

*АВТОМАТИЗАЦИЯ И УПРАВЛЕНИЕ
ТЕХНОЛОГИЧЕСКИМИ ПРОЦЕССАМИ
И ПРОИЗВОДСТВАМИ*

УДК 004

DOI: 10.17212/2782-2230-2022-2-9-20

**ИССЛЕДОВАНИЕ ЗАШУМЛЕННОЙ
АУДИОЗАПИСИ РЕЧИ ДЛЯ ОБРАБОТКИ
В АВТОМАТИЗИРОВАННОЙ СИСТЕМЕ***

Н.А. ВОЛКОВ¹, А.В. ИВАНОВ², Н.Е. КАРПОВА³, А.В. ЧУВАКОВ⁴

¹ 443100, РФ, г. Самара, ул. Молодогвардейская, 244, Самарский государственный технический университет, аспирант кафедры «Электронные системы и информационная безопасность». E-mail: volkovnikandr@gmail.com

² 630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, кандидат технических наук, заведующий кафедрой защиты информации. E-mail: andrej.ivanov@corp.nstu.ru

³ 443100, РФ, г. Самара, ул. Молодогвардейская, 244, Самарский государственный технический университет, кандидат технических наук, заместитель заведующего кафедрой «Электронные системы и информационная безопасность». E-mail: annuim@mail.ru

⁴ 443100, РФ, г. Самара, ул. Молодогвардейская, 244, Самарский государственный технический университет, кандидат химических наук, доцент кафедры «Электронные системы и информационная безопасность». E-mail: avch2105@gmail.com

В настоящей статье рассматривается возможность обработки звука от шума при помощи нейронной сети, которая работает с распознаванием изображений. Чтобы в этом убедиться, были рассмотрены спектрограммы записанного голоса диктора длительностью 10 секунд и спектрограммы с наложением белого шума на записанную аудиодорожку. После анализа зашумленной аудиодорожки субъективным методом (прослушивание аудиодорожки) и анализа спектрограмм зашумленной аудиодорожки было выявлено, что нейронная сеть сможет распознать отличия изображений, на которых видна зашумленность. Это необходимо для того, чтобы в дальнейшем обучить нейронную сеть распознавать интенсивность зашумленности аудиодорожки.

Ключевые слова: разборчивость речи, распознавание речи, обработка звука, нейронная сеть, шумоочистка, белый шум

* Статья получена 10 апреля 2022 г.

ВВЕДЕНИЕ

В области защиты речевой акустической информации большую роль играет применение методов шумоочистки. Существует значительное множество таких методов, которые применяются различными авторами в своих исследованиях. Также существуют программно-аппаратные средства, реализующие методы шумоочистки, но большая часть работы по применению этих методов сводится к работе эксперта. При применении последовательности цепочки методов шумоочистки эксперт добивается понижения отношения сигнал / шум в зашумленной аудиозаписи речи. Технологии и методы распознавания речи в современном мире позволяют задуматься об автоматизации процесса оценки качества шумоочистки аудиозаписей.

Для автоматизации процесса необходимо применить нейронную сеть, которая работает с распознаванием звука, она должна уметь определять зашумленность аудиодорожки. Но так как работа со звуком в нейронной сети – довольно сложный процесс, то для облегчения задачи необходимо применить нейронную сеть, которая работает с распознаванием изображения.

Целью настоящей работы является анализ полученных спектрограмм звуковой аудиодорожки без наложения белого шума на нее и с наложением белого шума. Задачами исследования являются рассмотрение необходимого программного обеспечения для обработки звука, применение данных программ при записи речи и использование функции генерации шума. Также задачей в данном исследовании является получение спектрограммы записанной аудиодорожки и спектрограмм с наложением белого шума на аудиодорожку.

