

7. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ И СИСТЕМЫ АВТОМАТИЗАЦИИ В ПРОМЫШЛЕННОМ ПРОИЗВОДСТВЕ

УДК 004.934

РАЗРАБОТКА И ИССЛЕДОВАНИЕ МНОГОСЛОЙНОГО ПЕРСЕПТРОНА ДЛЯ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧЕВЫХ СИГНАЛОВ ЙЕМЕНСКОГО ДИАЛЕКТА АРАБСКОГО ЯЗЫКА

Н.Х.А. Радан, К.В. Сидоров, Ю.В. Сидорова

© Радан Н.Х.А., Сидоров К.В.,
Сидорова Ю.В., 2026

***Аннотация.** В статье рассмотрена система автоматического распознавания речевых сигналов йеменского диалекта арабского языка с помощью многослойного персептрона. В качестве информативных признаков речи применяются мел-частотные кепстральные коэффициенты. Описаны результаты тестирования системы.*

***Ключевые слова:** многослойный персептрон, распознавание речи, йеменский диалект, мел-частотные кепстральные коэффициенты.*

В последние полтора-два десятилетия динамично разрабатываются и развиваются системы автоматического распознавания речи (САРР). Подобные системы (в сочетании с нейросетевыми моделями и технологиями искусственного интеллекта) массово применяются в телефонии, в сфере бизнеса для автоматизации функционирования кол-центров, а также для автоматического сбора и анализа разнотипной информации, в маркетинговых исследованиях и при решении многих других задач [1].

Арабский язык занимает 4-е место в мире по распространенности (после китайского, испанского и английского). Анализ научной литературы в области САРР позволил установить, что существует ряд опубликованных работ по алгоритмам, моделям, системам и приложениям для автоматического распознавания стандартного арабского языка, а также различных вариаций его диалектов, таких как египетский, марокканский, алжирский, суданский, ливийский, иракский, тунисский и др. [2, 3]. Йемен-

скому диалекту посвящено меньшее количество научных исследований в области САРР по сравнению с другими указанными диалектами.

В данной статье описывается САРР, позволяющая распознавать речевые сигналы йеменского диалекта арабского языка. Распознавание речевых сигналов осуществляется с помощью искусственной нейронной сети – многослойного персептрона (МСП). Сеть состоит из отдельных элементов (нейронов). Нейрон представляет собой вычислительную единицу, которая получает информацию и производит над ней простые вычисления, а затем передает ее дальше. Модель искусственного нейрона представлена на рис. 1.

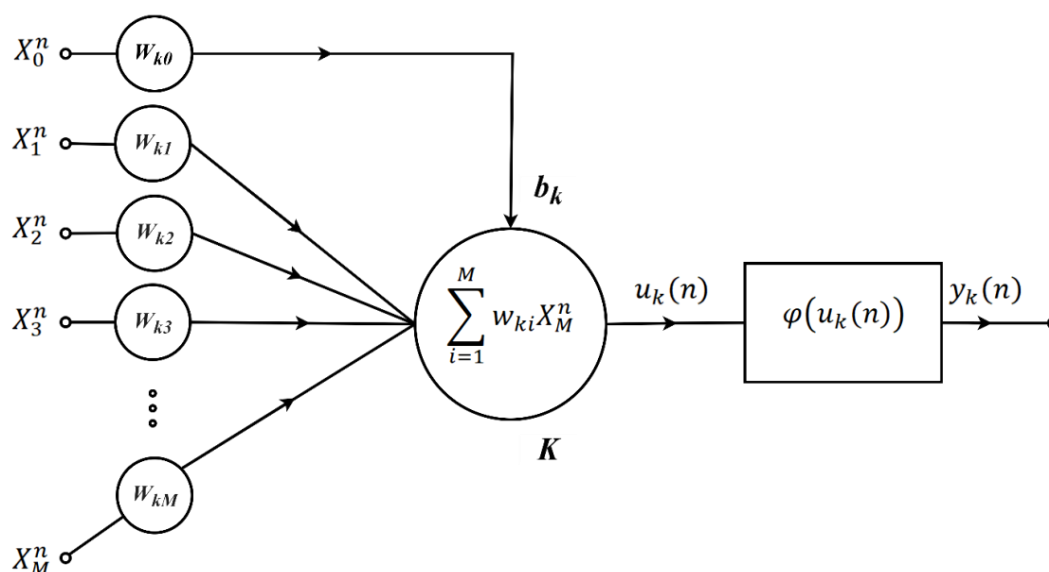


Рис. 1. Модель нейрона [4]

