



Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Московский государственный технический университет  
имени Н.Э. Баумана  
(национальный исследовательский университет)»  
(МГТУ им. Н.Э. Баумана)

---

ФАКУЛЬТЕТ ИУ «Информатика и системы управления»

КАФЕДРА ИУ-7 «Программное обеспечение ЭВМ и информационные технологии»

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА  
К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ  
НА ТЕМУ:**

***«Анализ и классификация известных методов  
распознавания музыкальных фрагментов на основе  
интерпретации ритма»***

Студент      ИУ7-52Б      \_\_\_\_\_ Карпова Е. О.

Консультант      \_\_\_\_\_ Гаврилова Ю. М.

Руководитель      \_\_\_\_\_ Строганов Ю. В.

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Московский государственный технический университет имени Н.Э. Баумана  
(национальный исследовательский университет)»  
(МГТУ им. Н.Э. Баумана)

---

УТВЕРЖДАЮ

Заведующий кафедрой ИУ-7  
(Индекс)

\_\_\_\_\_ И. В. Рудаков  
(И.О.Фамилия)

«16» сентября 2022 г.

**ЗАДАНИЕ**  
**на выполнение научно-исследовательской работы**

по теме

**«Анализ и классификация известных методов распознавания музыкальных  
фрагментов на основе интерпретации ритма»**

Студент группы ИУ7-52Б

**Карпова Екатерина Олеговна**

Направленность НИР

**учебная**

Источник тематики

**НИР кафедры**

График выполнения НИР: 25% к 6 нед., 50% к 9 нед., 75% к 12 нед., 100% к 15 нед.

**Техническое задание**

**Провести обзор существующих методов распознавания музыкальных фрагментов на основе интерпретации ритма и сформулировать критерии их сравнения. Провести анализ предметной области распознавания музыкальных фрагментов на основе интерпретации ритма. Классифицировать существующие методы распознавания музыкальных фрагментов на основе интерпретации ритма.**

**Оформление научно-исследовательской работы:**

Расчетно-пояснительная записка на **12-20** листах формата А4.

Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т. п.)

Презентация на **6-10** слайдах.

Дата выдачи задания «16» сентября 2022 г.

**Руководитель НИР**

\_\_\_\_\_ **Строганов Ю. В.**  
(Подпись, дата) (И.О.Фамилия)

**Консультант**

\_\_\_\_\_ **Гаврилова Ю. М.**  
(Подпись, дата) (И.О.Фамилия)

**Студент**

\_\_\_\_\_ **Карпова Е. О.**  
(Подпись, дата) (И.О.Фамилия)

# Содержание

Определения	4
Обозначения и сокращения	6
Реферат	7
Введение	8
<b>1 Анализ предметной области</b>	<b>9</b>
1.1 Актуальность задачи . . . . .	9
1.2 Постановка задачи и её приложения . . . . .	9
1.3 Этапы, предшествующие непосредственному распознаванию му- зыки . . . . .	10
1.3.1 Интерпретация ритма . . . . .	10
1.3.2 Транскрибирование . . . . .	12
1.4 Обзор методов распознавания музыкальных фрагментов на ос- нове ритмической интерпретации . . . . .	14
1.4.1 Поиск редакционного расстояния . . . . .	15
1.4.2 Поиск расстояния Хэмминга . . . . .	16
1.4.3 Поиск Евклидова интервального векторного расстояния .	16
1.4.4 Поиск интервально-разностного векторного расстояния .	17
1.4.5 Поиск обменного расстояния . . . . .	18
1.4.6 Поиск хронотонического расстояния . . . . .	18
1.5 Вывод . . . . .	21
<b>2 Классификация методов распознавания музыкальных фраг- ментов на основе интерпретации ритма</b>	<b>22</b>
2.1 Определение критерия классификации . . . . .	22
2.2 Классификация методов . . . . .	22
2.2.1 Классификация методов по способу записи ритмическо- го рисунка . . . . .	22
2.2.2 Классификация методов по качественным характери- стам . . . . .	23
2.3 Вывод . . . . .	24

<b>Заключение</b>	<b>25</b>
<b>Список использованных источников</b>	<b>26</b>

# Определения

В настоящей расчетно-пояснительной записке применяют следующие термины с соответствующими определениями.

Ритм — организация музыки во времени, последовательность длительностей — звуков и пауз (другими словами, последовательность ударов воображаемого метронома, согласованная в каждый момент времени с музыкальным рисунком проигрываемой композиции) [1].

Распознавание музыки — перевод звукового сигнала в цифровое представление, с которым, в дальнейшем, и происходит взаимодействие (цифровое) [2].

Интерпретация ритма — вариант восприятия ритмического паттерна музыкального фрагмента человеком.

Транскрибирование — это перевод содержания аудиофайла в символьный формат [3].

Вероятностная модель — это математическая модель реального явления, содержащего элементы принципиально неустранимой неопределённости (случайности).

Нейронная сеть — математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма.

*MIDI*-файл — стандарт цифровой звукозаписи на формат обмена данными между электронными музыкальными инструментами, интерфейс позволяет единообразно кодировать в цифровой форме такие данные как нажатие клавиш, настройку громкости и других акустических параметров, выбор тембра, темпа, тональности и др., с точной привязкой во времени [4].

