

Министерство образования Республики Беларусь

Учреждение образования  
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет компьютерных систем и сетей

Кафедра электронных вычислительных машин

Отчет по преддипломной практике

Студент

П.С. Арабей

Руководитель

А.Н. Ткаченко

Консультант от кафедры ЭВМ

Н.Н. Иванов

Нормоконтролёр

А.С. Сидорович

МИНСК 2017

Министерство образования Республики Беларусь

Учреждение образования  
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет компьютерных систем и сетей

Кафедра электронных вычислительных машин

К ЗАЩИТЕ ДОПУСТИТЬ:  
Зав. каф. ЭВМ  
\_\_\_\_\_ Д.И. Самаль

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА  
к дипломному проекту  
на тему  
МОДУЛЬ ВЫДЕЛЕНИЯ ИНФОРМАЦИОННЫХ ОБРАЗОВ ИЗ  
МУЗЫКАЛЬНОГО ПРОИЗВЕДЕНИЯ

БГУИР ДП 1-40 02 01 01 004 ПЗ

Студент	П.С. Арабей
Руководитель	Н.Н. Иванов
Консультанты:	
от кафедры ЭВМ	Н.Н. Иванов
по экономической части	И.В. Смирнов
Нормоконтролёр	А.С. Сидорович
Рецензент	

МИНСК 2017

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ . . . . .	6
1 ОБЗОР ЛИТЕРАТУРЫ . . . . .	8
1.1 Обзор методов выделения фрагментов из музыкальных произведений . . . . .	8
1.2 Обзор методов представления музыкального трека в спектрально-временном виде . . . . .	8
1.3 Обзор методов выделения информационных признаков . . . . .	9
1.4 Мел-кепстральные коэффициенты(MFCC) . . . . .	14
1.5 Обзор аналогов . . . . .	16
1.6 Выбор информационных образов для жанровой классификации музыкальных произведений . . . . .	20
2 СИСТЕМНОЕ ПРОЕКТИРОВАНИЕ . . . . .	22
3 ФУНКЦИОНАЛЬНОЕ ПРОЕКТИРОВАНИЕ . . . . .	25
3.1 Алгоритм работы модуля выделения информационных образов из музыкального произведения . . . . .	25
3.2 Реализация внутренних модулей . . . . .	26
4 РАЗРАБОТКА ПРОГРАММНЫХ МОДУЛЕЙ . . . . .	33
5 РЕЗУЛЬТАТЫ ЖАНРОВОЙ КЛАССИФИКАЦИИ И ВИЗУАЛИЗАЦИИ . . . . .	34
6 ПРОГРАММА И МЕТОДИКА ИСПЫТАНИЙ . . . . .	42
7 РУКОВОДСТВО ПОЛЬЗОВАТЕЛЯ . . . . .	49
7.1 Требования к аппаратному и программному обеспечению . . . . .	49
7.2 Руководство по использованию программного средства . . . . .	49
8 ЭКОНОМИЧЕСКОЕ ОБОСНОВАНИЕ РАЗРАБОТКИ МОДУЛЯ ВЫДЕЛЕНИЯ ИНФОРМАЦИОННЫХ ОБРАЗОВ ИЗ МУЗЫКАЛЬНОГО ПРОИЗВЕДЕНИЯ . . . . .	50
8.1 Описание функций, назначения и потенциальных пользователей ПО . . . . .	50
8.2 Расчёт затрат на разработку ПО . . . . .	50
8.3 Оценка результата (эффекта) от продажи ПО . . . . .	55
8.4 Расчёт показателей эффективности инвестиций в разработку ПО . . . . .	57
ЗАКЛЮЧЕНИЕ . . . . .	60
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ . . . . .	61
ПРИЛОЖЕНИЕ А . . . . .	63
ПРИЛОЖЕНИЕ Б . . . . .	64

## ВВЕДЕНИЕ

На сегодняшний день количество музыкальных треков в Интернете не поддаётся подсчёту. В одном только сервисе Яндекс.Музыка доступно 17 миллионов треков [1]. Сегодня выпускается тысячи песен каждый день [2]. Существует сотни жанров и поджанров музыки. Поэтому задача поиска новой музыки становится нетривиальной.

Для решение данной задачи используются рекомендательные сервисы, которые используют четыре подхода к анализу музыки для составления рекомендации:

1. Популярность трека. Использование данных о популярности трека и информации о пользователях. Этот способ доступен только большим онлайн сервисам. Качество рекомендации зависит от количества пользователей данного сервиса. Популярность треков определяется количеством и качеством отзывов и количеством покупок.

2. Метаинформация о треке. Это способ также доступен большим онлайн сервисам с большой библиотекой музыки. Качество рекомендации зависит от размеров библиотеки. Метаинформация о треке представляет собой тег с жанром, информация об исполнителе, альбом в который включён этот трек и т.д.

3. Семантика текста. Анализируя музыкальные блоги с применением технологий обработки естественных языков подход позволяет постоянно изучать веб и аналитически просматривать десятки миллионов страниц, имеющих отношения к музыке. Текст песни анализируется на основе взвешенных лексем. Произведения рекомендуется при близости его дескрипторов с дескрипторами пользователя.

4. Акустико-синтаксический анализ. На текущий момент нет методов достоверного распознавания музыкальных инструментов и музыкальных звуков. Однако, несмотря на это анализ сигнала играет очень важную роль используется в работе рекомендательных алгоритмов. Анализируются фрагменты произведения длительности от 200 мс до 4 с, в зависимости от вариативности мелодии. Для каждого сегментов произведения определяется громкость, тембр, выявляются наиболее громко звучащие музыкальные инструменты; устанавливается к какой части композиции (припев, куплет и т. д.) относится этот сегмент.

В первом способе информационные признаки о треке получаются на основе анализа поведения пользователя. Рекомендации при использовании данного метода делаются на основе коллаборативной фильтрации на основе соседей. Во втором методе является метаинформация о треке, а рекомендация делается на основе коллаборативной фильтрации на основе модели. В современных мультимедийных сервисах(youtube, lastfm, Яндекс.Музыка) используется гибридный подход, который объединяет в себе подход осно-

ванный на соседстве и основанный на модели [3] [4].

Целью данного дипломного проекта является создание модуля, который выделял информационные образы из музыкального трека только на основании акустического анализа. Информационный образ - это признаковое описание трека как музыкального сигнала. Такой сервис должен быть лишён субъективности и обеспечить качество рекомендации.

В соответствии с поставленной целью были определены следующие задачи:

- выделение спектральных, временных и иных признаков из музыкального трека;
- проверка значимости признаков путём использования их в задаче жанровой классификации (задача настройки классификатора решаться не будет);
- визуализация данных алгоритмом t-SNE.

## 1 ОБЗОР ЛИТЕРАТУРЫ

В части рассмотренных исследований решались смежные задачи, такие как жанровая классификация, идентификация музыкального трека, нахождение заимствований в музыкальном треке, распознавание звуковых источников. Для решения данных задач использовались разные подходы к выделения информационных образов из музыкальных треков. Среди рассмотренных приложений также стоит отметить разнообразие подходов к выделению информационных признаков. В данной дипломной работе стоит задача найти такие характерные информационные образы, которые позволили на основе их делать рекомендацию

### 1.1 Обзор методов выделения фрагментов из музыкальных произведений

В работе[5] информационные признаки, которые будут рассмотрены в следующем подразделе, рассчитываются по фрагментам музыкального трека в 1 секунду, состоящему из 40 непересекающихся окон в 25 миллисекунд. Для классификации используется 30 фрагментов. Полная сборка аудиоданных используемая в данной работе состоит из 15 жанров \* 50 файлов \* 30 секунд = 22500 секунд (то есть 6,25 часов аудио). Каждый фрагмент анализируется независимо. По каждому окну получают набор числовых признаков. Их математическое ожидание и дисперсию используют в качестве признаков для фрагмента. В работе[6] используется такой же подход. Трёхсекундный фрагмент состоит из 1500 окон по 0,02 секунды. В отличие от работ[5] по автоматической жанровой классификации музыкального трека, в качестве признаков фрагмента используют среднее значение, дисперсию, коэффициент асимметрии, коэффициент эксцесса информационных признаков окон. Эти четыре значения затем используются для измерения подобия. В работе[7], где описывается принцип работы большинства аудиоидентификационных систем, используются фрагменты музыкального трека размером от 10 до 500 миллисекунд с перекрытием фрагментов от 50 до 90 процентов. В этой же работе описываются алгоритмы распознавания ремиксов.

Каждый музыкальный трек проходит передискретизацию до частоты 44100 Гц. Затем берется четырёхсекундный фрагмент.

Недостатками данных методов является потеря информации о структурах в музыкальном треке, которые не помещаются в один фрагмент.

### 1.2 Обзор методов представления музыкального трека в спектрально-временном виде

В работе[8] решается задача идентификации(опознавания) музыкального трека, а также поиск заимствований в жанре музыки хип-хоп. Для

получения спектрограммы используется оконное преобразование Фурье с окном Хемминга и размером окна 64 миллисекунды. После получения спектрограммы используется только амплитудная составляющая с последующим логарифмированием и применением высокочастотного фильтра к спектральной кривой. Эти шаги предпринимаются для того, чтобы сделать спектрограмму менее зависимой от абсолютного уровня и грубой спектральной формы. Также в работе рассмотрено constant Q преобразование[9]. Выделение информационных признаков в работе состоит из трёх стадий. Конечным результатом является коррелограмма. Первая стадия моделирует влияние звуковой волны на базилярную мембрану ушной улитки с помощью полосового фильтра, который реализуется четырьмя каскадными секциями фильтра второго порядка, которые реализуют фильтр восьмого порядка с импульсной характеристикой «gammatone» с центральной частотой в 9,3 кГц. Вторая стадия моделирует движение волосковых клеток. Так как клетки реагируют только на отклонение ресничек в одном направлении, и это вводит этап выпрямления в цепочке обработки сигналов. На низких частотах волосковые клетки имеют тенденцию срабатывать в определенной фазе сигнала - процесс, называемый блокировкой фазы. По мере того как частота входного сигнала увеличивается, блокировка фазы начинает заканчиваться примерно на 1,5 кГц и исчезает на 5 кГц, но при отсутствии блокировки тонкой структуры формы сигнала волосковые клетки фиксируются в амплитудной огибающей сигнала. Этот эффект моделируется посредством операции свертки с импульсным откликом фильтра типа «приподнятого косинуса» [10] с  $T = 0,25$  мс. Третья стадия это построение коррелограммы. Коррелограмма моделирует человеческое восприятия основного тона. Для этого используется автокорреляция. Измерением средней энергии в зависимости от задержки можно определить базовый период сигнала. После FFT-преобразования считается оконная автокорреляция для частотной области. Полученный трёхмерный массив данных(время, частота, высота тона) используется для получения информационных признаков для классификации музыкальных инструментов. В работе симулируется человеческое восприятие. Механизмы распознавания в человеческом мозге гораздо лучше приспособлены к сложным акустическим средам, чем любой искусственный механизм (рисунок 1.1).

В работе для извлечения информационных образов о ритме используется вейвлет Добеши четвёртого порядка.

### 1.3 Обзор методов выделения информационных признаков

В работе различается два вида признаков: спектральные и временные. Временные признаки - это значения, которые вычисляются из значений самих аудиоданных. Спектральные - на основе спектра полученного преобразованием Фурье.