1. ВЫБОР ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ОБРАБОТКИ ЗВУКА

Как было выявлено ранее [1], удобной программой для обработки звука является Adobe Audition – это программное обеспечение, которое содержит набор инструментов для работы со звуком, включающий поддержку нескольких звуковых дорожек, отображение формы волны, спектральный анализ, и позволяет создавать, микшировать, редактировать и восстанавливать аудиозапись. Эта программа имеет мощный инструмент удаления шума: эффект шумоподавления, эффект адаптивного шумоподавления, фильтрация при помощи параметрического эквалайзера, динамическая обработка, удаление гула, использование диагностической панели, просмотр звукового сигнала и спектра [2].

Также было выявлено, что удобным языком программирования для написания программного обеспечения в решаемых нами задачах является Python – это мощный инструмент для создания программ самого разнообразного назначения. С его помощью можно решать задачи различных типов. Язык Python обладает некоторыми примечательными особенностями, которые обуславливают его широкое распространение [3].

Программное обеспечение, которое обрабатывает звуковую дорожку и выдает результат в виде спектрограммы, использует следующие библиотеки:

1) matplotlib – библиотека Python для визуализации данных двумерной (2D) и трехмерной (3D) графики. Библиотека является гибким, легко конфигурируемым пакетом, который вместе с NumPy, SciPy и IPython предоставляет возможности, подобные MATLAB [4];

2) SciPy – библиотека, предназначенная для обработки сигналов, изображений, решения обыкновенных дифференциальных уравнений. При работе с изображениями является дополнением к библиотеке matplotlib [5];

3) Wave – библиотека, которая позволяет обрабатывать файлы, имеющие формат .wav [6].

2. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

2.1. АНАЛИЗ СПЕКТРОГРАММЫ БЕЗ НАЛОЖЕНИЯ БЕЛОГО ШУМА НА АУДИОДОРОЖКУ

При помощи программы обработки звука Adobe Audition мы записали звуковую дорожку одного диктора длительностью 10 секунд в формате .wav. Звуковая дорожка была записана в одноканальном формате звука (моно). Затем, при помощи программного обеспечения, написанного на Python, мы получили спектрограмму диктора (рис. 1).

2.2. АНАЛИЗ СПЕКТРОГРАММ С НАЛОЖЕНИЕМ БЕЛОГО ШУМА НА АУДИОДОРОЖКУ

В функционале Adobe Audition имеется эффект наложения на аудиодорожку шума – коричневый, розовый и белый. Для эксперимента нам необходим только белый шум. Эта функция позволяет указать необходимую длительность по времени белого шума, а также интенсивность белого шума (рис. 2).

