

# ПОДХОД К ФОРМИРОВАНИЮ ВЕКТОРА ПРИЗНАКОВ ДЛЯ АЛГОРИТМА ФОРМИРОВАНИЯ ЦИФРОВЫХ ОТПЕЧАТКОВ АУДИОФАЙЛОВ

## AN APPROACH FOR CALCULATION OF A FEATURE VECTOR FOR AUDIO FINGERPRINTING TECHNIQUES

**A. Mansurov  
P. Ladygin**

*Summary.* Audio fingerprinting is the essential technique for automatic music identification and monitoring for unauthorized use of music content. The complexity of audio fingerprinting techniques lies in providing the identification tokens with enough discriminative power, distortion invariance, and computational simplicity. The most complex step is retrieving the discriminative features (or feature vectors) from the processed audio file. In this paper, an approach for calculation of a feature vector based on musical notation of a musical composition is proposed. The detailed description of the proposed technique and testing results on simple and complex music pieces are presented. The proposed approach seems to be insensitive to various alterations of musical compositions.

*Keywords:* audio fingerprinting, audio files, spectrogram, identification.

**Мансуров Александр Валерьевич**

*К.т.н., доцент, ФГБОУ ВО «Алтайский  
государственный университет»  
mansurov.alex@gmail.com*

**Ладыгин Павел Сергеевич**

*Преподаватель, ФГБОУ ВО «Алтайский  
государственный университет»  
pavel-ladygin@yandex.ru*

*Аннотация.* Цифровые отпечатки аудиофайлов являются незаменимым средством для идентификации музыкального контента и выявления нелегального использования музыкальных произведений. Процесс их формирования на основе результатов обработки аудиофайла является одним из самых сложных моментов во всей технологической цепочке создания отпечатка. В публикации предлагается подход к созданию основы цифрового отпечатка — «вектора признаков» на основе нотного представления музыкального произведения. Приводится описание предлагаемого способа и результаты его апробации при обработке примитивных и сложных музыкальных композиций. Предложенный способ характеризуется устойчивостью к различным модификациям композиций — изменению скорости исполнения, переходам к другим октавам и добавлению дополнительных музыкальных инструментов.

*Ключевые слова:* цифровой отпечаток, аудиофайлы, спектрограмма, идентификация.

## Введение

**Т**ехнологии формирования цифровых отпечатков аудиофайлов являются важными составляющими систем автоматической идентификации аудиопроизведений, которые могут быть использованы для мониторинга эфира радиостанций, организации музыкальных библиотек, а также выявления случаев нелегального использования музыкального контента, незаконного тиражирования аудиозаписей и иных случаев нарушения прав на интеллектуальную собственность. Среди наиболее известных подходов можно выделить способ, предложенный J. Haitsma и T. Kalker (Philipps) [1], технологию Shazam [1,2], а также способ, используемый компанией Google в своих сетевых сервисах (например, Youtube) [7].

В рассмотренной технологической цепочке работы с цифровыми отпечатками одним из самых сложных и комплексных процессов является сам процесс формирования цифрового отпечатка на основе обработанного музыкального фрагмента [8]. Сформированный отпечаток должен быть достаточно компактным для его удобного хранения и быстрого извлечения из базы, обладать инвариантностью по отношению к различным вноси-

мым искажениям и модификациям в аудиофрагмент, обладать характеристиками для гарантированного различения или обнаружения схожести с другими отпечатками, а его алгоритм вычисления не должен требовать больших вычислительных мощностей. Процесс формирования отпечатка схематично можно представить следующим образом (рис. 1):

Учитывая предъявляемые требования и сложность процесса формирования отпечатка, исследователями были разработаны и предложены множество различных подходов и способов к построению цифровых отпечатков для аудио композиций. Наиболее часто используются энергетические спектральные характеристики (Фурье-спектр и аналогичные) [4,9], мел-частотные кепстральные коэффициенты [10], характеристики вейвлет-спектра [7]. Последующие действия заключаются в обработке результатов выполненных преобразований и формирования т.н. уникального «вектора признаков» или хеш-строки, которые и будут являться цифровыми отпечатками. Сам процесс обработки во многом определяет новаторство и уникальность формируемой технологии, поскольку на данном этапе происходит синтез уникальных характеристик, присущих каждой аудио