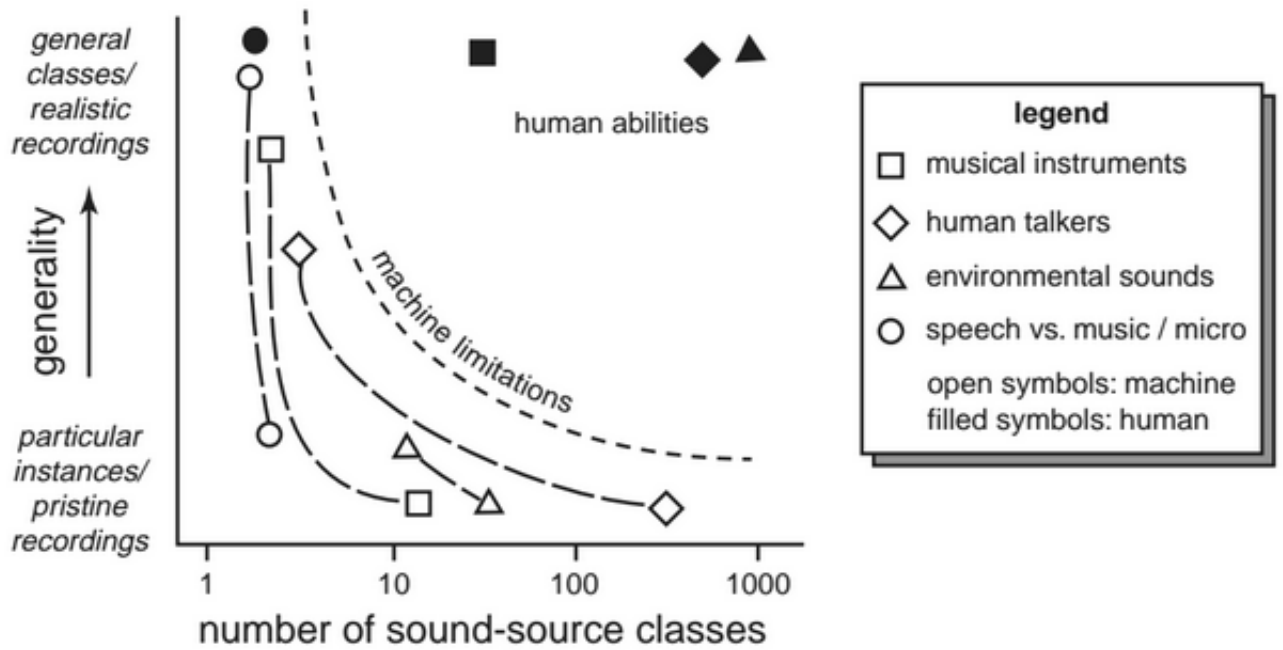


Рисунок 1.1 – График показывающий ограничения искусственных механизмов распознавания

Временные признаки:

1. Количество переходов сигнала через ноль.

$$ZeroCrossignRate = \frac{1}{N_S} \sum_{i=0}^{N_S-2} sgn(s_i) \oplus sgn(s_{i+1}) \quad (1.1)$$

где  $N_S$  – количество сэмплов в окне,  $sgn()$  – функция, которая возвращает 1 при положительном значении аргумента, и 0 при отрицательном.  $s$  – вектор, содержащий данные сигнала для одного окна

2. Автокорреляция первого порядка.

$$FirstOrderAutocorrelation = \frac{1}{N_S} \sum_{i=0}^{N_S-2} s_i * s_{i+1} \quad (1.2)$$

где  $N_S$  – количество сэмплов в окне,  $s$  – вектор, содержащий данные сигнала для одного окна.

3. Энергия сигнала.

$$Energy = \frac{1}{N_S} \sum_{i=0}^{N_S-1} (s_i)^2 \quad (1.3)$$

где  $N_S$  – количество сэмплов в окне,  $s$  – вектор, содержащий данные сигнала для одного окна.

Спектральные признаки:

1. Линейная регрессия спектра



$$\beta = \frac{N_B \sum_{i=1}^{N_B} (a_i F_i) - \sum_{i=1}^{N_B} a_i \sum_{i=1}^{N_B} F_i}{N_B \sum_{i=1}^{N_B} (F_i)^2 - \sum_{i=1}^{N_B} (F_i)^2} \quad (1.4)$$

где  $a_i$  - амплитуда,  $N_B$  - количество компонент разложения, а  $F$  - частота соответствующей амплитуды. Это статистическое приближение использовано для нахождения наклона ( $\beta$ ) спектра, который является мерой отношения высокочастотной составляющей к низкочастотной составляющей звука и, следовательно, тембра.

2. Среднее арифметическое взвешенное спектра.

$$SpectralCentroid = \frac{\sum_{i=1}^{N_B} a_i F_i}{\sum_{i=1}^{N_B} a_i} \quad (1.5)$$

Психоакустически это показатель воспринимаемой «яркости» звука, обеспечивающий лучшую оценку «яркого» звука, нежели тональность.

3. Гладкость спектра, как мера спектральной огибающей.

$$SSmoothness = \sum_{i=1}^{N_B} 20 \log a_i - \frac{20 \log a_{i-1} + 20 \log a_i + 20 \log a_{i+1}}{3} \quad (1.6)$$

где  $a_i$  - амплитуда, а  $N_B$  - количество компонент разложения. Белый шум, спектр которого имеет энергию на всех частотах будет иметь гладкость 1. Синусоидальный сигнал имеет только один пик в спектре и гладкость спектра будет равняться 0.

4. Дисперсия спектра относительно его среднего взвешенного. Оркестровые треки будут иметь большее значение, в отличие от сольных или монофонических треков. Вычисляется по следующей формуле:

$$SSpread = \sqrt{\frac{\sum_{i=1}^{N_B} a_i (F_i - SpectralCentroid)^2}{\sum_{i=1}^{N_B} a_i}} \quad (1.7)$$

где  $a_i$  - амплитуда,  $SpectralCentroid$  - центроида спектра,  $N_B$  - количество компонент разложения, а  $F$  - частота соответствующей амплитуды.

5. Коэффициент асимметрии сигнала, является мерой того, насколько искажён спектр относительно его среднего взвешенного.

$$SDissymmetry = \sqrt{\frac{\sum_{i=1}^{N_B} a_i (F_i - SpectralCentroid)^3}{\sum_{i=1}^{N_B} a_i}} \quad (1.8)$$

В работе[5] исследованы алгоритмы автоматической жанровой классификации. Предложены набор признаков для представления «музыкальных аспектов» и ритмических структур аудиосигнала.

Термин «музыкальные аспекты» используются для обозначения особенности музыки, связанные с текстурой, тембром и музыкальными инструментами. В работе используется 4 статистических информационных образов, представляющие собой мат. ожидания и СКО следующих числовых признаков:

1. Среднее арифметическое взвешенное спектра, как мера спектральной яркости (см. формулу 1.5).
2. Энергетические спектральные окна по уровню 0,85.

$$\sum_{i=1}^{Rolloff} a_i = \sum_{i=1}^{N_B} a_i \quad (1.9)$$

где  $a_i$  - амплитуда, а  $N_B$  - количество компонент разложения.

$$Flux = \|a_i - a_p\| \quad (1.10)$$

где  $a_i$  - амплитуда текущего окна, а  $a_p$  - амплитуда прошлого окна.

Дополнительно выделяется процент окон, чья спектральная энергия меньше чем средняя спектральная энергия по фрагменту.

Также выделяются ритмические признаки. Вычисление признаков для представления ритмической структуры музыки основано на вейвлет преобразовании с вейвлетом Добеши. К результату вейвлет преобразования применяется автокорреляционная функция. Фиксируются первые пять пиков автокорреляционной функции и их соответствующие периодичности в ударах в минуту рассчитываются и добавляются в «ритмическую» гистограмму (см. рисунок 1.2). Этот процесс повторяется итерацией по сигналу и накоплением периодичности в гистограмме.

Пики гистограммы соответствуют различным периодичности звукового сигнала и используются в качестве основы для расчета признаков ритма. Используются следующие признаки, основанные на «ритмической» гистограмме :

1. Период - 0: Периодичность в ударах в минуту первого пика.
2. Амплитуда - 0: Относительная амплитуда (деленная на сумму ам-

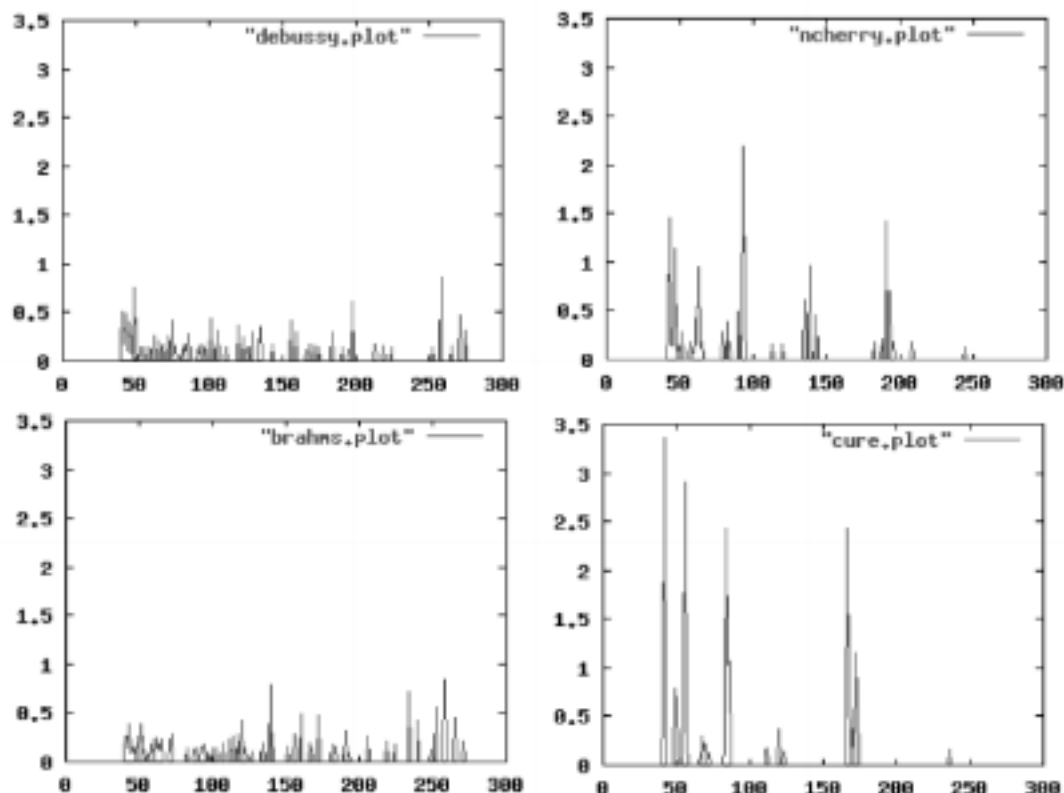


Рисунок 1.2 – Гистограмма ритма. Слева классическая музыка, справа поп-музыка

плитуд) первого пика.

3. Отношение периодичности - 1: отношение периодичности второго пика к периодичности первого пика.

4. Амплитуда - 1: Относительная амплитуда второго пика.

5. Отношение периодичности - 2: отношения периодичности третьего пика к периодичности второго пика.

6. Амплитуда - 2: Относительная амплитуда третьего пика.

7. Отношение периодичности - 3: отношения периодичности четвертого пика к периодичности третьего пика.

8. Амплитуда - 3: Относительная амплитуда третьего пика.

