

Разработка метода распознавания эмоций по речевому сигналу

Д. А. Кравчук¹ ✉

¹ Институт нанотехнологий, электроники и приборостроения Южного федерального университета
ул. Шевченко, д. 2/Е, г. Таганрог 347922, Российская Федерация

✉ e-mail: kravchukda@sfedu.ru

Резюме

Цель исследования – автоматическое распознавание эмоций говорящего, основанное на обработке звуко-записей, предназначенных для использования в системах сигнализации при работе с операторами локомотивных бригад и диспетчерских служб.

Методы. Распознавание эмоций человека в последние годы является быстроразвивающейся областью исследований. Особенности речевого тракта, например мощность звука, формантные частоты, применяются для обнаружения некоторых эмоций с хорошей точностью. Использован метод определения энергии сигнала с выделением доминирующей частоты. В работе разработан программный код, на основе которого приведен анализ четырех эмоций: гнев, радость, страх и спокойствие. Самый важный и сложный шаг – это определение признаков наиболее подходящих для различения эмоций и наличие баз данных. Сбор баз данных является сложной задачей, требующей проявления искренности эмоций. Зачастую сбор базы данных происходит в искусственной среде, и речь может звучать постановочно, для исключения таких проблем необходимо использовать записи колл-центров.

Результаты. Получены и обработаны записи базовых эмоциональных состояний, такие как гнев, радость, грусть, страх и удивление, которые являются наиболее частым случаем исследования. Разработанный программный код позволяет приблизиться в автоматическому определению эмоций по речевому сигналу. Для анализа речевых записей в выборках использовались показатели энергии сигнала и выделения доминирующей частоты.

Заключение. Реализуемый метод контроля эмоционального состояния человека оператора по речевому сигналу находит широкое применение в профилактике и улучшении показателей психофизиологической профпригодности работников локомотивных бригад, сохранения их профессионального здоровья. Отчетливые различия наблюдаются в характеристиках всех видов эмоций.

Ключевые слова: эмоции; речевой сигнал; спектр сигнала; метод распознавания.

Конфликт интересов: Авторы декларируют отсутствие конфликта интересов, связанных с публикацией данной статьи.

Для цитирования: Кравчук Д. А. Разработка метода распознавания эмоций по речевому сигналу // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. 2024. Т. 14, № 2. С. 72–80. <https://doi.org/10.21869/2223-1536-2023-14-2-72-80>

Поступила в редакцию 18.04.2024

Подписана в печать 11.05.2024

Опубликована 28.06.2024

Development of a method for recognizing emotions from a speech signal

Denis A. Kravchuk

¹ Institute of Nanotechnology, Electronics and Instrumentation of the Southern Federal University
2/E Shevchenko Str., Taganrog 347922, Russian Federation

✉ e-mail: kravchukda@sfnu.ru

Abstract

The purpose of research is automatic recognition of the speaker's emotions, based on the processing of sound recordings intended for use in alarm systems when working with operators of locomotive crews and dispatch services.

Methods. Human emotion recognition has been a rapidly developing area of research in recent years. Features of the vocal tract, such as sound power, formant frequencies, are used to detect certain emotions with good accuracy. A method was used to determine the signal energy by highlighting the dominant frequency. The work has developed a program code, on the basis of which an analysis of four emotions is given - anger, joy, fear and calm. The most important and difficult step is to determine the features most suitable for distinguishing emotions and the availability of databases. Collecting databases is a complex task requiring the manifestation of sincerity of emotions. Often, the collection of a database takes place in an artificial environment and the speech may sound staged; to eliminate such problems, it is necessary to use call center recordings.

Results. Recordings of basic emotional states, such as anger, joy, sadness, fear and surprise, which are the most common case of the study, were obtained and processed. The developed software code allows us to get closer to automatically determining emotions from a speech signal. To analyze speech recordings in samples, indicators of signal energy and identification of the dominant frequency were used.

