

УДК 004.62:623.76

DOI: [10.26102/2310-6018/2025.50.3.007](https://doi.org/10.26102/2310-6018/2025.50.3.007)

## Выбор акустических признаков в задачах обнаружения беспилотных летательных аппаратов

Д.Е. Прозоров<sup>1</sup>✉, В.А. Бызов<sup>1</sup>, Р.Е. Мышкин<sup>2</sup>

<sup>1</sup>Вятский государственный университет, Киров, Российская Федерация

<sup>2</sup>Научно-исследовательский институт средств вычислительной техники, Киров, Российская Федерация

**Резюме.** С ростом числа инцидентов, связанных с неправомерным использованием беспилотных летательных аппаратов (БПЛА), повышается актуальность разработки эффективных методов их автоматического обнаружения. В статье выполнен краткий обзор современных подходов к обнаружению БПЛА, особое внимание уделено методам акустического мониторинга, обладающим рядом преимуществ по сравнению с радиочастотными и визуальными системами. Рассмотрены основные акустические признаки, используемые для распознавания звуковых сигналов дронов, а также методы их выделения с помощью открытых библиотек Librosa и Essentia. Для исследования эффективности различных признаков сформирован и использован сбалансированный датасет, включающий аудиозаписи дронов и фоновых шумов. Апробирована методика многоэтапного отбора признаков с применением библиотеки Feature-engine, включающая удаление неизменяющихся (константных) и дублирующихся признаков, корреляционный анализ и оценку значимости признаков. В результате получено подмножество из 53 акустических признаков, обеспечивающее компромисс между качеством обнаружения БПЛА и вычислительными затратами. Описаны математические основы формирования спектральных признаков, в том числе различные типы спектрограмм (мел-, барк- и гамма-спектрограммы), векторные и скалярные акустические признаки. Полученные результаты могут быть использованы при построении систем автоматического акустического обнаружения БПЛА, базирующихся на методах машинного обучения.

**Ключевые слова:** беспилотный летательный аппарат, акустические сигналы, акустические признаки, спектральный анализ, машинное обучение.

**Для цитирования:** Прозоров Д.Е., Бызов В.А., Мышкин Р.Е. Выбор акустических признаков в задачах обнаружения беспилотных летательных аппаратов. *Моделирование, оптимизация и информационные технологии*. 2025;13(3). URL: <https://moitvvt.ru/ru/journal/pdf?id=1955> DOI: 10.26102/2310-6018/2025.50.3.007

## Selection of acoustic features in unmanned aerial vehicle detection tasks

D.E. Prozorov<sup>1</sup>✉, V.A. Byzov<sup>1</sup>, R.E. Myshkin<sup>2</sup>

<sup>1</sup>Vaytka State University, Kirov, the Russian Federation

<sup>2</sup>Scientific Research Institute of Computer Engineering, Kirov, the Russian Federation

**Abstract.** With the increasing number of incidents involving the unauthorized use of unmanned aerial vehicles (UAVs), the development of effective methods for their automatic detection has become increasingly relevant. This article provides a concise overview of current approaches to UAV detection, with particular emphasis on acoustic monitoring methods, which offer several advantages over radio-frequency and visual systems. The main acoustic features used for recognizing drone sound signals are examined, along with techniques for extracting these features using open-source libraries such as Librosa and Essentia. To evaluate the effectiveness of various features, a balanced dataset was compiled and utilized, containing audio recordings of drones and background noise. A multi-stage feature

selection methodology was tested using the Feature-engine library, including the removal of constant and duplicate features, correlation analysis, and feature importance assessment. As a result, a subset of 53 acoustic features was obtained, providing a balance between UAV detection accuracy and computational cost. The mathematical foundations of spectral feature extraction are described, including different types of spectrograms (mel-, bark-, and gammatone-spectrograms), as well as vector and scalar acoustic features. The results presented can be used to develop automatic UAV acoustic detection systems based on machine learning methods.

**Keywords:** unmanned aerial vehicle, acoustic signals, acoustic features, spectral analysis, machine learning.

**For citation:** Prozorov D.E., Byzov V.A., Myshkin R.E. Selection of acoustic features in unmanned aerial vehicle detection tasks. *Modeling, Optimization and Information Technology*. 2025;13(3). (In Russ.). URL: <https://moitvivr.ru/ru/journal/pdf?id=1955> DOI: 10.26102/2310-6018/2025.50.3.007

## Введение

С увеличением доступности беспилотных летательных аппаратов (БПЛА, дронов) для коммерческого и частного использования наблюдается рост инцидентов безопасности, связанных с их неправомерным применением [1, 2]. Упоминаются случаи, когда БПЛА использовались для шпионажа за коммерческими объектами [3], проникновения на закрытые территории [4, 5], съемки частной жизни людей. Еще более серьезную проблему представляют дроны в контексте терроризма и военных действий: для доставки взрывчатых веществ, проведения атак на людей или объекты критически важной инфраструктуры [6]. В сложившейся ситуации приобретают актуальность средства обнаружения БПЛА, которые могли бы использоваться в автоматизированных системах противодействия угрозам.

Среди методов обнаружения БПЛА можно выделить методы радиолокационного сопровождения, радиочастотного анализа, компьютерного зрения и акустического мониторинга [1, 2].

Радиолокационные системы в основном используются для наблюдения за большими объектами, могут работать в широком диапазоне погодных условий, но зачастую малопригодны для обнаружения низколетящих, относительно медленных и небольших объектов. Некоторые типы радаров – доплеровские радары – могут быть полезны для обнаружения сигнатур, вызванных работой роторов и винтов БПЛА [7]. Интересным направлением является пассивная радиолокация, позволяющая обнаруживать объекты в зоне действия лицензированных источников радиосигналов, например, ТВ и сотовых систем связи [8]. Дополнительного повышения эффективности радиолокационных систем можно достичь применением мультисенсорных платформ и их интеграцией с алгоритмами машинного обучения (МО) [9, 10], что позволяет компенсировать ограничения используемых технологий. В целом, применение радаров ограничивается высокой стоимостью оборудования и эксплуатации.

Мониторинг радиочастотного спектра является одной из самых распространенных техник обнаружения дронов [1]. Сканеры спектра могут обнаружить сигналы каналов управления и каналов передачи видео БПЛА. Основной уязвимостью таких сканеров является возможность пропуска БПЛА, летящих в автономном режиме (режиме радиомолчания) и БПЛА, использующих нестандартные частоты.