Коннекционизм — один из подходов в области искусственного интеллекта, когнитивной науки, нейробиологии, психологии и философии сознания, моделирующий мыслительные или поведенческие явления процессами становления в сетях из связанных между собой простых элементов.

Скрытая переменная — переменные, которые не могут быть измерены в явном виде, а могут быть только выведены через математические модели с использованием наблюдаемых переменных.

Скрытая модель Маркова — статистическая модель, имитирующая рабо-

ту процесса, похожего на марковский процесс с неизвестными параметрами, и задачей ставится разгадывание неизвестных параметров на основе наблюдаемых [5].

Марковский процесс — случайный процесс, эволюция которого после любого заданного значения временного параметра  $t$  не зависит от эволюции, предшествовавшей  $t$ , при условии, что значение процесса в этот момент фиксировано [6].

Онсет — начало ноты или удар.

# Обозначения и сокращения

*MIDI — Musical Instrument Digital Interface*

*HMM — Hidden Markov Model*

# Реферат

Целью данной работы является анализ известных методов распознавания музыкальных фрагментов на основе интерпретации ритма и их классификация.

В результате проведён анализ заданной предметной области, были рассмотрены существующие методы распознавания музыкальных фрагментов на основе интерпретации ритма и была проведена их классификация на основе сформулированных критериев сравнения.

Ключевые слова: распознавание музыкальных фрагментов, интерпретация ритма, транскрибирование, поиск расстояния, ноты, ритм, расстояние Хэмминга, расстояние Евклида, интервально-разностное расстояние, редакционное расстояние, хромотоническое расстояние.



# Введение

Уже долгое время музыка является неотъемлемой частью жизни человека: прослушивание музыки может повысить уровень концентрации, поднять настроение, настроить на определённую деятельность и способствовать продуктивности. Научно доказано, что прослушивание музыки оказывает значительное влияние на состояние человеческого организма [7].

Благодаря работе мозга и слухового аппарата, человек может запоминать ритм музыки и впоследствии распознавать её заново.

С развитием техники способностью распознавать музыку смог обладать не только человек, но и различные устройства, появились новые возможности обработки звуковых сигналов.

Музыка складывается из многих компонентов, таких как мелодия, темп, громкость и т.д. Одним из основных элементов, характеризующих музыкальный фрагмент, является также его ритм [8].

Целью данной работы является анализ известных методов распознавания музыкальных фрагментов на основе интерпретации ритма и их классификация.

Для достижения поставленной цели необходимо решить следующие задачи:

- провести анализ предметной области распознавания музыкальных фрагментов на основе интерпретации ритма;
- провести обзор существующих методов распознавания музыкальных фрагментов на основе интерпретации ритма;
- сформулировать критерии их сравнения;
- классифицировать данные методы.

# 1 Анализ предметной области

В данной части рассматривается актуальность задачи, излагаются общие сведения об этапах, предшествующих распознаванию музыки по ритмическому паттерну, рассматриваются существующие методы распознавания музыкальных фрагментов на основе интерпретации ритма.

## 1.1 Актуальность задачи

В настоящее время технология распознавания музыки становится всё более востребованной [9].

Ритм — один из наиболее фундаментальных элементов, позволяющих охарактеризовать музыкальный фрагмент [10]. Обобщая, можно сказать, что совокупность ритмических последовательностей формирует структуру всего музыкального фрагмента, поэтому извлечение и анализ этих паттернов позволяет идентифицировать музыкальную композицию.

Вычислительная теория музыки предлагает широкий спектр интересных геометрических, комбинаторных и алгоритмических задач [11]. Измерение сходства между ритмами является фундаментальной проблемой данной теории со многими приложениями, такими как поиск музыкальной композиции и решение проблем нарушений авторских прав [10].

## 1.2 Постановка задачи и её приложения

Под задачей распознавания музыки подразумевается перевод звукового сигнала в цифровое представление [7]. Данное представление преобразовывается в массив данных, которые однозначно идентифицируют заданный фрагмент среди остальных. Обычно, наборы таких данных хранятся в базах данных. При необходимости распознать очередной фрагмент полученное представление сравнивается с другими представлениями музыкальных сигналов, которые уже хранятся в базе данных. Если выявляется совпадение, пользователь получает информацию об интересующем музыкальном фрагменте.

Технология распознавания музыки может решать следующие задачи:

- задача поиска песни по музыкальному фрагменту;
- задача определения принадлежности песни к этнической культуре;
- задача нахождения сходства между музыкальными композициями;
- задача контроля соблюдения авторских прав.

## 1.3 Этапы, предшествующие непосредственному распознаванию музыки

### 1.3.1 Интерпретация ритма

Интерпретация ритма — вариант восприятия ритмического паттерна музыкального фрагмента человеком. Различные интерпретации будут далее рассмотрены на примере. Одним из самых простых и популярных ритмов является *Clave Son*, свойственный для танцевальной музыки, например, сальсы, и впервые появившийся на острове Куба [11]. Для музыкантов данный ритм зачастую записывается с использованием стандартной нотной записи, варианты которой представлены на рисунке 1.1.

