СРАВНЕНИЕ ВЕКТОРОВ ПРИЗНАКОВ АУДИОФАЙЛОВ, ПОЛУЧЕННЫХ С ПОМОЩЬЮ ХРОМАТОГРАММ И ПИТЧ-ТРЕКЕРА CREPE

Ладыгин П.С., Мансуров А.В., Рудер Д.Д.
Алтайский государственный университет, г. Барнаул
email: pavel-ladygin@yandex.ru

Аннотация. В данной работе сравниваются векторы признаков, получаемые с помощью анализа хроматограмм аудиофайла и с помощью питч-трекера CREPE, основанного на свёрточной нейронной сети. Поиск оптимального решения является дополнительной проработкой алгоритма формирования компьютерного инструментального подхода к осуществлению экспертных оценок музыкальных произведений на предмет нарушения прав на интеллектуальную собственность. Приведено описание алгоритма получения специального цифрового отпечатка аудиозаписи (аудиофайла). Выполняется сравнительный анализ вычисленных битовых последовательностей, полученных с помощью хроматограмм и CREPE. Анализируется эффективность подходов к получению более информативных векторов признаков для сравнительного анализа оригинала музыкального произведения с другими.

Ключевые слова: цифровой отпечаток, вектор признаков, хроматограмма, свёрточная нейронная сеть.

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. При этом современные компьютерные алгоритмы позволяют анализировать аудиальную информацию более эффективно и точно.

Одной из наиболее надёжных характеристик аудиофайлов являются цифровые отпечатки, используемые некоторыми популярными системами поиска и хранения информации об аудиоданных (Shazam [2], ContentID [3]). Построение качественного вектора признаков аудиофайла, позволяющего сравнивать цифровые отпечатки аудио и оценивать степень схожести их между собой, является важной задачей для создания современных экспертных систем по проверке музыки на плагиат.

Мелодия представляется важнейшим, первичным компонентом музыкальной композиции, на который распространяется защита интеллектуальной собственности. Под мелодией понимается [4] «осмысленно-выразительное и законченное по построению одноголосное последование звуков, которые объединены конкретными отношениями высоты, длительности и силы, а также является основой музыкального произведения» [5]. При этом высота отдельно звучащей ноты мелодии измеряется в Герцах, в большинстве случаев является табличным значением и подлежит измерению автоматическими средствами.

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.

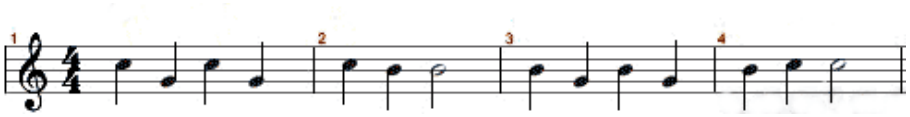

Рисунок 1. Нотная запись мелодии «В траве сидел кузнечик» (композитор – В.Я. Шаинский).

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

		До (C)	Ре (D)	Ми (E)	Фа (F)	Соль (G)	Ля (A)	Си (B)
f, 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

В общем случае предлагаемый ранее [6] метод формирования вектора признаков для получения цифровых отпечатков аудиофайлов представляет собой следующую последовательность действий (рис.2):

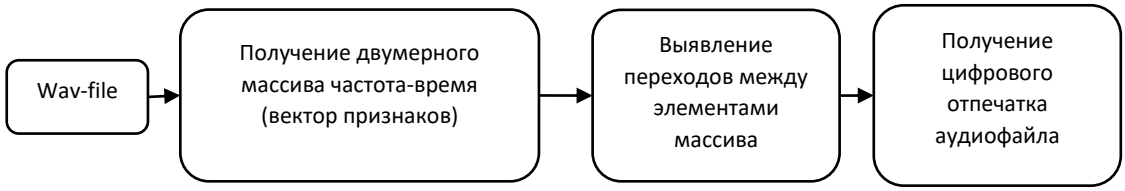

Рисунок 2. Алгоритм формирования вектора признаков для получения цифровых отпечатков аудиофайлов [6].