8-мерный вектор признаков, используемый для представления ритмической структуры и силы, комбинируется с 9-мерным вектором музыкальной поверхности( 4 статистических информационных образа \* 2 параметра), чтобы сформировать 17-мерный вектор признаков, который используется для автоматической классификации музыкального жанра. В работе [6] вводят меру спектральной плоскостности и коэффициент амплитуды.

$$SFM_k = \frac{[\prod S_k^2(f)]^{\frac{1}{N}}}{\frac{1}{N} \sum_f S_k^2(f)} \quad (1.11)$$

$$SCF_K = \frac{\max_k S_K^2(k)}{\frac{1}{N} \sum_k S_K^2(k)} \quad (1.12)$$

где  $S_K^2$  – спектральная плотность мощности. Чувственный эквивалент этих признаков можно охарактеризовать как шумоподобность и тоноподобие. Признаки с чувственным значением представляют характеристики звука, которые с большей вероятностью сохраняются в течении произведения и, следовательно, должны быть более надежными

#### 1.4 Мел-кепстральные коэффициенты(MFCC)

Мел - психофизическая единица высоты звука, применяется главным образом в музыкальной акустике. Количественная оценка звука по высоте основана на статистической обработке большого числа данных о субъективном восприятии высоты звуковых тонов. Звуковые колебания частотой 1000 Гц при эффективном звуковом давлении  $2 \cdot 10^{-3}$  Па (то есть при уровне громкости 40 фон), воздействующие спереди на наблюдателя с нормальным слухом, вызывают у него восприятие высоты звука, оцениваемое по определению в 1000 мел. Звук частоты 20 Гц при уровне громкости 40 фон обладает по определению нулевой высотой (0 мел). Зависимость нелинейна, особенно при низких частотах (для «низких» звуков). Преобразовать значение частоты звука (Гц) в значение высоты (мел) можно по формуле:

$$m = 1127,01048 \ln\left(1 - \frac{f}{700}\right) \quad (1.13)$$

При обработке звука мел-частотный кепстр (MFC) представляет собой кратковременный спектр мощности звука, основанный на косинус - преобразовании Фурье на логарифмической спектральной мощности на нелинейной мел-шкале частот.

Мел-кепстральные коэффициенты (MFCC) - это коэффициенты, которые в совокупности образуют MFC. Различие между кепстром и мел-кепстром в том, что в MFC полосы частот равномерно распределены по шкале мела, что приближается к восприятию слуховой системы человека более близко, чем линейные интервалы частот, используемые в нормальном кепстре.

MFCC обычно используются в качестве образов в системах распознавания речи, таких как системы, которые могут автоматически распознавать номера телефонов, произносимые в телефоне. MFCC также все чаще находят применение в приложениях поиска музыкальной информации, таких как классификация жанров[11], меры сходства звука[12] и т. д.

Для получения коэффициентов используют следующий алгоритм:

1. Используя преобразование Фурье получить спектр исходного сигнала.

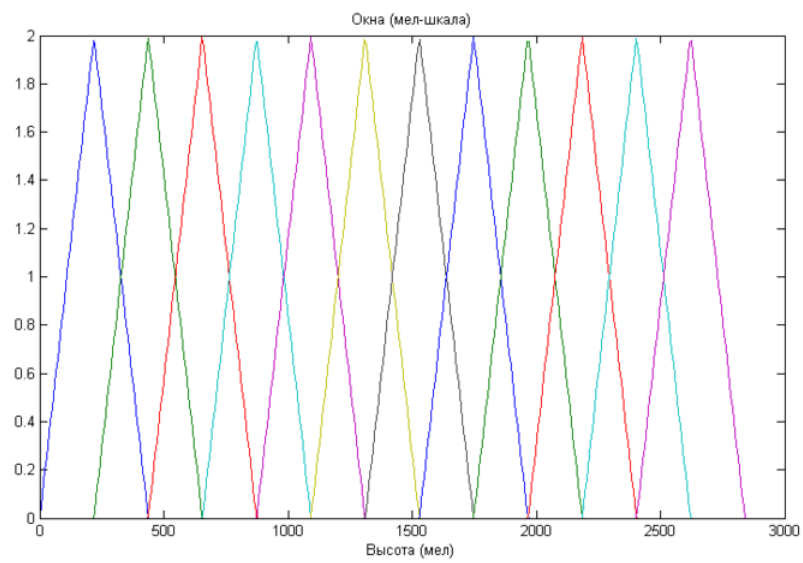


Рисунок 1.3 – Треугольные окна в мел-шкале

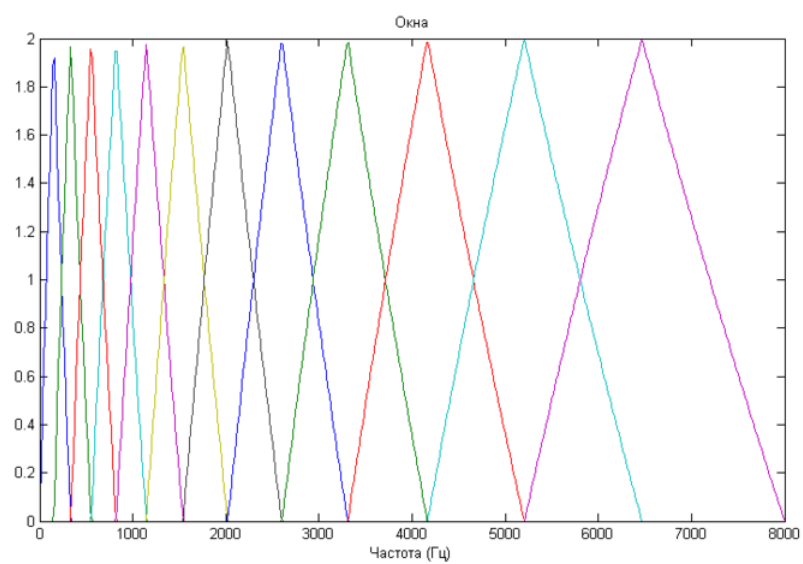


Рисунок 1.4 – треугольные окна в частотной шкале

2. Располагаем треугольные окна равномерно на мел-шкале (см. рисунок 1.3).

3. Переводим треугольные окна в частотную шкалу по формуле:

$$f = 700(e^{\frac{m}{1127}} - 1) \quad (1.14)$$

4. Делаем свёртку спектра с каждым окном (см. рисунок 1.4) и берём логарифм спектра мощности.

5. Полученные значения переводим во временную область дискретным косинусным преобразованием.

6. MFCC представляют собой амплитуды результирующего спектра.

## 1.5 Обзор аналогов

### 1.5.1 HOLO

Данный проект - это приложение, которое по анализу музыкальной коллекции позволяет составлять плейлисты похожих на заданные образцы композиции, а также визуализировать полученные данные о музыкальной коллекции[13]. Метод получения информационных признаков основан на преобразовании Фурье с последующим получением евклидова расстояния до набора готовых спектров, с последующим построением матрицы переходов. Спектры несут служебную функцию опорных значений, и их форма выбрана по принципу отличаться друг от друга как можно сильнее, как по громкости, так и по корреляции графика частот. Формирование базы данных реализовано следующим образом:

1. Из файла извлекается фрагмент параметризуемой длины, начиная с 20% общей длины.

2. Фрагмент нарезается на параметризуемое количество окон с параметризуемым процентом перекрытия.

3. Каждое окно подвергается преобразованию Фурье, сглаживанию и очистке.

4. Имея некоторый набор заранее подготовленных центроидов, оценивается расстояние каждого окна до каждого из этих центроидов.

5. На основе расстояния окно маркируется порядковым номером ближайшей центроиды.

6. Строится матрица переходов между номерами центроид(рисунок 1.5).

7. Матрица переходов используется как вектор информационных признаков на основе которого происходит рекомендация.

Также в приложении реализовано визуализация результатов сканирования локальной музыкальной библиотеки.

Преимущества данного приложение:

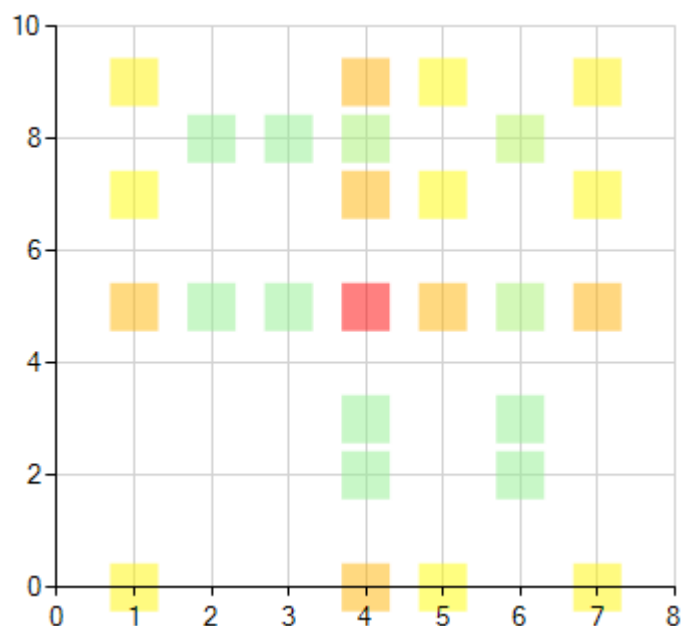


Рисунок 1.5 – Матрица переходов

- быстроедействие;
- использует акустический анализ.

Минусы:

- нет взаимодействия с другими сервисами воспроизведения музыки;
- используется только 20% музыкального трека;
- анализируются только MP3-файлы 44,1кГц 16 бит;

### 1.5.2 Athena-NeuroPlay

Данный проект - это приложение, которое использует генетический алгоритм, который на основе оценки пользователя качества исходной выборки строит топологию нейронной сети прямого распространения с несколькими скрытыми слоями [14].

Первый этап - нормализация данных. Для инструментов выделяются необходимые частоты. Также принимается в расчёт и чувствительность слуха в зависимости от частоты. Таким образом задача нормализации сводится к выделению некоторой информации о частотах, которая показывает:

- как часто в композиции звучит звук из данного диапазона частот;
- как громко он звучал;
- как долго он звучал;
- для каждого определенного диапазона частот (нужно разбить весь «слышимый» спектр на определенное число диапазонов(см. рисунок 1.6)).

Для выделения частотной насыщенности в треке на каждом временном интервале используется FFT. Временной интервал имеет размер 1024 сэмплов. На основе спектра получают следующие признаки: насыщенность звука определенными спектрами, частота возникновения различных спек-

тров звука, его громкость, его длительность. Затем данные нормализуются.

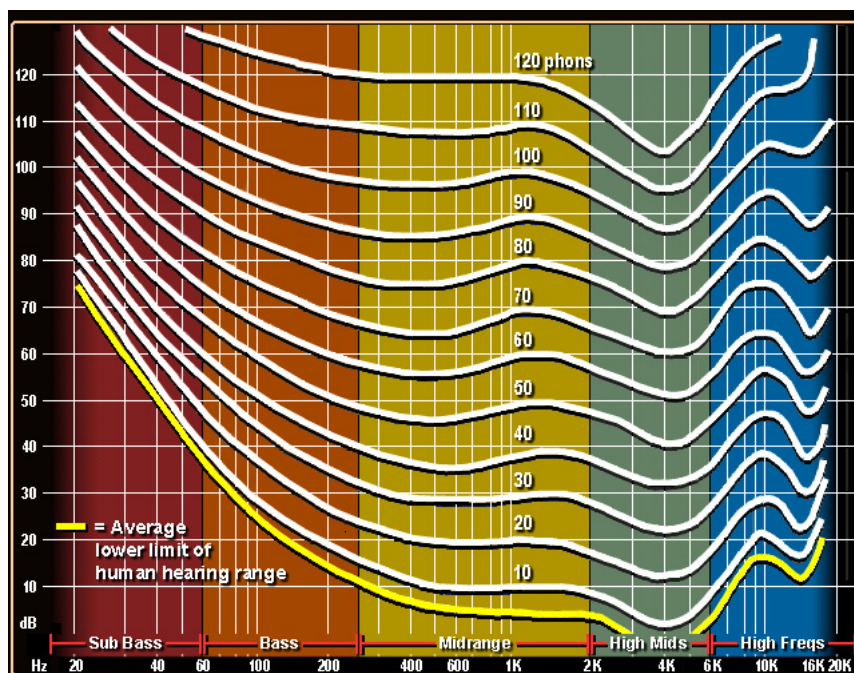


Рисунок 1.6 – Диапазоны частот

На втором этапе используется нейронная сеть прямого распространения с 1024 входными нейронами и  $24 \times 7$  выходными, где каждый выходной нейрон показывает «качество» трека для того или иного времени суток. В качестве исходной выборки используется плейлист, который делится на две части: обучающая и контрольная. Размер обучающей и контрольной выборке устанавливается пользователем. При прослушивании трека из обучающей выборке пользователь ставит оценку от 0 до 1, где 0 - «плохой трек», а 1 - «хороший». Оценка влияет на все оценки по времени суток, но больше на ту, в которое время было поставлена оценка. После того как обучающая выборка промаркирована, идёт процесс обучения и настройки топологии нейронной сети. Топология нейронной сети определяется генетическим алгоритмом. Для этого берётся стартовая конфигурация в 10 слоев, в каждом по 100 нейронов, которая обучается за 1000 эпох, после определяются качество ее обучения. Качество обучения - это суммарная оценка контролирующей и обобщающей способности нейронной сети. На следующем этапе создаётся 10 конфигураций, каждая из которых является «мутантом» исходной, то есть у нее изменена в случайную сторону либо количество слоев, либо количество нейронов на каждом или некоторых слоях. Далее идёт обучение каждой конфигурации тем же способом, что и исходную. Выбирается лучшая из них по обобщающей способности и определяется как исходная. Данный процесс продолжаем до тех пор, пока не наступает такой момент, что мы не можем найти конфигурацию, которая обучается лучше чем исходная. Данную конфигурацию считаем лучшей, она оказалась способной



лучше всех «запомнить» исходные данные и лучше всех предсказывает, то есть результат ее обучения наиболее качественный из возможных.

