

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ COMPUTER SIMULATION

УДК 004.8

DOI: 10.18413/2518-1092-2022-7-2-0-7

Самигулин Т.Р.
Смирнов И.З.
Лаушкина А.А.

**ОПРЕДЕЛЕНИЕ МАРКЕРОВ АГРЕССИВНОГО
ПОВЕДЕНИЯ ЧЕЛОВЕКА НА ОСНОВЕ АНАЛИЗА
АУДИО И ТЕКСТОВОГО КАНАЛОВ**

Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики», Кронверкский пр., д. 49, г. Санкт-Петербург, 197101, Россия
e-mail: timursamigulin98@gmail.com, Ivany1@mail.ru, nastasjalausckina@mail.ru

Аннотация

Взаимодействие участников цифровых платформ требует оперативной реакции на действия друг с другом. Важно понимать, что подобное взаимодействие носит многогранный характер, при котором, наряду с передачей информации, необходимо получать информацию о том, как изменяется состояние собеседника, а также особенности его текущего эмоционального состояния и личностных характеристик. В данной работе описана реализация методов извлечения маркеров агрессивного поведения человека по аудио и текстовому каналам. Также представлен подход к объединению вербальных и невербальных признаков речи для детектирования агрессивного поведения человека.

Сделан вывод, что использование аудио и текстового каналов передачи информации по отдельности является недостаточным для детектирования агрессивного поведения.

Ключевые слова: агрессивное поведение; вербальный анализ; микроэкспрессии; транскрибация; анализ текста; естественный язык; машинное обучение

Для цитирования: Самигулин Т.Р., Смирнов И.З., Лаушкина А.А. Определение маркеров агрессивного поведения человека на основе анализа аудио и текстового каналов // Научный результат. Информационные технологии. – Т.7, №2, 2022. – С. 55-57. DOI: 10.18413/2518-1092-2022-7-2-0-7

Samigulin T.R.
Smirnov I.Z.
Laushkina A.A.

**DETERMINATION OF MARKERS OF AGGRESSIVE BEHAVIOR
BASED ON THE ANALYSIS OF AUDIO AND TEXT CHANNELS**

Saint Petersburg National Research University of Information Technologies, Mechanics and Optics,
49 Kronverkskiy prospekt, St. Petersburg, 197101, Russia
e-mail: timursamigulin98@gmail.com, Ivany1@mail.ru, nastasjalausckina@mail.ru

Abstract

The interaction of participants in digital platforms requires a prompt response to actions from each other. At the same time, such interaction is multifaceted, in which, along with the transfer of information, knowledge about the state of the interlocutor, his current emotional state and personal characteristics is necessary. As part of a pilot study at ITMO University, together with psychologists from St. Petersburg State University, a list of markers of aggressive behavior was determined. And within the framework of this work, the implementation of methods for extracting markers of aggressive human behavior through audio and text channels is described. This paper also describes the approach of combining verbal and non-verbal features in speech. From the results obtained in the course of applying methods for extracting markers of aggressive behavior, we can conclude that the use of audio and text channels of information separately is not enough to detect aggressive behavior.

Keywords: aggressive behavior; verbal analysis; microexpressions; speech recognition; text analysis; natural language; machine learning

For citation: Samigulin T.R., Smirnov I.Z., Laushkina A.A. Determination of markers of aggressive behavior based on the analysis of audio and text channels // Research result. Information technologies. – Т.7, №2, 2022. – P. 55-57. DOI: 10.18413/2518-1092-2022-7-2-0-7

ВВЕДЕНИЕ

Количество информации непрерывно увеличивается, в связи с этим, от переизбытка информации возникает информационный шум, который негативно сказывается на состоянии человека. Так, перегрузка информацией искажает как эмоциональные, так и интеллектуальные способности человека, что сказывается на способности к эмпатии и принятию обдуманных решений. В частности, ухудшение эмоциональных и интеллектуальных способностей проявляется в более частом проявлении агрессии [1]. В работе [2] А.А. Реан под агрессией понимается любое намеренное действие, направленное на причинение ущерба другому живому существу. В данной работе под агрессией понимается враждебная реакция человека на некомфортную для него ситуацию.