Автоматическое обнаружение БПЛА путем видеонаблюдения можно свести к известной задаче компьютерного зрения – обнаружения и классификации объектов. Отмечается [11], что недостатками визуального обнаружения в оптическом и/или инфракрасном диапазоне являются: относительно малая дальность обнаружения,

высокая чувствительность к погодным условиям и невидимость объектов за препятствиями.

Одним из перспективных направлений разработок является обнаружение акустических сигналов дронов. Звуки, излучаемые двигателями и движущимися элементами БПЛА, могут быть обнаружены на расстояниях, превышающих дистанции достоверного визуального обнаружения, и достигают нескольких сотен метров [3]. При наличии микрофонной решетки может быть решена и задача определения направления на источник акустических сигналов [12].

Основными недостатками методов акустического обнаружения в отличие от радиочастотных методов являются: зависимость акустического спектра БПЛА от режима работы аппарата (взлет, посадка, маневрирование и изменение скорости, прямой полет) и сильное влияние акустической сцены – шумов техногенного и естественного происхождения. Это создает проблемы для традиционных методов обнаружения, основанных на фиксированных частотных шаблонах. В таких условиях эффективным способом обнаружения и классификации сигналов становятся методы МО [9, 10]. МО на текущий момент являются одним из ключевых инструментов для решения задач распознавания образов благодаря их способности анализировать сложные паттерны в данных, которые трудно обнаружить более традиционными способами на основе алгоритмов теории оптимального приема сигналов (изложенных, например, в работах В.И. Тихонова, Л.М. Финка, Дж. Проакиса и других ученых).

Важным этапом обработки данных в процессе обучения моделей МО или их работы в режиме предсказания является этап выделения признаков, поступающих на вход модели. Известно больше сотни признаков, используемых в задачах обработки акустических сигналов. В данной работе выполнено исследование акустических признаков сигналов БПЛА, с целью определения подмножества, позволяющего обеспечить компромисс между характеристиками качества обнаружения БПЛА и вычислительными ресурсами системы обнаружения.

## Материалы и методы

**Выбор акустических признаков.** В исследовании рассмотрены акустические признаки, которые можно выделить специализированными open source библиотеками Librosa и Essentia (Таблица 1). Librosa поддерживает высокоуровневый интерфейс для функций анализа акустических данных и предназначена для Python-приложений<sup>1</sup>. Essentia изначально реализована на языке C++, но также имеет API к Python и позволяет более детально контролировать процессы получения акустических признаков<sup>2</sup>.

Таблица 1 – Акустические признаки  
Table 1 – Acoustic features

№ п/п	Обозначение признака	Описание
1	mel_db	Мел-спектрограмма. Используются 24 полосы частот. Значения переведены в децибелы.
2	bark_db	Bark-спектрограмма. Используются 27 полос частот. Значения переведены в децибелы.

Таблица 1 (продолжение)

<sup>1</sup> Librosa. Audio and Music Processing in Python. URL: <https://librosa.org> (дата обращения: 06.02.2025).

<sup>2</sup> Essentia. Open-Source Library and Tools for Audio and Music Analysis, Description And Synthesis. URL: <https://essentia.upf.edu> (дата обращения: 06.02.2025).

Table 1 (continued)

3	erb_db	Гамматон-спектрограмма. Использовалось 40 полос частот, поэтому получены 40 признаков. Значения переведены в децибелы.
4	Mfcc	Мел-частотные кепстральные коэффициенты. Количество коэффициентов равно 13.
5	mfcc_delta	Коэффициенты Delta-MFCC. Эти коэффициенты определяются как разности между коэффициентами MFCC двух последовательных фреймов. Полученные значения можно рассматривать как аппроксимацию частной производной MFCC по времени. Количество признаков – 13.
6	mfcc_delta2	Коэффициенты Delta-Delta-MFCC. Эти коэффициенты определяются как разности между коэффициентами Delta-MFCC двух последовательных фреймов. Полученные значения можно рассматривать как аппроксимацию второй частной производной MFCC по времени. Количество признаков – 13.
7	gfcc	Гамматон-частотные кепстральные коэффициенты. Количество коэффициентов – 13.
8	gfcc_delta	Коэффициенты Delta-GFCC. Количество коэффициентов – 13.
9	gfcc_delta2	Коэффициенты Delta-Delta-GFCC. Количество коэффициентов – 13.
10	centroid	Спектральный центроид.
11	contrast	Спектральный контраст. Использованы 7 полос частот.
12	chroma	Хромограмма. Использованы 12 полос частот.
13	crest	Спектральный гребень.
14	decrease	Спектральное уменьшение.
15	entropy	Спектральная энтропия.
16	flatness	Спектральная плоскостность.
17	flux	Спектральный поток.
18	spread	Спектральный разброс.
19	skewness	Спектральная асимметрия.
20	kurtosis	Спектральный эксцесс.
21	rolloff	Спектральный спад.
22	pitch	Частота основного тона.
23	zero_crossing_rate	Скорость пересечения нуля.
24	rms	Среднеквадратичное значение.

Для проведения исследований использовался сформированный авторами датасет AudioDataset\_Detection\_Balanced, который содержит 4708 аудиозаписей звуков дронов и 4708 аудиозаписей звуков, не относящихся к шуму дронов. Все аудиозаписи имеют частоту дискретизации 44,1 кГц, продолжительность каждой записи – 1 сек. Данный датасет получен объединением корпусов: «Drone-detection-dataset»<sup>3</sup>, ACSAC [13], DREGON [14] и был разбит на обучающую и тестовую сбалансированные подвыборки в пропорции 8:2 (7532 аудиозаписи в обучающей выборке и 1884 – в тестовой).

Каждый аудиофайл был разбит на фреймы размером 1024 отсчетов с перекрытием в 512 отсчетов. Таким образом, количество фреймов для каждого файла равно  $[44100/512] = 86$ . Для построения набора признаков каждого аудиофайла применялся следующий подход: значение признака вычислялось для каждого из 86 фреймов, после чего усреднялось.

<sup>3</sup> DroneDetectionThesis. Drone-detection-dataset. GitHub. URL: <https://github.com/DroneDetectionThesis/Drone-detection-dataset> (дата обращения: 06.02.2025).

В соответствии с Таблицей 1 для каждого аудиофайла был построен вектор из 201 признака. Чтобы выбрать наиболее значимые для классификации признаки, была использована библиотека Feature-engine<sup>4</sup>. При помощи этой библиотеки были реализованы несколько этапов отбора признаков. Процедура отбора производилась на обучающей выборке датасета.

На первом этапе при помощи класса DropConstantFeatures был осуществлен поиск признаков, которые являются константными или почти константными. Признак признавался константным, если он имел одно и то же значение не менее чем в 99,8 % наблюдений. Константных признаков не было обнаружено. Далее при помощи класса DropDuplicateFeatures был осуществлен поиск дублирующихся признаков. Таких признаков также не было обнаружено.

