статьи, но фундаментальная часть этих алгоритмов одинакова. Она заключается в замене определенного пикселя оригинального ММИ из ИП (2) на соответствующий пиксель аппроксимированного ММИ из АП (3). При этом в ИП P^{s^0} необходимо наилучшим образом подобрать исходные подмножества пикселей $P_s^{s^0}$:

$$P_{s^a}^a = \left\{ p_{s_a^o}^o \mid s_a^o \in [s_{s^a+d_{s^a-1}}^{s^o}, s_{s^a+d_{s^a-1}}^{s^o} + d_{s^a} - 1], s^a \in [1, S^a] \right\}, \quad (4)$$

где d_{s^a} — количество тонов $p_{s_a^o}^o$ ИП, заменяемых тоном $p_{s^a}^a$ из АП.

При этом диапазоны яркостей, которые предназначены для замещения конкретными пикселями $p_{s^a}^a \in P^a$, а следовательно, и подмножества $P_{s^a}^o$, должны отвечать двум обязательным условиям:

$$P_{s^a}^o \subset P^o : \forall s^a \in [1, S^a] \rightarrow \bigcup_{s=1}^{S^a} P_{s^a}^o = P^o \ \& \ \bigcap_{s=1}^{S^a} P_{s^a}^o = \emptyset. \quad (5)$$

Свойства (5) подмножеств $P_{s^a}^o, s^a \in [1, S^0]$ обеспечивают корректность алгоритма замены пикселей и подразумевают, во-первых, полноту включения в их совокупность всех пикселей ИП, а во-вторых, невозможность принадлежности любого из них двум различным подмножествам одновременно.

Следует обратить внимание, что количество подмножеств $P_{s^a}^o$, определяемое индексными пределами в (5), равно размеру АП, то есть S^a . Это вызвано необходимостью использования всех пикселей АП $p_{s^a}^a \in P^a$ при аппроксимирующем преобразовании оригинального ММИ. Таким образом, механизм аппроксимации формируется следующим условием замены любого пикселя $p_{s^o}^o \in P^o$ из оригинального ММИ на один и тот же пиксель $p_{s^a}^a \in P^a$ из АП:

$$\forall s^o : p_{s^o}^o \in P^o \rightarrow p_{s^o}^o \cong p_{s^a}^o \in P_{s^a}^o \subset P^o. \quad (6)$$

Следовательно, вариативность и эффективность алгоритма преобразования оригинального ММИ в аппроксимированный ММИ целиком и полностью определяется двумя факторами:

- структурой разбиения ИП на подмножества $P_{s^a}^o$, то есть сколько и какие пиксели ИП входят в интервал, все элементы которого заменяются на $p_{s^a}^a$;
- количественными значениями $p_{s^a}^a$, составляющих АП, каждый из которых совпадает с одним из элементов $p_{s^a}^a = p_{s^o}^o \in P_{s^a}^o$,

поскольку значения элементов палитры всегда являются числами натурального ряда.

Конечный этап тоновой аппроксимации — оценка ее качества. Основой для расчета оценки является ММ изображения в цифровой среде. В растровой графике цифровое изображение представляется множеством P координатно привязанных к полю экрана точек (пикселей), содержащих информацию о цвете и/или о его тоне. Для монохромного (в смысле «одноцветного») изображения — ММИ — такое множество может быть представлено как матрица из r строк (row) и c столбцов (column) пикселей $\left(p_{sI}^I\right)_{ij}$:

$$P\left[s^I, i, j\right] = \begin{bmatrix} \left(p_{sI}^I\right)_{11} & \cdots & \left(p_{sI}^I\right)_{1c} \\ \vdots & \ddots & \vdots \\ \left(p_{sI}^I\right)_{r1} & \cdots & \left(p_{sI}^I\right)_{rc} \end{bmatrix}, \quad (7)$$

где символ p_{sI}^I — числовая характеристика яркости монохромного пикселя в палитре созданного изображения, отмеченного верхним индексом I , а нижние индексы i и j указывают на координатную привязку пикселей к строкам и столбцам соответственно.

Такая модель позволяет произвести «попиксельное» сравнение и расчет отклонения между пикселями аппроксимированного и оригинального ММИ. Наиболее часто используемым критерием оптимизации является минимум квадрата отклонения. Однако сравнение этого критерия с другими, в частности с критерием минимума модуля отклонения в работе [4], продемонстрировало преимущество использования именно этого нестандартного критерия. В работе он применяется в двух видах, приведенных далее. Так, для решения и исследования задач аппроксимации отдельных изображений, или нескольких изображений, имеющих одинаковый размер в пикселях, удобнее применять абсолютную форму выражения для вычисления критерия, так как тогда его значение всегда остается целым числом:

$$P_{\Sigma} = \sum_{i=1}^n \left[\sum_{j=1}^m \left| \left(p_{s^a}^a\right)_{ij} - \left(p_{s^o}^o\right)_{ij} \right| \right]. \quad (8a)$$

При обработке же, и особенно при сравнении качества аппроксимации разновеликих изображений, нужно использовать выражения для вычисления критерия в относительной форме:

$$\Delta p = \frac{1}{N} \cdot \sum_{i=1}^n \left[\sum_{j=1}^m \left| \left(p_{s^a}^a \right)_{ij} - \left(p_{s^o}^o \right)_{ij} \right| \right] \quad (86)$$

поскольку абсолютные значения суммарного отклонения тонов при разном количестве пикселей в изображениях несопоставимы. Тогда Δp является, в общем случае, вещественным, что требует при сравнении значений использовать максимальную точность представления числа.

Здесь $\left(p_{s^a}^a \right)_{ij}$ — пиксель созданного аппроксимированного ММИ, $\left(p_{s^o}^o \right)_{ij}$ — пиксель оригинального ММИ, $N = m \times n$ — количество пикселей изображения.

Принятие решений о качестве аппроксимации ММИ с привлечением объективно вычисляемых по формулам (8а) или (8б) оценок, а не визуальных критериев, позволяет проводить разработку алгоритмов, создаваемых для обработки изображений, на единой объективной критериальной основе.

3.2. Модифицированная под задачу аппроксимации ММИ генно-хромосомная структура ЭГА. Результативность алгоритма ЭГА в первую очередь определяется корректной интеграцией предметной задачи в эволюционное пространство на уровне генно-хромосомной структуры [23-38]. Многочисленные экспериментальные исследования различных модифицированных под сформулированную здесь предметную задачу вариантов этой структуры [17-19] показали целесообразность представления ее в следующем виде:

– особью (альтернативным решением) назначается аппроксимированный ММИ, то есть любой полученный аппроксимацией вариант ММИ — $P[s^l, i, j]$, представленный в виде (7);

– порождающая данную особь аппроксимирующая палитра P^a , заданная множеством (3), рассматривается как хромосома;

– гены — это отдельные тона ИП — $p_{s^a}^a$, из которых строится АП (3);

– подмножества $P_{s^a}^o$, заданные выражением (4), из которых выбираются варианты $p_{s^a}^a$, представляют собой аллели.

Особенность данной модели заключается в отсутствии непосредственного воздействия генетических операторов на матрицу аппроксимированного ММИ как таковую, что с точки зрения генетики является корректным, поскольку эта матрица рассматривается как некий конечный результат, полученный на основе структуры и элементов (генов) хромосомы, которая содержит всю наследственную информацию, позволяющую создать особь. Необходимо отметить, что в начальных исследованиях, например, подобного разделения не было, и в качестве хромосомы рассматривалась матрица аппроксимированного ММИ, что делало структуру алгоритма громоздкой, а потому малоэффективной. Исследовалась также двуххромосомная структура [18], которая оказалась избыточной. Исследования же новой модели гено-хромосомной структуры показали ее высокую эффективность в сравнении с первыми более абстрактными моделями [17, 18].

Хромосома, представленная АП, является с математической точки зрения вектором натуральных чисел размерностью s . В подразделе 2.1 показаны результаты выполненной с ее помощью аппроксимации выбранного для демонстрации изображения (рисунок 1). Использовалась хромосома размером $s = 8$, а аппроксимация осуществлялась различными алгоритмами.

Рассмотрим, например, вектор хромосомы, которая была получена в результате применения алгоритма медианного сечения для рисунка 1:

$$P_8^{AMC} = (15, 47, 80, 111, 142, 172, 205, 240)^T. \quad (9)$$

Среднее отклонение аппроксимированного изображения от исходного в тонах ИП, вычисленное по критерию (86) для хромосомы (9), составило 7,65. Для сравнения рассмотрим экстремальную хромосому, полученную посредством разработанного гибридного алгоритма для этого же изображения:

$$P_8^{GAC0} = (10, 25, 46, 67, 90, 119, 145, 191)^T. \quad (10)$$

Оценка качества хромосомы (10), используемой как АП, вычисленная по критерию (86), составляет 6,67. Это означает, что в среднем по изображению с помощью гибридного алгоритма субоптимизации получено улучшение аппроксимированного ММИ почти на тон по сравнению с результатом применения (9), и составило

12,8%. Совершенно очевидно, что структуры АП в виде хромосом (9) и (10) серьезно отличаются:

$$\Delta P_8^{AMC-ГАСО} = P_8^{AMC} - P_8^{ГАСО} = (5, 22, 34, 44, 52, 53, 60, 49)^T. \quad (11)$$

Показательно, что экстремальная хромосома (10) по всем тонам смещена относительно (9) в меньшую (т.е. в темную) сторону. Это связано с тем, что оригинальный ММИ (рисунок 1) использует преимущественно темные оттенки, что фиксируется на ее частотно-яркостной диаграмме (рисунок 6а).

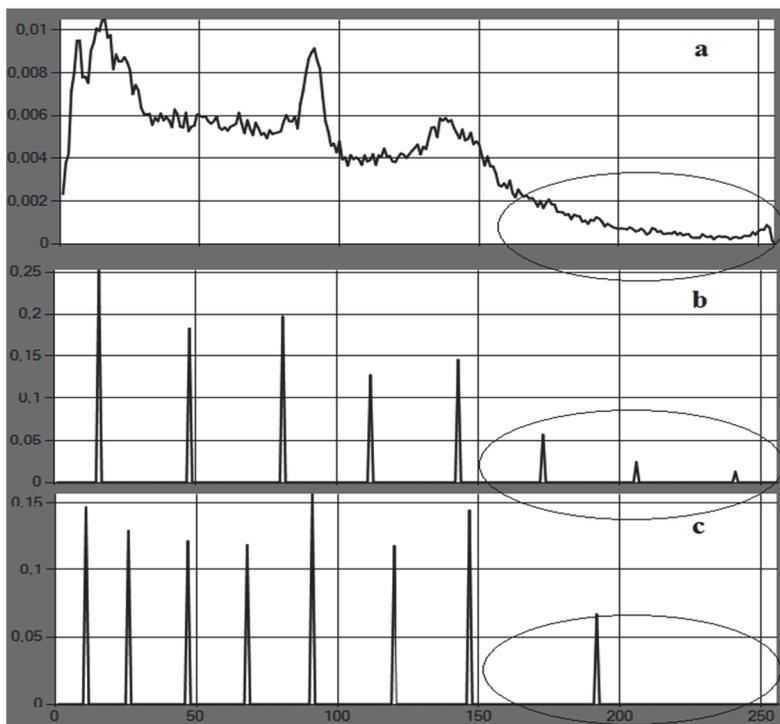


Рис. 6. Яркостные диаграммы: а) оригинального ММИ; б, с) двух аппроксимированных версий

Рассмотрим на рисунке 6б частотно-яркостную диаграмму аппроксимированного ММИ, порожденную хромосомой (9). Алгоритм равномерного распределения или медианного сечения, выделяет 3 тона АП участку, имеющему малое количество тонов (показан эллипсом на рисунке 6а, б, с), что создает нехватку тонов для достаточно

адекватного отображения большинства пикселей оригинального ММИ и приводит к визуальным отклонениям, представленным на демонстрационных примерах (см. рисунки 2, 4, 5). Отклонения рассчитаны по критерию оптимизации (8б).

Гибридный алгоритм субоптимизации в ходе поиска позволяет найти субоптимальную структуру АП критерию оптимизации (8а), что и позволяет минимизировать отклонения. На частотно-яркостной диаграмме аппроксимированного ММИ, порожденного хромосомой (10), видно, что большинство тонов АП сосредоточены на темной стороне палитры, а тот же участок светлых тонов аппроксимируется уже 1 тоном АП (рисунок 6с). Несомненно, при такой структуре АП основные потери приходятся на пиксели со светлыми оттенками, но для конкретного изображения (рисунок 1) эти отклонения не создают серьезных информативных потерь, так как не доминируют при передаче изображения.

Для дополнительной демонстрации эффективности предложенного метода гибридной субоптимизации АП над стандартным алгоритмом медианного сечения проведена обработка 30 разнородных по предметной области и размеру изображений, выбранных случайным образом. Анализ предполагает простое сравнение согласно критерию качества (8а) аппроксимированных изображений, полученных двумя методами.

Результаты проведенного сравнительного эксперимента продемонстрированы на гистограмме (рисунок 7), где по оси ординат отложен процент уменьшения отклонения (8а) аппроксимированного изображения от оригинального в результате применения гибридного алгоритма, а по оси абсцисс нумерация обработанных изображений. Изображения были отсортированы в порядке убывания процентной доли улучшения результата гибридного алгоритма в сравнении с алгоритмом медианного сечения.

В среднем предложенный гибридный алгоритм субоптимизации позволил обеспечить минимизацию отклонений согласно критерию (8а) на ~15% по сравнению с алгоритмом медианного сечения. Наибольшая минимизация отклонений составила ~58%, то есть по сравнению с медианным сечением гибридный алгоритм субоптимизации позволил улучшить качество тоновой аппроксимации более чем в два раза. Наименьшая минимизация отклонений составила 1%, что объясняется относительно равномерной частотно-яркостной диаграммой оригинального изображения, где медианное сечение, основанное на равномерном разбиении яркостной диаграммы на участки, способно давать эффективный результат.

Улучшение качества тоновой аппроксимации гибридным алгоритмом
в сравнении с медианным сечением

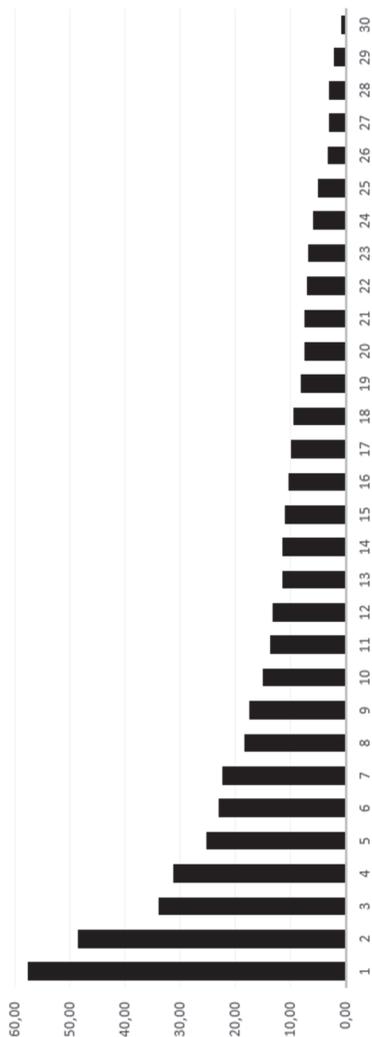


Рис. 7. Диаграмма с результатами проведенного сравнительного эксперимента

Для наглядности на рисунке 8 рассмотрим три частотно-яркостные диаграммы, где гибридным алгоритмом был получен наибольший выигрыш $\sim 58\%$ (рисунок 8а), средний $\sim 15\%$ (рисунок 8б) и наименьший $\sim 1\%$ (рисунок 8с). Частотно-яркостные диаграммы оригинальных изображений (рисунок 8) подтвердили взаимосвязь между низким процентом улучшения результата гибридным алгоритмом в сравнении с медианным сечением и относительной равномерностью яркостной диаграммы.

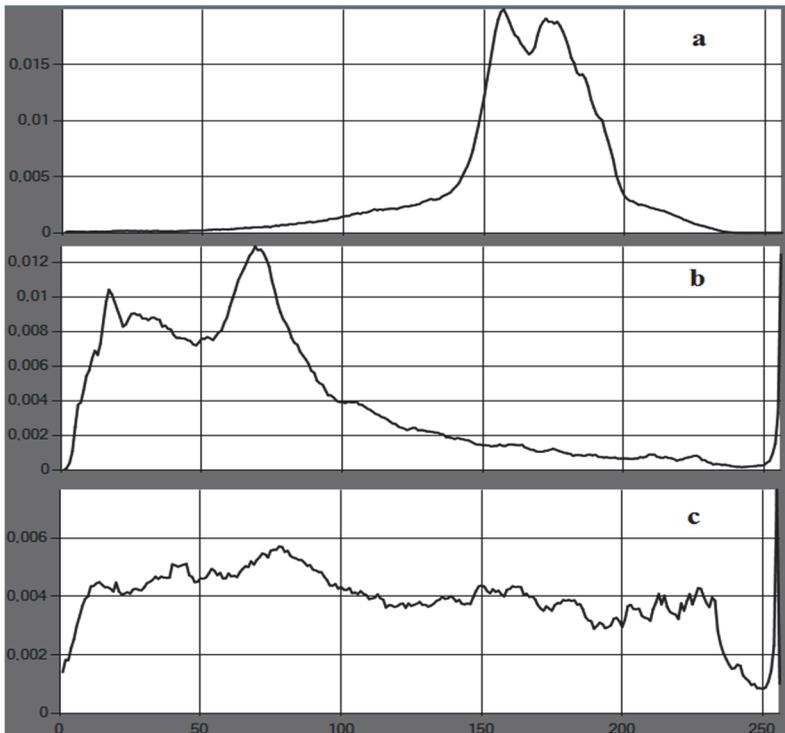


Рис. 8. Яркостные диаграммы трех оригинальных изображений (а, b, с), обработанных аппроксимирующими методами

Предложенная генно-хромосомная структура определяет одно из основных отличий модифицированного под решаемую задачу ЭГА от традиционно используемого алгоритма. Естественно, помимо генно-хромосомной весомое влияние на работу модифицированного ЭГА оказывают как разработанная структура алгоритма, так и реализация генетических операторов (см. раздел 2.3), а также настроечные параметры алгоритма (см. раздел 2.4).

3.3. Структура и традиционные механизмы модифицированного ЭГА. Работа ЭГА начинается с формирования исходной (материнской) популяции. Популяция в модифицированном ЭГА — это традиционно организуемая совокупность хромосом, создающих множество альтернативных решений — особой аппроксимированного ММИ. В решаемой задаче полностью случайное формирование популяции неэффективно, поскольку порождает множество заведомо плохих решений, что сильно удлиняет поиск.

Поэтому стартовая популяция формируется на основе некоей стартовой хромосомы, которая генерируется алгоритмом взвешенного распределения, представленного в работе [17]. В работе [19] показано, что оно в большинстве случаев дает по критерию (8а) лучший результат, нежели равномерное распределение.

После создания стартовой хромосомы в ее структуру начинают циклически (от первого до последнего гена) вводиться случайные изменения. При этом последние ограничиваются заданной долей возможного расстояния между соседними генами. Эта доля определяется специальной программной настройкой так называемого «исходного варьирования». Данная процедура повторяется до полного формирования стартовой популяции заданного размера.

Относящиеся к традиционным генетическим операторам скрещивание (кроссинговер) и мутация [23-38] также предусмотрены. Кроссинговер реализуется посредством обмена генов между двумя хромосомами, направленными на скрещивание, а мутация вносит случайные изменения в генную структуру хромосом. В качестве механизма селекции использован простой и эффективный, а потому популярный метод рулетки [23-38].

Результатом цикла воздействия селекции и генетических операторов является формирование нового поколения хромосом. Соответственно, заданное число поколений определяет количество этих циклов. Простейшим условием завершения алгоритма является достижение последнего поколения.

При исследованиях в работе [18] выявлено, что наибольший коэффициент значимости имеет показатель исходного варьирования, то есть формирование начальной популяции задает определяющее значение на ход поиска. Для эффективного использования данного аспекта было принято решение реализовать и внедрить «перезапускающуюся» модель, которая частично похожа на модель СНС, описанная Эшелманом [39]. Неполная реализация модели предполагает использование малочисленной популяции, а также малого количества поколений, но выполнение повторного запуска

алгоритма, где стартовой хромосомой для формирования исходной популяции назначается лучшая хромосома предыдущего цикла поиска. Данный подход позволил резко повысить эффективность алгоритма [19] по сравнению с предыдущей, более каноничной реализацией [17, 18]. Реализованная схема фактически производит обширное корректирующее изменение популяции через заданные промежутки поколений. Параметр, определяющий количество перезапусков, получил название «количество поисковых итераций».

Таким образом, приведенный краткий обзор модифицированного ЭГА, разработанного под задачу оптимизации тоновой аппроксимации ММИ, показывает, что многие параметры алгоритма являются настроечными и сами по себе требуют оптимизации. На уровне основных настроечных параметров генетических механизмов в данной работе (в текущем исследовании) рассматриваются возможности повышения эффективности работы всего гибридного алгоритма за счет настройки параметров модифицированного ЭГА. Под эффективностью гибридного алгоритма понимается достижение его оптимального или субоптимального быстродействия при сохранении гарантированной точности отыскания оптимума критерия точности. Это достигается такой настройкой ЭГА, чтобы он затрачивал минимальное количество поисковых итераций, достаточное для быстрого выхода гибридного алгоритма в экстремальную зону, что позволяет сократить общее время поиска. В результате общее вычислительное время остается единственным критерием оценки эффективности гибридного алгоритма (см. 2.2).

Основным условием оценки результатов таких исследований должна быть ее объективность, то есть она должна опираться не на визуальные, а на количественные методы с использованием критериев (8a) или (8б).

4. Экспериментальные исследования эффективности гибридного алгоритма субоптимизации при использовании модифицированного ЭГА. На основании описанной структуры (см. раздел 2.3) можно выделить две группы основных настроечных параметров модифицированного ЭГА — количественные факторы и вероятностные факторы.

4.1. Настроечные параметры модифицированного ЭГА в составе гибридного алгоритма субоптимизации. Природа поведения количественных факторов поддается анализу значительно легче, чем вероятностных, поскольку чем больше их значения, тем, как правило, более эффективен поиск модифицированного ЭГА с точки зрения

точности. Однако увеличение количественных факторов прямо пропорционально увеличивает вычислительное время модифицированного ЭГА. Это связано с увеличением числа вычислений функции приспособленности (8а) или (8б).

В разработанной модели модифицированного ЭГА под рассматриваемую предметную задачу использовано три количественных фактора:

- размер популяции;
- количество поколений;
- количество поисковых итераций алгоритма.

В настоящий момент самой актуальной работой является [22], где были получены субоптимальные значения размера популяции — 12, количества поколений 10 и количества поисковых итераций 12.

К вероятностным факторам в модифицированного ЭГА относятся следующие параметры:

- вероятность кроссинговера генов;
- вероятность мутации гена;
- амплитуда мутации гена.
- исходное варьирование генов.

В работе [40] был спланирован и проведен эксперимент по первичному исследованию потенциальной возможности субоптимизации вероятностных факторов модифицированного ЭГА с целью повышения эффективности работы гибридной модели в целом. Исследовались три различных набора значений вероятностных факторов, выбранных относительно субоптимальных значений, которые были получены до структурных изменений алгоритма. Условно эти наборы можно разделить по степени агрессивности их влияния: высокая, средняя и малая. Значения исследованных в работе [40] вероятностных факторов представлены в таблице 1.

Таблица 1. Исследованные значения вероятностных факторов

Степень агрессивности модифицированного ЭГА	Вероятность кроссинговера Генов	Вероятность мутации гена	Амплитуда мутации Гена	Исходное варьирование генов
Большая	20	20	3	0.3
Средняя	10	10	2	0.2
Малая	5	5	1	0.1

При проведении эксперимента значения количественных факторов, кроме количества поисковых итераций, были зафиксированы на их субоптимальных уровнях, полученных в работе [22]. Количество же итераций было увеличено до 50, чтобы иметь возможность оценить

поведение алгоритма и характер влияния вероятностных факторов на более длительной протяженности эволюционного процесса.

Анализ по трем изображениям продемонстрировал разнородность результатов в связи с сильным влиянием пространства решений конкретного изображения на поведение того или иного набора вероятностных факторов. Таким образом, выяснилось, что возможность выбора одной неизменной стратегии решения поставленной задачи маловероятна. В связи с этим в работе [40] была предложена идея разработки адаптивной или смешанной схемы изменения вероятностных факторов. Описываемое в данной статье исследование посвящено оценке ее результативности и перспективности.

4.2. Исследование влияния вероятностных факторов на точность аппроксимации изображений модифицированным ЭГА. Предварительное исследование предпринято для приближенной оценки влияния настроек вероятностных факторов на быстродействие модифицированного ЭГА при поиске экстремума. Суть исследования состояла в распределении между тремя популяциями значений вероятностных факторов, представленных в таблице 1, осуществляя пошаговый поиск экстремума этими популяциями в параллельном режиме. При этом для каждого последующего шага в качестве выходного значения использовалась наилучшая хромосома из трех параллельно обработанных популяций. Описанная схема получила условное название «трехходовки».

При исследовании «трехходовки» на каждом новом шаге осуществлялась запись порядкового номера того набора вероятностных факторов, который дал наилучший результат. Это позволило определить потенциальные возможности фиксированной схемы их изменения от шага к шагу.

Тестирование проведено на ПК с процессором Intel Core i7 4700MQ 2,4 GHz (up to 3,4 GHz). При этом время одновременной обработки трех популяций по сравнению с одной («одноходовкой») увеличилось на ~35%. Для эксперимента отобрано 6 ММИ с различными частотно-яркостными диаграммами. Все изображения обрабатывались алгоритмом оптимизационной тоновой аппроксимации стандартной палитры 256-ти тонов 8-тоновой палитрой. Последнее связано с тем, что в работе [19] было показано, что 8-тоновая палитра структурно близка к оптимальным многоцелевым аппроксимирующим свойствам.

Эксперимент состоял в тестировании алгоритма аппроксимации на 6 выбранных изображениях производственных помещений по

описанной схеме «трехходовки». Каждое изображение протестировано 30 раз. Для оценки эффективности трехходовки изображения подвергались такому же тестированию, но по схеме «одноходовки», то есть с реализацией модифицированного ЭГА, использующей только один фиксированный набор вероятностных факторов (без параллелизма).

Результаты исследования представлены на графиках изменения среднего значения критерия оптимизации (8а) по выборке из 30 опытов (рисунок 9). По оси ординат отложены значения критерия оптимизации, а по оси абсцисс итерация алгоритма.

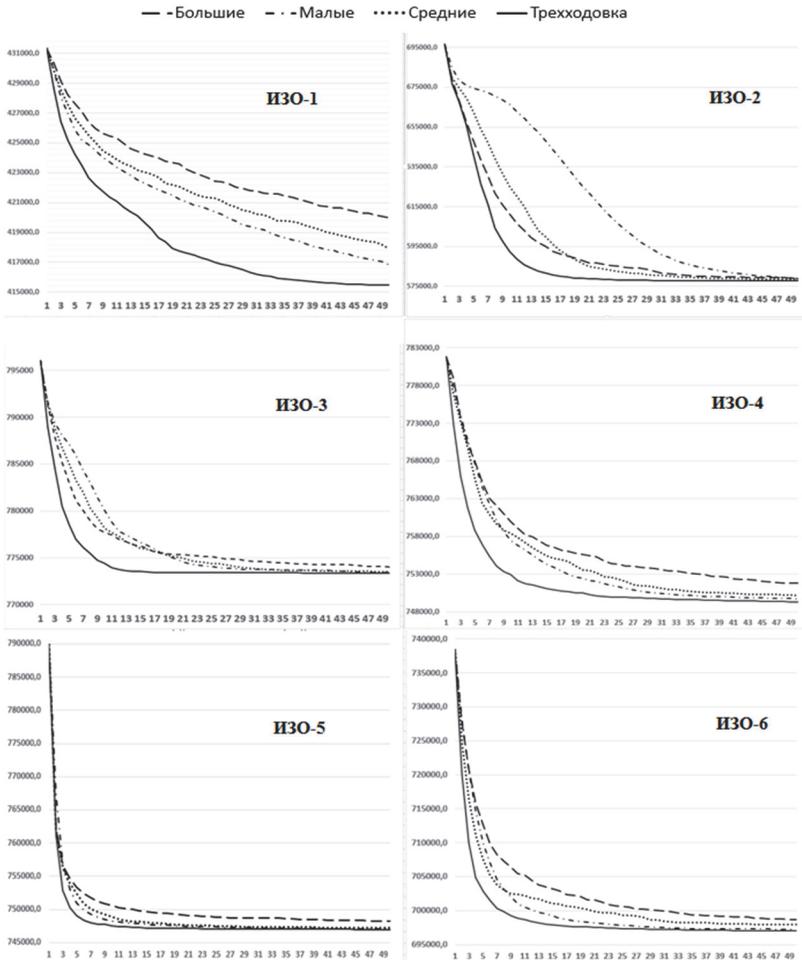


Рис. 9. Графики с результатами эксперимента

Графики отчетливо демонстрируют повышение эффективности работы модифицированного ЭГА при использовании трехходовки. Это проявилось как в получении наилучшего критерия точности аппроксимации, так и в существенно большем быстродействии. Алгоритм трехходовки во всех случаях, кроме ИЗО-1, позволил выйти в экстремальную область не хуже, чем к ~ 15 итерации модифицированного ЭГА (см. рисунок 9).

В связи с тем, что одноходовая схема применения модифицированного ЭГА, использующая только большие вероятностные факторы, показала наихудший результат по всем изображениям, она используется для оценки эффективности остальных схем модифицированного ЭГА с точки зрения быстродействия. Иными словами, предлагается взять за оценку шаг, на котором результаты остальных реализаций алгоритма, модифицированного ЭГА, превзошли финальный результат одноходовки с большими значениями вероятностных факторов. Номера шагов и значения показателя (8а), удовлетворяющих этому условию, приведены в таблице 2.

Таблица 2. Значение оптимизации (8а) и шаги, на которых указанные реализации алгоритма превзошли результат больших вероятностных факторов на 50 шаге

ВФ	ИЗО-1	ИЗО-2	ИЗО-3	ИЗО-4	ИЗО-5	ИЗО-6	Сред.
Больш.	420001	579230	774042	751804	748231	698644	661992
	50	50	50	50	50	50	50
Сред.	419801	579076	774001	751579	748213	698637	-107
	33	36	28	25	12	27	26,8
Малые	419949	579015	773960	751725	748164	698484	-109
	26	49	25	22	10	17	24,8
3-ходовка	419671	579015	773985	751748	748031	698639	-144
	14	19	10	11	6	10	11,6 (15,66)

Результаты таблицы 2 демонстрируют преимущество трехходового алгоритма. Средний шаг по всем изображениям при этой схеме равен 11,6 шагам, тогда как одноходовка с малыми вероятностными факторами имеет показатель 24,8 шага. Однако необходимо обратить внимание, что 11,6 шагов трехходовки является фактическим значением, тогда как для сравнения необходимо использовать эквивалентное значение к одноходовке, поскольку один шаг трехходовки требует на 35% больше времени. Эквивалентное значение указано в скобках (см. таблица 2,

ст. «Сред.») Таким образом, трехходовка в среднем позволяет сократить время обработки на ~37%.