Некомфортной ситуацией для человека, например, является увеличение информационного шума. Такая ситуация, в частности, возникает в случае, когда компании используют телефонные звонки как средство информирования пользователей о своем продукте. В качестве примера можно привести банки, которые предлагают пользователю открыть новый продукт, или мобильных операторов, информирующих пользователя о новом “выгодном” для него тарифе.

Обычно для подобных задач у компаний есть call-центры, которые обзванивают клиентов и предлагают свой продукт. У работников call-центров есть набор стратегий того, как необходимо вести общение с клиентом в зависимости от его ответов. Недостатком данных стратегий является то, что они ориентируются только на ответы клиента, при том, что речь можно представить в виде вербальных и невербальных сигналов. В таком случае слова отнесутся к вербальным сигналам, а звуковая волна, создаваемая голосом, к невербальным [3]. При учитывании только ответа клиента внимание переходит только на вербальные признаки.

Учитывание невербальных признаков позволяет оценить психоэмоциональное состояние человека и таким образом детектировать агрессивное поведение. Игнорирование данного аспекта создает для компании риски, такие как интерпретация и перенос негативных эмоций пользователя на бренд, что несет имиджевые риски для компании. Таким образом анализ невербальных признаков дает возможность подобрать более подходящую стратегию общения с клиентом. В связи с этим, задача выделения маркеров агрессивного поведения из аудио и текстового каналов речи является актуальной задачей.

Работа с аудиоканалом позволяет получить просодическую информацию (частотные, амплитудные, динамические, темпоральные характеристики [4]) о речи, которая широко применяется в задачах анализа психоэмоционального состояния человека [5]. В частности, аудиоканал широко применяется при решении задачи детекции эмоций [6,7].

Также в дополнение к аудио каналу можно анализировать текстовый канал, используя транскрипцию аудио в текст. Таким образом мы сможем учесть морфологические и семантические признаки такие как: повелительное наклонение, уменьшительно-ласкательные суффиксы, нецензурная лексика, токсичная речь, оскорбления и угрозы.

Данная работа направлена на соотнесение параметров, получаемых при работе с речью, представленной в виде текста и аудиосигнала с целью исследования взаимодействия между вербальными и невербальными признаками в задаче детектирования агрессивного поведения.

ОСНОВНАЯ ЧАСТЬ

В рамках пилотажного исследования в качестве гипотезы было принято наличие зависимости между уровнем агрессии личности и количественными и качественными характеристиками вербальных и невербальных маркеров агрессии. Был выполнен сбор данных для определения поведенческих маркеров агрессивного поведения. Данные представляют из себя

набор видеозаписей психологических тестирований студентов, возраст которых варьируется в пределах 20-25 лет.

Добровольцам предоставлялся контент, направленный на активацию агрессии и нейтрального состояния, после чего требовалось выполнить определенные задания: прохождение теста Розенцвейга и описание своих действий при нахождении в мысленной ситуации «Игра». В первом случае участнику предоставлялся стимульный материал (изображения), которые описывают фрустрационную ситуацию и содержат заполненный квадрат со словами одного участника ситуации и пустой для слов, которые предполагается заполнить испытуемому [8]. В данном случае ответы чаще всего были непродолжительными, но разнообразными. Описание своих действий при прохождении «Игры» было более развернутым и аргументированным, что позволяло получить более продолжительные записи речи.

Для анализа записей каждый фрагмент описывался набором невербальных и вербальных маркеров поведения. Далее описаны подходы, применяемые для их извлечения.

ИЗВЛЕЧЕНИЕ МАРКЕРОВ АГРЕССИВНОГО ПОВЕДЕНИЯ ИЗ РЕЧИ

К невербальным признакам в эксперименте отнесены параметры речи, применяемые для описания аудиоканала. В данном случае были выбраны следующие признаки:

- Резкое повышение/понижение громкости речи;
- Темп речи;
- Эмоциональный окрас речи: злость, отвращение, спокойствие, радость, страх, удивление, печаль.

Видеозаписи собранного набора данных представляли из себя записи людей, находящихся в разных условиях и использующих различные устройства звукозаписи. Поэтому необходима была предварительная обработка данных, которая заключалась в удалении шумов и нормализации громкости. Аналогичный подход к предобработке записей применяется в работе [9].

