

УДК 004.056

МЕТОДИКА СНИЖЕНИЯ ВЛИЯНИЯ МАЛОИНФОРМАТИВНЫХ УЧАСТКОВ АУДИОФАЙЛОВ НА ПОЛУЧАЕМЫЙ ЦИФРОВОЙ ОТПЕЧАТОК

Ладыгин Павел Сергеевич, Лепендин Андрей Александрович

Алтайский государственный университет, г. Барнаул
e-mail: pavel-ladygin@yandex.ru, andrey.lependin@gmail.com

A TECHNIQUE FOR REDUCING THE INFLUENCE OF UNINFORMATIVE SECTIONS OF AUDIO FILES ON THE RESULTING DIGITAL FINGERPRINT

Ladygin Pavel S., Lependin Andrey A.

Altai State University, Barnaul

Аннотация. В данной работе предложена методика снижения влияния малоинформативных участков музыкальных аудиозаписей, содержащих мелодические конструкции, на качество цифровых отпечатков по оригинальному алгоритму. Предварительно проанализированы фрагменты хроматограмм аудиофайлов в области отсутствия полезного сигнала. Показана эффективность применения предварительной фильтрации сигналов на отрезках времени, не содержащих частоты основного тона, в отношении мелодии, исполненной на духовом музыкальном инструменте. Разработанный подход позволил сократить длину вычисляемого отпечатка, что в свою очередь приводит к ускорению работы алгоритмов сопоставления мелодических конструкций между собой. Проведена апробация методики на модификациях аудиозаписей по темпу звучания для различных музыкальных инструментов. Приведены результаты вычислений медианного значения степени схожести цифровых отпечатков до применения фильтрации и после неё, что показало улучшение качества сопоставления мелодических конструкций в среднем на 1,5%, а в отдельных случаях и до 3%. Эмпирически подобран порог фильтрации для аккордовых и безаккордовых мелодий. Данная методика может быть применена в задачах автоматического проведения экспертных оценок при установлении нарушения прав

Abstract. In this paper, a technique is proposed to reduce the influence of uninformative sections of musical audio recordings containing melodic constructions on the quality of digital prints according to the original algorithm. Fragments of chromatograms of audio files in the absence of a useful signal were previously analyzed. The effectiveness of the application of pre-filtering of signals on time intervals that do not contain the frequency of the fundamental tone in relation to the melody performed on a wind musical instrument is shown. The developed approach made it possible to reduce the length of the calculated print, which in turn leads to an acceleration of the algorithms for matching melodic constructions with each other. The methodology was tested on modifications of audio recordings according to the tempo of sound for various musical instruments. The results of calculations of the median value of the degree of similarity of digital prints before and after filtering are presented, which showed an improvement in the quality of matching melodic constructions by an average of 1.5%, and in some cases up to 3%. The filtering threshold for chord and chordless melodies is empirically selected. This technique can be applied in the tasks of automatic expert assessments when determining the violation of intellectual property rights for musical works.

Keywords: chromatogram, degree of similarity, melodic constructions, digital fingerprint, filtration.

на интеллектуальную собственность для музыкальных произведений.

Ключевые слова: хроматограмма, степень схожести, мелодические конструкции, цифровой отпечаток, фильтрация.

Для цитирования: Ладыгин П.С., Лепендин А.А. Методика снижения влияния малоинформативных участков аудиофайлов на получаемый цифровой отпечаток // Проблемы правовой и технической защиты информации. 2023. №11. С. 29-34.

For citation: Ladygin P.S., Lependin A.A. A technique for reducing the influence of uninformative sections of audio files on the resulting digital fingerprint // Legal and Technical Problems Information Protection. 2023. No. 11. P. 29-34.

В современном мире информация и данные имеют огромное значение. Они являются основой для принятия решений, улучшения качества продуктов и услуг, а также защиты интеллектуальной собственности. Один из способов идентификации цифровых данных – это цифровые отпечатки. Цифровой отпечаток аудиофайла – это уникальный набор характеристик, которые описывают аудиофайл и позволяют идентифицировать его среди других файлов [1]. Цифровой отпечаток создается с помощью специальных алгоритмов [2-6], которые анализируют различные параметры аудио, такие как частота, амплитуда, спектральные характеристики и др. в т. ч. с использованием нейронных сетей [7, 8].

Предложенная ранее методика формирования цифрового отпечатка

аудиофайла, содержащего мелодическую конструкцию [9], показала свою надежность при сравнении оригинальных экземпляров с экземплярами, подвергшимся наиболее распространенным аугментациям, используемых злоумышленниками при обходе алгоритмов популярных интернет-площадок по недопущению нарушений в отношении интеллектуальной собственности. Однако, имела свои недостатки в области сильно зашумленных участков аудиофайлов или участков, которые содержат тишину. Причиной тому является особенность построения хроматограмм библиотекой librosa, используемой для анализа аудиоданных в языке программирования Python. Пример спектрограммы такого участка аудиофайла представлен на рисунке 1.