Анализ всех архивных файлов трехходовки с отмеченным набором вероятностных факторов, дававшим лучший результат на конкретном шаге, показал, что в процессе работы алгоритма по мере приближения к экстремуму все более эффективным становится настройка с малыми вероятностными факторами. Это связано с необходимостью сужать область поиска. Однако из-за разнородности пространства решений у каждого изображения фиксированная система изменения вероятностных факторов не позволяет универсально получать улучшения для любых изображений. Это объясняет преимущество трехходовки в проведенном эксперименте и дает основание ориентироваться в решаемой задаче на параллельно конкурирующие алгоритмы. Конкретные решения, как по их количеству, так и по значениям конкурирующих вероятностных факторов, являются предметом дополнительных исследований.

Что касается описываемого в данной статье исследования, то ввиду очевидного и устойчивого преимущества трехходовки по сравнению с использованными наборами вероятностных факторов в отдельности, а также с учетом обеспечения схемой трехходовки выхода на экстремальную зону уже к ~15 шагу принято решение внедрить разработанную схему трехходовки в гибридный алгоритм, чтобы сократить количество итераций модифицированного ЭГА, что, как ожидалось, обеспечит серьезный выигрыш в быстродействии.

4.3. Субоптимизация количества итераций модифицированного ЭГА в гибридном алгоритме субоптимизации. Полученные результаты исследования эффективности трехходовой схемы работы модифицированного ЭГА дали основания на проведение дополнительного исследования уже модернизированного гибридного алгоритма с целью субоптимизации количества поисковых итераций. Исходя из анализа графиков работы трехпопуляционного алгоритма (рисунок 9) выбраны 3 значения — 12, 16 и 20 шагов. Эксперимент предусматривал 10-кратную обработку изображений гибридным алгоритмом.

Главным критерием оценки в новом эксперименте выбрано общее время работы гибридного алгоритма, поскольку экстремальность результата гарантируется этапом детерминированного поиска. Это связано с тем, что основной временной ресурс приходится на работу детерминированного алгоритма, и общее время выполнения гибридного алгоритма зависит

от точности результата работы модифицированного ЭГА. Данное обстоятельство подразумевает приоритет близости этого результата к экстремуму по сравнению с количеством поисковых итераций.

Результаты эксперимента продемонстрированы в таблице 3, где указано среднее время обработки гибридного алгоритма для каждого изображения при соответствующем количестве поисковых итераций по данным 10 тестов.

Рассмотрение среднего показателя времени по всем изображениям (см. таблица 3, строка 8) позволяет остановиться на выборе 20 итераций модифицированного ЭГА для сокращения времени обработки гибридного алгоритма в среднем по совокупности изображений. Проведенный эксперимент показывает, что выбранная настройка позволяет сократить время работы гибридной модели в среднем на 14,5 секунд по сравнению с 12 итерациями, что составляет 13,9% выигрыша во времени, а по сравнению с 16 на 13,2 секунды, то есть сокращение общего времени на 12,8%. Что же касается СКО, то оно закономерно уменьшается с ростом числа итераций. При более тонком исследовании задачи этот показатель может быть уточнен.

Таблица 3. Результаты эксперимента по исследованию количества поисковых итераций (КПИ) модифицированного ЭГА

КПИ модифицированного ЭГА	12	16	20	24	28
Среднее время по ИЗО-1 (сек)	106,3	107,4	75,3	94,4	85,4
Среднее время по ИЗО-2 (сек)	173,1	153,4	114,1	97,2	97,4
Среднее время по ИЗО-3 (сек)	75,1	78,7	80,8	86,9	98,2
Среднее время по ИЗО-4 (сек)	112,7	107,1	101,2	100,4	89,6
Среднее время по ИЗО-5 (сек)	70,5	74,1	76,2	75,9	82,6
Среднее время по ИЗО-6 (сек)	85,3	94,3	88,1	88,8	96,8
Среднее время по всем ИЗО	103,8	102,5	89,3	90,6	91,7
СКО по всем ИЗО	37,8	28,5	15,4	8,8	6,7

Среднее время гибридного алгоритма при 24 итерациях возросло на ~ 1,3 секунды по сравнению с 20 шагами, хотя СКО уменьшилось. Дальнейшее повышение количества поисковых итераций приводит к незначительному увеличению общего времени алгоритма и сокращению СКО. Такое поведение полностью согласуется с ранее описанным характером распределения временной составляющей между эвристическим и детерминированным методом. Вероятно, при более тонком исследовании задачи показатель 20 необходимо уточнить.

Следует отметить, что по изображениям ИЗО-3, ИЗО-5 и ИЗО-6 при 12 итерациях был получен несколько лучший результат (таблица 3, строки 4, 6, 7) в сравнении с другими. Однако выигрыш в

этих опытах был незначительным, а по остальным изображениям увеличение количества поисковых итераций сильно ухудшало время обработки. Причина выявленного феномена становится ясной при детальном анализе графиков на рисунке 2. В упомянутых изображениях выход в экстремальную область происходит быстрее других, и 12 итераций оказывается вполне достаточно. Но так как вычислительное время модифицированного ЭГА мало относительно всей гибридной модели, то увеличение до 20 итераций не приводит к сильному возрастанию общего времени поиска. При этом в случае неблагоприятной пиксельной структуры изображений малое количество поисковых итераций уже не обеспечивает выхода в экстремальную область, и общее время работы гибридного алгоритма резко возрастает на этапе детерминированного поиска экстремума.

Учитывая значительное различие топологий пространств решения разных изображений, в последующих исследованиях целесообразно рассмотреть вопрос разработки адаптивных систем изменения, как количества поисковых итераций, так и других настроечных параметров модифицированного ЭГА.

5. Заключение. Основные результаты проведенного экспериментального исследования можно оценить следующими положениями:

1. Экспериментально показана невозможность построения универсального и эффективного алгоритма решения задачи тоновой аппроксимации монохромных растровых изображений с использованием ЭГА с фиксированными значениями вероятностных факторов его настройки.

2. Реализованная при модификации ЭГА параллельно работающая трехпопуляционная модель при использовании экспериментально подобранных настроек позволяет обеспечить уменьшение количества поисковых итераций в среднем на ~37% по сравнению с наилучшей однопопуляционной моделью.

3. Как результат этого в предложенном гибридном алгоритме модифицированный ЭГА решает задачу поиска близкой к оптимальной стартовой структуры аппроксимирующей палитры для быстрого поиска ее ближайшего экстремума, обеспечивая оптимальность или субоптимальность тоновой аппроксимации, повышая быстродействие гибридного алгоритма в целом на ~14%.

Литература

1. *Brun L., Treméau A.* Digital Color Imaging Handbook // The Electrical Engineering and Applied Signal Processing Series. NYC: CRC Press. 2003. 764 p.
2. *Shapiro L., Stockman G.* Computer Vision // Prentice Hall PTR. 2001. 609 p.

3. *Emre C.* Improving the Performance of K-Means for Color Quantization // *Image and Vision Computing*. 2011. vol. 29. pp. 260–271.
4. *Burger W., Burger M.* Color Quantization // *Digital Image Processing. Texts in Computer Science*. 2016. pp. 329–339.
5. *Yue X.D. et al.* An efficient color quantization based on generic roughness measure // *Pattern Recognition*. 2014. vol. 47. pp. 1777–1789.
6. *Kanan C., Cottrell G.* Color-to-Grayscale: Does the Method Matter in Image Recognition? // *PLoS One*. 2012. vol. 7. 7 p. URL: <https://doi.org/10.1371/journal.pone.0029740> (дата обращения: 02.07.2018).
7. *Князь В.А. и др.* Технологии интеллектуальной обработки информации для задач навигации и управления беспилотными летательными аппаратами // *Труды СПИИРАН*. 2016. Вып. 2(45). С. 26–44.
8. *Padmavathi K., Thangadurai K.* Implementation of RGB and Grayscale Images in Plant Leaves Disease Detection – Comparative Study // *Indian Journal of Science and Technology*. 2016. vol. 9. pp. 1–6.
9. *Hu Y.-C., Chen W.-L., Lj C.-C., Chuang J.-C.* Improved vector quantization scheme for grayscale image compression // *Opto-Electronics Review*. 2012. vol. 20. pp. 187–193.
10. *Ramirez E., Jimenez O., Perez A., Pogrebnyak O.* Grayscale Image Segmentation Based on Associative Memories // *Computations in Systems*. 2011. vol. 15. pp. 149–162.
11. *Харинов М.В.* Обобщение трех подходов к оптимальной сегментации цифрового изображения // *Труды СПИИРАН*. 2013. Вып. 2(25). С. 294–316.
12. *Kharinov M.* Reclassification formula that provides to surpass K-means method // *Computer Vision and Pattern Recognition*. 2012. 10 p. URL: <https://arxiv.org/ftp/arxiv/papers/1209/1209.6204.pdf> (дата обращения: 02.07.2018).
13. *Деревянкина А.А.* Автоматизация исследования изображений методом s-аппроксимации // *Математические методы в технике и технологиях (ММТТ)*. 2010. Т. 6. С. 37–43.
14. *Patel T., Srivastava R.* Hierarchical visual cryptography for grayscale image // *Online International Conference on Green Engineering and Technologies (IC-GET)*. 2016. pp. 1–4.
15. *Zhang S., Wu Y., Bao Y., Bai J.* Watermarking Algorithm for Bas-Relief Based on Depth Grayscale Image // *International Conference on Computing Intelligence and Information System (CIIS)*. 2017. pp. 294–297.
16. *Heckbert P.* Color image quantization for frame buffer display // *Proceedings of the 9th annual conference on Computer graphics and interactive techniques (SIGGRAPH'82)*. 1982. pp. 297–307.
17. *Neydorf R.A., Aghajanyan A.G., Vucinic D.* Monochrome Multitone Image Approximation on Lowered Dimension Palette with Sub-optimization Method based on Genetic Algorithm // *Improved Performance of Materials*. 2018. pp. 144–154.
18. *Neydorf R.A., Aghajanyan A.G., Vucinic D.* Monochrome multitone image approximation with low-dimensional palette // *IEEE East-West Design & Test Symposium (EWDTS)*. 2016. pp. 1–4.
19. *Нейдорф Р.А., Агаджанян А.Г.* Исследование аспектов возможного применения субоптимальной тоновой аппроксимации изображений в задачах технического зрения средств автономной навигации // *Известия ЮФУ. Технические науки*. 2017. № 1-2(186-187). С. 133–145.
20. *Нейдорф Р.А., Агаджанян А.Г., Нейдорф А.П.* Оптимизация результатов аппроксимации растровых изображений и оценка их экстремальности // *Математические Методы в Технике и Технологиях*. 2017. Т. 1. С. 19–26.

21. *Neydorf R.A., Aghajanyan A.G., Vucinic D.* A high-speed hybrid algorithm of monochrome multitone images approximation // IEEE East-West Design & Test Symposium (EWDTS). 2017. pp. 1–4.
22. *Neydorf R.A., Aghajanyan A.G., Vucinic D.* Improved Bi-optimal Hybrid Approximation Algorithm for Monochrome Multitone Image Processing // The Eleventh International Conference on Advanced Engineering Computing and Applications in Sciences (ADVCOMP 2017). 2017. pp. 20–25.
23. *Luke S.* Essentials of Metaheuristics // Lulu. 2013. 227 p. URL: <https://cs.gmu.edu/~sean/book/metaheuristics/Essentials.pdf> (дата обращения: 17.03.2018).
24. *Mitchell M.* An Introduction to Genetic Algorithms // MIT press 1999. 221 p.
25. *Eiselt H.A., Sandblom C-L.* Heuristic Algorithms // Integer Programming and Network Models. 2000. pp. 229–258.
26. *Cook S.A.* An overview of computational complexity // Communication of the ACM. 1983. vol. 26. no. 6. pp. 401–408.
27. *Deb K., Pratap A.* A fast and elitist multiobjective genetic algorithm: NSGA-II // IEEE Transactions on Evolutionary Computation. 2002. vol. 6. pp. 182–197.
28. *Deb K., Himanshu J.* An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach. Part I: Solving Problems with Box Constraints // IEEE Transactions on Evolutionary Computation. 2014. vol. 18. pp. 577–601.
29. *Sun J.Q., Schütze O.* A hybrid evolutionary algorithm and cell mapping method for multi-objective optimization problems // 2017 IEEE Symposium Series on Computational Intelligence (SSCI). 2017. pp. 1–9.
30. *Fogel D.B.* An introduction to simulated evolutionary optimization // IEEE Transactions on Neural Networks. 1994. vol. 5. pp. 3–14.
31. *Roeva O.* Genetic Algorithm and Firefly Algorithm Hybrid Schemes for Cultivation Processes Modelling // Transactions on Computational Collective Intelligence XVII. 2014. vol. 8790. pp. 196–211.
32. *Neumüller C., Wagner S.* Parameter Meta-optimization of Metaheuristic Optimization Algorithms // International Conference on Computer Aided Systems Theory. 2012. vol. 6927. pp. 367–374.
33. *Chekanin V.A., Chekanin A.V.* Design of Library of Metaheuristic Algorithms for Solving the Problems of Discrete Optimization // Advances in Mechanical Engineering. 2018. pp. 25–32.
34. *Podorozhniak A., Lubchenko N., Balenko O., Zhuikov D.* Neural network approach for multispectral image processing // Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET). Ukraine. Lviv-Slavvansk. 2018. pp. 978–981.
35. *Grishkin V., Lakushkin O., Stepanenko N.* Biofouling detection based on image processing technique // 2017 Computer Science and Information Technologies (CSIT). 2017. pp. 158–161.
36. *Gillette A., Wilson C., George A.* Efficient and autonomous processing and classification of images on small spacecraft // 2017 IEEE National Aerospace and Electronics Conference (NAECON). 2017. pp. 135–141.
37. *Nikam P.A., Sawant S.D.* Circuit board defect detection using image processing and microcontroller // 2017 International Conference on Intelligent Computing and Control Systems (ICICCS). 2017. pp. 1096–1098.
38. *Schütze O., Hernandez V.* The hypervolume based directed search method for multi-objective optimization problems // Journal of Heuristics. 2016. vol. 22. pp. 273–300.

39. *Eshelman L.J.* The CHC adaptive search algorithm: how to safe search when engaging in non traditional genetic recombination // *Foundations of Genetic Algorithms*. 1991. pp. 265–283.
40. *Агаджанян А.Г.* Анализ влияния вероятностных параметров настройки эволюционно-генетического алгоритма на результативность оптимизационной аппроксимации изображений // *Техника и технологии: курс на инновации: сборник материалов международной научно-практической конференции*. Иркутск: «Научное партнерство Алпекс». 2017. С. 49–57.

Нейдорф Рудольф Анатольевич — д-р техн. наук, профессор, профессор кафедры программного обеспечения вычислительной техники и автоматизированных систем, Донской Государственный Технический Университет. Область научных интересов: математическое моделирование, методы оптимизации и преобразования информации, эвристические методы исследований. Число научных публикаций — 379. rudolf.neydorf.44@mail.ru; пл. Гагарина 1, Ростов-на-Дону, 344000; р.т.: +79034722292.

Агаджанян Альберт Грантович — аспирант кафедры программного обеспечения вычислительной техники и автоматизированных систем, Донской Государственный Технический Университет. Область научных интересов: эвристические алгоритмы, эволюционные методы оптимизации, обработка цифровых данных. Число научных публикаций — 15. foralbert92@gmail.com; пл. Гагарина 1, Ростов-на-Дону, 344000; р.т.: +79515050716.

R.A. NEYDORF, A.G. AGHAJANYAN
**DUAL OPTIMIZATION OF MONOCHROME IMAGES TONE
APPROXIMATION USING PARALLEL EVOLUTIONARILY
GENETIC SEARCH**

Neydorf R.A., Aghajanyan A.G. Dual Optimization of Monochrome Images Tone Approximation using Parallel Evolutionarily Genetic Search.

Abstract. The paper considers the optimization problem of tone approximation for monochrome (for example: in grayscale palette) images. The procedure of tone approximation implies the reduction of approximated image's number of tones, which are used in image displaying, compared to number of tones in the original image. The point of the procedure optimization consists of minimization of visual quality losses that estimated according to total or mean deviation between the same pixels of original image and approximated one. As a tool of the optimization the hybrid algorithm is used. It was developed and investigated by authors. The hybrid algorithm combines heuristic and deterministic algorithms of searching the best structure of approximating palette according to criterion of deviations minimization. The heuristic algorithm is based on evolutionarily-genetic paradigm. The main goal of heuristic stage is the reduction of search area of approximating palette's structures that are the closest to optimum. Such role for heuristic stage was defined according to its fast computational time. The goal of deterministic algorithm of directed exhaustive search is to find the nearest extreme for the result that was obtained by previous algorithm. The developed hybrid algorithm allows to provide dual optimization of tone approximation. It means that the algorithm provides a result, in which two different criteria become optimal relative to each other. The current investigation is devoted to consideration of possibility to increase the effectiveness of hybrid algorithm on the level of heuristic stage. The possibility of implementation the parallel model of evolutionarily-genetic algorithm with different settings is considered. The results of initial experiments are discussed and compared with known algorithm of tone approximation.

Keywords: tone approximation, hybridization, parallel model, optimization, adaptive scheme, evolutionarily-genetic algorithm, approximating palette.

Neydorf Rudolf Anatolyevich — Ph.D., Dr. Sci., professor, professor of hardware and software engineering department, Don State Technical University. Research interests: mathematical modeling, optimization and processing of information, heuristic methods of investigation. The number of publications — 379. rudolf.neydorf.44@mail.ru; 1, Gagarin square, Rostov-on-Don, 344000, Russia; office phone: +79034722292.

Aghajanyan Albert Grantovich — Ph.D. student of hardware and software engineering department, Don State Technical University. Research interests: heuristic algorithms, evolutionary methods of optimization, processing of digital data. The number of publications — 15. foralbert92@gmail.com; 1, Gagarin square, Rostov-on-Don, 344000, Russia; office phone: +79515050716.

References

1. Brun L., Tremeau A. Digital Color Imaging Handbook. The Electrical Engineering and Applied Signal Processing Series. NYC: CRC Press. 2003. 764 p.
2. Shapiro L., Stockman G. Computer Vision. Prentice Hall PTR. 2001. 609 p.
3. Emre C. Improving the Performance of K-Means for Color Quantization. *Image and Vision Computing*. 2011. vol. 29. pp. 260–271.
4. Burger W., Burge M. Color Quantization. Digital Image Processing. Texts in Computer Science. 2016. pp. 329–339.

5. Yue X.D. et al. Chen An efficient color quantization based on generic roughness measure. *Pattern Recognition*. 2014. vol. 47. pp. 1777–1789.
6. Kanan C., Cottrell G. Color-to-Grayscale: Does the Method Matter in Image Recognition? *PLoS One*. 2012. vol. 7. 7 p. Available at: <https://doi.org/10.1371/journal.pone.0029740> (accessed: 02.07.2018).
7. Knyaz V.A. et al. [Intelligent Data Processing Technologies for Unmanned Aerial Vehicles Navigation and Control]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2016. vol. 2(45). pp. 26–44. (In Russ.).
8. Padmavathi K., Thangadurai K. Implementation of RGB and Grayscale Images in Plant Leaves Disease Detection – Comparative Study. *Indian Journal of Science and Technology*. 2016. vol. 9. pp. 1–6.
9. Hu Y.-C., Chen W.-L., Lo C.-C., Chuang J.-C. Improved vector quantization scheme for grayscale image compression. *Opto-Electronics Review*. 2012. vol. 20. pp. 187–193.
10. Ramirez E., Jimenez O., Perez A., Pogrebnyak O. Grayscale Image Segmentation Based on Associative Memories. *Computations in Systems*. 2011. vol. 15. pp. 149–162.
11. Kharinov M. [A generalization of three approaches to an optimal segmentation of digital image]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2013. vol. 2(25). pp. 294–316. (In Russ.).
12. Kharinov M. Reclassification formula that provides to surpass K-means method. *Computer Vision and Pattern Recognition*. 2012. 10 p. Available at: <https://arxiv.org/ftp/arxiv/papers/1209/1209.6204.pdf> (accessed: 02.07.2018).
13. Derevyankina A.A. [Automation of images investigation by the s-approximation method] *Matematicheskie Metodi v Texnike i Texnologiyax (MMTT)* [Mathematical methods in technics and technologies (MMTT)]. 2010. Issue 6. pp. 37–43. (In Russ.).
14. Patel T., Srivastava R., Hierarchical visual cryptography for grayscale image. Online International Conference on Green Engineering and Technologies (IC-GET). 2016. pp. 1–4.
15. Zhang S., Wu Y., Bao Y., Bai J. Watermarking Algorithm for Bas-Relief Based on Depth Grayscale Image. International Conference on Computing Intelligence and Information System (CIIS). 2017. pp. 294–297.
16. Heckbert P. Color image quantization for frame buffer display. Proceedings of the 9th annual conference on Computer graphics and interactive techniques (SIGGRAPH'82). 1982. pp. 297–307.
17. Neydorf R.A., Aghajanyan A.G., Vucinic D. Monochrome Multitone Image Approximation on Lowered Dimension Palette with Sub-optimization Method based on Genetic Algorithm. *Improved Performance of Materials*. 2018. pp. 144–154.
18. Neydorf R.A., Aghajanyan A.G., Vucinic D. Monochrome multitone image approximation with low-dimensional palette. IEEE East-West Design & Test Symposium (EWDTS). 2016. pp. 1–4.
19. Neydorf R.A., Aghajanyan A.G. [The Research Of The Application Possibilities Of Tones Approximation In A Technical Vision For The Autonomous Navigation Objects]. *Izvestia SFEDU. Texnicheskie nauki – Izvestiya SFedU. Engineering Sciences*. 2017. vol. 1-2(186-187). pp. 133–145. (In Russ.).
20. Neydorf R.A., Aghajanyan A.G., Neydorf A.R. [Optimization of approximation results of halftone images and assessment of their extremality] *Matematicheskie Metodi v Texnike I Texnologiyax* [Mathematical models in technics and technologies]. 2017. Issue 1. pp. 19–26. (In Russ.).
21. Neydorf R.A., Aghajanyan A.G., Vucinic D. A high-speed hybrid algorithm of monochrome multitone images approximation. IEEE East-West Design & Test Symposium (EWDTS). 2017. pp. 1–4.
22. Neydorf R.A., Aghajanyan A.G., Vucinic D. Improved Bi-optimal Hybrid Approximation Algorithm for Monochrome Multitone Image Processing. The Eleventh International Conference on Advanced Engineering Computing and Applications in Sciences (ADVCOMP 2017). 2017. pp. 20–25.

23. Luke S. Essentials of Metaheuristics. Lulu. 2013. 227 p. Available at: <https://cs.gmu.edu/~sean/book/metaheuristics/Essentials.pdf> (accessed: 17.03.2018).
24. Mitchell M. An Introduction to Genetic Algorithms. MIT press 1999. 221 p.
25. Eiselt H.A., Sandblom C-L. Heuristic Algorithms. Integer Programming and Network Models. 2000. pp. 229–258.
26. Cook S.A. An overview of computational complexity. *Communication of the ACM*. 1983. vol. 26. no. 6. pp. 401–408.
27. Deb K., Pratap A. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*. 2002. vol. 6. pp. 182–197.
28. Deb. K., Himanshu J. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach. Part I: Solving Problems with Box Constraints. *IEEE Transactions on Evolutionary Computation*. 2014. vol. 18. pp. 577–601.
29. Sun J.Q., Schütze O. A hybrid evolutionary algorithm and cell mapping method for multi-objective optimization problems. 2017 IEEE Symposium Series on Computational Intelligence (SSCI). 2017. pp. 1–9.
30. Fogel D.B. An introduction to simulated evolutionary optimization. *IEEE Transactions on Neural Networks*. 1994. vol. 5. pp. 3–14.
31. Roeva O. Genetic Algorithm and Firefly Algorithm Hybrid Schemes for Cultivation Processes Modelling. Transactions on Computational Collective Intelligence XVII. 2014. vol. 8790. pp. 196–211.
32. Neumüller C., Wagner S. Parameter Meta-optimization of Metaheuristic Optimization Algorithms. International Conference on Computer Aided Systems Theory. 2012. vol. 6927. pp. 367–374.
33. Chekanin V.A., Chekanin A.V. Design of Library of Metaheuristic Algorithms for Solving the Problems of Discrete Optimization. *Advances in Mechanical Engineering*. 2018. pp. 25–32.
34. Podorozhniak A., Lubchenko N., Balenko O., Zhuikov D. Neural network approach for multispectral image processing. Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET). Ukraine. Lviv-Slavyansk. 2018. pp. 978–981.
35. Grishkin V., Lakushkin O., Stepanenko N. Biofouling detection based on image processing technique. 2017 Computer Science and Information Technologies (CSIT). 2017. pp. 158–161.
36. Gillette A., Wilson C., George A. Efficient and autonomous processing and classification of images on small spacecraft. 2017 IEEE National Aerospace and Electronics Conference (NAECON). 2017. pp. 135–141.
37. Nikam P.A. Sawant S.D. Circuit board defect detection using image processing and microcontroller. 2017 International Conference on Intelligent Computing and Control Systems (ICICCS). 2017. pp. 1096–1098.
38. Schütze O. Hernandez V. The hypervolume based directed search method for multi-objective optimization problems. *Journal of Heuristics*. 2016. vol. 22. pp. 273–300.
39. Eshelman L.J. The CHC adaptive search algorithm: how to safe search when engaging in non-traditional genetic recombination. *Foundations of Genetic Algorithms*. 1991. pp. 265–283.
40. Aghajanyan A.G. [Analysis the influence of probabilistic parameters on results of evolutionary-genetic algorithm optimization of images approximation] *Техника и технологии: курс на инноватии: сборник материалов международной научно-практической конференции* [Technics and technologies; the road to innovations: collected papers of international science conference]. Irkutsk: «Арех». 2017. pp. 49–57. (In. Russ.).

R. YOSHINOV, O. ILIEV
**THE STRUCTURAL WAY FOR BINDING A LEARNING
MATERIAL WITH PERSONAL PREFERENCES OF LEARNERS**

Yoshinov R., Iliev O. The Structural Way for Binding a Learning Material with Personal Preferences of Learners.

Abstract. Learning content creation process requires more than just collection and presentation of set of information. In order to gain knowledge, the learning content should be designed in such a way to meet predefined learning goals. Learning goals determine the entire process of learning. Bloom's Taxonomy provides a description of a cognitive process with six hierarchical levels, each containing specific learning goal to achieve. It could be adapted into a model by which tutors create learning materials. However, when it comes to productivity of learning, it is important to consider the personalization of the presented content according to the learning style of the individual. This article analyzes the correlation between Bloom's Taxonomy and Honey & Mumford's learning cycle, providing a way to bind the structure of learning material to the personal preferences of learners. This novel way of creating learning materials is integrated into a model that is used for automatic generation of personalized learning materials. The effectiveness of the model is further verified through an experiment with real participants. The results of the experiment show promising potential in the way of how a learner's capabilities may be enriched. However, while experimenting and rest of the work on the model outline some challenges before the model's application and future work.

Keywords: learning goals, bloom's taxonomy, Honey & Mumford learning cycle, learning materials, gamification, personalized learning process, A/B testing.

1. Introduction. Learning goals determine the entire process of learning. Goal priorities and goal dependencies when deciding what to learn, and how to coordinate multiple learning strategies improve the effectiveness of learning often changing the context in which the process of learning is being performed, as described in Section 2 "Importance of learning goals in the process of learning".

The process of learning content creation requires more than just information grouping. The learning content should be designed in such a way to meet predefined learning goals. Section 3 "Segregation of the learning content according to Bloom's Taxonomy model" provides a novel way to adapt the Bloom's Taxonomy to segregate a learning material to six parts, each of them setting a specific learning goal to achieve. The granulation of the learning content is made following the Wagner's [26] model and provides a way to reuse the created content.

In order to create a learning material, it is necessary to have some kind of source of learning content that is properly structured and described in advance so that the information there to be reusable. However, the existing repositories are not following a common standard and this

interference their reusability. The advantages and disadvantages of the modern learning content management systems are overviewed into Section 4 “Source of learning content” as well as a novel data structure that can solve their limitation.

The process of learning cannot be universally applicable for all learners. Different learners perceive the information transmitted in various ways, and their performance is influenced by how the learning content is being served. Learning styles, presented into Section 4 “Personalizing the learning materials according to the cognitive abilities and the preferences of the learners”, aim to make complex tasks seem easy-looking, simply adapting the method of presentation of the information.

The segregation of the learning material following the adapted Bloom’s model would increase the productivity of the learners in the process of acquiring new knowledge. This productivity could be further enhanced when the segregated learning material is re-arranged in such a way as to follow a specific learning style appropriate for each of the different types of learners. A novel way to bind the structure of learning material to the personal preferences of learners is provided into Section 6 “Binding the structure of learning material to the personal preferences of learners”.

The model of segregation the learning content and re-arrange it to satisfy the preferences of the individuals having different learning styles, can be automated as described into Section 7 “Automatically generated learning materials”. When it comes to automation of a process, however, it is important to get feedback from the people who use the results of this process. In this way, the model that describes the process can be self-correcting in order to provide more accurate results. Section 8 “Ways to collect the learner’s feedback and increasing their motivation” provides a way to both motivate all the participants in the learning process and collect their feedback using so called “Gamification” strategy.

Section 9 “Example of the application of the model for automatically generated personalized learning materials” presents the model’s application into a real scenario — lesson in Informatics, in order to show its benefits and the way it is working.

The effectiveness of the presented model was verified through an experiment with real people described into Section 10 “Active experiments — A/B Testing approach”. These experiments show the practical potential of the presented model, but also together with the rest of the work on the model outline some “challenges before the model’s application and future work” as it can be seen in Section 11.

2. Importance of learning goals in the process of learning.

Learning goals determine the entire process of learning. Goal priorities and goal dependencies when deciding what to learn, and how to coordinate multiple learning strategies improve the effectiveness of learning often changing the context in which the process of learning is being performed. It is also explicit in the formulation of the learning process, the search for information, hypothesis evaluation, and other aspects of learning. Learning strategies, represented as methods for achieving learning goals, can be chained, composed, and optimized, resulting in learning plans that are created dynamically, and pursued in a flexible manner [21]. Identifying or determining the learner's goals and analyzing them into lower level learning goals is a very challenging task that is very difficult to be performed by the learner, and it is usually performed through the instructor's intervention, based on the appropriate methods, and decomposing learner's goals into lower level learning goals in order to facilitate the learning process. Previous knowledge, as well as previously acquired skills, are closely related to the learning objectives, measured directly by the use of tests, concept maps, portfolios, auditions, etc. — or indirectly, by means of self-reports, inventory of prior courses, experiences, and so on.

The learner's goals should be taken into account both in the organization of a learning experience, and the selection of its underlying content — the learning objects. In order to be able to be selected during the personalization procedure, the learning objects should also be dependent on difficulty, domain, and the learning experience and educational level of the learner. Participation in learning activities which correspond with the proper use of associated learning resources and materials, explore the ability for their transformation into knowledge for the target learner, with regards to a domain, and preferred difficulty is highly dependent on his learning experience aligned with the corresponding educational level.

3. Segregation of the learning content according to Bloom's Taxonomy model. Bloom's Taxonomy [4], and its revised version developed by Lorin Anderson [1] present a 6-level hierarchical classification of cognitive processes, taking place during the acquirement of a new piece of knowledge or skill, to gain expertise in a topic. Each level of Taxonomy presents a specific learning goal, described with key verbs (Table 1), helping the tutors to formulate questions, tasks, examples, definitions, etc. [1]. This model could be transferred into model by which tutors create learning materials — each such material should be composed of six parts needed to complete the process of

learning some newly acquired knowledge, which should be relevant to the specific learning goal.