Из плюсов данного приложения стоит отметить:

- использует акустический анализ;
- запускается локально на компьютере пользователя.

Из недостатков:

- время работы;
- качество рекомендации.

### 1.5.3 Pandora Radio

Pandora (Пандора) - служба потокового воспроизведения музыки в Интернете, основанное на системе «Music Genome Project» [15]. Пользователь медиапроигрывателя Pandora выбирает музыкального исполнителя, после чего система ищет похожие композиции, используя около 400 музыкальных характеристик такие как жанр, тип инструментов, тип вокала, темп, синкопа, тональность, гармония и т. д. Используя функции «нравится» или «не нравится», слушатель может настроить «радиостанцию» по своему вкусу. В базе данных системы более миллиона композиций и более ста тысяч исполнителей [16]. Зарегистрированный пользователь может создать в своём профиле до 100 различных «радиостанций», транслирующих музыку в тех или иных жанрах. Медиапроигрыватель Pandora доступен пользователям с персональными компьютерами, смартфонами, планшетами с различными операционными системами.

Проект «Music Genome Project» - это набор более 450 атрибутов для описания песен и сложный математический алгоритм для их организации. Проект в настоящее время состоит из 5 суб-геномов : Pop / Rock, Hip-Hop / Electronica, Jazz, World Music и Classical. Песня представляется вектором, содержащим значения приблизительно для 450 «генов». Каждый ген соответствует характеристике музыки, например, пол ведущего вокалиста, уровень искажения на электрогитаре, тип фонового вокала и так далее. Рок и поп-песни имеют 150 генов, рэп-песни имеют 350 генов, а джазовые песни - приблизительно 400. Другие жанры музыки, такие как мировая и классическая музыка, имеют 300-450 генов. Система зависит от достаточного количества генов для получения полезных результатов. Учитывая вектор одной или нескольких песен, список других подобных песен построен с использованием того, что компания называет своим «алгоритмом сопоставления» [17]. Атрибуты для каждой песни выставляются музыкантом в процессе, который занимает от 20 до 30 минут на песню [18]. Десять процентов песен анализируются более чем одним музыкантом, чтобы обеспечить соответствие внутренним стандартам и статистическую надежность.

Преимущества данного сервиса:

- большая база данных музыки;

- имеются приложения для десктопов и мобильных платформ;
- быстродействие;
- взаимодействия с другими сервисами.

Из недостатков стоит отметить:

- сервис недоступен за пределами США, Австралии и Новой Зеландии;
- требуется регистрация;
- акустический анализ представлен не в чистом виде.

## 1.6 Выбор информационных образов для жанровой классификации музыкальных произведений

На основе анализа литературы были выбраны четыре типа информационных образов, которые будут извлекаться из музыкального трека:

- временной образ;
- спектральный образ;
- ритмический образ;
- мел-кепстральный образ.

Временные образ (здесь и далее представления музыкального трека как набора отсчётов будет называться сигналом).

1. Энергия сигнала, как мера яркости и громкости мелодии (см. формулу 1.3).

2. Количество переходов сигнала, через ноль, как мера зашумлённости (см. формулу 1.1).

3. Автокорреляция сигнала, как мера изменения резкости тембра (см. формулу 1.2).

Спектральный образ.

1. Среднее арифметическое взвешенное спектра. С точки зрения восприятия, оно имеет робастную связь с впечатлением «яркости» звука (см. формулу 1.5).

2. Линейная регрессия спектра, как мера отношения высокочастотной составляющей к низкочастотной составляющей звука и, следовательно, тембра (см. формулу 1.4).

3. Гладкость спектра, как мера гармоничности сигнала (см. формулу 1.6).

4. Дисперсия спектра относительно среднего взвешенного. С точки зрения восприятия, определяет «ширину» тембра (см. формулу 1.7).

5. Коэффициент асимметрии, как мера того, насколько искажён спектр относительно среднего взвешенного спектра, и, следовательно наклон к высоким или низким частотам (см. формулу 1.8).

6. Энтропия Винера или мера спектральной плоскостности. Определяет зашумлённость сигнала. Чем меньше значение, тем больше спектральной мощности сосредоточено в относительно небольшом числе полос (см.

формулу 1.4).

7. Энергетическое спектральное окно по уровню 0,85, как мера спектральной формы (см. формулу 1.9)

Ритмический образ. Признаки считаются по коррелограмме.

1. Амплитуда первого пика.
2. Отношение частот первого пика к частоте второго пика.
3. Амплитуда второго пика.
4. Отношение частот второго пика к частоте третьего пика.
5. Амплитуда третьего пика.
6. Отношение частот третьего пика к частоте четвёртого пика.
7. Амплитуда четвёртого пика.
8. Частота первого пика.

Мел-кепстральный образ представляет из себя 16 мел-кепстральных коэффициентов.

## 2 СИСТЕМНОЕ ПРОЕКТИРОВАНИЕ

Изучив теоретические аспекты разрабатываемого модуля и выработав список требований необходимых для разработки модуля, разбиваем систему на компоненты. Компоненты в виде блоков и их взаимосвязи указаны на чертеже. В разрабатываемом модуле можно выделить следующие блоки:

- модуль чтения музыкального произведения;
- модуль препроцессинга и нарезки музыкального трека на фрагменты;
- модуль получения частотно-временного представления сигнала;
- модуль извлечения информационных образов;
- модуль обработки информационных образов;
- база данных информационных образов;
- модуль жанровой классификации музыкального произведения.
- модуль визуализация.

Структурная схема, иллюстрирующая перечисленные блоки и связи между ними приведена на чертеже ГУИР.400201.004 С1. Для решения задачи выделения спектральных, временных и иных признаков из музыкального трека будут использоваться следующие блоки:

- модуль препроцессинга и нарезки музыкального трека на фрагменты;
- модуль получения частотно-временного представления сигнала;
- модуль извлечения информационных образов;
- модуль обработки информационных образов.

Для решения задачи проверки значимости признаков путём использования их в задаче жанровой классификации будет использоваться модуль жанровой классификации музыкального произведения.

Для решения задачи визуализации данных алгоритмом t-SNE используется модуль визуализации.

Модуль чтения музыкального произведения состоит из двух частей. Первая часть это консольная утилита Lame. Lame свободное приложение для кодирования аудио в формат MP3 (MPEG-1 audio layer 3) и декодирования аудио в WAV формат, который наиболее удобен для чтения и представления музыкальных треков в виде массива. Библиотека Scipy позволяет преобразовать WAV аудио в NumPy массив.

Модуль препроцессинга обеспечивает первоначальную обработку музыкальных треков. На вход модуля музыкальный трек подаётся как набор сэмплов (NumPy массив). В данном блоке происходит нормализация, фильтрация и, если необходимо, приведения стереозвука к монозвuku .

Модуль получения частотно-временного представления сигнала представляет собой набор преобразований такие как оконное преобразование Фурье и вейвлеты, которые представляются в виде многомерного NumPy - массива. Так подобные вычисления требуют высокой производительности,

то для более эффективного вычисления используются параллельные вычисления на центральном процессоре.

Модуль извлечения информационных образов. Данный модуль представляет из себя набор методов получения информационных образов из временной и частотной области музыкального произведения. Также в этом модуле считаются мел-кепстральные коэффициенты. На вход данному модулю подаётся многомерные Numpy массивы. На выходе получается одномерный Numpy массив информационных признаков для каждого музыкального трека.

Модуль обработки информационных образов. В данном модуле признаки нормализуются по МО и СКО. Удаляются выбросы и некорректные значения.

Модуль хранения информационных образов хранит в себе результат работы всего модуля и хранит в себе набор информационных образов музыкальных треков и принадлежность к тому или иному кластеру.

При выборе базы данных были сформулированы следующие требования:

- производительность;
- объектный язык запросов;
- возможность параллельной записи и чтения.

Так как главный модуль является частью сервиса рекомендации музыки, то появляются дополнительные требования к базе данных:

- масштабируемость;
- репликация;
- балансировка нагрузки.

С учётом всех этих требований и того факта, что данные для хранения представляют собой простую структуру, выбор пал на базу данных MongoDB. MongoDB — документоориентированная система управления базами данных (СУБД) с открытым исходным кодом, не требующая описания схемы таблиц. Классифицирована как NoSQL, использует JSON-подобные документы и схему базы данных. Написана на языке C++. Имеется подробная и качественная документация, большое число примеров и драйверов под популярные языки Java, JavaScript, Node.js, C++, C#, PHP, Python, Perl, Ruby. MongoDB может работать с набором реплик. Набор реплик состоит из двух и более копий данных. Каждый экземпляр набора реплик может в любой момент выступать в роли основной или вспомогательной реплики. Все операции записи и чтения по умолчанию осуществляются с основной репликой. Вспомогательные реплики поддерживают в актуальном состоянии копии данных. В случае, когда основная реплика дает сбой, набор реплик проводит выбор, который из реплик должен стать основным. Второстепенные реплики могут дополнительно являться источником для операций чтения. MongoDB масштабируется горизонтально используя шардинг. Пользователь

выбирает ключ шарда, который определяет как данные в коллекции будут распределены. Данные разделяются на диапазоны (в зависимости от ключа шарда) и распределяются по шардам. Из преимуществ MongoDB:

1. Объектный язык запросов.
2. Поддержка индексации.
3. Поддержка Map/Reduce для распределенных операций над данным.
4. Документы, не требующие определения схемы. Одно из самых важных преимуществ. Преимущество заключается в том, что нет нужды хранить пустые ячейки данных в каждом документе.
5. Поддержка сложных массивов. Каждый элемент массива может представлять из себя объект.
6. Поддержка шардинга на уровне платформы.
7. Атомарность гарантируется только на уровне целого документа, то есть частичного обновления документа произойти не может.
8. Любые данные, которые считываются одним клиентом, могут параллельно изменяться другим клиентом.

СУБД управляет наборами JSON-подобных документов, хранимых в двоичном виде в формате BSON. Хранение и поиск файлов в MongoDB происходит благодаря вызовам протокола GridFS.

Модуль жанровой классификации необходим для проверки значимости выделенных образов. Для этого используется набор стандартных алгоритмов классификации из библиотеки Scikit-learn с параметрами по умолчанию. На вход модуля поступает двумерный Numpy-массив. На выходе матрица ошибок. В качестве исходной выборки используется репозиторий музыки GZTAN.