Для расчета резкого изменения громкости был применен эвристический подход, который заключается в расчете угла наклона между соседними отсчетами амплитудной огибающей и абсолютным расстоянием до максимального значения. Таким образом, изменение громкости было посчитано резким, если амплитудное значение отсчета близко к максимальному значению по всему речевому сигналу и достигается быстрее, чем происходят переходы к другим значениям внутри сигнала. За темп речи было принято число произносимых слов в минуту.

Для извлечения класса эмоций была обучена модель сверточной нейронной сети с полносвязным классификатором. Модель принимает на вход аудио-сигнал в виде мелкепестральных коэффициентов (MFCC – краткосрочный спектр мощности звука, расчет мощности основан на особенностях восприятия звука органами слуха) [4]. Для обучения были объединены следующие наборы данных CaFE [10], Emo-DB [11], RAVDESS [12], SAVEE [13] для того, чтобы не было переобучения на языке одного из наборов данных [14]. Полученный набор данных включают в себя короткие аудиозаписи речи, каждая из которых отнесена к одному из семи классов: злость, скука, тревога, радость, печаль, отвращение, нейтральное состояние (отсутствие эмоции). Классификатор принимает на вход фрагменты длительностью 3 секунды в случае, если фрагмент имеет меньшую длительность, происходит заполнение фрагмента нулями до необходимого размера.

В данном эксперименте классификатор использовался для получения вероятностных оценок каждого класса эмоции на заданном промежутке времени вместо получения одного значения – класса. Далее значения усреднялись в границах одного задания. Таким образом можно наблюдать динамику палитры эмоций между заданиями. На рисунке 1 представлены график эмоционального состояния одного из участников при прохождении теста Розенцвейга.

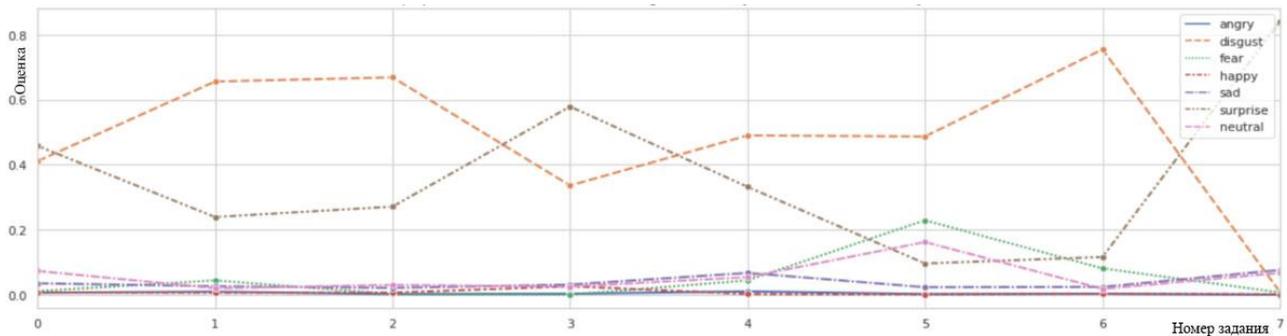


Рис. 1. График эмоционального состояния участника эксперимента при прохождении теста Розенцвейга. Вероятностные оценки эмоций представлены для 8 заданий в рамках теста
Fig. 1. Graph of the emotional state of the participant of the experiment when passing the Rosenzweig test. Probabilistic estimates of emotions are presented for 8 tasks within the test

Для перехода к текстовому каналу следовало решить задачу транскрибации (расшифровка информации из аудио или видео в текстовую форму) речи. Был выбран набор инструментов Vosk, так как модель, предоставляемая для транскрибации, поддерживает более 20 языков и обучена на большой базе наборов, а также позволяет работать с записями без ограничения по длительности.

ИЗВЛЕЧЕНИЕ МАРКЕРОВ АГРЕССИИ ИЗ ТЕКСТА

В дополнение к невербальным признакам из аудио канала были сформулированы следующие вербальные признаки, которые позволяют детектировать агрессию по тексту [15]:

- наличие нецензурной лексики и жаргона в тексте;
- наличие повелительного наклонения;
- наличие уменьшительно-ласкательных\уменьшительно-унизительных конструкций;
- наличие оскорблений;
- наличие угроз;
- наличие токсичности.