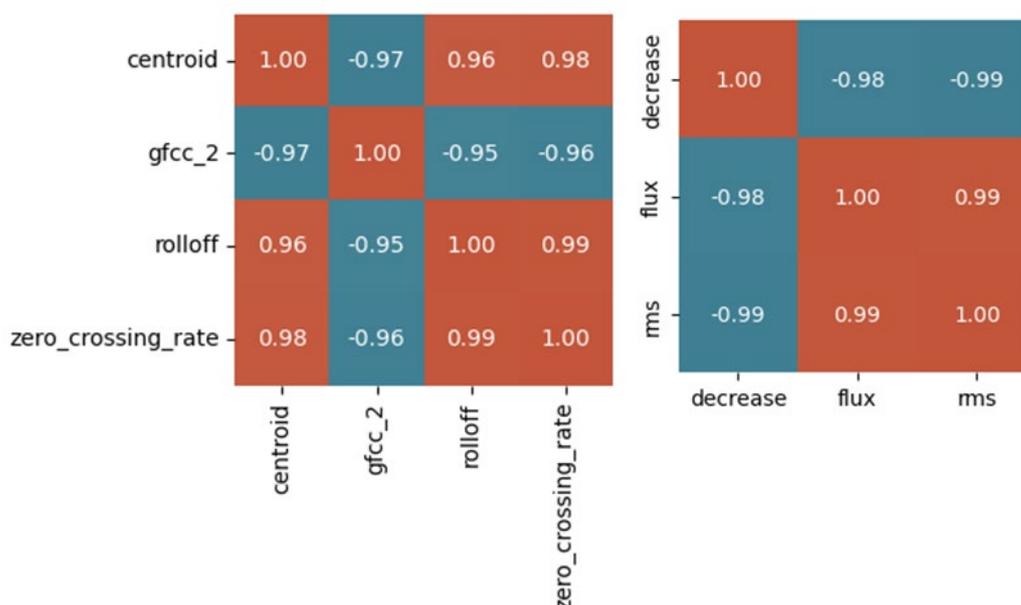


Рисунок 1 – Корреляционные матрицы двух групп признаков  
Figure 1 – Correlation matrices of two groups of features

На следующем этапе при помощи класса SmartCorrelatedSelection были определены группы сильно коррелирующих признаков, для которых модуль коэффициента корреляции Пирсона превышает 0,95. Было обнаружено 16 групп таких признаков. Большинство из этих групп содержат признаки, соответствующие разным полосам частот для трех типов спектрограмм. Отметим две выделяющиеся группы коррелирующих признаков: {centroid, gfcc\_2, rolloff, zero\_crossing\_rate} и {decrease, flux, rms}. Корреляционные матрицы для этих групп признаков приведены на Рисунке 1. В каждой из 16 групп коррелирующих признаков был оставлен один признак, для которого модель случайного леса (с десятью деревьями), основанная на одном этом признаке, давала лучший результат по метрике ROC-AUC. После выполнения этого этапа осталось 122 признака.

На последнем этапе при помощи класса SelectBySingleFeaturePerformance на каждом из оставшихся признаков по отдельности была обучена модель случайного леса (с десятью деревьями). Использовалась трехкратная (3-fold) кросс-валидация. Было посчитано среднее значение метрики ROC-AUC для этих 122 одномерных моделей и

<sup>4</sup> Feature-engine. URL: <https://feature-engine.trainindata.com/en/latest/index.html> (дата обращения: 06.02.2025).

оставлены те признаки, для которых значение ROC-AUC превысило данное среднее значение.

В результате было выбрано 66 признаков: bark\_db[2,7,9,14,18,26,27], chroma[1,...,12], contrast[1,...,7], crest, decrease, entropy, erb\_db[4,7,9,13,15,17,40], flatness, gfcc[3,5,...,13], kurtosis, mel\_db[15,21,23], mfcc[1,...,5,7,...,10,12,13], pitch, skewness, spread, zero\_crossing\_rate.

**Математические модели акустических признаков.** Рассмотрим выбранные акустические признаки более подробно.

Большая часть параметров (Таблица 1) являются производными от оценок спектральной плотности мощности (СПМ) сигнала, дискретные отсчеты которой также можно использовать в качестве одной из компонент вектора признаков (ВП). Известны классы как непараметрических, так и параметрических методов получения статистически устойчивых, сглаженных оценок СПМ [15]. К наименее требовательным к вычислительным ресурсам можно отнести непараметрические коррелограммные и периодограммные методы оценки СПМ<sup>5</sup> [16], к которым, в частности, относится метод Уэлча. Помимо них можно упомянуть не рассматриваемые в данной работе и более вычислительно затратные методы авторегрессионного оценивания СПМ (на основе, например, алгоритма Юла-Уолкера), метод Кэйпона (метод минимума дисперсии), алгоритм гармонического разложения Писаренко и алгоритм классификации множественных сигналов (MUSIC).

В качестве компромиссного решения для расчета СПМ взят метод Уэлча. Уэлч модифицировал известный метод Бартлетта, добавив взвешенную обработку данных с помощью оконных функций и перекрытие сегментов для вычисления выборочных оценок СПМ [15]. Цель оконной обработки – ослабление эффектов боковых лепестков СПМ и уменьшение смещения оценки. Цель перекрытия сегментов – снижение дисперсии оценки спектра. Массив данных из  $N$  отсчетов разбивается на  $P$  перекрывающихся сегментов. Максимальное количество сегментов равно целой части отношения  $\text{int}[(N - D)/(S + 1)]$ , где  $S$  – смещение сегментов относительно друг друга,  $D$  – количество отсчетов в сегменте. Отсчеты  $p$ -го взвешенного сегмента:

$$x^{(p)}[n] = w[n] \cdot x[n + pS], \quad (1)$$

где  $0 \leq n \leq D - 1$ ;  $0 \leq p \leq P - 1$ .

При наличии оконной фильтрации нормированная оценка СПМ  $p$ -го сегмента данных будет равна:

$$\begin{aligned} \hat{P}^{(p)}[k] &= \frac{1}{UD} |\mathcal{F}(x^{(p)}[n])|^2 = \\ &= \frac{1}{UD} \left| \sum_{n=0}^{D-1} x^{(p)}[n] \exp(-j2\pi kn/D) \right|^2, \quad 0 \leq k \leq D - 1, \end{aligned} \quad (2)$$

где  $\mathcal{F}(x^{(p)}[n])$  – дискретно-временное преобразование Фурье (ДПФ)  $p$ -го сегмента, и

$$U = \frac{1}{D} \left| \sum_{n=0}^{D-1} w[n] \right|^2$$

– нормирующий множитель, компенсирующий энергетический эффект непрямоугольного окна  $w[n]$ .

