УДК 004.622 DOI:10.35803/1694-5298.2021.4.570-575

ЛАДЫГИН П.С., ПОЛЯКОВ В.В.
[1]Алтайский государственный университет, г. Барнаул, Российская Федерация

LADYGIN P.S. , POLYAKOV V.V.
[1]Altay State University, Barnaul, Russian Federation
pavel-ladygin@yandex.ru, pvv@asu.ru

ИДЕНТИФИКАЦИЯ ЗАПИСЕЙ ЗВУКОВЫХ СИГНАЛОВ С ПРИМЕНЕНИЕМ НЕЙРОННЫХ СЕТЕЙ

IDENTIFICATION OF RECORDINGS OF SOUND SIGNALS USING NEURAL NETWORKS

Бул эмгекте аудио маалыматтарды талдоо жана классификациялоо үчүн жазылган аудио сигналдардын окшоштук пайызын эсептөөнүн актуалдуу көйгөйү каралган. Аудио жазуунун (аудио файлдын) атайын санариптик манжа изин алуу алгоритминин сүрөттөлүшү келтирилген. Функциялардын векторун түзүү ыкмасы жана салыштыруу процедурасы сыналды, бул окшош булактардан чыккан үн сигналдарынын жазууларын салыштырып, алардын идентификациясын 61% -65% тактык менен орнотууга, ошондой эле ар башка үн сигналдарын туура аныктоого мүмкүндүк берет. . Окутулган нейрон тармагы болгон CREPE чайыр трекеринин жардамы менен алынган эсептелген бит ыраатттуулугуна салыштырмалуу талдоо жүргүзүлдү. Аудио сигналдарда, жазылган сигналдардан өзгөчөлүктөрдү бөлүп алуу үчүн, сигналдар башкаруу объектинин өзгөчөлүктөрүн гана эмес, бурмалоолорду, ар кандай мүнөздөгү туш келди процесстерди жана башка тоскоолдук кылган факторлорду чагылдыраары эске алынат.

Өзөк сөздөр: *идентификация, санариптик манжа изи, нейрон тармагы, хромаграмма, өзгөчөлүк вектору.*

В данной работе рассматривается актуальная проблема расчета в процентном соотношении сходства записанных звуковых сигналов для задач анализа и классификации аудиоданных. Приведено описание алгоритма получения специального цифрового отпечатка аудиозаписи (аудиофайла). Апробирована методика формирования вектора признаков и процедура сравнения, которая позволяет производить сопоставление записей звуковых сигналов схожих источников и устанавливать их идентичность с точностью 61%–65%, а также корректно идентифицировать заведомо разные звуковые сигналы. Выполнен сравнительный анализ вычисленных битовых последовательностей, полученных с помощью питч-трекера CREPE, который представляет собой обученную нейронную сеть. В случае аудиосигналов для извлечения из регистрируемых сигналов признаков учитывается, что сигналы отражают не только особенности объекта контроля, но и искажения, случайные процессы различной природы и другие мешающие факторы.

Ключевые слова: *идентификация, цифровой отпечаток, нейросеть, хромаграмма, вектор признаков.*

This paper considers the actual problem of calculating the percentage of similarity of recorded audio signals for the analysis and classification of audio data. The description of the algorithm for obtaining a special digital fingerprint of an audio recording (audio file) is given. The technique of forming a vector of features and a comparison procedure have been tested, which allows comparing the recordings of sound signals from similar sources and establishing their identity with an accuracy of 61% –65%, as well as correctly identifying obviously different sound

ABBYY PDF Transformer 2.0 Click here to buy www.ABBYY.com

signals. A comparative analysis of the calculated bit sequences obtained using the CREPE pitch tracker, which is a trained neural network, has been performed. In the case of audio signals, for the extraction of features from the recorded signals, it is taken into account that the signals reflect not only the features of the control object, but also distortions, random processes of various nature and other interfering factors.