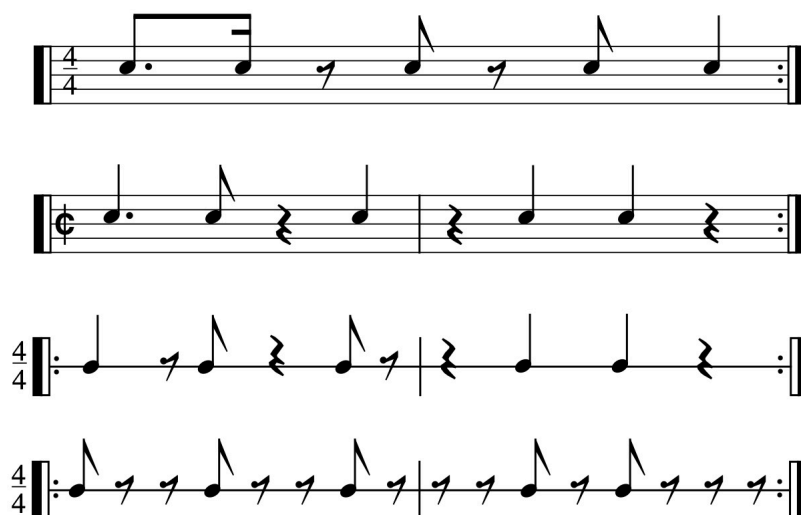


Рисунок 1.1 – Стандартная нотная запись ритма *Clave Son* [11]

Для многих ударников, не умеющих читать ноты, или обывателей, также не знакомых с музыкальной нотацией, данный ритм будет более понятен в записи, указанной на рисунке 1.2.



Рисунок 1.2 – *Box Notation Method* записи ритма *Clave Son* [11]

Данный метод известен как *Box Notation Method*, разработанный Филипом Харландом из Калифорнийского университета в Лос-Анджелесе в 1962 году и также известный как *TUBS (Time Unit Box System)*. Наличие выбранного символа в данной записи обозначает ноту, а отсутствие — паузу.

Аналогичны по смыслу и интерпретации на рисунке 1.3.

**X . . X . . X . . . X . X . . .**  
**1 0 0 1 0 0 1 0 0 0 1 0 1 0 0 0**

Рисунок 1.3 – Иные варианты интерпретации ритма без музыкальной нотации [11]

Немного отличным от предыдущих является вариант представления, при котором записываются в ряд числа, обозначающие длины интервалов между последовательными ударами. Данная интерпретация представлена на рисунке 1.4.

**3 3 4 2 4**

Рисунок 1.4 – Интервальный вариант интерпретации ритма без музыкальной нотации [11]

Существуют также геометрические варианты интерпретации ритма, например, в форме выпуклого многоугольника, что представлено на рисунке 1.5.

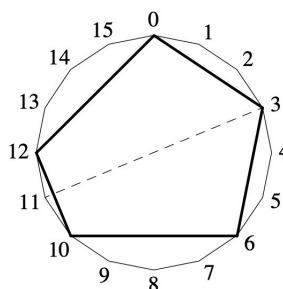


Рисунок 1.5 – Геометрический вариант интерпретации ритма без музыкальной нотации [11]

Данное представление более реалистично, потому что временные линии ритма цикличны, а все другие представления, рассмотренные выше, имеют начало и конец паттерна, расположенные физически на расстоянии друг от друга. При таком рассмотрении, ритм может быть представлен как вектор  $X = (x_1, x_2, \dots, x_n)$ , где  $x_i$  — число вершин, пропущенных  $i$ -м ребром многоугольника, начиная с вершины с номером 0.

### 1.3.2 Транскрибирование

Транскрибирование — это перевод содержания аудиофайла в символьный формат.

Могут быть выделены следующие виды транскрибирования:

1. классифицирующие, среди которых использование скрытых марковских моделей и вероятностный анализ скрытых компонентов;
2. нейросетевые, подразделяющиеся на одноуровневые и многоуровневые;
3. комбинированные, например, с использованием адаптивных осцилляторов и нейросетей.

Согласно источникам [12] и [13], большинство перечисленных ниже методов транскрибирования применимы для ритмического транскрибирования.

#### **Автоматическое транскрибирование музыки с использованием моделей событий нот**

В источнике [14] описано автоматическое транскрибирование музыки с использованием моделей событий нот. Данный метод предполагает наличие трёх вероятностных моделей — модели тишины и модели событий нот, представляющих собой скрытые марковские модели (НММ), и музыковедческой модели. Изначальные данные для транскрибирования обычно извлекаются из *MIDI*-файлов. Модель тишины моделирует временные области, в которых не звучат ноты, а события нот описываются с помощью скрытой модели с тремя состояниями, соответствующими типичным значениям признаков для

нот. Музыковедческая модель, используя метод оценки музыкальной тональности, находит наиболее вероятные переходы между моделями нот и тишины. Схема взаимодействия моделей проиллюстрирована рисунком 1.6.