Оценка спектра методом периодограммы Уэлча:

$$\hat{P}_W[k] = \frac{1}{P} \sum_{p=0}^{P-1} \hat{P}^{(p)}[k]. \quad (3)$$

<sup>5</sup> Шахтарин Б.И., Ковригин В.А. *Методы спектрального оценивания случайных процессов*. Москва: Горячая Линия–Телеком; 2011. 256 с.

Для гауссовых процессов наилучшим окном является окно Ханна, рекомендуется перекрытие сегментов 65 % [15].

Примеры расчета спектров методом Уэлча представлены на Рисунке 2.

Спектрограмма представляет собой график изменения спектра акустического сигнала во времени. Спектрограмму можно получить объединением  $k$  СПМ на разных отрезках времени [17]:

$$S[t_i, k] = \begin{bmatrix} \hat{P}[t_i, k] \\ \dots \\ \hat{P}[t_j, k] \end{bmatrix}, \quad (4)$$

где  $t_i$  – момент начала первого временного фрейма спектрограммы,  $t_j$  – момент начала последнего временного фрейма. При этом фреймы, на которых вычисляются выборочные спектры  $\hat{P}[t_i, k]$ , в общем случае могут пересекаться.

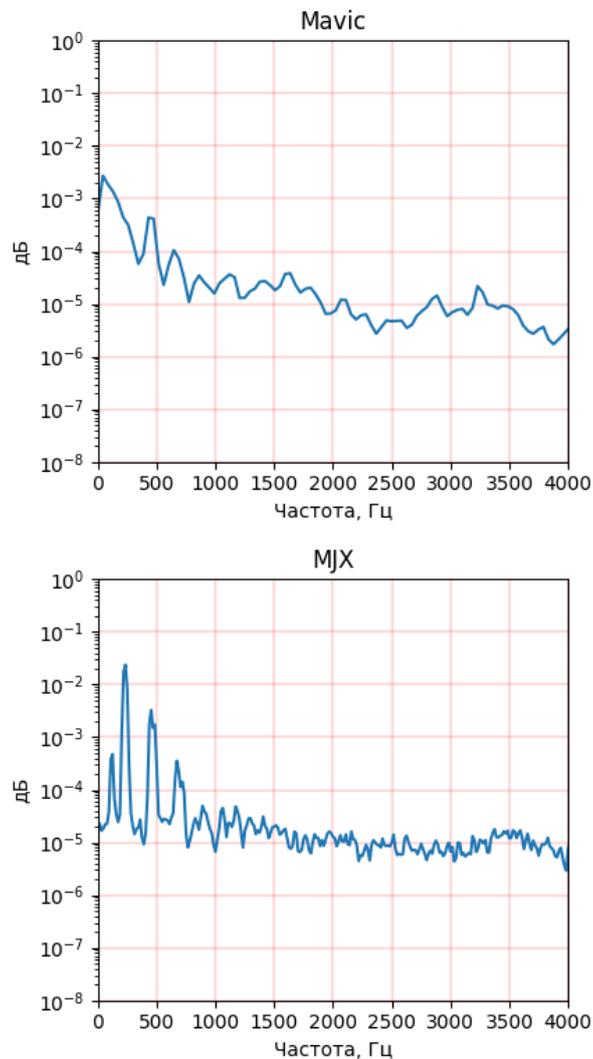


Рисунок 2 – СПМ малых БПЛА мультироторного типа  
 Figure 2 – PSD of small multirotor UAVs

Для построения мел-спектрограммы используется т. н. мел-шкала (существует множество вариантов мел-шкал). Часто для преобразования частоты в мел-коэффициенты используется формула О'Шонесси [18]:

$$mel[n] = 2595 \log_{10} \left( 1 + \frac{f_n}{700} \right) \cong 1127 \ln \left( 1 + \frac{f_n}{700} \right). \quad (5)$$

Количество полос (банков) частотных фильтров мел-шкал и их размер могут отличаться в зависимости от реализации. В Librosa и Essentia при расчете мел-шкал могут использоваться способы преобразования, реализованные в библиотеках НТК (Hidden Markov Model Toolkit)<sup>6</sup> и Slaney (MATLAB Auditory Toolbox)<sup>7</sup>. В библиотеке Slaney коэффициенты, получаемые в частотных полосах, нормализуются на площадь частотных полос. В НТК фильтры настроены таким образом, чтобы пиковое значение коэффициентов не превышало 1,0. Кроме того, в НТК фильтры распределяются так, чтобы их центры были линейно распределены в мел-шкале ниже 1 кГц и логарифмично выше; часто используется 26 фильтров. В библиотеке Slaney фильтры распределяются более равномерно, реализация может включать 40 фильтров. Мел-фильтры представляют собой треугольные функции с максимумом на центральной частоте полосы  $n$ . В результате суммирования весовых коэффициентов фильтров с отсчетами СПМ формируются оценки энергии сигнала  $\hat{E}_{mel}[n]$ . При необходимости полученные оценки могут логарифмироваться для сжатия динамического диапазона результата.

Мел-спектрограмма формируется конкатенацией полученных оценок для последовательности временных фреймов:

$$S_{mel}[t_i, k] = \begin{bmatrix} \hat{E}_{mel}[t_i, k] \\ \dots \\ \hat{E}_{mel}[t_j, k] \end{bmatrix}, \quad (6)$$

где  $\hat{E}_{mel}[t_i, k]$  –  $k$ -й мел-признак, полученный на интервале  $[t_i, t_i + \Delta T]$ .

Шкала Барка (Bark spectrogram) была предложена Э. Цвикером в 1961 г. Шкала основана на психоакустической концепции критических полос слуха и представляет собой неравномерную частотную шкалу, отражающую особенности восприятия звука человеческим ухом, и на текущий момент считается несколько устаревшей. Одна из простых аппроксимаций, позволяющая преобразовывать частоту в барк-коэффициент и обратно, представлена в работе [19]:

$$bark[n] = 13 \arctan(0,00076f_n) + 3,5 \arctan \left( \left( \frac{f_n}{7500} \right)^2 \right). \quad (7)$$

В каждой  $n$ -й барк-полосе формируется оценка энергии сигнала  $\hat{E}_{bark}[n]$  путем суммирования отсчетов СПМ. Барк-спектрограмма:

$$S_{bark}[t_i, k] = \begin{bmatrix} \hat{E}_{bark}[t_i, k] \\ \dots \\ \hat{E}_{bark}[t_j, k] \end{bmatrix}. \quad (8)$$

В библиотеке Essentia используется набор фиксированных полос<sup>8</sup>.

