DOI 10.37882/2223–2966.2020.08.21

СПОСОБ ФОРМИРОВАНИЯ ЦИФРОВОГО ОТПЕЧАТКА АУДИОФАЙЛА НА ОСНОВЕ ВЕКТОРА ПРИЗНАКОВ, ПОЛУЧАЕМОГО С ИСПОЛЬЗОВАНИЕМ CONSTANT-Q И ФУРЬЕ ПРЕОБРАЗОВАНИЙ

AN APPROACH TO CALCULATE A FEATURE VECTOR USING FOURIER AND CONSTANT-Q TRANSFORMS FOR AUDIO FINGERPRINTING

A. Mansurov
P. Ladygin

Summary. This paper presents an approach for audio fingerprinting of audio files based on the calculation of a feature vector using Fourier and Constant-Q transforms. A comparative analysis of spectrograms and chromatograms obtained with the Fourier and Constant-Q transforms is performed, and peculiarities of the spectrograms and chromatograms are discussed. It is concluded that the calculation of feature vectors by processing chromatograms to identify roots of chords (or "pure" tones) is more effective than using techniques that require processing spectrograms. The proposed approach for the calculation of feature vectors allows their simple comparison and identification of complex musical compositions with an accuracy of 97%.

Keywords: audio fingerprinting, audio files, spectrogram, identification, chromatogram, wavelet, constant-q transform.

Мансуров Александр Валерьевич
К.т.н., доцент, ФГБОУ ВО «Алтайский государственный университет», г. Барнаул
mansurov.alex@gmail.com
Ладыгин Павел Сергеевич
Старший преподаватель, ФГБОУ ВО «Алтайский государственный университет», г. Барнаул
pavel-ladygin@yandex.ru

Аннотация. В публикации рассматривается процесс формирования цифровых отпечатков аудиофайлов на основе векторов признаков, получаемых путем анализа аудиофайла (фрагмента или целого файла) с использованием преобразования Фурье и Constant-Q преобразования. Выполняется сравнительный анализ спектрограмм и хроматограмм, полученных с помощью указанных преобразований, отмечаются особенности получаемых спектрограмм и хроматограмм, делается заключение об эффективности хроматограмм для построения вектора признаков на основе выделенных сигналов основных («чистых») тонов. Предложенный способ формирования вектора признаков позволяет использовать простую функцию для их последующего сравнения, что дает возможность производить идентификацию сложных композиций с точностью более 97%.

Ключевые слова: цифровой отпечаток, аудиофайлы, спектрограмма, хроматограмма, идентификация, вейвлет, constant-q преобразование.

Введение

Вопрос идентификации и сопоставления (сравнения) аудиофайлов, музыкальных композиций и их фрагментов друг с другом является важной проблемой в случае проведения экспертных исследований, установления факта нелегального использования или нарушения прав на интеллектуальную собственность. Традиционные подходы подразумевают непосредственное исследование экспертом музыкального произведения «на слух» или путем анализа нотных партитур [1], что существенно ограничивает эффективность проводимого исследования и может вносить долю субъективности. Также, существенно ограничивается сфера применения подходов исключительно музыкальными произведениями, исключая большое количество аудиальной информации, не являющейся музыкальной (например, данные акустической эмиссии при исследовании прочностных характеристик материалов).

Современные компьютерные алгоритмы позволяют анализировать аудиальную информацию более эффективно и точно, и их применение не лимитировано исключительно музыкальными композициями. К продвинутым решениям можно отнести, например, технологию Shazam [1,2], а также способ, используемый компанией Google в своих сетевых сервисах (например, Youtube) [3]. Основой таких решений является осуществление спектрального анализа аудиофайла (фрагмента или целиком), получение спектральных характеристик и вычисление (составление) на их основе цифрового отпечатка, который потом используется для сопоставления с другими подобными отпечатками [4]. Наиболее часто используются энергетические спектральные характеристики и анализ спектрограмм Фурье-спектра [4,5] и мел-частотные кепстральные коэффициенты [6]. Однако, не меньшей популярностью пользуются и характеристики, получаемые при обработке вейвлет-спектра анализируемого сигнала [7–10]. Во многом это объясняется адаптивностью к частотному диапазону, селективностью в детализации вычисляемых характеристик и свободой выбора вейвлет-функции в применяемом преобразовании [10].

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

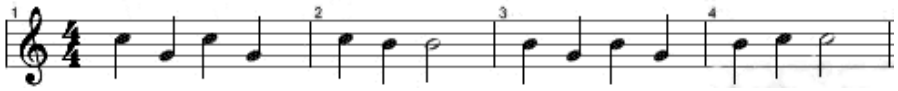


Рис. 1. Нотная запись мелодии «В траве сидел кузнечик»

Рис. 2. Алгоритм формирования вектора признаков для получения цифровых отпечатков аудиофайлов из работы [14].

Эффективным и успешным для получения спектра исследуемого сигнала и последующего вычисления характеристик и получения цифрового отпечатка является Constant-Q преобразование [8–10, 11–13]. Constant-Q преобразование похоже на преобразование Фурье и также осуществляет преобразование серии данных в частотной области, но в своей основе оно тесно связано с непрерывным Morlet вейвлет-преобразованием и имеет ряд достоинств [11,12]. В частности, поддерживая постоянную $Q = f_k / \Delta f_k$ обеспечивается автоматическая адаптация ширины каждого фильтра Δf_k в соответствии с центральной частотой f_k (как в случае вейвлет-преобразований). В этом случае возможно прямое установление соответствий между серией применяемых фильтров и музыкальными нотами при условии идентификации соответствующих центральных частот [13].

