

УДК 004.93

DOI: 10.18413/2518-1092-2022-7-4-0-8

Урсол Д.В.¹
Болгова Е.В.²
Черноморец Д.А.²
Черноморец А.А.²**ОБ АЛГОРИТМЕ РАСПОЗНАВАНИЯ ЗВУКОВ
НА ОСНОВЕ КОСИНУС ПРЕОБРАЗОВАНИЯ**¹⁾ ООО «Промышленные электронные системы», ул. Михайловское шоссе 121а, г. Белгород, 308000, Россия²⁾ Белгородский государственный национальный исследовательский университет,
ул. Победы д. 85, г. Белгород, 308015, Россия*e-mail: chernomorets@bsu.edu.ru***Аннотация**

Данная статья посвящена решению задачи распознавания различных звуков в окружающей среде, что широко применяется в системах наблюдения и контроля и позволяет идентифицировать объекты различной природы, например, автомобиль, катер, самолет, животных, птиц и др. В работе предложен алгоритм распознавания звуков в звуковом сигнале на основе анализа частотных составляющих сигнала, соответствующих коэффициентам дискретного косинус преобразования фрагментов исследуемого сигнала. Дискретное косинус преобразование обеспечивает в отличие от преобразования Фурье разложение сигнала на вещественные частотные составляющие, что позволяет снизить вычислительные затраты при реализации алгоритма. В разработанном алгоритме на основе частотного анализа звукового сигнала, в качестве примера, определяются ноты различных октав. На этапе предварительной обработки в исходном сигнале выделяются фрагменты, соответствующие паузам, и формируются информативные фрагменты звукового сигнала, при анализе которых на следующем этапе алгоритма осуществляется распознавание нот. Вычислительные эксперименты с модельным звуковым сигналом продемонстрировали работоспособность разработанного алгоритма.

Ключевые слова: звуковой сигнал; дискретное косинус преобразование; частота дискретизации; частота сигнала; ноты; октавы

Для цитирования: Урсол Д.В., Болгова Е.В., Черноморец Д.А., Черноморец А.А. Об алгоритме распознавания звуков на основе косинус преобразования // Научный результат. Информационные технологии. – Т.7, №4, 2022. – С. 67-75. DOI: 10.18413/2518-1092-2022-7-4-0-8

Ursol D.V.¹
Bolgova E.V.²
Chernomorets D.A.²
Chernomorets A.A.²**ABOUT THE SOUNDS RECOGNITION ALGORITHM
BASED ON THE COSINE TRANSFORM**¹⁾ Industrial Electronic Systems LLC, 121a Mikhailovskoe shosse, Belgorod, 308000, Russia²⁾ Belgorod State National Research University, 85 Pobedy St., Belgorod, 308015, Russia*e-mail: chernomorets@bsu.edu.ru***Abstract**

This article is devoted to solving the problem of recognizing various sounds in the environment, which is widely used in surveillance and control systems and allows you to identify objects of various nature, for example, a car, a boat, an airplane, animals, birds, etc. The paper proposes an algorithm for recognizing sounds in an audio signal based on the analysis of the signal frequency components corresponding to the discrete cosine transform coefficients of signal fragments. The discrete cosine transform provides, in contrast to the Fourier transform, the decomposition of the signal into real frequency components, which reduces computational costs when implementing the algorithm. In the developed algorithm, based on the frequency analysis of the audio signal, as

an example, notes of different octaves are determined. At the stage of preprocessing, fragments corresponding to pauses are allocated in the initial signal and the informative audio signal fragments are formed, during the analysis of which, at the next stage of the algorithm, notes are recognized. Computational experiments with a model sound signal demonstrated the developed algorithm.

Keywords: sound signal; discrete cosine transform; sampling rate; signal frequency; notes; octaves

For citation: Ursol D.V., Bolgova E.V., Chernomoretz D.A., Chernomoretz A.A. About the sounds recognition algorithm based on the cosine transform // Research result. Information technologies. – Т.7, №4, 2022. – P. 67-75. DOI: 10.18413/2518-1092-2022-7-4-0-8

Распознавание звуков в окружающей среде является важным в различных сферах деятельности человека. Решение задачи распознавания звуков позволяет идентифицировать объекты различной природы (или оценить их состояние), такие как человек, животное, птица, отдельные устройства (автомобиль, катер, самолет) и др., что широко применяется в системах наблюдения и контроля [1-6].

