

П.Г. Букина, А.А. Меринов, С.С. Харченко, Е.Ю. Костюченко

**РАСПОЗНАВАНИЕ ЭМОЦИОНАЛЬНЫХ СОСТОЯНИЙ В РУССКОЯЗЫЧНОЙ РЕЧИ С ИСПОЛЬЗОВАНИЕМ MFCC-ПРИЗНАКОВ И BLSTM-МОДЕЛИ ДЛЯ НАБОРА ДАННЫХ «DUSHA»**

*Рассматривается задача автоматического распознавания эмоциональных состояний по речевому сигналу с использованием современных методов глубокого обучения. Актуальность исследования обусловлена растущей потребностью в интеллектуальных системах, способных оценивать эмоциональное состояние человека. Такие технологии могут найти широкое применение в медицине, психологии, информационных системах и управлении персоналом. Основной целью работы является разработка эффективной нейросетевой модели для распознавания эмоций в русскоязычной речи с точностью, превосходящей показатели современных архитектур. В качестве основы для экспериментов использован открытый русскоязычный набор данных «Dusha», содержащий 300000 аудиозаписей, из которых для обучения взято 183055 аудиозаписей подвыборки «Crowd», размеченной по четырём эмоциональным категориям: радость, грусть, злость и нейтральное состояние. Для подачи речевых сигналов в нейронную сеть были получены мел-частотные кепстральные коэффициенты (MFCC), которые демонстрируют высокую эффективность в задачах анализа и классификации речи, – 20 коэффициентов, извлеченные при использовании временного окна 20 мс и перекрытия 10 мс, с последующей нормализацией значений. Базовой архитектурой нейросетевой модели выступила двунаправленная рекуррентная сеть с долгой краткосрочной памятью (BLSTM), способная учитывать как прошлый, так и будущий контекст во временной последовательности. Для борьбы с переобучением и повышения обобщающей способности модель была модифицирована: в неё были добавлены свёрточные слои (CNN), слои MaxPooling для снижения размерности, а также механизмы Dropout и BatchNormalization. В результате предложенная гибридная CNN-BLSTM архитектура продемонстрировала точность 62,9% на тестовой выборке, что на 6,7% выше исходной модели (56,2%). Полученные результаты были сопоставлены с результатами современных архитектур, а именно MobileNetV2, HuBERT и WavLM. Анализ позволил обозначить перспективы дальнейшего повышения качества исследования за счет комплексной работы над структурой и параметрами используемой модели, а также балансировки классов и рассмотрения дополнительных акустических признаков.*

*Эмоции; набор данных; нейронная сеть; алгоритм; речевой сигнал; аудиозапись; классификация.*

P.G. Bukina, A.A. Merinov, S.S. Kharchenko, E.Y. Kostyuchenko

**RECOGNITION OF EMOTIONAL STATES IN RUSSIAN SPEECH USING MFCC FUNCTIONS AND THE BLSTM MODEL FOR THE DUSHA DATASET**

*This paper investigates the task of automatic emotion recognition from speech signals using contemporary deep learning techniques. The relevance of this study arises from the increasing demand for intelligent systems capable of assessing human emotional states, with potential applications in medicine, psychology, information systems, and personnel management. The primary objective is to develop an efficient neural network model for emotion recognition in Russian speech that outperforms existing state-of-the-art architectures. The experiments were conducted using the open-source Russian-language dataset Dusha, which contains 300,000 audio recordings. A total of 183,055 samples from the Crowd subset, annotated with four emotional categories—joy, sadness, anger, and neutral state—were used for training. Mel-frequency cepstral coefficients (MFCCs) were extracted as input features (20 coefficients with a 20 ms window and 10 ms overlap), followed by normalization. The baseline architecture employed a bidirectional long short-term memory network (BLSTM), capable of modeling both past and future temporal dependencies. To improve generalization and mitigate overfitting, the model was enhanced with convolutional layers (CNN), MaxPooling layers, and regularization mechanisms including Dropout and Batch Normalization. The resulting hybrid CNN-BLSTM architecture achieved 62.9% accuracy on the test set, exceeding the baseline performance (56.2%) by 6.7%. The results were further compared with state-of-the-art architectures such as MobileNetV2, HuBERT, and WavLM. The analysis highlights future directions for improving model performance through structural optimization, class balancing, and incorporation of additional acoustic features.*