В данной работе рассматривается перспективность формирования цифрового отпечатка в виде «вектора признаков», получаемого при анализе фрагмента аудиофайла музыкальной композиции (или, в перспективе, любого произвольного аудиосигнала) с использованием «традиционного» Фурье-спектра и спектрограмм Constant-Q преобразования. В работе изложен непосредственно сам подход к получению «вектора признаков», приведен анализ полученных результатов. Работа является развитием предложенного ранее в [14] метода формирования вектора признаков для получения цифровых отпечатков аудиофайлов.

Метод формирования вектора признаков для получения цифровых отпечатков аудиофайлов

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 представлению и последующую работу при построении «вектора признаков» с частотной сеткой, соответствующей частоте каждой чистой ноты своей октавы.

На рис. 1. представлена нотная запись фрагмента мелодии «В траве сидел кузнечик». Таблица 1 содержит соответствие нот, присутствующих в мелодии, частотам, на которых они звучат.

Предложенный ранее в [14] метод формирования вектора признаков для получения цифровых отпечатков аудиофайлов представляет собой следующую последовательность действий рис. 2:

В работе [14] показана устойчивость получаемого цифрового отпечатка к различным модификациям аудиофайла (замедление, ускорение, смена «питча»), поэтому в данной работе не рассматривается.

В продолжение исследования принято, что в среде графического программирования Labview затруднён

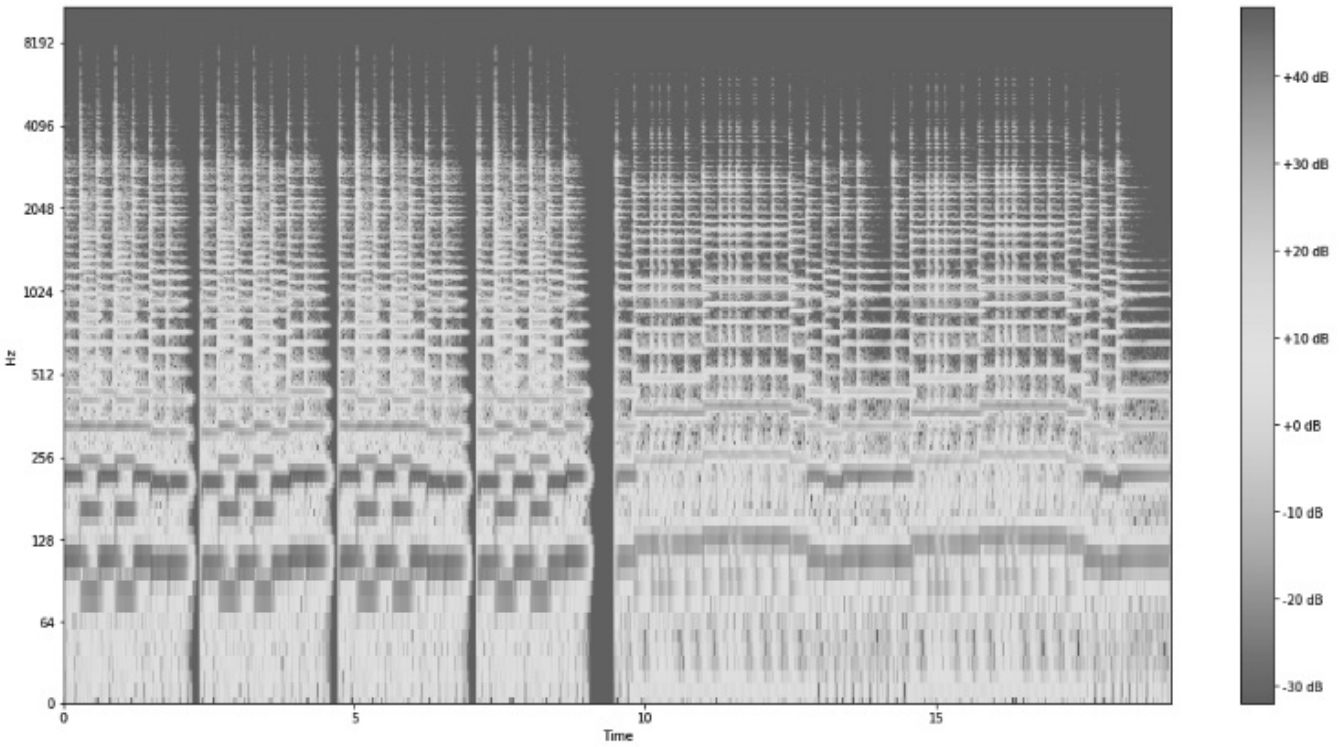


Рис. 3а. STFT-спектрограмма аудиосигнала с мелодией «В траве сидел кузнечик» в 3-й октаве, виртуальный инструмент FL Studio 11– Guitar.