Для проверки наличия нецензурной лексики и жаргона был собран словарь из 140 тысяч различных слов, включающий в себя все возможные морфологические модификации для охвата как можно большего количества вариантов слов, относящихся к нецензурной лексике. Каждое слово входящего текста проверяется по данному словарю и подсчитывается количество нецензурной лексики в тексте.

Находится слово в повелительном наклонение или нет определяется с помощью библиотеки `rumorphy2`, позволяющей производить морфологический разбор слова [16].

Чтобы определить уменьшительно-ласкательные и уменьшительно-унизительные конструкций, необходимо:

–обработать слова помощью стеммера. Стеммер позволяет выделить основу слова, то есть убрать окончания;

–проверить наличие уменьшительно-ласкательных и уменьшительно-унизительных суффиксов: -ик, -ек, -к, -ец, -иц, -оск, -ечк, -оньк, -еньк, -ышк, -инш, -ушк, -юшк [17].

Для детекции оскорблений и угроз были дообучены модели нейронной сети на основе архитектуры `RuBertTiny` [18]. Модели дообучались для задачи бинарной классификации, то есть текст относился к одному из двух классов.

Для детекции угроз:

- тесты, содержащие угрозу;
- тексты, не содержащие угроз.

Для детекции оскорблений:

- тексты, содержащие оскорбления;
- тексты, не содержащие оскорбления.

Для обучения использовался набор данных, содержащий 248290 комментария из социальной сети “Одноклассники”. Из них:

- 203685 нормальных сообщений, не содержащих ни оскорблений, ни угроз;
- 36836 сообщений, содержащих оскорбления;
- 11777 сообщений с угрозами.

Также была дообучена модель RuBertTiny для детекции токсичности в тексте [19]. К токсичному тексту относятся любые дискриминирующие и оскорбительные сообщения по отношению к человеку, группе лиц, организации или любому другому одушевленному или неодушевленному субъекту. Для этого было объединено несколько различных наборов данных и на них обучена модель классификации [20].

По итогам обучения моделей были получены следующие результаты, представленные в таблице 1.

Таблица 1

Оценка метрик

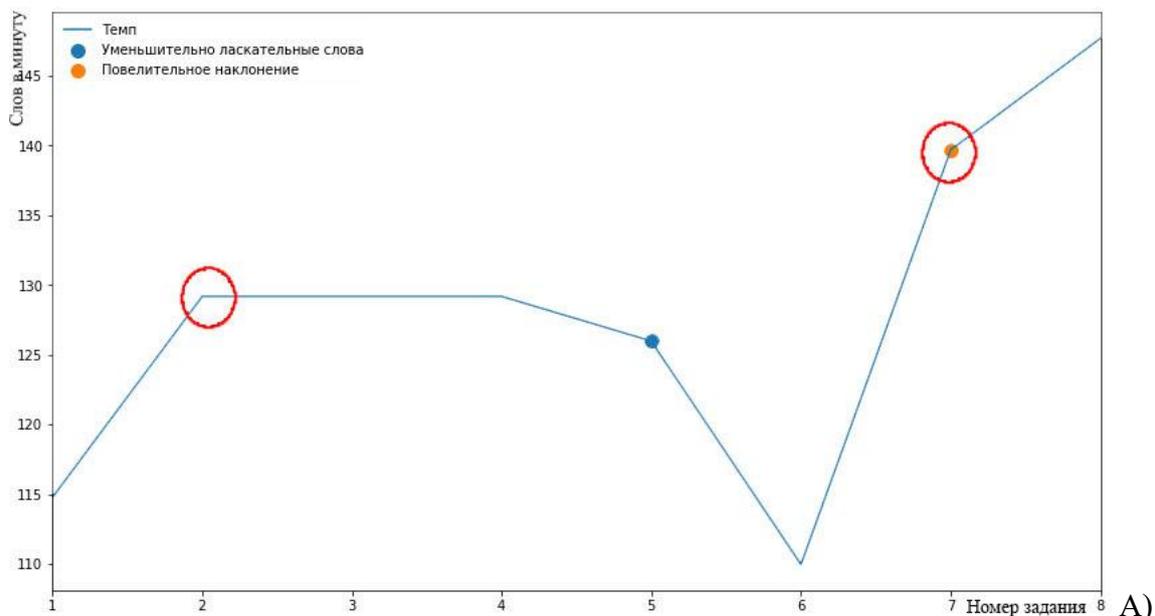
Table 1

Evaluation of metrics

Метрика\Модель	Детекция угроз	Детекция оскорблений	Детекция токсичности
Precision	0.953	0.95	0.945
Recall	0.959	0.948	0.95
F1 score	0.96	0.949	0.948