*Emotion recognition; dataset; neural networks; algorithm; speech signal; audio recording; classification.*

**Введение.** Современные технологии обработки аудиоданных дают возможность автоматически определять психоэмоциональное состояние человека. Анализ речевых сигналов как способ определения психоэмоционального состояния является важным направлением в развитии систем искусственного интеллекта [1]. Эмоции отражаются в акустических характеристиках речи, связанных с физиологическим и когнитивным состоянием человека, что позволяет использовать голос как индикатор внутреннего состояния [2]. Такие технологии находят применение в медицине, психологии, информационных системах и управлении персоналом. Например, в корпоративной среде отслеживание эмоционального состояния сотрудников может способствовать повышению продуктивности и снижению уровня истощения при решении профессиональных задач на работе [3].

Исследуя зависимость между речью и стрессом в 1988 году Национальный институт истины США разработал систему CVSA (Computer Voice Stress Analyzer), способную анализировать уровень стресса по голосу с целью выявления лжи, что нашло широкое применение в правоохранительной практике [4]. Также в настоящее время системы распознавания эмоций активно внедряются в службы клиентской поддержки и колл-центры для оценки качества взаимодействия и автоматического выявления конфликтных ситуаций [5].

Реализация системы распознавания эмоций требует двух этапов: формирования репрезентативного набора данных и выбора модели классификации. Современные исследования опираются на аудиоданные с эмоционально окрашенной речью, однако существует дисбаланс в распределении по языкам – англоязычные наборы данных преобладают, тогда как для других языков, включая русский, наблюдается заметный дефицит [6]. На этапе проектирования исследователи сталкиваются с необходимостью выбора набора эмоций. При выборе набора часто используется модель с 6 базовыми эмоциями, предложенная Полом Экманом: радость, гнев, грусть, страх, удивление и отвращение [7]. При этом если опираться на более широкую профессиональную практику, то поздние работы, в частности исследование Калифорнийского университета в Беркли, демонстрируют существование 27 базовых эмоций и их непрерывный переход друг в друга [8].

Другой распространённый подход к классификации – использование двумерного пространства «валентность–возбуждение». Возбуждение коррелирует с физиологическими параметрами, такими как сердечный ритм и кровяное давление, что отражается в речевой динамике: радость и гнев проявляются через энергичную речь, а грусть через медленную и монотонную. При этом валентность является характеристикой положительной или отрицательной окраски эмоций, которая не имеет однозначной речевой корреляции. Исследование Йоханнеса Вагнера указывает на эффективность использования трансформеров для извлечения скрытой лингвистической информации, улучшающей классификацию по шкале валентности [9].

Также при исследовании эмоций данных ключевым аспектом остаётся степень естественности речи в аудиоданных [10]. Выделяют три уровня: естественная (в реальных ситуациях), полустественная (в кино, постановках), и искусственная (дикторская речь с заданными фразами). Первые редко используются из-за этических ограничений, третьи страдают недостаточной реалистичностью эмоций. Оптимальный компромисс предлагают полустественные данные. Также необходимо учитывать параметры говорящего: пол, возраст и язык. Половые и возрастные различия в экспрессии эмоций требуют сбалансированных выборок для обеспечения универсальности модели.

В данной работе ставится задача разработки эффективной модели автоматического распознавания эмоций в русскоязычной речи. Актуальность модели обусловлена малым количеством решений, адаптированных под русский язык. Большинство современных нейросетевых моделей обучаются и тестируются на англоязычных наборах данных, что ограничивает их применимость в русскоязычной среде. Следовательно, существует необходимость в создании специализированной модели, учитывающей особенности русскоязычной речи и не уступающей по точности имеющимся архитектурам.