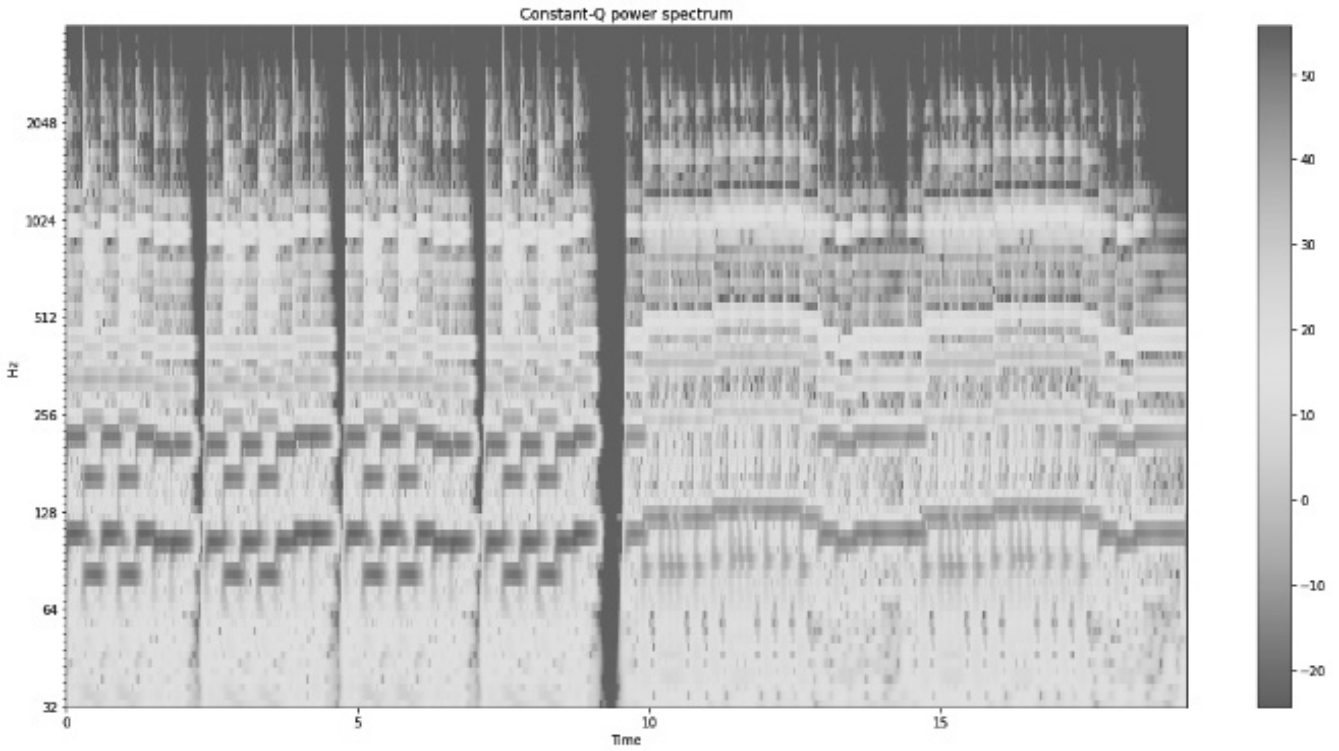


Рис. 3б. CQT-спектрограмма аудиосигнала с мелодией «В траве сидел кузнечик» в 3-й октаве, виртуальный инструмент FL Studio 11– Guitar.