Conclusion. The implemented method of monitoring the emotional state of a human operator using a speech signal is widely used in the prevention and improvement of indicators of the psychophysiological professional suitability of locomotive crew workers and the preservation of their professional health. Distinct differences are observed in the characteristics of all types of emotions.

Keywords: emotions; speech signal; signal spectrum; recognition method.

Conflict of interest: The authors declares no conflict of interest related to the publication of this article.

For citation: Kravchuk D. A. Development of a method for recognizing emotions from a speech signal. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. Seriya: Upravlenie, vychislitel'naja tekhnika, informatika. Meditsinskoe priborostroenie* = Proceedings of the Southwest State University. Series: Control, Computer Engineering, Information Science. Medical Instruments Engineering. 2024;14(2):72–80. (In Russ.) <https://doi.org/10.21869/2223-1536-2024-14-2-72-80>

Received 18.04.2024

Accepted 11.05.2024

Published 28.06.2024

Введение

В последние годы были проведены различные исследования, чтобы выявить особенности, которые будут полезны при распознавании эмоций. Ряд исходных особенностей, особенностей речевого тракта, например мощность звука, формантные частоты, кепстральные коэффициенты, используются для обнаружения некоторых эмоций с хорошей точностью. Также анализируются параметры качества голоса, такие как коэффициент гармонического шума (HNR), мерцание.

Анализ речевого сигнала необходим для определения эмоций, использования в качестве биометрии [1], даже для изучения частоты сердечных сокращений говорящего [2], обнаружения усталости.

Основные эмоции подразделяются на печаль, страх, счастье, отвращение, удивление и гнев [3]. В литературе по анализу эмоций предлагаются две предпочтительные характеристики, известные как статистические и временные характеристики [4; 5].

Система распознавания речевых эмоций может быть структурирована путем анализа функций, которые эффективно выявляют каждую эмоцию в речевых сигналах [6].

Набор методов классификации речевых эмоций основан на скрытой модели Маркова [7], модели гауссовой смеси (GMM) [8], самоорганизующейся карте [9] и нейронной сети [10]. Классификатор разложения сингулярных значений (SVD) используется в [8], тогда как в [7] для классификации эмоций предлагается модель ансамблевой программной регрессии, также предлагается глубокая сеть убеждений, основанная на функциях высокого и низкого уровня [6]. Существует метод, основанный на машине опорных векторов и нейронных сетях, для классификации

пяти эмоций, таких как гнев, удивление, нейтральность, счастье и печаль [11]. Предложен в [12] метод, использующий несколько подклассификаторов для установления семи типов эмоций.

Материалы и методы

Каждый человек обладает уникальным голосом. Однако можно безошибочно отличить мужской голос от женского и наоборот, даже не видя собеседника, это происходит из-за различного биологического строения голосового аппарата у женщин и мужчин. Несмотря на это, голоса мальчиков и девочек в детстве похожи между собой.

Однако из-за большого диапазона размеров связок встречаются как мужчины с очень высоким голосом, так и женщины с крайне низким. Другим важным биологическим отличием, оказывающим влияние на голос, является длина речевого тракта, которая у мужчин больше, что еще сильнее снижает тон голоса, помимо размера голосовых связок.

На основе этих различий были записаны два голоса – мужской и женский. Была выбрана фраза Мартышки из мультфильма «38 попугаев»: «Все время думать одну и ту же мысль нельзя! Это очень вредно! От этого можно соскучиться и заболеть».

В работе [12] авторы выделяют четыре наиболее часто используемых эмоциональных состояний: нормальное, гнев, горе и страх. В речевом сигнале содержатся признаки, которые не только характеризуют индивидуальные особенности говорящего, но и его психофизическое состояние [13]. Отечественными авторами неоднократно проводились работы, посвященные связи эмоционального состояния человека и его речи [14].