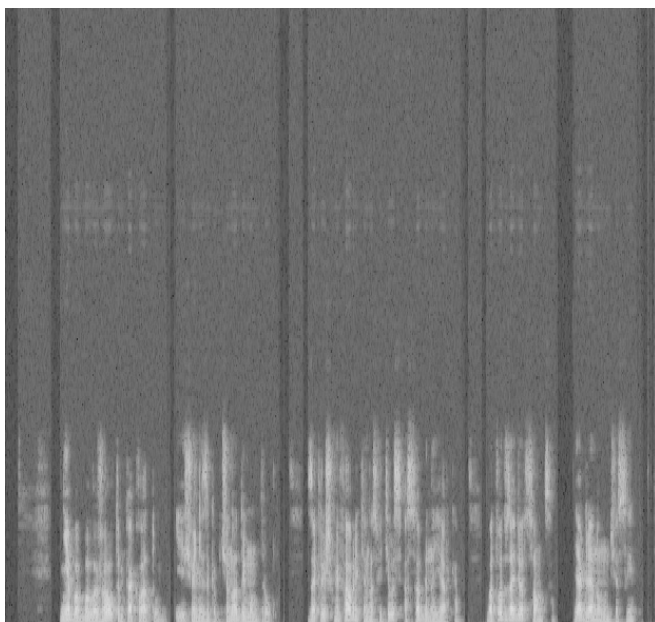


Рис. 1. Спектрограмма записанной речи диктора без наложения белого шума

Fig. 1. Spectrogram of the speaker's recorded speech without white noise overlay

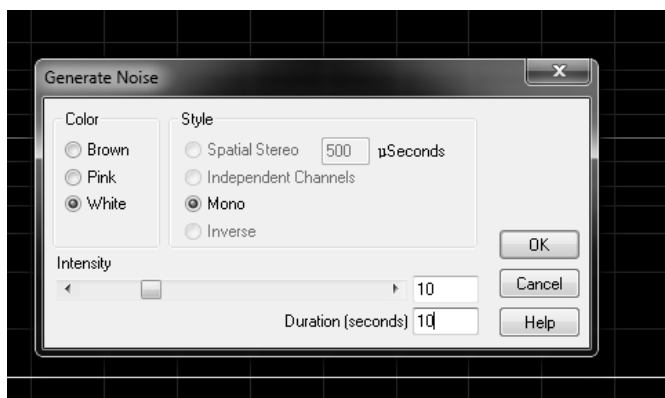


Рис. 2. Окно применения функции «Генерация шума» в Adobe Audition

Fig. 2. Окно применения функции «Генерация шума» в Adobe Audition

Интенсивность белого шума позволяет выбрать значение от 0 до 50, что позволяет нам тщательно рассмотреть, как будет меняться спектрограмма записанной речи при наложении белого шума. Воспользуемся значениями с шагом в 10. При этом также будем давать субъективную оценку слышимости речи диктора.

При интенсивности белого шума, равной 10, все слова диктора были различимы человеческим ухом. Спектрограмма данной аудиодорожки показана на рис. 3.

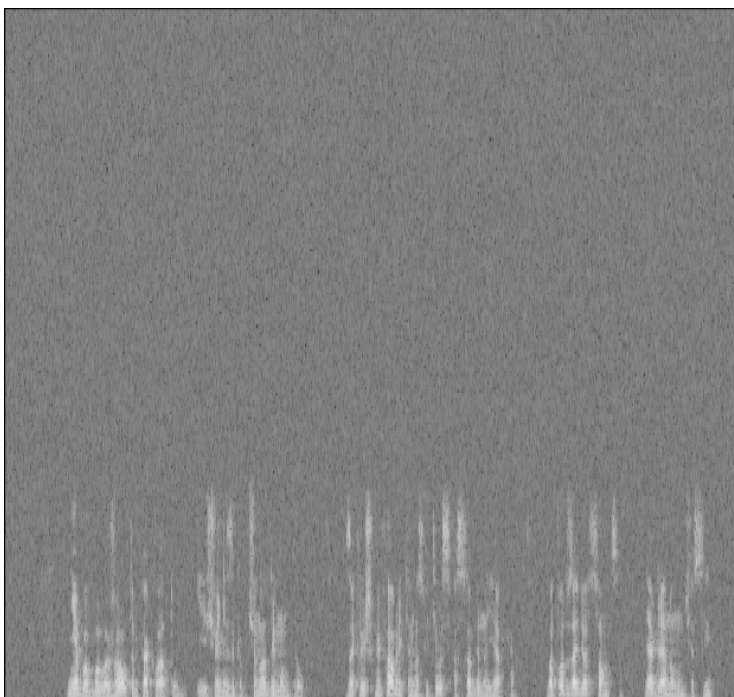


Рис. 3. Спектрограмма записанной речи диктора при наложении белого шума с интенсивностью, равной 10

Fig. 3. The spectrogram of the speaker's recorded speech when white noise is applied with an intensity equal to 10

При интенсивности белого шума, равной 20, было сложнее различить слова диктора, но все слова были различимы человеческим ухом. Спектрограмма этой аудиодорожки показана на рис. 4.