Модуль визуализации визуализирует данные с помощью алгоритма t-SNE. На вход подаётся двумерный Numpy-массив. На выходе изображение в формате JPEG.

Для реализации модулей был выбран язык программирования Python, так как для него существует множество библиотек выполняющих математические расчеты, и облегчающих решение задач связанных с анализом данных и машинным обучением. В работе используются библиотека NumPy - это расширение языка Python, добавляющее поддержку больших многомерных массивов и матриц, вместе с большой библиотекой высокоуровневых математических функций для операций с этими массивами. Также существует библиотека SciKit-learn, которая содержит реализацию алгоритмов машинного обучения инструменты для работы с данными. В дипломной работе также используется библиотека Skipy, которая предназначена для выполнения научных и инженерных расчётов.

### 3 ФУНКЦИОНАЛЬНОЕ ПРОЕКТИРОВАНИЕ

#### 3.1 Алгоритм работы модуля выделения информационных образов из музыкального произведения

На ход модуля чтения музыкального произведения подаётся путь к музыкальному произведению. Файл с расширением “.wav” считывается в оперативную память в формате массива отсчётов. Файл с иным расширением конвертируются в WAV формат.

В модуле препроцессинга и нарезки от музыкального трека отрезается 10% длины с начала и конца трека. В случае, если трек является многоканальным, то он приводится к одноканальному посредством чередованием правого и левого канала. Данные нормализуются по МО и СКО. Затем используется экспоненциальное сглаживание с коэффициентом сглаживания 0,99. Результат нарезается на фрагменты по 5 секунд с перекрытием в 0,5 секунды.

В модуле получения частотно-временного представления используется оконное преобразование Фурье с окном Хэмминга шириной в 10 миллисекунд. Также в модуле для получения ритмического образа используется вейвлет Добеши каскадным алгоритмом с количеством каскадов 4. Результат каждого каскада сглаживается экспоненциальным сглаживанием с коэффициентом сглаживания 0,97, передискретизируется с уменьшением дискретизации в 16 раз, нормализуется по МО. Затем результаты поэлементно складываются и вычисляется автокорреляция. Результатом работы модуля является спектрограмма и коррелограмма.

В модуле выделения информационных образов из временного представления сигнала выделяется временной, из спектрограммы - спектральный образ по каждому срезу спектрограммы, из коррелограммы - ритмический. Мел-кепстральные коэффициенты считаются по окнам в 5 миллисекунд.

В модуле обработки информационных образов по всем образам кроме временного считается МО, СКО, коэффициент асимметрии, коэффициент эксцесса.

В базу данных записывается информационный образ в формате: название произведения, номер фрагмента и результат работы модуля обработки информационных образов.

В модуле классификации информационных образов используются набор стандартных алгоритмов классификации из библиотеки scikit-learn с параметрами по умолчанию:

- а) метод ближайших соседей с  $k = 3$ ;
- б) метод опорных векторов с полиномиальным ядром;
- в) дерево принятия решений;
- г) случайный лес;

д) нейронная сеть прямого распространения с 1000 нейронов на скрытом слое;

е) `adaBoost`;

ж) наивный баесовский классификатор;

з) квадратичный дискриминант.

Для оценки качества классификации использовался скользящий контроль количеством разбиений равным 10 и алгоритмом разбиения `stratified k-fold`. Также для каждого классификатора строится матрица ошибок.

В модуле визуализации используется алгоритм `t-sne`. Это нелинейный метод уменьшения размерности, который особенно хорошо подходит для вложения высокоразмерных данных в пространство двух или трех измерений, которое затем можно визуализировать на диаграмме рассеяния. В частности, он моделирует каждый высокоразмерный объект с помощью двух- или трехмерной точки таким образом, что аналогичные объекты моделируются соседними точками, а разнородные объекты моделируются удаленными точками. Данные визуализируются в 2д и 3д графики.

### 3.2 Реализация внутренних модулей

Рассмотрим в деталях функциональные части системы: для этого произведем детальный анализ компонентов, модулей, составляющих их классов и отдельных методов, реализующих логику программы.

Структурно система подразделяется на ряд модулей, объединенных под общим модулем с названием `MainModule`. Дальнейшее разделение ведется по следующим модулям:

1. Модуль чтения музыкального произведения - `WavModule`.
2. Модуль препроцессинга и нарезки - `PreprocessingModule`.
3. Модуль получения частотно-временного представления сигнала - `SpectralTransformerModule`.
4. Модуль извлечения информационных образов - `FeatureExtractorModule`.
5. Модуль обработки информационных образов - `FeatureProcessingModule`.
6. База данных информационных образов - `DatabaseModule`.
7. Модуль жанровой классификации музыкального произведения - `GenreClassificationModule`.
8. Модуль визуализации - `VisualizeDataModule`.

#### 3.2.1 Класс `Track`

Класс `Track` является абстракцией для хранения трека в виде массива отсчетов, частоты дискретизации и метаданных о треке.

Класс `Track` имеет следующие поля:



- Поле `data` – массив `ndarray`, который хранит в трек в виде отсчётов.
- Поле `sample_rate` – частота дискретизации трека.
- Поле `label` – поле хранящее метайнформацию о треке.

### 3.2.2 Класс `SpectralTrack`

Класс `SpectralTrack` наследуется от класса `Track` и является абстракцией для хранения спектра трека и его коррелограммы

Класс `Track` имеет следующие поля:

- Поле `spectra_data` – массив `ndarray`, который хранит спектрограмму трека, полученную с помощью оконного преобразования Фурье.
- Поле `percussion_data` – массив `ndarray`, который хранит коррелограмму трека.

### 3.2.3 Класс `TrackModel`

Класс `TrackModel` является абстракцией для хранения временного, спектрального, ритмического и мел-кепстрального образа.

Класс `TrackModel` имеет следующие методы и поля:

- Поле `timing_features` – массив `ndarray`, который хранит временной образ трека.
- Поле `spectral_features` – массив `ndarray`, который хранит спектральный образ трека.
- Поле `percussion_features` – массив `ndarray`, который хранит ритмический образ трека.
- Поле `mfcc_features` – массив `ndarray`, который хранит мел-кепстраьный образ трека.
- Поле `label` – поле хранящее метайнформацию о треке .
- `to_vector()` – метод, который конкатенирует образы. И возвращает одномерный массив типа `ndarray`.

### 3.2.4 Класс `WavModule`

Класс `WavModule` предназначен для преобразования MP3 к формату WAV и считывании данных в Numpy-массив.

Класс `WavModule` имеет следующие методы:

- `create_wav(file_name)` процедура, которая принимает на вход путь к MP3-файлу и с помощью стандартной библиотеки Python вызывает либо `bash`-скрипт в случае запуска в операционной системе Linux, либо `batch`-скрипт в случае запуска в операционной системе Windows, который использует консольное приложение Lame для преобразования MP3-файла в WAV формат
- `read_wav(filename, label)` - метод, который считывает файл формата WAV и сохраняет данные в класс `Track`

### 3.2.5 Класс PreprocessingModule

Класс `PreprocessingModule` предназначен для первичной обработки трека, которая включает в себя приведения стерео звука к моно, нормализация по МО и СКО, удаления заданного процента трека с начала и с конца, экспоненциальное сглаживание и нарезка трека на фрагменты. В конструкторе задаётся коэффициент сглаживания, процент пересечения фрагментов, процент отсечения трека с начала, процент отсечения трека с конца, размер фрагмента в секундах.

Класс `PreprocessingModule` имеет следующие методы и поля:

- Поле `alpha` – коэффициент сглаживания, который может принимать значения от 0 до 1.
- Поле `overlap` – процент пересечения фрагментов.
- Поле `cut_start` – процент отсечения трека с начала трека
- Поле `cut_end` – процент отсечения трека с конца трека
- Поле `frame_size_sec` – размер фрагментов в секундах
- `stereo_to_mono(track)` – метод, который преобразует стерео звук к моно с помощью чередования правого и левого канала
- `scale(track)` – метод, который нормализует данные по МО и СКО
- `filter(track)` – метод, который преобразовывает данные используя экспоненциальное сглаживание:

$$s_t = \begin{cases} c_1, & t_1 = 0. \\ s_t + \alpha * (c_t - s_{t-1}), & t > 0 \end{cases} \quad (3.1)$$

где -  $\alpha$  – коэффициент сглаживания заданный в поле `alpha`  
• `cutting(track)` – метод, который отсекает процент данных трека с начала и конца, который задаётся полями `cut_start` и `cut_end`

- `framing(track)` – метод, который нарезает трека на фрагменты длиной заданной полем `frame_size_sec` и перекрытием заданным полем `overlap`.

### 3.2.6 Класс SpectralTransformer

Класс `SpectralTransformer` предназначен для получения спектрально-временного представления трека, а также для получения коррелограммы. Спектрально-временное представление получается путём оконного преобразования Фурье (см. формулу 3.2) с окном Хемминга размером 10 миллисекунд (см. формулу 3.3)

$$F(m, \omega) = \sum_{n=-\infty}^{\infty} f[n]w[n-m]e^{-j\omega n} \quad (3.2)$$

$$w_i = 0.54 - 0.46 \cos \frac{2\pi i}{n-1} \quad (3.3)$$

Также в модуле получает ритмические(перкуSSIONные) признаки.

- Поле `alpha` – коэффициент сглаживания, который может принимать значения от 0 до 1.
- Поле `window` – массив, который содержит в себе окно Хемминга.
- Поле `level` – количество каскадов в
- Поле `rate` – частота передискретизации.
- `short_time_fourier(track)` – метод, который преобразует трек оконным преобразованием Фурье с окном Хемминга, возвращает двумерный массив амплитуд типа `ndarray`
- `wavelet_daubechies(data)` – метод, который делает вейвлет преобразования Добеши.
- `filter(track)` – метод, который преобразовывает данные используя экспоненциальное сглаживание.
- `resampling(data)` – метод, который передискретизирует данные в `rate` раз
- `normalize_and_sum(track)` – метод, который

### 3.2.7 Класс `FeatureExtractor`

Класс `FeatureExtractor` предназначен для

- Поле `time_feature_models` – коэффициент сглаживания, который может принимать значения от 0 до 1.
- Поле `spectre_feature_models` – массив, который содержит в себе окно Хемминга.
- Поле `results` – количество каскадов в
- Поле `nceps` – частота передискретизации.
- Поле `frame_size` – частота передискретизации.
- `extract_feature(track)` – метод,
- `extract_percussion_feature(track)` – метод, который делает вейвлет преобразования Добеши.
- `extract_mfcc(track)` – метод, который преобразовывает данные используя экспоненциальное сглаживание.
- `extract_spectra_feature(track)` – метод, который передискретизирует данные в `rate` раз
- `normalize_and_sum(track)` – метод, который

### 3.2.8 Класс `FeatureExtractorModel`

Класс `FeatureExtractorModel` абстрактный класс, который используется в качестве базового для классов, которые извлекают признак из временной или из спектральной области.

Класс `FeatureExtractorModel` имеет следующие методы:

- `get(data, params)` – метод, который принимает на вход `ndarray` и массив параметров, а на выход выдаёт признак типа `float`. В данном классе метод абстрактный
- `normalize(result, data)` – метод, который нормализует данные по размер входного массива

### 3.2.9 Класс `SpectralFeature`

Класс `SpectralFeature` абстрактный класс, который используется в качестве базового для классов, который извлекают признак из спектральной области. Класс наследуется от `FeatureExtractorModel`.

### 3.2.10 Класс `TimingFeature`

Класс `TimingFeature` абстрактный класс, который используется в качестве базового для классов, который извлекают признак из временной области. Класс наследуется от `FeatureExtractorModel`.