Key words: *identification, digital fingerprint, neural network, chromagram, feature vector.*

Введение. В современном мире одной из самых развивающихся технологий является аудиоанализ на основе алгоритмов машинного обучения. Он включает в себя: автоматическое распознавание речи, цифровую обработку сигналов, классификацию, поиск, тегирование, обнаружение событий и генерацию акустической информации. Логическая, эмоциональная, описательная или иная релевантная интерпретация звука всё меньше доверяется человеческому уху, и всё чаще становится задачей для различного рода приложений на повсеместных устройствах.

Возросшие вычислительные мощности, доступность больших объёмов аудиоматериалов и тенденции в решении различных задач аудиоанализа предполагают рассмотрение звукового сигнала в качестве сложного комплекса синусоид с различной частотой, амплитудой и фазой, подобно мозгу человека, имеющих определённый смысл и значение. Современные популярные и распространенные системы (Алиса, Siri, Маруся и т.д) — это продукты, созданные на основе моделей, извлекающих информацию из аудиосигналов.

Тем не менее на сегодняшний день вопрос идентификации и сопоставления (сравнения) аудиосигналов, и их фрагментов друг с другом не является повсеместным автоматизированным процессом. Например, традиционные подходы, применяемые привлекаемыми экспертами в ходе судебных разбирательств, включают в себя непосредственное исследование экспертом аудиоданных «на слух» [1]. За редким исключением применение экспертами технических средств ограничивается наложением одной записи на другую или выяснением процесса формирования аудиозаписи и аудиофайла как конечного продукта.

Подобные действия несут в себе существенную долю субъективизма, зависят от квалификации эксперта и далеко не всегда позволяют справиться с техническими приемами, применяемыми на этапе записи аудиоинформации (например, изменение питча, ритма, внесение эффектов и искажений), а также с естественным преобразованием звука под влиянием распространения акустических волн в разных средах.

Существуют технологии создания звуковых отпечатков (цифровых отпечатков аудиофайлов), которые уже используются для детектирования музыкальных композиций в приложении «Shazam» компании Shazam Entertainment [2], а также при добавлении видеофайлов и аудиофайлов в видеохостинг Youtube [3]. Однако зачастую, самые простые программные средства позволяют обходить такое препятствие с помощью ускорения, замедления композиции на доли секунды, «смены высоты тона» [4]. Такая модификация не заметна для человеческого уха, однако, вычисление цифрового отпечатка приводит к его другому значению, относительно цифрового отпечатка первоначальной версии аудиофайла, что позволяет обходить стороной препятствие по добавлению неуникального файла на стороне хостинга.

Таким образом, проблема автоматизированного анализа зарегистрированных акустических сигналов в задачах акустического контроля различных объектов является актуальной.

Исследование выполнено в рамках реализации Программы поддержки научно-педагогических работников ФГБОУ ВО «Алтайский государственный университет», проект «Разработка метода выявления неправомерного воздействия на аудиофайлы на основе многомерного анализа амплитудно-частотных характеристик аудиосигналов».

Методы и материалы. Для проведения исследования по выявлению информативных признаков из аудиосигналов была использована библиотека Librosa [5] интерактивной оболочки Jupyter Notebook для языка программирования Python.

В качестве сопоставляемых файлов были использованы акустические сигналы следующего формата:

1. Wav_1 – запись акустического сигнала с заданной амплитудой в частотном диапазоне 512–1024 Гц длительностью 3 с. (сирена скорой помощи №1)
2. Wav_2 – запись акустического сигнала с заданной амплитудой в частотном диапазоне 512–1024 Гц длительностью 3 с. (сирена скорой помощи №2)
3. Wav_3 – запись акустического сигнала с заданной амплитудой в частотном диапазоне 256–1024 Гц длительностью 2 с. (звонок телефона №1)
4. Wav_4 – запись акустического сигнала с заданной амплитудой в частотном диапазоне 256–1024 Гц длительностью 2 с. (звонок телефона №2)
5. Wav_5 – запись акустического сигнала с заданной амплитудой в частотном диапазоне 20–4098 Гц длительностью 2 с. (звук выстрела №1)
6. Wav_6 – запись акустического сигнала с заданной амплитудой в частотном диапазоне 20–4098 Гц длительностью 2 с. (звук выстрела №2)

Для построения вектора признаков аудиофайлов использовалась глубокая сверточная нейронная сеть CREPE (Convolutional Representation for Pitch Estimation) [6].