Table 1. Verbs describing the learning goals, typical for each cognitive domain, described in Bloom's Taxonomy model

Taxonomy Level	Verbs Describing the Learning Goal
Remembering	define, describe, identify, know, label, list, match, name, outline, recall, recognize, reproduce, state
Understanding	understand, comprehend, convert, defend, distinguish, estimate, explain, extend, generalize, give an example, paraphrase, summarize, translate
Applying	apply, change, compute, construct, demonstrate, discover, manipulate, operate, predict, prepare, produce, relate, show, solve, use
Analyzing	analyze, break down, diagram, deconstruct, differentiate, discriminate, illustrate, infer, select, separate
Evaluating	evaluate, appraise, conclude, compare, contrast, criticize, critique, interpret, justify, support
Creating	create, categorize, combine, compile, compose, devise, design, generate, modify, organize, plan, rearrange, reconstruct, reorganize, revise, rewrite, tell, write

The verbs that characterize the learning objectives defined in Bloom's Taxonomy levels are not universally applicable when it comes to drawing up the content of a certain learning material. The idea of hierarchical classification of learning as a cognitive process is to describe the way of acquiring new knowledge. When it comes to building up learning material, however, these verbs do not always work. For example, if a tutor prepares a job-specific assignment, he could use the "define" verb characteristic feature of the Taxonomy "remembering" process, but when the lesson is to be composed, it would be more appropriate to use the word "definition" that will clearly describe this part of the lesson associated with the "remembering" process. Moreover, the use of the set of verbs related to specific levels of Taxonomy, supplemented by nouns

and overriding syllables, would increase the number of search terms that an automated content collection system could use. Such a system could search ready-made repositories with learning content for “definition” or the task of “defining” a specific problem.

The main components of the learning content model are as follows [26]:

- Content Asset: Content assets include “raw” media such as images, clippings, audio and video clips, and more.
- Information Object: Text passage, web page, etc. that focus on a single piece of information. Such a piece can explain a concept, illustrate a principle or describe a process.
- Learning Object: In the learning content model, the learning object is a collection of information objects that are assembled together to meet a learning goal.
- Learning Component: The learning component is a basic concept of things like lessons or courses that are related to meeting multiple learning goals at a higher level. They are a combination of several training sites.
- Learning Environment: The learning environment is a combination of learning content and technology that the learner interacts with. The combination of training components with communication tools and/or other functionalities that aim to provide online learning experience can be aggregated in a learning environment (like LCMS).

The individual parts that could be divided into a learning material following Bloom’s Taxonomy model can be considered as information objects (IOs) in the sense of Wagner’s model [26] (Figure 1), and the whole group forms a learning object (LO). This granulation of information objects will allow easy aggregation, and re-use of individual “pieces” of information (IOs) into complete lessons (LOs).

It is generally accepted that there is a relation between the size of the learning object and the possibility of its reuse. Well-granulated learning objects and components have the potential to be flexibly assembled into new learning objects, while whole courses are not suitable for use in different contexts [2]. This fact is also illustrated by the Figure 1. This article is based on this fact. One of the main shortcomings of modern learning systems is the use of ready-made training materials by teachers who upload entire lessons and/or exams in different finished files. This leads to limitations, both on the technological level – the content of the files cannot be easily indexed and searched by users of the training systems, which limits and even makes its reuse impossible, as well as in terms of conceptually granular level, as reuse a whole lesson is very difficult.

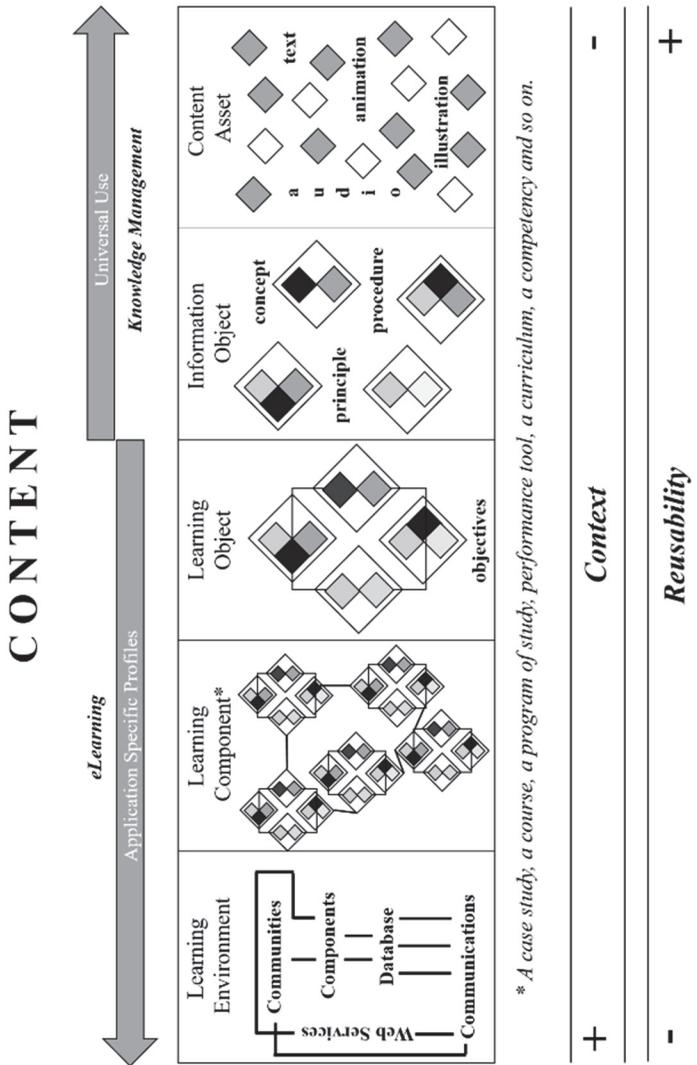


Fig. 1. Component based content model [26]

4. Source of learning content. In order to automate the creation of learning material, it is necessary to have some kind of source (repository) of information that is properly structured and described in advance so that the information there to be reusable. There are different kinds of sources of learning content determined by the type of information in them, how it is granulated and described. Unfortunately, most of these repositories do not follow a common standard for describing available content and their use is very difficult without a thorough knowledge of the repository's architecture. For the purposes of this article, the so-called Learning Content Management System (LCMS) will be addressed. In the context of the article, they will be referred to as repositories of information that could be reused to create new learning materials [28].

The advantage of LCMS kind of systems is that they already have content stored in them. Groups of lecturers, experts and trainees participate in their development — generating learning materials and assessing on the basis of feedback the credibility and relevance of these learning materials. In other words, the use of a ready set of information could be the fastest and most convenient way to create an automatic content generation system, as content will already be collected. However, their use has its limitations and disadvantages. Otherwise, teachers and experts have yet to fill in a content or create a “smart” automated system to crawl up a similar type of repository and “dig” information from there, structuring it and describing it more convenient for the automated learning system standard [28].

Before examining the advantages and disadvantages of LCMS, a review of the concept of the eLearning system and the components from which its architecture is designed will be reviewed [28].

4.1. eLearning System. Information and communication technologies have opened new horizons and opportunities for training and teaching, they overcome the problems and limitations of traditional approaches. In e-learning systems traditional forms of learning are enriched with new opportunities that have a strong technological foundation. They are complete eLearning infrastructures that allow the development, management and provision of advanced learning services at any time and everywhere [2].

The eLearning Systems can be divided primarily into two types, defined by their infrastructure, Learning Content Management Systems (LCMS) and Learning Management Systems (LMS). In modern electronic learning systems, however, these two functions are often mixed into a common “super” system.

LCMS focus on the creation, overfilling and management of learning content. They cover the complete cycle of collecting, delivering, managing and reusing training content in many different ways [17]. All these training systems use their own repository of learning objects/materials. They allow

users to create, manage, search and reuse ready-made training materials [2]. These training materials can be small pieces of information — created within the user's system, following its standards of structuring and describing information. Such “pieces of information” are easy to index, search and reuse in different contexts from the same and/or other users. They can contain media files, tests, simulations, plain text, graphics, references to external sources [2]. Describing this kind of information related to learning material is achieved through standardized metadata structures that allow not only the re-use of the pieces” of information by the same or other users but also the sharing of information between different repositories of learning content. The truth, however, is that the use of this functionality of the LCMS is not a frequent practice by educators and experts. They typically do not create their training materials through the system where they are well described and structured, and more often use ready-made files created by them and just upload them to the system and complete the course content. Such file types are most often PowerPoint presentations, PDF/Word/Excel files, or archives from other files. Unfortunately, indexing, searching and reusing available information in these files is almost impossible. A number of obstacles face such a challenge, including the coding of the different formats, the lack of a metadata description. As a result, any kind of searching or indexing software cannot use the repository as a base.

In the training systems, the materials are granulated into small independent pieces that can be used alone or in combination with other materials to form higher level objects and meet the needs of the user [2]. Fundamental idea of learning objects is the lesson designer to create small components that can be reused many times in a different learning context [11]. Many publications claim that re-use not only saves money and time for trainers, but also enhances the quality of training materials. Just like the LEGO blocks, the idea is to create something small that can only be complete for itself but also easily combined with other components [11]. Learning Object (LO) should follow the rule that each unit should do only one thing and minimize the link with other units [5]. There is a general consensus that the learning object must be *Reused* (can be modified and used in different courses), *Accessible* (indexable and accessible by descriptive metadata), *Compatible* (operates on different hardware/software), *Durable* (to maintain proper operation after software or hardware upgrades) [18].

An important feature of the reusability and personalizability of training objects is their granularity [2]. However, the structure and content of learning objects is still unclear and possible to interpret in different ways [6, 20]. There is still an incomplete understanding of what a learning object is and how it differs from simple objects such as files, photos, videos,

or a whole scientific report. When and at what level does an object become a learning object? What distinguishes learning objects from other learning materials? Because of the openness of these issues, there are different implementations of LOs [2]. An overview of existing models for identifying learning objects can be found in [3, 24, 25, 34].

4.2. A novel structure of a leaning repository that provides an appropriate level of granularity to allow easy re-use of the learning content. In order to meet the model requirements, set out in this article, a specific structure is proposed for describing learning materials with a set of metadata, described into Table 2 and Table 3. Thanks to such descriptions of information with descriptors it is possible to achieve full reusability of the teaching materials, their indexing and searching, as well as the automated generation of teaching materials, in relation to a predetermined topic [28].

Table 2 contains the descriptors that are describing learning material at the highest possible level. Every learning material (group of small pieces of learning content to form a IO) should be described with such kind of descriptors by its creator. On the other hand, Table 3 describes with metadata all the singles pieces (LO) of the IO on a lower level.

Table 2. Descriptors describing learning material at the highest possible level

Descriptor	Field Type	Example
Title	Text field	<i>Motion and rest of objects</i>
Key words	Text field	<i>motion of objects, motion and rest</i>

Table 3. Descriptors describing the individual learning objects associated with some learning material

Descriptor	Field Type	Example
Language	Language code selected from ISO 639-1	<i>en-US</i>
Learning goal	Value selected from a predefined list	<i>Definition</i>
Content	Text area field with the option of adding an image/video/audio or an external reference	<i>An object moves if it changes its position in time compared with another object. The object is at rest (it is still) if it does not change its position relative to the orientation.</i>
Complexity level	Low/Normal/High	<i>Normal</i>
Level of education	Value selected from a predefined list	<i>VI-th grade</i>
Learning context	Value selected from a predefined list	<i>Human and Nature</i>

The purpose of this metadata structure is to enable the teacher to create learning content on a particular topic in which to add different types of learning objectives. For example, if the teacher creates a learning content and gives him the title “Motion and rest of objects”, he/she can add a definition to it, for example — descriptive information related to the learning content. Consequently, another tutor or the same one could search for previously created content and use it in another or the same context, but on a topic he or she defined. For example, the teacher could look for a “definition” and submit some keywords — search criteria. This will find all the existing content if there is one [28].

Filling in the content according to this structure of meta descriptors could also be done automatically by searching for existing repositories of learning materials from a “smart system” that recognizes the content of the information and classifies it according to the proposed architecture [28].

5. Personalizing the learning materials according to the cognitive abilities and the preferences of the learners. The process of learning cannot be universally applicable for all learners. Different learners perceive the information transmitted in various ways, and their performance is influenced by how the learning content is being presented. Some of them prefer to be engulfed in theory completely before continuing with the application of knowledge, while others prefer to gain new knowledge by solving specific problems. The different types of learners, and the type of cognitive processes that take place within their minds when acquiring new knowledge can be described by the so-called “learning styles”. Learning styles are characterized by different methods of learning, organizing and understanding the information received from other people [7]. They do not deal with everyone’s ability or the level of learners’ intelligence, but they aim to make complex tasks seem easy-looking, simply adapting the method of presentation of the information [29].

In literature, there can be found several definitions of the concept learning style [15]. Learning styles might be generally defined as: the preferred attitude of the individual for organizing and presenting the information [22]; the various ways in which learners acquire, process, store, and recall knowledge. [14]; the distinctive types of behavior that serve as indicators of the way a person learns from and adapts to the environment, and also provides signs of how the human operates [9]; the attitudes and behavior that determine the preferable way of learning for the individual concerned [12]. As an example, a student learning how to program would rather start writing a code straight away in order to learn a new programming language, while another would prefer reading and learning the

new language before he actually approaches real programming. Choosing a learner-friendly learning style provides an opportunity to customize the learning material in order to enhance productivity.

One way of classifying learners into their appropriate learning styles is to use the Kolb's learning cycle [16]. In his theory, Kolb considers the so-called experiential learning process, which is presented as a cycle of stages through which the learner passes in order to reach complete knowledge. In this way, Kolb divides the learners into four distinctive types – accommodators, divergers, convergers, and assimilators. According to him, effective learning can only be accomplished once the learner had passed through the entire four-stage cycle, but each learner can start his transition from each of the four stages, and follow the logical sequence of events to complete the circle [19].

Kolb's cycle was additionally developed by Honey & Mumford (1992), where apart from accepting changes to the correct model of introducing the correct type of learning content for the specific type of learner, a Learning Styles Questionnaire (LSQ) has been presented — serving as a blueprint for classifying the learners according to their most appropriate type of learning [30, 31, 33]. According to Honey & Mumford's learning cycle (Figure 2), the learners might be divided into four types — activists, reflectors, theorists, and pragmatists.

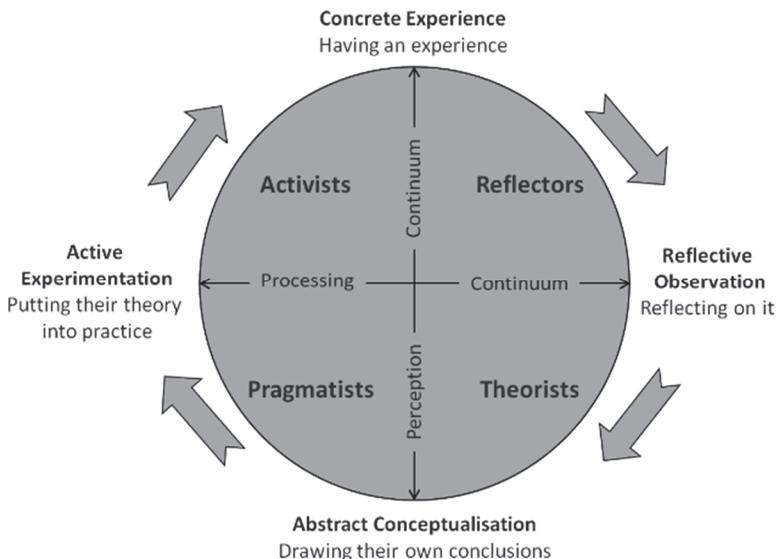


Fig. 2. Honey & Mumford's Learning Cycle

6. Binding the structure of learning material to the personal preferences of learners. The segregation of the learning material following the adapted Bloom's model would increase the productivity of the learners in the process of acquiring new knowledge. However, the productivity could be further enhanced when the segregated learning material is re-arranged in such a way as to follow a specific learning style appropriate for each of the different types of learners. In other words, the consistency of goals that should be placed on learners must be tied to the most appropriate for them learning style [32].

In one of his articles, James Gallagher presents the relationship between Kolb's learning cycle, and Bloom's Taxonomy [8]. He presents the idea that the different learning styles should go through all levels of the learning process — as defined by Bloom — even though in somewhat different sequence. The principle of crawling through the levels in a clockwise manner has been used, and a specific level appropriate to the various learning styles has also been used in order to start the cycle. This article uses Gallagher's work by updating the inherent model. Instead of using Bloom's original Taxonomy, its updated version was being implemented, and instead of Kolb's learning cycle, it integrates the one developed by Honey & Mumford (Figure 3).

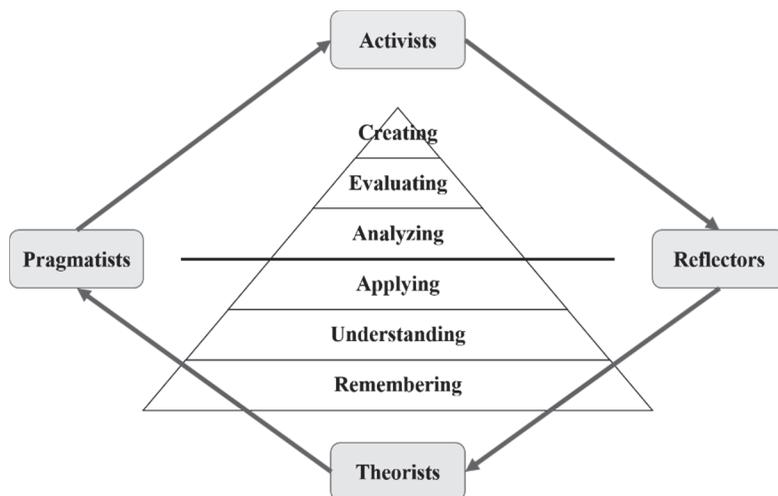


Fig. 3. An adapted Model of Gallagher [8] illustrating the relationship between the learning style, according to the Honey & Mumford's learning cycle, and the learning goals, according to the revised version of Bloom's Taxonomy

The sequence of processes defined in Bloom's Taxonomy through which the learners with different learning styles must pass, according to Honey & Mumford, is graphically presented in Figure 4.

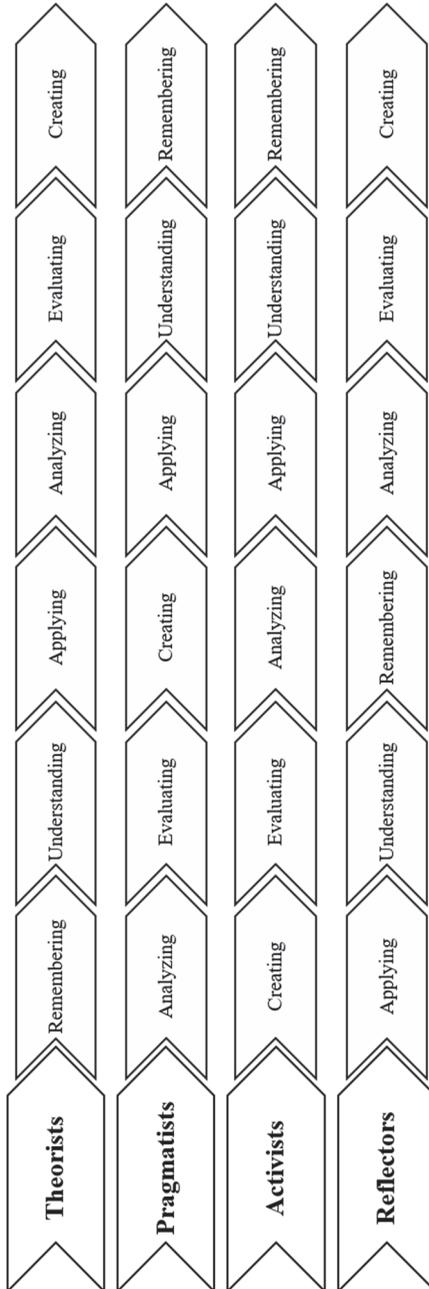


Fig. 4. The sequence of processes as defined in Bloom's taxonomy, through which people must pass for each learning style defined by Honey & Mumford's learning cycles

7. Automatically generated learning materials. The segregation of learning content, represented as a collection of different IOs, according to Wagner's classification [26], which should meet each learner's specific learning goal in the process of acquiring new knowledge, allows the easy generation of learning content in the form of LOs. The model presented in this article aims to aggregate all necessary "pieces" of information to compile a new learning material following the model of cognitive processes as distinguished in Bloom's Taxonomy — a lesson should be composed of separate objects provoking the invocation of processes into the learner's mind as: remembering, understanding, applying, analyzing, evaluating, and creating.

When it comes to aggregating individual IOs, there must be some source of learning content. The easiest option would be to have a repository of learning materials, created by the tutors, properly separated into individual components, allowing the reuse of each part, independently of other parts, in different lessons within different contexts. The tutors have to create the learning content in such a way that each individual part of it only refers to one type of cognitive process. For example, a learner could define a specific problem; to illustrate a definition by way of an example; to provoke learner to search for different options for its application in different situations; to set a task of analyzing the situation related to the particular problem; to stimulate learners to evaluate a specific solution to the problem based on their own knowledge, and to offer their own solution to a certain study case.

Another option is to use an automated search engine to look through the already-created arrays of information. The search engine should use a set of keywords describing the learning objectives per each level of Bloom's Taxonomy in a combination with the relevant context, and thus automatically generate the required small "pieces" of learning content. For example, the search engine could search for text containing keywords such as: definition, example, apply, analyze, rate, create, etc., in predefined context. Once the learning material is granulated according to the principles set out in the model presented in the article, it can be reused multiple times. Moreover, the sequence of presenting the individual IOs composing an LO can easily be altered. This allows the link between consistency in achieving the learning objectives, and the learning style discussed in the Section 6 "Binding the structure of learning material to the personal preferences of learners".

The model represented in the article provides an opportunity for its thorough automation – from collecting the items of segregated learning content to meet a specific learning goal, to the principle of rearranging the individual parts of learning content in such a way as the final learning material meets the needs of each style of learning.

8. Ways to collect the learner's feedback and increasing their motivation. When it comes to automation of a particular process, it is

important to get feedback from the people who use the results of this process. In this way, the model that describes the process can be corrected in order to provide more accurate results. For this reason, the model presented in the article should provide a process of collecting feedback from the learners in order to evaluate each part of the lesson presented, as well as its complete set-up (LO). If a certain section of the lesson receives poor feedback, then it should not be reused again at the expense of other IOs that meet the same learning objective. What's more, this part of the lesson should be replaced ASAP in the overall architecture of the lesson, in order to achieve better productivity in the process of acquiring new knowledge.

One particular challenge encountered by each model for presenting new knowledge to the individual learners is their motivation. Of utmost importance for the learners is their productivity in acquiring new knowledge. Moreover, in a model that could require the manual generation of learning content by tutors, it is also important to pay attention to the level of tutors' motivation to generate IOs. One possible solution to these two problems is the so-called process of Gamification. Gamification can be defined as the application of game elements and principles in a non-game context [13]. These elements and principles have been typically used to increase consumer engagement with a company, website or idea, increase the productivity of the learning process, and much more. Many studies have shown that the implementation of gaming elements in a ready-made concept could have a very positive effect on its performance [10].

We can divide gamification into two types — structured and meaningful. Structural gamification is more widespread and easy to apply, so it's included in the model presented in this article. With it, different gaming elements are directly applied in a specific context. This is done in order to pass the user through some process or content, providing various stimulating elements in the passage steps. In education, such structural gamification is used to translate learners through learning content and learning process management. The gaming elements that are commonly used are points, levels, badges, leadership lists, and achievements. These elements can be directly applied to the learning context by providing learner points for every correct test response, for example. Different badges for the best essay in the whole class or even the whole school can be provided. Through these points and badges, learners climb to levels that are presented in a leader list — public for the whole school. The other type of gamification — meaningful, is more difficult to apply. It aims at adapting a process or content to gaming elements, gaming mechanics, and game way of thinking. Using it, the content or process is changed to be more gaming like. The most commonly used elements are adding history, challenges, mystery, and characters to a context. For example, in the educational context, the content could change by introducing a learner's

goal, the challenges that he has to go through in order to reach it, a story that has mysteries and characters in it [23].

When gamification elements are introduced into the model, the motivation and knowledge of both learners and tutors could be enhanced. The first ones will obtain certain features symbolizing their success in going through different stages of learning, while the latter will be encouraged to improve their learning content. The public presentation of the results, which can be made visible to all participants in the process, is meant to provoke them to express naturally their innate desires for achieving success, competition, and/or cooperation with others [23].

9. Example of the application of the model for automatically generated personalized learning materials. Table 4 presents a lesson in Informatics — Flowcharts. The lesson considered as a LO has been decomposed into separate IOs responsible for each specific learning objective facing the learner.

Table 4. Separation of the LOs, partly needed to complete the process of getting knowledge, according to Bloom’s Taxonomy, for a lesson presented in "Flowcharts"

Learning Objective	Keyword	Learning Content
Remembering	Definition	<p>Flowcharts are diagrams that are used to make and record algorithms to be executed and achieve a corresponding result. They consist of geometric shapes, each of which has a definite meaning. In these figures, the data that the algorithm performer has to perform is entered.</p> <p><i>Main elements:</i></p> <ul style="list-style-type: none"> – <i>Terminal</i> — Indicates the beginning and ending of a program or sub-process. Represented as a stadium, oval or rounded (fillet) rectangle. They usually contain the word "Start" or "End", or another phrase signalling the start or end of a process, such as "submit inquiry" or "receive product". – <i>Process</i> — Represents a set of operations that changes value, form, or location of data. Represented as a rectangle. – <i>Condition</i> — Shows a conditional operation that determines which one of the two paths the program will take. The operation is commonly a yes/no question or true/false test. Represented as a diamond (rhombus). – <i>Input/Output</i> – Indicates the process of inputting and outputting data, as in entering data or displaying results. Represented as a parallelogram (extra images).

Continuation of Table 4.

Understanding	Example	<pre> graph TD A[Does the device work?] --> B{Does it has battery?} B -- No --> C[Put on a battery] B -- Yes --> D{Is the battery charged?} D -- No --> E[Charge the battery] D -- Yes --> F[The devise is broken] </pre>
Applying	Application	Find more information about flowcharts over the internet and write down three applications of flowcharts/algorithms in the real life?
Analyzing	Analysis	Analyze the results of the two algorithms displayed below. Why the are different? (two extra images)
Evaluating	Evaluation	Analyze the sorting algorithm displayed bellow using Big O Notation. (extra image)
Creating	Creation	Create an algorithm and design a flowchart, which finds a number in a set of numbers.

The individual IOs of the LO (the completed lesson) should be rearranged according to the sequence presented in Figure 4.

When presenting the learning material to the learner, continuous feedback about the information objects should be provided, as well as its replacement in the absence of accuracy. Moreover, the learner’s achievement is distinguished for each successfully completed lesson, which aims to increase his/her dedication and motivation.

Using the model provided in the article, not only LOs that aim to generate new knowledge in learners can be created (like the one in the example presented in this section), but also assignments in order for knowledge to be verified. In the second case, the model should work with IOs that are not so much of an explanatory nature, but formulating questions, and provoking independent action on behalf of learners. For example, for

“remembering” learning goal, the automated system should not search for keyword “application” in a specific context, instead, the system should search for “apply” keyword.

10. Active experiments — A/B Testing approach. The effectiveness of the presented model was verified through an experiment with real people. To setup the experiment, a group of 112 participants were involved. They were at age between 13 and 14 — all of them in 7-th grade. The children were asked to read a learning material in flow charts, to make some examples, to research more about the problem over the Internet and finally to take an exam. They were rated with points between 1 and 30. The following section describes the experiment and analyse the accumulated data.

In order to provide more objective data, the experiment was made into a controlled environment and mostly dependent variables were used. Initial setup of the experiment:

- Children were in the same age group and in the same grade. Potentially, their background knowledge and skills are close to each other.

- 112 participants were involved. This large number of participants in the experiment provides the opportunity to reveal more accurate results.

- All the participants were asked to complete a questioner that defines their learning style, according to Honey & Mumford Learning Cycle.

- The learning material that the participants had to read is the one from the example given in this article. It is well granulated, following the principles by the Blooms Taxonomy, when convert the cognitive learning process to creation of a learning material. This allows both its reusability as a full and the reusability of each individual piece. Moreover, as the model presented into the article offers the content can be reordered in such way to satisfy the needs of every learning style.

- The evaluation of the participants was done thought a test, containing 30 questions. Every right question was giving 1 point to the participants. The wrong answers are not taking points out from the participant score.

- All the participants had the same maximum amount of time to read the learning material — 20 minutes and the same maximum amount of time to complete the test after reading it — 20 minutes.

To conduct the experiment, we decided to adapt so called A/B testing approach, also known as bucket testing or split-run testing. The testing approach is initially developed as a tool in web analytics, however, its main advantage is that it is easily integrated into web-based software environment. Such software environment was developed to make the experiment presented into this article.

A/B testing (Figure 5) provides a simple approach to compare two variants of a content. The participants of the experiment are divided into two groups – group A and group B. The first, group is called a control group. This

group received the traditional (well known) variant of the content. The second, group is called treatment group. This group received the novel (opportunity) variant of the content. Since, this approach is technically adapted it allows the users to be separated into the groups randomly using their session with the server. It is also easy to manage what percentage of the users will be into the treatment group and what rests for the control group. Finally, the system can measure specific parameters that are later used to decide if the novel variant is better or worse than the traditional one.

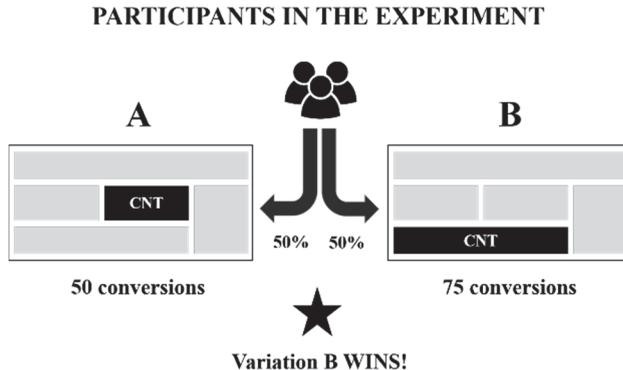


Fig. 5. Illustration of A/B Testing approach

Once the appropriate learning style of every participant were defined, further setup of the experiment was done:

- The controlled group size was set to 25% of all the participants – 28 participants. Pragmatically it was set this group to have equal number of representatives of each learning style — 7 theorists, 7 reflectors, 7 activists, and 7 pragmatists.

- The rest — 84, of the participants were into the treatment group. In this group, there were 16 theorists, 27 pragmatists, 24 activists, and 17 reflectors, according to their learning style.

Once the learning style of all the participants were defined and they were randomly separated into the groups, the actual experiment, were done:

- All the participants received the same learning material to read. However, the control group receive the learning material in its traditional order inducing the cognitive processes described by Bloom's Taxonomy into their regular order: *remember, understand, apply, analyse, evaluate, and create*. The participants into the treatment group received the material reordered, according to their learning style.

- All the participants were asked to take the same test after they read the learning material. The questions were into the same order for every participant.