Формально функционирование такого нейрона с индексом k можно описать в виде следующих уравнений [4]:

$$u_k(n) = \sum_{i=1}^M w_{ki} X_i^n, \quad y_k(n) = \varphi(u_k(n) + b_k), \quad (1)$$

где X_i^n – компоненты входного вектора $x_n = \{X_1^n, X_2^n, \dots, X_M^n\}$ размерностью M в момент времени n ; $w_{ki} = \{w_{k1}, w_{k2}, \dots, w_{kM}\}$ – вектор синаптических весов k -го нейрона; b_k – порог чувствительности нейрона; $\varphi(\cdot)$ – функция активации; $y_k(n)$ – выходной сигнал нейрона.

Функция активации $\varphi(\cdot)$ (1) – функция, которая принимает на входе результат сумматора и выполняет преобразование, чтобы превратить сумму взвешенных входов в адекватный выход, который можно интерпретировать с точки зрения решения поставленной задачи.

Нейроны делятся на три основных типа: входной, скрытый и выходной слой. Входной слой получает информацию, n скрытых слоев ее

обрабатывают, выходной слой выводит результат анализа информации. Чем больше скрытых слоев, тем более сложные задачи может решать сеть. При отсутствии скрытых слоев возможна простая классификация данных. Один скрытый слой позволяет реализовать сложную функциональную зависимость между входными и выходными данными. Если число слоев увеличивать, то появляется возможность осуществлять распознавание сигналов. Простой считается нейронная сеть, имеющая лишь входной и выходной слои, без наличия скрытых слоев. В такой сети сигналы попадают сразу с входного слоя на выходной. Сеть подобного рода способна настроиться не под все сложные зависимости, очень чувствительна к предварительной обработке признаков и требует их тщательного отбора. Преимуществом является достаточно быстрая скорость обучения. Одной из самых известных и наиболее распространенных моделей нейронной сети является МСП. Его структурная схема представлена на рис. 2.

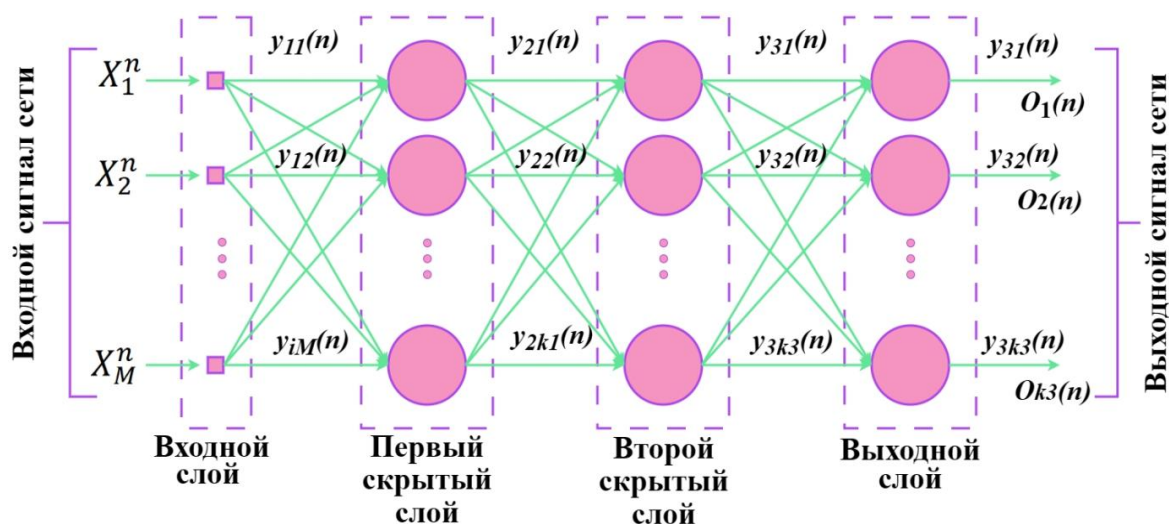


Рис. 2. Модель МСП [4]

Топология МСП – иерархическая сетевая структура, в которой нейроны разделены на несколько слоев. Внутри слоя нейроны можно считать линейно упорядоченными и не взаимодействующими между собой. Каждый нейрон сети получает входной сигнал от каждого нейрона предыдущего слоя, т. е. МСП – сеть прямого распространения, в которой входной сигнал распространяется только в прямом направлении от слоя к слою. Самый известный алгоритм обучения МСП – алгоритм обратного распространения ошибки. Сети с прямым распространением сигнала решают множество задач классификации и регрессии на табличных данных, а также многие задачи на изображениях и часть задач обработки естественного языка [4].