Для распознавания эмоций по голосовому сигналу нужно количественно охарактеризовать сигнал [15] и выделить

важные параметры, которые отвечают за эмоциональное состояние [16]. К признакам, несущим информацию о психофизическом состоянии человека, относят [17]:

- громкость (интенсивность) речи;
- изменение частоты основного тона;
- изменение темпа речи;
- смещение энергетического спектра речи;
- спектральные характеристики (интенсивность, высота);
- временные характеристики (длительность слов и фраз, паузы между ними и др.).

Результаты и их обсуждение

Используя микрофон и программу «Запись голоса» из стандартных программ Windows, был записан текст с различными эмоциональными состояниями. В записи участвовали пятеро мужчин и женщин, каждый испытуемый записывал по десять фраз для последующей обработки. После этого записи были конвертированы в формат wav.

После был написан код для обработки аудиосигнала в Matlab, для чего

сначала был присвоен переменной путь к записанному аудиофайлу, а после, используя функции `audioread` и `timetable`, произведена обработка сигнала, начальный этап представлен ниже (рис. 1).

```
clc;
clear all;
Filename = '\\tsclient\C\filename.wav';
[data,fs] = audioread (Filename);
Sound= timetable(seconds((0:length(data)-1)/fs),data);
```

Рис. 1. Начальный этап обработки записи речи испытуемого

Fig. 1. Initial stage of record processing the subject's speech

С помощью встроенного модуля обработки сигнала MATLAB 2017b (signal analyzer) представлен спектр аудиосигнала. Полученные графики четырех различных эмоциональных состояний представлены в спокойном состоянии (рис. 2), в грустном состоянии (рис. 3), в гневе (рис. 4) и в состоянии страха (рис. 5). При этом верхние графики представляют собой запись речевого сигнала, на нижних графиках представлены результаты обработки в виде спектральной плотности мощности.

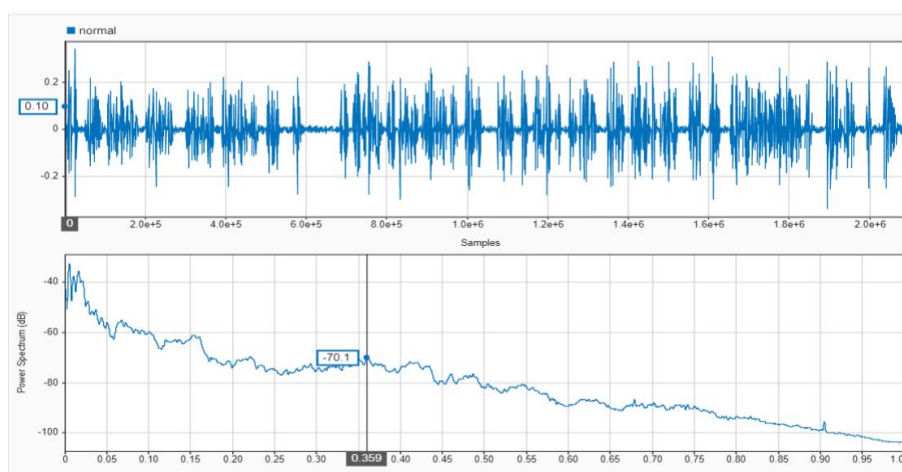


Рис. 2. Результат обработки сигнала фрагмента речи «Спокойное состояние говорящего»: сверху – запись сигнала; внизу – спектральная плотность мощности сигнала

Fig. 2. The result of signal processing in a quiet state: at the top – signal recording; at the bottom – signal power spectral density