### 3.2.11 Класс `Energy`

Класс `Energy` класс, который получает значение энергии сигнала из временной области (см. формулу 1.3). Класс наследуется от `TimingFeature`.

Класс `Energy` имеет следующие методы:

- `get(data, params)` – метод, который принимает на вход массив отсчётов, а на выход выдаёт энергию сигнала.

### 3.2.12 Класс `ZeroCrossingRate`

Класс `ZeroCrossingRate` класс, который получает значение количества переходов сигнала через ноль из временной области (см. формулу 1.1). Класс наследуется от `TimingFeature`.

Класс `ZeroCrossingRate` имеет следующие методы:

- `get(data, params)` – метод, который принимает на вход массив отсчётов, а на выход выдаёт количества переходов сигнала через ноль .

### 3.2.13 Класс `Autocorrelation`

Класс `Autocorrelation` класс, который получает значение автокорреляцию первого рода из временной области (см. формулу 1.2). Класс наследуется от `TimingFeature`.

Класс `Autocorrelation` имеет следующие методы:

- `get(data, params)` – метод, который принимает на вход массив отсчётов, а на выход выдаёт автокорреляцию первого рода.

### 3.2.14 Класс SpectralCentroid

Класс SpectralCentroid класс, который получает значение арифметического среднего взвешенного из спектра (см. формулу 1.5). Класс наследуется от SpectralFeature.

Класс SpectralCentroid имеет следующие методы:

- `get(data, params)` – метод, который принимает на вход массив отсчётов, а на выход значение арифметического среднего взвешенного.

### 3.2.15 Класс SpectralSmoothness

Класс SpectralSmoothness класс, который получает значение гладкости спектра (см. формулу 1.6). Класс наследуется от SpectralFeature.

Класс SpectralSmoothness имеет следующие методы:

- `get(data, params)` – метод, который принимает на вход массив отсчётов, а на выход значение гладкости спектра .

### 3.2.16 Класс LinearRegression

Класс LinearRegression класс, который получает значение линейной регрессии спектра (см. формулу 1.4). Класс наследуется от SpectralFeature.

Класс LinearRegression имеет следующие методы:

- `get(data, params)` – метод, который принимает на вход массив отсчётов, а на выход значение линейной регрессии спектра .

### 3.2.17 Класс SpectralSpread

Класс SpectralSpread класс, который получает значение дисперсии спектра относительно арифметического среднего взвешенного (см. формулу 1.7). Класс наследуется от SpectralFeature.

Класс SpectralSpread имеет следующие методы:

- `get(data, params)` – метод, который принимает на вход массив отсчётов и массив params, где первым элементом идёт вычисленное арифметическое среднее взвешенное , а на выход выдаёт значение дисперсии спектра относительно арифметического среднего взвешенного .

### 3.2.18 Класс SpectralDissymmetry

Класс SpectralDissymmetry класс, который получает коэффициент асимметрии спектра (см. формулу 1.7). Класс наследуется от SpectralFeature.

Класс SpectralDissymmetry имеет следующие методы:

- `get(data, params)` – метод, который принимает на вход массив отсчётов и массив `params`, где первым элементом идёт вычисленное арифметическое среднее взвешенное, а на выход выдаёт значение коэффициент асимметрии спектра.

### 3.2.19 Класс `FeatureProcessing`

Класс `FeatureProcessing` предназначен для

- Поле `with_mean` – коэффициент сглаживания, который может принимать значения от 0 до 1.
- Поле `with_std` – массив, который содержит в себе окно Хемминга.
- Поле `with_skew` – количество каскадов в
- Поле `with_kurtosis` – частота передискретизации.
- `mean(data)` – метод,
- `std(data)` – метод, который делает вейвлет преобразования Добеши.
- `skew(data)` – метод, который преобразовывает данные используя экспоненциальное сглаживание.
- `kurtosis(data)` – метод, который передискретизирует данные в `rate` раз
- `process_feature(track)` – метод, который

### 3.2.20 Класс `GenreClassificationModule`

Класс `GenreClassificationModule` предназначен для

- Поле `classifiers` – коэффициент сглаживания, который может принимать значения от 0 до 1.
- Поле `labels_name` – массив, который содержит в себе окно Хемминга.
- Поле `cv` – количество каскадов в
- `cross_validation_predict(clf, data, labels)` – метод,
- `std(data)` – метод, который делает вейвлет преобразования Добеши.
- `skew(data)` – метод, который преобразовывает данные используя экспоненциальное сглаживание.
- `kurtosis(data)` – метод, который передискретизирует данные в `rate` раз
- `process_feature(track)` – метод, который

## 4 РАЗРАБОТКА ПРОГРАММНЫХ МОДУЛЕЙ

## 5 РЕЗУЛЬТАТЫ ЖАНРОВОЙ КЛАССИФИКАЦИИ И ВИЗУАЛИЗАЦИИ

### 5.0.1 Результаты жанровой классификации

AdaBoost

classical	70	13	2	1	2	1	10	0	0	0
jazz	18	35	4	1	6	17	16	2	1	0
country	4	6	24	9	3	12	18	4	15	6
pop	1	1	8	11	1	9	3	5	49	12
rock	2	7	13	7	2	26	7	5	22	9
metal	2	9	7	9	3	47	1	4	12	4
blues	1	6	16	6	2	14	25	3	15	12
disco	0	1	4	9	2	27	2	4	38	12
hiphop	0	1	8	11	0	10	3	2	49	15
reggae	0	0	12	14	1	8	10	1	38	14

Рисунок 5.1 – Матрица ошибок полученная с помощью метода классификации AdaBoost SAMME, где используется «комитет» деревьев принятия решений.

### 5.0.2 Результаты визуализации



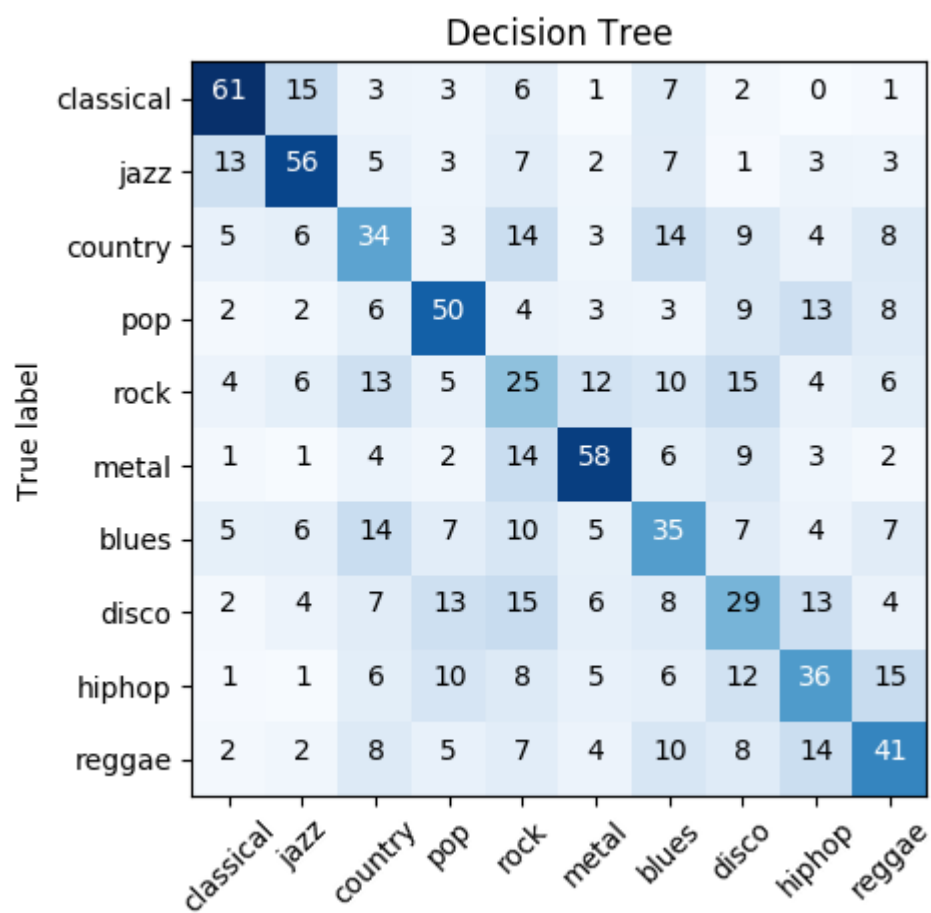


Рисунок 5.2 – Матрица ошибок полученная с помощью дерева принятия решения

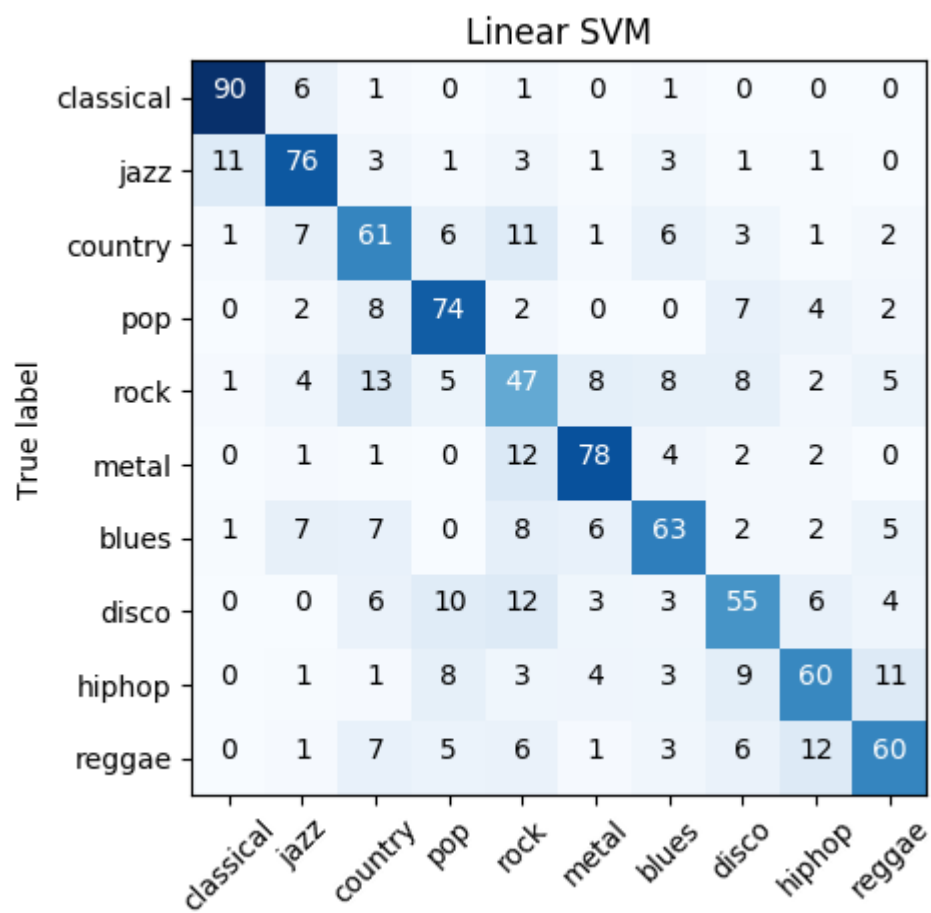


Рисунок 5.3 – Матрица ошибок полученная с помощью метода опорных векторов

Gaussian Naive Bayes

classical	81	4	7	0	5	0	2	1	0	0
jazz	23	39	8	2	5	2	11	8	0	1
country	2	1	55	3	3	5	11	16	0	2
pop	1	4	7	51	2	3	4	17	2	9
rock	1	0	21	5	11	21	4	26	1	10
metal	0	0	3	3	4	77	1	9	2	1
blues	4	5	18	2	2	5	36	17	0	10
disco	0	1	2	11	3	6	6	59	1	11
hiphop	1	1	2	8	2	6	8	26	10	36
reggae	0	2	9	6	2	2	8	10	2	58
	classical	jazz	country	pop	rock	metal	blues	disco	hiphop	reggae