Многослойной (глубокой) нейронной сетью считается сеть не менее чем с двумя скрытыми слоями (входным и выходным). Нейроны в сети

связываются синапсами. У синапсов есть один параметр – вес W (1), благодаря чему входная информация изменяется при передаче от одного нейрона к другому.

Основными видами функции активации $\phi(.)$ (1) являются сигмоидная (логистическая) и функция гиперболического тангенса. Сигмоидная функция представляет собой самую распространенную функцию активации, диапазон ее выходных значений равен $[0; 1]$. Функция гиперболического тангенса используется в случае, когда выходные значения могут быть как отрицательными, так и положительными. Диапазон значений функции равен $[-1; 1]$.

Обучение обычно осуществляется по методу обратного распространения ошибки, в котором сеть получает множества входных и выходных данных. Ошибка определяется разностью между реальным и желаемым выходами. Этот процесс называется обучением с учителем. Он отличается от обучения без учителя тем, что во втором случае множество выходных данных сеть составляет самостоятельно.

В данной работе решается задача оценки возможности применения МСП с прямым распространением данных для разработки САРР (на примере речевых сигналов йеменского диалекта арабского языка). Для проведения исследований была сформирована база данных, которая содержит 5 000 структурированных звукозаписей (аудиоданных в формате wav) йеменского диалекта. В качестве дикторов выступили 10 мужчин (носителей йеменского диалекта), возраст которых варьировался от 25 до 30 лет. Все звукозаписи систематизированы и разделены на 10 классов речевых объектов, соответствующих арабским цифрам от 0 до 9 (табл. 1). От каждого диктора получено по 50 звукозаписей (частота дискретизации 16 кГц) для соответствующего класса речевых объектов.

Таблица 1

Арабские цифры

Цифра	Арабское написание	Фонетическое название	Произношение
0	صِفْر	Sifr	Sifr
1	واحد	Wahid	Wahid
2	إِثْنَيْن	?iθnajn	Ithnaiyn
3	ثَلَاثَة	θala:θih	Thalathah
4	أَرْبَعَة	?arbʕah	Arbaah
5	خَمْسَة	xamsih	Khamsih

Цифра	Арабское написание	Фонетическое название	Произношение
6	سِتَّة	Sittih	Sittih
7	سَبْعَة	sabʕah	Sabaah
8	ثَمَانِيَّة	θamani:h	Thamaniyah
9	تِسْعَة	tissʕah	Tisaah

На вход МСП подаются информативные признаки речи (арабских цифр от 0 до 9). На выходе формируются результаты классификации цифр. В качестве признаков речевых сигналов выбраны мел-частотные кепстральные коэффициенты (МЧКК). Длительность сегментов речи определяется интервалом квазистационарности, который равен 10–30 мс. Для каждого сегмента извлекались 12 коэффициентов МЧКК. С помощью программной среды Matlab исследовался двухслойный МСП. В качестве функции активации была использована сигмоида

$$S(z) = 1/(1 + e^{-z}), \quad (2)$$

где z – вход функции активации.

Результат работы функции активации (2) определяется числом, значение которого находится в пределах от 0 до 1. При создании сети применялся метод обучения *trainrp*, который использует ускоренный алгоритм обратного распространения ошибки, т. е. функция обучения нейронной сети обновляет значения весов и смещений на основе алгоритма устойчивого обратного распространения. Точность обучения сети определяется числом эпох обучения, количество которых можно скорректировать в процессе обучения при наблюдении за величиной отклонения результатов классификации от требуемых результатов. При исследовании двуслойной сети рассматриваются случаи, когда число нейронов входного слоя принимает следующие значения: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100.

Алгоритм проведения экспериментов показан на рис. 3.