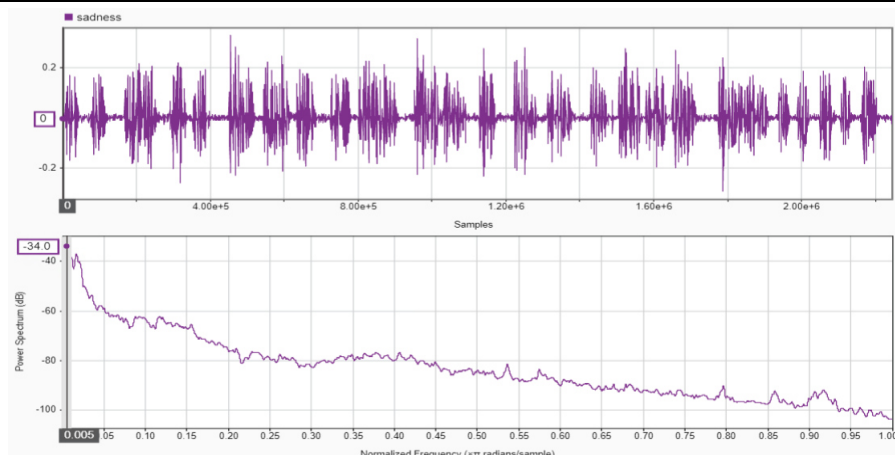


Рис. 3. Результат обработки сигнала в грустном состоянии: сверху – запись сигнала; внизу спектральная плотность мощности сигнала

Fig. 3. The result of signal processing in a sad state: at the top – signal recording; at the bottom – signal power spectral density

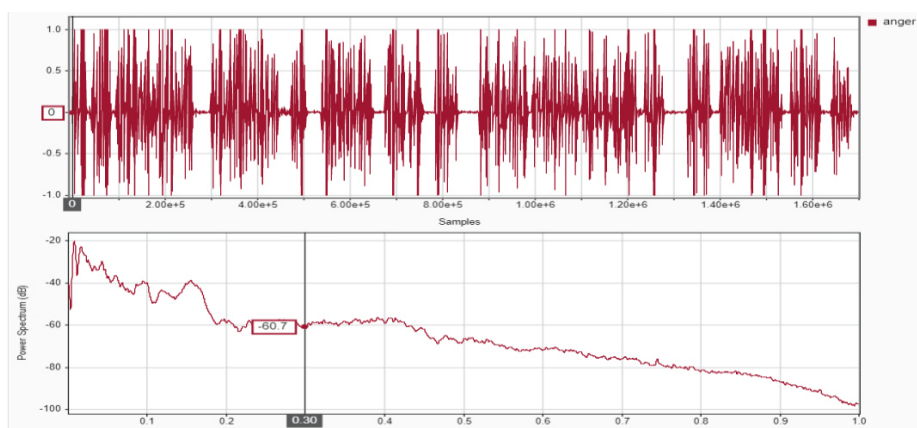


Рис. 4. Результат обработки сигнала в гневе: сверху – запись сигнала; внизу – спектральная плотность мощности сигнала

Fig. 4. Result of signal processing in anger: at the top – signal recording; at the bottom – signal power spectral density)

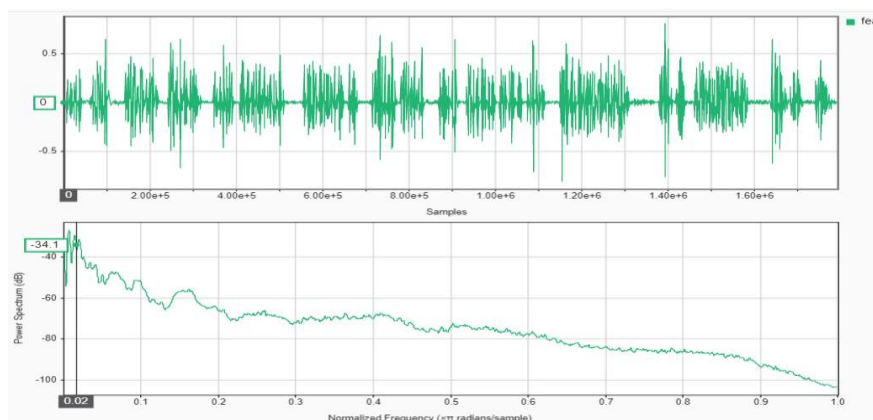


Рис. 5. Результат обработки сигнала в состоянии страха: сверху – запись сигнала; внизу – спектральная плотность мощности сигнала;