Рисунок 5.4 – Матрица ошибок полученная с помощью наивного баесовского классификатора

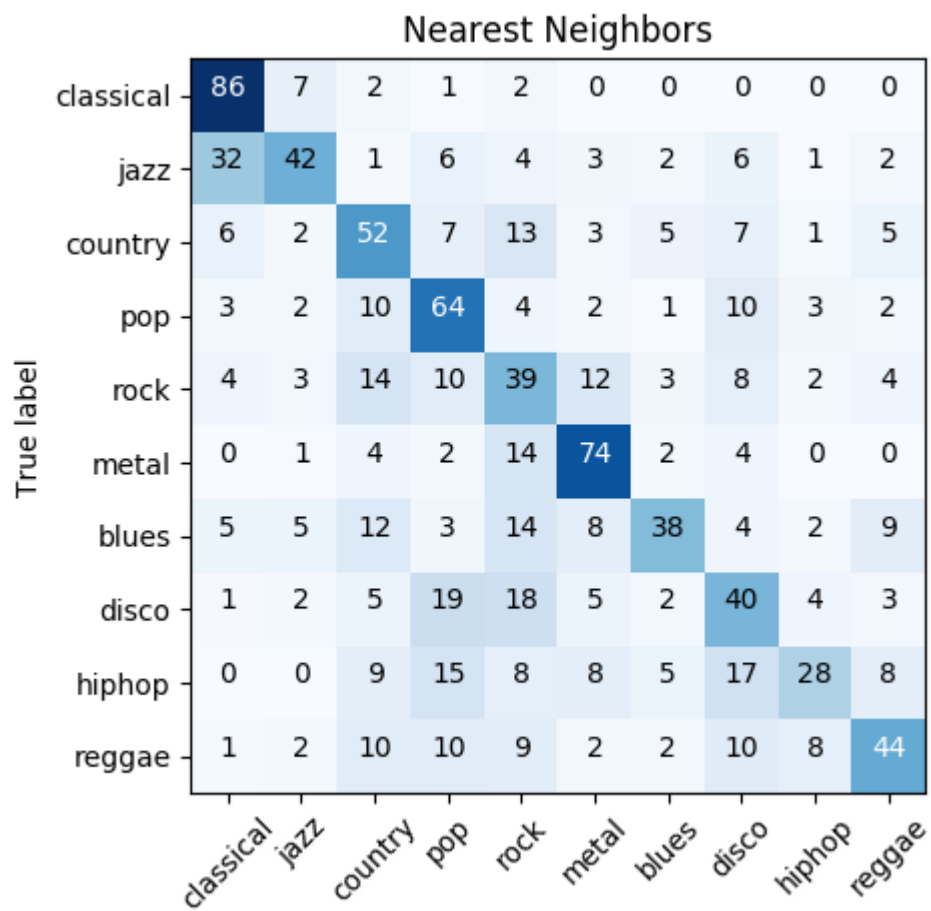


Рисунок 5.5 – Матрица ошибок полученная с помощью метода k ближайших соседей

Neural Net

True label \	classical	jazz	country	pop	rock	metal	blues	disco	hiphop	reggae
classical	90	5	1	0	1	0	2	0	0	1
jazz	8	83	1	1	2	0	3	0	0	1
country	2	3	67	5	9	1	7	2	0	3
pop	1	1	6	76	3	0	0	6	4	3
rock	3	3	8	4	50	6	5	11	3	6
metal	0	1	0	0	8	85	2	2	3	0
blues	0	4	6	0	4	6	71	2	2	4
disco	0	1	4	9	9	4	2	61	7	4
hiphop	0	0	1	5	2	4	2	8	66	12
reggae	0	1	5	5	3	1	6	4	11	63

Рисунок 5.6 – Матрица ошибок полученная с помощью многослойного перцептрона

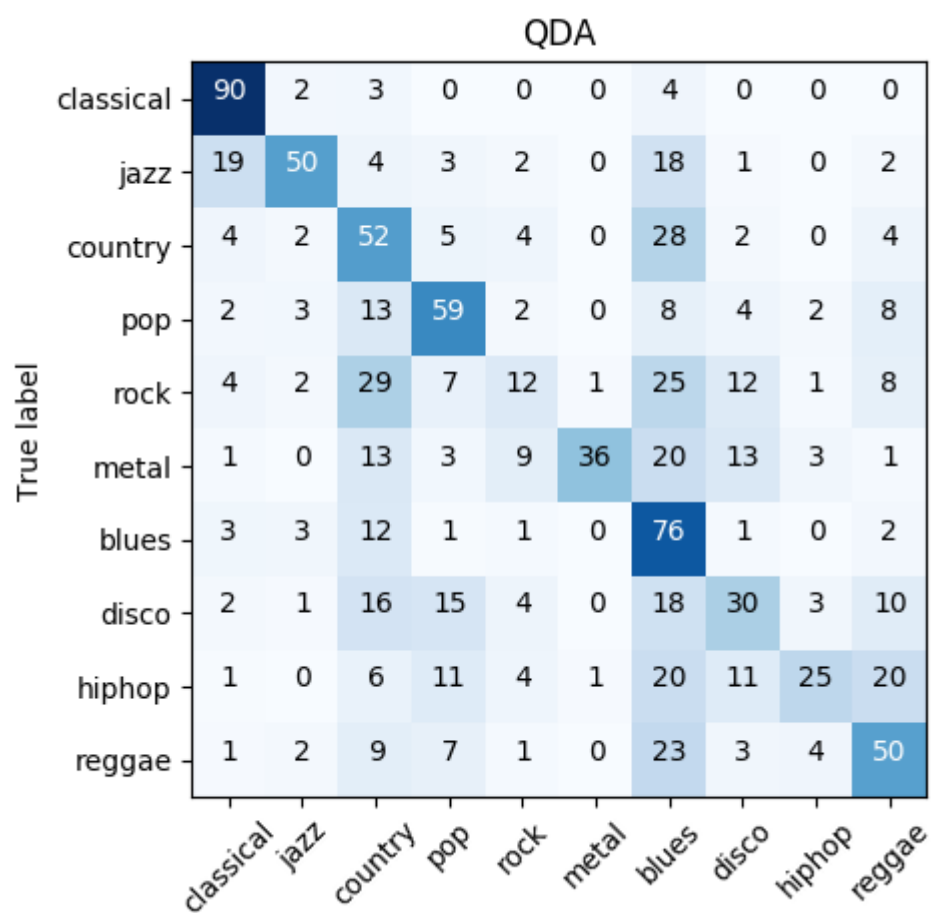


Рисунок 5.7 – Матрица ошибок полученная с помощью QDA

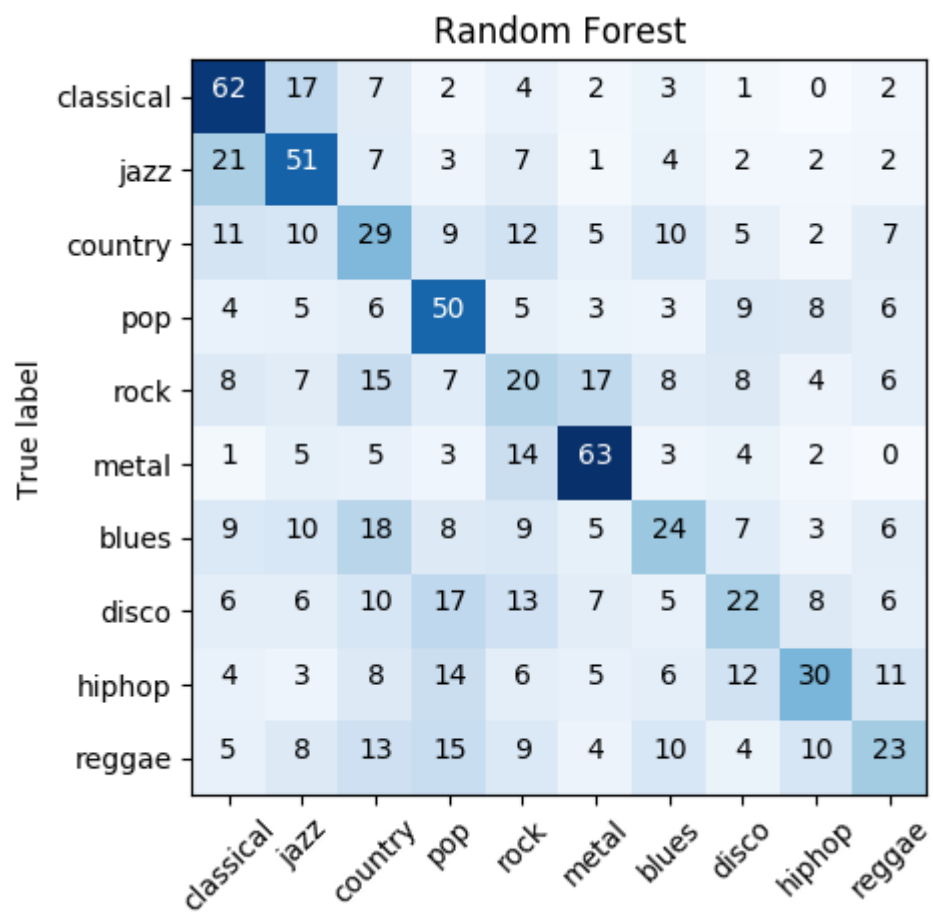


Рисунок 5.8 – Матрица ошибок полученная с случайного леса

## 6 ПРОГРАММА И МЕТОДИКА ИСПЫТАНИЙ

Тестирование программного обеспечения - это анализ, проводимый с целью предоставления заинтересованным сторонам информации о качестве тестируемого продукта или услуги. Тестирование программного обеспечения также может обеспечить объективное независимое представление программного обеспечения, которое позволит бизнесу оценить и понять риски внедрения программного обеспечения. Методы тестирования включают в себя процесс выполнения программы или приложения с целью обнаружения ошибок программного обеспечения (ошибок или других дефектов) и проверки пригодности программного продукта для использования.

Тестирование программного обеспечения включает в себя выполнение программного компонента или системного компонента для оценки одного или нескольких интересующих свойств. Как правило, эти свойства указывают, в какой степени испытываемый компонент или система:

- отвечает требованиям, которые руководствовались его дизайном и разработкой;
- правильно реагирует на все виды входных данных;
- работает в течение приемлемого времени;
- является достаточно удобной для использования;
- может быть установлена и запущена в своих рабочих средах;
- достигает общего результата, к которому стремятся его заинтересованные стороны.

Поскольку количество возможных тестов для даже простых программных компонентов практически бесконечно, всё тестирование программного обеспечения использует некоторую стратегию для выбора тестов, которые являются выполнимыми для доступного времени и ресурсов. В результате тестирование программного обеспечения обычно (но не исключительно) пытаются выполнить программу или приложение с целью обнаружения ошибок программного обеспечения (ошибок или других дефектов). Задача тестирования – это итеративный процесс, когда исправление одной ошибки может привести к нахождению другой ошибки или даже создать новую.

Тестирование программного обеспечения может обеспечить объективную независимую информацию о качестве программного обеспечения и риск его отказа пользователям или спонсорам.

Общий подход к разработке программного обеспечения часто определяет, когда и как проводится тестирование. Например, при поэтапном процессе разработки большинство тестов выполняется после того, как системные требования определены, а затем реализованы в тестируемых программах. Напротив, в рамках подхода (agile) требования, программирование и тестирование часто выполняются одновременно.

В данном проекте для повышения качества разрабатываемой продукции тестирование совершалось по принципу гибкой (agile) модели:



- люди и взаимодействие важнее процессов и инструментов;
- работающий продукт важнее исчерпывающей документации;
- сотрудничество с заказчиком важнее согласования изначальных условий контракта;
- готовность к изменениям важнее следования первоначальному плану разработки.