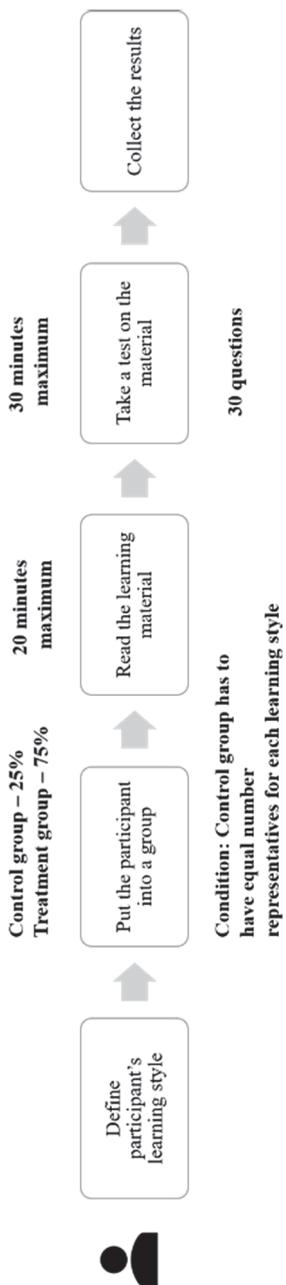


Fig. 6. Illustration of the experiment process

Result from the experiment:

After the experiment was made and the results (Figure 7) were analysed, it was noticed that the participants from the treatment group (the one using the model offered into the article) achieved about 12% better results than the participants from the control group (the one using the traditional way of structuring of learning material). This concludes that the model shows promising potential in the way of how a learner's capabilities may be enriched.

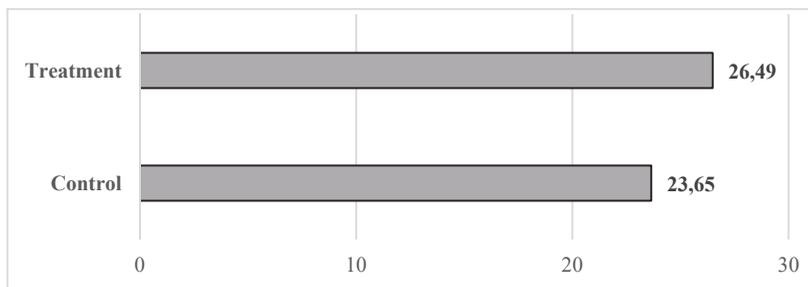


Fig. 7. Average results of the participants

11. Potential challenges before the model's application and future work. The main problem with the use of the model for automatically generating customized learning materials is the lack of pre-introduced content on the basis on which training materials to be created. Without the availability of data in the system's local repository, the preparation of teaching materials is much more difficult. This problem is caused by the lack of popularity of the model and its integration into existing learning systems. It could be overcome by creating algorithms for automated waging, sifting and classifying educational information in external repositories. Another important condition to overcome the above mentioned problem is finding appropriate repositories as well as configuring the system to make it possible to use the information in these repositories. The most relevant are the repositories of already-in-the-real-world content management systems LCMS that have a broad community of users who use and evaluate both content-based and content-generating content [27].

Another potential problem with the model is the possible generation of insufficiently accurate training material as a result of the search of external sources where the information is not classified in advance according to the context, the expected educational level and the level of complexity of the content contained. This problem is envisaged and the model is potentially integrated in the model – the user's ability to request the system to re-search and replace a specific information object (IO), part of the training presented. Moreover, the system has the obligation to reflect the

user's request to replace the information in order to “teach” the system and store relevant content in the local repository [27].

A problem faced by the model is shared by all existing LMS and LCMS – we cannot be sure of the relevance of the curriculum. Because users who can create learning content themselves determine whether their input is correct, users who use this training information can also not be sure of the content they provide. A potential solution to this problem is embedded in the model with the introduction of an opportunity to provide feedback from trainees on the training they are offered. This feedback is further tied to gamification affiliates. The model adapts it to the results of the gamification analysis and provides these information objects, which are part of the most positive feedback trainings. In this way, the “maturity” of the introduced content is evaluated [27].

12. Conclusion. Identifying or determining the learner’s goals and analyzing them into lower level learning goals is a very challenging task that is very difficult to be performed by the learner. It is usually performed through the instructor’s intervention, based on the appropriate methods.

The adaptation of the Bloom’s Taxonomy in a type of model to create learning materials, composed by six parts needed to complete the process of learning newly acquired knowledge, which should be relevant to the specific learning goal, seems to be potential way to granulate the content into small individual, independent and reusable pieces. However, the right granulation and creating learning materials that achieve the right learning goals is not enough when it comes to individuals with different learning styles. This article analyzes the correlation between Bloom’s Taxonomy and Honey & Mumford’s learning cycle, providing a way to bind the structure of learning material to the personal preferences of learners.

The model described into the article may be fully automated. It is able to generate learning material from scratch using a proper search engine and filling in a repository of learning content. The collected information could be classified by its type in sense of a learning goal and then arranged, following the right order, according a specific learning style, into ready learning material. When it comes to automation of a particular process, however, it is important to get some feedback from the individuals who use the results of this process. In this way, the model that describes the process can be self-correcting in order to provide more accurate results. This article provides a way to both motivate all the participants in the learning process and collect their feedback using so called “Gamification”.

The example of model’s application with a lesson in Informatics – Flowcharts, presents all the processes which takes part of the model and shows its potential. The same example is used as a learning content base of

a real experiment with more than 100 participants. This experiment verifies the effectiveness of the model, showing promising potential in the way of how a learner's capabilities may be enriched. However, while experimenting and rest of the work on the model outline some challenges before the model's application and future work.

References

1. Lee Y.J. et al. East-Asian Primary Science Curricula: An Overview Using Revised Bloom's Taxonomy. Springer. 2016. 81 p.
2. Arapi P. Supporting Personalized Learning Experiences on top of Multimedia Digital Libraries. Ph.D. Thesis. 2017. 277 p.
3. Balatsoukas P., Moris A., O'Brien A. Learning objects update: Review and critical approach to content aggregation. *Journal of Educational Technology & Society*. 2008. vol. 11. no. 2. pp. 119–130.
4. Morton D., Colbert-Getz J. Measuring the impact of the flipped anatomy classroom: The importance of categorizing an assessment by Bloom's taxonomy. *Anatomical sciences education*. 2017. vol. 10. no. 2. pp. 170–175.
5. Boyle T., Cook J. Learning objects, pedagogy and reuse. Learning technology in transition. From individual enthusiasm to institutional implementation. 2003. pp. 31–44.
6. Knight C., Gašević D., Richards G. Ontologies to integrate learning design and learning content. *Journal on Interactive Media in Education*. 2005. vol. 2005. pp. 1–24.
7. Coffield F., Moseley D., Hall E., Ecclestone K. Learning styles and pedagogy in post-16 learning: a systematic and critical review. Learning and Skills Research Centre 2004. 173 p.
8. Gallagher J. The Business Case Study: A Suitable Candidate For Blended Learning? *Journal of Business Case Studies*. 2006. vol. 2. no. 4. 14 p.
9. Gregoric A. Learning/teaching styles: Potent forces behind them. *Educational Leadership*. 1979. pp. 36–40.
10. Hamari J., Koivisto L., Sarsa H. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS). 2014. pp. 3025–3034.
11. Hodgins W. The future of learning objects. The Instructional Use of Learning Objects Bloomington: IN: AECT. 2002. pp. 281–298.
12. An D., Carr M. Learning styles theory fails to explain learning and achievement: Recommendations for alternative approaches. *Personality and Individual Differences*. 2017. vol. 116. pp. 410–416.
13. Huotari K., Hamari J. Defining gamification: a service marketing perspective. Proceedings of the 16th International Academic MindTrek Conference. 2012. pp. 17–22.
14. James W.B., Gardner D.L. Learning styles: Implications for distance learning. *New directions for adult and continuing education*. 1995. vol. 1995. no. 67. pp. 19–31.
15. Karagiannidis C., Sampson D. Adaptation rules relating learning styles research and learning objects meta-data. Workshop on Individual Differences in Adaptive Hypermedia. 3rd International Conference on Adaptive Hypermedia and Adaptive Web-based Systems (AH2004). 2004. pp. 66–73.
16. Kolb D.A. *Experiential Learning: Experience as the Source of Learning and Development*: 2nd ed. Pearson FT Press. 2004. 416 p.
17. Lennox D. Managing Knowledge with Learning Objects. WBT Systems White Paper. 2001. 12 p.
18. Mason R., Rehak D. Keeping the learning in learning objects. Reusing online resources: a sustainable approach to e-learning. 2003. pp. 20–34.

19. McLeod S.A. Kolb's Learning Styles and Experiential Learning Cycle. 2017. 5 p. Available at: <https://www.simplypsychology.org/simplypsychology.org-Kolb-Learning-Styles.pdf> (accessed: 06.07.2018.).
20. Metros S.E. Learning objects: A rose by any other name. *Educause Review*. 2005. vol. 40. no. 4. pp. 12–13.
21. Ram A., Leake D. Goal-Driven Learning. The MIT Press. 1995. 61 p.
22. Riding R., Rayner S. Cognitive styles and learning strategies: Understanding style differences in learning and behaviour. David Fulton Publishers. 1998. 217 p.
23. Robson K. et al. Is it all a game? Understanding the principles of gamification. *Business Horizons*. 2015. vol. 58. no. 4. pp. 411–420.
24. Verbert K., Duval E. Towards a Global Component Architecture for Learning Objects: A Comparative Analysis of Learning Object Content Models. EdMedia: World Conference on Educational Media and Technology. Association for the Advancement of Computing in Education (AACE). 2004. pp. 202–208.
25. Verbert K., Duval E. ALOCOM: a generic content model for learning objects. *International Journal on Digital Libraries*. 2008. vol. 9. no. 1. pp. 41–63.
26. Wagner E.D. Steps to creating a content strategy for your organization”. *The e-Learning developers' journal*. 2002. 9 p.
27. Yoshinov R., Iliev O. “Controlled self-study” in thematic educational community environment. Proceedings of the Forty-seventh Spring Conference of the Union of Bulgarian Mathematicians. 2018. pp. 200–213.
28. Yoshinov R., Iliev O. Content reuse - a major problem with modern content storage systems. Eleventh National Conference with International Participation “Education and Research in the Information Society”. 2018.
29. Yoshinov R., Kotseva M., Pavlova D. Specifications for Centralized DataCenter serving the educational cloud for Bulgaria. International conference ETAI. 2015. pp. 1–6.
30. Yoshinov R., Kotseva M. The steps for elaboration of the “Rosetta stone” demonstrator. Proceedings of International Conference Inspiring Science Education. 2016. pp. 91–96.
31. Yoshinov R., Arapi P., Kotseva M., Christodoulakis S. Supporting Personalized Learning Experiences on top of Multimedia Digital Libraries. *International journal of education and information technologies*. 2016. vol. 10. pp. 152–158.
32. Yoshinov R., Pavlova D., Kouzov O. Reflection of ISE idea for linking school education and scientific research in the National Strategy for effective implementation of ICT in education and science in the Republic of Bulgaria. Proceedings of International Conference Inspiring Science Education. 2016. pp. 129–134.
33. Yoshinov R., Kotseva M. Vision for the Engagement of the e-Facilitator in School in the Inspiring Science Education Environment. *Serdica Journal of Computing*. 2015. vol. 9. no. 3-4. pp. 241–256.
34. Trifonov R., Yoshinov R., Jekov B., Pavlova G. Methodology for Assessment of Open Data. *International Journal of Computers*. 2017. vol. 2. pp. 28–37.

Yoshinov Radoslav Dakov — Ph.D., professor, head of laboratory of telematics, Bulgarian Academy of Sciences (BAS). Research interests: computer science, medical systems, computer networks and communication, E-Government cybersecurity of computer networks. The number of publications — 191. yoshinov@cc.bas.bg; 8 bl., Akad. G. Bonchev Str., 1113, Sofia, Republic of Bulgaria; office phone: +359888627190.

Iliev Oleg Petrov — junior researcher of laboratory of telematics, Bulgarian Academy of Sciences (BAS). Research interests: information technologies, computer science, IT components to support education process. The number of publications — 4. iliev.oleg@gmail.com; 8 bl., Akad. G. Bonchev Str., 1113, Sofia, Republic of Bulgaria; office phone: +359884381052.

Р.Д. Йошинов, О.П. Илиев
**СТРУКТУРНЫЙ СПОСОБ И МОДЕЛЬ АВТОМАТИЧЕСКОЙ
ГЕНЕРАЦИИ ПЕРСОНАЛИЗИРОВАННЫХ УЧЕБНЫХ
МАТЕРИАЛОВ**

Йошинов Р.Д., Илиев О.П. Структурный способ и модель автоматической генерации персонализированных учебных материалов.

Аннотация. Определение целей обучения и их анализ — довольно сложная задача, которую учащемуся трудно решить самостоятельно. Расстановка целевых приоритетов того, что нужно изучать и какие стратегии обучения сочетать между собой, повышают эффективность получения новых знаний, часто изменяя контекст учебного процесса. Создание учебного контента требует не только сбора и представления информации — для приобретения знаний контент должен быть разработан таким образом, чтобы соответствовать заранее определенным целям обучения. Чтобы подготовить учебный материал, необходимо иметь источник учебного контента, который правильно структурирован и описан заранее, однако существующие хранилища не соответствуют общему стандарту, что мешает их повторному использованию. Таксономия Блума описывает когнитивный процесс с шестью иерархическими уровнями, каждый из которых содержит определенную цель обучения. Он может быть адаптирован к модели, посредством которой преподаватели подготавливают учебные материалы, однако когда дело доходит до продуктивности обучения, важно учитывать персонализацию представленного контента в соответствии со стилем обучения человека. В статье анализируется модель создания учебных материалов, основанная на таксономии Блума и цикле обучения Хони и Мамфорда. Описанную модель можно полностью автоматизировать и приспособить к самостоятельной генерации учебных материалов, используя подходящую для этого поисковую систему и закрытые репозитории учебного контента. Собранная информация может быть классифицирована по цели обучения, а затем упорядочена в соответствии со стилем обучения в готовый учебный материал. Эффективность модели дополнительно подтверждается экспериментом с реальными участниками, и результаты эксперимента показывают многообещающий потенциал расширения возможностей учащегося.

Ключевые слова: цели обучения, таксономия Блума, учебный цикл Хони и Мамфорда, учебные материалы, игрофикация, персонализированный процесс обучения, тестирование А/В.

Йошинов Радослав Даков — к-т техн. наук, профессор, заведующий лабораторией телематики, Болгарская академия наук (БАН). Область научных интересов: информатика, информационные технологии, модели, связанные с обучением, решения для поддержки электронного управления. Число научных публикаций — 191. yoshinov@cc.bas.bg; ул. Академика Георги Бончев, бл. 8, 1113, София, Республика Болгария; р.т.: +359888627190.

Илиев Олег Петров — младший научный сотрудник лаборатории телематики, Болгарская академия наук (БАН). Область научных интересов: информационные технологии, компьютерные технологии, ИТ-компоненты для поддержки образовательного процесса. Число научных публикаций — 4. iliev.oleg@gmail.com; ул. Академика Георги Бончев, бл. 8, 1113, София, Республика Болгария; р.т.: +359884381052.

Литература

1. *Lee Y.J. et al.* East-Asian Primary Science Curricula: An Overview Using Revised Bloom's Taxonomy // Springer. 2016. 81 p.
2. *Arapı P.* Supporting Personalized Learning Experiences on top of Multimedia Digital Libraries // Ph.D. Thesis. 2017. 277 p.
3. *Balatsoukas P., Moris A., O'Brien A.* Learning objects update: Review and critical approach to content aggregation // Journal of Educational Technology & Society. 2008. vol. 11. no. 2. pp. 119–130.
4. *Morton D., Colbert-Getz J.* Measuring the impact of the flipped anatomy classroom: The importance of categorizing an assessment by Bloom's taxonomy // Anatomical sciences education. 2017. vol. 10. no. 2. pp. 170–175.
5. *Boyle T., Cook J.* Learning objects, pedagogy and reuse // Learning technology in transition. From individual enthusiasm to institutional implementation. 2003. pp. 31–44.
6. *Knight C., Gašević D., Richards G.* Ontologies to integrate learning design and learning content // Journal on Interactive Media in Education. 2005. vol. 2005. pp. 1–24.
7. *Coffield F., Moseley D., Hall E., Ecclestone K.* Learning styles and pedagogy in post-16 learning: a systematic and critical review // Learning and Skills Research Centre 2004. 173 p.
8. *Gallagher J.* The Business Case Study: A Suitable Candidate For Blended Learning? // Journal of Business Case Studies. 2006. vol. 2. no. 4. 14 p.
9. *Gregoric A.* Learning/teaching styles: Potent forces behind them // Educational Leadership. 1979. pp. 36–40.
10. *Hamari J., Koivisto L., Sarsa H.* Does Gamification Work? – A Literature Review of Empirical Studies on Gamification // Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS). 2014. pp. 3025–3034.
11. *Hodgins W.* The future of learning objects // The Instructional Use of Learning Objects Bloomington: IN: АЕСТ. 2002. pp. 281–298.
12. *An D., Carr M.* Learning styles theory fails to explain learning and achievement: Recommendations for alternative approaches // Personality and Individual Differences. 2017. vol. 116. pp. 410–416.
13. *Huotari K., Hamari J.* Defining gamification: a service marketing perspective // Proceedings of the 16th International Academic MindTrek Conference. 2012. pp. 17–22.
14. *James W.B., Gardner D.L.* Learning styles: Implications for distance learning // New directions for adult and continuing education. 1995. vol. 1995. no. 67. pp. 19–31.
15. *Karagiannidis C., Sampson D.* Adaptation rules relating learning styles research and learning objects meta-data // Workshop on Individual Differences in Adaptive Hypermedia. 3rd International Conference on Adaptive Hypermedia and Adaptive Web-based Systems (AH2004). 2004. pp. 66–73.
16. *Kolb D.A.* Experiential Learning: Experience as the Source of Learning and Development: 2nd ed. // Pearson FT Press. 2004. 416 p.
17. *Lennox D.* Managing Knowledge with Learning Objects // WBT Systems White Paper. 2001. 12 p.
18. *Mason R., Rehak D.* Keeping the learning in learning objects // Reusing online resources: a sustainable approach to e-learning. 2003. pp. 20–34.
19. *McLeod S.A.* Kolb's Learning Styles and Experiential Learning Cycle. 2017. 5 p. URL: <https://www.simplypsychology.org/simplypsychology.org-Kolb-Learning-Styles.pdf> (дата обращения: 06.07.2018.).
20. *Metros S.E.* Learning objects: A rose by any other name // Educause Review. 2005. vol. 40. no. 4. pp. 12–13.
21. *Ram A., Leake D.* Goal-Driven Learning // The MIT Press. 1995. 61 p.

22. *Riding R., Rayner S.* Cognitive styles and learning strategies: Understanding style differences in learning and behaviour // David Fulton Publishers. 1998. 217 p.
23. *Robson K. et al.* Is it all a game? Understanding the principles of gamification // Business Horizons. 2015. vol. 58. no. 4. pp. 411–420.
24. *Verbert K., Duval E.* Towards a Global Component Architecture for Learning Objects: A Comparative Analysis of Learning Object Content Models // EdMedia: World Conference on Educational Media and Technology. Association for the Advancement of Computing in Education (AACE). 2004. pp. 202–208.
25. *Verbert K., Duval E.* ALOCOM: a generic content model for learning objects // International Journal on Digital Libraries. 2008. vol. 9. no. 1. pp. 41–63.
26. *Wagner E.D.* Steps to creating a content strategy for your organization” // The e-Learning developers’ journal. 2002. 9 p.
27. *Yoshinov R., Iliev O.* “Controlled self-study” in thematic educational community environment // Proceedings of the Forty-seventh Spring Conference of the Union of Bulgarian Mathematicians. 2018. pp. 200–213.
28. *Yoshinov R., Iliev O.* Content reuse - a major problem with modern content storage systems // Eleventh National Conference with International Participation “Education and Research in the Information Society”. 2018.
29. *Yoshinov R., Kotseva M, Pavlova D.* Specifications for Centralized DataCenter serving the educational cloud for Bulgaria // International conference ETAI. 2015. pp. 1–6.
30. *Yoshinov R., Kotseva M.* The steps for elaboration of the “Rosetta stone” demonstrator // Proceedings of International Conference Inspiring Science Education. 2016. pp. 91–96.
31. *Yoshinov R., Arapi P., Kotseva M., Christodoulakis S.* Supporting Personalized Learning Experiences on top of Multimedia Digital Libraries // International journal of education and information technologies. 2016. vol. 10. pp. 152–158.
32. *Yoshinov R., Pavlova D., Kouzov O.* Reflection of ISE idea for linking school education and scientific research in the National Strategy for effective implementation of ICT in education and science in the Republic of Bulgaria // Proceedings of International Conference Inspiring Science Education. 2016. pp. 129–134.
33. *Yoshinov R., Kotseva M.* Vision for the Engagement of the e-Facilitator in School in the Inspiring Science Education Environment // Serdica Journal of Computing. 2015. vol. 9. no. 3-4. pp. 241–256.
34. *Trifonov R., Yoshinov R., Jekov B., Pavlova G.* Methodology for Assessment of Open Data // International Journal of Computers. 2017. vol. 2. pp. 28–37.

S.B. SUZIĆ, T.V. DELIĆ, S.J. OSTROGONAC, S.V. ĐURIĆ, D.J. PEKAR
**STYLE-CODE METHOD FOR MULTI-STYLE PARAMETRIC
TEXT-TO-SPEECH SYNTHESIS**

Suzić S.B., Delić T.V., Ostrogonac S.J., Đurić S.V., Pekar D.J. Style-Code Method for Multi-Style Parametric Text-to-Speech Synthesis.

Abstract. Modern text-to-speech systems generally achieve good intelligibility. The one of the main drawbacks of these systems is the lack of expressiveness in comparison to natural human speech. It is very unpleasant when automated system conveys positive and negative message in completely the same way. The introduction of parametric methods in speech synthesis gave possibility to easily change speaker characteristics and speaking styles. In this paper a simple method for incorporating styles into synthesized speech by using style codes is presented.

The proposed method requires just a couple of minutes of target style and moderate amount of neutral speech. It is successfully applied to both hidden Markov models and deep neural networks-based synthesis, giving style code as additional input to the model. Listening tests confirmed that better style expressiveness is achieved by deep neural networks synthesis compared to hidden Markov model synthesis. It is also proved that quality of speech synthesized by deep neural networks in a certain style is comparable with the speech synthesized in neutral style, although the neutral-speech-database is about 10 times bigger. DNN based TTS with style codes are further investigated by comparing the quality of speech produced by single-style modeling and multi-style modeling systems. Objective and subjective measures confirmed that there is no significant difference between these two approaches.

Keywords: text-to-speech synthesis, expressive speech synthesis, deep neural networks, speech style, style code, one-hot vector.

1. Introduction. Text-to-speech (TTS) synthesis, a set of techniques that enable computers to convert text to human voice, has been a popular research area during the last few decades. This technology has a wide range of possible usage scenarios. Initially, it was used as a reading aid for blind people. It was also successfully applied in call centers for reading different types of information to the customers. Nowadays, audiobooks are generated by TTS systems and personal assistant applications use this technology to deliver information to its users. There are few different approaches to converting text to speech. Concatenative synthesis [1], a method based on concatenation of authentic speech segments from some prerecorded database, produces high-quality speech, the naturalness of which is still considered to be state-of-the-art. However, in some cases, audible glitches appear, usually in contexts that are not covered by the speech database. Furthermore, creating a new synthetic voice can only be done by obtaining a whole new speech database and spending a significant amount of resources on

the database preparation. A first technique in which some of these drawbacks were overcome is parametric synthesis based on hidden Markov models (HMM) [2]. It produces speech of constant quality, and even smaller speech databases can be used for getting a voice of decent quality. However, because of some drawbacks in modeling approach the speech synthesized by HMM system sounds muffled [3].

In recent years, prevalent research methods for text-to-speech synthesis have been based on deep neural networks (DNN). Reason for this prevalence is considered to be related to immanent characteristics of DNNs that are so-called *deep structures*, in contrast with the already mentioned HMMs that are so-called *shallow structures* [4]. As deep structures are proven to be more appropriate for modeling complex relations between input and output data [5], it was expected that DNNs would be suitable for modeling relations between linguistic features and acoustic parameters. Different papers examined this approach [6-7] and concluded that DNNs are appropriate for usage in TTS, since they provide synthesized speech of high quality. Furthermore, it has been proven that DNNs are better than HMMs in this context, since synthesized voice has even higher quality, comparable to that of concatenative synthesis [4].

The two most important requirements that synthesized speech should fulfill are intelligibility and naturalness. The research community mostly agrees that modern TTS systems achieve good performance regarding these two requirements [8]. The major critique of TTS systems is the uniformity of synthesized speech. Namely, most of the speech is generated in same speaking style and yet modern applications require not only high-quality naturally sounding speech, but also the possibility of changing speaking style, thus allowing users to exchange subtext information. For example, the style in which some news is generated should be different from the style in which commercials or warnings are synthesized, as explained in [9]. In [10] it is stated that some aircraft accidents investigators think that neutral voice, in which warnings in critical situation are generated, are the reason why the passengers do not perceive these situations as potentially dangerous.

Although different speaking styles can be associated with some emotional states, the term speaking style is more general. Emotional state is usually associated with speaker's inner state which affects speech characteristics. The term speaking style in this paper is used to mark any deviation in speech characteristics compared to neutral speech and does not consider the cause or intention for this change.

This paper presents an expansion of [11], the goal of which was to make a multi-style model using style codes in order to enable speech synthesis in different speaking styles, both for HMM and DNN synthesis. Although it is already proven that DNN synthesis achieves better results than HMM [12-14], it is yet to be tested if proposed approach can be used in both synthesis technics and if there is difference in their performance. Since it is assumed that only a small amount of speech material with new speaking style is available, the aim is to test if the speech synthesized in target style has similar quality as the speech synthesized in neutral style, for which much more training material is used. The main contribution of this paper is extensive analysis of the performance of the style code approach in DNN synthesis framework.

The rest of the paper is organized as follows. In section II, parametric approaches to speech synthesis are explained. The review of methods used in style modeling is given in section III, which is followed by an explanation of technic used for creation of multi-style DNN model in section IV. In the section V, the results are presented and discussed. Finally, conclusions are drawn and directions of further research are mentioned.

2. Parametric text-to-speech synthesis. Parametric speech synthesis consists of two phases. First, in training phase, acoustic features are extracted with a vocoder and models are trained on the extracted features. In synthesis phase, models are used to generate acoustic parameters which are converted to the speech samples by a vocoder.

HMM based TTS.

In order to overcome some of the shortcomings of concatenative synthesis (complicated implementation of a new voice, occasional glitches in speech, memory space requirements), statistical parametric methods emerged as the best solution. [15]. These systems model the parameters extracted from the speech by using some generative modeling approaches. The most widely used systems are based on hidden Markov models (HMM). Actually, the terms HMM synthesis and statistical speech synthesis are being used as synonyms in literature. However, one of the parameters that need to be modeled is fundamental frequency, which is not defined in unvoiced regions and therefore usage of standard HMM modeling is not fully applicable. Bearing that in mind, an extended HMM model called multi-space distribution hidden Markov model (MSD-HMM) is proposed and successfully applied in speech synthesis [16].

In the field of automatic speech recognition (ASR) based on HMM, the speech units that are modeled are the triphones, phonemes with known preceding and succeeding phoneme identities. For the purposes of speech synthesis, modeling unit takes a much wider context. Besides the preceding and succeeding phoneme identity, the context factors that are taken into account are different phonetic, linguistic and prosodic information. Since there is not enough training data to adequately model all the contexts, tree-based context clustering is used, meaning that similar contexts share distribution parameters [17]. This, naturally, introduces some unwanted smoothing of the parameters and influences naturalness of synthesized speech. Various techniques have been developed in order to address this problem [18].

In the synthesis stage, the most probable parameter sequence should be generated based on input text and known models' parameters, λ . If the HMM state sequence, q , is known, the solution to this problem can be found by solving the following optimization equation:

$$\hat{O} = \arg \max_o P(O|q, \lambda). \quad (1)$$

The solution to the optimization problem from Equation 1, as well as some other algorithms for generating speech parameters, can be found in [19]. It is proven that inclusion of dynamic features (first and second derivatives) improve overall quality of synthesized speech.

DNN based TTS.

In order to improve modeling of a layered, hierarchical structure of a human speech production system, deep neural networks (DNNs) are becoming dominant in speech synthesis. DNNs manage to achieve better generalization and thus synthesized speech is less smoothed compared to the one synthesized using HMM based approach.

When neural networks are applied to speech synthesis linguistic features are used as inputs. Extraction of linguistic features is carried out on the phoneme level. Therefore, linguistic features most often contain information on phoneme identity, as well as the identity of phoneme's contexts, phoneme's accent, etc. Those features are extracted by a separate front-end module and delivered into two neural networks (Figure 1). The first one is trained to predict phonemes' states durations. In order to get durations on a state level, as targets during the training, initial alignment has to be performed. This procedure is usually done by using monophone models and a couple of rounds of Baum-Welch

algorithm [20]. Inputs and targets of the duration network are phoneme-aligned. The second network is trained independently from the first one, and its task is to predict the acoustic features. Beside the features used as inputs of the duration network, it requires additional inputs specifying within-phone positional information (including state-level durations). Target features for this network are extracted from the training dataset by the vocoder. Inputs and targets are frame-aligned. Networks are trained using back-propagation algorithm.

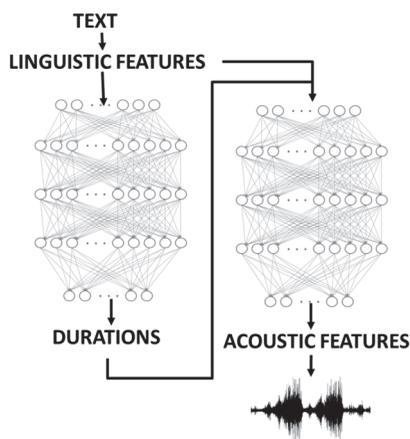


Fig. 1. DNN based TTS model

In the stage of synthesis, state-level durations predicted from the first network are used to extract additional features for the second network, which predicts acoustic features required by vocoder to produce waveforms. It was found that better results are achieved when dynamic acoustic features are used along with the static ones. For this reason, first and second derivatives of acoustic features are also used as targets during the training of the second network. In the synthesis stage, after those are predicted, they are only used by MLPG algorithm [19] in order to slightly correct static acoustic features trajectories. Those static features are smoothed this way and after that propagated to the vocoder.

3. Style modeling in TTS. In [21] approaches to an expressive speech synthesis are categorized in three different groups: implicit control, explicit control and playback approach. As it was mentioned before, expressive speech can be considered as an example of using different styles in speech.

Explicit control approaches are based on using some transformation rules to the prosody of a sentence synthesized in neutral style in order to get speech in some other style. These transformation rules are learned on a database of expressive speech. The examples of these transformations can be found in [22].

Playback approaches to an expressive speech synthesis create separate synthesizers for desired styles based only on expressive speech data. In [23] concatenative speech synthesis system capable of producing speech in three different styles is presented. Each style is synthesized using approximately 1 hour of corresponding speech data. HMM based TTS system that uses different acoustic models for each desired style is presented in [24]. In the same paper, another technique for expressive speech synthesis, which embeds style information in input linguistic features, is proposed. The authors have compared these approaches and concluded that they produce the same results regarding expressive speech quality.