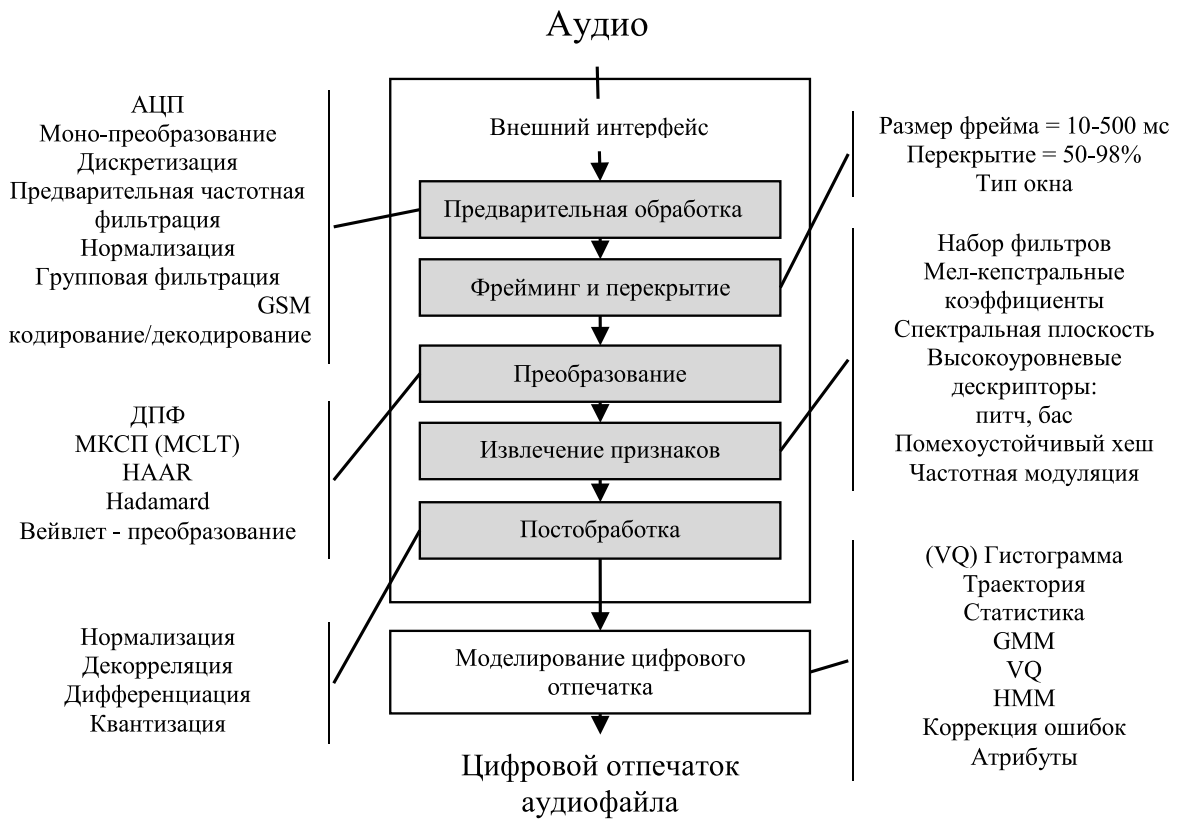


Рис. 1. Процесс формирования цифрового отпечатка аудиофайла [8]

композиции. При этом используются либо известные особенности работы эксплуатируемых алгоритмов или процессов обработки, либо предлагается особенный эвристический подход, связанный с особенностями самой аудиозаписи [11,12].

Использование особенностей аудио композиций при формировании «вектора признаков» в алгоритмах цифровых отпечатков является закономерным ответом на различные способы принудительного изменения композиции с помощью ускорения, замедления на доли секунды, или смены высоты тона. Такая модификация не заметна для человеческого уха, однако, вычисление цифрового отпечатка приводит к его другому значению относительно цифрового отпечатка первоначальной версии аудиофайла [13]. Согласно исследованию [14], наиболее важными в процессе восприятия и идентификации музыкальной композиции человеком являются интервалы и переходы между нотами, длительность, тональность и интенсивность звучания нот, а также «созвучие» — т.е. соседство нот между собой. Эффективность учета подобных характеристик в способах идентификации аудио композиций [11,12,15,16] позволяет сделать вывод о перспективности подобных

подходов и необходимости дополнительного изучения способов оптимизации и повышения качества формирования «вектора признаков» или хеш-строки, на основании которой в дальнейшем будет строиться цифровой отпечаток.

В данной работе предлагается способ формирования «вектора признаков» для технологии получения цифровых отпечатков аудиофайлов, основанный на особенностях восприятия музыкальных композиций человеком и нотном представлении любого музыкального произведения. В работе изложен сам подход, приведены схема работы предлагаемого алгоритма и результаты тестирования на нескольких музыкальных фрагментах.