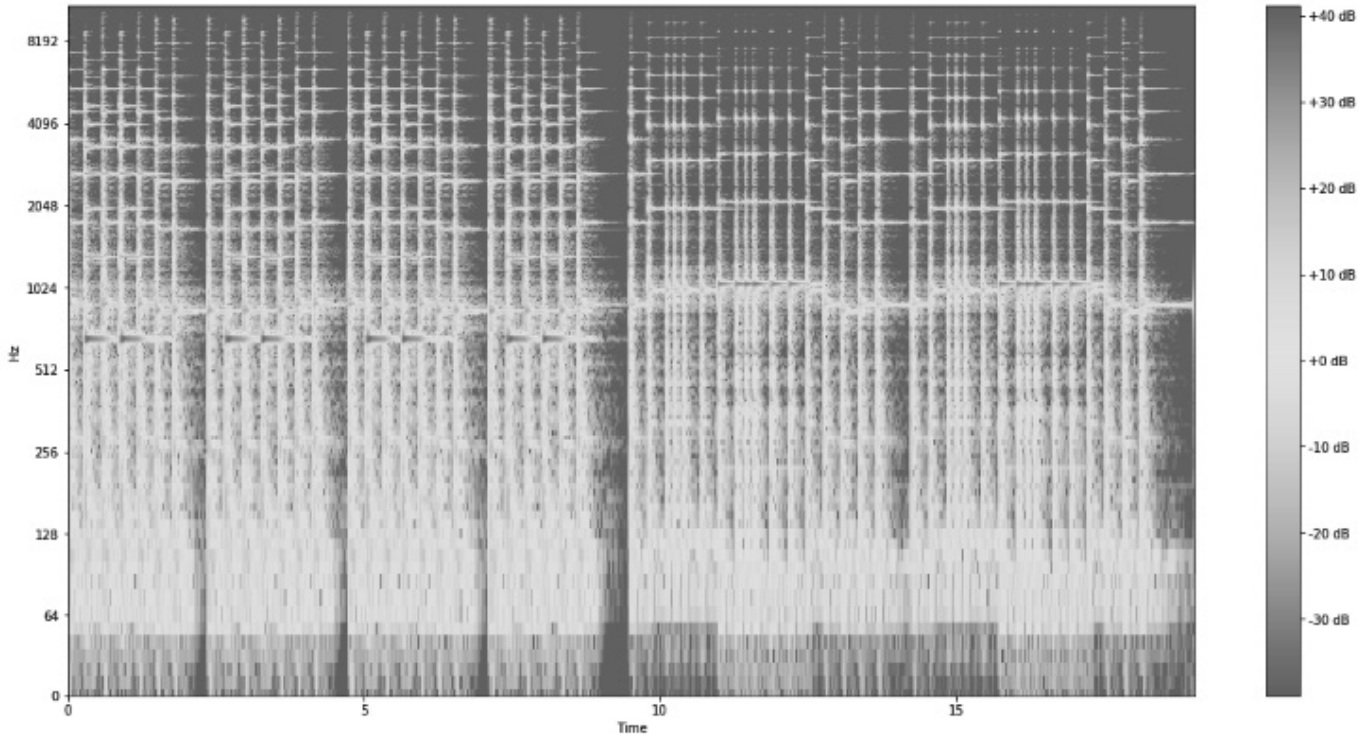


Рис. 4а. STFT-спектрограмма аудиосигнала с мелодией «В траве сидел кузнечик» в 6-й октаве, виртуальный инструмент FL Studio 11– Piano.

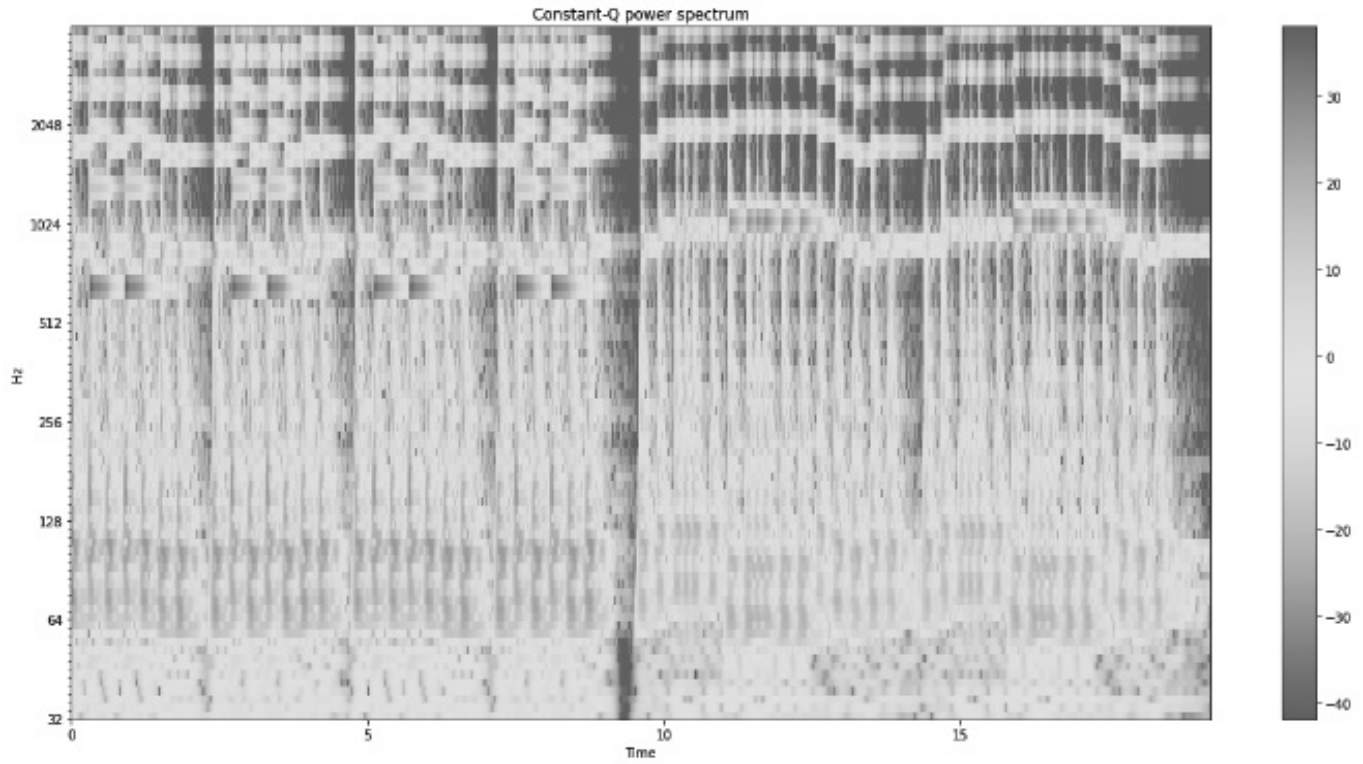


Рис. 4б. STFT-спектрограмма аудиосигнала с мелодией «В траве сидел кузнечик» в 6-й октаве, виртуальный инструмент FL Studio 11– Piano.

процесс исследования аудиосигналов на предмет дополнительных признаков, важных для создания цифровых отпечатков аудиофайлов. Также было принято решение сделать алгоритм независимым от ширины окна Фурье-преобразования, которая зависела от темпа исполнения мелодии и не всегда попадала удачно на начало и конец ноты. Продолжение исследования предложенного метода происходило на языке Python, как на наиболее удобном и полном в отношение инструментария для исследования аудиосигналов. Для поиска частоты основного тона принято решение исследовать спектрограммы, полученные с помощью Short-time Fourier (STFT) и Constant-Q (CQT) — преобразований, как наиболее полно отражающие содержимое аудиосигналов.

Short-time Fourier преобразование

Оконное преобразование Фурье — это разновидность преобразования Фурье, определяемая следующим образом:

$$F(t,w) = \int_{-\infty}^{+\infty} f(\tau)W(\tau - t)e^{-iw\tau}d\tau \quad (1)$$

где $W(\tau–t)$ — некоторая оконная функция [15].