**1. Исследование наборов данных.** Набор данных «Dusha» является открытым русскоязычным набором, он состоит из 300000 аудиозаписей в формате «WAV», сопровождаемых транскрипциями, и предназначен для задач распознавания эмоций по речевым признакам [11]. Набор состоит из двух групп: Crowd и Podcast. Первая группа представ-

ляет озвученные актерами сгенерированные тексты на базе разговоров людей с виртуальными ассистентами. Вторая группа содержит маленькие выжимки из подкастов, классифицированные по эмоциям. В данном наборе присутствует четыре класса эмоций: печаль, гнев, счастье и нейтральность.

В табл. 1 представлены русскоязычные наборы данных и сравнительные характеристики. Преимуществами набора данных «Dusha» перед другими русскоязычными наборами данных являются внушительный размер, большое количество задействованных говорящих, доступность, актуальность и современность. Для наборов данных «Ruslana» [12] и «RESD» [13] ограничен доступ.

Таблица 1

#### Русскоязычные наборы данных

Название	Объем	Количество говорящих	Метод создания	Количество эмоций
Dusha	300000 записей	Более 2000	Чтение реплик, подкасты	4
Ruslana	3660 записей	61	Чтение предложений	5
RESD	Более 3000 записей	200	Чтение диалогов	7

В дальнейшем исследовании используется обучающая часть «Crowd» набора данных «Dusha», содержащая 183055 аудиозаписей, поскольку часть с подкастами недоступна по причине авторского права. В табл. 2 представлено количество аудиозаписей в обучающем наборе для каждой эмоции.

Таблица 2

#### Содержание обучающего набора

Наименование эмоции	Количество аудиофайлов	Длительность, час
Радость	22007	29
Грусть	56378	73
Злость	29798	39
Нейтральность	74872	88

Для получения входных данных было решено использовать алгоритм вычисления мел-частотных кепстральных коэффициентов (MFCC).

**2. Извлечение входных признаков.** В качестве входных признаков для нейронной сети были выбраны мел-частотные кепстральные коэффициенты (MFCC). Они представляют вектор из определенного количества вещественных чисел, отражающий энергию спектра речевого сигнала. Данный признак учитывает волновую природу сигнала и выделяет наиболее существенные частоты, воспринимаемые человеком. MFCC устойчивы к шуму, а также они широко используются в распознавании речи [14–16].

Процесс извлечения мел-частотных кепстральных коэффициентов представлен на рис. 1 и включает несколько этапов. Аудиосигнал разбивается на короткие перекрывающиеся кадры с применением оконной функции. После чего для каждого кадра вычисляется спектр мощности при помощи быстрого преобразования Фурье. Полученная энергия спектра пропускается через банк треугольных мел-фильтров, имитирующих нелинейное восприятие частот человеческим ухом. Далее применяются логарифмирование для усиления низкоамплитудных компонентов и дискретное косинусное преобразование (DCT) для сжатия информации в вектор коэффициентов.

Для извлечения MFCC из аудиофайла использовалась библиотека «Librosa» [17]. Она содержит функцию «mfcc», позволяющую вычислять заданное количество MFCC. При помощи данной функции были получены 20 признаков с размером окна 20 мс и перекрытием 10 мс. Полученные массивы коэффициентов были нормализованы, в результате чего все значения были приведены к диапазону от 0 до 1.



Рис. 1. Алгоритм извлечения MFCC из аудиосигнала

**3. Разработка архитектуры нейронной сети.** В задаче распознавания эмоций по речевому сигналу зачастую отдают предпочтение искусственному интеллекту. В основном исследователи используют сверточную или рекуррентную архитектуры нейронной сети. В данной работе отдано предпочтение рекуррентной нейронной сети (RNN) на основе длительной кратковременной памяти (LSTM), так как она имеет преимущество в связи с возможностью извлечения эволюции долгосрочных временных рядов за счет памяти [18]. Сверточные нейронные сети (CNN) предназначены для решения задач распознавания объектов, однако имеют место в распознавании эмоций в речи, поскольку они быстрее обучаются и содержат меньшее число параметров для оптимизации.