Предлагаемый ранее подход предполагает кодирование переходов вверх по ноте (рис.1) как 01, вниз – 10, а отсутствие перехода – 00, то «В траве сидел кузнечик» при делении партитуры на четвертные ноты будет представлен в виде битовой последовательности: 10011001100000 0010011001010000. Ранее [6] показана устойчивость получаемого цифрового отпечатка к различным модификациям аудиофайла (замедление, ускорение, смена «питча»), поэтому в данной работе не рассматривается. Такой отпечаток может являться вспомогательной характеристикой для быстрого поиска схожих по переходам мелодий в базах данных крупных систем, а также частью автоматической системы, востребованной для проведения искусствоведческих экспертиз.

Для получения такого признака мелодии в автоматическом режиме возможно использование хроматограмм, показавших свою эффективность в данном подходе по сравнению с извлечением признаков из спектрограмм [7]. В данной работе происходит сравнение векторов признаков, полученных с помощью хроматограмм и с помощью системы CREPE.

Хроматограммы.

Наиболее устойчивыми к изменениям тембра и инструментовки характеристиками аудиосигналов, содержащих мелодические конструкции, являются хроматограммы [8]. 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. Для построения хроматограмм может быть использована функция библиотеки Librosa языка Python librosa.feature.chroma_stft(y=x, sr=sr).

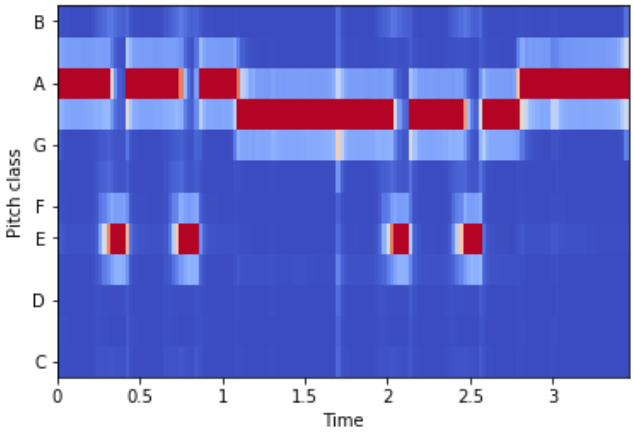

Рисунок 3. STFT-Хроматограмма мелодии «В траве сидел кузнечик» в 5 октаве, инструмент – Piano.

На языке Python написан алгоритм, извлекающий полезную информацию из хроматограммы, как из изображения. Получаемый двумерный массив частота (нота) – время необходимый для создания цифрового отпечатка принимает следующий вид:

$[(k_0, X), (k_1, X), \dots (k_n, X)]$, (1)

в котором kn – номер отсчёта в сигнале, а X – значение из набора: 0, 1,.. z, где z=11, что соответствует номеру ноты в 12-ступенчатом звуковом ряде C, C ♯, D, D♯ , E, F, F♯, G, G♯, A, A♯, B.

На рис. 4 представлен участок массива, полученного для аудиофайла продолжительностью 3 секунды, содержащего фрагмент мелодии «В траве сидел кузнечик»:

[(0, 9),
(1, 9),
(2, 9),
(3, 9),
…,
(147, 9),
(148, 9)]

Рисунок 4. Вектор признаков фрагмента мелодии «В траве сидел кузнечик», полученного из хроматограммы.

Если объединить отсчёты, в которых нота не меняется, получаем вектор: [(0-13, 9), (14-17, 4), (18-31, 9), (32-36, 4), (37-46, 9), (47-87, 8), (88-91, 4), (92-105, 8), (106-110, 4), (111-120, 8), (121-148, 9). Что согласно предлагаемому алгоритму кодирования представляет собой битовую последовательность, представленную в Таблице 2.

Таблица 2. Сравнение цифровых отпечатков (теория и хроматограмма).