#### Способ формирования «вектора признаков»

В предлагаемом способе мелодия представляется важнейшим, первичным компонентом музыкальной композиции. Поскольку любая мелодия может быть представлена в виде нотной записи, описываемый способ предлагает выполнение перехода к нотному пред-

Таблица 1. Соответствие «название ноты» — «частота звучания»

		Нота						
		До	Ре	ми	фа	Соль	ля	си
Частота, Гц	1 октава	261,63	293,66	329,63	349,23	392	440	493,88
	2 октава	523,25	587,32	659,26	698,46	784	880	987,75



Рис. 2. Нотная запись мелодии «В траве сидел кузничек»

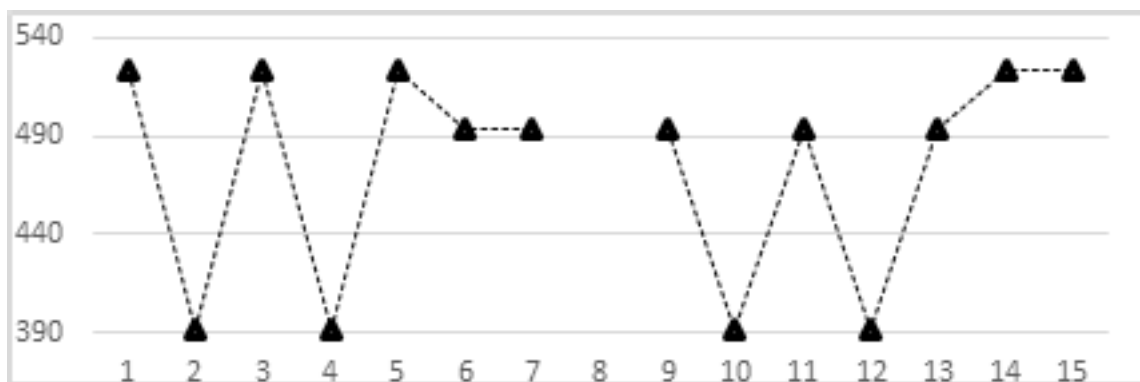


Рис. 3. Разложение мелодии по частотам.

ставлению и последующую работу при построении «вектора признаков» с частотной сеткой, соответствующей частоте каждой чистой ноты своей октавы. В этом случае используемая частотная сетка будет отличаться от частотных «гребенков», используемых при обработке спектра аудио композиций в работах [4–6,9].

На рис. 2. представлена нотная запись фрагмента мелодии «В траве сидел кузничек». Таблица 1 содержит соответствие нот, присутствующих в мелодии, частотам, на которых они звучат. На рис. 3 представлено разложение мелодии по частотам, где по шкале Y отмечена частота звука, по оси X отмечены отсчеты времени.

Переходы между частотами (нотами) предлагается использовать в качестве основы для формирования «вектора признаков». Переход от более высокой частоты к более низкой («сверху-вниз») или отсутствие такого перехода кодируется нулём, а переход от более низкой частоты к более высокой («снизу-вверх») — единицей. Создаваемая битовая последовательность при этом должна оставаться инвариантной по отношению к способу исполнения (используемому музыкальному инструменту), а также к ускорению или замедлению аудиофайла, поскольку факт перехода от частоты к частоте при такой модификации останется без изменений.

В основе работы предлагаемого способа лежит Фурье – спектр анализируемой аудио композиции. Спектр обрабатывается для выделения чистых тонов, соответствующих частотам нот в своих октавах. Процесс обработки базируется на алгоритмах распознавания нотных последовательностей, изложенных в работах [17,18]. Схематичная иллюстрация процесса приведена на рис. 4 и описывается следующим образом:

1. На вход подаётся wav-file (фрагмент музыкальной композиции).

2. Вычисление дискретного преобразования Фурье для каждого участка мелодии, равного ширине окна, полученного в предыдущем пункте. Число N — это количество разбиений мелодии на фрагменты. Вычисляется по формуле:

$$N = \frac{T \cdot F_d}{W} \quad (1)$$

где  $T$  — длительность мелодии в секундах,  $F_d$  — частота дискретизации аудиофайла,  $W$  — ширина окна ПФ.  $N$  округляется в большую сторону.

3. Поиск гармоник с максимальной амплитудой (чистый тон) в каждом участке мелодии.