В данной работе предлагается алгоритм распознавания звуков в звукозаписи, соответствующих отдельным нотам, на основе анализа частотных составляющих сигнала, соответствующих коэффициентам дискретного косинус преобразования [7-11] фрагментов исследуемого звукового сигнала.

Дискретное косинус преобразование (ДКП) связано с дискретным преобразованием Фурье [12-15]. Однако, ДКП обеспечивает в отличие от преобразования Фурье разложение сигнала на вещественные частотные составляющие (гармоники), что позволяет снизить вычислительные затраты при реализации разработанного алгоритма.

Для сигнала $X = (x_n)$, $n = 1, 2, \dots, n_x$, содержащего n_x отсчетов, коэффициенты $W = (w_m)$, $m = 1, 2, \dots, n_x$, дискретного косинус преобразования, в большинстве случаев, определяются следующим соотношением:

$$w_m = \sqrt{\frac{2}{n_x}} \sum_{n=1}^{n_x} x_n \cos \frac{\pi}{n_x} (n - \frac{1}{2})(m - 1), \quad m = 1, 2, \dots, n_x. \quad (1)$$

Предлагаемый алгоритм распознавания звуков состоит в следующем.

Исходными данными в предлагаемом алгоритме являются следующие величины:

- звуковой сигнал $Y = (y_n)$, $n = 1, 2, \dots, n_s$, – вектор, содержащий значения n_s отсчетов сигнала (будем считать, что значения отсчетов сигнала находятся в интервале $[-1, 1]$),

- частота дискретизации f_s – частота взятия отсчетов непрерывного по времени сигнала при его дискретизации (Гц).

Прецедентами в решаемой задаче распознавания являются ноты До, Ре, Ми, Фа, Соль, Ля и Си, которым соответствуют различные частоты $F = \{f_{ij}\}$, $i = 1, 2, \dots, 7$, $j = 1, 2, \dots, 9$, в девяти октавах: Субконтроктава, Контроктава, Большая октава, Малая октава, Первая октава, Вторая октава, Третья октава, Четвертая октава и Пятая октава.

Частоты нот $F = \{f_{ij}\}$, $i = 1, 2, \dots, 7$, $j = 1, 2, \dots, 9$, в соответствующих октавах имеют значения, приведенные в таблице 1 [16].

Таблица 1

Соответствие частот (Гц) и нот в различных октавах

Table 1

Matching of frequencies (Hz) and notes in different octaves

Октава	До	Ре	Ми	Фа	Соль	Ля	Си
Субконтроктава	16,352	18,354	20,602	21,827	24,500	27,500	30,368
Контроктава	32,703	36,708	41,203	43,654	48,999	55,000	61,735
Большая октава	65,406	73,416	82,407	87,307	97,999	110,000	123,470
Малая октава	130,810	146,830	164,810	174,610	196,000	220,000	246,940
Первая октава	261,630	293,660	329,630	349,230	392,000	440,000	493,880
Вторая октава	523,250	587,330	659,260	698,460	783,990	880,000	987,770
Третья октава	1046,3	1174,7	1318,5	1396,9	1568,000	1760,000	1975,500
Четвертая октава	2093,0	2349,3	2637,0	2793,8	3136,000	3520,000	3951,100
Пятая октава	4186,0	4698,6	5274,0	5587,7	6271,900	7040,000	7902,100

Исходными параметрами разработанного алгоритма также являются следующие величины:

- размер n_w отсчетов анализируемого на наличие пауз фрагмента сигнала (скользящее окно),
- пороговое значение h , используемое для обнаружения пауз при анализе значений отсчетов сигнала,
- погрешность ε обнаружения нот (применяется для оценивания отличия частот выделяемых в сигнале гармоник от значений, приведенных в таблице 1).