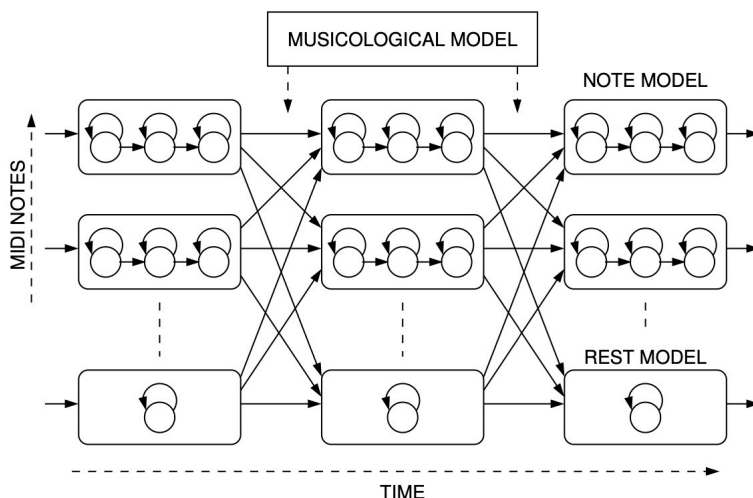


Рисунок 1.6 – Схема взаимодействия моделей при автоматическом транскрибировании музыки с использованием модели событий нот [14]

## Построение инвариантной к сдвигам модели скрытых переменных

В источнике [15] используется подход, представляющий собой построение инвариантной к сдвигам модели скрытых переменных. Основой подхода является теория вероятностного анализа скрытых компонентов [16], при котором входные спектрограммы аппроксимируются вероятностным распределением, зависящим от времени и частоты, а затем распределение раскладывается на отдельные спектральные компоненты.

## Транскрибирование с помощью модели слухового восприятия

В источнике [17] использован "коннекционистский подход". Это означает, что музыкальный фрагмент рассматривается как набор элементов и сетевых связей между ними. Происходит переход от осциллограммы звукового сигнала к спектральному частотно-временному представлению с помощью модели слухового восприятия, полученной на основе бионических исследований слуха человека, и фильтрации звукового сигнала с частотным преобразованием. Элементами композиции, в рамках данного подхода, являются отдельные адаптивные осцилляторы, представляющие собой модели гармонических синусоидальных сигналов, адаптирующиеся под входящий сигнал с помощью

нейросетевых методов. Сеть осцилляторов выделяет пики частот основного тона и обертонов. Каждый осциллятор отслеживает определённую ноту, подстраиваясь под входящий поток событий.

### **Транскрибирование с помощью выявления во входящем звуковом сигнале онсетов**

В источнике [18] используется выделение во фрагменте начал нот — онсетов. Метод основан на использовании сонограммы звуковой записи, которая в дальнейшем исследуется с помощью нейросетевых методов. Онсеты определяются в два этапа: сначала определяются те частотные полосы, в которых появились компоненты ноты, а потом определяется нота или ноты, способные породить такой набор частотных компонент. Поэтому для решения этой задачи строится двухуровневая нейронная сеть. Используемая нейронная сеть имеет два уровня. Первый уровень состоит из отдельных нейросетей, обучаемых на эталонных произведениях, для которых вручную описаны начало каждой ноты. Благодаря двухуровневой структуре верхний уровень позволяет корректировать результаты работы нейросетей нижнего уровня и получать более точные результаты.

## **1.4 Обзор методов распознавания музыкальных фрагментов на основе ритмической интерпретации**

После выбора метода интерпретации и транскрибирования музыкального фрагмента необходимо распознать музыкальный фрагмент на основе полученной ритмической картины. Для этого рассмотрим методы вычисления ритмического сходства музыкальных произведений.

В основе любого алгоритма распознавания музыкального фрагмента лежит задача определения сходства ритмической составляющей. Выбор метода решения этой задачи частично предопределен способом представления ритма. В соответствии с рассмотренными вариантами интерпретации ритма, рассмотрим методы измерения сходства двух ритмов, представленных последо-

вательностями символов.

### 1.4.1 Поиск редакционного расстояния

Одним из наиболее распространённых методов распознавания музыкального фрагмента в ритмической записи последовательностью символов является поиск редакционного расстояния. Редакционное расстояние — это минимальное количество вставок, удалений и замен, необходимых для преобразования одной строки в другую. Для каждой операции над строкой вводится условная цена. Суммарная стоимость всех произведённых операций и будет являться редакционным расстоянием между данными строками.

Цены за редакционные операции можно обозначить и определить следующим образом:

- $\omega(\alpha, \beta) = 1$  — цена замены при  $\alpha \neq \beta$ ;
- $\omega(\alpha, \alpha) = 0$  — цена эквивалентной замены;
- $\omega(\emptyset, \beta) = 1$  — цена вставки;
- $\omega(\alpha, \emptyset) = 1$  — цена удаления.

Для описания алгоритма вычисления редакционного расстояния, например, Левенштейна, применяется рекуррентное (т.е. использующее результаты предшествующих вычислений на следующем шаге) соотношение

$$D(s_1[1..i], s_2[1..j]) = \begin{cases} 0, & \text{если } i = 0, j = 0, \\ i, & \text{если } i > 0, j = 0, \\ j, & \text{если } i = 0, j > 0, \\ \min( & D(s_1[1..i], s_2[1..j-1]) + 1, \\ & D(s_1[1..i-1], s_2[1..j]) + 1, \\ & D(s_1[1..i-1], s_2[1..j-1]) + m), \end{cases} \quad (1.1)$$

где

$$m = \begin{cases} 0, & \text{если } s_1[i] = s_2[j], \\ 1, & \text{иначе.} \end{cases} \quad (1.2)$$



Таким образом, сходство с музыкальным фрагментом будет зафиксировано, если редакционное расстояние между их ритмической записью окажется минимальным.