Матрица распознавания цифр при использовании МСП представлена в табл. 2. Анализируя матрицу, можно сделать вывод о том, что наибольшая точность получена при распознавании цифр «ноль», «один», «два», «четыре» и «пять». Аналогично наименьшая точность получена при распознавании цифр «три», «шесть», «семь», «восемь» и «девять».

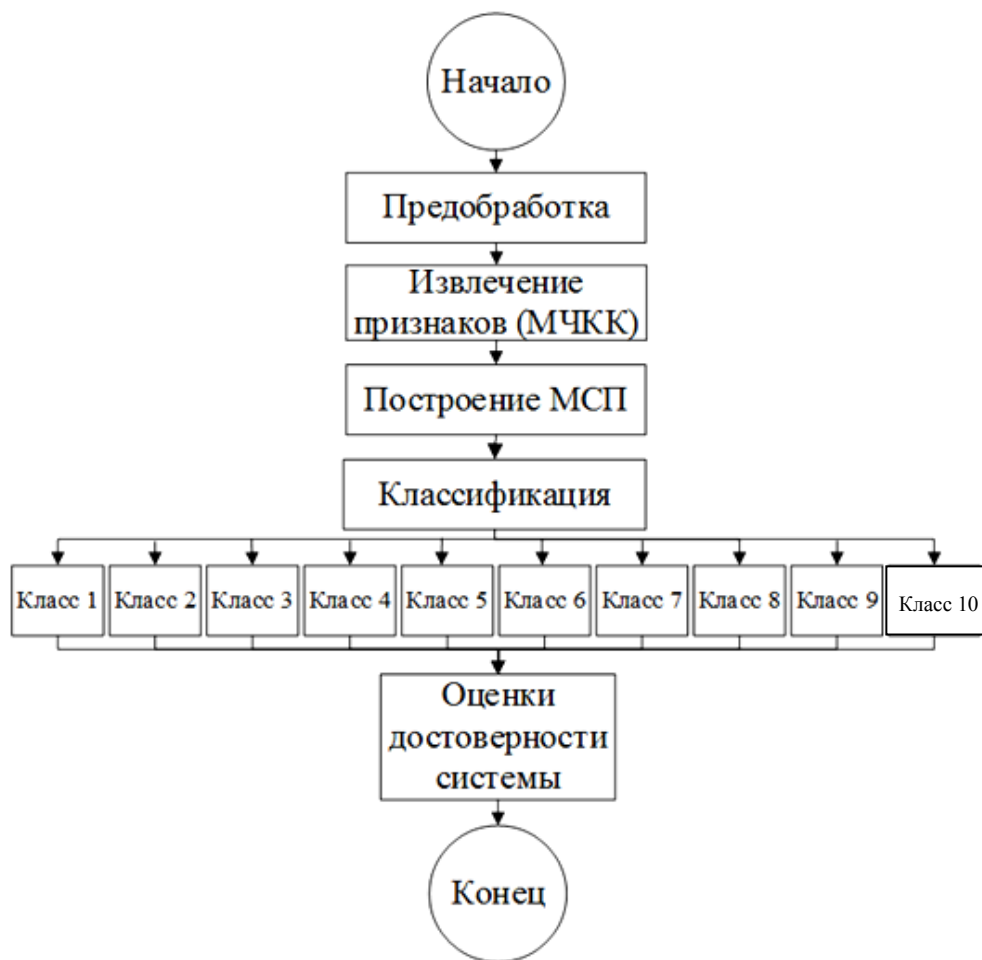


Рис. 3. Алгоритм проведения экспериментов

Таблица 2

Матрица автоматического распознавания речи

Модели (цифры)	0	1	2	3	4	5	6	7	8	9
0	482	0	2	5	1	0	1	0	0	0
1	4	490	0	29	0	1	0	1	4	3
2	0	2	489	2	1	2	5	0	7	7
3	8	7	1	458	1	3	1	7	6	3
4	0	1	0	0	491	2	3	3	1	0
5	0	0	0	0	0	492	1	1	2	1
6	4	0	4	0	0	0	486	0	0	2
7	2	0	0	5	6	0	2	488	0	6
8	0	0	2	0	0	0	0	0	480	0
9	0	0	2	1	0	0	1	0	0	478

В табл. 3 приведены детализированные оценки системы автоматического распознавания йеменского диалекта, где False Rejection Rate (FRR) – ошибка первого рода («ложный пропуск»), False Acceptance Rate (FAR) – ошибка второго рода («ложное срабатывание»).