Первоначально реализованная архитектура имеет четыре двунаправленных слоя (рис. 2). Основная идея двунаправленного слоя заключается в учете нейронной сетью прошлого и будущего контекста для каждой точки во входной последовательности за счет подачи сигнала как вперед, так и назад во времени. Первый двунаправленный LSTM-слой имеет по 256 скрытых нейронов в прямом и обратном направлении, второй – 128 нейронов, в третьем слое количество нейронов уменьшено до 64. Четвертый слой содержит также 64 скрытых нейрона на каждое направление, однако возвращает только последний временной шаг выходной последовательности, когда предыдущие слои возвращают выход для каждого временного шага последовательности. Выходной слой имеет 4 нейрона, соответствующих количеству классов. Поскольку рассматривается многоклассовая классификация, в качестве функции активации выбрана функция «softmax».



Рис. 2. Архитектура исходной модели

Представленная на рис. 2 модель нейронной сети обучалась на 50 эпохах и показывала стабильные результаты.

Графики точности (рис. 3) и функции потерь (рис. 4) исходной модели при обучении свидетельствуют о том, что модель переобучается, так как доля верных ответов на валидационной выборке значительно ниже доли верных ответов на обучающей выборке и стабилизируется на данном уровне (около 0,56), не достигая точности обучающей выборки.

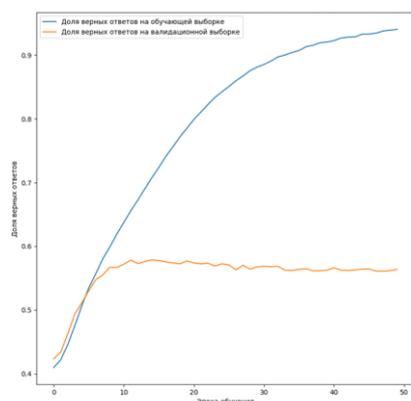


Рис. 3. График точности исходной модели

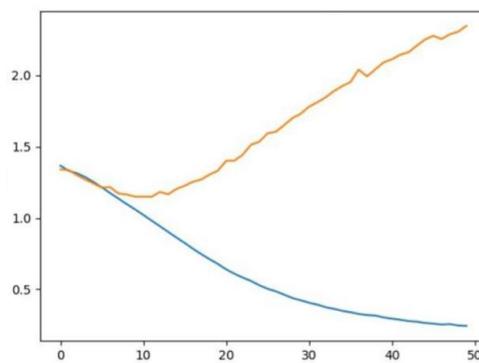


Рис. 4. График функции потерь исходной модели

На основе полученных результатов было принято решение избежать переобучение за счет изменения архитектуры модели, сама архитектура модели была изменена при помощи эмпирического метода тестирования, который включал в себя постепенное изменение параметров и структуры (рис. 5).

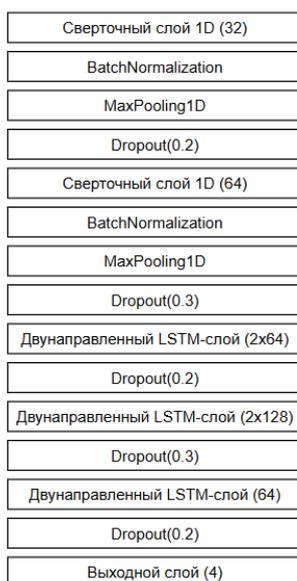


Рис. 5. Архитектура итоговой модели

В результате в итоговой модели сокращено количество нейронов и двунаправленных LSTM-слоев по причине малой скорости обучения и высокого риска переобучения за счет большого количества параметров. Также итоговая модель содержит сверточные слои, поскольку они применяют фильтры, способные вычлнять локальные признаки внутри временных окон, например, резкие изменения частоты, связанные с определенными эмоциональными состояниями. Использование слоев MaxPooling позволяет сокра-

тить размерность данных, тем самым улучшив обобщающую способность модели при помощи выделения наиболее значимых признаков и сокращения несущественной информации. Для борьбы с переобучением были добавлены слои Dropout, которые во время обучения «отключают» определенную долю нейронов на каждой итерации, что позволяет избежать сильной зависимости между нейронами. Слои BatchNormalization позволяют ускорить настройку модели за счет нормализации не только входных признаков, но и активаций промежуточного слоя.