Implicit approaches can be applied in statistical speech synthesis and are based on interpolation between different models. In [25] method which adapts a model of neutral speech to some desired style is described.

The comparison of concatenative and HMM based synthesis regarding expressive speech is given in [26]. It is concluded that concatenative synthesis performs better regarding overall emotion intensity, whereas the HMM approach is better for emotion intensity manipulation.

4. Style-code modeling in DNN based TTS. When we initially started experiments with proposed techniques there were almost no attempts at the expressive speech synthesis using DNN. Meanwhile, several papers were published where similar techniques are proposed [27-29].

Some research related to manipulation of voice intended for speaker modeling includes voice conversion [30], speaker adaptation [31] and multi-speaker synthesis [32]. Voice conversion approaches are based on parallel corpus of source and target speaker. The aim is to make conversion function, which when applied on speech of the source speaker should make it sound as the target speaker. The speaker adaptation approaches try to adapt already trained models towards some target speaker and do not require parallel corpus. Starting models are usually trained on a large training corpus and a small amount of target speaker speech material is used for model adaptation. Multi-speaker synthesis requires a collection of smaller datasets for training, and then, in the synthesis phase, decides on the voice that is to be synthesized.

In [31], [32] and [33] multi-speaker model is made by training a DNN on a speech material from a few to more than 100 speakers. One of the methods involves standard DNN-based single-speaker system, but extended with the feature that represents *speaker code* (Figure 2) [31]. The speaker code can be represented as a one-hot vector, or extended with additional information about a certain speaker, like gender, age, etc. [33]. This extension provides even better results, resembling on the idea of using *i-vectors* as speaker code [34]. In synthesis stage, the network will synthesize the speech in some speaker's voice by setting appropriate speaker code.

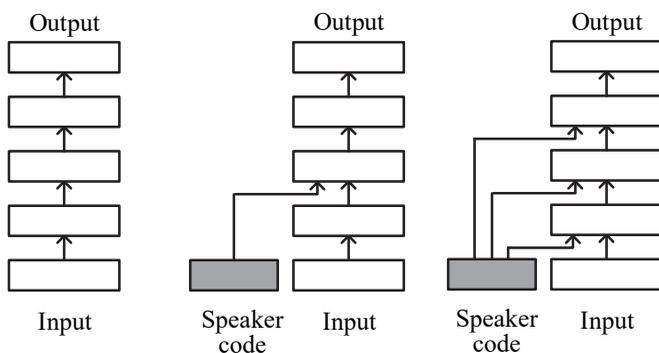


Fig. 2. The idea presented in [30] for multi-speaker DNN TTS model

The idea presented in this paper is to use a single one-hot vector to represent speaking style. It is solved by adding lexical questions of type “*is_style_x*”, where *x* can be neutral, angry, happy, etc. In case of using DNN model, since the input of the neural network represents binary label with value 1 on places representing questions on which answers are positive, part of the input will be one-hot vector indicating speaking style (Figure 3). In case of using HMM modeling, similar idea is used. Input label is extended with information that indicates if a phone belongs to speech of a certain style. In HMM modeling this actually corresponds to the idea presented in [24].

Usually, just a part of the sentence will be said with high expression of emotions and that is the reason why the style code is given per word, although the whole sentences were labeled with a marker of just one style in used database. This way, it should be possible to produce just a part of a sentence with an emotion, and the rest in neutral style.

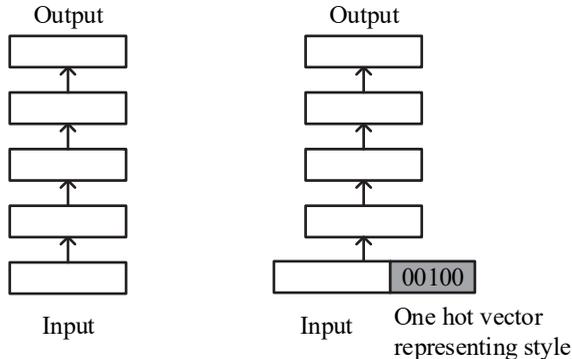


Fig. 3. The idea presented in [34] but modified for multi-style DNN TTS

In multi-style modeling, it is expected that the database consists of a lot more material of neutral speech than any other style. In case of multi-speaker modeling, such big differences between databases of speakers can badly influence final model (e.g. if some features of different speakers are averaged). On the other hand, this effect should have smaller impact in multi-style approach since all sentences are uttered by the same speaker. This could be actually beneficial for this model since the states are not covered with small database of certain speaking style can be compensated by neutral style and it will not badly influence the final model, since the voice is the same in the two databases.

5. Experiments and results.

DNN and HMM based systems.

DNN system is built using Merlin toolkit [35]. It consists of two neural networks, one for duration and one for acoustic modeling. Both networks contain 4 hidden layers with 1024 units per layer. Neurons in first 3 layers use tangent hyperbolic activation function, while the 4th layer is based on long short-term memory (LSTM) architecture [36]. The output layer is linear. Input for duration model consists of answers to 554 closed lexical questions. These include questions about phoneme identity, number of phonemes and syllables in a word, ToBI tags [37] attached to a phoneme, etc. One-hot vector representing style is not included in the overall question number. Outputs are durations per HMM state. Input for acoustic model, beside the answers to the lexical questions, contains 9 additional features regarding state and phoneme durations, which were also introduced in Merlin toolkit. The acoustic

features are extracted by WORLD vocoder [38]. Since it produces smoothed envelopes, they are converted to mel-generalized cepstral coefficients (MGCs). At the end, the output feature vector for acoustic network contains 40 MGCs, 1 band aperiodicity (BAP) feature, logarithm of fundamental frequency, first and second derivatives of previously mentioned features and one feature representing information if a given frame is voiced or unvoiced (VUV).

HMM system is built by using HTS toolkit [39]. HMM models that are used in the system are 5-state left-to-right models with no skip, where each state is represented by a single Gaussian with the diagonal covariance matrix. Same acoustical features as for DNN system are used, except for the VUV feature. Input label consists of lexical features extracted in similar way as for DNN synthesis, but the number of available features was higher. Namely, HMM contained a combination of some features and initial experiments in DNN synthesis showed that usage of these complex questions does not improve the quality of synthesis. This can be explained by the DNN's capability to model some complex relations that HMMs cannot.

Databases.

In order to compare performance of the proposed technic in case of using HMM and DNN, 4h and 20 min of neutral speech and 20 min of speech in angry style were used. The same database is used in case of investigating quality of the synthesized speech, while in case of further investigation of the proposed technic for the DNN based synthesizer, the database is expanded with two more speaking styles – happy and apologetic. More precisely, 2h of neutral speech and 10 min per style are used for these experiments. The whole database is pronounced by a native American English male voice talent. It is recorded in a professional studio. Some statistics can be seen in Table 1.

Table 1. Statistics of the database

Style	Speech rate [phoneme/s]	Average f_0 [Hz]	std f_0 [Hz]
Neutral	12.7	98.7	34.1
Apologetic	10.8	101.9	25.1
Happy	11.4	170.2	71.4
Angry	10.9	103.9	30.3

It can be noticed that the happy style has significantly higher average fundamental frequency (f_0) than other styles, as well as more than two times higher standard deviation of f_0 in comparison to any other style. The neutral style is the fastest and has the lowest average f_0 . Apologetic and angry styles have very similar characteristics with around 15% smaller standard deviation of f_0 in case of apologetic style.

Performance of the proposed technic for different synthesis method.

The two systems were first compared objectively. All objective measures were calculated as differences between acoustic features predicted by neural network and acoustic features extracted directly from original recordings. The objective measures that were used include:

- mel-cepstral distance (MCD) — mean square error of mel generalized cepstral coefficients calculated in decibels,
- mean square error of band aperiodicities (BAP) calculated in decibels,
- root mean square error of fundamental frequency (RMSE f_0) calculated in hertz,
- correlation of fundamental frequency (CORR f_0),
- percentage of correctly predicted voiced frames (V/UV).

The objective measures are calculated for test files only, for both systems and they are shown in Table 2. During the feature generation the original durations created during the DNN alignments procedure were used. Based on the results from Table 2 it can be concluded that DNN system outperforms HMM regarding all objective measures.

Table 2. Objective comparison between HMM and DNN systems

	MCD [dB]	BAP [dB]	RMSE f_0 [Hz]	CORR f_0	V/UV [%]
HMM	6.73	0.18	23.60	0.5	8.64
DNN	4.29	0.15	20.84	0.63	5.52

One example of fundamental frequency trajectory created by the tested systems is shown in Figure 4. Both trajectories are compared with trajectory extracted from the original recording. It can be seen that DNN-predicted trajectory resembles much better the original one than the HMM-predicted trajectory.

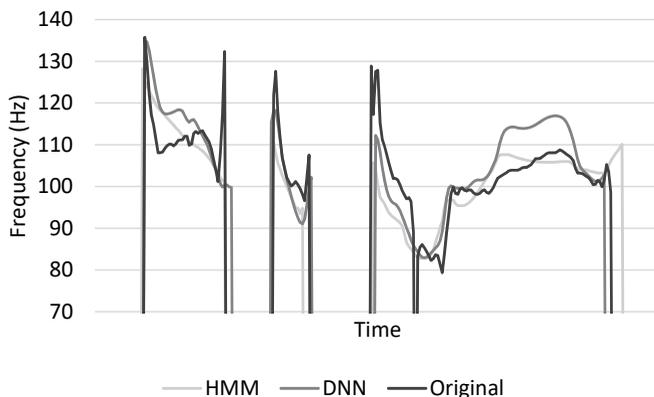


Fig. 4. Comparison of fundamental frequency trajectory created by the HMM and DNN systems with original trajectory

But it is known that objective measures are not always fully correlated with subjective assessment of overall speech quality. Subjective listening tests are considered as the most reliable. Due to this reason, for the comparison of the obtained results from HMM and DNN synthesizers by the proposed style-code method, preference tests were conducted among 24 amateur non-native listeners. They were asked to recognize which sentence, in each of 20 pairs, was pronounced in angry style. One of the possible options was also *No preference* (i.e. none of the sentences stands out). The first 10 pairs consisted of sentences pronounced in neutral and angry style, synthesized by DNN synthesizer, while the remaining 10 pairs (also pronounced in both styles per pair) were synthesized by HMM synthesizer. These results are given in Figure 5. It clearly shows much better results in case of DNN synthesis. As many as 39% of answers were *No preference* in case of HMM synthesis, and even in 1% listeners chose neutral sentence as the one where the angry emotion was better expressed. That leads to the conclusion that HMM model is not able to produce entirely clear difference between styles in case of the proposed technic. On the other hand, only 12% answers were *No preference*, in case of DNN synthesis, while remaining 88% of answers accurately recognized angry style. Since even in spontaneous human communication it is not always easily recognizable if the speaker intended to express some emotion, 88% can be considered as high accuracy and the proposed technic can be considered as very powerful although pretty simple.

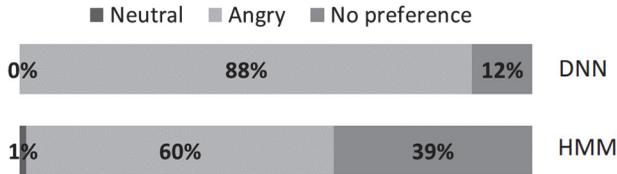


Fig. 5. Results of preference test regarding style expression

In order to directly compare DNN and HMM synthesis in angry style, the second listening test was conducted. It consisted of 10 pair of sentences where each pair of sentences was synthesized by both synthesizers. Listeners were asked to choose sentence in each pair in which angry emotion was more emphasized. As shown in Figure 6, exceptionally high preference for DNN synthesized sentences clearly confirms that the proposed technic has better performances in case of DNN.

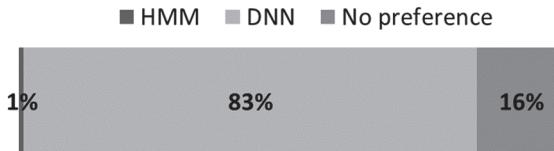


Fig. 6. Direct comparison of expressive speech synthesized by HMM and DNN based synthesizers

Quality of synthesized speech.

As previously mentioned, the key requirements for synthesized speech is intelligibility and naturalness. Intelligibility seized to be an issue a long time ago, but the naturalness is still a big issue in the field. Naturalness is defined as resemblance to human speech. Since for the experiment, almost 10 times bigger amount of neutral speech than angry-style speech was used, it was important to investigate if mismatch in the amount of the used material caused some loss in quality for speech synthesized in angry style, compared to the neutral one. In order to do that, additional listening test was conducted. In this test listeners were asked to grade the quality of 10 sentences synthesized by DNN synthesizer by giving grades between 1 (very bad) and 5 (very good). Among these, 5 are pronounced in angry and 5 in neutral style. Quality of

synthesis is defined to listeners as resemblance to human speech and the absence of the artifacts.

The obtained average grades for the synthesis in neutral style and the synthesis in angry style are almost the same, 3.9 against 3.8 in favor of neutral style. This proves the hypothesis that all states not covered by the small angry-style database are compensated by neutral style.

Objective comparison of results obtained with three different styles.

Since all of the previous experiments were performed by using only a single style (other than neutral), we also tested the performance of the approach in reproduction of other styles. In order to test the proposed method performance, separate synthesizers were constructed, where each one is capable of producing a single style. These synthesizers were created by using 2h of neutral speech and 10 minutes of intended style (angry, apologetic or happy). The results were compared only objectively since subjective comparison of different speaking styles does not represent an adequate approach, due to significant differences in the manners in which the styles are expressed.

For each synthesizer, objective measures were calculated by using 30 test sentences of corresponding speaking style, which were not a part of the training data. The results are given in Table 3. Among the presented styles, the best objective measures are achieved with apologetic style and mostly the worst for happy style. It is the most emphasized for RMSE of f_0 which is 41.48 Hz in case of happy style. This can be explained by its characteristics — mean frequency as well as standard deviation are the highest among all the styles (see Table 1). On the other hand, correlation of f_0 is the best among all three styles.

Table 3 Objective measures 1-style modeling

	MCD [dB]	BAP [dB]	RMSE f_0 [Hz]	CORR f_0	V/UV [%]
Happy	5.50	0.19	41.48	0.79	5.59
Apologetic	4.70	0.13	16.85	0.73	4.88
Angry	4.79	0.17	18.67	0.62	5.68

The obtained results show that some differences exist in modeling different styles, which can be explained by the differences in original part of databases.

It should also be noticed that the objective measures in Tables 3 and 4 for angry style are worse compared to ones from Table 2. This can be

explained by the amount of material of this style being used in training process. Specifically, the amount of angry style material used in experiments whose results are presented in Table 2 was 20 minutes, while in other 2 experiments only 10 minutes of angry speech was used.

Table 4 Objective measures 3-style modeling

	MCD [dB]	BAP [dB]	RMSE f_0 [Hz]	CORR f_0	V/UV [%]
Happy	5.46	0.19	42.85	0.77	5.64
Apologetic	4.67	0.13	17.11	0.72	4.92
Angry	4.75	0.17	18.46	0.63	5.70

Comparison of single-style and multi-style modeling.

As already mentioned in previous sections, the style-code approach can be applied to simultaneous modeling of an arbitrary number of styles. Therefore, we wanted to compare single-style modeling with multi-style approach. For these purposes we constructed a new synthesizer, which was trained on full previously introduced database — 2 hours of neutral speech and 10 minutes of speech of each of the three additional styles. The objective measures calculated in the same manner as for single-style are given in Table 4. As in the case of single-style modeling, the best results are obtained for the apologetic style, while the measures for the happy style again were the worst among all three styles.

The average measures per all three styles for those two approaches are presented in Table 5. These results indicate that measures for both approaches differ only in minor.

Table 5 Average objective measures for 1-style and 3-style modeling

	MCD [dB]	BAP [dB]	RMSE f_0 [Hz]	CORR f_0	V/UV [%]
1-style	5.00	0.16	25.55	0.71	5.38
3-style	4.96	0.16	26.14	0.71	5.42

We also conducted one subjective test, in which 15 non-native listeners were included. The task consisted of 30 pairs of sentences. In each pair there was one sentence synthesized with 1-style approach and another synthesized with 3-style approach. Each of three styles was represented with 10 pairs. Participants had to choose in which of the two sentences the

presented style was better expressed. The intended style of each pair of sentences was clearly presented to the participants. There was also *No preference* as a possible answer.

The results are given in Figure 7. It can be seen that, on average, multi-style model is preferred over the other one. Only for angry style percentage is the same. *No preference* answer is chosen most often in case of apologetic and angry style, while in case of the happy style, multi-style model is chosen significantly more often than the other two possible answers.

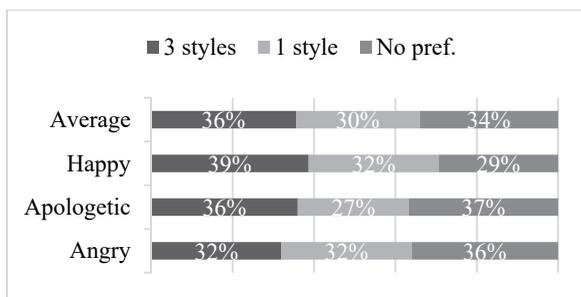


Fig.7. Results of subjective tests for comparing single-style and multi-style modeling

As conclusion it can be stated that there is no significant difference between modeling one style per model or multiple at once, although the slight advantage can be given to multi-style model, probably because of better generalization.

Analysis based on parallel corpus.

There are a few sentences in our database originally recorded in different styles but with the same content spoken (i.e. parallel corpus). We have chosen one of these sentences and performed further analysis in order to check the impact of style code to parameter generation, as well as the impact of some linguistic differences to generated parameters. The chosen sentence is originally recorded in happy and apologetic style and annotated in accordance to the actual prosodic events. In this particular sentence, there are 2 more phrase breaks in apologetic sentence in comparison to the happy one. All other prosodic annotations are the same.

In this experiment, the focus was on predicted fundamental frequency trajectories. Although the same content was spoken in both sentences the length of corresponding phonemes is not the same. Namely, according to Table 1 the phoneme rate is slightly higher for apologetic style.

This makes the direct comparison not possible. In order to make direct comparison feasible original durations were used.

Figure 8 shows generated trajectories when linguistic annotations and durations of original apologetic sentence are used. It can be seen that fundamental frequency of the original recording (black line) does not vary much (mostly it is between 60 Hz and 120 Hz) and its average is around 95 Hz, which is in accordance to the style characteristic (Table 1). It can also be noticed that the synthesis in apologetic style follow the original trajectory very well. On the other hand, synthesis of the same sentence with the same durations and annotations and switching only the style code to happy produced different curve. The average frequency is increased for about 65 Hz, and it varies a lot more than the other two curves in Figure 8, reaching even 300 Hz. This behaviour is also in accordance to happy style characteristics presented in Table 1.

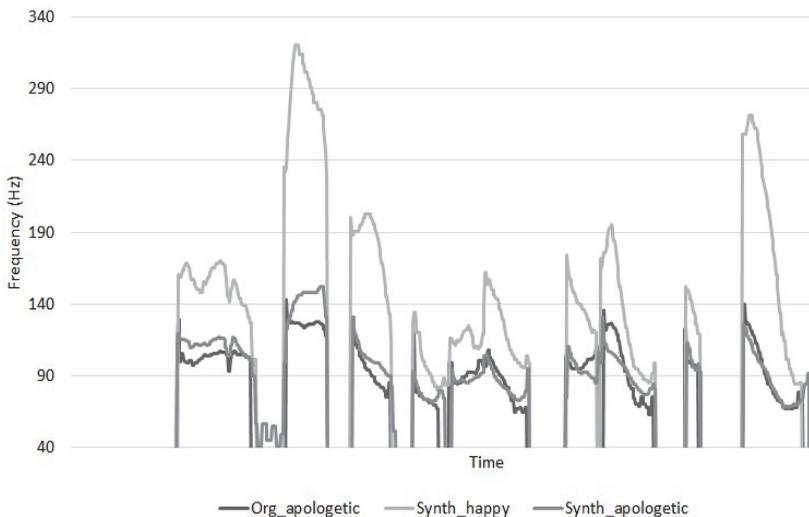


Fig. 8. Fundamental frequency trajectories of the sentences originally anotated as apologetic, synthesized in happy and apologetic style

The case when annotations and durations for original happy sentence are used are shown in Figure 9. Again, it can be seen that the sentence synthesized in apologetic style, has a lot lower fundamental frequency compared to both, original and sentence synthesized in happy style. Synthesized happy sentence follows the original trajectory of fundamental

frequency very well, but sometimes fails to track some rush changes from original recording. This can be explained by certain smoothing that is introduced by the model.

The case when annotations and durations for original happy sentence are used are shown in Figure 9. Again, it can be seen that the sentence synthesized in apologetic style, has a lot lower fundamental frequency compared to both, original and sentence synthesized in happy style. Synthesized happy sentence follows the original trajectory of fundamental frequency very well, but sometimes fails to track some rush changes from original recording. This can be explained by certain smoothing that is introduced by the model.

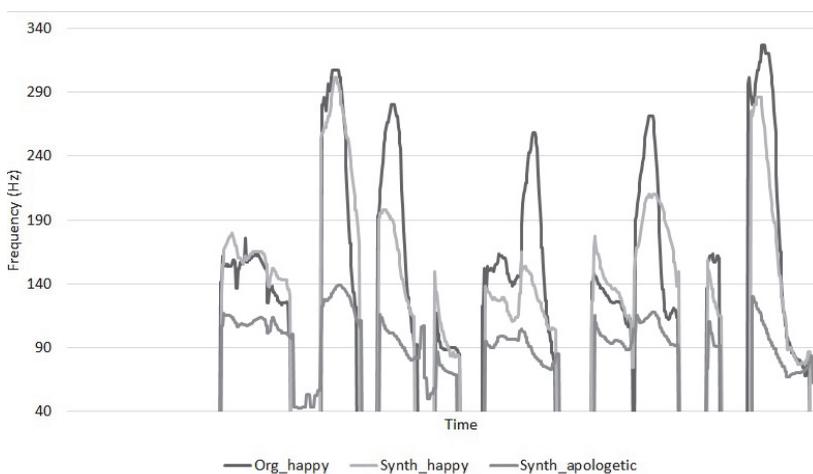


Fig. 9. Fundamental frequency trajectories of the sentences originally anoted as happy, synthesized in happy and apologetic style

Both examples prove that the style code itself succesfully predicts important style characteristics and that these characteristics are not highly dependent on linguistic input.

The impact of linguistic differences on the generated trajectories can be analyzed in Figure 10. These trajectories are generated using linguistic features from both original recordings, which have the same spoken content but differ in their annotations. The biggest differences in the generated trajectories can be observed in the middle of the sentence where the phrase brakes are annotated differently.

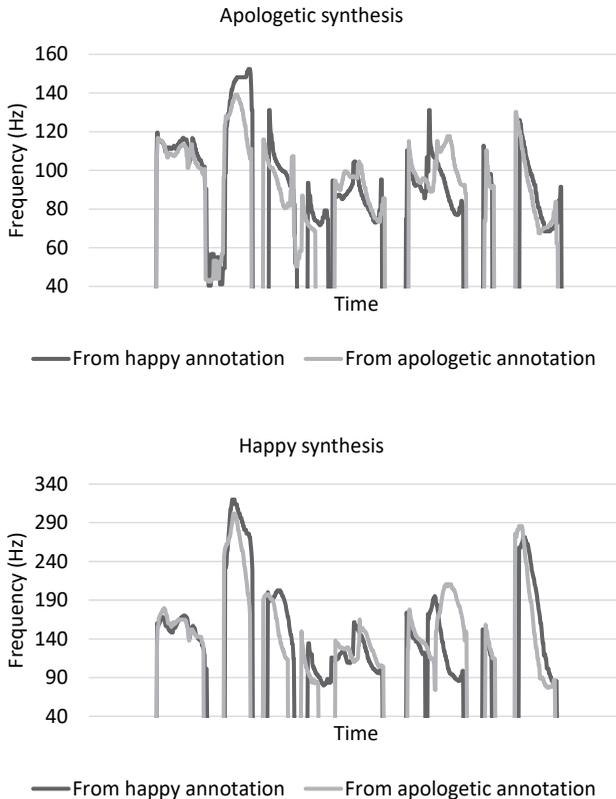


Fig. 10. Comparison of linguistic features impact on fundamental frequency trajectory

One more characteristic that the model learned, but cannot be seen from figures, is loudness of certain styles. In case of apologetic style, the speech is noticeable quieter, e.g. in this particular sentence, the difference between synthesized happy and apologetic sentence is around 10 dB.

4. Conclusion. In this paper we present a simple but very effective method for incorporating styles into statistical speech synthesis. It is based on style codes, similar to earlier introduced speaker codes, and consists of adding one-hot vector representing style to the auxiliary inputs used in NN-based speech synthesis.

The proposed method is also applicable to HMM based speech synthesis. Objective and subjective results show that the proposed method achieves higher performances in case of DNN systems. Objective and

subjective results also suggest that, although the amount of speech material of certain style is much smaller compared to material in neutral style, quality of speech synthesized in certain style is preserved. It is proved that the method is applicable to any speaking style and that there is no significant difference if the model learns multiple styles at once or one by one.

Although the duration is important feature in expressing some style, the focus of this paper was on analysis of the acoustic features. Some analysis on the duration prediction performance should be performed.

Bearing in mind that humans are able to control the level of expressed emotion in speech and rarely one emotion is completely distinguished from another, some future work should investigate possibilities of controlling level of expressed emotion as well as mixing them.

References

1. Hunt A.J., Black A.W. Unit selection in a concatenative speech synthesis system using a large speech database. *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP-96)*. 1996. vol. 1. pp. 373–376.
2. Tokuda K. et al. Speech synthesis based on hidden Markov models. *Proceedings of the IEEE*. 2013. vol. 101. no. 5. pp. 1234–1252.
3. Watts O. et al. From HMMs to DNNs: where do the improvements come from? *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2016. pp. 5505–5509.
4. Ling Z.H. et al. Deep learning for acoustic modeling in parametric speech generation: A systematic review of existing techniques and future trends. *IEEE Signal Processing Magazine*. 2015. vol. 32. no. 3. pp. 35–52.
5. Yu D., Deng L. Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*. 2014. vol. 7. no. 3-4. pp. 198–387.
6. Qian Y., Fan Y., Hu W., Soong F.K. On the Training Aspects of Deep Neural Network (DNN) for Parametric TTS Synthesis. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2014. pp. 3829–3833.
7. Delić T., Sečujski M., Sinteza govora na srpskom jeziku zasnovana na veštačkim neuralnim mrežama. *Telecommunication forum (TELFOR 2016)*. 2016. pp. 403–406.
8. Solomennik A.I., Chistikov P.G., Evaluation of naturalness of synthesized speech with different prosodic models. *Proceedings International conference on Computational Linguistics and Intellectual Technologies “Dialogue 2013”*. 2013. 7 p.
9. Abe M. Speaking Styles: Statistical Analysis and Synthesis by a Text-to-Speech System. *Progress in speech synthesis*. 1997. pp. 495–510.
10. Rusko M. et al. Expressive Speech Synthesis for Critical Situations. *Computing and Informatics*. 2015. vol. 33. no. 6. pp. 1312–1332.
11. Delić T. et al. Multi- style Statistical Parametric TTS. *Proceedings Digital speech and image processing (DOGS 2017)*. 2017. pp. 5–8.
12. Wu Z., Valentini-Botinhao C., Watts O., King S. Deep Neural Networks employing multi-task learning and stacked bottleneck features for speech synthesis. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2015. pp. 4460–4464.
13. Watts O. et al. From HMMs to DNNs: Where do the improvements come from? *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2016. pp. 5505–5509.

14. Delić T., Sečujski M., Suzić S. A review of Serbian parametric speech synthesis based on deep neural networks. *Telfor Journal*. 2017. vol. 9. no. 1. pp. 32–37.
15. Zen H., Tokuda K., Black A.W. Statistical parametric speech synthesis. *Speech Communication*. 2009. vol. 51. no. 11. pp. 1039–1064.
16. Zen H. et al. A hidden semi-Markov model-based speech synthesis system. *IEICE transactions on information and systems*. 2007. vol. 90. no. 5. pp. 825–834.
17. Yoshimura T. et al. Simultaneous modeling of spectrum, pitch and duration in HMM-based speech synthesis. Sixth European Conference on Speech Communication and Technology. 1999. 4 p.
18. Toda T., Tokuda K. A speech parameter generation algorithm considering global variance for HMM-based speech synthesis, *IEICE transactions on information and systems*. 2007. vol. E90-D. no. 5. pp. 816–824.
19. Tokuda K. et al. Speech Parameter Generation Algorithms for HMM-based Speech Synthesis. International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2000. pp. 1315–1318.
20. Rabiner L.R. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE. 1989. vol. 77. no. 2. pp. 257–286.
21. Schröder M. Expressive speech synthesis: Past, present, and possible futures. Affective information processing. 2009. pp. 111–126.
22. Tao J., Kang Y., Li A. Prosody conversion from neutral speech to emotional speech. *IEEE Transactions on Audio, Speech, and Language Processing*. 2006. vol. 14. no. 4. pp. 1145–1153.
23. Iida A., Campbell N., Higuchi F., Yasumura M. A corpus-based speech synthesis system with emotion. *Speech Communication*. 2003. vol. 40. no. 1–2. pp. 161–187.
24. Yamagishi J., Onishi K., Masuko T., Kobayashi T. Modeling of various speaking styles and emotions for HMM-based speech synthesis. Eighth European Conference on Speech Communication and Technology. 2003. pp. 2461–2464.
25. Yamagishi J. et al. Model adaptation approach to speech synthesis with diverse voices and styles. International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007). 2007. vol. 4. p. IV-1233–IV-1236.
26. Barra-Chicote R. et al. Analysis of statistical parametric and unit selection speech synthesis systems applied to emotional speech. *Speech Communication*. 2010. vol. 52. no. 5. pp. 394–404.
27. Inoue K. et al. An investigation to transplant emotional expressions in DNN-based TTS synthesis. Proc. APSIPA Annual Summit and Conference. 2017. pp. 1253–1258.
28. An S., Ling Z., Dai L. Emotional statistical parametric speech synthesis using LSTM-RNNs. Proc. APSIPA Annual Summit and Conference. 2017. pp. 1613–1616.
29. Lorenzo-Trueba J. et al. Investigating different representations for modeling and controlling multiple emotions in DNN-based speech synthesis. *Speech Communication*. 2018. vol. 99. pp. 135–143.
30. Stylianou Y., Cappe O., Moulines E. Statistical Methods for Voice Quality Transformation. Fourth European Conference on Speech Communication and Technology. 1995. pp. 447–450.
31. Hojo N., Ijima Y., Mizuno H. An Investigation of DNN-Based Speech Synthesis Using Speaker Codes. INTERSPEECH. 2016. pp. 2278–2282.
32. Fan Y., Qian Y., Soong F.K., He L. Multi-speaker modeling and speaker adaptation for DNN-based TTS synthesis. International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2015. pp. 4475–4479.
33. Luong H.T., Takaki S., Henter G.E., Yamagishi J. Adapting and controlling DNN-based speech synthesis using input codes. International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2017. pp. 4905–4909.