РЕЗУЛЬТАТ РАБОТЫ АЛГОРИТМА

На рисунке 2 изображены результаты извлечения меток из видеозаписей прохождения теста Розенцвейга двух участников эксперимента.



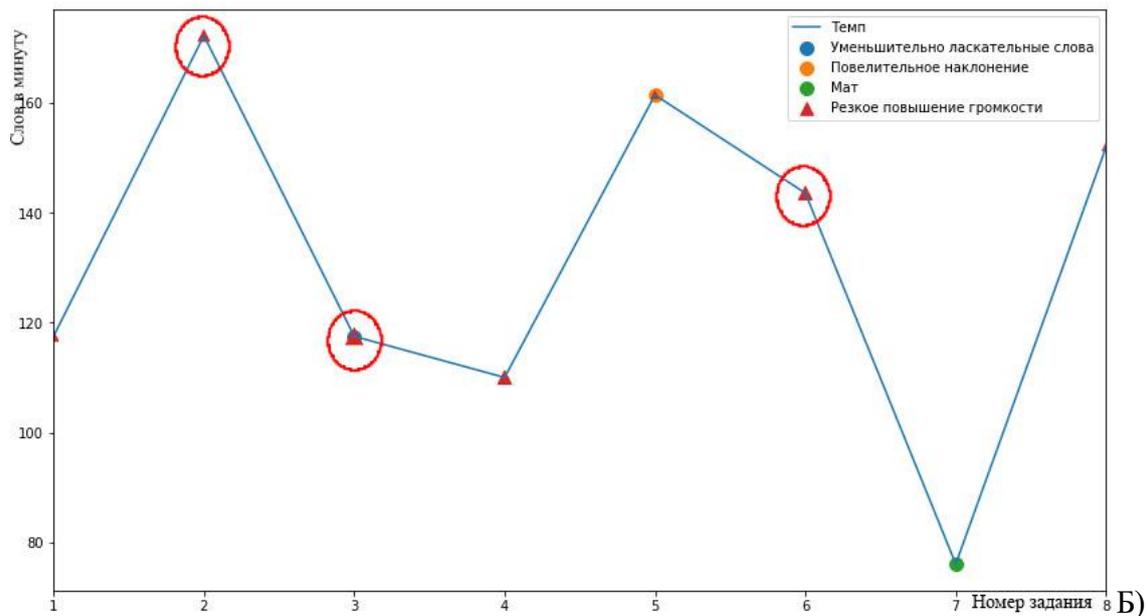


Рис. 2. График отображающий изменение значения темпа, а так же наличия вербальных и невербальных меток во время ответа участника на задания теста Розенцвейга. Размеры обозначений (круг в случае вербальных признаков, треугольник для невербального) зависят от количества появления маркеров. Окружностью красного цвета выделены фрагменты, в которых поведение участника было агрессивным А) Первый участник Б) Второй участник

Fig. 2. Graph showing the change in the tempo value, as well as the presence of verbal and nonverbal marks during the participant's response to the tasks of the Rosenzweig test. The size of the symbols (circle in the case of verbal signs, triangle for nonverbal) depends on the number of markers appearing. The red circle highlights the fragments in which the participant's behavior was aggressive А) The first participant Б) The second participant

Можно наблюдать, что параметры меняются в зависимости от контента. Также анализировать необходимо не сами статичные значения маркеров, а характер их изменения. Это позволяет менять стратегию ведения диалога с человеком в зависимости от того, в какой момент его поведение становится агрессивным.