**4. Обучение разработанной модели.** Обучающий набор Crowd был разбит на три части: 70% взято на обучение модели, 20% – на валидацию, 10% – на тестирование. После многочисленных прогонов было выявлено оптимальное количество эпох, необходимых модели для обучения, а именно 300. Также для модели были автоматически настроены веса для классов, чтобы минимизировать влияние несбалансированности набора. В табл. 3 представлены результаты обучения исходной модели и итоговой. В качестве функции потерь (loss) использовалась категориальная кросс-энтропия.

Таблица 3

**Результаты обучения на валидационном наборе**

Модель	Акcuracy, %	F1-Score, %	Loss
Исходная	56,39	56,2	2,3082
Итоговая	63,41	63,68	0,8923

График точности (рис. 6) модели на обучающем и валидационном наборах отражает способность модели к обучению, в отличие от первоначальной архитектуры.

Также функция потерь итоговой модели стремится к нулю в процессе обучения, что свидетельствует об отсутствии переобучения (рис. 7).

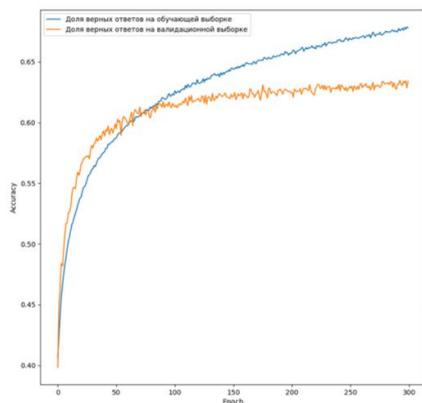


Рис. 6. График точности итоговой модели

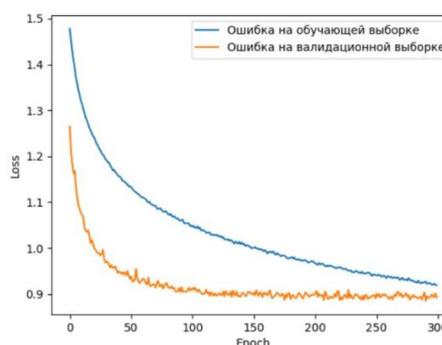


Рис. 7. График функции потерь итоговой модели

В результате изменения архитектуры и параметров модели удалось исключить переобучение и повысить точность на валидационной выборке.

**5. Тестирование разработанных моделей.** В табл. 4 представлено количество аудиофайлов для каждой эмоции в тестовом наборе данных.

Таблица 4

**Содержание тестового набора**

Наименование эмоции	Количество аудиофайлов
Злость	2404
Грусть	4459
Нейтральность	6036
Радость	1746

В табл. 5 представлены полученные метрики оценки качества исходной и итоговой моделей на тестовых данных.

Таблица 5

#### Результаты тестирования созданных моделей

Модель	Accuracy	Precision	Recall	F1-Score
Исходная	0,562	0,573	0,563	0,570
Итоговая	0,629	0,636	0,629	0,631

В результате тестирования итоговая модель показала увеличение значений метрик качества на 6-7% в сравнении с исходной моделью.

Полученные в результате тестирования метрики точности разработанных моделей были сравнены с результатами, представленными в статьях [19, 20]. В первой статье используется архитектура, основанная на MobileNetV2, которая в качестве входных данных принимает мел-частотные признаки. Во второй статье набор данных тестируется на моделях HuBERT и WavLM, которые используются в задачах обработки разговорной речи и учитывают контекст.