На этапе предварительной обработки предлагается выполнить сегментацию сигнала на фрагменты между паузами.

Первоначально на основе пороговой обработки с порогом h в сигнале выделяются промежутки, соответствующие паузам. Для этого анализируются различные положения скользящего окна при его перемещении по отсчетам сигнала.

В очередном положении скользящего окна размером n_w отсчетов предлагается вычислить среднее арифметическое значений модулей амплитуды отсчетов сигнала $Y = (y_n)$, $n = 1, 2, \dots, n_s$, которые соответствуют текущему положению k окна (k – номер отсчета сигнала Y , которому соответствует первый отсчет скользящего окна в заданном положении), $k = 1, 2, \dots, n_s - n_w + 1$:

$$m_k = \frac{1}{n_w} \sum_{i=k}^{k+n_w-1} |y_i|.$$

Если выполняется условие:

$$m_k < h,$$

то выделенный с помощью скользящего окна фрагмент считается паузой (звук отсутствует).

Иначе – выделенный с помощью скользящего окна фрагмент считается информативным (содержит звук) и добавляется к списку информативных фрагментов, содержащих различные звуки.

Если в сигнале не обнаружены информативные фрагменты, то необходимо уменьшить размер скользящего окна и/или уменьшить значение порога и повторить этап удаления пауз из исходного сигнала.

На следующем этапе осуществляется распознавание нот, соответствующих звукам, содержащимся в выделенных информативных фрагментах звукового сигнала. Распознавание нот основано на частотном анализе сигналов в рамках дискретного косинус преобразования (1).

Для каждого выделенного фрагмента $X = (x_n)$, $n = 1, 2, \dots, n_x$, содержащего n_x отсчетов, необходимо выполнить следующие действия:

1. Вычислить коэффициенты ДКП $W = (w_m)$, $m = 1, 2, \dots, n_x$, на основании соотношения (1).

2. Построить график энергетического спектра фрагмента сигнала, указав вдоль оси абсцисс значения частот от нуля до $f_s/2$ – половина частоты дискретизации сигнала, вдоль оси ординат – квадраты коэффициентов ДКП.

3. Упорядочить значения квадратов коэффициентов ДКП по убыванию. В данной работе предлагается анализировать не более первых 5 коэффициентов, что позволяет распознать в выделенном фрагменте сигнала до 5 различных нот.

4. Определить частоту f_l , $l=1,2,\dots,5$, которая соответствует очередному упорядоченному коэффициенту ДКП, имеющему номер m_l , $l=1,2,\dots,5$:

$$f_l = \frac{f_s}{2n_x}(m_l - 1). \quad (2)$$

5. Классифицировать звук по значению его частоты f_l (определить ноту, соответствующую звуку):

а) определить в матрице $F = \{f_{ij}\}$, $i=1,2,\dots,7$, $j=1,2,\dots,9$, (таблица 1) частоту f_{ij}^* , ближайшую к частоте f_l ,

$$f_{ij}^* = \min_{\substack{i=1,2,\dots,7 \\ j=1,2,\dots,9}} |f_{ij} - f_l|, \quad (3)$$

б) будем считать, что частота f_l определяет ноту, соответствующую частоте f_{ij}^* , если для погрешности σ ,

$$\sigma = |f_{ij}^* - f_l|, \quad (4)$$

выполняется следующее условие:

$$\sigma < \varepsilon. \quad (5)$$

Для проверки работоспособности разработанного алгоритма были проведены вычислительные эксперименты. В качестве анализируемого звукового сигнал был использован модельный сигнал длительностью 7 с, содержащий отдельные ноты, которые разделены паузами. Частота дискретизации анализируемого звукового сигнал – 8192 отсчетов/с.

На рисунке 1 приведен модельный звуковой сигнал и его энергетический спектр, вычисленный на основе дискретного косинус преобразования.