### 1.4.2 Поиск расстояния Хэмминга

Для двоичных числовых последовательностей мерой сходства может являться расстояние Хэмминга, широко используемое в теории программирования. Расстояние Хэмминга — это количество позиций, где элементы не совпадают. Следовательно, это один из самых простых видов сопоставления шаблонов. При этом подходе каждый ритм представлен вектором  $X = (x_1, x_2, \dots, x_n)$ , где  $x_i$  представляют собой бинарные характеристики ритма. Если нота проигрывается в  $i$ -ю единицу времени, то  $x_i = 1$ , иначе  $x_i = 0$ . Таким образом, ритмы представляются точками в  $n$ -мерном бинарном векторном пространстве (гиперкубе). Расстояние Хэмминга между двумя точками  $X = (x_1, x_2, \dots, x_n)$  и  $Y = (y_1, y_2, \dots, y_n)$  в этом пространстве определяется выражением:

$$d(X, Y) = \sum_{i=1}^n |x_i - y_i|, \quad (1.3)$$

Расстояние Хэмминга легко вычисляется за время  $O(n)$ . Данный метод выявляет наличие несоответствия, но не его величину.

### 1.4.3 Поиск Евклидова интервального векторного расстояния

Некоторые алгоритмы распознавания ритма и музыкальных паттернов используют интервалы между онсетами, то есть последовательными началами нот в ритме, в качестве основных признаков для измерения сходства. Этот подход и его варианты подходит для измерения подобия ритмов, больше, чем рассмотренные ранее, чем расстояние Хэмминга.

Каждая ритмическая структура может быть представлена вектором чисел, совпадающих с длиной этих интервалов. Более конкретно, ритм может

быть представлен вектором  $X = (x_1, x_2, \dots, x_n)$ , где  $x_i$  — число вершин, пропущенных  $i$ -м ребром многоугольника, начиная с вершины, помеченной 0.

По сути, это то же самое, что и последовательность временных интервалов между онсетами, поскольку временной интервал равен количеству пропущенных вершин, увеличенному на 1. Различие между двумя ритмами  $X = (x_1, x_2, \dots, x_n)$  и  $Y = (y_1, y_2, \dots, y_n)$  может быть вычислено с помощью Евклидова расстояния между двумя интервальными векторами  $X$  и  $Y$ , в  $n$ -мерном векторном пространстве, как:

$$d(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (1.4)$$

Векторы интервалов могут быть вычислены из заданных двоичных последовательностей за время  $O(n)$ , после чего Евклидово расстояние может быть вычислено за время  $O(m)$ , где  $m$  — количество онсетов, что приводит к итоговой оценке в  $O(n)$ .

#### 1.4.4 Поиск интервально-разностного векторного расстояния

Музыкальный паттерн может быть представлен вектором разности ритмов [19]. Если  $T = (t_1, t_2, \dots, t_n)$  — вектор интервалов времени между началами нот ритма (онсетами), то вектор разности ритма определяется как  $X = (x_1, x_2, \dots, x_{n-1})$ , где  $x_i = \frac{t_{i+1}}{t_i}$ .

В контексте рассматриваемых здесь циклических ритмов  $x_n = \frac{t_1}{t_n}$  размерность вектора разности ритмов будет равна  $n$ , а не  $n - 1$ . Тогда пусть  $X = (x_1, x_2, \dots, x_n)$  и  $Y = (y_1, y_2, \dots, y_n)$  обозначают два интервально-разностных вектора. Различия между  $X$  и  $Y$  определяются следующим образом:

$$d(X, Y) = \sum_{i=1}^n \frac{\max(x_i, y_i)}{\min(x_i, y_i)} - n. \quad (1.5)$$

Данная мера расстояния также вычисляется за время  $O(n)$ .

### 1.4.5 Поиск обменного расстояния

Задача сравнения двух двоичных строк одинаковой длины с одинаковым количеством единиц предполагает возможность проведения операции обмена. Обмен — это смена местами соседних единицы и нуля в двоичной строке. Расстояние перестановки между двумя ритмами — это минимальное количество обменов, необходимых для преобразования одного ритма в другой. Например, ритм, заданный последовательностью  $[x.x..x.x.x]$ , где  $x$  обозначает позицию онсета, можно преобразовать в ритмическую последовательность  $[x.xxx.xx.x.x]$  четырьмя обменами, а именно заменой третьей, пятой, шестой и седьмой доли на предшествующие им паузы. Разностное расстояние можно вычислить путем фактического выполнения перестановки, но это неэффективно [20]. Если  $X$  имеет единицы в первых  $\frac{n}{2}$  позициях и нули в других местах, и если  $Y$  имеет единицы в последних  $\frac{n}{2}$  позициях и нули в других, то потребуется по крайней мере квадратичное число обменов. С другой стороны, если вместо этого мы сравним вычисленные расстояния, получится гораздо более эффективный алгоритм. Сначала необходимо проанализировать двоичную последовательность и сохранить вектор позиций онсетов. Например,  $X$  и  $Y$  из примера выше дают векторы  $U = (u_1, u_2, \dots, u_7) = (1, 3, 5, 6, 8, 10, 12)$  и  $V = (v_1, v_2, \dots, v_7) = (1, 3, 4, 6, 7, 9, 11)$  соответственно. Разница между  $u_i$  и  $v_i$  равна количеству обменов, которые необходимо выполнить, чтобы привести два вектора в соответствие. Следовательно, в общем случае расстояние перестановки между двумя векторами позиций онсетов  $U$  и  $V$  с  $k$  онсетами определяется как:

$$d(U, V) = \sum_{i=1}^n |u_i - v_i|. \quad (1.6)$$

Вычисление  $U$  и  $V$  из  $X$  и  $Y$  выполняется за время  $O(n)$ .