Поскольку в статьях используется метрика accuracy, результаты тестирования разработанных моделей были сравнены по ней с результатами моделей, представленных в статьях (табл. 6). Также модели были сравнены с моделью на основе MobileNetV2 при помощи метрики f-меры (f1-score), т.к. набор данных является несбалансированным. По другим двум моделям оценка данной метрикой не проводилась.

Таблица 6

#### Сравнение результатов с современными моделями

Модель	F1-Score	Accuracy
Исходная модель	0,573	0,562
Итоговая модель	0,636	0,629
MobileNetV2	0,770	0,760
HuBERT	-	0,866
WavLM	-	0,872

**Заключение.** Экспериментально показано, что модифицированная модель на основе BLSTM с добавлением сверточных слоев обеспечивает улучшение точности классификации более чем на 6% по сравнению с исходной архитектурой, что подтверждает эффективность внесенных изменений. Полученные результаты рассмотрены в сравнении с современными архитектурами, такими как MobileNetV2, HuBERT и WavLM, что позволяет выделить возможности для дальнейшего повышения качества исследования путём оптимизации структуры и параметров модели.

*Данная работа выполнялась в рамках Программы развития ТУСУР на 2025-2036 годы Программы стратегического академического лидерства «Приоритет 2030».*

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Lian H., Lu C., Li S., Zhao Y. A Survey of deep learning-based multimodal emotion recognition: speech, text, and face // Entropy. – 2023. – No. 25 (10).
2. Богданова Д.Р., Акушев А.Т. Распознавание эмоций по речевому сигналу // E-Scio: электрон. научн. журн. – 2021. – № 6 (57).
3. Назарова Е.К. Влияние психики на производительность труда // Universum: психология и образование. – 2024. – № 7. – С. 53-56.
4. Горшков Ю.Г. Визуализация эмоциональной напряженности человека по речевому сигналу // Научная визуализация. – 2023. – № 2. – С. 102-112.
5. Никифоров А.А. Разработка модуля распознавания эмоций разговора колл-центра с использованием рекуррентных искусственных нейронных сетей, для выявления нежелательного контента // Вестник науки. – 2023. – № 7. – С. 226-231.
6. Малыгина Ю.П. Нейронные сети: особенности, тенденции, перспективы развития // Молодой исследователь Дона: электронный научный журнал. – 2018. – № 5 (14). – С. 79-82.

7. Ekman P., Oster H. Facial expressions of emotions // *Annual Review of Psychology*. – 1979. – No. 30. – P. 527-554.
8. Cowen A.S., Keltner D. Self-report captures 27 distinct categories of emotion bridged by continuous gradients // *Proceedings of the National Academy of Sciences, PNAS*. – 2017. – No. 114 (38). – P. E7900-E7909.
9. Wagner J., Triantafyllopoulos A., Wierstorf H., Schmitt M., Burkhardt F., Eyben F., Schuller B.W. Dawn of the transformer era in speech emotion recognition: closing the valence gap // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2023.
10. Ayadi M., Kamel M., Karray F. Survey on speech emotion recognition: Features, classification schemes, and databases // *Pattern Recognition*. – 2011. – No. 44. – P. 572-587.
11. Датасет для распознавания эмоций – Dusha. – Режим доступа: <https://developers.sber.ru/portal/products/dusha?ysclid=m7hrx3w5t3967089717>.
12. Makarova V., Petrushin V. Ruslana: a database of Russian emotional utterances // *7th International Conference on Spoken Language Processing*. – 2002. – P. 2041-2044.
13. Russian emotional speech dialogs (RESL). – Режим доступа: <https://www.kaggle.com/datasets/ar4ikov/resd-dataset>, свободный.
14. Sahidullah M., Saha G. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition // *Speech Communication*. – 2012. – Vol. 54, Issue 4. – P. 543-565.
15. Jagtap S., Desai K., Patil J. A Survey on Speech Emotion Recognition Using MFCC and Different Classifier. – 2022.
16. Badr Y., Mukherjee P., Thumati S. Speech Emotion Recognition using MFCC and Hybrid Neural Networks // *13<sup>th</sup> International Conference on Neural Computation Theory and Applications*. – 2021.
17. Librosa – librosa 0.10.2 documentation. – Режим доступа: <https://librosa.org/doc/latest/index.html>.
18. Hochreiter S., Schmidhuber J. Long Short-Term Memory // *Neural Computation*. – 1997. – No. 9 (8). – P. 1735-1780.
19. Kondratenko V., Sokolov A., Karpov N., Kutuzov O., Savushkin N., Minkin F. Large raw emotional dataset with aggregation mechanism // *ArXiv (Cornell University)*. – 2022.
20. Лемаев В.И., Лукашевич Н.В. Автоматическая классификация эмоций в речи: методы и данные // *Litera. Nota bene*. – 2024. – № 4. – С. 159-173.