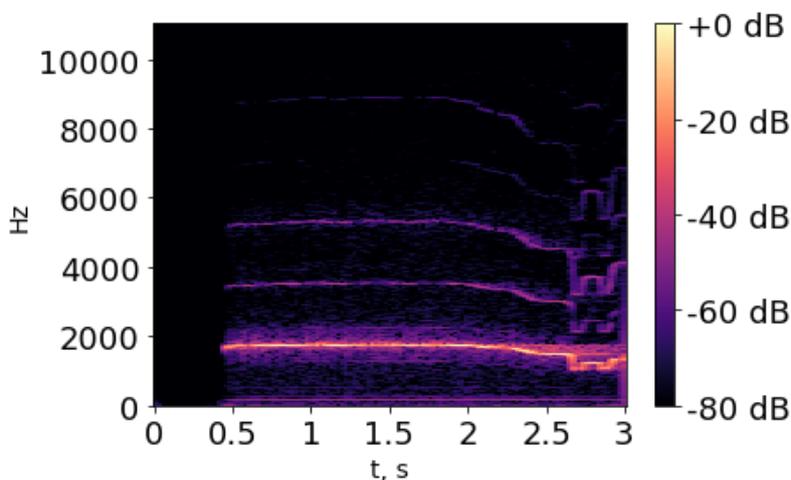


Рисунок 1. Спектрограмма рассматриваемого аудиофайла

На участке с 0–0,4 с. наблюдается фрагмент, при котором исполнитель данной композиции, сыгранной в рассматриваемом случае на флейте, берёт дыхание, что не является информативным признаком в задаче сравнения двух мелодий между

собой. При этом хроматограмма аудиофайла определяет данный участок в качестве сигнала, не имеющего однозначно определенной частоты основного тона (ноты). Хроматограмма рассматриваемого аудиофайла представлена на рисунке 2.

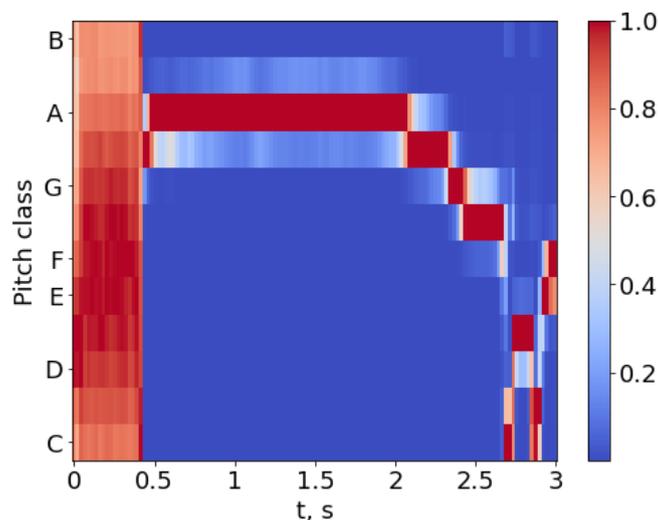


Рисунок 2. Хроматограмма рассматриваемого аудиофайла

Удаление подобных участков автоматическими средствами без предварительного анализа не представляется оптимальным, так как подобные сигналы могут обнаруживаться в любом участке мелодий, особенно исполненных на духовых инструментах. При этом, как известно, подобные сигналы (в т.ч. участки тишины) могут сами по себе являться неотъемлемой частью музыкальной композиции независимо задействованного источника сигнала. В связи с данными аргументами представляется необходимым предварительный анализ каждого

фрагмента хроматограммы на предмет интенсивности всего диапазона частот в каждый отсчёт времени перед дальнейшими действиями по формированию цифрового отпечатка.

Анализ вектора признаков рассматриваемой мелодии после спектрального фурье-преобразования и до построения хроматограмм на различных участках показывает, что в области тишины и в значимый момент времени имеют следующие значения весов цветности из набора {C, C#, D, D#, E, F, F#, G, G#, A, A#, B}, которыми обозначаются ноты в 12-ступенном звуковом ряде (таблица 1).