34. Yang S., Wu Z., Xie L. On the Training of DNN-based Average Voice Model for Speech Synthesis. Proc. Signal and Information Processing Association Annual Summit and Conference (APSIPA). 2016. pp. 1–6.
35. Wu Z., Watts O., King S. Merlin: An Open Source Neural Network Speech Synthesis System. Proc. 9th ISCA Speech Synthesis Workshop (SSW9). 2016. pp. 218–223.
36. Fan Y., Qian Y., Xie F.L., Soong F.K. TTS synthesis with bidirectional LSTM based recurrent neural networks. Fifteenth Annual Conference of the International Speech Communication Association INTERSPEECH. 2014. pp. 1964–1968.
37. Silverman K. et al. ToBI: A standard for labeling English prosody. Proc. International Conference on Spoken Language Processing (ICSLP). 1992. pp. 867–870.
38. Morise M., Yokomori F., Ozawa K. WORLD: a vocoder-based high-quality speech synthesis system for real-time applications. *IEICE transactions on information and systems*. 2016. vol. E99-D. no. 7. pp. 1877–1884.
39. Zen H. et al. The HMM-based speech synthesis system (HTS) version 2.0. Proc. Sixth ISCA Workshop on Speech Synthesis. 2007. pp. 294–299.

Suzić Siniša — researcher of Laboratory of Acoustics and Speech Technology of Faculty of Technical Sciences, University of Novi Sad. Research interests: expressive speech synthesis, digital signal processing, dialogue systems, machine learning, deep neural networks. The number of publications — 22. sinisa.suzic@uns.ac.rs; 6, Trg Dositeja Obradovića, 21000, Novi Sad, Serbia; office phone: +381-21-485-2521.

Delić Tijana Vlado — researcher of Laboratory of Acoustics and Speech Technology of Faculty of Technical Sciences, University of Novi Sad. Research interests: expressive speech synthesis, digital signal processing, dialogue systems, machine learning, deep neural networks. The number of publications — 18. tijanadelic@uns.ac.rs; ; office phone: +381(21)485-2521.

Ostrogonač Stevan — senior researcher, AlfaNum – Speech Technologies, software developer, AlfaNum – Speech Technologies. Research interests: text-to-speech synthesis, automatic speech recognition, natural language processing, dialogue systems, development of speech and language resources, machine learning, neural networks. The number of publications — 39. ostrogonač.stevan@alfanum.co.rs; 40, Bulevar Vojvode Stepe, 21000, Novi Sad, Serbia; office phone: +381-64-845-5302.

Durić Simona — researcher of Laboratory of Acoustics and Speech Technology of Faculty of Technical Sciences, University of Novi Sad. Research interests: expressive speech synthesis, digital signal processing, dialogue systems, machine learning, deep neural networks. The number of publications — 6. simona.djuric@uns.ac.rs; 6, Trg Dositeja Obradovića, 21000, Novi Sad, Serbia; office phone: +381(21)485-2521.

Pekar Darko Jovan — research assistant of the Department for Power, Electronic and Telecommunications Engineering of the Faculty of Technical Sciences, University of Novi Sad, CEO (Chief Executive Officer), AlfaNum Speech Technologies. Research interests: human-computer interaction, speech recognition and synthesis, speaker identification, emotion recognition, speech morphing, numerical simulations, artificial intelligence. The number of publications — 100. darko.pekar@alfanum.co.rs; 40, Bulevar Vojvode Stepe, 21000, Novi Sad, Serbia; office phone: +381-21-485-2521.

Acknowledgements. The research is supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (grant TR32035).

С. Сузич, Т.В. Делич, С. Острогонац, С. Джурич, Д.Й. ПЕКАР
**МЕТОД СТИЛЕВЫХ КОДОВ ДЛЯ МНОГОСТИЛЕВОГО
ПАРАМЕТРИЧЕСКОГО СИНТЕЗА РЕЧИ ПО ТЕКСТУ**

Сузич С., Делич Т.В., Острогонац С., Джурич С., Пекар Д.Й. Метод стилевых кодов для многостилевого параметрического синтеза речи по тексту.

Аннотация. Современные системы преобразования текста в речь обычно обеспечивают хорошую разборчивость. Один из главных недостатков этих систем — отсутствие выразительности по сравнению с естественной человеческой речью. Особенно неприятно воспринимается на слух, когда автоматическая система передает утвердительные и отрицательные предложения совершенно одинаково. Введение параметрических методов в синтезе речи дало возможность легко изменять характеристики говорящего и стили речи. В статье представлен простой способ включения стилей в синтезированную речь, используя стилевые коды.

Предлагаемый метод требует не более пары минут заданного стиля и некоторый объем данных нейтральной речи. Он успешно применяется в синтезе речи на глубоких нейронных сетях и в скрытых марковских моделях, предоставляя стилевой код как дополнительный вклад в модель. Аудирование подтвердило, что наибольшая выразительность достигается за счет синтеза глубоких нейронных сетей по сравнению с синтезом скрытых марковских моделей. Также доказано, что качество речи, синтезированное глубокими нейронными сетями в определенном стиле, сопоставимо с речью, синтезированной в нейтральном стиле, хотя база данных нейтральной речи примерно в 10 раз больше. Глубокие нейронные сети на основе синтеза речи по тексту со стилевыми кодами изучаются путем сравнения качества речи, создаваемой системами одностилевого моделирования и многостилевого моделирования. Объективные и субъективные измерения подтвердили, что между этими двумя подходами нет существенной разницы.

Ключевые слова: синтез речи по тексту, экспрессивный синтез речи, глубокие нейронные сети, стиль речи, стилевой код, прямой унитарный вектор.

Сузич Синиша — научный сотрудник лаборатории акустики и речи факультета технических наук, Нови-Садский университет. Область научных интересов: синтез выразительной речи, обработка цифровых сигналов, диалоговая система, машинное обучение, глубокие нейронные сети. Число научных публикаций — 22. sinisa.suzic@uns.ac.rs; Трг Доситея Обрадовича, 6, 21000, Нови Сад, Сербия; р.т.: +381(21)485-2521

Делич Тийана Владо — научный сотрудник лаборатории акустики и речи факультета технических наук, Нови-Садский университет. Область научных интересов: синтез выразительной речи, обработка цифровых сигналов, диалоговая система, машинное обучение, глубокие нейронные сети. Число научных публикаций — 18. tjanadelic@uns.ac.rs; Трг Доситея Обрадовича, 6, 21000, Нови Сад, Сербия; р.т.: +381(21)485-2521.

Острогонац Стеван — старший научный сотрудник, AlfaNum – Speech Technologies Ltd, разработчик программного обеспечения, AlfaNum – Speech Technologies Ltd. Область научных интересов: синтез речи, автоматическое распознавание речи, обработка естественного языка, диалоговая система, разработка речевых и языковых ресурсов, машинное обучение, нейронные сети. Число научных публикаций — 39. ostrogonac.stevan@alfanum.co.rs; бул. Войводе Степе, 40, 21000, Нови Сад, Сербия; р.т.: +381-64-845-5302.

Джурич Симона — научный сотрудник лаборатории акустики и речи факультета технических наук, Нови-Садский университет. Область научных интересов: синтез выразительной речи, обработка цифровых сигналов, диалоговая система, машинное обучение, глубокие нейронные сети. Число научных публикаций — 6. simona.djuric@uns.ac.rs; Трг Доситея Обрадовича, 6, 21000, Нови Сад, Сербия; р.т.: +381(21)485-2521.

Пекар Дарко Йован — младший научный сотрудник департамента энергетики, электроники и телекоммуникационного инжиниринга факультета технических наук, Нови-Садский университет, главный исполнительный директор, AlfaNum Speech Technologies. Область научных интересов: человеко-машинное взаимодействие, распознавание и синтез речи, идентификация диктора, морфинг речи, статистический анализ, искусственный интеллект. Число научных публикаций — 100. darko.pekar@alfanum.co.rs; бул. Войводе Степе, 40, 21000, Нови Сад, Сербия; р.т.: +381-21-485-2521.

Поддержка исследований. Работа выполнена при финансовой поддержке Министерства образования, науки и технологического развития Республики Сербия (грант TR32035).

Литература

1. *Hunt A.J., Black A.W.* Unit selection in a concatenative speech synthesis system using a large speech database // *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP-96)*. 1996. vol. 1. pp. 373–376.
2. *Tokuda K. et al.* Speech synthesis based on hidden Markov models // *Proceedings of the IEEE*. 2013. vol. 101. no. 5. pp. 1234–1252.
3. *Watts O. et al.* From HMMs to DNNs: where do the improvements come from? // *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2016. pp. 5505–5509.
4. *Ling Z.H. et al.* Deep learning for acoustic modeling in parametric speech generation: A systematic review of existing techniques and future trends // *IEEE Signal Processing Magazine*. 2015. vol. 32. no. 3. pp. 35–52.
5. *Yu D., Deng L.* Deep learning: methods and applications // *Foundations and Trends® in Signal Processing*. 2014. vol. 7. no. 3-4. pp. 198–387.
6. *Qian Y., Fan Y., Hu W., Soong F.K.* On the Training Aspects of Deep Neural Network (DNN) for Parametric TTS Synthesis // *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2014. pp. 3829–3833.
7. *Delić T., Sečujski M.* Sinteza govora na srpskom jeziku zasnovana na veštačkim neuralnim mrežama // *Telecommunication forum (TELFOR 2016)*. 2016. pp. 403–406.
8. *Solomennik A.I., Chistikov P.G.* Evaluation of naturalness of synthesized speech with different prosodic models // *Proceedings International conference on Computational Linguistics and Intellectual Technologies “Dialogue 2013”*. 2013. 7 p.
9. *Abe M.* Speaking Styles: Statistical Analysis and Synthesis by a Text-to-Speech System // *Progress in speech synthesis*. 1997. pp. 495–510.
10. *Rusko M. et al.* Expressive Speech Synthesis for Critical Situations // *Computing and Informatics*. 2015. vol. 33. no. 6. pp. 1312–1332.
11. *Delić T. et al.* Multi- style Statistical Parametric TTS // *Proceedings Digital speech and image processing (DOGS 2017)*. 2017. pp. 5–8.
12. *Wu Z., Valentini-Botinhao C., Watts O., King S.* Deep Neural Networks employing multi-task learning and stacked bottleneck features for speech synthesis // *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2015. pp. 4460–4464.

13. *Watts O. et al.* From HMMs to DNNs: Where do the improvements come from? // International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2016. pp. 5505–5509.
14. *Delić T., Sečujski M., Suzić S.* A review of Serbian parametric speech synthesis based on deep neural networks // Telfor Journal. 2017. vol. 9. no. 1. pp. 32–37.
15. *Zen H., Tokuda K., Black A.W.* Statistical parametric speech synthesis // Speech Communication. 2009. vol. 51. no. 11. pp. 1039–1064.
16. *Zen H. et al.* A hidden semi-Markov model-based speech synthesis system // IEICE transactions on information and systems. 2007. vol. 90. no. 5. pp. 825–834.
17. *Yoshimura T. et al.* Simultaneous modeling of spectrum, pitch and duration in HMM-based speech synthesis // Sixth European Conference on Speech Communication and Technology. 1999. 4 p.
18. *Toda T., Tokuda K.* A speech parameter generation algorithm considering global variance for HMM-based speech synthesis // IEICE transactions on information and systems. 2007. vol. E90-D. no. 5. pp. 816–824.
19. *Tokuda K. et al.* Speech Parameter Generation Algorithms for HMM-based Speech Synthesis // International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2000. pp. 1315–1318.
20. *Rabiner L.R.* A tutorial on hidden Markov models and selected applications in speech recognition // Proceedings of the IEEE. 1989. vol. 77. no. 2. pp. 257–286.
21. *Schröder M.* Expressive speech synthesis: Past, present, and possible futures // Affective information processing. 2009. pp. 111–126.
22. *Tao J., Kang Y., Li A.* Prosody conversion from neutral speech to emotional speech // IEEE Transactions on Audio, Speech, and Language Processing. 2006. vol. 14. no. 4. pp. 1145–1153.
23. *Iida A., Campbell N., Higuchi F., Yasumura M.* A corpus-based speech synthesis system with emotion // Speech Communication. 2003. vol. 40. no. 1-2. pp. 161–187.
24. *Yamagishi J., Onishi K., Masuko T., Kobayashi T.* Modeling of various speaking styles and emotions for HMM-based speech synthesis // Eighth European Conference on Speech Communication and Technology. 2003. pp. 2461–2464.
25. *Yamagishi J. et al.* Model adaptation approach to speech synthesis with diverse voices and styles // International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007). 2007. vol. 4. p. IV-1233–IV-1236.
26. *Barra-Chicote R. et al.* Analysis of statistical parametric and unit selection speech synthesis systems applied to emotional speech // Speech Communication. 2010. vol. 52. no. 5. pp. 394–404.
27. *Inoue K. et al.* An investigation to transplant emotional expressions in DNN-based TTS synthesis // Proc. APSIPA Annual Summit and Conference. 2017. pp. 1253–1258.
28. *An S., Ling Z., Dai L.* Emotional statistical parametric speech synthesis using LSTM-RNNs // Proc. APSIPA Annual Summit and Conference. 2017. pp. 1613–1616.
29. *Lorenzo-Trueba J. et al.* Investigating different representations for modeling and controlling multiple emotions in DNN-based speech synthesis // Speech Communication. 2018. vol. 99. pp. 135–143.
30. *Stylianou Y., Cappe O., Moulines E.* Statistical Methods for Voice Quality Transformation // Fourth European Conference on Speech Communication and Technology. 1995. pp. 447–450.
31. *Hojo N., Ijima Y., Mizuno H.* An Investigation of DNN-Based Speech Synthesis Using Speaker Codes // INTERSPEECH. 2016. pp. 2278–2282.

32. *Fan Y., Qian Y., Soong F.K., He L.* Multi-speaker modeling and speaker adaptation for DNN-based TTS synthesis // International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2015. pp. 4475–4479.
33. *Luong H.T., Takaki S., Henter G.E., Yamagishi J.* Adapting and controlling DNN-based speech synthesis using input codes // International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2017. pp. 4905–4909.
34. *Yang S., Wu Z., Xie L.* On the Training of DNN-based Average Voice Model for Speech Synthesis // Proc. Signal and Information Processing Association Annual Summit and Conference (APSIPA). 2016. pp. 1–6.
35. *Wu Z., Watts O., King S.* Merlin: An Open Source Neural Network Speech Synthesis System // Proc. 9th ISCA Speech Synthesis Workshop (SSW9). 2016. pp. 218–223.
36. *Fan Y., Qian Y., Xie F.L., Soong F.K.* TTS synthesis with bidirectional LSTM based recurrent neural networks // Fifteenth Annual Conference of the International Speech Communication Association INTERSPEECH. 2014. pp. 1964–1968.
37. *Silverman K. et al.* ToBI: A standard for labeling English prosody // Proceedings of International Conference on Spoken Language Processing (ICSLP). 1992. pp. 867–870.
38. *Morise M., Yokomori F., Ozawa K.* WORLD: a vocoder-based high-quality speech synthesis system for real-time applications // IEICE transactions on information and systems. 2016. vol. E99-D. no. 7. pp. 1877–1884.
39. *Zen H. et al.* The HMM-based speech synthesis system (HTS) version 2.0 // Proceedings of Sixth ISCA Workshop on Speech Synthesis. 2007. pp. 294–299.

А.А. АШИМОВ, А.С. ГЕЙДА, И.В. ЛЫСЕНКО, Р.М. ЮСУПОВ
**ЭФФЕКТИВНОСТЬ ФУНКЦИОНИРОВАНИЯ И ДРУГИЕ
ОПЕРАЦИОННЫЕ СВОЙСТВА СИСТЕМ: ЗАДАЧИ И МЕТОД
ОЦЕНИВАНИЯ**

Ашимов А.А., Гейда А.С., Лысенко И.В., Юсупов Р.М. **Эффективность функционирования и другие операционные свойства систем: задачи и метод оценивания.**

Аннотация. Представлен краткий обзор исследования эффективности целенаправленных процессов функционирования систем, выполнен обзор некоторых актуальных задач исследования операционных свойств, решаемых за рубежом. На основе анализа особенностей исследования эффективности целенаправленных процессов функционирования систем и ряда других операционных свойств отечественными и зарубежными авторами сделан вывод об актуальности решения современных задач исследования операционных свойств на основе аналитического оценивания их показателей. Описаны новые направления исследований операционных свойств систем и процессов их функционирования, таких как потенциал систем и операционные свойства использования информационных технологий. Приведены особенности совершенствуемых систем и процессов их функционирования, а также роль информационных технологий при функционировании совершенствуемых систем.

Описаны особенности целенаправленных изменений систем и переходных процессов при функционировании совершенствуемых систем. Объяснена роль информационных технологий при изменениях функционирования систем. Полученные результаты позволили исследовать операционные свойства совершенствуемых систем и использования информационных технологий. Предложены примеры моделей формирования эффектов функционирования совершенствуемой системы. Модели разработаны с учетом реализации информационных и неинформационных действий при функционировании такой системы. Для унифицированного оценивания показателей операционных свойств предложен метод аналитического оценивания, который основан на применении последовательности вводимых унифицированных схем оценивания. Предложена последовательность из трех таких схем, позволяющая оценить все описанные операционные свойства. Раскрыты особенности оценивания с использованием предложенных схем. Полученные результаты должны позволить перейти к решению задач исследования на математических моделях. Полученные результаты проиллюстрированы на примере расчета показателей операционных свойств использования информационных технологий.

Ключевые слова: эффективность, потенциал системы, операционные свойства, информационные технологии, цели, показатели, модели, моделирование, методы.

1. Введение. Операционные свойства систем и процессов их функционирования (далее — ОС) составляют ядро свойств, определяющих качество систем и процессов их функционирования (а также их изменений), поскольку описывают основные результаты использования систем на практике и их соответствие требованиям в разных условиях. Совершенствование систем и процессов их функционирования перспективно осуществлять аналитически, оценивая показатели

ОС в зависимости от переменных и параметров в решаемых задачах. В статье приведены основные особенности исследования ОС в приложении к задачам, решаемым при исследовании систем, функционирование которых имеет технологический вид (задано технологической документацией). К ним относятся, например, производственные системы, системы реализации целевых программ и проектов. Сделан вывод об актуальности аналитического оценивания ОС указанных систем и их функционирования при решении ряда практических задач. Для аналитического оценивания ОС совершенствуемых систем необходимо выполнить моделирование таких систем с учетом ряда его особенностей. А именно на основе заданных технологической документацией закономерностей формирования результатов элементарных технологических операций разных видов необходимо рассчитать характеристики результатов функционирования системы. Затем на их основе, зная возможные требования к функционированию, необходимо рассчитать показатели ОС. Приведены схемы такого оценивания и рассмотрен пример. Показано, что полученные результаты позволяют перейти к решению ряда актуальных практических задач: математических задач аналитического оценивания, анализа ОС исследуемых систем, а затем и задач синтеза по оцениваемым показателям ОС.

2. Особенности исследования операционных свойств отечественными и зарубежными авторами. Труды отечественных ученых, работающих в рамках научного направления «Исследование эффективности целенаправленных процессов» традиционно перечисляют, начиная с известной работы академика А.Н. Колмогорова [1], в работе которого впервые был изложен ряд методов оценивания эффективности стрельбы на основе введенной автором вероятностной меры. Эти методы затем были развиты Е. С. Вентцель [2], много лет проработавшей в военно-воздушной инженерной академии имени Н. Е. Жуковского. После войны она заинтересовалась задачей [3] оценивания эффективности различных видов вооружения, боеприпасов и способов организации огневых средств при стрельбе по летящим объектам. По предложению академика А. Н. Колмогорова, за показатель эффективности ПВО была взята вероятность поражения воздушной цели, вычисляемая с помощью распределения вероятностей точек разрыва снаряда в окрестности цели и условных вероятностей уничтожения цели, зависящих от точки, в которой мог бы произойти разрыв.

В 70 годах XX века эти методы получили дальнейшее развитие в трудах кафедры боевой эффективности военно-космической академии имени А. Ф. Можайского, руководимой в то время Р. М. Юсуповым, ныне научным руководителем Санкт-Петербургского института информатики и автоматизации Российской академии

наук [4, 5]. Там же, в академии имени А. Ф. Можайского, Г. Б. Петуховым и его учениками разрабатывалась теория эффективности целенаправленных процессов (ТЭЦП) [6, 7]. В ней показателем эффективности является вероятностная мера соответствия характеристик случайных эффектов целенаправленного процесса требуемым (директивно заданным) значениям этих характеристик.

Примерно в это же время на основе методов, предложенных А. Н. Колмогоровым, были выполнены исследования повышения эффективности процессов создания, производства и применения различных видов военной техники [8-18]. Так, например, исследования эффективности применения ракетных комплексов [8, 9] проводились коллективом ученых военной академии РВСН имени Ф. Э. Дзержинского [11-13]. Выполнялись исследования эффективности функционирования сложных систем разной природы Н. А. Северцевым [8], А. В. Ильичевым [16], В. В. Подиновским [17]. Коллектив ученых военно-морской академии им. Н. Г. Кузнецова [18] под руководством И. А. Рябина за создание и внедрение комплекса методов обеспечения высокой надежности и эффективности систем вооружения в 1979 году присудили Государственную премию СССР в области науки и техники.

В последующие годы ТЭЦП получила развитие по ряду направлений исследований, которые велись в военно-космической академии имени А. Ф. Можайского, в том числе на кафедре боевой эффективности Военно-космических Сил [11] и в Санкт-Петербургском институте информатики и автоматизации Российской академии наук (СПИИРАН) [19-22]. В настоящее время по развитию ТЭЦП ведутся исследования новых операционных свойств систем [19] и процессов их функционирования, например потенциала систем [20], эффективности использования информационных технологий [21] при функционировании систем.

За рубежом эффективность функционирования систем и другие ОС (effectiveness, efficacy, efficiency, potentiality, capability), ОС использования информационных технологий (ИТ) традиционно исследовались на основе эконометрических и эконометрических методов и моделей, тесно связанных с функционированием рынка, а также путем разработки эвристик, лучших (по операционным свойствам) практик, таких как РМВОК [22], СММИ [23], ВАВОК [24], ВРМ [25], SWeBOK [30], COBIT [31]. Описание некоторых из зарубежных исследований по направлению операционных свойств и сопоставление их с отечественными результатами приведено в [19] и [21] (в части исследования ОС ИТ).

Ряд полученных за рубежом результатов исследований различных видов операционных свойств систем, таких как efficiency, effec-

tiveness, operational capabilities, dynamic capabilities, в том числе и с учетом использования ИТ, приведен в [33-55].

Важно отметить, что в результате различий используемых моделей и методов оценивание ОС в рамках российской школы исследования эффективности реализуется, как правило, количественно, путем оценивания меры соответствия прогнозируемых результатов требованиям на аналитических моделях целенаправленного функционирования, опирающихся на закономерности формирования эффектов функционирования в зависимости от переменных и параметров в решаемых задачах. Такое прогнозирование ОС в случае исследования систем, функционирование которых носит технологический вид (далее СТС), опирается на закономерности формирования эффектов функционирования СТС в результате выполнения элементарных технологических операций с ее элементами. Для элементарных технологических операций технологической документацией заданы закономерности получения их результатов. Как следует из практики, эти закономерности представлены в документации, как дискретно-непрерывные функциональные зависимости (например, таблицы и функции зависимостей результатов элементарных операций от особенностей способа реализации этой операции). Для оценивания основных результатов (эффектов) функционирования СТС следует использовать указанные в документации зависимости результатов элементарных операций и закономерности комплексирования результатов с учетом возможных случайных событий для расчета (в общем случае случайных) эффектов функционирования СТС. Таким образом, характеристики эффектов функционирования СТС получают выполнением ряда математических операций (таких как сложение, умножение, взятие максимума, минимума) с характеристиками результатов элементарных операций. Операции описаны дискретно-непрерывными функциональными зависимостями характеристик их результатов от переменных характеристик с учетом возможных случайных событий. В результате оценивания эффектов функционирования СТС и сравнения их с требуемыми значениями можно оценить ОС, а затем решить задачи исследования ОС как задачи аналитического оценивания, анализа по показателям ОС, а затем как задачи синтеза по показателям этих свойств. Однако из-за сложности разработки моделей формирования эффектов функционирования исследование ОС в зависимости от переменных и параметров в решаемых задачах может быть трудным. Такие трудности возникают, когда в решаемых задачах требуется построение: аналитических моделей функционирования систем по достижению изменяющихся целей; аналитических моделей функционирования совер-

шенствуемых систем и процессов их функционирования; моделей переходных процессов (от достижения одной цели к другой) при функционировании; аналитических моделей использования ИТ при функционировании СТС. Общей особенностью, присущей таким задачам, является необходимость вскрытия закономерности формирования эффектов функционирования СТС из эффектов элементарных операций с учетом случайных изменений среды, целей функционирования, случайных изменений СТС и отношений между элементами при функционировании.

Зарубежными учеными указанные задачи часто решаются эвристически: решения часто получают путем обобщения опыта, экспертного оценивания, на основе обработки уже полученных результатов (апостериори) функционирования эконометрическими и экспертными методами. Тем не менее потребность в использовании аналитических моделей, опирающихся на априорные законы и закономерности формирования эффектов при функционировании СТС, сохраняется. Например, такая необходимость возникает при решении задач использования систем поддержки принятия решений (Decision Management Systems [26]); при управлении комплексными программами [27]; при использовании ИТ бизнес анализа новых поколений [29]; при использовании Agile технологий [30]; при разработке информационных и бизнес-архитектур систем [31, 32]. Использование аналитических моделей позволяет научно обоснованно решать прогностические задачи анализа и синтеза характеристик СТС и их функционирования как математические задачи (например, оптимизационные).

Общей особенностью, присущей такого вида задачам, представляется необходимость учитывать изменения среды, последующие изменения цели функционирования СТС, и в результате — изменения СТС и ее функционирования по причине воздействий среды (далее ИСФ). ИСФ реализуется для перехода к достижению изменившейся цели и для устранения несоответствий функционирования и требований к нему. При ИСФ требуется использование информационных действий для проверки состояний и для предписаний о дальнейших (в том числе информационных) действиях.

Использование аналитического оценивания показателей ОС систем и их функционирования при ИСФ, основанное на вскрытии законов и закономерностей формирования эффектов функционирования (из эффектов элементарных действий) должно позволить устранить указанные недостатки методов оценивания, основанных на эвристиках. Это позволит научно обоснованно решать и автоматизировать решение ряда актуальных практических задач исследования ОС СТС с

учетом ИСФ как математических задач аналитического оценивания, анализа и синтеза по показателям ОС СТС.

Цель моделирования ОС СТС с учетом ИСФ — обеспечить возможность решения актуальных задач практики (решаемых в настоящее время эвристически) на основе аналитического оценивания ОС СТС или с использованием как аналитического оценивания ОС, так и эвристики [14]. Рассмотрим примеры задач, в которых необходимо моделирование ОС СТС с учетом ИСФ.

3. Совершенствование систем и процессов их функционирования на основе исследования операционных свойств. Основные сведения об ОС СТС как о свойствах, характеризующих различные аспекты приспособленности систем и их функционирования к достижению целей (в частности при изменении целей функционирования), были описаны в [20]. Там же были описаны особенности оценивания ОС при ИСФ и показатели оценивания таких ОС. Достижение изменяющихся (действительных и возможных) целей при функционировании означает получать соответствующие изменяющимся требованиям эффекты (важнейшие результаты, к которым предъявлены требования).

В зависимости от состояний системы и среды могут меняться как цели (требования к эффектам), так и достигнутые состояния системы при функционировании. В результате возникает необходимость изменить дальнейшие способы действий для получения требуемых эффектов. Также может возникнуть необходимость изменения функционирования (переходных действий к началу достижения новой цели, действий по устранению недостатков и по изменению СТС).

Особенность введенных ранее концептов и принципов оценивания ОС при ИСФ для решения задач совершенствования систем и процессов их функционирования состоит, во-первых, в том, что такие ОС (например, потенциал системы) [21] учитывают результаты переходных (к достижению изменившейся цели) процессов и результаты функционирования по достижению изменившихся целей.

Во-вторых, введенные ОС [21] позволяют переходить к оцениванию показателей ОС с учетом ИСФ на основе вскрытия аналитических закономерностей формирования эффектов при функционировании СТС с учетом ИСФ. Такое аналитическое оценивание ОС с учетом ИСФ осуществляется на основе имеющихся закономерностей формирования эффектов при реализации действий. С использованием таких закономерностей возможно вскрыть зависимости показателей ОС от: выбираемых характеристик СТС; характеристик действий по изменению СТС; характеристик актуализации целей; характеристик используемых в соответствии с заданными ИТ информационных действий.

Целесообразно оценивать ОС аналитически, чтобы затем на этой основе решать практические задачи исследования ОС как математические задачи оценивания, анализа и синтеза по показателям ОС с учетом ИСФ. К таким практическим задачам (которые решаются сейчас в основном эвристически) относятся: задачи совершенствования комплексных программ и проектов; задачи бизнес анализа нового поколения; задачи использования Agile технологий.

Примерами ОС [19] являются оперативность, результативность, ресурсоемкость, объемлющее их свойство эффективности функционирования системы, потенциал системы, ОС использования информационной технологии.

Исследование ОС с учетом ИСФ включает решение задач оценивания показателей ОС, анализа ОС и синтеза характеристик системы, ее функционирования, по показателям ОС с учетом ИСФ. Результатом решения задачи синтеза по показателям ОС является план функционирования — решение математической задачи планирования. При использовании цифровых ИТ решение получают в виде цифровых моделей (например, цифровой модели плана технологических операций и планов переходных процессов при необходимости).

Из практики известно, что необходимость исследовать ОС при совершенствовании СТС и процессов их функционирования появляется регулярно, в связи с актуализацией ИСФ.

При этом актуализируются задачи совершенствования СТС и их функционирования разного вида. К показателям ОС с учетом ИСФ, использующимся при решении таких задач, предъявляются следующие требования:

- показатели должны позволять оценивать ОС при ИСФ и совершенствовании СТС и процессов их функционирования на основе прогнозирования того, насколько эффекты будут соответствовать требованиям;

- показатели должны учитывать возможности выполнения технологических операций функционирования СТС при ИСФ (устранения недостатков, модернизации, совершенствования СТС и их функционирования, переходных технологических операций), в том числе технологических информационных операций при реакции на ИСФ;

- показатели должны учитывать случайности при ИСФ. То, какие ИСФ произойдут и какие действия при функционировании будут затем реализованы, зависит от случайностей разной природы, от решений людей выполнить различные действия и от результатов этих действий. Сначала требуется выполнить информационные действия по проверке состояния СТС и среды, и затем, в зависимости от этих действий, может потребоваться решать задачи (задача — вопрос, на кото-

рый надо дать ответ) разного вида по поводу того, следует ли реализовать действия по изменению СТС и ее функционирования, какие и как.

Укажем ряд особенностей функционирования СТС и связанные с ними особенности задач исследования ОС СТС при ИСФ.

Первая особенность. Проявляется при проверке состояния СТС и ее среды в случае необходимости ответа на вопрос: нужно ли, исходя из полученных (проверенных) состояний СТС и ее среды, изменить состав и способы действий с элементами СТС. Такое изменение может быть вызвано, например, зафиксированным в результате проверки состояния СТС и ее среды изменением предоставляемых ресурсов, отказами, выявлением брака, сменой цели, занятостью рабочих мест. Необходимость изменения действий — случайное событие. Если необходимость изменения действий наступила, может потребоваться изменить: состав СТС, характеристики элементов СТС, действия с элементами СТС, способы действий, характеристики действий, их последовательности, характеристики последовательностей действий. Затем может возникнуть вопрос: следует ли реализовать переходные действия к состоянию, из которого допустимо начинать действия для достижения актуализированной «новой» цели.