В гамматон-спектрограмме используются гамматон-фильтры, которые (как и в предыдущем случае) моделируют характеристики человеческой слуховой системы. Изначально гамматон-фильтр был разработан для моделирования частотно-избирательной характеристики базилярной мембраны внутреннего уха человека. В последнее время этот тип спектрограмм приобрел популярность и в других областях обработки звуковых сигналов, например, в голосовой биометрии и обнаружении

<sup>6</sup> Young S., et al. The HTK Book. HTK. URL: <http://htk.eng.cam.ac.uk/docs/docs.shtml> (дата обращения: 06.02.2025).

<sup>7</sup> Slaney M. *Auditory Toolbox: A Matlab Toolbox for Auditory Modeling Work*. Interval Research Corporation; 1998. 52 p.

<sup>8</sup> BarkBands. Essentia. URL: [https://essentia.upf.edu/reference/streaming\\_BarkBands.html](https://essentia.upf.edu/reference/streaming_BarkBands.html) (дата обращения: 06.02.2025).

фальсификаций аудио. Гамматон-фильтры частотно-зависимы: обеспечивается более высокое разрешение на низких частотах и более низкое разрешение на высоких частотах.

Импульсная характеристика гамматон-фильтра [20]:

$$h_{\text{gammatone}}[k] = at^{n-1} \exp(-2\pi bt) \cos(2\pi f_c t_k + \varphi), \quad (9)$$

где  $n$  – порядок фильтра,  $b$  – ширина полосы частот,  $f_c$  – центральная частота полосы. Параметр  $\varphi$  (фаза) при реализации часто игнорируется.

Рассмотрим так называемую ERB-шкалу (Equivalent Rectangular Bandwidth)<sup>9</sup>:

$$\text{erb}[n] = 24,7 \left( \frac{4,37f_c}{1000} + 1 \right). \quad (10)$$

Рекомендуемые значения полос гамматон-фильтров равны  $b_n = 1,019 \text{ erb}[n]$ , количество фильтров 64–128.

Оценки энергии сигнала  $\hat{E}_{\text{erb}}[n]$  формируются в результате оконной обработки сигналов внутри частотных полос (10) фильтрами (9).

$$S_{\text{erb}}[t_i, k] = \begin{bmatrix} \hat{E}_{\text{erb}}[t_i, k] \\ \dots \\ \hat{E}_{\text{erb}}[t_j, k] \end{bmatrix}. \quad (11)$$

Мел-частотные кепстральные коэффициенты (мел-кепстр – MFCC) были впервые предложены в 1980-х годах Дэвисом и Мермельштейном [21]. Они разработали этот метод для улучшения распознавания речи, и с тех пор MFCC фактически стали стандартным подходом в области обработки акустических данных.

Мел-частотные кепстральные коэффициенты рассчитываются на основе оценок СПМ в мел-шкале (5), формируемой фильтрами с треугольной полосой пропускания. Банк фильтров обычно состоит из 12 треугольных фильтров. Итоговые коэффициенты MFCC рассчитываются следующим образом [22]:

$$c_{\text{MFCC}}[n] = \sqrt{\frac{2}{M}} \sum_{k=1}^M \left\{ \log(\hat{E}_{\text{mel}}[k]) \cos \left[ \frac{\pi n(k-0,5)}{M} \right] \right\}, \quad n = 1, \dots, N, \quad (12)$$

$N$  – количество вычисляемых коэффициентов.

Расчет гамматон-кепстральных коэффициентов похож на процедуру вычисления мел-кепстра. После формирования банка гамматон-фильтров (см. (9), (10)), выполняется дискретное косинусное преобразование в заданных частотных полосах [23]:

$$c_{\text{GFCC}}[n] = \sqrt{\frac{2}{M}} \sum_{k=1}^M \left\{ \frac{1}{3} \log(\hat{E}_{\text{erb}}[k]) \cos \left( \frac{\pi n(k-0,5)}{M} \right) \right\}, \quad n = 1, \dots, N, \quad (13)$$

$N$  – количество вычисляемых коэффициентов.

Хромограмма является отображением распределения энергии аудиосигнала по различным тональным классам (нотам) с течением времени. При этом тональные сигналы, отличающиеся ровно на целое число октав, относятся к одному классу. Обычно используется равномерно темперированная шкала с 12 классами. Аналогом такой классификации в музыке является (одинаковое) обозначение нот, отличающихся на октаву.

Существует несколько способов преобразования аудиосигнала в хромограмму. В работе [24] предложен способ расчета хромограммы через быстрое преобразование Фурье (БПФ), в то время как в [25] используется набор эллиптических фильтров для вычисления энергии сигнала в заданных частотных полосах.

<sup>9</sup> Slaney M. *An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank*. Apple Computer; 1993. 42 p.

При использовании БПФ элементы вектора коэффициентов хромаграммы могут быть получены через уравнение [24]:

$$chroma[t_i, k] = \sum_{n \in S_k} \frac{\mathcal{F}(x[t_i, n])}{N_k}, \quad (14)$$

где  $x[t_i, n]$  – выборки сигнала на  $i$ -м временном фрейме;  $S_k$  – множество тональных элементов (частот), входящих в один класс;  $N_k$  – количество этих элементов;  $\mathcal{F}(\cdot)$  – БПФ. Для уменьшения дисперсии оценок СПМ можно использовать периодограммные оценки, подобные (3).

Спектральный контраст позволяет отразить относительное распределение гармонических и негармонических компонентов в спектре сигнала и изначально предназначался для классификации музыкальных произведений.