ЗАКЛЮЧЕНИЕ

В работе предложен подход для объединения информации, получаемой из вербальной и невербальной составляющей речи для детектирования агрессивного поведения. Были описаны методы для извлечения маркеров агрессивного поведения. На основании результатов, полученных в процессе применения методов для извлечения маркеров агрессивного поведения, можно сделать вывод, что использование каналов информации по отдельности недостаточно для детектирования агрессивного поведения. В дальнейшем можно дополнить текущее исследование видео каналом и оценить корреляцию между всеми тремя каналами коммуникации. Так же полученные маркеры позволяют создать интерпретируемую автоматизированную модель, детектирующую агрессивное поведение.

Список литературы

1. Федорук А.В., Зиматкина Т.И. О проблеме психогенного стресса у студенческой молодежи в связи с информационной перегрузкой // УО "Гродненский государственный медицинский университет" г. Гродно, Республика Беларусь.
2. Rean A.A. Prevention of aggression and antisocial behaviour in adolescents // National Psychological Journal. – 2018. – № 2. – P. 3-12.
3. Lefter I., Rothkrantz L.J.M., Burghouts G.J. Aggression Detection in Speech Using Sensor and Semantic Information. – 2012. – P. 665–672.

4. Goupil L. et al. Listeners' perceptions of the certainty and honesty of a speaker are associated with a common prosodic signature // *Nature Communications*. – 2021. – Vol. 12. – № 1. – P. 861.
5. Giddens C.L. et al. Vocal Indices of Stress: A Review // *Journal of Voice*. – 2013. – Vol. 27. – № 3. – P. 390.e21-390.e29.
6. Atassi H., Smekal Z., Esposito A. Emotion recognition from spontaneous Slavic speech // 2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom). IEEE, 2012. – P. 389–394.
7. Speech Emotion Recognition For Kazakh And Russian Languages // *Applied Mathematics & Information Sciences*. – 2020. – Vol. 14. – № 1. – P. 65–68.
8. Тарабрина Н.В. Экспериментально-психологическая методика изучения фрустрационных реакций // *Методические рекомендации*. 1984.
9. Marcolla F., de Santiago R., Dazzi R. Novel Lie Speech Classification by using Voice Stress // *Proceedings of the 12th International Conference on Agents and Artificial Intelligence*. SCITEPRESS – Science and Technology Publications, 2020. – P. 742–749.
10. Gournay P., Lahaie O., Lefebvre R. A canadian french emotional speech dataset // *Proceedings of the 9th ACM Multimedia Systems Conference*. New York, NY, USA: ACM, 2018. – P. 399–402.
11. Burkhardt F. et al. A Database of German Emotional Speech // *Interspeech*. 2005. P. 1517–1520.
12. Luna-Jiménez C. et al. A Proposal for Multimodal Emotion Recognition Using Aural Transformers and Action Units on RAVDESS Dataset // *Applied Sciences*. 2021. – Vol. 12. – № 1. – P. 327.
13. S. Haq P.J.B.J. Multimodal Emotion Recognition / ed. Wang W. IGI Global, 2010. 398–423 p.
14. Kossaifi J. et al. SEWA DB: A Rich Database for Audio-Visual Emotion and Sentiment Research in the Wild. 2019.
15. Девяткин Д.А. и др. Интеллектуальный анализ проявлений вербальной агрессивности в текстах сетевых сообществ // *Искусственный интеллект и принятие решений*. – 2014. – №. 2. – С. 27-41.
16. Korobov M. Morphological analyzer and generator for Russian and Ukrainian languages // *International conference on analysis of images, social networks and texts*. – Springer, Cham, 2015. – С. 320-332.
17. Евсеева И.В. К вопросу о важности уменьшительно-ласкательных суффиксов в русском языке // *Взаимодействие языка и культуры в коммуникации и тексте: сборник научных статей*. – 2012. – Т. 3. – № 3. – С. 73-77.
18. Tenney I., Das D., Pavlick E. BERT rediscovers the classical NLP pipeline // *arXiv preprint arXiv:1905.05950*. – 2019.
19. Saeed H. H., Shahzad K., Kamiran F. Overlapping toxic sentiment classification using deep neural architectures // 2018 IEEE international conference on data mining workshops (ICDMW). – IEEE, 2018. – P. 1361-1366.
20. Maslej-Krešňáková V. et al. Comparison of deep learning models and various text pre-processing techniques for the toxic comments classification // *Applied Sciences*. – 2020. – Т. 10. – № 23. – P. 8631.