Для вычисления STFT в Python может быть применена функция librosa.stft(X), где X — аудиосигнал [16].

Constant-Q преобразование

В соответствии с теоретической базой и рекомендациями [9,12], Constant-Q преобразование тесно связано с Фурье-преобразованием и для его вычисления используется аналогичный подход. Для Constant-Q преобразования k-я спектральная компонента $X[k]$ вычисляется следующим образом:

$$X[k] = \frac{1}{N[k]} \sum_{n=0}^{N[k]-1} W[k,n]x[n]e^{\frac{-j2\pi Qn}{N[k]}} \quad (2)$$

где N[k] — длина окна для каждого «бина», W[k, n] — функция окна, Q — постоянная преобразования, x[n] — исходный сигнал.

Для вычисления CQT в Python может быть применена функция librosa.cqt(x, sr), где x — аудиосигнал, sr — частота дискретизации аудиосигнала.

Сравнительный анализ спектрограмм, полученных с помощью STFT и CQT

Для анализа работы двух преобразований CQT и STFT были сформированы 23 звуковые дорожки, содержащие как реальные записи музыкальных инструментов, так и сгенерированные в программе FL Studio [17]. В данной работе представлены наиболее интересные результаты.

Визуальный анализ полученных спектрограмм показывает следующее:

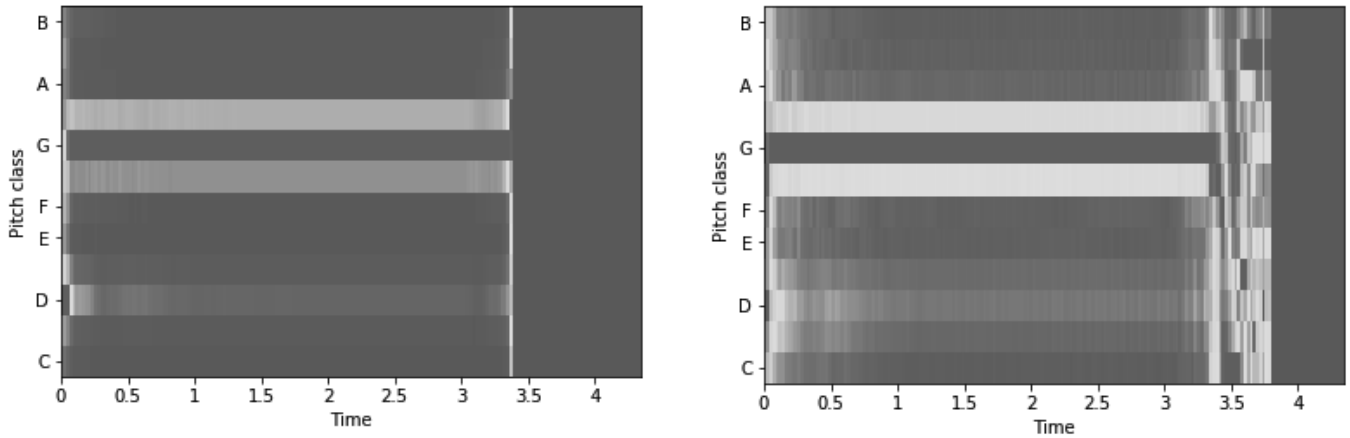
1. ширина и яркость основного тона на спектрограммах STFT меняется в зависимости от его значения: в нижней полосе частот основной тон более широк и размазан, в то время как в верхней полосе частот узок и мало заметен;
2. ширина основного тона на спектрограммах CQT практически не меняется в зависимости от полосы частот;
3. спектрограммы CQT имеют чуть более читаемый вид в нижней полосе частот;
4. спектрограммы STFT содержат меньше «шумов», характеризующих тембральные характеристики музыкального инструмента. Такие шумы больше свойственны спектрограммам CQT.

Исходя из вышеизложенного можно заключить, что спектрограммы CQT и STFT содержат информацию о длительности аудиосигнала, его основных наиболее интенсивных частотах (в случае монофонической записи — частотах основного тона), однако излишнее зашумление обертонами затрудняет процесс извлечения информации о наиболее значащих частотах. В задачах идентификации основной мелодии не важно, на каком музыкальном инструменте мелодия сыграна, значит не важны «лишние» гармоники. Также частота основного тона на данных изображениях размазана по полосе частот, что затрудняет определение её точного значения для формирования вектора признаков аудиосигнала.