Частотный диапазон при вычислении октавного спектрального контраста делится на 6 октавных поддиапазонов [26], в каждом из которых оценивается амплитудный спектр с помощью БПФ. Для каждого поддиапазона  $k$ :

– амплитуды сортируются по убыванию:  $x_{k,1} > x_{k,2} > \dots > x_{k,N}$ , где  $N$  – число компонент в поддиапазоне;

– вычисляются «пики» как среднее значение наибольших амплитудных компонент:  $peak_k = \log\left(\frac{1}{\alpha^N} \sum_{i=1}^{\alpha N} x_{k,i}\right)$ ; параметр  $\alpha$  (обычно 0,02) определяет размер усредняемой окрестности;

– вычисляются «впадины», как среднее значение наименьших амплитудных компонент:  $valley_k = \log\left(\frac{1}{\alpha^N} \sum_{i=1}^{\alpha N} x_{k,N-i+1}\right)$ ;

– вычисляется спектральный контраст:

$$sc[k] = peak[k] - valley[k]. \quad (15)$$

Спектральный гребень (Spectral Crest Factor – SCF) вычисляется как отношение максимального значения СПМ к его среднему значению в  $n$ -й полосе частот<sup>10</sup>:

$$csf[n] = \frac{\max(\hat{P}[k])}{\frac{1}{N} \sum_k \hat{P}[k]}. \quad (16)$$

Спектральное уменьшение (Spectral Decrease) отражает степень снижения спектральных компонент<sup>10</sup>:

$$decrease = \frac{1}{\sum_{k=2, \dots, K} \hat{P}[k]} \sum_{k=2, \dots, K} \frac{\hat{P}[k] - \hat{P}[1]}{K-1}. \quad (17)$$

Спектральная энтропия может быть вычислена (после нормализации спектра) по формуле, аналогичной формуле Шеннона:

$$entropy = - \sum_k \hat{P}[k] \log(\hat{P}[k]). \quad (18)$$

Предварительно спектр нормализуется так, чтобы его интеграл по всем частотам был равен 1. Это делается для того, чтобы спектральная плотность могла быть интерпретирована как распределение вероятностей.

Спектральная плоскостность (Spectral Flatness) – это мера зашумленности (плоскостности, декорреляции) спектра или его части. Она вычисляется как отношение среднего геометрического к среднему арифметическому значения спектра:

<sup>10</sup> Peeters G. A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project. ResearchGate. URL: [https://www.researchgate.net/publication/200688649\\_A\\_large\\_set\\_of\\_audio\\_features\\_for\\_sound\\_description\\_similarity\\_and\\_classification\\_in\\_the\\_CUIDADO\\_project](https://www.researchgate.net/publication/200688649_A_large_set_of_audio_features_for_sound_description_similarity_and_classification_in_the_CUIDADO_project) (дата обращения: 06.02.2025).

$$sfm[n] = \frac{(\prod_k \hat{P}[k])^{1/N}}{\frac{1}{N} \sum_k \hat{P}[k]}. \quad (19)$$

Спектральный разброс (Spectral Spread) определен<sup>10</sup>, по существу, как дисперсия спектральных бинов:

$$spread = \sigma^2(\hat{P}[k]). \quad (20)$$

Спектральная асимметрия (Spectral Skewness) спектральных бинов является аналогом центрального момента третьего порядка случайных процессов:

$$skewness = \frac{m_3(\hat{P}[k])}{\sigma^3(\hat{P}[k])}. \quad (21)$$

Спектральный эксцесс (Spectral Kurtosis), в свою очередь, вычисляется как центральный момент четвертого порядка:

$$kurtosis = \frac{m_4(\hat{P}[k])}{\sigma^4(\hat{P}[k])}. \quad (22)$$

Существует довольно много методов определения частоты основного тона (ЧОТ, pitch). Условно, можно выделить два основных подхода, на основе оценки максимумов автокорреляционной функции [27] и на основе преобразования Фурье и оценке спектра [28]. В библиотеке Essentia реализованы оба варианта.

Частота пересечения нуля (zero crossing rate) характеризует частоту изменения знака сигнала:

$$zcr = \frac{1}{N-1} \sum_{k=1}^{N-1} |sign(x[k]) - sign(x[k-1])|. \quad (23)$$

### Результаты и обсуждение

Выполнен анализ временных затрат на вычисление отобранных признаков. Для этого измерено время вычисления вектора признаков для 100 аудиофайлов, каждый продолжительностью в одну секунду. Эксперимент осуществлялся на компьютере с процессором Intel Core i5-7400 (3.00 ГГц), объем оперативной памяти равен 32 ГБ. Отметим, что признаки chroma и contrast были вычислены с использованием библиотеки Librosa, все остальные признаки находились при помощи функций библиотеки Essentia. Результаты представлены на Рисунке 3 (использовалась логарифмическая шкала по оси ординат).

На рисунке также указано время нахождения спектра сигнала (столбец spectrum), поскольку вычисление спектра является необходимым этапом нахождения многих других используемых признаков. Для признаков spread, skewness, kurtosis представлено одно время, так как данные признаки вычислялись при помощи одной функции Essentia.

Проведенные измерения показывают, что использование признаков pitch и chroma не целесообразно в системах реального времени из-за долгого времени вычислений.

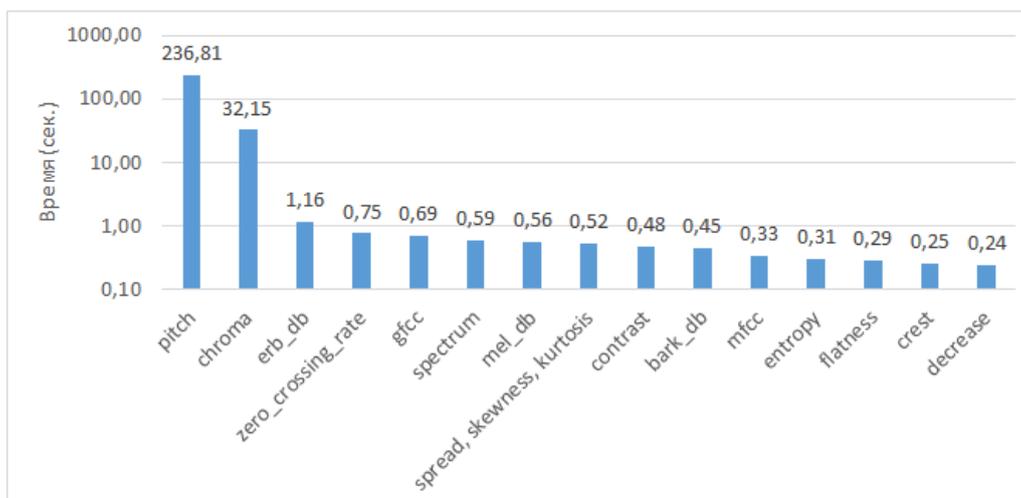


Рисунок 3 – Время вычисления акустических признаков для 100 аудиофайлов  
Figure 3 – Time for acoustic feature extraction for 100 audio files

Чтобы проверить корректность проведенной выше процедуры отбора признаков обучение четырех моделей МО повторялось трижды: на множестве всех признаков (Таблица 1), на множестве 66 выбранных признаков и на множестве выбранных признаков без признаков pitch и chroma (53 признака). Используемые модели – это MLPClassifier, SVC, RandomForestClassifier (из библиотеки sklearn<sup>11</sup>) и XGBClassifier (из библиотеки xgboost<sup>12</sup>). Признаки были приведены к нулевому среднему и единичной дисперсии при помощи метода StandardScaler из библиотеки sklearn; в моделях использовались установленные по умолчанию параметры. Ниже в Таблице 2 приведены результаты данных экспериментов. Для удобства в таблице используются следующие сокращения: MLPClassifier – MLP, RandomForestClassifier – RFC, XGBClassifier – XGB.