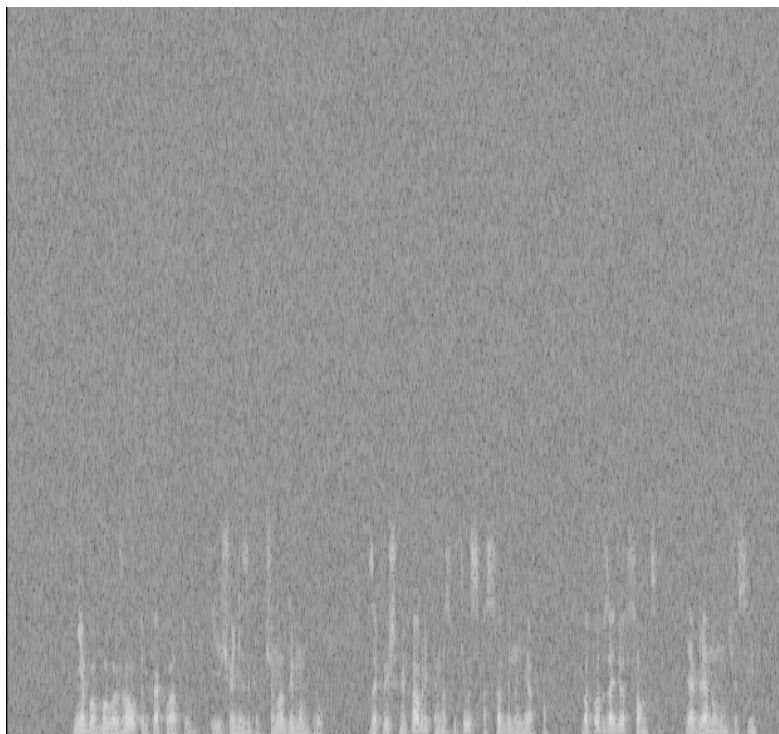


Рис. 4. Спектрограмма записанной речи диктора при наложении белого шума с интенсивностью, равной 20

Fig. 4. The spectrogram of the speaker's recorded speech when white noise is applied with an intensity equal to 20

При интенсивности белого шума, равной 30, распознать удалось половину слов, сказанных диктором. Спектрограмма этой аудиодорожки показана на рис. 5.

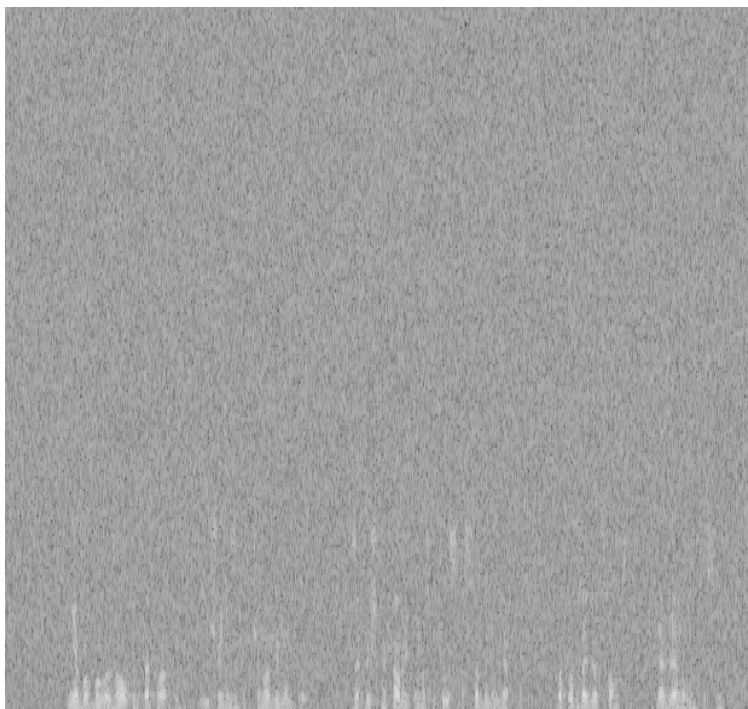


Рис. 5. Спектрограмма записанной речи диктора при наложении белого шума с интенсивностью, равной 30

Fig. 5. The spectrogram of the speaker's recorded speech when white noise is applied with an intensity equal to 30

При интенсивности белого шума, равной 40, были слышны слова диктора, но распознать их человеческим ухом не удалось. Спектрограмма этой аудиодорожки показана на рис. 6.