Сравнительный анализ хроматограмм, полученных с помощью STFT и CQT

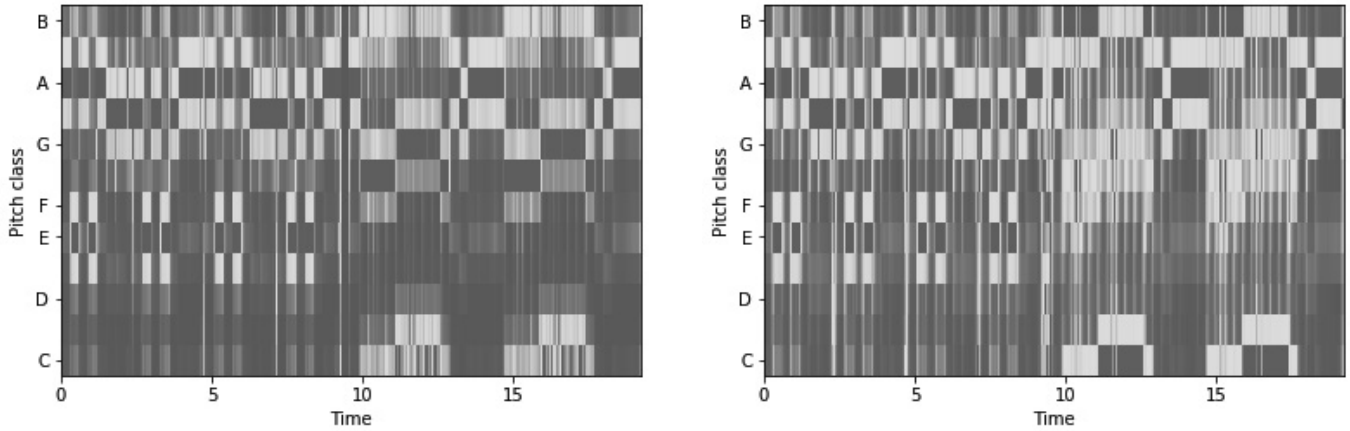
Наиболее устойчивыми к изменениям тембра и инструментовки характеристиками аудиосигналов, содержащих мелодические конструкции, являются хроматограммы [18]. Построение хроматограмм основывается на условном разделении частотного диапазона на 12 полутонов (согласно западной музыкальной нотации), соответствующих стандартной октаве (C, C♯, D, D♯, E, F, F♯, G, G♯, A, A♯, B), что позволяет отражать гармонические и мелодические характеристики аудиосигнала.

Хроматограмма «собирает» все гармоники, объединяясь с частотой основного тона, что нормирует участки сигнала к оси частот, удобной для идентификации аудиосигнала. По оси Y отображается нота, по оси X — время.



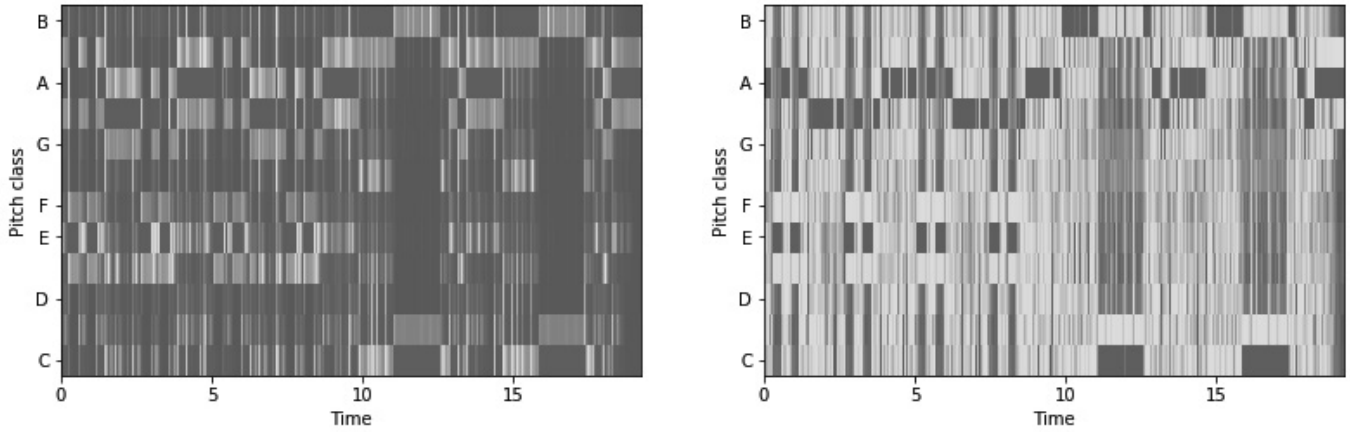
а) б)

Рис. 5. Нота G4 — «соль» 4-й октавы: а) STFT-хроматограмма, б) CQT-хроматограмма



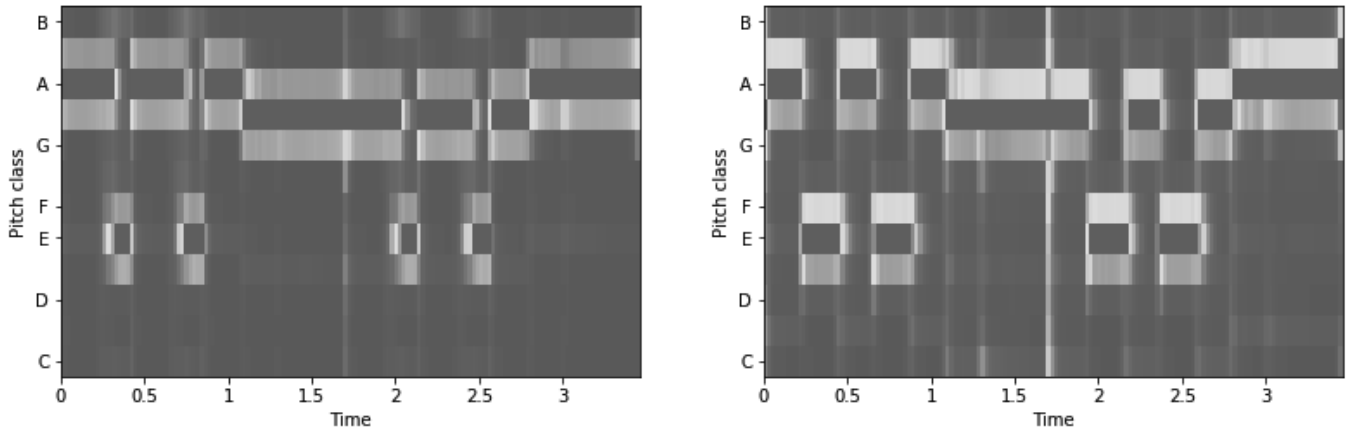
а) б)