Таблица 2 – Сравнение моделей, обученных на всех признаках, и моделей, обученных на 66 признаках и на 53 признаках  
Table 2 – Comparison of models trained on all features versus models trained on 66 features and on 53 features

Метрика	Количество признаков	Модель			
		MLP	SVC	RFC	XGB
F1	201	0,9974	0,9989	0,9989	0,9963
	66	0,9995	0,9995	1,0000	0,9979
	53	0,9995	0,9995	0,9995	0,9995
Accuracy	201	0,9973	0,9989	0,9989	0,9963
	66	0,9995	0,9995	1,0000	0,9979
	53	0,9995	0,9995	0,9995	0,9995
Precision	201	0,9958	1,0000	1,0000	0,9958
	66	1,0000	1,0000	1,0000	1,0000
	53	1,0000	1,0000	1,0000	1,0000
Recall	201	0,9989	0,9979	0,9979	0,9968
	66	0,9989	0,9989	1,0000	0,9958
	53	0,9989	0,9989	0,9989	0,9989

<sup>11</sup> Scikit-Learn. Machine Learning in Python. URL: <https://scikit-learn.org/stable/> (дата обращения: 06.02.2025).

<sup>12</sup> XGBoost. URL: <https://xgboost.ai/> (дата обращения: 06.02.2025).

Из данной таблицы видно, что отбор признаков действительно повысил качество моделей МО в задаче обнаружения БПЛА на основе акустических данных. При этом удаление признаков pitch и chroma не повлияло значительно на качество моделей.

### Заключение

В процессе исследования использована методика отбора наиболее значимых признаков, включающая определение коррелирующих групп признаков и оценку их важности для задачи обнаружения БПЛА мультироторного типа. Анализ значимости акустических признаков (включая спектральные коэффициенты), получаемых с помощью библиотек Librosa и Essentia, показывает, что:

- без потери качества можно ограничиться 66 наиболее информативными признаками из 201 исходных, обеспечивающими баланс между качеством обнаружения БПЛА и вычислительными затратами;

- наиболее требовательными к вычислительным ресурсам являются «pitch» (частота основного тона) и «chroma» (хромаграмма); удаление этих признаков уменьшает общее количество признаков до 53 и снижает значение F1-меры менее, чем на 0,1 пункта при использовании тестовых классификаторов MLP, SVC, RFC, XGB.

Исследование подтверждает перспективность акустических методов обнаружения БПЛА на основе МО. Оптимизированный набор из 53 акустических признаков обеспечивает эффективное решение задачи выявления дронов при приемлемой вычислительной нагрузке. Следует отметить, что использованные классификаторы являются «тестовыми», применены для оценки значимости акустических признаков, и в процессе разработки систем обнаружения БПЛА могут быть заменены на более эффективные в зависимости от вычислительной мощности системы.

### СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Seidaliyeva U., Iipbayeva L., Taissariyeva K., Smailov N., Matson E.T. Advances and Challenges in Drone Detection and Classification Techniques: A State-of-the-Art Review. *Sensors*. 2023;24(1). <https://doi.org/10.3390/s24010125>
2. Lee H., Han S., Byeon J.-I., et al. CNN-Based UAV Detection and Classification Using Sensor Fusion. *IEEE Access*. 2023;11:68791–68808. <https://doi.org/10.1109/ACCESS.2023.3293124>
3. Tejera-Berengue D., Zhu-Zhou F., Utrilla-Manso M., Gil-Pita R., Rosa-Zurera M. Analysis of Distance and Environmental Impact on UAV Acoustic Detection. *Electronics*. 2024;13(3). <https://doi.org/10.3390/electronics13030643>
4. Patel K., Ramirez L., Canales D., Rojas E. Unmanned Aerial Vehicles Detection Using Acoustics and Quantum Signal Processing. In: *2024 AIAA Science and Technology Forum and Exposition, 08–12 January 2024, Orlando, FL, USA*. American Institute of Aeronautics and Astronautics; 2024. <https://doi.org/10.2514/6.2024-1740>
5. Taha B., Shoufan A. Machine Learning-Based Drone Detection and Classification: State-of-the-Art in Research. *IEEE Access*. 2019;7:138669–138682. <https://doi.org/10.1109/ACCESS.2019.2942944>
6. Najafi Ja., Mirzakuchaki S., Shamaghdari S. Autonomous Drone Detection and Classification Using Computer Vision and Prony Algorithm-Based Frequency Feature Extraction. *Journal of Intelligent & Robotic Systems*. 2025;111(1). <https://doi.org/10.1007/s10846-024-02216-x>
7. Zhang Yi.D., Xiang X., Li Yi, Chen G. Enhanced Micro-Doppler Feature Analysis for Drone Detection. In: *2021 IEEE Radar Conference (RadarConf21), 07–14 May 2021*,