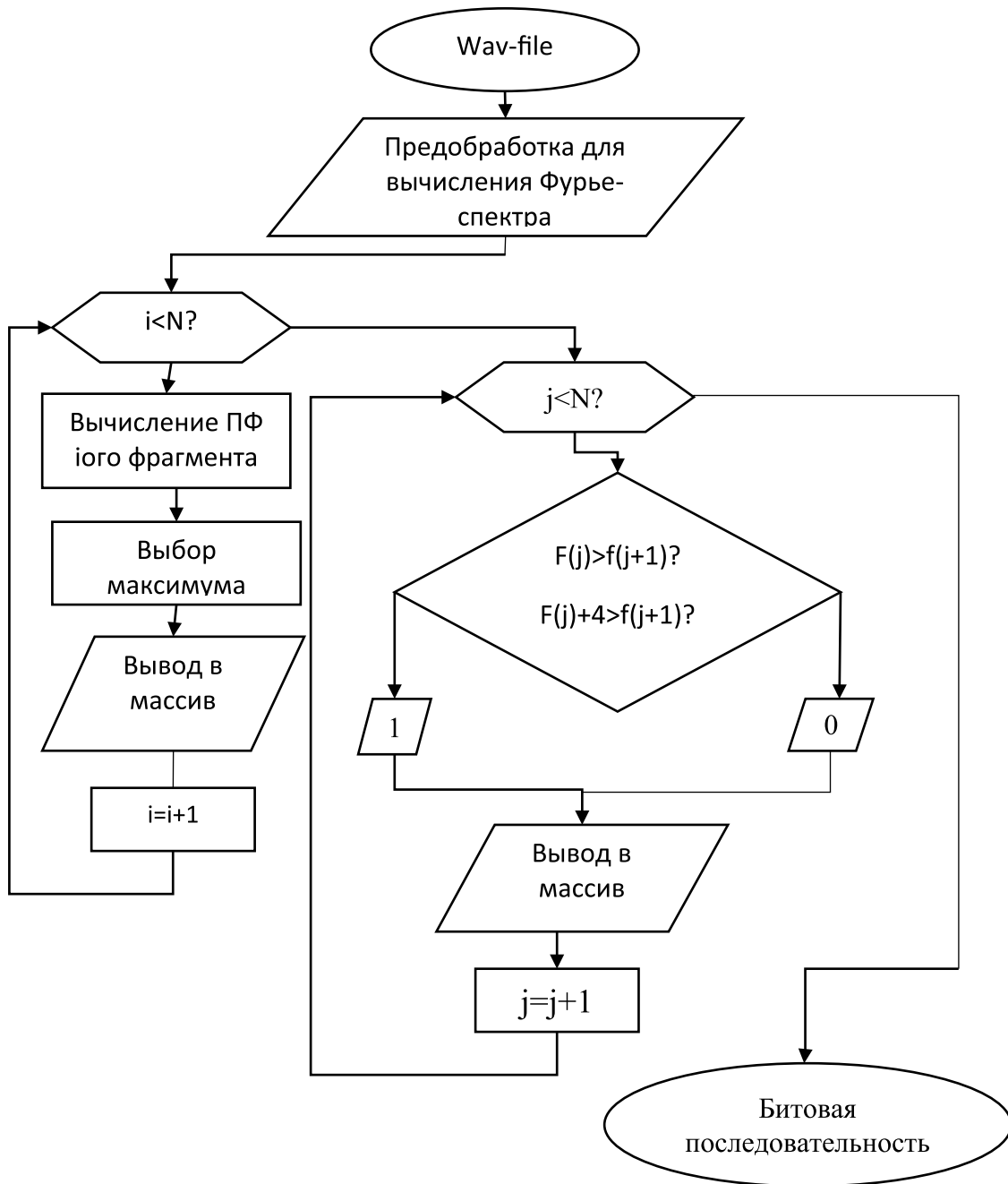


Рис. 4. Схематичная иллюстрация формирования битовой последовательности

4. Выявление переходов между частотами гармоник чистого тона. При этом сравнение двух ближайших частот происходит с некоторым перекрытием.

5. Вывод в массив битовой последовательности.

Данный алгоритм реализован в среде графического программирования Labview, поскольку пакет специально разработан для программирования различных измерений, анализа данных и оформления результатов. Алгоритм анализа переходов нотной последова-

тельности в среде LabVIEW [19] представлен на рис. 5 и рис. 6.

Код, написанный в графической среде разработки Labview полностью повторяет вышеизложенный алгоритм. На рис. 6. изображён участок программы, осуществляющий Преобразование Фурье с выбором гармоники с максимальной амплитудой.