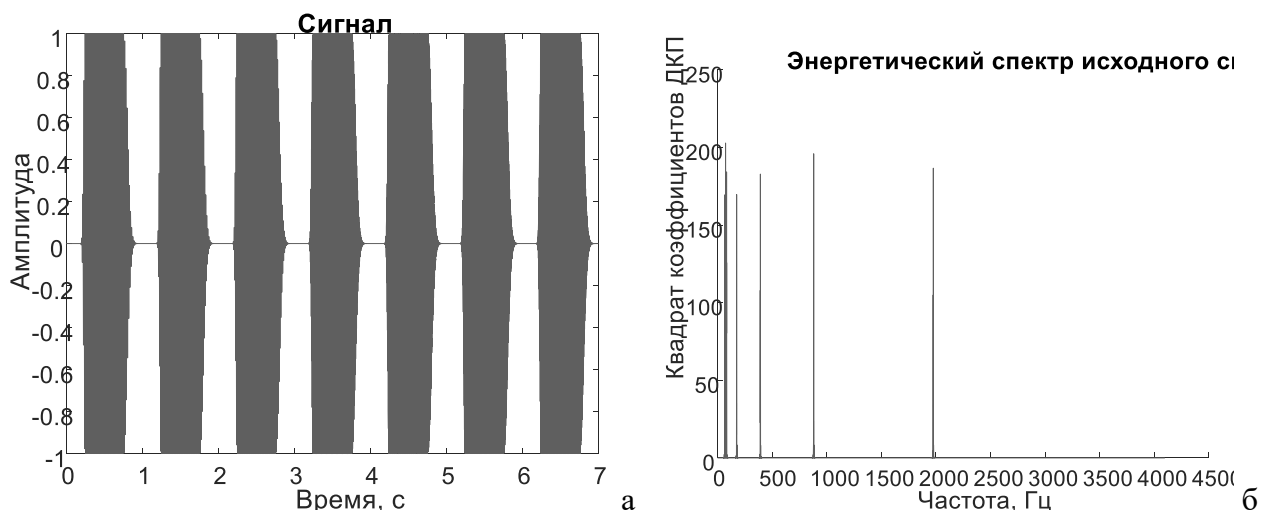


Рис. 1. Модельный сигнал:

а – значения отсчетов, б – энергетический спектр на основе ДКП

Fig. 1. Model signal: a – the samples values, b – the energy spectrum based on the DCT

Следует отметить, что энергетический спектр, приведенный на рисунке 1б, не позволяет определить фрагменты сигнала, которым соответствуют частоты, отображенные на графике энергетического спектра. В работе предложено выделить информативные фрагменты сигнала, соответствующие различным нотам.

Значения параметров алгоритма имеют следующие значения:

- размер n_w анализируемого на наличие пауз фрагмента сигнала – 128 отсчетов,
- пороговое значение h , применяемое для обнаружения пауз, – 0,2,
- погрешность ε обнаружения нот – 0,01.

Результат анализа исходного сигнала на наличие пауз приведен на рисунке 2. На данном рисунке отсчеты, соответствующие паузам, отмечены линиями.