Алгоритм сопоставления информативных признаков из аудиофайла и сопоставления их между собой представлен на рис. 1.

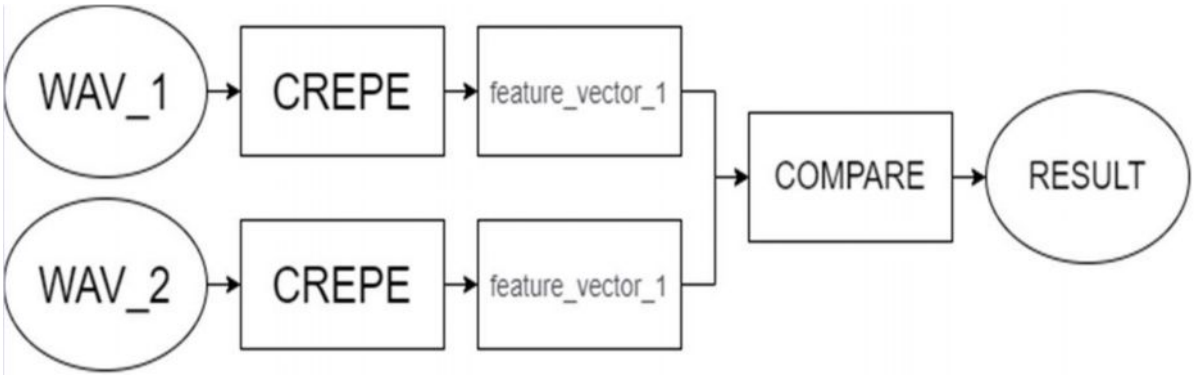


Рис.1. Алгоритм выявления неправомерного воздействия на аудиофайл

Два сравниваемых аудиофайла подаются на вход свёрточной нейронной сети CREPE, получающие векторы признаков (feature_vector), которые представлены тремя столбцами:
- момент времени (в данной работе эмпирически использован шаг в 100 мс);
- частота основного тона в Гц, определённая в этот момент времени;
- вероятность точности определённого значения частоты.

Фрагмент получаемого вектора представлен на рис.2.

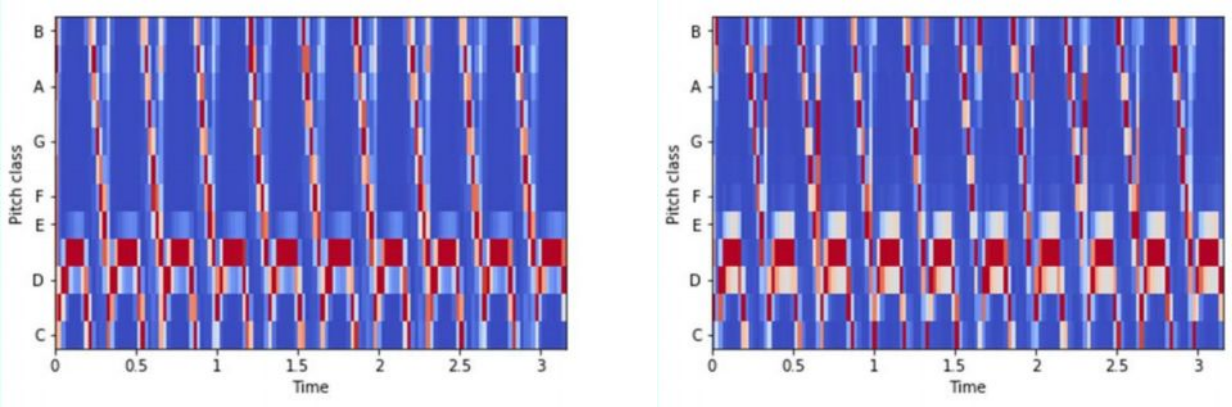
[(0.0, 440.69, 0.89),
(0.1, 440.68, 0.97),
(0.2, 440.03, 0.96),
(0.3, 330.93, 0.94),
…,
(3.3, 440.23, 0.97),
(3.4, 441.35, 0.92)]