Сформированная битовая последовательность является текущим «вектором признаков», характеризую-

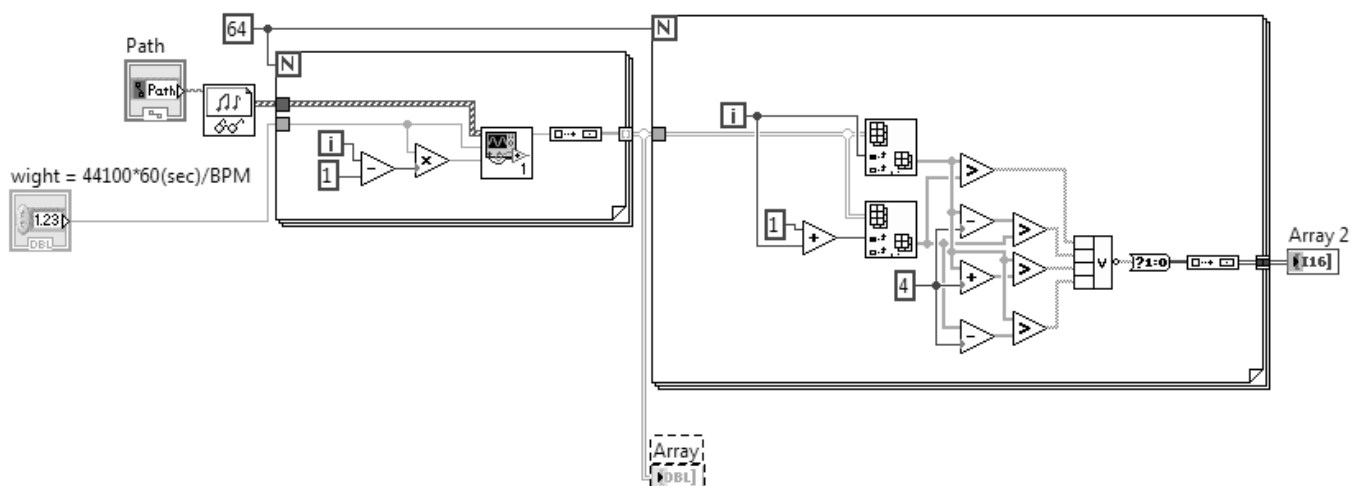


Рис. 5. Алгоритм анализа переходов нотной последовательности в среде LabVIEW

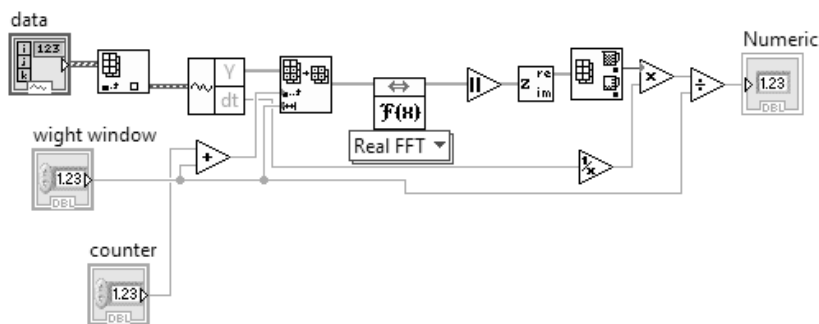


Рис. 6. Алгоритм выделения частоты ноты

щим аудио композицию. Текущий «вектор признаков» в дальнейшем может быть подвергнут постобработке для оптимизации использования в последующих процессах хранения, поиска и сравнения с аналогичными векторами. Сам процесс сравнения может базироваться на любом алгоритме или любой методике. В данной работе для сравнения используется простая функция сравнения, для определения схожести двух векторов между собой:

$$Q = 100\% - \frac{N1(O1 \wedge O2)}{N0} * 100 \quad (2)$$

где  $Q$  — степень схожести,  $O_1, O_2$  — битовые последовательности отпечатков,  $N_1()$  — количество единиц в результате операции XOR,  $N_0$  — длина отпечатков.

#### Тестирование предлагаемого способа

Для тестирования предлагаемого подхода используются монофонические записи музыкальных инструментов, исполняющих мелодию «В траве сидел кузнечик», с частотой дискретизации 44100 Гц, сохранённые в формате.wav. В тестировании участвовали три варианта

исполнения мелодии: Piano4–5окт — звук синтезатора, воспроизводящий мелодию в 4ой и 5 ой октаве, Piano2–3окт — звук синтезатора, воспроизводящий мелодию во 2ой и 3ей октаве, GuitarMalayaOkt — звук гитары, воспроизводящий мелодию в Малой октаве. Анализ спектров аудиофайлов, на которых записана одна и та же мелодия, но в разных октавах, показывает, что при более низких частотах прослеживается больше тембровых характеристик музыкального инструмента (обертонов). Стандартные методики формирования цифровых отпечатков воспримут данные аудиофайлы как содержащие различные мелодии, принадлежащие разным правообладателям.