### 1.4.6 Поиск хронотонического расстояния

Далее рассматривается поиск хронотонического расстояния на примере. Пусть музыкальный фрагмент задан, как показано на рисунке 1.7.



Рисунок 1.7 – Музыкальный фрагмент для поиска хронотонического расстояния [21]

Тогда хронотоническая цепочка может быть записана как  $ch = [\frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{8}, \frac{3}{8}]$ . Видно, что кратчайшее время в цепочке —  $\frac{1}{16}$  удара. Необходимо квантовать время на атомарные удары по  $\frac{1}{16}$  удара. Теперь получается цепочка  $ch = [4, 4, 4, 4, 2, 2, 1, 1, 2, 2, 6, 6, 6, 6, 6, 6] \cdot (\frac{1}{16})$ . Данную цепочку можно изобразить на графике, по оси  $x$  откладывая номер атомарного удара в последовательности, а по оси  $y$  — хронотоническое значение, как показано на рисунке 1.8.

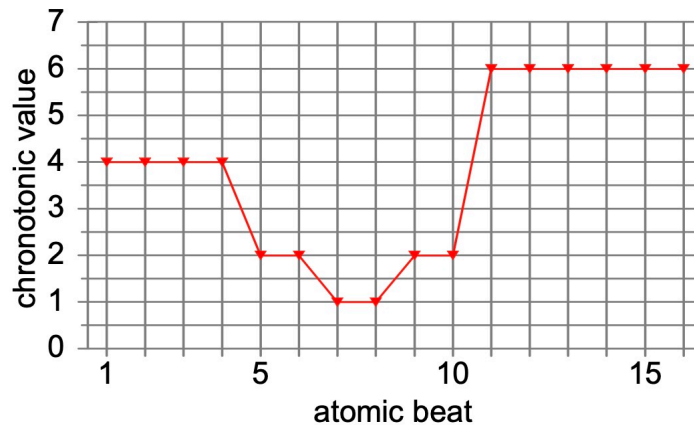


Рисунок 1.8 – Зависимость хронотонического значения от номера удара [21]

Если взять другую последовательность, как на рисунке 1.9, то можно выполнить отображение первой последовательности на вторую, используя отражение графика одной из последовательностей относительно оси  $x$ .



Рисунок 1.9 – Второй музыкальный фрагмент для поиска хронотонического расстояния [21]

Данную последовательность можно записать цепочкой  $ch' = [4, 4, 4, 4, 2, 2, 1, 3, 3, 3, 1, 1, 4, 4, 4, 4] \cdot (\frac{1}{16})$ . Проведённое отображение показано на рисунке 1.10 тёмно-синим цветом.

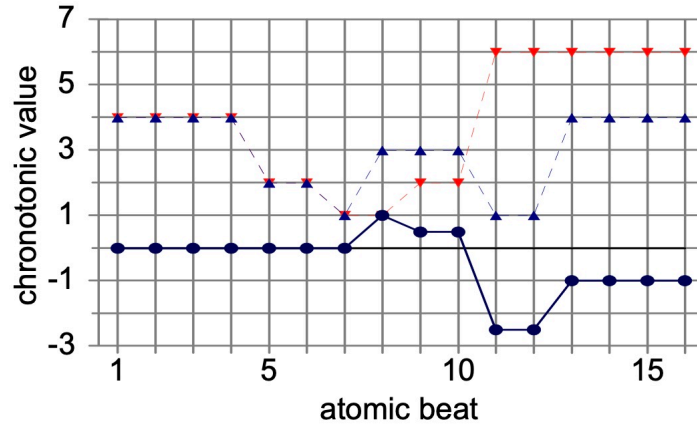


Рисунок 1.10 – Хронотоническое преобразование [21]

Такой механизм преобразования позволяет нам соотносить любые цепи одинаковой длины. Если две цепи имеют разную длину, необходимо увеличить более короткую цепь, чтобы она соответствовала более длинной цепи.

Степень подобия между двумя цепочками определяется близостью цепочки отражения  $R = [0, 0, 0, 0, 0, 0, 0, 1, 0.5, 0.5, -2.5, -2.5, 1, 1, 1, 1]$  и оси  $0\ x$ : чем они ближе, тем сильнее сходство.