Таблица 3

Оценки системы автоматического распознавания речи

Модели (цифры)	Точность, %	FRR, %	FAR, %	Общая ошибка, %	Взаимная точность, %
0	96,4	3,6	1,8	5,4	94,6
1	98	2	8,4	10,4	89,6
2	97,8	2,2	5,2	7,4	92,6
3	91,6	8,4	7,4	15,8	84,2
4	98,2	1,8	2	3,8	96,2
5	98,4	1,6	1	2,6	97,4
6	97,2	2,8	2	4,8	95,2
7	97,6	2,4	4,2	6,6	93,4
8	96	4	0,4	4,4	95,6
9	95,6	4,4	0,8	5,2	94,8
MEAN	96,68	3,32	3,32	6,64	93,36

Результаты оценки точности МСП представлены на рис. 4. Результаты оценки ошибки ложного пропуска и ложного срабатывания, а также общей ошибки МСП проиллюстрированы на рис. 5.

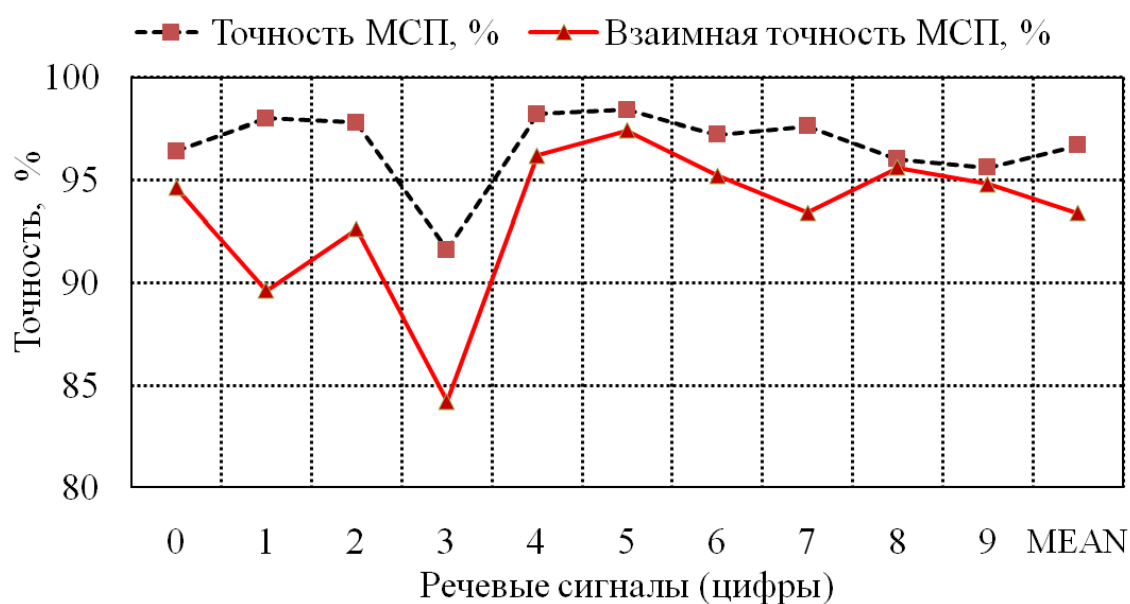


Рис. 4. Оценка точности распознавания

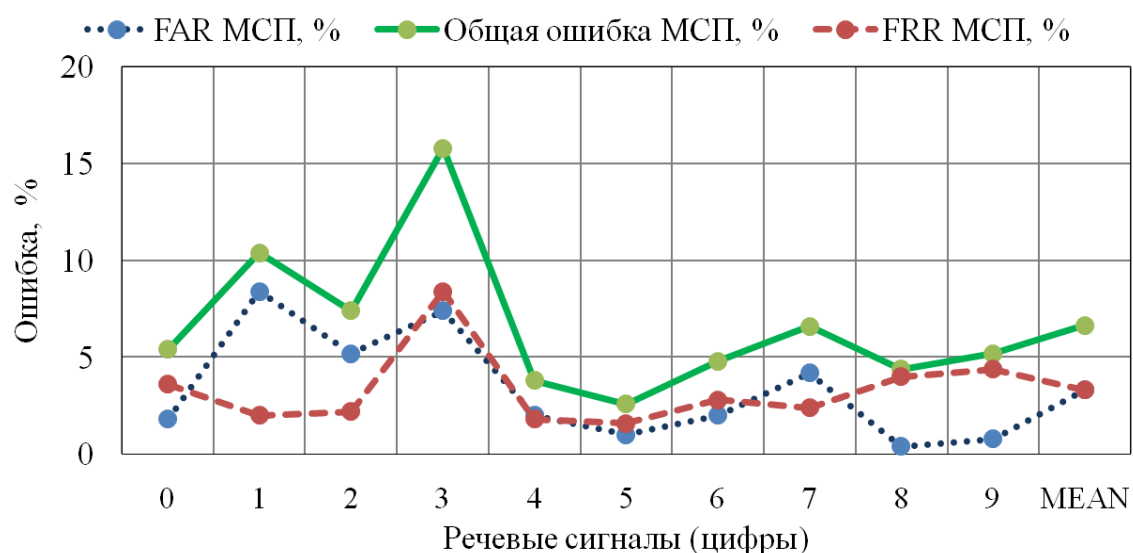


Рис. 5. Оценка ошибки распознавания

Таким образом, авторами работы предложен первичный вариант САРР йеменского диалекта арабского языка (на примере речевых сигналов арабских цифр). Система реализована с использованием МСП. Для нахождения информативных признаков используется математический аппарат МЧКК. Общая точность работы на текущем этапе исследований САРР составляет более 95 %. В дальнейшем планируется сотрудничество с автором работы [5] с целью проверки работоспособности предложенной САРР на записях телефонной речи в Республике Йемен.

Библиографический список

1. Методы и модели автоматического распознавания речи: учебное пособие / И.С. Кипяткова [и др.]. СПб.: СПб ФИЦ РАН, 2021. 116 с.
2. Arabic Speech Recognition with Deep Learning: A Review / W. Alghibab [et al.] // Social Computing and Social Media. Design, Human Behavior and Analytics. Lecture Notes in Computer Science. Vol. 11578. 2019. P. 15–31.
3. Радан Н.Х.А. Системы автоматического распознавания арабской речи и йеменского диалекта // Вестник Санкт-Петербургского университета государственной противопожарной службы МЧС России. 2023. № 2. С. 194–212.