Таблица 1. Значения весов цветности хроматограммы в различные моменты времени

T, с	C	C#	D	D#	E	F
0,04	0.72	0.82	0.98	1	0.94	0.86
1,5	2.53×10^{-5}	2.05×10^{-5}	1.05×10^{-5}	1.37×10^{-5}	4.91×10^{-5}	3.09×10^{-5}
T, с	F#	G	G#	A	A#	B
0,04	0.78	0.71	0.65	0.62	0.58	0.63
1,5	2.51×10^{-5}	4.84×10^{-4}	1.41×10^{-1}	1	1.4×10^{-1}	4.63×10^{-4}

Результаты, приведенные в таблице 1 показывают, что «тишина» (или момент вдоха исполнителя перед взятием «ноты») на 0,04 секунде рассматриваемого аудиофайла вносит высокую неопределенность в идентификацию частоты основного тона. При этом на значимом участке мелодии (1,5 секунды от начала аудиофайла) частота основного тона, выраженная весами цветности хроматограммы, определяется однозначно и отличается от остальных значений в 101-105 раз.

Анализ всего получаемого вектора признаков из рассматриваемой мелодии показывает, что сумма весов цветности в моменты тишины находится в диапазоне от 9,3 до 10,96, в то время как в более информативных участках не превышает 3.



Рисунок 3. Индексы пиков хроматограммы до применения фильтрации

Далее к полученному массиву значений применялся оригинальный алгоритм формирования цифрового отпечатка, заключающийся в кодировании переходов от индекса к индексу по следующему правилу: переход вверх по индексу кодировался как 01, вниз – 10, отсутствие перехода игнорировалось. Получаемый цифровой отпечаток данного аудиофайла до применения фильтрации получил следующий вид: 0101 1001 1001 0110 0110 1010 0101 1010 1010 0110 0101.

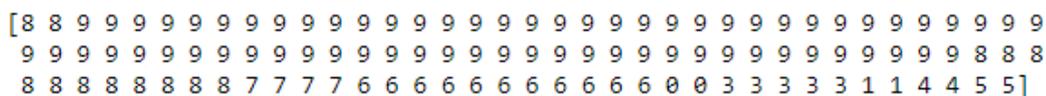


Рисунок 4. Индексы пиков хроматограммы после применения фильтрации

Получаемый цифровой отпечаток данного аудиофайла в таком случае принимает следующий вид: 01 1010 1010 0110 0101. Как видно из результата, фильтрация привела к уменьшению длины отпечатка на 26 бит, которыми в данном

Из данного анализа представляется важной предварительное исключение из хроматограммы сигнала таких участков, сумма весов цветности в каждый отсчёт времени которых превышает некоторый порог.

Для проверки данной гипотезы и подбора порога исключения малоинформативных участков были найдены индексы пиков хроматограммы рассматриваемой мелодии, содержащейся в аудиофайле длительностью 3 секунды в wav-формате (рисунок 3). Индексы пиков при этом соответствуют порядковым номерам из набора {C, C#, D, D#, E, F, F#, G, G#, A, A#, B}, которыми обозначаются ноты в 12-ступенном звуковом ряде, т. е. 0 – это нота «до», 1 – «до диез» и т.д., определенных в окне, равном 23 мс.

После применения фильтрации по порогу равному 4 количество индексов пиков хроматограммы уменьшилось, что более точно стало соответствовать содержанию частот основного тона в рассматриваемой мелодии (рисунок 4). Данный порог позволил избавиться от начальных индексов, не принадлежащих участкам основной мелодии, и подобран эмпирически.

случае характеризуется рассматриваемая мелодия, что в свою очередь ускоряет процесс сопоставления двух отпечатков для задач быстрого поиска мелодии в базах данных популярных интернет-платформ и

анализа нарушения прав на интеллектуальную собственность.

Применение рассмотренной методики фильтрации было осуществлено к разработанной базе данных аудиофайлов [10], содержащих записи мелодий, которые исполнялись на различных музыкальных

инструментах. К оригинальным аудиофайлам применялись ускорения и замедления темпа на 1%, 2%, 3% и 10%, что расширило набор данных: всего по 8 версий для каждой аудиозаписи. Структура используемого набора данных представлена в таблице 2.

Таблица 2. Структура набора данных для проверки качества работы предложенного метода

Вид инструмента	Вид мелодии	Оригинальные образцы	Образцы с изменением скорости
Духовые	без аккордов	10	80
Клавишные	без аккордов	19	152
	с аккордами	12	96
Струнные	без аккордов	6	48
	с аккордами	5	40
Всего	-	52	416

Для оценки устойчивости степени схожести аудиофайлов к изменению темпа проводилось сопоставление цифровых отпечатков каждого оригинального образца со всеми его версиями с отличающимся темпом. Степень схожести при этом рассчитывалась по следующей формуле:

$$Q(O_1, O_2) = (1 - N(O_1, O_2)/N_0) \cdot 100 \%, \quad (1)$$

где O_1, O_2 – битовые последовательности цифровых отпечатков, функция $N(,)$ вычисляла число совпадений битов во

входных последовательностях, N_0 – длина наименьшего из отпечатков.