## REFERENCES

1. Lian H., Lu C., Li S., Zhao Y. A Survey of deep learning-based multimodal emotion recognition: speech, text, and face, *Entropy*, 2023, No. 25 (10).
2. Bogdanova D.R., Akushev A.T. Raspoznavanie emotsiy po rechevomu signalu [Emotion recognition based on speech signal], *E-Scio: elektron. nauchn. zhurn* [E-Scio: electronic scientific journal], 2021, No. 6 (57).
3. Nazarova E.K. Vliyanie psikhiki na proizvoditel'nost' truda [Influence of the psyche on labor productivity], *Universum: psikhologiya i obrazovanie* [Universum: Psychology and Education], 2024, No. 7, pp. 53-56.
4. Gorshkov Yu.G. Vizualizatsiya emotsional'noy napryazhennosti cheloveka po rechevomu signalu [Visualization of human emotional tension based on speech signal], *Nauchnaya vizualizatsiya* [Scientific Visualization], 2023, No. 2, pp. 102-112.
5. Nikiforov A.A. Razrabotka modulya raspoznavaniya emotsiy razgovora koll-tsentra s ispol'zovaniem rekurrentnykh iskusstvennykh neyronnykh setey, dlya vyavleniya nezhelatel'nogo kontenta [Development of a module for recognizing emotions in call center conversations using recurrent artificial neural networks to identify unwanted content], *Vestnik nauki* [Herald of Science], 2023, No. 7, pp. 226-231.
6. Malygina Yu.P. Neyronnye seti: osobennosti, tendentsii, perspektivy razvitiya [Neural networks: features, trends, prospects for development], *Molodoy issledovatel' Dona: elektronnyy nauchnyy zhurnal* [Young researcher of the Don: electronic scientific journal], 2018, No. 5 (14), pp. 79-82.
7. Ekman P., Oster H. Facial expressions of emotions, *Annual Review of Psychology*, 1979, No. 30, pp. 527-554.
8. Cowen A.S., Keltner D. Self-report captures 27 distinct categories of emotion bridged by continuous gradients, *Proceedings of the National Academy of Sciences, PNAS*, 2017, No. 114 (38), pp. E7900-E7909.
9. Wagner J., Triantafyllopoulos A., Wierstorf H., Schmitt M., Burkhardt F., Eyben F., Schuller B.W. Dawn of the transformer era in speech emotion recognition: closing the valence gap, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
10. Ayadi M., Kamel M., Karray F. Survey on speech emotion recognition: Features, classification schemes, and databases, *Pattern Recognition*, 2011, No. 44, pp. 572-587.
11. Dataset dlya raspoznavaniya emotsiy – Dusha [Dataset for emotion recognition – Dusha]. Available at: <https://developers.sber.ru/portal/products/dusha?ysclid=m7hrx3w5t3967089717>.