Результат тестирования предложенного подхода к формированию «вектора признаков» представлен на рис. 7:

Результат тестирования демонстрирует работоспособность предлагаемого подхода, а также позволяет убедиться в том, что не имеет значения, на каком инструменте и в каком диапазоне частот исполнена анализируемая мелодия. Замедление/растяжение всей мелодии

Таблица 2. Битовые последовательности тестируемого участка

№	Музыкальные инструменты	Битовые последовательности / «вектор признаков»
1.	гитара	D4206452508C0C392248488A8652A222098898104E45048B38940A28C514B22522208104120D29C28458932046A59C1AC912ED4A9D32C92244516B42B2594350996C18425D44A201D4891549C0909248B54604AAB09224E04CA1052292466085480001
2.	Гитара+перкуссия	D4206452508C0C392248688A8652A422098898104E45048B38940A28C514B22522208504120D29C2A458932246A59C2AC892AD4A9532D12144D16B4292594350986C18423D44A201D4891549C0905248B34204AAB11224E04CA1052252C46085080001
3.	Гитара+синтезатор	D4206452508C0C392248488A8652A2220C88A8108E56248D21148A202614CA89512A8848CC984AB081A2954282A6A10AC210025214565068744A98514A8249A28CD5485301452A0424512890D095AA4B840204AAB09224E04CA1052466085480001
4.	«Monday» Ludovico Einaudi	202405C40080119090A20048C82000080104460810004114000483004040A261200C200000300520001100120011085812004912080215AC009206881020488840082000000802106431041802088210201000090211081112002092041001A102211428224128

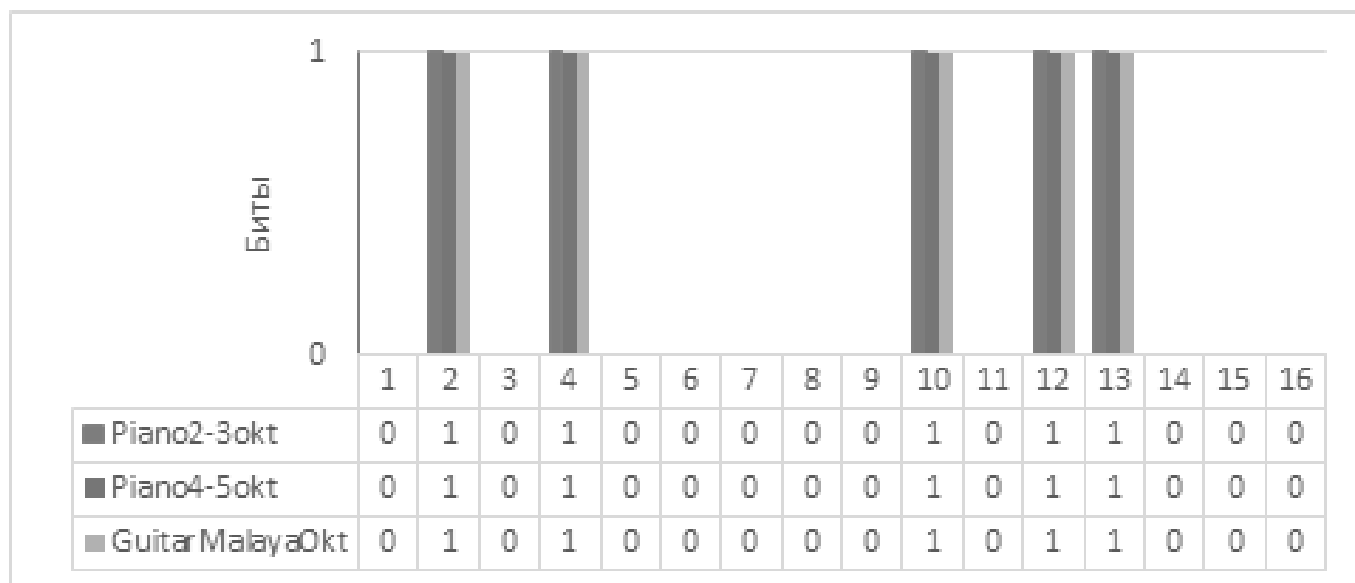


Рис. 7. Битовые последовательности мелодии «В траве сидел кузнечик».

также не влияет на результат, т.к. факт перехода между частотами основных тонов останется неизменным, лишь изменится ширина окна для преобразования Фурье.

Тестирование предлагаемого подхода при анализе более сложной композиции выполняется на основе инструментальной композиции группы «Горизонталь» — «Мы», в которой звучат перкуSSIONные инструменты,

акустическая гитара и синтезатор, а также композиции Ludovico Einaudi «Monday», исполненной на синтезаторе. Усложнение первой композиции производится путем поэтапного наложения музыкальных партий друг на друга в программе Reaper [20]. Сравнение осуществляется при помощи функции (2). Полученные битовые последовательности и результаты сравнения приведены в Таблице 2.