4. Маковкин К.А. Гибридные модели – скрытые марковские модели / Многослойный персептрон – и их применение в системах распознавания речи. Обзор // Речевые технологии. 2012. Вып. 3. С. 58–83.
5. Аль-Дайбани А.М.С. Исследование методов и разработка алгоритмов обработки сигналов для систем автоматического распознавания

телефонной речи в Республике Йемен: дис. ... канд. техн. наук.
Владимир: ВлГУ, 2019. 150 с.

**DEVELOPMENT AND RESEARCH OF A MULTILAYER
PERCEPTRON FOR AUTOMATIC RECOGNITION
OF SPEECH SIGNALS OF THE YEMENI DIALECT
OF THE ARABIC LANGUAGE**

N.H.A. Radan, K.V. Sidorov, Yu.V. Sidorova

Abstract. *The paper considers the system of automatic recognition of speech signals of the yemeni dialect of the arabic language. Speech signals are recognized using a multilayer perceptron. Small-frequency cepstral coefficients are used as informative features of speech. The results of testing the system are described.*

Keywords: *multilayer perceptron, speech recognition, yemeni dialect, low-frequency cepstral coefficients.*

Об авторах:

РАДАН Наим Хуссейн Али – аспирант, ФГБОУ ВО «Тверской государственный технический университет», Тверь. E-mail: naeem.radan@gmail.com

СИДОРОВ Константин Владимирович – кандидат технических наук, доцент, доцент кафедры автоматизации технологических процессов, ФГБОУ ВО «Тверской государственный технический университет», Тверь. E-mail: bmsidorov@mail.ru

СИДОРОВА Юлия Владимировна – студентка, ФГБОУ ВО «Тверской государственный технический университет», Тверь. E-mail: us7129798@gmail.com

About the authors:

RADAN Naeem Hussein Ali – Postgraduate, Tver State Technical University, Tver. E-mail: naeem.radan@gmail.com

SIDOROV Konstantin Vladimirovich – Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Technological Process Automation, Tver State Technical University, Tver. E-mail: bmsidorov@mail.ru

SIDOROVA Yulia Vladimirovna – Student, Tver State Technical University, Tver. E-mail: us7129798@gmail.com