Рис. 6. Мелодия «В траве сидел кузнечик» в 3 октаве, инструмент — Guitar:
а) STFT-хроматограмма, б) CQT-хроматограмма



а) б)

Рис. 7. Мелодия «В траве сидел кузнечик» в 6 октаве, инструмент — Piano: а) STFT-хроматограмма, б) CQT-хроматограмма



а) б)

Рис. 8. STFT-хроматограма участка мелодии «В траве сидел кузнечик», инструмент — Piano: а) в 4-й октаве, б) в 5-й октаве

Для построения хроматограмм может быть использована функция библиотеки Librosa языка Python librosa. feature.chroma_cqt(y=x, sr=sr).

Для ранее рассмотренных аудиосигналов были построены хроматограммы, изображенные на рис. 5а, рис. 5б, рис. 6а, рис. 6б, рис. 7а, рис. 7б.

Визуальный анализ полученных хроматограмм показывает, что:

1. ширина и яркость основного тона на хроматограммах STFT и CQT не меняется в зависимости от его положения в воспроизводимом диапазоне частот;
2. хроматограммы CQT имеют намного более читаемый вид в нижней полосе частот, чем хроматограммы от STFT;
3. хроматограммы STFT содержат меньше «шумов» вокруг основной ноты. Такие шумы больше свойственны спектрограммам CQT.
4. хроматограммы CQT содержат «зашумлённые хвосты» и не равны нулю в тех областях аудиосигнала, в которых ноты нет. Здесь мы видим некоторый шум, который сложно как-либо идентифицировать.

По сравнению со спектрограммами, хроматограммы содержат больше информации о длительностях тех или иных нот, присутствующих в аудиосигнале, что позволяет строить вектор признаков, позволяющий сравнивать аудиосигналы с мелодическими конструкциями между собой. Однако, видно, что весь частотный диапазон спектрограмм приводится к 12-ступенчатой оси ординат. Когда мелодия могла пойти вверх, в другую октаву, хроматограмма отобразит эту ноту всё равно внутри своего диапазона. Данную особенность необходимо учитывать при построении векторов признаков аудиосигналов.

Построение вектора признаков аудиосигнала

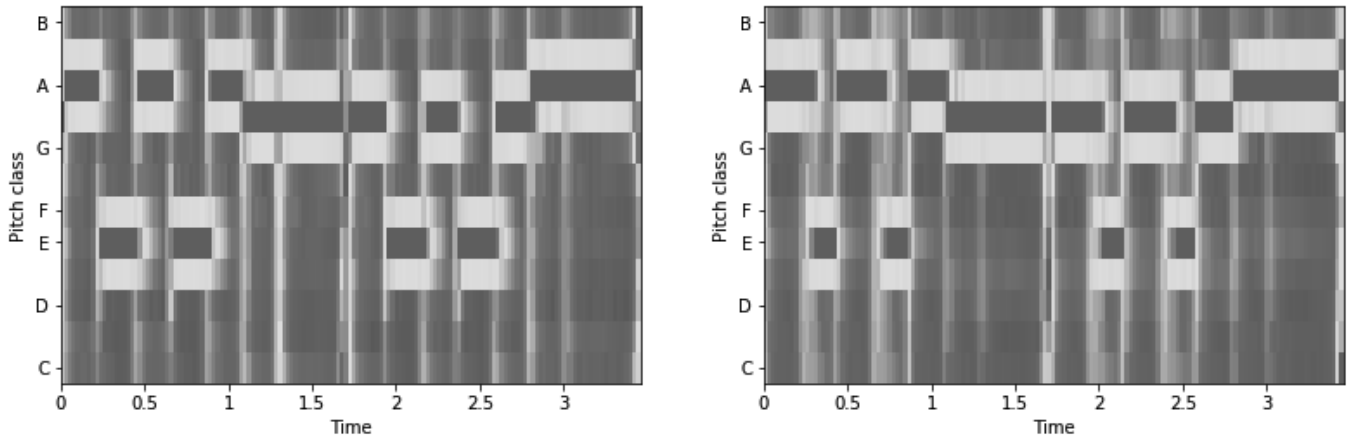
При экспертной оценке двух мелодических конструкций между собой на предмет похожести учитывается определённое количество одинаково идущих нот подряд. По различным судебным решениям можно обнаружить, что «плагиатом» считалось осознанное или неумышленное заимствование от 7 до 11 нот подряд [19]. При этом эксперт визуально сравнивает партитуры музыкальных партий, а также «на слух» определяет их степень похожести.

Для более быстрой оценки мелодий на предмет заимствования предлагается формирование вектора признаков спорных аудиосигналов и сравнение их между собой. Используя вышеприведённые характеристики аудиосигналов, предлагается формирования вектора признаков следующего содержания:

$$[(k0, X), (k1, X), \ldots (kn, X)], \quad (3)$$

в котором kn — номер отсчёта в сигнале, а X — значение из набора: $0, 1,.. z$, где $z = 12$, что соответствует номеру ноты в 12-ступенчатом звуковом ряде C, C♯, D, D♯, E, F, F♯, G, G♯, A, A♯, B.