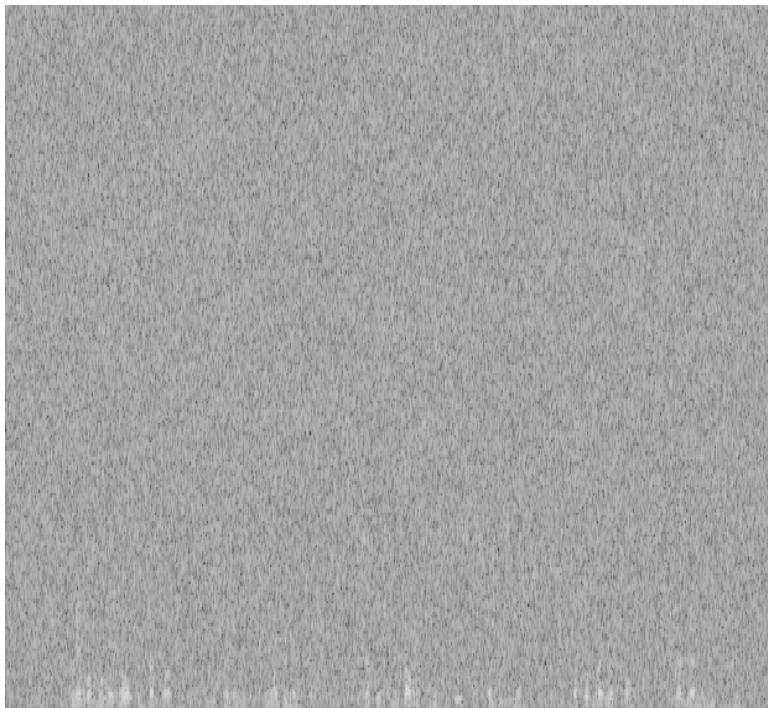


Рис. 6. Спектрограмма записанной речи диктора при наложении белого шума с интенсивностью, равной 40

Fig. 6. The spectrogram of the speaker's recorded speech when white noise is applied with an intensity equal to 40

При интенсивности белого шума, равной 50, слова диктора распознать человеческим ухом стало невозможным. Спектрограмма этой аудиодорожки показана на рис. 7.

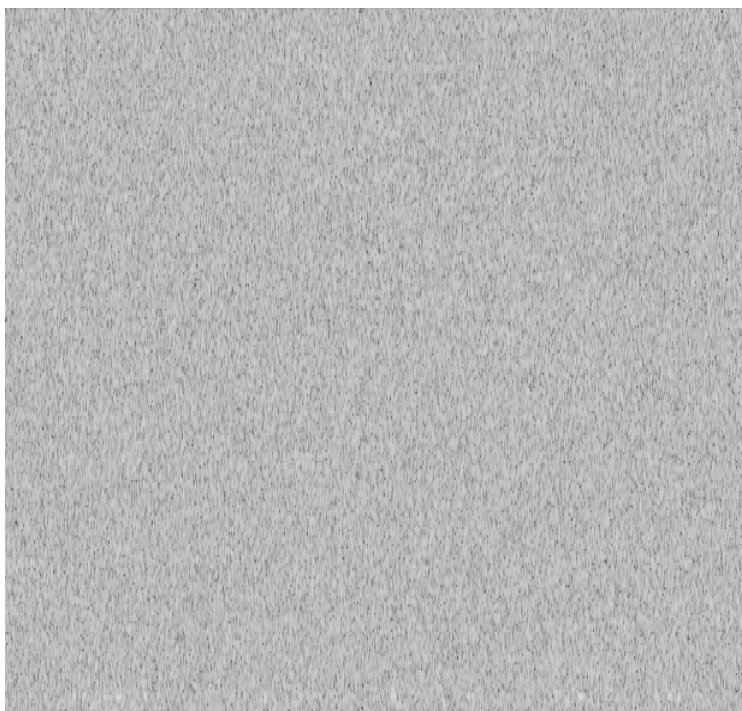


Рис. 7. Спектрограмма записанной речи диктора при наложении белого шума с интенсивностью, равной 50