Теоретическая последовательность	Полученная последовательность
10011001100000 0010011001010000	1001100110 1001100101

Как видно из таблицы, полученная последовательность короче теоретической на 10 бит в связи с тем, что в хроматограмме сливается в одну полосу конец и начало одинаковых нот. Такие участки представляются единой нотой, что не соответствует действительности. Помимо этого, в записи использованы ноты одинаковой длительности, однако в хроматограмме этого не наблюдается. Приведение к 12-ти ступенчатому нотному ряду также не является корректным в задачах анализа мелодических конструкций, так как зачастую композиторы руководствуются использованием более чем одной октавы.

Convolutional Representation for Pitch Estimation (CREPE)

CREPE состоит из глубокой сверточной нейронной сети, которая работает непосредственно со звуковым сигналом во временной области для получения оценки основного тона. Блок-схема предлагаемой архитектуры системы представлена на рис.3. Входные данные представляют собой выдержку из 1024 отсчетов из временной области звукового сигнала с частотой дискретизации 16 кГц. Здесь шесть сверточных слоев, которые приводят к 2048-мерному скрытому представлению, который затем плотно соединен с выходным слоем с сигмоидальными активациями, соответствующими 360-мерному выходному вектору. При этом результирующая оценка высоты тона рассчитывается детерминировано [9].

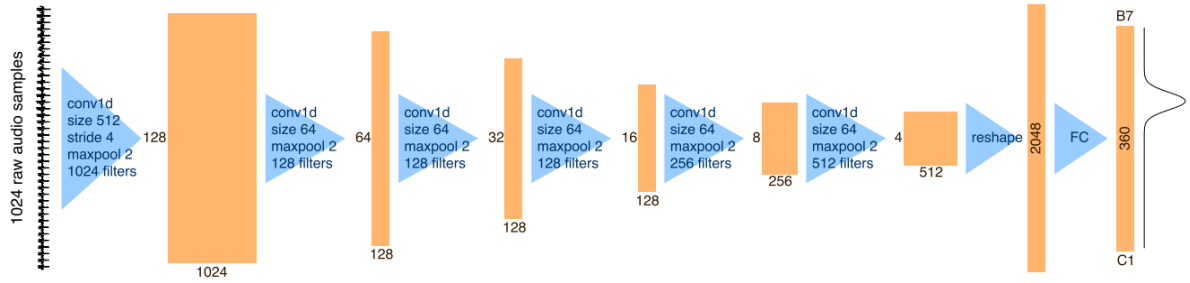

Рисунок 5. Архитектура питч-трекера CREPE.

На рис.6 представлен участок массива, полученного для аудиофайла продолжительностью 3 секунды, содержащего фрагмент мелодии «В траве сидел кузнечик» с помощью функции crepe.predict(audio, sr, viterbi=False, model_capacity='full',step_size=100):

[(0.0, 440.69, 0.89),
(0.1, 440.68, 0.97),
(0.2, 440.03, 0.96),
(0.3, 330.93, 0.94),
…,
(3.3, 440.23, 0.97),
(3.4, 441.35, 0.92)]

Рисунок 6. Вектор признаков фрагмента мелодии «В траве сидел кузнечик», полученного с помощью CREPE.

Здесь вектор признаков по умолчанию представлен тремя столбцами: первый содержит момент времени (здесь эмпирически использован шаг в 100 мс), второй содержит частоту основного тона в Гц, определённую в этот момент времени, третий столбец – вероятность точности определённого значения частоты. Такой вектор признаков несёт более полную информацию об аудиофайле.

К полученному вектору признаков применены следующие преобразования средствами языка python:

1. Удаление элементов с низкой вероятностью точности. Эмпирически установлен порог в 0.90.

2. Приведение найденной частоты основного тона к табличному значению (Таблица 1).

Полученный в результате преобразований вектор признаков позволил получить битовую последовательность, представленную в Таблице 3:

Таблица 3. Сравнение цифровых отпечатков (теория и CREPE).

Теоретическая последовательность	Полученная последовательность (CREPE)
10011001100000 0010011001010000	1001100110 1001100101

Последовательности, полученные с помощью хроматограмм и питч-трекера Crepe полностью совпадают для данной мелодии, однако:

1. Значения частот, получаемых с помощью CREPE изначально не приведены к 12-ступенчатому звуковому ряду, что позволяет анализировать мелодии, исполненные в пределах двух и более октав и получать более корректный цифровой отпечаток.