Если взять две цепочки разной длины, то для уравнивания двух цепочек надо умножить хроноты (музыкальные длительности) одной цепочки на коэффициент (в случае, если, например, восьмые доли дополнить до четвертных,  $= 2$ ). Модель сходства не должна зависеть от соотношения длин цепочек, поэтому запишем формулу, соответствующую методу Шепарда:

$$F_1 = e^{-k_1 \cdot \ln^2(a)}, \quad (1.7)$$

где  $F_1$  — показатель подобия, когда длина двух цепочек различна, то есть  $|ch| = a \cdot |ch'|$ , а  $k_1$  — эмпирическая константа.

Далее нужно записать показатель подобия, связанный с цепочкой отражения  $R = [r_1, r_2, \dots, r_n]$ , учитывая, что чем выше сходство, тем длиннее вектор:

$$\|F_2\| = \left\| \left( \frac{e^{r_1^2 - \frac{k_2}{n}}}{\sqrt{n}} \quad \frac{e^{r_2^2 - \frac{k_2}{n}}}{\sqrt{n}} \quad \dots \quad \frac{e^{r_n^2 - \frac{k_2}{n}}}{\sqrt{n}} \right)^T \right\|, \quad (1.8)$$

что можно также записать как

$$\|F_2\| = \sqrt{\sum_{i=1}^n \left( \frac{e^{r_i^2 - \frac{k_2}{n}}}{\sqrt{n}} \right)^2}, \quad (1.9)$$

где  $n$  — длина цепочек в атомарных ударах (после выравнивания),  $r_i$  —  $i$ -я компонента цепочки отражения,  $k_2$  — эмпирическая постоянная.

Хронтоническое сходство будет выражаться в виде:

$$S = F_1 \cdot ||F_2||. \quad (1.10)$$

Описанная гипотеза была проверена в эксперименте [21].

## 1.5 Вывод

В данной части была рассмотрена актуальность задачи, изложены общие сведения об этапах, предшествующих распознаванию музыки по ритмическому паттерну, рассмотрены существующие методы распознавания музыкальных фрагментов на основе интерпретации ритма.

## **2 Классификация методов распознавания музыкальных фрагментов на основе интерпретации ритма**

В данной части определяется критерий классификации методов распознавания музыкальных фрагментов на основе интерпретации ритма и проводится их классификация.

### **2.1 Определение критерия классификации**

Согласно [20], в основе любого алгоритма распознавания музыкального фрагмента лежит задача определения сходства ритмической составляющей. Методы измерения сходства двух ритмов, представленных последовательностями символов, рассматривались, исходя из представленных в работе способов интерпретации ритма, поэтому будет проведена классификация именно по этому признаку.

### **2.2 Классификация методов**

#### **2.2.1 Классификация методов по способу записи ритмического рисунка**

Исходя из полученного критерия, была получена схема классификации, которая выглядит следующим образом (2.1).

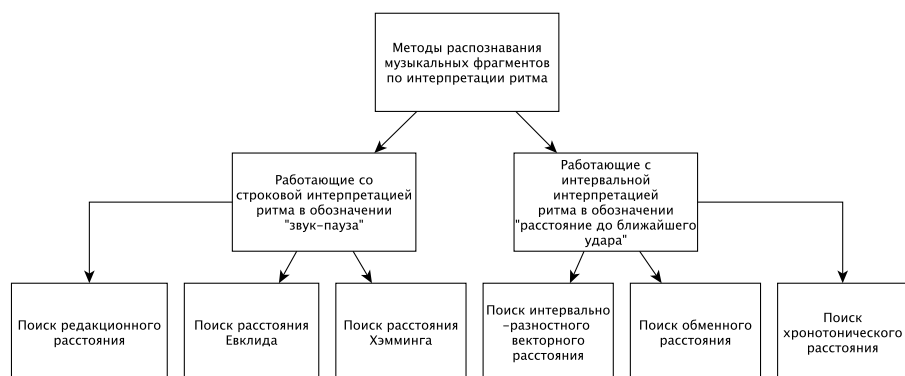


Рисунок 2.1 – Классификация методов распознавания музыкальных фрагментов на основе интерпретации ритма

## 2.2.2 Классификация методов по качественным характеристикам

Кроме наиболее очевидного критерия классификации, можно также провести классификацию и сравнение перечисленных методов по качественным критериям, указанным ниже. Автор источника [20] проводит исследования данных методов распознавания музыкальных фрагментов на основе эксперимента по построению филогенетических деревьев. Данные деревья в контексте проводимого эксперимента отражают эволюционные взаимосвязи между бинарными и троичными ритмами двух семейств.

Сравнение проводится по следующим критериям:

1. средний процент соответствия результатов работы методов действительной близости ритмов для двух семейств (avg fit %);
2. вычислительная сложность (complexity);
3. соответствие соотношениям близости внутренних групп (clustering).

В таблице 2.1 представлена классификация и сравнение существующих методов распознавания музыкальных фрагментов на основе интерпретации ритма. Обозначения: РР — редакционное расстояние, РЕ — расстояние Евклида, РХ — расстояние Хэмминга, ИРВР — интервально-разностное векторное расстояние, ОР — обменное расстояние, ХР — хронотоническое расстояние.