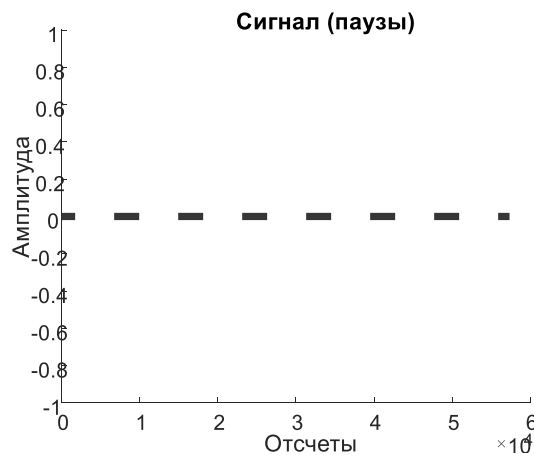
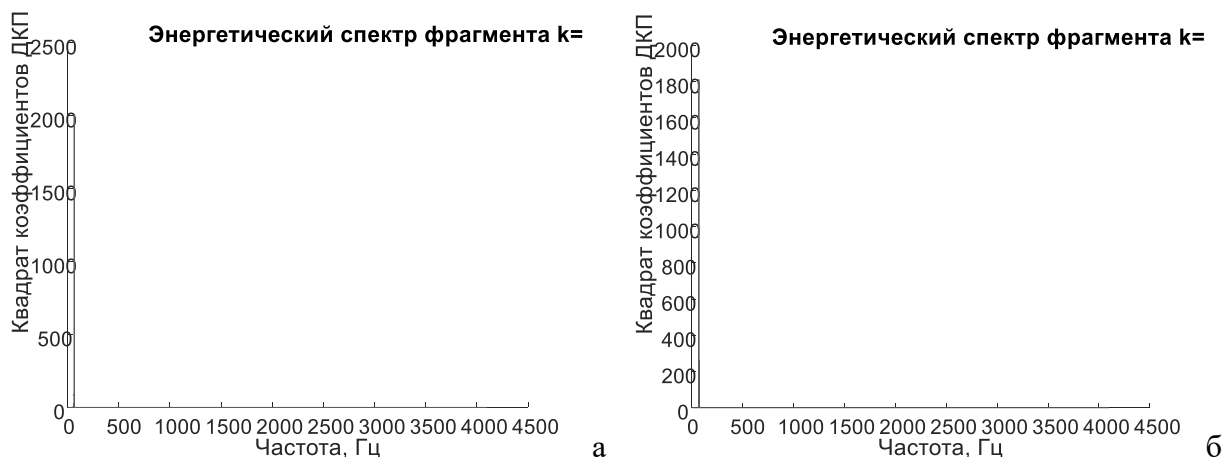


Рис. 2. Результат выделения пауз
Fig. 2. The result of the pauses selection

В результате выделения пауз в исходном сигнале было выявлено 7 информативных фрагментов, содержащих звуки.

На рисунке 3 приведены графики спектров 7 выделенных информативных фрагментов анализируемого сигнала ($k = 1, 2, \dots, 7$).