- Atlanta, GA, USA. IEEE; 2021. P. 1–4. <https://doi.org/10.1109/RadarConf2147009.2021.9455228>
8. Souli N., Theodorou I., Kolios P., Ellinas G. Detection and Tracking of Rogue UASs Using a Novel Real-Time Passive Radar System. In: *2022 International Conference on Unmanned Aircraft Systems (ICUAS), 21–24 June 2022, Dubrovnik, Croatia*. IEEE; 2022. P. 576–582. <https://doi.org/10.1109/ICUAS54217.2022.9836054>
  9. McCoy J., Rawat D.B. Optimized Machine Learning Based Multimodal UAV Detection Using Ensemble Stacking. In: *2024 IEEE 6<sup>th</sup> International Conference on Cognitive Machine Intelligence (CogMI), 28–31 October 2024, Washington, DC, USA*. IEEE; 2024. P. 40–49. <https://doi.org/10.1109/CogMI62246.2024.00016>
  10. Zahid Rao A., Shahid Siddique S., Danish Mujib M., Abul Hasan M., Alokaily A.O., Tahira T. Sensor Fusion and Machine Learning for Seated Movement Detection with Trunk Orthosis. *IEEE Access*. 2024;12:41676–41687. <https://doi.org/10.1109/ACCESS.2024.3377111>
  11. Wang Ye, Chen Yu., Choi J., Kuo C.-C.J. Towards Visible and Thermal Drone Monitoring with Convolutional Neural Networks. *APSIPA Transactions on Signal and Information Processing*. 2019;8(1). <https://doi.org/10.1017/ATSIP.2018.30>
  12. Guo Ju., Ahmad I., Chang K. Classification, Positioning, and Tracking of Drones by HMM Using Acoustic Circular Microphone Array Beamforming. *EURASIP Journal on Wireless Communications and Networking*. 2020;2020(1). <https://doi.org/10.1186/s13638-019-1632-9>
  13. Diao Yu., Zhang Yi., Zhao G., Khamis M. Drone Authentication via Acoustic Fingerprint. In: *ACSAC '22: Proceedings of the 38<sup>th</sup> Annual Computer Security Applications Conference, 05–09 December 2022, Austin, TX, USA*. New York: Association for Computing Machinery; 2022. P. 658–668. <https://doi.org/10.1145/3564625.3564653>
  14. Deleforge A., Carlo D.D., Strauss M., Serizel R., Marcenaro L. Audio-Based Search and Rescue with a Drone: Highlights from the IEEE Signal Processing Cup 2019 Student Competition. *IEEE Signal Processing Magazine*. 2019;36(5):138–144. <https://doi.org/10.1109/MSP.2019.2924687>
  15. Marple S.L., Jr. *Digital Spectral Analysis*. Mineola, New York: Dover Publications; 2019. 432 p.
  16. Haykin S., Liu K.J.R. *Handbook on Array Processing and Sensor Networks*. Hoboken: John Wiley & Sons; 2009. 924 p.
  17. Flanagan J.L. *Speech Analysis Synthesis and Perception*. Berlin, Heidelberg: Springer; 1972. 446 p. <https://doi.org/10.1007/978-3-662-01562-9>
  18. O’Shaughnessy D. *Speech Communication: Human and Machine*. Reading: Addison-Wesley; 1990. 548 p.
  19. Traunmüller H. Analytical Expressions for the Tonotopic Sensory Scale. *The Journal of the Acoustical Society of America*. 1990;88(1):97–100.
  20. Van Gisbergen J.A.M., Grashuis J.L., Johannesma P.I.M., Vendrik A.J.H. Neurons in the Cochlear Nucleus Investigated with Tone and Noise Stimuli. *Experimental Brain Research*. 1975;23(4):387–406. <https://doi.org/10.1007/BF00238022>
  21. Davis S., Mermelstein P. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. 1980;28(4):357–366. <https://doi.org/10.1109/TASSP.1980.1163420>
  22. Xu M., Duan L.-Yu, Cai J., Chia L.-T., Xu Ch., Tian Q. HMM-Based Audio Keyword Generation. In: *Advances in Multimedia Information Processing – PCM 2004: 5<sup>th</sup> Pacific Rim Conference on Multimedia: Proceedings: Part III, 30 November – 03 December*

- 2004, Tokyo, Japan. Berlin, Heidelberg: Springer; 2004. P. 566–574. [https://doi.org/10.1007/978-3-540-30543-9\\_71](https://doi.org/10.1007/978-3-540-30543-9_71)
23. Qi J., Wang D., Xu J., Tejedor J. Bottleneck Features Based on Gammatone Frequency Cepstral Coefficients. In: *INTERSPEECH 2013: 14<sup>th</sup> Annual Conference of the International Speech Communication Association, 25–29 August 2013, Lyon, France*. ISCA; 2013. P. 1751–1755. <https://doi.org/10.21437/Interspeech.2013-435>
24. Bartsch M.A., Wakefield G.H. Audio Thumbnailing of Popular Music Using Chroma-Based Representations. *IEEE Transactions on Multimedia*. 2005;7(1):96–104. <https://doi.org/10.1109/TMM.2004.840597>
25. Müller M., Kurth F., Clausen M. Audio Matching via Chroma-Based Statistical Features. In: *ISMIR 2005: 6<sup>th</sup> International Conference on Music Information Retrieval: Proceedings, 11–15 September 2005, London, UK*. 2005. P. 288–295. <https://doi.org/10.5281/zenodo.1416799>
26. Jiang D.-N., Lu L., Zhang H.-J., Tao J.-H., Cai L.-H. Music Type Classification by Spectral Contrast Feature. In: *IEEE International Conference on Multimedia and Expo: Proceedings, 26–29 August 2002, Lausanne, Switzerland*. IEEE; 2002. P. 113–116. <https://doi.org/10.1109/ICME.2002.1035731>
27. De Cheveigné A., Kawahara H. YIN, a Fundamental Frequency Estimator for Speech and Music. *The Journal of the Acoustical Society of America*. 2002;111(4):1917–1930.
28. Klapuri A. Qualitative and Quantitative Aspects in the Design of Periodicity Estimation Algorithms. In: *2000 10<sup>th</sup> European Signal Processing Conference, 04–08 September 2000, Tampere, Finland*. IEEE; 2000. P. 1–4.

#### ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

**Прозоров Дмитрий Евгеньевич**, доктор технических наук, профессор, профессор Вятского государственного университета, Киров, Российская Федерация.

*e-mail*: [prozorov.de@gmail.com](mailto:prozorov.de@gmail.com)

ORCID: [0000-0002-3577-8838](https://orcid.org/0000-0002-3577-8838)

**Dmitriy E. Prozorov**, Doctor of Engineering Science, Professor, Professor of Vyatka State University, Kirov, the Russian Federation.

**Бызов Виктор Александрович**, кандидат физико-математических наук, доцент Вятского государственного университета, Киров, Российская Федерация

*e-mail*: [vbyzov@yandex.ru](mailto:vbyzov@yandex.ru)

ORCID: [0000-0002-3613-5949](https://orcid.org/0000-0002-3613-5949)

**Viktor A. Byzov**, Candidate of Physical and Mathematical Sciences, Docent of Vyatka State University, Kirov, the Russian Federation.

**Мышкин Роман Евгеньевич**, заместитель генерального директора по научно-техническому развитию, Научно-исследовательский институт средств вычислительной техники, Киров, Российская Федерация.

*e-mail*: [myshkin.r@niisvt.ru](mailto:myshkin.r@niisvt.ru)

**Roman E. Myshkin**, Deputy CEO for Research and Technical Development, Scientific Research Institute of Computer Engineering, Kirov, the Russian Federation.

*Статья поступила в редакцию 13.05.2025; одобрена после рецензирования 17.06.2025; принята к публикации 23.06.2025.*

*The article was submitted 13.05.2025; approved after reviewing 17.06.2025; accepted for publication 23.06.2025.*