Рис.2. Фрагмент получаемого вектора признаков с помощью Crepe

В блоке Compare происходит преобразование полученного вектора в битовую последовательность, где переходы между частотами в векторе кодируются, как 01, если происходит подъём по частоте, 10 – если вниз, 00 – если нет изменения. Для рис.2 такая последовательность будет иметь следующий вид: 101010…01. Работоспособность предложенного преобразования показана в работах [7, 8].
На этапе Compare далее рассчитывается степень схожести двух отпечатков аудиофайлов по следующей формуле:

$$Q = 100\% - \frac{\sum(N1(O1\hat{}O))}{N0} * 100, \qquad (1)$$

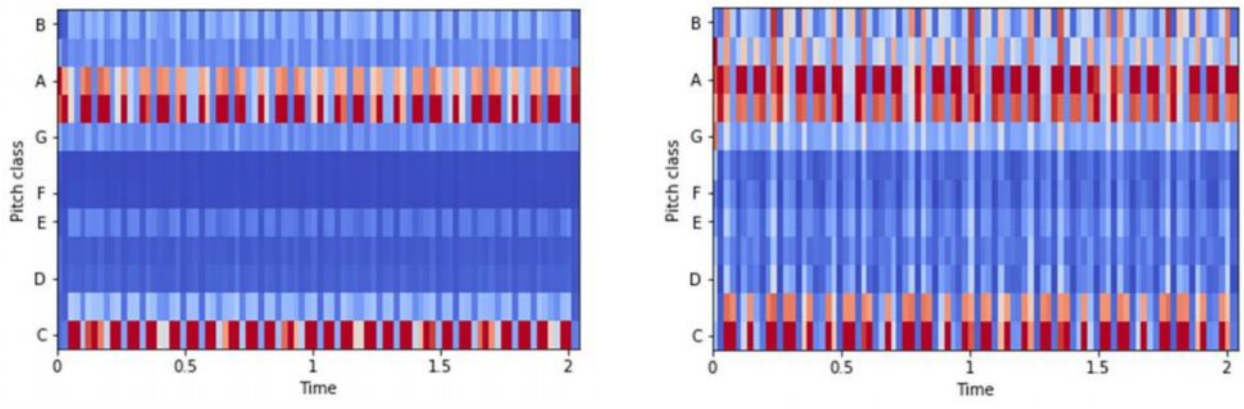
где Q – степень похожести, O1, O2 – БП, N1() – количество единиц в результате операции XOR, N0 – длина наименьшего из отпечатков.
Для визуализации записи звуковых сигналов в данной работе использованы хромаграммы. Хромаграмма «собирает» все гармоники, объединяясь с частотой основного тона, что нормирует участки сигнала к оси частот, удобной для идентификации аудиосигнала. По оси Y отображается 12 полутонов (согласно западной музыкальной нотации), соответствующих стандартной октаве (C, C ♯ , D, D ♯ , E, F, F ♯ , G, G ♯ , A, A ♯ , B), по оси X – время, что наглядно демонстрирует как изменяется основная частота сигнала в каждый момент времени.
Результаты исследований. *Сирены скорой помощи*
Хромаграммы сигналов WAV_1 и WAV_2 (длительность 3 секунды) представлены на рис.3a, 3b: Процент сходства битовых отпечатков составил 65,57%.



a b

Рис. 3. STFT -хромаграмма WAV.
a - WAV_1, b - WAV_2

Звонки
Хромаграммы сигналов (длительность 2 секунды) представлены на рис.4a, 4b. Процент сходства битовых отпечатков составил 64.10%.