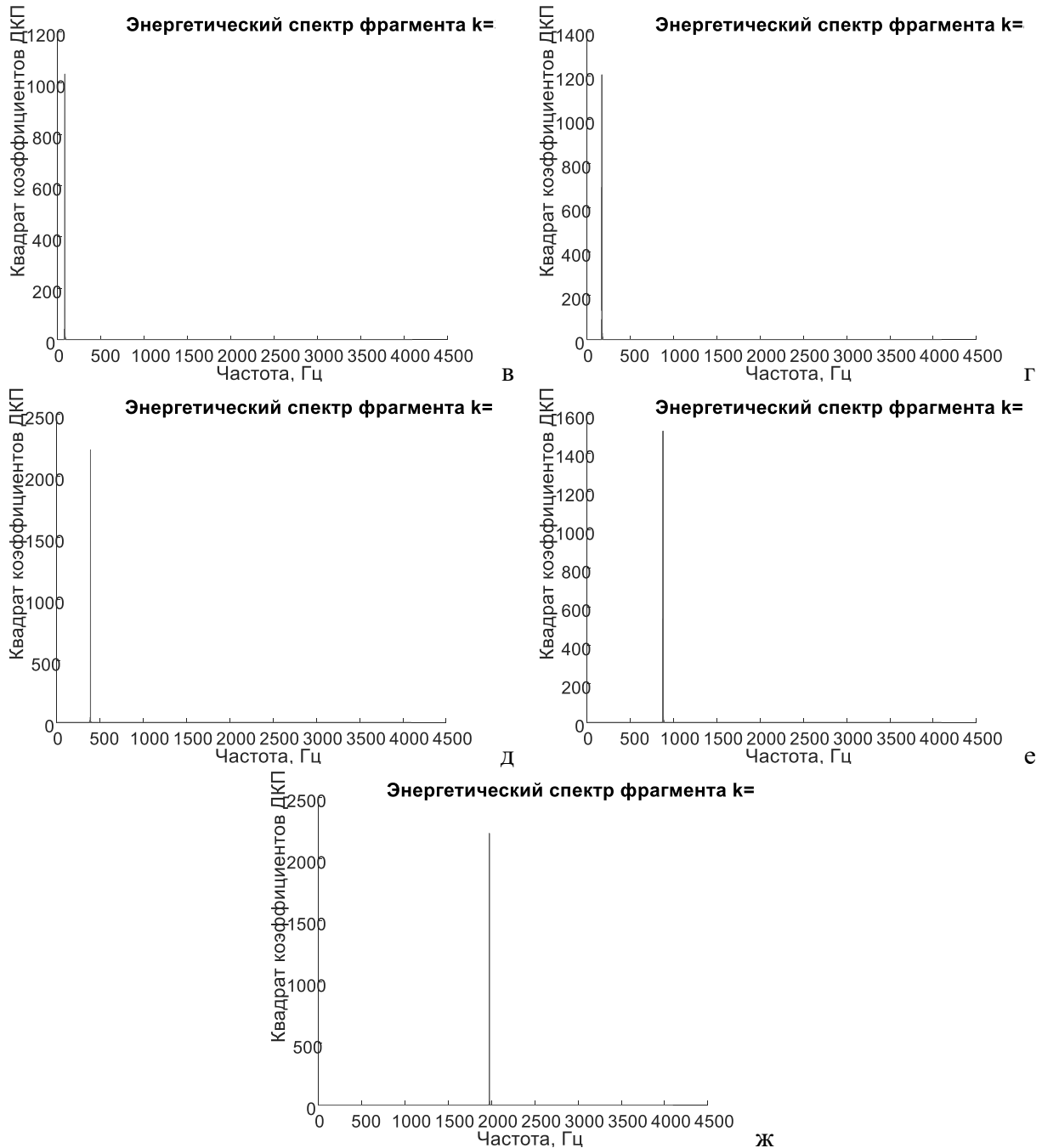


Рис. 3. Энергетические спектры выделенных информативных фрагментов анализируемого сигнала: а – для фрагмента 1, б – для фрагмента 2, в – для фрагмента 3,

г – для фрагмента 4, д – для фрагмента 5, е – для фрагмента 6, ж – для фрагмента 7

Fig. 3. Energy spectra of the analyzed signal selected informative fragments:

а – for fragment 1, б – for fragment 2, в – for fragment 3,

д – for fragment 4, е – for fragment 5, ф – for fragment 6, г – for fragment 7

Результаты распознавания – выявления соответствия частот гармоник, приведенных на рисунке 3, и нот различных октав на основании соотношений (2)-(5), приведены в таблице 2.

Таблица 2

Результаты распознавания нот в модельном сигнале

Table 2

The notes recognition results in the model signal

Номер фрагмента	Частота гармоник в фрагменте, Гц	Нота	Погрешность σ
1	65,641	Большая октава, нота До	0,0035
2	73,026	Большая октава, нота Ре	0,0053
3	82,872	Большая октава, нота Ми	0,0056
4	173,949	Малая октава, нота Фа	0,0038
5	392,205	Первая октава, нота Соль	0,0005
6	879,589	Вторая октава, нота Ля	0,0004
7	1975,795	Третья октава, нота Си	0,0001

Результаты распознавания нот в модельном сигнале, приведенные в таблице 2, демонстрируют работоспособность предложенного алгоритма.

Таким образом, в работе предложен алгоритм распознавания звуков на основе частотного анализа в рамках дискретного косинус преобразования. В разработанном алгоритме на основе анализа гармоник исследуемого сигнала, соответствующих коэффициентам ДКП, в качестве примера, определяются ноты различных октав, содержащихся в звуковом сигнале. Вычислительные эксперименты продемонстрировали работоспособность разработанного алгоритма. Следует отметить, что данный алгоритм при незначительных доработках может быть обобщен для решения задачи распознавания звуков произвольной природы.