Fig. 7. The spectrogram of the speaker's recorded speech when white noise is applied with an intensity equal to 50

На рис. 3–7 представлены спектрограммы записанной речи диктора при наложении белого шума с различной интенсивностью. Выявлен порог распознавания речи, равный 50 единицам интенсивности по 100-балльной шкале.

ЗАКЛЮЧЕНИЕ

Таким образом, было выявлено, что при наложении белого шума на звуковую аудиодорожку с записанной речью диктора можно увидеть, что спектрограмма звуковой дорожки отчетливо меняется. Из этого следует, что нейронную сеть можно обучить для распознавания зашумленности аудиодорожки. Дальнейшая работа будет заключаться в том, что необходимо создать

набор данных, состоящий из спектрограмм без наложения белого шума на звуковую дорожку и спектрограмм с наложением белого шума на аудиодорожку с градацией от 10 до 50 общим объемом 10 000 изображений. Затем необходимо обучить нейронную сеть созданным набором данных. Результатом дальнейшей работы будет являться то, что нейронная сеть должна научиться определять зашумленность аудиодорожки по градациям интенсивности белого шума. Эти результаты помогут в развитии автоматизации процесса оценки качества шумоочистки аудиозаписей.

СПИСОК ЛИТЕРАТУРЫ

1. Волков Н.А., Иванов А.В. Анализ существующих средств разработки для решения задачи распознавания речи в области оценки защищенности речевой информации // XXI Всероссийский конкурс-конференция студентов и аспирантов по информационной безопасности «SIBINFO-2021» / Томский государственный университет систем управления и радиоэлектроники. – Томск, 2021. – С. 22–25.
2. Adobe Audition. Профессиональная студия звукозаписи: web-сайт. – URL: <https://www.adobe.com/ru/products/audition.html> (дата обращения: 01.06.2022).
3. Beginner's Guide to Python // Python: web-сайт. – URL: <https://wiki.python.org/moin/BeginnersGuide> (accessed: 01.06.2022).
4. Matplotlib Tutorials: web-сайт. – URL: <https://matplotlib.org/stable/tutorials/index> (accessed: 01.06.2022).
5. SciPy. About us. – URL: <https://scipy.org/about.html> (accessed: 01.06.2022).
6. Read and write WAV files // Python: web-сайт. – URL: <https://docs.python.org/3/library/wave.html> (accessed: 01.06.2022).

Волков Никита Андреевич, аспирант кафедры «Электронные системы и информационная безопасность» Самарского государственного технического университета. Основное направление научных исследований – обработка звука от шума, техническая защита информации от утечки по каналам связи. E-mail: volkovnikandr@gmail.com

Иванов Андрей Валерьевич, кандидат технических наук, заведующий кафедрой защиты информации Новосибирского государственного технического университета. Область научных интересов – обработка звука от шума, техническая защита информации от утечки по каналам связи. E-mail: andrej.ivanov@corp.nstu.ru

Карпова Надежда Евгеньевна, кандидат технических наук, заместитель заведующего кафедрой «Электронные системы и информационная безопасность» Самарского государственного технического университета. Область научных интересов – автоматизация процессов, нейронные сети, техническая защита информации от утечки по каналам связи. E-mail: annuin@mail.ru

Чуваков Александр Владимирович, кандидат химических наук, доцент кафедры «Электронные системы и информационная безопасность» Самарского государственного технического университета. Область научных интересов – автоматизация процессов, нейронные сети, техническая защита информации от утечки по каналам связи. E-mail: avch2105@gmail.com