Вторая особенность. Проявляется в случае, если в результате ИСФ событие, состоящее в необходимости изменения действий, наступило и затем оказалось необходимым изменить СТС и последующие действия или выполнить переходные действия по достижению изменившейся цели. При этом действия необходимо реализовать полученными на этапе планирования составом и способами мероприятий. Это означает, что проверенное состояние предугадано на этапе планирования и ему было сопоставлено решение в виде состава и способов действий. Такие действия предполагаются запланированными заранее в зависимости от возможных состояний СТС и среды, одно из которых и было выявлено путем проверки. При этом необходимо найти (например, в базе данных) полученное на этапе планирования решение, соответствующее результатам проверки.

Третья особенность. Проявляется, если проверенное состояние не может быть отнесено к множеству, для элементов которого решения о дальнейших действиях были получены на этапе планирования. При этом необходимо найти неизвестный до начала функционирования, лучший состав СТС, состав и (или) способы действий из множества допустимых.

При проявлении каждой из этих особенностей или их комбинации необходимо использовать какую-либо (одну или несколько) ИТ для проверки состояний СТС и ее среды и затем для принятия решений о дальнейших действиях. Использование ИТ при проявлении ука-

занных особенностей часто связано с действиями людей по формированию целей, по прогнозированию, по совершенствованию функционирования и, наконец, по реализации решений на практике.

Исследование ОС СТС с учетом ИСФ составляет основу совершенствования СТС и их функционирования. Однако такое исследование может быть выполнено как на основе использования математических методов и моделей, так и эвристически, путем систематизации опыта — как в некоторых из рассмотренных задач совершенствования систем и функционирования. Решение задач оценивания, анализа и синтеза СТС по рассчитываемым на основе аналитических закономерностей функционирования и совершенствования функционирования показателей ОС позволяет перейти к постановке и решению практических задач совершенствования СТС и их функционирования как математических задач.

4. Исследование операционных свойств использования информационных технологий. Концепты и принципы исследования ОС использования ИТ были рассмотрены в работе [21]. Важно отметить, что польза, извлекаемая из ИТ, — это возможность выбрать лучшие информационное и неинформационное действия (из множества возможных или из множества допустимых действий соответствующего вида), чтобы затем их реализовать. При функционировании СТС такая возможность реализуется за счет информационных действий. Возможности получения лучших результатов функционирования СТС с учетом ИСФ, с одной стороны, и необходимость выбора действий и их характеристик для получения лучших результатов, с другой стороны, ведет к необходимости использования ИТ при функционировании СТС в условиях ИСФ. Сравнить использование разных ИТ при функционировании таких СТС целесообразно по показателям ОС СТС в условиях использования одной из ИТ при ИСФ.

Использование ИТ при функционировании СТС в условиях ИСФ состоит в проверке достигнутого (действительного) состояния и затем прогнозирования соответствия будущих состояний СТС актуальным целям. Такие проверки и прогнозирование происходят при реализации информационного действия. Поэтому ИТ используется сначала для оценивания состояний СТС и их соответствий целям, затем при необходимости информационные действия используются для изменения последующих действий (например, для начала реализации действий по переходу к достижению новой цели). Использование информационных действий одной (первой) ИТ, например, может привести к решению о продолжении функционирования СТС по имеющемуся плану, а второй — к решению о переходных действиях. Тем самым с использованием информационных действий второй ИТ будут реализованы переходные

действия, а затем — изменится функционирование СТС. В результате при использовании второй ИТ получают состояния, отличные от состояний при использовании первой ИТ. При этом информационные действия так чередуются с действиями неинформационными, что следующие за ними неинформационные действия, полученные в результате информационного действия, функционально зависят от него.

ИСФ могут затрагивать состав, связи и характеристики: элементов системы; действий при функционировании системы; актуализаций действий. В зависимости от того, какие ИТ будут использованы при реализации информационных действий, будут реализованы разные зависимые от них действия (разными способами) и получены разные эффекты функционирования при ИСФ. По тому, насколько такие эффекты отличаются при использовании разных информационных действий (разных ИТ) при ИСФ и насколько при таком использовании будут достигаться цели функционирования, можно судить об ОС использования ИТ.

Развернем последовательности зависимых информационных и неинформационных действий (с зависимыми способами реализации) в дерево последовательностей действий разными способами. В зависимости от используемых информационных действий в соответствии с какой-либо ИТ будут получены цепочки зависимых от них разных (с разными характеристиками) способов действий и разные конечные результаты. Использованию одной ИТ соответствует поддерево, в котором все информационные действия выполняются по этой ИТ.

Общим в двух рассмотренных примерах исследования ОС при ИСФ представляется постановка задачи исследования, а именно дано: характеристики эффектов для каждого способа действий, их зависимости от переменных, зависимости неинформационных действий от информационных, характеристики возможных ИСФ. Найти: значения показателей ОС с учетом возможных ИСФ. Затем, используя полученные результаты, решить задачи анализа и синтеза по этим показателям.

ОС следует оценивать таким методом, чтобы:

– различные ОС объектов (систем, их функционирования) при разных ИСФ, при достижении разных целей оценивались бы по подобным схемам;

– метод и схемы оценивания были бы подобными для СТС с разными особенностями: ИСФ; эффектов функционирования СТС; используемых технологий; переменных в решаемых задачах; достигаемых СТС целей;

– получаемые оценки показателей ОС с учетом ИСФ и зависимости показателей ОС от переменных имели бы явную практическую

интерпретацию, позволяющую научно обоснованно выбирать лучшие для исследователя объекты оценивания.

Такой метод оценивания был разработан на основе использования *схем оценивания ОС* — устойчивых (к изменениям видов реализуемых действий, к их особенностям, к условиям реализации действий и к достигаемым целям) математических выражений, задающих меру соответствия характеристик прогнозируемых эффектов изменяющимся требованиям к ним.

5. Метод аналитического оценивания операционных свойств на основе схем их оценивания. Метод оценивания ОС основан на последовательном введении подобных схем оценивания ОС разных видов. Такая последовательность отличается тем, что она позволяет удовлетворить указанные выше требования к методу аналитического оценивания в задачах оценивания ОС. В последовательности каждая последующая схема использует предыдущую. В результате последовательность приспособлена к продолжению порождения схем, используемых для решения перспективных задач исследования ОС.

Опишем последовательность из трех схем оценивания ОС, составляющих основу метода оценивания ОС.

Первая из таких схем — *схема оценивания эффективности функционирования*, разработанная на основе концепции оценивания эффективности целенаправленных процессов [6, 7].

При оценивании в соответствии с этой схемой предполагается, что в процессе реализации действий решения о совершенствовании СТС не влияют на функционирование, действия по достижению цели не прерываются, цель не меняется.

Вторая схема обобщает первую путем учета достижения изменяющихся целей. При использовании этой схемы предполагается, что возможные изменения целей функционирования СТС и соответствующие им переходные процессы к функционированию по достижению изменившейся цели заданы заранее, соответствующие решения уже разработаны, неизвестен лишь момент прерывания функционирования для перехода к одному из заданных новых функционирований в связи с изменившейся целью.

Наконец, *третья схема* обобщает две первые для учета возможных целенаправленных действий в результате ИСФ (совершенствования, переходных процессов), которые выбирают и реализуют с использованием информационных действий при функционировании, в зависимости от состояний СТС и ее среды при функционировании. Состояния СТС и ее среды проверяют в заданные моменты времени, число которых конечно.

Выполним последовательную формализацию указанных схем.

1. *Схема оценивания эффективности.* Схема описывает оценивание величины I_p — значения предиката p , задаваемого следующим образом:

$$p: \hat{Y}_{\langle 3 \rangle} \times R_{\langle 3 \rangle}^0 \times Y_{\langle 3 \rangle}^0 \rightarrow [0, 1], \text{ где } p \text{ — неопределенный предикат;}$$

$\hat{Y}_{\langle 3 \rangle}$ — это множество векторов случайных значений характеристик эффектов: $\hat{Y}_{\langle 3 \rangle} = \{\hat{y}_{\langle 3 \rangle_i} : i = \overline{1, n}\}$, где i — номер вектора;

Здесь случайные величины и события задаются с использованием меры, удовлетворяющей аксиомам вероятностной меры. Предполагается, что такая мера также может иметь смысл меры возможности.

Вектора значений характеристик описаны для примера в традиционном трехкомпонентном [5] виде характеристик результативности (1), оперативности (2) и ресурсоемкости (3). В общем случае эти векторы могут иметь любую конечную размерность.

$Y_{\langle 3 \rangle}^0$ — это множество векторов требуемых (директивных, заданных) значений характеристик эффектов:

$$Y_{\langle 3 \rangle}^0 = \{y_{\langle 3 \rangle_k}^0 : k = \overline{1, K}\} \text{ где } k \text{ — номер вектора;}$$

$R_{\langle 3 \rangle}^0$ — это множество векторов требуемых отношений между случайными значениями характеристик эффектов и их требуемыми значениями: $R_{\langle 3 \rangle}^0 = \{r_{\langle 3 \rangle_j}^0 : j = \overline{1, M}\}$, j — номер отношения.

Таким образом, $I_p = p(\hat{y}_{\langle 3 \rangle_i}, r_{\langle 3 \rangle_j}^0, y_{\langle 3 \rangle_k}^0)$ — это значение вероятности того, что указанный в скобках предикат примет значение «истина» или это значение того, что соответствующее ему (предикату) случайное событие наступит.

2. *Схема оценивания потенциала I_{oc} СТС.* Первая рассмотренная схема оценивания I_p — вид (второй) схемы I_{oc} — значения отображения p_{oc} :

$$p_{oc} : \hat{Y}_{\langle 3 \rangle}(t) \times R_{\langle 3 \rangle}^0 \times \hat{Y}_{\langle 3 \rangle}(t), t_0, T \rightarrow [0, 1], \text{ в котором:}$$

$\hat{Y}_{\langle 3 \rangle}(t)$ — случайный процесс, моделирующий изменения эффектов $\hat{y}_{\langle 3 \rangle_i}(t)$ функционирования СТС во времени t .

$\hat{Y}_{\langle 3 \rangle}^0(t)$ — случайный процесс, моделирующий возможные случайные изменения $Y_{\langle 3 \rangle}^0$ во времени (что соответствует возможным из-

менениям цели функционирования системы в результате изменений среды СТС):

$$\hat{y}_{<3>k}^0 = \hat{y}_{<3>k}^0(t_0), t_0 \text{ — момент начала функционирования,}$$

T — плановая длительность функционирования системы.

При изменении требований (переходе от достижения одной цели к другой) предполагается, что СТС переходит к достижению новой цели без существенных затрат ресурсов на переходные действия u .

Таким образом, $I_{oc} = P_{oc}(\hat{y}_{<3>i}(t;u), r_{<3>j}^0, y_{<3>k}^0(t)), t \in [t_0, t_0 + T]$ —

это значение вероятностной меры того, что прогнозируемые (виртуальные) значения эффектов функционирования заданной СТС с учетом изменяющихся целей будут соответствовать требуемым значениям эффектов на плановом отрезке времени функционирования СТС.

Если $t = t_0$ — константа, то изменения целей не реализуются и поэтому $I_p = I_{oc}$.

3. *Схема оценивания операционных свойств I_{ocn} СТС, совершенствуемых при ИСФ.* В этом случае при реализации функционирования в условиях ИСФ в зависимости от изменяющихся состояний s СТС и ее среды, проверяемых с помощью тех или иных информационных действий (в соответствии с заданной ИТ) в заданные моменты времени, число которых конечно, инициируются и реализуются действия $u_z(s)$ по совершенствованию СТС и ее функционирования и действия переходных процессов. Такие действия инициируются и уточняются (выбираются) в результате информационных действий (в соответствии с той или иной ИТ) в связи с изменениями состояний СТС (зафиксированными в результате проверки состояний) или (и) для достижения изменившейся цели с учетом изменившихся состояний s среды СТС, проверенными в заданные моменты времени при достижении целей. Действия переходных процессов дают два вида целевых эффектов: (1) информация о состояниях и инициированных (выбранных) действиях «целевого» процесса и по его изменению — при реализации информационных действий и (2) выполнения переходных действий к измененному функционированию — при реализации неинформационных действий. При этом на реализацию таких переходных действий тратятся ресурсы. Функция $u_z(s)$ реализуется с использованием ИТ z и при изменении аргумента s (состояния СТС и среды) дает значение, описывающее состав и характеристики действий.

При этом рассмотренная выше (вторая) схема оценивания потенциала I_{oc} — вид (третьей) схемы оценивания $I_{ocn} = p_{ocn}(t, u_z(s))$ — значения отображения p_{ocn} :

$p_{ocn} : \hat{Y}_{<3>}(u_z(s), t) \times R_{<3>}^0 \times \hat{Y}_{<3>}^0(t), t_0, T \rightarrow [0, 1]$, в котором $u_z(s)$ — характеристики переходных действий в зависимости от состояния $s \in S$ СТС и ее среды при функционировании, проверяемые с использованием заданной ИТ Z .

S — совокупность возможных состояний СТС и ее среды при функционировании.

На основе S задаются деревья исходов при функционировании СТС и ее среды $T_z \subseteq S \times S$, с вершинами которых ассоциированы возможные состояния s СТС и ее среды при функционировании с использованием заданной ИТ z , а с дугами — переходы между этими состояниями в результате действий, в том числе переходных $u_z(s)$.

$\hat{Y}_{<3>}(u_z(s), t)$ — характеристики случайного процесса получения эффектов в результате реализации последовательностей действий целевого функционирования и переходных процессов от одного функционирования к другому при заданных возможных состояниях (характеризуемых деревом T_z , введенным выше). Таким образом,

$I_{ocn} = p_{ocn}(\hat{Y}_{<3>}(u_z(s), t; z), r_{<3>}^0, y_{<3>}^0(t)), t \in [t_0, t_0 + T]$ — значение вероятности того, что прогнозируемые значения эффектов функционирования СТС в заданные моменты времени, с учетом реализации переходных процессов в зависимости от состояний СТС и ее среды при функционировании будут требуемым образом соответствовать (изменяющимся) требуемым значениям эффектов.

6. Примеры использования схем оценивания операционных свойств. Рассмотрим три примера использования схем оценивания ОС — так, чтобы полученные в одном примере результаты использовались бы в последующих примерах. При этом первый пример соответствует оцениванию ОС функционирования СТС для достижения одной заданной цели, которая не меняется, как не меняется и СТС или технологические операции (ГлОп) ее функционирования. А именно исход функционирования среды на границе СТС один и имеет вид одного требуемого состояния к моменту окончания функционирования. Планируемый исход функционирования СТС тоже один. Функционирования планируется и оценивается к своему окончанию в соответствии с календарным планом функционирования для достижения заданной цели. При этом возможные воздействия среды на элементы при функционировании среды тоже заданы и не ме-

няются, от этих воздействий зависят результаты (эффекты) функционирования, но не состав ТлОп. Тем самым рассматривается одно функционирование по заданному календарному плану. ОС такого функционирования (эффективность) и исследуется. Показатель эффективности такого функционирования рассчитывается, как вероятность случайного события, состоящего в соответствии прогнозируемых (для заданного календарного плана функционирования) эффектов к окончанию функционирования, с одной стороны, требованиям к эффектам по окончании функционирования — с другой. Функционирование среды задано законами распределения характеристик результатов ТлОп и характеристиками требований к эффектам при достижении заданной цели.

Рассмотрим второй пример, соответствующий оцениванию ОС СТС, которая может функционировать в зависимости от исходов функционирования среды на границе СТС (которых может быть несколько) для достижения разных целей в разных последовательностях их достижения. При этом каждый действительный исход функционирования среды и СТС (последовательность состояний и событий на границе с СТС) построен для одной последовательности частично достижимых целей (на границе среды и СТС). Одной такой последовательности на границе среды с СТС сопоставлено несколько возможных исходов функционирования, каждый из которых проявляется при каких-либо событиях, вызванных средой на границе с СТС и (или) на границе с элементами СТС.

Каждому исходу функционирования соответствует заданное календарным планом функционирование по достижению нескольких смежных друг друга целей. Тем самым исходу функционирования сопоставлена последовательность из «целевых» функционирований и «переходных» функционирований с учетом событий, вызванных средой на границе с СТС. Исход функционирования оценивается аналогично тому, как и в первом примере — с использованием эффективности функционирования, но по достижению заданной последовательности целей (не только одной, как в первом примере), и так же, как и в первом примере, к заданному моменту окончания функционирования СТС. Множеству же исходов функционирования сопоставлен показатель потенциала СТС.

Наконец, третий пример соответствует оцениванию ОС СТС в том случае, когда исходу функционирования сопоставлен вектор показателей, каждый из которых характеризует приспособленность функционирования давать требуемые результаты к заданным моментам времени проверки состояний СТС и ее среды.

Элементы таких векторов рассчитываются аналогично первому примеру как меры соответствия прогнозируемых эффектов (при из-

вестном исходе функционирования и к заданному моменту) требованиям к ним (в тех же условиях).

При этом:

$$W_i(C_i, C_u^0) = \prod_{j=1, J} P(\tilde{Y}_{j,i}(C_i) \leq Y_{j,i}^0(C_i, C_u^0)), \text{ где:}$$

P — вероятность наступления случайного события;

$\tilde{Y}_{j,i}(C_i)$ — значение j -го эффекта функционирования к окончанию C_i . Например, затраты времени персонала на реализацию комплекса работ, затраты электроэнергии.

$\tilde{Y}_{j,i}^0(C_i, C_u^0)$ — директивные (требуемые) значения эффектов функционирования при исходе функционирования СТС C_i и исходе функционирования среды C_u^0 ;

$W_i(C_i, C_u^0)$ — вероятность удовлетворения требований при исходе функционирования СТС C_i и единственном заданном исходе функционирования среды — C_u^0 .

Обозначим C_i — i -й исход функционирования СТС. Под ним будем понимать возможную последовательность реализаций мероприятий функционирования СТС, он ассоциирован с ветвью дерева исходов;

$T_c(C_u^0, h)$ — дерево возможных исходов, ветви которого $\{C_i, i = \overline{1, I}\}$ — исходы функционирования СТС и ее среды. Дерево построено для заданной последовательности C_u^0 директивных состояний, сформированных на границе СТС и ее среды (исхода функционирования среды на границе с СТС) и для заданной ИТ h из множества H возможных ИТ.

Таким образом, известно отображение $C_u^0 \rightarrow T_c(C_u^0, h)$ для любой C_u^0 и, следовательно, отображение $T_i^{Cp}(h) \rightarrow \{T_c(C_u^0, h), u = \overline{1, U}\}$.

Здесь $T_i^{Cp}(h) = \{C_u^0(h), u = \overline{1, U}\}$ — дерево сценариев функционирования на границе среды и системы, состоящее из U веток дерева, каждая из которых ассоциирована своими C_u^0 .

Пусть исходы C_u^0 функционирования среды на границе с СТС предполагаются заданными возможными воздействиями среды на элементы ГлОп и требуемыми по завершении C_i эффектами.

Деревья исходов функционирования среды на границе СТС и исходов функционирования СТС строятся на этапе планирования в зависимости от: возможных ТИО, используемой h -й ИТ и состояний на границе среды; связей ТИО и ТНИО, в соответствии с используемой технологией; изменений состояний СТС.

В результате реализации каждого исхода функционирования среды C_u^o на границе с СТС и каждого исхода функционирования СТС C_i с использованием h -й ИТ складывается разное соответствие (на границе СТС и ее среды). Те ИТ, которые обеспечивают лучшее (из возможных) соответствие на границе СТС и среды и следует считать предпочтительными. Пусть: $p_i(C_i, C_u^o) = \prod_{k: e_{i,k} \in C_i(C_u^o)} p_{i,k}(C_i)$ — вероятность p_i

реализации заданного при h -й ИТ исхода C_i функционирования СТС при условии, что исход на границе среды и СТС C_u^o тоже задан h -й ИТ.

В примере предполагается, что эта вероятность может быть рассчитана как произведение вероятностей $p_{i,k}(C_i)$ реализации исходов мероприятий, входящих в C_i при C_u^o и заданной h -й ИТ, то есть исходы мероприятий независимы.

Оценивать результаты комплекса ТлОп с использованием заданной h -й ИТ следует по всем возможным исходам функционирования СТС и всем возможным исходам функционирования на границе СТС и среды. Этому требованию, в частности, удовлетворяет скалярный показатель потенциала СТС — функция потенциала [1] при использовании h -й ИТ:

$$\psi(T_l^{Cp}(h)) = \sum_{C_u^o \in T_l^{Cp}(h)} \psi_u(T_c(C_u^o, h)) \cdot p_u(C_u^o, h);$$

$$\psi_u(T_c(C_u^o, h)) = \sum_{C_i \in T_c(C_u^o, h)} (W_i(C_i, C_u^o) \cdot p_i(C_i, C_u^o)).$$

Пусть теперь имеется две ИТ, h («новая») и f («старая»). Для них построены c -е и p -е деревья исходов $T_c(C_u^o, h)$ и $T_p(C_u^o, f)$ для каждого исхода C_u^o функционирования среды на границе, ассоциированного с деревьями $T_l^{Cp}(h)$ и $T_m^{Cp}(f)$.

Определим разность значений функции потенциала СТС при реализации новой и старой ИТ:

$$\Delta\psi(h, f) = \psi(T_l^{Cp}(h)) - \psi(T_m^{Cp}(f)).$$

Эта разность характеризует ОС использования «новой» ИТ по сравнению со «старой». Она может быть использована как *показатель результативности внедрения новой ИТ. Целевой эффект (результат) внедрения* — лучшая приспособленность $\psi(T_l^{Cp}(h))$ к достижению целей в условиях изменяющихся воздействий со стороны среды.

В рассматриваемом случае приспособленность измеряется скалярном (рассчитываемым как вероятностная смесь). В общем случае векторный показатель потенциала может быть представлен как векторная случайная величина $\tilde{\Psi}(T_l^{Cp}(h))$, математическое ожидание которой и задает скалярный показатель $\psi(T_l^{Cp}(h))$.

Тогда, если задано директивное значение $\Psi^{\partial}(T_l^{Cp}(h))$, то показатель результативности внедрения h -й ИТ может быть задан как:

$$W^{Pez}(h) = P(\tilde{\Psi}(T_l^{Cp}(h)) > \Psi^{\partial}(T_l^{Cp}(h))).$$

Показатель результативности должен быть сопоставлен с *обеспечивающими* (получение целевого эффекта внедрения ИТ) *эффектами*. Это целесообразно сделать с использованием показателя *эффективности внедрения* новой ИТ.

Пусть в показателе эффективности внедрения новой ИТ целевой эффект описывается скалярным показателем результативности h -й ИТ по сравнению с f -й ИТ — $\Delta\psi(h, f)$. Для последующего расчета показателя эффективности внедрения h -й ИТ следует рассчитать еще и показатели оперативности и ресурсоемкости ее внедрения.

Рассмотрим пример обеспечивающих эффектов, необходимых для внедрения h -й ИТ на смену f -й ИТ. Для упрощения примера ограничимся двумя обеспечивающими эффектами (по одному для расчета показателя оперативности и ресурсоемкости):

– случайным моментом времени $\tilde{T}(h)$, к которому может быть внедрена h -я ИТ;

– случайной величиной стоимости $\tilde{C}(h)$ внедрения h -й ИТ.

Требования к этим результатам будем для простоты считать детерминированными константами T^o и C^o соответственно, а отношение соответствия примем имеющим вид «обеспечивающий эффект не больше требования». Тогда эффективность внедрения h -й ИТ может быть оценена с использованием следующего показателя [1]:

$$W(h, f) \equiv \Delta\psi(h, f) \cdot Poss\left(\left(\tilde{C}(h) \leq C^o\right) \cap \left(\tilde{T}(h) \leq T^o\right)\right).$$

Здесь $Poss\left(\tilde{T}(h) \leq T^o\right)$ — показатель оперативности внедрения в практику h -й ИТ; $Poss\left(\tilde{C}(h) \leq C^o\right)$ — показатель ресурсоемкости внедрения этой ИТ на смену f -й ИТ.

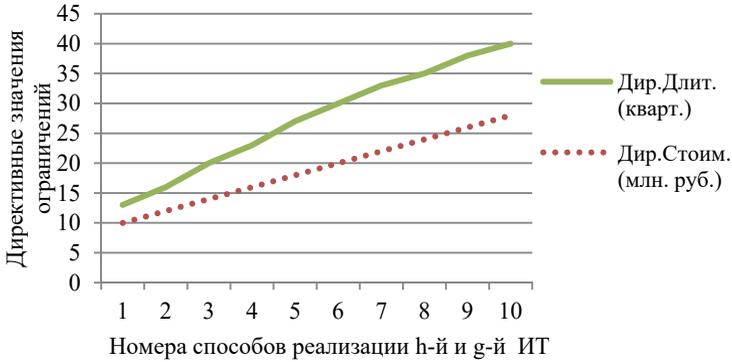
Рассмотрим примеры расчетных значений показателей ОС внедрения h -й и g -й ИТ при разных ограничениях на внедрение ИТ для случая, когда показатель результативности внедряемой ИТ — скаляр и требования к результатам внедрения детерминированы. В примере рассмотрены 10 способов выполнения двух ИТ (g -й и h -й). Каждый из 10 способов характеризуется одними и теми же ограничениями для g -й и h -й ИТ (рисунок 1а) соответственно.

Расчетные значения показателей оперативности, ресурсоемкости, результативности и комплексный показатель эффективности внедрения ИТ для 10-ти способов внедрения h -й ИТ показаны на рисунке 1б, для g -й ИТ — на рисунке 1с. Показатели эффективности внедрения для всех 10 способов h -й и g -й ИТ показаны на рисунке 1д. Как видно из полученных результатов, для разных ограничений по ресурсам лучшей эффективностью внедрения могут характеризоваться разные внедряемые ИТ.

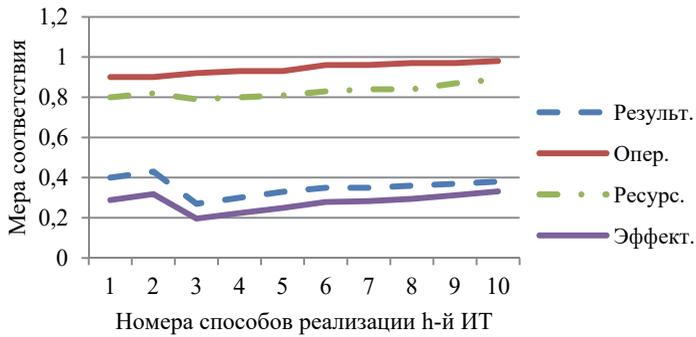
Для случая векторного показателя результативности внедрения h -й ИТ показатель эффективности внедрения h -й ИТ примет вид:

$$W(h, f) \equiv Poss\left(\tilde{\Psi}(h) > \Psi^o\right) \cap \left(\tilde{C}(h) \leq C^o\right) \cap \left(\tilde{T}(h) \leq T^o\right).$$

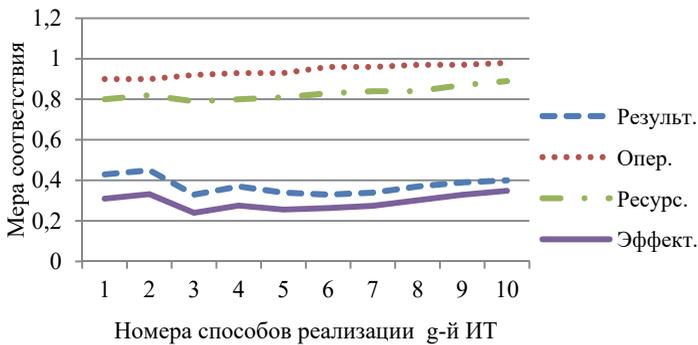
Возможны разные виды особенностей предъявления требований к результатам функционирования в зависимости от особенностей исходов функционирования среды и соответствующих им исходов функционирования СТС C_i .



a)

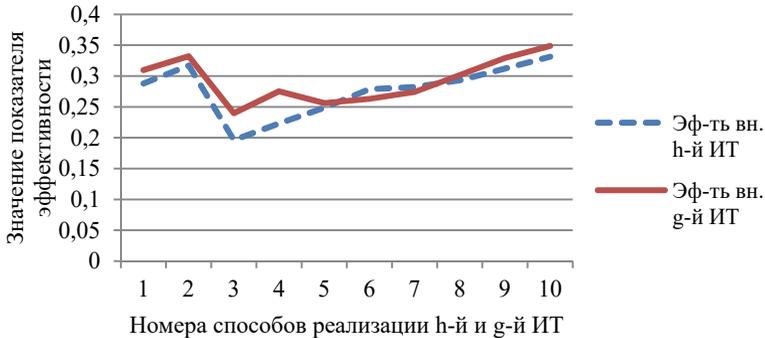


b)



c)

Рис. 1. Примеры расчетных значений показателей ОС внедрения h-й и g-й ИТ при разных ограничениях (а, b, с)



d)

Рис. 1. Примеры расчетных значений показателей ОС внедрения h-й и g-й ИТ при разных ограничениях (d)

Рассмотрим сначала вид предъявления требований к результатам, когда требования предъявляются единовременно, по завершении функционирования. В этом случае достижение цели оценивается как соответствие требуемых средой состояний СТС прогнозируемому состоянию СТС *по завершении функционирования СТС*.

В этом случае мера соответствия рассчитывается как *вероятность попадания случайного вектора эффектов (к окончанию функционирования) в требуемую область*.

Пусть теперь следует учитывать *частичные достижения цели* — такие состояния, которые требуются для дальнейшего достижения цели. Такая необходимость, например, возникает в случае исследования изменяющихся целей.

Частичное достижение цели описывают как соответствие состояний СТС и среды в такие моменты времени, когда реализуется проверка состояния — информационная операция на границе СТС и среды. Оценивание соответствия в текущий момент времени — оценка действительного соответствия. Прогнозирование оценки соответствия в будущие моменты времени — оценка возможного соответствия. Последовательность таких соответствий — реализация соответствия на границе СТС и среды. Каждому возможному исходу функционирования на границе СТС и среды и исходу C_i функционирования СТС сопоставлена одна последовательность мер соответствия. Вероятностную меру реализации такой последовательности можно использовать, как меру удовлетворения требований. В результате можно корректно оценить показатели потенциала для случая изменяющихся целей.

7. Заключение. На основе краткого анализа особенностей исследования эффективности целенаправленных процессов функционирования систем и ряда других ОС отечественными и зарубежными авторами сделан вывод об актуальности решения современных задач исследования ОС с учетом изменений: среды, последующих изменений цели функционирования СТС, изменений СТС и ее функционирования в результате воздействий среды, а затем — изменений функционирования для достижения изменившейся цели и устранения несоответствий функционирования требованиям. Показано, что такое исследование следует выполнять на основе оценивания показателей ОС за счет вскрытия аналитических закономерностей проявления эффектов в условиях указанных изменений. Приведены особенности актуальных задач исследования ОС: задач исследования ОС совершенствуемых СТС и задач исследования ОС использования ИТ. Предложена последовательность схем оценивания ОС, составляющая основу метода аналитического оценивания ОС. Рассмотрен пример расчета показателей операционных свойств использования информационных технологий.