2. Гибкость настройки питч-трекера позволяет искать значения частот без анализа изображения, а только исходя из готового массива числовых данных, что уточняет значение определённой частоты и делает её независимой от музыкальной октавы. Как было изучено ранее [7], хроматограммы плохо характеризуют высокочастотную область мелодических конструкций.

3. Значения времени, хранимые в векторе признаков CREPE, соотносятся с реальными данными в отличие от отсчётов, представленных в векторе признаков от хроматограмм.

4. Вероятность точности определённых значений частот позволяют понять, можно ли доверять конкретной ячейке вектора признаков. Зачастую в окно преобразований попадает момент перехода от ноты к ноте в мелодии. В таком случае система усредняет значения частот и выводит результат, который в векторе признаков от хроматограмм невозможно оценить автоматически.

Построение экспертной системы сравнения аудиоданных между собой с использованием нейронных сетей, цифровых отпечатков и сравнения векторов признаков представляется перспективным направлением деятельности для задач защиты прав на интеллектуальную собственность. Дальнейшая работа предполагает анализ работы питч-трекера CREPE на большем объёме музыкальных данных. Предполагается автоматизация поиска оптимальной ширины окна (step_size) для функции crepe.predict, использованной в данной работе, на основе вычисления bpm аудиофайла, что позволит более точно «попадать» в начало и конец звучания каждой отдельно исполненной ноты.

Библиографический список

1. Экспертное заключение по информационным материалам запроса от 30.03.2017 / Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет»

[Электронный ресурс] // режим доступа: https://spbu.ru/sites/default/files/20171206_zakl.pdf (дата обращения 02.05.2020).

2. Shum S. The Basics of Audio Fingerprinting / MIT Computer Science and Artificial Intelligence Laboratory [Электронный ресурс] // режим доступа: http://people.csail.mit.edu/sshum/talks/audio_fingerprinting_sls_24Oct2011.pdf (дата обращения 25.06.2020).

3. Эволюция Content ID: как Youtube совершенствует свою самую спорную функцию [Электронный ресурс] / режим доступа: http://www.air.io/content-id-evolution/ (дата обращения 22.06.2020).

4. Горшунов, А. А. Музыкальные произведения как объекты авторского права / А. А. Горшунов. — Текст : непосредственный // Новый юридический вестник. — 2020. — № 4 (18). — С. 24-29.

5. Должанский А. Н. Краткий музыкальный словарь // Л.: 1964. - С. 40.

6. 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.

7. 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

8. Shepard, Roger N. Circularity in judgments of relative pitch // Journal of the Acoustical Society of America. №36 (212). – pp.2346–2353.

9. Jong Wook Kim, Justin Salamon, Peter Li, Juan Pablo Bello. CREPE: A Convolutional Representation for Pitch Estimation // Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) – 2018.

COMPARISON OF THE FEATURE VECTOR'S OF AUDIO FILES OBTAINED WITH THE USE OF CHROMATOGRAMS AND THE PITCH-TRACKER CREPE

Ladygin P.S., Mansurov A.V., Ruder D.D.
Altai State University, Barnaul
email: pavel-ladygin@yandex.ru

Abstract. This paper compares feature vectors obtained by analyzing the chromatograms of an audio file and using the CREPE pitch tracker based on a convolutional neural network. The search for an optimal solution is an additional study of the algorithm for the formation of a computer instrumental approach to the implementation of expert assessments of musical works for infringement of intellectual property rights. The description of the algorithm for obtaining a special digital fingerprint of an audio recording (audio file) is given. Comparative analysis of the calculated bit sequences obtained using chromatograms and CREPE is performed. The effectiveness of approaches to obtaining more informative vectors of features for a comparative analysis of the original musical work with others is analyzed.

Keywords: fingerprint, feature vector, chromatogram, convolutional neural network.