Построение вектора признаков происходит с помощью анализа хроматограмм, как более удобных изображений мелодических конструкций, и преобразования изображения в массив. Для построения такого вектора написана функция fingerprint:



а) б)

Рис. 9. CQT-Хроматограмма участка мелодии «В траве сидел кузнечик», инструмент — Piano: а) в 4-й октаве, б) в 5-й октаве

```
def fingerprint(arr2D):
    local_maxima = get_2D_peaks(arr2D, plot=True)
    local_maxima_sorted = sorted(local_maxima, key=lambda tup: tup[0])
    return local_maxima_sorted
```

В данной функции метод get_2D_peaks(arr2D, plot=True) занимается поиском локальных пиков на изображении хроматограммы. Все гармоники с интенсивностью меньше 95% от максимальной фильтруются.

Для тестового формирования векторов признаков аудиофайлов и сравнения их между собой были сформированы две мелодические конструкции, исполненные в 4й и 5й октавах с одинаковой скоростью, их STFT-хроматограммы представлены на рис. 8а и рис. 8б.

Поэлементное сравнение, полученных векторов признаков из хроматограмм STFT-преобразования по типу (2), показывает совпадение мелодий в 85%.

Хроматограммы CQT-преобразования представлены на рис. 9а и рис. 9б.

Поэлементное сравнение, полученных векторов признаков из хроматограмм CQT-преобразования по типу (2), показывает совпадение мелодий в 97%.

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.

Показано, что спектрограммы аудиофайлов не позволяют достоверно определять частоту основного тона и менее эффективны для построения векторов признаков по предложенному алгоритму, чем хроматограммы. Анализ хроматограмм позволяет находить частоту с максимальной амплитудой на каждом фрагменте мелодии, которая в большинстве случаев является «чистым тоном» и является табличным значением, имеющим соответствующее нотное обозначение.

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 97%. Применение более эффективного метода для выполнения сравнения в данной публикации не рассматривается и является вопросом для последующих исследований.

Данный подход наиболее полезен для автоматического анализа аудиоданных, который может облегчить работу эксперта, удешевить и ускорить процесс искусствоведческой экспертизы, а также расширить существующую доказательную базу в задачах проверки мелодий на плагиат. 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. Экспертное заключение по информационным материалам запроса от 30.03.2017 / Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет» [Электронный ресурс]. URL: https://spbu.ru/sites/default/files/20171206_zakl.pdf (дата обращения 02.05.2020).
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 25.06.2020).
3. Эволюция Content ID: как Youtube совершенствует свою самую спорную функцию [Электронный ресурс] / Air. URL: http://www.air.io/content-id-evolution/ (дата обращения 22.06.2020).
4. 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.
5. Haitsma J, Kalker T. A Highly Robust Audio Fingerprinting System / Journal of New Music Research, Vol. 32(2003), No. 2, p. 211–222.
6. 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.
7. 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
8. Schörkhuber C., Klapuri A. Constant-Q transform toolbox for music processing. // 7th Sound and Music Computing Conference 2010, Barcelona, Spain, 2010. URL: https://iem.kug.ac.at/fileadmin/media/iem/projects/2010/smc10_schoerkhuber.pdf (дата обращения 22.06.2020).
9. Cancela P., Rocamora M., Lopez E. An Efficient Multi-Resolution Spectral Transform for Music Analysis. // 10th International Society for Music Information Retrieval Conference, 2009, pp. 309–314. URL: http://ismir2009.ismir.net/proceedings/PS2–20.pdf (дата обращения 22.06.2020).
10. Huzaifah M. Comparison of time-frequency representations for environmental sound classification using convolutional neural networks. 2017. URL: https://arxiv.org/abs/1706.07156 (дата обращения 22.06.2020).
11. Chettri S.R., Ishiwaka Y., Kimura H., Nagano I. Harmonic wavelets, constant Q transforms, and the cone kernel TFD. // Proc. SPIE2762, Wavelet Applications III, (22 March 1996), 1996, pp. 446–451. DOI: https://doi.org/10.1117/12.236016
12. Brown J. C. Calculation of a constant q spectral transform. // The Journal of the Acoustical Society of America, V. 89, 1991, pp. 425–434. URL: http://academics.wellesley.edu/Physics/brown/pubs/cq1stPaper.pdf (дата обращения 22.06.2020).
13. Purwins H., Blankertz B., Obermayer K. A new method for tracking modulations in tonal music in audio data format. // Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN2000. Neural Computing: New Challenges and Perspectives for the New Millennium, Como, Italy, 2000, V.6, pp. 270–27. DOI: 10.1109/IJCNN.2000.859408.
14. Мансуров А.В., Ладыгин П. С. Подход к формированию вектора признаков для алгоритма формирования цифровых отпечатков аудиофайлов // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и Технические Науки.—2017. -№ 09. -С. 27–34.
15. Ervin Sejdić, Igor Djurović, Jin Jiang. Time–frequency feature representation using energy concentration: An overview of recent advances // Digital Signal Processing. Volume 19, Issue 1, January 2009, Pages 153–183.
16. Librosa [Электронный ресурс] / librosa development team. URL: https://librosa.org/ (дата обращения 22.06.2020).
17. FL Studio [Электронный ресурс] / Официальный сайт. URL: https:// www.image-line.com/ (дата обращения 22.06.2020).
18. Shepard, Roger N. Circularity in judgments of relative pitch // Journal of the Acoustical Society of America. 36 (212): 2346–2353.
19. Количество нот плагиата в музыке / Copyright © КОПИРАЙТ [Электронный ресурс]. URL: https://www.copyright.ru/news/main/2015/2/2/plagiat/ (дата обращения 27.07.2020).
20. 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 (дата обращения 25.06.2020).