Полученные результаты должны позволить перейти к решению задач исследования ОС на аналитических математических моделях в соответствии с предложенными схемами оценивания ОС.

Литература

1. *Колмогоров А.Н.* Число попаданий при нескольких выстрелах и общие принципы оценки эффективности системы стрельбы // Сборник статей по теории стрельбы. Тр. Матем. ин-та им. В. А. Стеклова. М.-Л. 1945. № 12. С. 7–25.
2. *Вентцель Е.С., Лихтерев Я.М., Мильграм Ю.Г., Худяков И.В.* Основы теории боевой эффективности и исследования операций // М.: ВВИА имени Н.Е. Жуковского. 1961. 524 с.
3. *Аверкина Г., Эпштейн Г.* Писатель И. Грекова — профессор Е. С. Вентцель // Новый Мир. 2008. № 4.
4. *Иоффе А.Я. и др.* Вероятностные методы в прикладной кибернетике // Л. 1976. 424 с.
5. *Морозов Л.М., Петухов Г.Б., Сидоров В.Н.* Методологические основы теории эффективности // Л. 1979. 174 с.
6. *Петухов Г.Б.* Основы теории эффективности целенаправленных процессов. Часть 1. Методология, методы, модели // 1989. 647 с.
7. *Петухов Г.Б., Якунин В.И.* Методологические основы внешнего проектирования целенаправленных процессов и целеустремленных систем // М.: АСТ. 2006. 504 с.
8. *Авдуевский В.С.* Надежность и эффективность в технике: справочник в 10 т. // М.: Машиностроение. 1986. Т. 3. 328 с.
9. *Дубровский В., Ковбич М.* Дела и дни академика Уткина, конструктора «Сатань». Человек, создавший наивысшее техническое достижение XX века // Независимая газета. 2013. URL: http://www.ng.ru/science/2013-10-23/10_utkin.html (дата обращения 18.04.2018).
10. *Крочков Ю.И., Кузнецов В.И.* Исследования эффективности вооружения // М.: ВА РВСН им. Ф.Э. Дзержинского. 1980.
11. *Волков Е.Б. и др.* Технические основы эффективности ракетных систем // М.: Машиностроение. 1989.

12. *Ролдугин В.Д.* Моделирование и оценка эффективности боевых действий РВСН. // М.: ВА РВСН имени Ф.Э. Дзержинского. 2005.
13. *Крючков Ю.В., Кузнецов В.И., Охотников Г.Н.* Моделирование и оценка эффективности боевых действий и вооружения ракетных войск стратегического назначения // М.: ВА РВСН имени Ф.Э. Дзержинского. 1986. 627 с.
14. *Лысенко И.В.* Анализ и синтез сложных технических систем. Часть 1. Анализ и синтез систем обеспечения готовности ракет-носителей и космических аппаратов к запуску (основы теории) // М.: Воениздат. 1995. 368 с.
15. *Викулов С.Ф.* Военно-экономический анализ: история, методология, проблемы // Вооружение и экономика. 2012. № 4(20). С. 86–98.
16. *Груцанский В.А., Ильичев А.В., Нгуен К.Т.* Формализация показателей эффективности и безопасности комплексных программ в условиях неопределенности // Нелинейный мир. 2014. Т. 12. № 7. С. 33–37.
17. *Podinovskiy V.* Non-dominance and potential optimality for partial preference relations // European Journal of Operational Research. 2013. vol. 229. pp. 482–486.
18. *Ryabinin I.A.* Logical probabilistic analysis and its history // International journal of risk assessment & management. Proquest ABI/INFORM. 2015. vol. 18. no. 3-4. pp. 256–265.
19. *Гейда А.С., Исмаилова З.Ф., Клитный И.В., Лысенко И.В.* Задачи исследования операционных и обменных свойств систем // Труды СПИИРАН. 2014. Вып. 4(35). С. 136–160.
20. *Гейда А.С., Лысенко И.В.* Задачи исследования потенциала социально-экономических систем // Труды СПИИРАН. 2009. № 10. С. 63–84.
21. *Гейда А.С., Лысенко И.В., Юсупов Р.М.* Основные концепты и принципы исследования операционных свойств использования информационных технологий // Труды СПИИРАН. 2015. Вып. 5(42). С. 5–36.
22. *Юсупов Р.М., Мусаев А.А.* К оцениванию эффективности информационных систем: методологические аспекты // Информационные технологии. 2017. Т. 23. № 5. С. 323–332.
23. Руководство к своду знаний по управлению проектами – PMBOK 5 (финальная полная русская версия) // М. Олимп-Бизнес. 2014. 590 с.
24. *McMahon P.* 15 Fundamentals for Higher Performance in Software Development: Includes discussions on CMMI, Lean Six Sigma, Agile and SEMAT's Essence Framework // Pem Systems Publ. 2014. 336 p.
25. ИБА. A Guide to Business Analysis Body of Knowledge (BABOK 2.0) // 2015. 502 p.
26. *Шадунова М.* Свод знаний по управлению бизнес-процессами: BPM СВОК 3.0/BPM СВОК Version 3.0: Guide to the Business Process Management Common Body of Knowledge // М. Альпина Паблишер. 2016. 640 с.
27. *Taylor J., Purchase J.* Real-World Decision Modeling with DMN // Meghan-Kiffer Press. 2016. 437 p.
28. *Kendrick T.* How to Manage Complex Programs // AMACOM. 2016. 336 p.
29. *Dinsmore T.W.* Disruptive Analytics: Charting Your Strategy for Next-Generation Business Analytics // Apress. 2016. 276 p.
30. The Agile Extension to the BABOK® Guide. International Institute of Business Analysis (ИБА). 2015. 128 p.
31. Guide to the Software Engineering Body of Knowledge (SWEBOK). Version 3.0 // IEEE Computer society. 2004. 335 p.
32. *Blokdyk G.* COBIT 5: The Ultimate Step-By-Step Guide // 5STARCOoks. 2018. 124 p.
33. *Zidane Y., Olsson N.* Defining project efficiency, effectiveness and efficacy // International Journal of Managing Projects in Business. 2017. vol. 10. Issue 3. pp. 621–641.
34. *Carvalho M., Takao S., Silva D.* Efficiency and Effectiveness Analysis of Public Transport of Brazilian Cities // Journal of Transport Literature. 2015. vol. 9(3). pp. 40–44.

35. *Mesterova J., Prochazka J., Vaculik J., Smutny P.* Relationship between Self-Efficacy, Transformational Leadership and Leader Effectiveness // *Journal of Advanced Management Science*. 2015. vol. 3. no. 2. pp. 109–122.
36. *Nordon C. et al.* The «Efficacy-Effectiveness Gap»: Historical Background and Current Conceptualization // *Value in Health*. 2016. vol. 19. Issue 1. pp. 75–81.
37. *Docekalova M., Kocmanova A.* Evaluation of the Effectiveness of Manufacturing Companies by Financial and Non-financial Indicators // *Procedia – Social and Behavioral Sciences*. 2015. vol. 213(1). pp. 491–496.
38. *Hwangbo H., Johnson A.L., Ding Y.* A Production Economics Analysis for Quantifying the Efficiency of Wind Turbines // *Wind Energy*. 2017. vol. 20(9). pp. 1501–1513.
39. *Lee C-Y., Johnson A.L.* Effective production: measuring of the sales effect using data envelopment analysis // *Annals of Operations Research*. 2015. vol. 235(1). pp. 453–486.
40. *Lanzaa G. et al.* Measuring Global Production Effectiveness // *Procedia CIRP* 7. 2013. pp. 31–36.
41. *Eswaramurthi K., Mohanram P.* Improvement of Manufacturing Performance Measurement System and Evaluation of Overall Resource Effectiveness // *American Journal of Applied Sciences*. 2013. vol. 10(2). pp. 131–138.
42. *Azad M., Jafarian M., Farzipoor R., Mirhedayatian S.* A new fuzzy DEA model for evaluation of efficiency and effectiveness of suppliers in sustainable supply chain management context // *Computers&Operations Research*. 2015. vol. 54. pp. 274–285.
43. *Førsund F.* Measuring effectiveness of production in the public sector // *Omega*. 2017. vol. 73. pp. 93–103.
44. *Scott M., DeLone W., Golden W.* Measuring eGovernment success: a public value approach // *European Journal of Information Systems*. 2016. vol. 25(3). pp. 187–208.
45. *Bayne L., Schepis D., Purchase S.* A framework for understanding strategic network performance: Exploring efficiency and effectiveness at the network level // *Industrial Marketing Management*. 2017. vol. 67. pp. 134–147.
46. *Stojic N., Hashi I., Orlic E.* Creativity, innovation effectiveness and productive efficiency in the UK // *European Journal of Innovation Management*. 2018. pp. 1460–1060.
47. *Yang Z., Yong S., Hong Y.* Scale, congestion, efficiency and effectiveness in e-commerce firms // *Electronic Commerce Research and Applications*. 2016. vol. 20. pp. 171–182.
48. *Matinheikki Y., Pesonen T., Artto K., Peltokorpi A.* New value creation in business networks: The role of collective action in constructing system-level goals // *Industrial Marketing Management*. 2017. vol. 67. pp. 122–133.
49. *Yang Z., Yong S., Hong Y.* Analysis on pure e-commerce congestion effect, productivity effect and profitability in China // *Socio-Economic Planning Sciences*. 2017. vol. 57. pp. 35–49.
50. *Garza-Reyes J.* From measuring overall equipment effectiveness (OEE) to overall resource effectiveness (ORE) // *Journal of Quality in Maintenance Engineering*. 2015. vol. 21(4). pp. 506–527.
51. *Shelly P., Straub D., Liang T.* How Information Technology Governance Mechanisms and Strategic Alignment Influence Organizational Performance: Insights from a Matched Survey of Business and IT Managers // *MIS Quarterly*. 2015. vol. 39. pp. 497–518.
52. *Mikalef P., Pateli A.* Information technology-enabled dynamic capabilities and their indirect effect on competitive performance: Findings from PLS-SEM and fsQCA // *Journal of Business Research*. 2017. vol. 70. pp. 1–16.
53. *Wetering V., Mikalef P., Adamantia P.* A strategic alignment model for IT flexibility and dynamic capabilities: Toward an assessment tool // *Twenty-Fifth European Conference on Information Systems (ECIS)*. 2017. pp. 1–17.
54. *Laaksonen O., Peltoniemi M.* The essence of dynamic capabilities and their measurement // *International Journal of Management Reviews*. *British Academy of Management*. 2018. vol. 20(2). pp. 184–205.

55. *Schilke O., Hu S., Helfat C.* Quo vadis, dynamic capabilities? A content-analytic review of the current state of knowledge and recommendations for future research // Academy of Management Annals. 2018. vol. 12. no. 1. pp. 390–439.

Ашимов Абдыкаппар Ашимович — д-р техн. наук, профессор, академик Национальной академии наук Республики Казахстан, заведующий лабораторией системного анализа и управления, Институт проблем информатики и управления Министерства образования и науки Республики Казахстан. Область научных интересов: теория систем управления, техническая кибернетика, теория идентификации, теория усреднения стандартных дифференциальных уравнений. Число научных публикаций — 200. ashimov37@mail.ru; ул. Пушкина, 125, Алматы, 480100; р.т.: +7(727)272-83-05, Факс: +7(727)291-18-28.

Гейда Александр Сергеевич — к-т техн. наук, доцент, старший научный сотрудник лаборатории информационно-аналитических технологий в экономике, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: анализ и синтез организационно-технических, социально-экономических систем, оценивание эффективности их функционирования, потенциала организационно-технических и социально-экономических систем в условиях риска. Число научных публикаций — 131. geida@iias.spb.su, http://www.mathnet.ru/php/person.phtml?personid=58804&option_lang=rus; 14-я линия В.О., 39, Санкт-Петербург, 199178; р.т.: +7(812)328-3257, Факс: +7(812)328-4450.

Лысенко Игорь Васильевич — д-р техн. наук, профессор, заведующий лабораторией информационно-аналитических технологий в экономике, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: моделирование, информационно-аналитические технологии, экономический анализ функционирования организационно-технических систем, программно-целевое планирование и управление, разработка теории нечетких чисел и функций с приложениями. Число научных публикаций — 255. ilys@iias.spb.su; 14-я линия В.О., 39, Санкт-Петербург, 199178; р.т.: +7(812)328-3257, Факс: +7(812)328-4450.

Юсупов Рафаэль Мидхатович — д-р техн. наук, профессор, член-корреспондент РАН, заслуженный деятель науки и техники РФ, научный руководитель, Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), президент, НП Национальное общество имитационного моделирования («НОИМ»). Область научных интересов: теория управления, информатика, моделирование, теоретические основы информатизации и информационного общества, информационная безопасность. Число научных публикаций — 390. spiiiran@iias.spb.su; 14 линия В.О., 39, Санкт-Петербург, 199178; р.т.: +7-812-328-3311, Факс: +7(812)328-4450.

Поддержка исследований. Работа выполнена при поддержке грантов РФФИ № 16-08-00953 и 15-08-01825.

A.A. ASHIMOV, A.S. GEIDA, I.V. LYSENKO, R.M. YUSUPOV
**SYSTEM FUNCTIONING EFFICIENCY AND OTHER SYSTEM
OPERATIONAL PROPERTIES: RESEARCH PROBLEMS,
EVALUATION METHOD**

Ashimov A.A., Geida A.S., Lysenko I.V., Yusupov R.M. System Functioning Efficiency and Other System Operational Properties: Research Problems, Evaluation Method.

Abstract. The short historical sketch of researches of efficiency of systems functioning purposeful processes is given in this paper. The review of some relevant research problems of the operational properties solved abroad is provided.

On the basis of the analysis of features of efficiency research of systems functioning purposeful processes, as well as research of other operational properties, such as effectiveness, performance, operational and dynamic capabilities by domestic and foreign authors, the conclusion is drawn on relevance of the solution of some modern research problems on the basis of analytical estimation of operational properties indicators. A number of new systems and processes of their functioning operational properties researches directions is given. Among them, there are system capability (potentiality) and information technologies capability. Features of agile (dynamical, improved due to environment impact) systems and processes of their functioning, a role of information technologies are considered during functioning of such agile systems.

Main features of systems improvement and details of transition processes of such systems functioning improvement are described. The role of information technologies for systems improvement is discussed. The obtained results allowed to conduct research of operational properties of the improved systems, research of information technologies usage during system functioning. Examples of models of the improved system functioning effects formation are offered. Such models are developed taking into account the realization of information and non-information actions during the improved system functioning.

For the unified estimation of indicators of operational properties the method of analytical estimation of operational properties is offered. This method is based on the sequence of operational properties estimation schemes use. The sequence of the three schemes of operational properties estimation, which allows to estimate all described operational properties. Features of estimation of systems operational properties using the offered estimation method are revealed. The obtained results should allow to proceed to the solution of research problems of systems operational properties based on mathematical models use. An example of operational properties of information technology use indicators calculation is provided.

Keywords: efficiency, effectiveness, potential, operational properties, capability, dynamic capability, information technology capability, agile systems, evolving systems, purposes, indicators, models, modeling, method.

Ashimov Abdykappar Ashimovich — Ph.D., Dr. Sci., professor, Academician of the National Academy of Sciences of the Republic of Kazakhstan, head of systems analysis and control laboratory, Institute of Problems of Informatics and Control of the Ministry of Education and Science of the Republic of Kazakhstan. Research interests: theory of control systems, technical cybernetics, theory of identification, theory of the standard differential equations averaging. The number of publications — 200. ashimov37@mail.ru; 125 Pushkin str., Almaty, 480100, Republic of Kazakhstan; office phone: +7(727)272-83-05, Fax: +7(727)291-18-28.

Geida Alexander Sergeevich — Ph.D., assistant professor, senior researcher of laboratory for information-analytic technologies for economics, St. Petersburg Institute for Informatics and

Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: analysis and synthesis of techno-organizational, socio-economical systems, their functioning efficiency estimation, estimation of techno-organizational, socio-economical systems capabilities under risk conditions. The number of publications — 131. geida@iias.spb.su, http://www.mathnet.ru/php/person.phtml?personid=58804&option_lang=rus; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)328-3257, Fax: +7(812)328-4450.

Lysenko Igor Vasilievich — Ph.D., Dr. Sci., professor, head of laboratory for information-analytic technologies for economics, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: modeling, information-analytic technologies, economic analysis of techno-organizational systems functioning, fuzzy numbers theory and applications. The number of publications — 255. ilyas@iias.spb.su; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)328-3257, Fax: +7(812)328-4450.

Lysenko Igor Vasilievich — Ph.D., Dr. Sci., professor, head of laboratory for information-analytic technologies for economics, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: modeling, information-analytic technologies, economic analysis of techno-organizational systems functioning, fuzzy numbers theory and applications. The number of publications — 255. ilyas@iias.spb.su; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)328-3257, Fax: +7(812)328-4450.

Acknowledgements. This research is supported by RFBR (grants 16-08-00953 and 15-08-01825).

References

1. Kolmogorov A.N. [The number of hits for several shots and general principles of assessment of firing system effectiveness]. *Sbornik statej po teorii strel'by. Trudy Matemat. instituta im. V.A. Steklova – Collection of articles on the theory of firing. Steklov Mathematical Institute*. Moscow – Leningrad. 1945. vol. 12. pp. 7–25. (In Russ.).
2. Wentsel E.S, Likhterev Y.M, Milgram U. G, Khudyakov I.V. *Osnovy teorii boevoj ehffektivnosti i issledovaniya operatsij* [Basics of fighting efficiency and operations research]. Moscow: Zhukovskiy Air Force Engineering Academy. 1961. 524 p. (In Russ.).
3. Averkina G., Epshtein G. [Writer I. Grekova — professor E. S. Wentsel]. *Novyy Mir–New World*. 2008. vol. 4. (In Russ.).
4. Ioffe A.Ya. et al. *Veroyatnostnye metody v prikladnoj kibernetike* [Probabilistic Methods in Applied Cybernetics]. Leningrad. 1976. 424 p. (In Russ.).
5. Morozov L.M, Petukhov G.B, Sidorov V.N. *Metodologicheskie osnovy teorii ehffektivnosti* [Methodological basis of the effectiveness theory]. Leningrad. 1979. 174 p. (In Russ.).
6. Petukhov G.B. *Osnovy teorii ehffektivnosti tselenapravlennykh protsessov. Chast' 1. Metodologiya, metody, modeli* [Basics of the theory of efficiency of targeted processes. Part 1. Methodology, methods and models]. 1989. 647 p. (In Russ.).
7. Petukhov G.B, Yakunin V.I. *Metodologicheskie osnovy vneshnego proektirovaniya tselenapravlennykh protsessov i tselestremlennykh sistem* [Methodological basis of external designing of targeted processes and purposeful systems]. Moscow. AST. 2006. 504 p. (In Russ.).
8. Avduyevskiy V.S. *Nadezhnost' i ehffektivnost' v tekhnike: spravochnik v 10 t* [Reliability and efficiency in engineering: a guidebook in 10 volumes]. Moscow: Mechanical Engineering. 1988. Issue 3. 328 p. (In Russ.).

9. Dubrovskij V., Kovbich M. [Affairs and days of Academician Utkin, the designer of "Satan". The man who created the highest technical achievement of the twentieth century]. Available at: www.ng.ru/science/2013-10-23/10_utkin.html (accessed: 18.04.2018). (In Russ.).
10. Kryuchkov. U.I, Kuznetsov V.I. *Issledovaniya ehffektivnosti vooruzheniya* [Research on effectiveness of weapons]. Moscow: Dzerzhinsky Military Academy of Russian Strategic Missile Forces. 1980. (In Russ.).
11. Volkov E.B. et al. *Tekhnicheskie osnovy ehffektivnosti raketnykh sistem* [Engineering bases of efficiency of missile systems]. Moscow: Mechanical Engineering. 1989. (In Russ.).
12. Roldugin V.D. *Modelirovanie i otsenka ehffektivnosti boevykh dejstvij RVSN* [Simulation and evaluation of the Strategic Missile Forces fighting. Textbook]. Moscow: Dzerzhinsky Military Academy of Russian Strategic Missile Forces. 2005. (In Russ.).
13. Kryuchkov. Y.V, Kuznetsov. V.I, Okhotnikov G.N. *Modelirovanie i otsenka ehffektivnosti boevykh dejstvij i vooruzheniya raketnykh vojsk strategicheskogo naznacheniya* [Simulation and evaluation of military operations and weapons of Strategic Missile Forces]. Moscow. Dzerzhinsky Military Academy of Russian Strategic Missile Forces. 1986. 627 p. (In Russ.).
14. Lysenko I.V. *Analiz i sintez slozhnykh tekhnicheskikh sistem. Chast' 1. Analiz i sintez si-stem obespecheniya gotovnosti raket-nositelej i kosmicheskikh apparatov k zapusku (osnovy teorii)* [Analysis and synthesis of elaborate engineering systems. Part 1: Analysis and synthesis of systems providing readiness to launch and space vehicles for launching (theory basics)]. Moscow. Voenizdat. 1995. 368 p. (In Russ.).
15. Vikulov S.F. [Military-economic analysis: history, methodology, problems]. *Vooruzhenie i ehkonomika – Armament and Economics*. Moscow. 2012. vol. 4(20). pp. 86–98. (In Russ.).
16. Grushanskiy V.A., Ilyichev A.B., Thyong N. [Efficiency and Security indicators of public programs formalization under uncertainty conditions]. *Nelinejnij mir – Nonlinear world*. 2014. vol. 12(7). pp. 33–37. (In Russ.).
17. Podinovskiy V. Non-dominance and potential optimality for partial preference relations. *European Journal of Operational Research*. 2013. vol. 229. pp. 482–486.
18. Ryabinin I.A. Logical probabilistic analysis and its history. *International journal of risk assessment & management*. 2015. vol. 18. no. 3-4. pp. 256–265.
19. Geyda A.S., Ismahilova Z.F., Clitnyu I.V., Lysenko I.V. [Operational and exchange properties of systems research problems]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2014. vol. 4(35). pp. 136–160. (In Russ.).
20. Geyda A.S., Lysenko I.V. [Research problems of socio-economical systems capabilities]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2009. vol. 10. pp. 63–84. (In Russ.).
21. Geida A.S., Lysenko I.V., Yusupov R.M. [Main Concepts and Principles for Information Technologies Operational Properties Research]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2015. vol. 5(42). pp. 5–36. (In Russ.).
22. Yusupov R.M., Musaev A.A. [By Estimating the Effectiveness of Information Systems. Methodological Aspects]. *Informacionnye tehnologii – Information technologies*. 2017. vol. 23(5). pp. 323–332. (In Russ.).
23. *Rukovodstvo k svodu znaniy po upravleniyu proektami — PMBOK* [A Guide to the Project Management Body of Knowledge – PMBOK 5]. Moscow. Olymp-Business. 2014. 590 p. (In Russ.).
24. McMahon P. 15 Fundamentals for Higher Performance in Software Development: Includes discussions on CMMI, Lean Six Sigma, Agile and SEMAT's Essence Framework. Pem Systems Publ. 2014. 336 p.
25. ИВА. A Guide to Business Analysis Body of Knowledge (BABOK 2.0). 2015. 502 p.

26. Shalunova M. *Svod znaniy po upravleniyu biznes-protsessami. BPM CBOK 3.0* [BPM CBOK Version 3.0: Guide to the Business Process Management Common Body of Knowledge]. M.: Alpina Publisher. 2016. 640 p. (In Russ.).
27. Taylor J., Purchase J. *Real-World Decision Modeling with DMN*. 2016. Meghan-Kiffer Press. USA. 437 p.
28. Kendrick T. *How to Manage Complex Programs*. AMACOM. 2016. 336 p.
29. Dinsmore T.W. *Disruptive Analytics: Charting Your Strategy for Next-Generation Business Analytics*. Apress. 2016. 276 p.
30. *The Agile Extension to the BABOK® Guide*. International Institute of Business Analysis (IIBA). 2015. 128 p.
31. *Guide to the Software Engineering Body of Knowledge (SWEBOK) v.3*. IEEE Computer society. 2004. 335 p.
32. Blokdyk G. *COBIT 5: The Ultimate Step-By-Step Guide*. 5STARCOoks. 2018. 124 p.
33. Zidane Y., Olsson N. Defining project efficiency, effectiveness and efficacy. *International Journal of Managing Projects in Business*. 2017. vol. 10. Issue 3. pp. 621–641.
34. Carvalho M., Takao S., Silva D. Efficiency and Effectiveness Analysis of Public Transport of Brazilian Cities. *Journal of Transport Literature*. 2015. vol. 9(3). pp. 40–44.
35. Mesterova J., Prochazka J., Vaculik J., Smutny P. Relationship between Self-Efficacy, Transformational Leadership and Leader Effectiveness. *Journal of Advanced Management Science*. 2015. vol. 3. no. 2. pp. 109–122.
36. Nordon C. et al. The «Efficacy-Effectiveness Gap»: Historical Background and Current Conceptualization. *Value in Health*. 2016. vol. 19. Issue 1. pp. 75–81.
37. Docekalova M., Kocmanova A. Evaluation of the Effectiveness of Manufacturing Companies by Financial and Non-financial Indicators. *Procedia – Social and Behavioral Sciences*. 2015. vol. 213(1). pp. 491–496.
38. Hwangbo H., Johnson A.L., Ding Y. A Production Economics Analysis for Quantifying the Efficiency of Wind Turbines. *Wind Energy*. 2017. vol. 20(9). pp. 1501–1513.
39. Lee C-Y., Johnson A.L. Effective production: measuring of the sales effect using data envelopment analysis. *Annals of Operations Research*. 2015. vol. 235(1). pp. 453–486.
40. Lanzaa G. et al. Measuring Global Production Effectiveness. *Procedia CIRP*. 2013. pp. 31–36.
41. Eswaramurthi K., Mohanram P. Improvement of Manufacturing Performance Measurement System and Evaluation of Overall Resource Effectiveness. *American Journal of Applied Sciences*. 2013. vol. 10(2). pp. 131–138.
42. Azad M., Jafarian M., Farzipoor R., Mirhedayatian S. A new fuzzy DEA model for evaluation of efficiency and effectiveness of suppliers in sustainable supply chain management context. *Computers & Operations Research*. 2015. vol. 54. pp. 274–285.
43. Forsund F. Measuring effectiveness of production in the public sector. *Omega*. 2017. vol. 73. pp. 93–103.
44. Scott M., DeLone W., Golden W. Measuring eGovernment success: a public value approach. *European Journal of Information Systems*. 2016. vol. 25(3). pp. 187–208.
45. Bayne L., Schepis D., Purchase S. A framework for understanding strategic network performance: Exploring efficiency and effectiveness at the network level. *Industrial Marketing Management*. 2017. vol. 67. pp. 134–147.
46. Stojcic N., Hashi I., Orlic E. Creativity, innovation effectiveness and productive efficiency in the UK. *European Journal of Innovation Management*. 2018. pp.1460–1060.
47. Yang Z., Yong S., Hong Y. Scale, congestion, efficiency and effectiveness in e-commerce firms. *Electronic Commerce Research and Applications*. 2016. vol. 20. pp. 171–182.
48. Matinheikki Y., Pesonen T., Arto K., Peltokorpi A. New value creation in business networks: The role of collective action in constructing system-level goals. *Industrial Marketing Management*. 2017. vol. 67. pp. 122–133.

49. Yang Z., Yong S., Hong Y. Analysis on pure e-commerce congestion effect, productivity effect and profitability in China. *Socio-Economic Planning Sciences*. 2017. vol. 57. pp. 35–49.
50. Garza-Reyes J. From measuring overall equipment effectiveness (OEE) to overall resource effectiveness (ORE). *Journal of Quality in Maintenance Engineering*. 2015. vol. 21(4). pp. 506–527.
51. Shelly P., Straub D., Liang T. How Information Technology Governance Mechanisms and Strategic Alignment Influence Organizational Performance: Insights from a Matched Survey of Business and IT Managers. *MIS Quarterly*. 2015. vol. 39. pp. 497–518.
52. Mikalef P., Pateli A. Information technology-enabled dynamic capabilities and their indirect effect on competitive performance: Findings from PLS-SEM and fsQCA. *Journal of Business Research*. 2017. vol. 70. pp. 1–16.
53. Wetering V., Mikalef P., Adamantia P.A strategic alignment model for IT flexibility and dynamic capabilities: Toward an assessment tool. Twenty-Fifth European Conference on Information Systems (ECIS). 2017. pp. 1–17.
54. Laaksonen O., Peltoniemi M. The essence of dynamic capabilities and their measurement. *International Journal of Management Reviews. British Academy of Management*. 2018. vol. 20(2). pp. 184–205.
55. Schilke O., Hu S., Helfat C. Quo vadis, dynamic capabilities? A content-analytic review of the current state of knowledge and recommendations for future research. *Academy of Management Annals*. 2018. vol. 12. no. 1. pp. 390–439.

Signed to print 24.09.2018

Printed in Publishing center GUAP, 67, B. Morskaya, St. Petersburg, 190000, Russia

The journal is registered in Russian Federal Agency for Communications
and Mass-Media Supervision, certificate ПИ № ФС77-41695 dated August 19, 2010 г.
Subscription Index П5513, Russian Post Catalog

Подписано к печати 24.09.2018. Формат 60×90 1/16. Усл. печ. л. 14,35. Заказ № 412.

Тираж 150 экз., цена свободная.

Отпечатано в Редакционно-издательском центре ГУАП, 190000, Санкт-Петербург, Б. Морская, д. 67

Журнал зарегистрирован Федеральной службой по надзору в сфере связи
и массовых коммуникаций,
свидетельство ПИ № ФС77-41695 от 19 августа 2010 г.

Подписной индекс П5513 по каталогу «Почта России»

РУКОВОДСТВО ДЛЯ АВТОРОВ

Взаимодействие автора с редакцией осуществляется через личный кабинет на сайте журнала «Труды СПИИРАН» <http://www.proceedings.spiiras.nw.ru>. При регистрации авторам рекомендуется заполнить все предложенные поля данных.

Подготовка статьи ведется с помощью текстовых редакторов MS Word 2007 и выше. Объем основного текста – от 15 до 25 страниц включительно. Формат страницы документа – А5 (148 мм ширина, 210 мм высота); ориентация – портретная; все поля – 20 мм. Верхний и нижний колонтитулы страницы – пустые. Основной шрифт документа – Times New Roman, основной кегль (размер) шрифта – 10 pt. Переносы разрешены. Абзацный отступ устанавливается размером в 10 мм. Межстрочный интервал – одинарный. Номера страниц не проставляются.

В основную часть допускается помещать рисунки, таблицы, листинги и формулы. Правила их оформления подробно рассмотрены на нашем сайте в разделе «Руководство для авторов».

AUTHOR GUIDELINES

Interaction between each potential author and the Editorial board is realized through the personal account on the website of the journal "Proceedings of SPIIRAS" <http://www.proceedings.spiiras.nw.ru>. At the registration the authors are requested to fill out all data fields in the proposed form.

The submissions should be prepared using MS Word 2007 text editor or higher versions, at that, only manuscripts in *.docx format will be considered. The text of the paper in the main part of it should be from 15 – 25 pages of A5 size that is 210 X 148 mm; orientation – portrait; all margins – 20 mm. The font of the main paper text is Times New Roman of 10 pt font size. The pages' headers and footers should be empty; indentation – 10 mm; line spacing – single; pages are not numbered; hyphenations are allowed.

Certain figures, tables, listings and formulas are allowed in the main section, and their typography is considered by the paper template in more detail in journal web.