Таблица 3. Сравнительный анализ предлагаемого способа и особенностей работы популярных технологий Shazam и Philips

	Технология Shazam	Технология Philips	Предложенный способ
Отпечаток зависит от ширины окна преобразования Фурье	нет	да	да
Изменяется после растяжения/ускорения мелодии	да	да	нет
Изменяется после транспонирования мелодии в другую октаву	да	да	нет

Степень схожести между 1 и 2 «вектором признаков»: 93,2%. ПеркуSSIONные инструменты можно относить к шуму, поскольку перкуSSIONные партии не относят к мелодии. Тем не менее, перкуSSION вносит коррективы в спектр, но их влияние достаточно невелико для того, чтобы сделать вывод о том, что аудиофайлы имеют высокую степень схожести.

Степень схожести между 1 и 3 «вектором признаков»: 77,3%. Синтезатор — музыкальный инструмент, имеющий широкий спектр. На этом инструменте зачастую играют основные мелодии музыкальных произведений, поэтому вносимые им коррективы в цифровой отпечаток столь существенны. Необходимо дополнительное изучение аудиофайлов для принятия решения о наличии плагиата в таких файлах.

Степень схожести между 2 и 3 «вектором признаков»: 77,1%. В данном случае и предыдущем нужен дополнительный анализ, поскольку при отсутствии Базы Данных отпечатков и накопленного опыта невозможно установить какой процент считать определяющим в задачах идентификации аудиофайлов. Тем не менее, процент схожести достаточно высок, чтобы судить о том, что аудиофайлы содержат схожие музыкальные дорожки.

Сравнение «вектора признаков», полученного для произведения «Monday» и «вектора признаков» произведения «Мы» под номером 3 выполняется с условием, что композиции имеют разную длительность. Для сравнения векторов разной длины необходимо рассматривать одну композицию как часть другой и искать наибольшую схожесть во всех возможных наложениях. В связи этим композиция «Monday» укорачивается до длительности композиции «Мы». При таком условии степень схожести составляет 25%, что говорит о том,

что произведения с высокой степенью вероятности различны и предлагаемый подход работоспособен и в случаях сравнения абсолютно разных музыкальных композиций.

Сравнительный анализ предлагаемого способа и особенностей работы популярных технологий Shazam и Philips приведены в Таблице 3.

Предложенный способ получения «вектора признаков» аудиофайла может служить дополнительным инструментом в задачах выявления не уникального контента в сети. В ходе тестирования на примере простой мелодии, исполненной разными музыкальными инструментами, было показано гарантированное совпадение их битовых последовательностей, что характеризует разработанный способ как устойчивый к существованию возможных интерпретаций одного и того же произведения. Данное преимущество отличает разработанный способ от известных методик получения «вектора признаков» аудиофайлов.

### Заключение

В статье предлагается подход к формированию «вектора признаков», который получается в результате обработки аудиофайла и является основой для генерирования цифрового отпечатка данного аудиофайла. Разработанный подход базируется на том, что большинство регистрируемых музыкальных объектов интеллектуальной собственности являются в своей основе мелодией. Такая мелодия представляется как набор пауз и звуков, каждые из которых характеризуется спектрами, получаемыми с помощью преобразования Фурье. Анализ спектра позволяет находить частоту с максимальной амплитудой на каждом фрагменте мелодии, которая в большинстве случаев явля-

ется «чистым тоном» и является табличным значением, имеющим соответствующее нотное обозначение. Кодирование переходов между такими частотами в битовую последовательность позволяет получить «вектор признаков» для аудиофайла.

Выполненное тестирование предложенного подхода показало его убедительную эффективность в случае анализа простых мелодий. Применение простой функции сравнения двух векторов позволяет выполнять идентификацию сложных композиций с точностью более 77%. При этом сравнение двух разных композиций демонстрирует их совпадение на уровне

26%. Применение более эффективного метода для выполнения сравнения в данной публикации не рассматривается и является вопросом для последующих исследований.

Данный подход наиболее полезен при регистрации авторских прав на отдельные мелодии, что позволит правообладателям перед опубликованием проверять свой труд на наличие непреднамеренного плагиата. Помимо этого, предлагаемый способ может быть включен в существующие методики, предложенные в работах [4] и [5], для повышения точности идентификации аудиофайлов в различных Интернет-сервисах.