Список литературы

1. Абрамов Г.В., Коробова Л.А., Ивашин А.Л., Матыцина И.А. Анализ и использование математических методов для распознавания звуковых сигналов // Вестник Воронежского государственного университета инженерных технологий. 2015. № 2 (64). С. 61-65.
2. Данилов В.В., Салех Х.М. Исследование неисправностей беспилотных летательных аппаратов на основе распознавания звука агрегатов // Решение. 2018. Т. 1. С. 126-129.
3. Митянок В.В., Коновалова Н.В. Применение фазового анализа звуков речи для распознавания человека по его голосу // Техническая акустика. 2013. Т. 13. С. 4.
4. Припадчев А.А., Черный Е.В. Разработка системы для автоматизированного распознавания звуков критических ситуаций в потоке аудиосигнала / В сборнике: Альманах научных работ молодых ученых университета ИТМО. в 5 томах. 2016. С. 188-190.
5. Васильев Д.Е. Повышение эффективности распознавания звуков специального назначения с использованием искусственного интеллекта / В сборнике: Современные тенденции развития науки и мирового сообщества в эпоху цифровизации. Сборник материалов VII Международной научно-практической конференции. Редколлегия: Бабаева З.Ш. [и др.]. Москва, 2022. С. 143-145.
6. Балабаев С.А., Лупин С.А. Ускорение работы метода определения голосов птиц / В сборнике: Передовое развитие современной науки как драйвер роста экономики и социальной сферы. Сборник II Всероссийской научно-практической конференции. Петрозаводск, 2020. С. 51-56.
7. Иванов А.Н., Киселев А.М. Защита аудиофайлов методом цифрового маркирования на основе дискретного косинусного преобразования и дискретного вейвлет-преобразования // Ученые заметки ТОГУ. 2019. Т. 10. № 3. С. 42-52.
8. Ракицкий В.А. Дискретное косинусное преобразование как средство компьютерной обработки информации // Проблемы информатизации и управления. 2019. Т. 2. № 62. С. 52-56.
9. Черноморец А.А., Болгова Е.В., Черноморец Д.А. Обобщенный субполосный анализ на основе унитарных преобразований // Научные ведомости Белгородского государственного университета. Серия: Экономика. Информатика. 2015. № 7 (204). С. 97-104.
10. Новиков К.Д. Программа для осуществления дискретного косинусного преобразования средствами GPU / Свидетельство о регистрации программы для ЭВМ 2021611111, 21.01.2021. Заявка № 2021610330 от 13.01.2021.

11. Бумагин А.В., Гондарь А.В., Прудников А.А., Стешенко В.Б. Устройство для вычисления дискретного косинусного преобразования // Патент на изобретение RU 2430407 C1, 27.09.2011. Заявка № 2010115396/08 от 20.04.2010.
12. Паршин Б.Я., Жуков Д.О. Сравнение дискретного преобразования Фурье и модифицированного косинус-преобразования при сжатии аудиоинформации // Вестник компьютерных и информационных технологий. 2010. № 5 (71). С. 12-18.
13. Жилияков Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным: моногр. / Е.Г. Жилияков. – Белгород: Изд-во БелГУ, 2007. – 160 с.
14. Жилияков Е.Г., Черноморец А.А., Болгова Е.В. О субинтервальных матрицах на основе унитарных преобразований // Научный результат. Информационные технологии. 2017. Т. 2. № 1. С. 55-63.
15. Лапаев Н.Г., Седов А.Н., Шевченко О.В. Определения параметров сигналов с использованием дискретного преобразования Фурье и вейвлет-преобразования // Методы и устройства передачи и обработки информации. 2004. № 6. С. 140-148.
16. Конев А.А., Онищенко А.А., Костюченко Е.Ю., Якимук А.Ю. Автоматическое распознавание музыкальных нот // Научный вестник Новосибирского государственного технического университета. 2015. № 3 (60). С. 32-47.