В таблице 3 для удобства приведены медианные значения Q для различных изменений темпа в двух вариантах – до применения выше рассмотренной методики избавления мелодий от влияния тишины на вычисляемый цифровой отпечаток и после него. Эмпирически был подобран порог фильтрации, равный 5, который не допустил исключения из тестирования методики тех мелодий, которые содержали аккорды.

Таблица 3. Медианные значений степеней схожести для ускорения/замедления темпа

Изменение темпа, %	Медианное Q, % до устранения малоинформативных участков	Медианное Q, % после устранения малоинформативных участков
-10	86.5	88
-3	86.5	83
-2	85	87.5
-1	95	89
1	96	100
2	91	98
3	85	87.5
10	83	84

Из анализа таблицы видно, что фильтрация приводит к увеличению

показателя степени схожести аудиофайлов, содержащих одну и ту же мелодическую

последовательность, но воспроизводимых в различных темпах даже в отношении сложных композиций, содержащих аккорды, в среднем на 1,5%. Более высоких показателей для задач выявления фактов нарушения интеллектуальной собственности можно достичь, исключив из базы данных аудиозаписи, содержащие аккордовые последовательности, т.к. авторское право на них не распространяется.

Таким образом результаты работы демонстрируют полезный эффект предварительного устранения малоинформативных сигналов аудиофайлов от фрагментов, не содержащих частот основного тона, а также позволяют избавиться от зашумлённых участков хроматограмм для задач формирования цифровых отпечатков аудиофайлов, устойчивым к различного рода аугментациям и независимым от формата хранения самих аудиофайлов.

Исследование выполнено за счет гранта Российского научного фонда № 22–21–00199, <https://rscf.ru/project/22-21-00199/>

Библиографический список

1. Веялкин И. А., Шиман Д. В. Обзор структур алгоритмов поиска на основе отпечатков аудиоданных // Труды БГТУ. Серия 3: Физико-математические науки и информатика. 2015. №6 (179). URL: <https://cyberleninka.ru/article/n/obzor-struktur-algoritmov-poiska-na-osnove-otpechatkov-audiodannyh> (дата обращения: 01.12.2023).
2. Ладыгин П.С., Мансуров А.В., Рудер Д.Д. Сравнение векторов признаков аудиофайлов, полученных с помощью хроматограмм и пич-трекера CREPE // Проблемы правовой и технической защиты информации. 2020 № 8 С. 29–34.
3. Sonnleitner, R., & Widmer, G. (2015). Robust quad-based audio fingerprinting. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(3), 409-421.
4. Cancela, P., Rocamora, M., & López, E. (2009, October). An Efficient Multi-Resolution Spectral Transform for Music Analysis. In *ISMIR* (pp. 309-314).
5. Müller, Meinard; Kurth, Frank; Clausen, Michael (2005). "Audio Matching via Chroma-Based Statistical Features". *Proceedings of the International Conference on Music Information Retrieval*: 288–295.
6. Van Nieuwenhuizen, H. A., Venter, W. C., & Grobler, L. M. (2012). The study and implementation of shazam's audio fingerprinting algorithm for advertisement identification. In *Proceedings of SATNAC 2011*.
7. Разработка сверточной нейронной сети для классификации амплитудно-частотных характеристик аудиосигналов / В. Н. Попов, П. С. Ладыгин, В. В. Карев, Я. И. Борцова // Известия Алтайского государственного университета. – 2022. – № 1(123). – С. 116–120. – DOI 10.14258/izvasu(2022)1-19. – EDN IQQVEZ.
8. Ладыгин, П. С. Идентификация записей звуковых сигналов с применением нейронных сетей / П. С. Ладыгин, В. В. Поляков // Вестник Кыргызского государственного университета строительства, транспорта и архитектуры им. Н.Исанова. – 2021. – № 4(74). – С. 570–575. – DOI 10.35803/1694-5298.2021.4.570-575. – EDN JKGGHT.
9. Мансуров А. В., Ладыгин П. С. Способ формирования цифрового отпечатка аудиофайла на основе вектора признаков, получаемого с использованием Constant-Q и Фурье преобразований // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и Технические Науки. - 2020–№08. -С. 79–87 DOI 10.37882/2223-2966.2020.08.21
10. Ладыгин, П. С. Определение подлинности музыкальных аудиозаписей с помощью цифровых отпечатков на основе STFT- и CQT-хроматограмм / П. С. Ладыгин, А. А. Лепендин, А. В. Мансуров // Высокопроизводительные вычислительные системы и технологии. – 2023. – Т. 7, № 1. – С. 46–52. – EDN MEXHSR.