12. Makarova V., Petrushin V. Ruslana: a database of Russian emotional utterances, *7th International Conference on Spoken Language Processing*, 2002, pp. 2041-2044.
13. Russian emotional speech dialogs (RESL). Available at: <https://www.kaggle.com/datasets/ar4ikov/resd-dataset>.
14. Sahidullah M., Saha G. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition, *Speech Communication*, 2012, Vol. 54, Issue 4, pp. 543-565.
15. Jagtap S., Desai K., Patil J. A Survey on Speech Emotion Recognition Using MFCC and Different Classifier, 2022.
16. Badr Y., Mukherjee P., Thumati S. Speech Emotion Recognition using MFCC and Hybrid Neural Networks, *13<sup>th</sup> International Conference on Neural Computation Theory and Applications*, 2021.
17. Librosa – librosa 0.10.2 documentation. – Режим доступа: <https://librosa.org/doc/latest/index.html>.
18. Hochreiter S., Schmidhuber J. Long Short-Term Memory, *Neural Computation*, 1997, No. 9 (8), pp. 1735-1780.
19. Kondratenko V., Sokolov A., Karpov N., Kutuzov O., Savushkin N., Minkin F. Large raw emotional dataset with aggregation mechanism, *ArXiv (Cornell University)*, 2022.
20. Lemaev V.I., Lukashevich N.V. Avtomaticheskaya klassifikatsiya emotsiy v rechi: metody i dannye [Automatic classification of speech emotions: methods and data], *Litera. Nota bene*, 2024, No. 4, pp. 159-173.

**Букина Полина Германовна** – Томский государственный университет систем управления и радиоэлектроники; e-mail: bukina.polina2014@gmail.com; г. Томск, Россия; кафедра безопасности информационных систем; студент.

**Мерин Арсен Арзуманович** – Томский государственный университет систем управления и радиоэлектроники; e-mail: merinovarsen@mail.ru; г. Томск, Россия; кафедра безопасности информационных систем; студент.

**Харченко Сергей Сергеевич** – Томский государственный университет систем управления и радиоэлектроники; e-mail: kss@fb.tusur.ru; г. Томск, Россия; кафедра безопасности информационных систем; к.т.н.; доцент.

**Костюченко Евгений Юрьевич** – Томский государственный университет систем управления и радиоэлектроники; e-mail: key@keva.tusur.ru; г. Томск, Россия; кафедра безопасности информационных систем; к.т.н.; и.о. зав. кафедрой.

**Bukina Polina Germanovna** – Tomsk State University of Control Systems and Radioelectronics; e-mail: bukina.polina2014@gmail.com; Tomsk, Russia; the Department of Information Security Systems; student.

**Merinov Arsen Arzumanovich** – Tomsk State University of Control Systems and Radioelectronics; e-mail: merinovarsen@mail.ru; Tomsk, Russia; the Department of Information Security Systems; student.

**Kharchenko Sergey Sergeevich** – Tomsk State University of Control Systems and Radioelectronics; e-mail: kss@fb.tusur.ru; Tomsk, Russia; the Department of Information Security Systems; cand. of eng. sc.; associate professor.

**Kostyuchenko Evgeny Yurievich** – Tomsk State University of Control Systems and Radioelectronics; e-mail: key@keva.tusur.ru; Tomsk, Russia; the Department of Information Security Systems; cand. of eng. sc.; acting head of department.

УДК 004.89

DOI 10.18522/2311-3103-2025-6-248-262

**Ж.Х. Мохаммад**

## **МЕТОДИКА ПОСТРОЕНИЯ И ОЦЕНКИ ОНТОЛОГИЧЕСКОГО ПРОФИЛЯ ДЛЯ СИСТЕМ ПЕРСОНАЛИЗАЦИИ КОНТЕНТА**

*Данная статья посвящена разработке и апробации методики построения онтологического профиля, предназначенного для использования в системах персонализации контента. В работе детально описана модульная архитектура веб-системы персонализации, иллюстрирующая методы и алгоритмы обработки и анализа текста на каждом этапе, а также представлен пошаговый алгоритм создания онтологии. Методика включает первичную обработку данных: извлечение ключевых слов и словосочетаний, их иерархическую кластеризацию для выявления семантической*