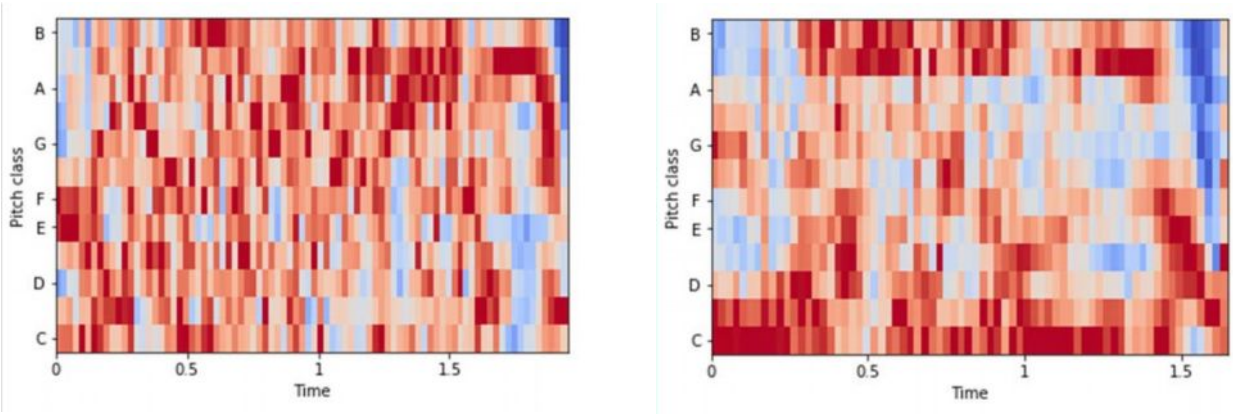


a b

Рис. 4. STFT -хромаграмма WAV.
a - WAV_3, b - WAV_4

Выстрелы

Хромаграммы сигналов (длительность 2 секунды) представлены на рис.5a, 5b. Процент сходства битовых отпечатков составил 61.29%.



a b

Рис. 4. STFT -хромаграмма WAV.
a - WAV_5, b - WAV_6

При этом сравнение битовых отпечатков WAV_1 и WAV_3 показало степень сходства - 27.97%.

Выводы. В работе предложена методика получения «векторов признаков», характеризующих информативные свойства записей акустических сигналов, основанная на использовании нейронных сетей. Данная методика апробирована на примере реальных звуковых сигналов с различными амплитудно-частотными характеристиками. Описанный подход обеспечил возможность адекватного сопоставления цифровых отпечатков акустических сигналов и позволил оценивать их сходство с точностью не ниже 60%.

Результаты работы могут быть использованы для идентификации записей акустических сигналов различной физической природы, в том числе применены к задачам голосовой биометрии. Они могут также использоваться при построении автоматизированных экспертных систем для сравнения различных аудиоданных между собой.

Список литературы

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. 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).

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. Audacity / [Электронный ресурс]. URL: http://www.audacityteam.org/about/features/ (дата обращения 22.03.2021).

5. Librosa [Электронный ресурс] / librosa development team. URL: https://librosa.org/ (дата обращения 22.06.2020).

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

7. Мансуров А. В. Способ формирования цифрового отпечатка аудиофайла на основе вектора признаков, получаемого с использованием Constant-Q и Фурье преобразований [Текст] / А.В. Мансуров, П.С.Ладыгин // 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

8. Ладыгин П.С. Сравнение векторов признаков аудиофайлов, полученных с помощью хроматограмм и питч-трекера CREPE [Текст] / П.С.Ладыгин, А.В.Мансуров, Д.Д.Рудер // Проблемы правовой и технической защиты информации. – 2020. - № 8. - С. 29- 34.