References

1. Fedoruk A.V. Zimatkina.T.I. On the problem of psychogenic stress among students in connection with information overload // *UO "Grodno State Medical University" Grodno, Republic of Belarus*.
2. Rean A.A. Prevention of aggression and antisocial behaviour in adolescents // *National Psychological Journal*. – 2018. – № 2. – P. 3-12.
3. Lefter I., Rothkrantz L.J.M., Burghouts G.J. Aggression Detection in Speech Using Sensor and Semantic Information. – 2012. – P. 665–672.
4. Goupil L. et al. Listeners' perceptions of the certainty and honesty of a speaker are associated with a common prosodic signature // *Nature Communications*. – 2021. – Vol. 12. – № 1. – P. 861.
5. Giddens C.L. et al. Vocal Indices of Stress: A Review // *Journal of Voice*. – 2013. – Vol. 27. – № 3. – P. 390.e21-390.e29.
6. Atassi H., Smekal Z., Esposito A. Emotion recognition from spontaneous Slavic speech // 2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom). IEEE, 2012. – P. 389–394.
7. Speech Emotion Recognition For Kazakh And Russian Languages // *Applied Mathematics & Information Sciences*. – 2020. – Vol. 14. – № 1. – P. 65–68.
8. Tarabrina N.V. Experimental psychological methodology for studying frustration reactions // *Methodological recommendations*. 1984.
9. Marcolla F., de Santiago R., Dazzi R. Novel Lie Speech Classification by using Voice Stress // *Proceedings of the 12th International Conference on Agents and Artificial Intelligence*. SCITEPRESS – Science and Technology Publications, 2020. – P. 742–749.

10. Gournay P., Lahaie O., Lefebvre R. A canadian french emotional speech dataset // Proceedings of the 9th ACM Multimedia Systems Conference. New York, NY, USA: ACM, 2018. – P. 399–402.
11. Burkhardt F. et al. A Database of German Emotional Speech // Interspeech. 2005. P. 1517–1520.
12. Luna-Jiménez C. et al. A Proposal for Multimodal Emotion Recognition Using Aural Transformers and Action Units on RAVDESS Dataset // Applied Sciences. 2021. – Vol. 12. – № 1. – P. 327.
13. S. Haq P.J.B.J. Multimodal Emotion Recognition / ed. Wang W. IGI Global, 2010. 398–423 p.
14. Kossaifi J. et al. SEWA DB: A Rich Database for Audio-Visual Emotion and Sentiment Research in the Wild. 2019.
15. Devyatkin D. A. et al. Intellectual analysis of verbal aggressiveness in the texts of online communities // Artificial intelligence and decision-making. – 2014. – №. 2. – Pp. 27-41.
16. Korobov M. Morphological analyzer and generator for Russian and Ukrainian languages // International conference on analysis of images, social networks and texts. – Springer, Cham, 2015. – С. 320-332.
17. Evseeva I. V. On the importance of diminutive suffixes in the Russian language // The interaction of language and culture in communication and text: a collection of scientific articles. – 2012. – Vol. 3. – No. 3. – pp. 73-77.
18. Tenney I., Das D., Pavlick E. BERT rediscovers the classical NLP pipeline // arXiv preprint arXiv:1905.05950. – 2019.
19. Saeed H. H., Shahzad K., Kamiran F. Overlapping toxic sentiment classification using deep neural architectures // 2018 IEEE international conference on data mining workshops (ICDMW). – IEEE, 2018. – P. 1361-1366.
20. Maslej-Krešňáková V. et al. Comparison of deep learning models and various text pre-processing techniques for the toxic comments classification // Applied Sciences. – 2020. – Т. 10. – № 23. – С. 8631.

Самигуллин Тимур Русланович, студент 2-го курса магистратуры, инженер Национального центра когнитивных разработок

Смирнов Иван Захарович, студент 1-го курса магистратуры, инженер Национального центра когнитивных разработок

Лаушкина Анастасия Александровна, аспирант, инженер Национального центра когнитивных разработок

Samigullin Timur Ruslanovich, 2nd year Master's student, engineer at National Center for Cognitive Research

Smirnov Ivan Zakharovich, 1st year Master's student, engineer at National Center for Cognitive Research

Laushkina Anastasia Alexandrovna, postgraduate student, engineer, National Center for Cognitive Development