Fig. 5. The result of signal processing in a state of fear: at the top – signal recording; at the bottom – signal power spectral density)

Выводы

Исходя из полученных данных можно сделать вывод о том, что в речевом сигнале отражаются изменения, происходящие в эмоциональном спектре:

1. *Горе* или *печаль*. Заметное снижение темпа речи на 15–45% и частоты основного тона на 5–15%. В результате обработки 10 записей сигнала от каждого испытуемого установлено уменьшение интенсивности сигнала и мощности спектра в высоких частотах.

2. *Гнев*. Происходит увеличение темпа речи, повышение частоты основного тона на 140–210%, наблюдается повышение мощности высоких частот в спектре на 8–12 дБ.

3. *Страх*. Характеризуется увеличением темпа речи, низкой интенсивностью

сигнала, повышенной частотой основного тона и незначительной мощностью спектра в высоких частотах. Также речевой сигнал может быть дополнен тремором в голосе и заиканиями.

4. *Радость* – это состояние характеризуется увеличением темпа речи на 15–25% за счет сокращения пауз между отдельными словами и заметным повышением частоты основного тона. Интенсивность речевого сигнала возрастает на 3,5–6 дБ.

Проведенные исследования являются продолжением работ в области определения психоэмоционального состояния человека. Работа проводится в рамках создания системы контроля психоэмоционального состояния локомотивных бригад.

Список литературы

1. Adjectives grouping in a dimensionality affective clustering model for fuzzy perceptual evaluation / W. Huang, Q. Wu, N. Dey, A. Ashour, S. J. Fong, R. González-Crespo // Int. J. Interact. Multimedia. Artif. Intell. 2020. Vol. 6, N 2. P. 10. <https://doi.org/10.9781/ijimai.2020.05.002>
2. Xusheng Wang, Xing Chen, Congjun Cao. Human emotion recognition by optimally fusing facial expression and speech feature, Signal Processing // Image Communication. 2020. Vol. 84, N 10. P. 115831. <https://doi.org/10.1016/j.image.2020.115831>
3. Akçay M. B. OğuzK. Speech emotion recognition: emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers // Speech. Commun. 2020. N 116. P. 56–76. <https://doi.org/10.1016/j.specom.2019.12.001>
4. Speech Emotion Recognition with Dual-Sequence LSTM Architecture / J. Wang, M. Xue, R. Culhane, E. Diao, J. Ding, V. Tarokh // 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain, 2020. P. 6474–6478. <https://doi.org/10.1109/ICASSP40776.2020.9054629>
5. Recognize basic emotional states in speech by machine learning techniques using mel-frequency cepstral coefficient features / N. Yang, N. Dey, R. S. Sherratt, F. Shi // J. Intell. Fuzzy. Syst. 2020. N 39 (2). P. 1925–1936. <https://doi.org/10.3233/jifs-179963>