DOI: 10.17212/2782-2230-2022-2-9-20

Investigation of noisy audio recordings of speech for processing in an automated system*

N.A. Volkov¹, A.V. Ivanov², N.E. Karpova³, A.V. Chuvakov⁴

¹ Samara State Technical University, 244 Molodogvardeyskaya Street, Samara, 443100, Russian Federation, graduate student of the Electronic Systems and Information Security Department. E-mail: volkovnikandr@gmail.com

² Novosibirsk State Technical University, 20 Karl Marx Prospekt, Novosibirsk, 630073, Russian Federation, candidate of technical sciences, head of the Information Security Department. E-mail: andrej.ivanov@corp.nstu.ru

³ Samara State Technical University, 244 Molodogvardeyskaya Street, Samara, 443100, Russian Federation, candidate of technical sciences, associate head of the Electronic Systems and Information Security Department. E-mail: annuin@mail.ru

⁴ Samara State Technical University, 244 Molodogvardeyskaya Street, Samara, 443100, Russian Federation, candidate of chemical sciences, associate professor of the Electronic Systems and Information Security Department. E-mail: avch2105@gmail.com

This article discusses the possibility of processing sound from noise using a neural network that works with image recognition. To make sure of this, spectrograms of the recorded voice of the speaker with a duration of 10 seconds and spectrograms with white noise superimposed on the recorded audio track were considered. After analyzing the noisy audio track by a subjective method (listening to the audio track) and analyzing the spectrograms of the noisy audio track, it was revealed that the neural network will be able to recognize the differences in the images on which the noise is visible. This is necessary in order to further train the neural network to recognize the noise intensity of the audio track.

Keywords: speech intelligibility, speech recognition, sound processing, neural network; noise cleaning, white noise

* Received 10 April 2022.

REFERENCES

1. Volkov N.A., Ivanov A.V. [Analysis of existing development tools for solving the problem of speech recognition in the field of assessing the security of speech information]. *XXI Vserossiiskii konkurs-konferentsiya studentov i aspirantov po informatsionnoi bezopasnosti «SIBINFO-2021»* [XXI All-Russian competition-conference of students and postgraduates on information security "SIBINFO-2021"]. Tomsk State University of Control Systems and Radioelectronics, Tomsk, 2021, pp. 22–25. (In Russian).
2. *Adobe Audition. Professional recording studio*. website. Available at: <https://www.adobe.com/ru/products/audition.html> (accessed 01.06.2022).
3. *Beginner's Guide to Python. Python*: website. Available at: <https://wiki.python.org/moin/BeginnersGuide> (accessed 01.06.2022).
4. *Matplotlib Tutorials*: website. Available at: <https://matplotlib.org/stable/tutorials/index> (accessed 01.06.2022).
5. *SciPy. About us*. Available at: <https://scipy.org/about.html> (accessed 01.06.2022).
6. Read and write WAV files. *Python*: website. Available at: <https://docs.python.org/3/library/wave.html> (accessed 01.06.2022).

Для цитирования:

Исследование зашумленной аудиозаписи речи для обработки в автоматизированной системе / Н.А. Волков, А.В. Иванов, Н.Е. Карпова, А.В. Чуваков // Безопасность цифровых технологий. – 2022. – № 2 (105). – С. 9–20. – DOI: 10.17212/2782-2230-2022-2-9-20.

For citation:

Volkov N.A., Ivanov A.V., Karpova N.E., Chuvakov A.V. Issledovanie zashumlennoi audiozapisi rechi dlya obrabotki v avtomatizirovannoi sisteme [Investigation of noisy audio recordings of speech for processing in an automated system]. *Bezopasnost' tsifrovyykh tekhnologii = Digital Technology Security*, 2022, no. 2 (105), pp. 9–20. DOI: 10.17212/2782-2230-2022-2-9-20.