## ЛИТЕРАТУРА

1. Shum S. The Basics of Audio Fingerprinting [Электронный ресурс] / MIT Computer Science and Artificial Intelligence Laboratory. URL: [http://people.csail.mit.edu/sshum/talks/audio\\_fingerprinting\\_sls\\_24Oct2011.pdf](http://people.csail.mit.edu/sshum/talks/audio_fingerprinting_sls_24Oct2011.pdf) (дата обращения 25.06.2017).
2. Raphi Z. Audio Fingerprinting. [Электронный ресурс] / Zafar Rafii. URL: <http://www.zafarrafii.com/doc/Rafii%20-%20Audio%20Fingerprinting%20-%20NU%20EECS%20352%202014.pdf> (дата обращения 25.06.2017).
3. Goenka K. V., Patil P. K. Overview of Audio Watermarking Techniques // International Journal of Emerging Technology and Advanced Engineering. February 2012. Vol.2, Issue 2, pp. 67–70.
4. Haitsma J, Kalker T. A Highly Robust Audio Fingerprinting System / Journal of New Music Research, Vol. 32(2003), No. 2, p. 211–222.
5. Wang, Avery Li-Chun. An Industrial-Strength Audio Search Algorithm / [Электронный ресурс] / CSIE. URL: <http://www.csie.ntu.edu.tw/~r95162/An%20Industrial-Strength%20Audio%20Search%20Algorithm.pdf> (дата обращения 25.06.2017).
6. Van Nieuwenhuizen H.A, Venter W.C, Grobler M. J. The Study and Implementation of Shazam's Audio Fingerprinting Algorithm for Advertisement Identification [Электронный ресурс] / SATNAC, September 2012. URL: <http://www.satnac.org.za/proceedings/2011/papers/Software/181.pdf> (дата обращения 25.06.2017).
7. Эволюция Content ID: как Youtube совершенствует свою самую спорную функцию [Электронный ресурс] / Air. URL: <http://www.air.io/content-id-evolution/> (дата обращения 22.03.2017).
8. Cano P., Batlle E., Kalker T., Haitsma J. A review of audio fingerprinting // The Journal of VLSI Signal Processing, 2005. Vol.41, pp. 271–284.
9. Baluja S., Covell M. Content fingerprinting using wavelets, Proc // 3rd European Conference on Visual Media Production (CVMP 2006). Part of the 2nd Multimedia Conference 2006, 2006, pp. 198–207. DOI: 10.1049/cp:20061964
10. Sonnleitner R. Widmer G. Robust quad-based audio fingerprinting. IEEE/ACM Trans. Audio, Speech and Lang. Proc. 24, 3 (March 2016), pp. 409–421. DOI=<http://dx.doi.org/10.1109/TASLP.2015.2509248>.
11. Chen N., Xiao H. D., Zhu H. Robust audio fingerprinting based on GammaChirp frequency cepstral coefficients and chroma // in Electronics Letters, vol. 50, no. 4, pp. 241–242, February 13 2014. doi: 10.1049/el.2013.3554.
12. Malekesmaeili, Mani; Ward, Rabab K. A local fingerprinting approach for audio copy detection // Signal Processing, 2014, V. 98, pp.308–321.
13. Marolt M. A Mid-Level Representation for Melody-Based Retrieval in Audio Collections // in IEEE Transactions on Multimedia, vol. 10, no. 8, pp. 1617–1625, Dec. 2008. doi: 10.1109/TMM.2008.2007293.
14. Schulkind M. D., Posner R. J., Rubin D. C. Musical features that facilitate melody identification: How do you know it's «your» song when they finally play it? // Music Perception, 21, 217–249.
15. Serra J., Gomez E., Herrera P., Serra X. Chroma Binary Similarity and Local Alignment Applied to Cover Song Identification // in IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 6, pp. 1138–1151, Aug. 2008. doi: 10.1109/TASL.2008.924595.
16. Ning Hu, Dannenberg R. B., Tzanetakis G. Polyphonic audio matching and alignment for music retrieval // 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684), 2003, pp. 185–188. doi: 10.1109/ASPAA.2003.1285862.
17. Fujishima, T. Realtime Chord Recognition of Musical Sound: A System Using Common Lisp Music // In Proceedings of the International Computer Music Conference, Beijing, China, 22–27 October 1999; pp. 464–467.
18. Poliner G. E., Ellis D. P. W., Ehmann A. F., Gomez E., Streich S. Ong B. Melody Transcription From Music Audio: Approaches and Evaluation // in IEEE Transactions on Audio, Speech, and Language Processing, May 2007. Vol. 15(4), pp. 1247–1256. doi: 10.1109/TASL.2006.889797.
19. LabVIEW // [Электронный ресурс]. URL: <http://labview.ru> (дата обращения 07.05.2016)
20. Reaper // [Электронный ресурс]. URL: <https://www.reaper.fm/> (дата обращения 07.05.2017)