6. Daneshfar F., Kabudian S. J., Neekabadi A. Speech emotion recognition using hybrid spectral-prosodic features of speech signal/ glottal waveform, metaheuristic-based dimensionality reduction, and Gaussian elliptical basis function network classifier // *Appl. Acoust.* 2020. N 166. P. 107360. <https://doi.org/10.1016/j.apacoust.2020.107360>
7. Palo H. K., Behera D., Rout B. C. Comparison of classifiers for speech emotion recognition (SER) with discriminative spectral features // *Advances in Intelligent Computing and Communication. Proceedings of ICAC.* Singapore: Springer, 2020. P. 78–85. https://doi.org/10.1007/978-981-15-2774-6_10
8. Speech emotion classification using attention-based lstm / Y. Xie, R. Liang, Z. Liang, C. Huang., C. Zou, B. Schuller // *IEEE/ACM Trans. Audio Speech. Lang. Proc.* 2019. Vol. 27, N 11. P. 1675–1685. <https://doi.org/10.1109/TASLP.2019.2925934>
9. Hassouneh A., Mutawa A. M., Murugappan M. Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods // *InformMed. Unlock.* 2020. N 20. P. 100372. <https://doi.org/10.1016/j.imu.2020.100372>
10. Automatic speech emotion recognition using an optimal combination of features based on EMD-TKEO / L. Kerkeni, Y. Serrestou, K. Raoof, M. Mbarki, M. A. Mahjoub, C. Cleder // *Speech. Commun.* 2019. N 114. P. 22–35. <https://doi.org/10.1016/j.specom.2019.09.002>
11. Uddin Md. Zia, Nilsson E. G., Emotion recognition using speech and neural structured learning to facilitate edge intelligence // *Engineering Applications of Artificial Intelligence.* 2020. Vol. 94. P. 103775. <https://doi.org/10.1016/j.engappai.2020.103775>
12. Excitation Features of Speech for Emotion Recognition Using Neutral Speech as Reference / S. R. Kadiri, P. Gangamohan, S. V. Gangashetty [et al.] // *Circuits. Syst. Signal. Process.* 2020. N 39. P. 4459–4481. <https://doi.org/10.1007/s00034-020-01377-y>
13. Gorshkov Y. G. Visualization of Lung Sounds Based on Multilevel Wavelet Analysis // *Scientific Visualization.* 2022. Vol. 14, N 2. P. 18–26. <https://doi.org/10.26583/sv.14.2.02>
14. Кравчук Д. А. Ультразвуковая система контроля психофизиологического состояния машиниста поезда // *Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение.* 2020. Т. 10, № 1. С. 134–142.
15. Распознавание эмоций по характеристикам речевого сигнала (лингвистический, клинический, информационный аспекты) / Л. П. Прокофьева, И. Л. Пластун, Н. В. Филиппова, Л. Ю. Матвеева, Н. С. Пластун // *Сибирский филологический журнал.* 2021. № 2. С. 325–336.
16. Горшков Ю. Г. Визуализация эмоциональной напряженности человека по речевому сигналу // *Научная визуализация.* 2023. Т. 15, № 2. С. 102–112.

17. Acoustocardiography with Assessment of Emotional Tension from the Voice / Y. G. Gorshkov, A. K. Volkov, N. A. Voinova [et al.] // Biomed. Eng. 2020. N 53. P. 383–387. <https://doi.org/10.1007/s10527-020-09948-8>

References

1. Huang W., Wu Q., Dey N., Ashour A., Fong S.J., González-Crespo R. Adjectives grouping in a dimensionality affective clustering model for fuzzy perceptual evaluation. *Int. J. Interact. Multimedia. Artif. Intell.* 2020;6(2):10. <https://doi.org/10.9781/ijimai.2020.05.002>
2. Xusheng Wang, Xing Chen, Congjun Cao. Human emotion recognition by optimally fusing facial expression and speech feature, *Signal Processing. Image Communication.* 2020;84(10):115831. <https://doi.org/10.1016/j.image.2020.115831>
3. Akçay M. B., Oğuz K. Speech emotion recognition: emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech. Commun.* 2020;(116):56–76. <https://doi.org/10.1016/j.specom.2019.12.001>
4. Wang J., Xue M., Culhane R., Diao E., Ding J., Tarokh V. Speech Emotion Recognition with Dual-Sequence LSTM Architecture. *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona, Spain; 2020. P. 6474–6478. <https://doi.org/10.1109/ICASSP40776.2020.9054629>
5. Yang N., Dey N., Sherratt R. S., Shi F. Recognize basic emotional states in speech by machine learning techniques using mel-frequency cepstral coefficient features. *J. Intell. Fuzzy. Syst.* 2020;(39):1925–1936. <https://doi.org/10.3233/jifs-179963>
6. Daneshfar F., Kabudian S.J., Neekabadi A. Speech emotion recognition using hybrid spectral-prosodic features of speech signal/glottal waveform, metaheuristic-based dimensionality reduction, and Gaussian elliptical basis function network classifier. *Appl. Acoust.* 2020;(166):107360. <https://doi.org/10.1016/j.apacoust.2020.107360>
7. Palo H.K., Behera D., Rout B.C. Comparison of classifiers for speech emotion recognition (SER) with discriminative spectral features. *Advances in Intelligent Computing and Communication: Proceedings of ICAC*. Singapore: Springer; 2020. P. 78–85. https://doi.org/10.1007/978-981-15-2774-6_10
8. Xie Y., Liang R., Liang Z., Huang C., Zou C., Schuller B. Speech emotion classification using attention-based lstm. *IEEE/ACM Trans. Audio. Speech. Lang. Proc.* 2019;27(11):1675–1685. <https://doi.org/10.1109/TASLP.2019.2925934>
9. Hassouneh A., Mutawa A.M., Murugappan M. Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. *InformMed. Unlock.* 2020;(20):100372. <https://doi.org/10.1016/j.imu.2020.100372>