Таблица 2.1 – Классификация и сравнение существующих методов распознавания музыкальных фрагментов на основе интерпретации ритма

Критерий	PP	PE	PX	IPBP	OP	XP
avg fit %	–	85.15	75.1	43.3	100	100
complexity	$O(n)$	$O(n)$	$O(n)$	$O(n)$	$O(n)$	$O(n)$
clustering	–	Средне	Средне	Средне	Средне	Хорошо

В некоторых строках, соответствующих информации о методе поиска редакционного расстояния, нет информации о результатах исследования, так как автор источника [20] не включал этот метод в рассмотрение.

## 2.3 Вывод

В данной части был определен критерий классификации существующих методов распознавания музыкальных фрагментов на основе интерпретации ритма и проведена их классификация и сравнение по качественным критериям.

# Заключение

В ходе данной работы была выполнена поставленная цель: были рассмотрены существующие методы распознавания музыкальных фрагментов на основе интерпретации ритма и была проведена их классификация.

Задачи, решенные для достижения поставленной цели:

1. проведён анализ предметной области распознавания музыкальных фрагментов на основе интерпретации ритма;
2. проведен обзор существующих методов распознавания музыкальных фрагментов на основе интерпретации ритма;
3. сформулированы критерии их сравнения;
4. классифицированы данные методы.

## Список использованных источников

1. Хачай МЮ, Кобылкин КС. МОДИФИКАЦИЯ ОДНОГО АЛГОРИТМА ОПРЕДЕЛЕНИЯ ТЕМПА МУЗЫКАЛЬНОЙ МЕЛОДИИ // СОВРЕМЕННЫЕ ПРОБЛЕМЫ МАТЕМАТИКИ. с. 167.
2. Зацарин Евгений Андреевич, Головин Артём Валерьевич, Ульби Тимур Валерьевич. РАСПОЗНАВАНИЕ МУЗЫКАЛЬНЫХ ФРАГМЕНТОВ, ПОДХОДЫ И НЮАНСЫ // Актуальные научные исследования в современном мире. 2021. № 4-3. С. 45–48.
3. Дудырев ЕО. Обзор современных средств и методов компьютерного транскрибирования музыки // Молодежный научно-технический вестник. 2017. № 6. С. 30–30.
4. Caldwell Dustin. Standard MIDI File Format. 1995.
5. Eddy Sean R. What is a hidden Markov model? // Nature biotechnology. 2004. Т. 22, № 10. С. 1315–1316.
6. Stroock Daniel W. An introduction to Markov processes. Springer Science & Business Media, 2013. Т. 230.
7. Зацарин Е. А. Головин А. В. Ульби Т. В. Распознавание музыкальных фрагментов, подходы и нюансы // Актуальные научные исследования в современном мире. 2021. № 4.
8. Foroughmand Hadrien, Peeters Geoffroy. Deep-rhythm for tempo estimation and rhythm pattern recognition // International Society for Music Information Retrieval (ISMIR). 2019.
9. Verbitskiy Sergey, Berikov Vladimir, Vyshegorodtsev Viacheslav. Eranns: Efficient residual audio neural networks for audio pattern recognition // Pattern Recognition Letters. 2022. Т. 161. С. 38–44.
10. Tsunoo Emiru, Ono Nobutaka, Sagayama Shigeki. Rhythm map: Extraction of unit rhythmic patterns and analysis of rhythmic structure from music acoustic signals // 2009 IEEE International Conference on Acoustics, Speech and Signal Processing / IEEE. 2009. С. 185–188.

11. Toussaint Godfried T. Algorithmic, geometric, and combinatorial problems in computational music theory // Proceedings of X Encuentros de Geometria Computacional. 2003. C. 101–107.
12. Raphael Christopher. Automated Rhythm Transcription. // ISMIR. T. 2001. 2001. C. 99–107.
13. Nakamura Eita, Itoyama Katsutoshi, Yoshii Kazuyoshi. Rhythm transcription of MIDI performances based on hierarchical Bayesian modelling of repetition and modification of musical note patterns // 2016 24th European Signal Processing Conference (EUSIPCO) / IEEE. 2016. C. 1946–1950.
14. Ryyanen Matti P, Klapuri Anssi. Polyphonic music transcription using note event modeling // IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005. / IEEE. 2005. C. 319–322.
15. Benetos Emmanouil, Dixon Simon. A shift-invariant latent variable model for automatic music transcription // Computer Music Journal. 2012. T. 36, № 4. C. 81–94.
16. Smaragdis Paris, Raj Bhiksha. Shift-invariant probabilistic latent component analysis // Journal of Machine Learning Research. 2007. T. 5.
17. Marolt Matija. A connectionist approach to automatic transcription of polyphonic piano music // IEEE Transactions on Multimedia. 2004. T. 6, № 3. C. 439–449.
18. Андреев С.А. Определение моментов начала нот (онсетов) при анализе музыкальных произведений // Сборник докладов Всероссийской конференции Математические методы распознавания образов-13 (ММРО-13).—Макс Пресс Москва. 2007. С. 277–279.
19. Coyle Edward J, Shmulevich Ilya. A system for machine recognition of music patterns // Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181) / IEEE. T. 6. 1998. C. 3597–3600.
20. Toussaint Godfried T [и др.]. A Comparison of Rhythmic Similarity Measures. // ISMIR. 2004.

21. Hofmann-Engl Ludger. Rhythmic similarity: A theoretical and empirical approach // Proceedings of the seventh international conference on music perception and cognition. 2002. C. 564–567.