References

1. Abramov G.V., Korobova L.A., Ivashin A.L., Matytsina I.A. Analysis and use of mathematical methods for recognition of sound signals // Bulletin of the Voronezh State University of Engineering Technologies. 2015. No. 2 (64). pp. 61-65.
2. Danilov V.V., Salekh Kh.M. Investigation of malfunctions of unmanned aerial vehicles based on the recognition of the sound of aggregates // Decision. 2018. Vol. 1. P. 126-129.
3. Mityanok V.V., Konovalova N.V. Application of phase analysis of speech sounds to recognize a person by his voice // Technical Acoustics. 2013. Vol. 13. P. 4.
4. Pripadchev A.A., Cherny E.V. Development of a system for automated recognition of sounds of critical situations in an audio signal stream / In the collection: Almanac of Scientific Works of Young Scientists of ITMO University. in 5 volumes. 2016. P. 188-190.
5. Vasiliev D.E. Improving the efficiency of special purpose sound recognition using artificial intelligence / In the collection: Modern trends in the development of science and the world community in the era of digitalization. Collection of materials of the VII International scientific-practical conference. Editorial board: Babaeva Z.Sh. [and etc.]. Moscow, 2022, pp. 143-145.
6. Balabaev S.A., Lupin S.A. Accelerating the work of the method for determining the voices of birds / In the collection: Advanced development of modern science as a driver for the growth of the economy and the social sphere. Collection of the II All-Russian Scientific and Practical Conference. Petrozavodsk, 2020. P. 51-56.
7. Ivanov A.N., Kiselev A.M. Protection of audio files by digital marking method based on discrete cosine transform and discrete wavelet transform // Uchenye zametki TOGU. 2019. V. 10. No. 3. P. 42-52.
8. Rakitsky V.A. Discrete cosine transform as a means of computer processing of information // Problems of informatization and control. 2019. V. 2. No. 62. P. 52-56.
9. Chernomorets A.A., Bolgova E.V., Chernomorets D.A. The generalized subband analysis on the basis of unitary transformations // Belgorod State University Scientific Bulletin. Economics. Information technologies. 2015. No. 7 (204). pp. 97-104.
10. Novikov K.D. Program for implementing discrete cosine transformation by means of GPU / Certificate of registration of the computer program 2021611111, 01/21/2021. Application No. 2021610330 dated 01/13/2021.
11. Bumagin A.V., Gondar A.V., Prudnikov A.A., Steshenko V.B. Device for calculating discrete cosine transform // Patent for invention RU 2430407 C1, 27.09.2011. Application No. 2010115396/08 dated 04/20/2010.
12. Parshin B.Ya., Zhukov D.O. Comparison of Discrete Fourier Transform and Modified Cosine Transform in Audio Information Compression. Bulletin of Computer and Information Technologies. 2010. No. 5 (71). pp. 12-18.
13. Zhilyakov E.G. Variational methods of analysis and construction of functions based on empirical data: monograph. / E.G. Zhilyakov. - Belgorod: Publishing House of BelSU, 2007. – 160 p.
14. Zhilyakov E.G., Chernomorets A.A., Bolgova E.V. About subinterval matrices based on unitary transformations // Research Result. Information Technology. 2017. V. 2. No. 1. P. 55-63.

15. Lapaev N.G., Sedov A.N., Shevchenko O.V. Determining the parameters of signals using the discrete Fourier transform and wavelet transform // Methods and devices for transmitting and processing information. 2004. No. 6. P. 140-148.

16. Konev A.A., Onishchenko A.A., Kostyuchenko E.Yu., Yakimuk A.Yu. Automatic recognition of musical notes // Scientific Bulletin of the Novosibirsk State Technical University. 2015. No. 3 (60). pp. 32-47.

Урсол Денис Владимирович, кандидат технических наук, инженер-программист

Болгова Евгения Витальевна, кандидат технических наук, доцент кафедры прикладной информатики и информационных технологий

Черноморец Дарья Андреевна, аспирант кафедры информационно-телекоммуникационных систем и технологий

Черноморец Андрей Алексеевич, доктор технических наук, доцент, профессор кафедры прикладной информатики и информационных технологий

Ursol Denis Vladimirovich, Candidate of Technical Sciences, Software Engineer

Bolgova Evgeniya Vitalievna, Candidate of Technical Sciences, Associate Professor of the Department of Applied Informatics and Information Technologies

Chernomorets Darya Andreevna, postgraduate student of the Department of Information and Telecommunications Systems and Technologies

Chernomorets Andrey Alekseevich, Doctor of Technical Sciences, Associate Professor, Professor of the Department of Applied Informatics and Information Technologies