10. Kerkeni L., Serrestou Y., Raoof K., Mbarki M., Mahjoub M.A., Cleder C. Automatic speech emotion recognition using an optimal combination of features based on EMD-TKEO. *Speech. Commun.* 2019;(114):22–35. <https://doi.org/10.1016/j.specom.2019.09.002>
11. Uddin Md. Zia, Nilsson E.G., Emotion recognition using speech and neural structured learning to facilitate edge intelligence. *Engineering Applications of Artificial Intelligence.* 2020;94:103775. <https://doi.org/10.1016/j.engappai.2020.103775>
12. Kadiri S.R., Gangamohan P., Gangashetty S.V., et al. Excitation Features of Speech for Emotion Recognition Using Neutral Speech as Reference. *Circuits. Syst. Signal. Process.* 2020;39:4459–4481. <https://doi.org/10.1007/s00034-020-01377-y>
13. Gorshkov Y.G. Visualization of Lung Sounds Based on Multilevel Wavelet Analysis. *Scientific Visualization.* 2022;14(2):18–26. <https://doi.org/10.26583/sv.14.2.02>
14. Kravchuk D.A. Ultrasonic system for monitoring the psychophysiological state of a train driver. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. Seriya: Upravlenie, vychislitel'naya tekhnika, informatika. Meditsinskoe priborostroenie = Proceedings of the Southwest State University. Series: Control, Computer Engineering, Information Science. Medical Instruments Engineering.* 2020;10(1):134–142. (In Russ.)
15. Prokofieva L.P., Plastun I.L., Filippova N.V., Matveeva L.Yu., Plastun N.S. Recognition of emotions based on the characteristics of the speech signal (linguistic, clinical, information aspects). *Sibirskii filologicheskii zhurnal = Siberian Journal of Philology.* 2021;(2):325–336. (In Russ.)
16. Gorshkov Yu.G. Visualization of human emotional tension using a speech signal. *Nauchnaya vizualizatsiya = Scientific Visualization.* 2023;15(2):102–112 (In Russ.)
17. Gorshkov Y.G., Volkov A.K., Voinova N.A., et al. Acoustocardiography with Assessment of Emotional Tension from the Voice. *Biomed. Eng.* 2020;(53):383–387. <https://doi.org/10.1007/s10527-020-09948-8>

Информация об авторе / Information about the Author

Кравчук Денис Александрович доктор технических наук, Институт нанотехнологий, электроники и приборостроения Южного федерального университета, г. Таганрог, Российская Федерация, e-mail: kravchukda@sfedu.ru, ORCID: 0000-0002-4220-3928

Denis A. Kravchuk, Doctor of Sciences (Engineering), Institute of Nanotechnology, Electronics and Instrumentation of the Southern Federal University, Taganrog, Russian Federation, e-mail: kravchukda@sfedu.ru, ORCID: 0000-0002-4220-3928