Благодаря концепциям, вложенным в основу гибкой модели тестирования, в результате их применения в проекте наблюдалось максимальная адаптации процесса разработки программного обеспечения к мгновенным изменениям требований заказчика.

В целом тестирование было разделено на два этапа:

- первый этап – тестирование отдельных модулей программы в процессе написания программного кода – т.н. модульное тестирование;
- второй этап – тестирование программного продукта целиком – т.н. интеграционное тестирование.

Оба этапа являются достаточно важными, так как работа каждого из них влияет на процесс проверки качества. Например, без модульного тестирования, при анализе работы программы в целом, будет происходить достаточное количество сбоев, выявить и локализовать которые может оказаться достаточно сложным заданием, в то время как при анализе работы одного модуля неисправность оказывается достаточно очевидной. И обратный случай, работоспособность каждого компонента в отдельности не гарантирует корректное поведение всей программы в целом.

Для модульного тестирования был выбран входящий в стандартную библиотеку фреймворк для модульного тестирования под названием unittest. Unittest поддерживает автоматизацию тестов, разделение кода установки и завершения тестов, агрегацию тестов в коллекции и независимость тестов от структуры отчетности. Модуль unittest предоставляет классы, которые позволяют легко поддерживать эти качества для набора тестов.

При построении модульных тестов программы используется так называемый подход «ААА», что означает «Arrange, Act, Assert». Сущность подхода состоит в том, что модульный тест включает в себе три фундаментальных действия (некоторые из которых могут быть разделены между различными тестами одного набора, покрывающими один компонент).

- часть «Arrange» подразумевает установку компонента в требуемое исходное состояние путем инициализации компонента нужными значениями, вызовами мутирующих внутреннее состояние методов и т.д;
- часть «Act» подразумевает выполнение тестируемого действия. Как правило заключается в вызове тестируемого метода или группы методов;
- часть «Assert» заключается в проверке результатов вызова тестируемого действия, соответствия внутреннего состояния системы ожидаемому по завершению вызова, факте наличия либо отсутствия необработанных ис-

ключений и т.д;

```
import unittest
import HelloWorld
# TestModule test suite
class TestStringMethods(unittest.TestCase):
    # Initialization - Arrange
    hello_world = HelloWorld()

    # Unit test
    def test_upper(self):
        # Act & Assert
        self.assertEqual('foo'.upper(), 'FOO')

    # Unit test
    def test_split(self):
        # Arrange
        s = self.hello_world.get_name()
        # Act & Assert
        self.assertEqual(s.split(), ['hello', 'world'])

if __name__ == '__main__':
    unittest.main()

...
```

Следуя вышеприведенному образцу был реализован набор тестов, покрывающий различные классы которые наследуются от `FeatureExtractorMode`.

Одной из сложностей модульного тестирования является написание такого модуля, чьи зависимости от других модулей системы максимально абстрагированы и отделены друг от друга. Иногда ввиду наличия сложных связей между разными модулями, их невозможно тестировать по отдельности (например, в случаях когда существует зависимость одного модуля от другого на этапе компиляции, такая как зависимость в конструкторе). В таких случаях принято использовать так называемые mock-объекты (ложные объекты), реализующие тот же интерфейс, что и модуль, от которого зависит тестируемый, но реализующий логику данного интерфейса достаточно просто, чтобы с гораздо большей долей вероятности не иметь неисправностей в своей реализации.

Произведя тестирование командой `python unittest -v unittests` получаем следующий вывод результатов тестирования:

```
python -m unittest -v unittests
```

```

test_energy (unittests.TestFeatureExtractor) ... ok
test_first_order_autocorrelation (unittests.
    TestFeatureExtractor) ... ok
test_zero_crossing_rate (unittests.
    TestFeatureExtractor) ... ok
test_spectral_centroid (unittests.TestFeatureExtractor)
    ... ok
test_spectral_smoothness (unittests.
    TestFeatureExtractor) ... ok
test_spectral_spread (unittests.TestFeatureExtractor)
    ... ok
test_spectral_dissymmetry (unittests.
    TestFeatureExtractor) ... ok
test_linear_regression (unittests.TestFeatureExtractor
    ) ... ok
test_rolloff (unittests.TestFeatureExtractor) ... ok
test_sfm (unittests.TestFeatureExtractor) ... ok
test_scf (unittests.TestFeatureExtractor) ... ok

```

---

```

Ran 11 tests in 0.04s

```

OK

Для тестирования классов которые наследуются от FeatureExtractorModel в качестве входных данных использовался кусок синусоиды и её спектр.

```

import unittest

```

```

class TestFeatureExtractor(unittest.TestCase):

```

```

    def setUp(self):
        # create seq
        self.test_data = np.linspace(-np.pi * 100,
                                      np.pi * 100,
                                      500)

        # get spectral data
        self.test_data_spectre = scipy.fft(self.
            test_data)

```

```

# test class Energy
def test_energy(self):
    energy = Energy()
    # get seq energy
    result = energy.get(self.seq)
    # energy must be greater than zero
    self.assertGreater(result, 0)

# test class ZeroCrossingRate
def test_zero_crossing_rate(self):
    zcr = ZeroCrossingRate()
    # get zero crossing rate
    rate = zcr.get(np.sin(self.test_data))
    real_rate = 200.0 / len(self.test_data)
    self.assertGreater(rate, real_rate)

...

```

Модульное тестирование наиболее эффективно, когда оно является неотъемлемой частью рабочего процесса разработки программного обеспечения. После написания функции или другого блока кода приложения создаются модульные тесты, которые проверяют поведение кода в ответ на стандартные, граничные и некорректные случаи ввода данных, также проверяются любые явные или предполагаемые допущения, сделанные кодом. В практике разработки программного обеспечения, известной как разработка, управляемая тестом, создается модуль тестов перед написанием кода, поэтому модульные тесты используются в качестве технической документации и спецификации функциональности.

После того, как модульные тесты были закончены для большей части функциональности и реализованные методы прошли все необходимые пункты тестирования, началось функциональное тестирование.

Функциональное тестирование является одним из ключевых видов тестирования, задача которого – установить соответствие разработанного программного обеспечения исходным функциональным требованиям заказчика. То есть проведение функционального тестирования позволяет проверить способность информационной системы в определенных условиях решать задачи, нужные пользователям.

В зависимости от степени доступа к коду системы можно выделить два типа функциональных испытаний:

- тестирование black box (черный ящик) – проведение функционального тестирования без доступа к коду системы;
- тестирование white box (белый ящик) – функциональное тестирование с доступом к внутренней структуре и коду приложения.

Таблица 6.1 – Тестирование программы

Модуль	Содержание теста	Ожидаемый результат	Тест пройден
WavModule	Конвертация MP3 в WAV	WAV-файл лежит в той же директории, что и MP3	да
WavModule	Считывание WAV-файла	Файл считывается в массив	да
Preprocessing Module	Разбивка трека на фрагменты	Трек разбивается на заданное количество фрагментов	да
Spectral Transformer	Получение коррелограммы на основе фрагмента трека	Коррелограмма получена и сохранена в массив	да
Feature Processing	Вычисления МО и СКО	МО и СКО вычислены и сохранены в вектор признаков	да
Genre Classification Module	Запуск каждого метода классификации в своём процессе	Выполнения процесса классификации на всех ядрах	да
Visualize DataModule	Сохранения графика в pdf	Файл с графиком находится в директории	да

Тестирование black box проводится без знания внутренних механизмов работы системы и опирается на внешние проявления ее работы. При этом тестировании проверяется поведение программного обеспечения при различных входных данных и внутреннем состоянии систем.

Основной информацией для создания тест-кейсов выступает документация (особенно — требования (requirements-based testing)). В случае тестирования white box создаются тест-кейсы, основанные преимущественно на коде системы программного обеспечения. Также существует расширенный тип black-box тестирования, включающего в себя изучение кода, — так называемый grey box (серый ящик). Концепция серого ящика заключается в комбинации методов белого ящика и чёрного ящика, состоящая в том, что к части кода и архитектуре доступ есть, а к части — нет.

Тестирование функциональности может проводиться в двух аспектах:

требования и бизнес-процессы. Тестирование в перспективе «требования» используют спецификацию функциональных требований к системе как основу для дизайна тестовых случаев (Test Cases). В этом случае необходимо сделать список того, что будет тестироваться, а что нет, приоритезировать требования на основе рисков, а на основе этого приоритезировать тестовые сценарии. Это позволяет сфокусироваться и не упустить при тестировании наиболее важный функционал.

Тестирование в перспективе «бизнес-процессы» используют знание этих самых бизнес-процессов, которые описывают сценарии ежедневного использования системы.

Тестирование данного программного продукта производилось несколькими пользователями на разных компьютерах:

- настольный компьютер Intel Core i7, 16 Гб RAM, Ubuntu 17.04;
- настольный компьютер Intel Core i5, 32 Гб RAM, Windows 10;
- ноутбук Intel Core i3, 4 Гб RAM, Windows 7.

В таблице 6.1 предоставлена информация о прохождении функциональных тестов данным программным обеспечением.

Как видно из таблицы, приложение хорошо справилось с тестами, что говорит о высокой работоспособности.

## 7 РУКОВОДСТВО ПОЛЬЗОВАТЕЛЯ

Модуль выделения информационных признаков из музыкального произведения представляет собой API на языке программирования Python, который решает задачи выделения спектральных, временных и иных признаков из музыкального трека, проверка значимости признаков путём использования их в задаче жанровой классификации, визуализация данных алгоритмом t-SNE.

### 7.1 Требования к аппаратному и программному обеспечению

Минимальные требования для полноценного функционирования программного обеспечения:

- операционные системы Windows XP с пакетом обновления 2 +, Windows Vista, Windows 7, Windows 8, Windows 10, Mac OS X 10.6 или более поздней версии Ubuntu 10.04 +, Debian 6 +, OpenSuSE 11.3 +, Fedora Linux 14;
- установленный интерпретатор Python 2.7;
- пакетный менеджер pip;
- pip-пакет Scipy версии 19.0 +;
- pip-пакет scikit-learn версии 0.18.1 +;
- pip-пакет PeakUtils версии 1.1.0 +;
- pip-пакет scilits.talkbox версии 0.2.5+;
- pip-пакет pymongo;
- установленный MongoDB;
- процессор Intel Pentium 4 / Athlon 64 или более поздней версии с поддержкой SSE2;
- свободное место на жестком диске 350 Мб;
- оперативная память 512 Мб.

### 7.2 Руководство по использованию программного средства

## **8 ЭКОНОМИЧЕСКОЕ ОБОСНОВАНИЕ РАЗРАБОТКИ МОДУЛЯ ВЫДЕЛЕНИЯ ИНФОРМАЦИОННЫХ ОБРАЗОВ ИЗ МУЗЫКАЛЬНОГО ПРОИЗВЕДЕНИЯ**

### **8.1 Описание функций, назначения и потенциальных пользователей ПО**

В рамках дипломного проекта был разработан модуль выделения информационных признаков из музыкального произведения.

Программный модуль решает задачи выделения спектральных, временных и иных признаков из музыкального трека, проверка значимости признаков путём использования их в задаче жанровой классификации, визуализация данных алгоритмом t-SNE.

Целью данного дипломного проекта является создание модуля, который выделял информационные образы из музыкального трека только на основании акустического анализа

Разрабатываемый программный продукт относится к категории программного обеспечения, разрабатываемого по индивидуальному заказу для использования внутри организации-заказчика.

Актуальность разработки данного ПО объясняется тем, существует потребность в наличии доступного инструмента, позволяющего выделять информационные образы из музыкального трека только на основании акустического анализа.

Целью экономического обоснования программных средств является определение экономической выгоды создания данного продукта и дальнейшего его применения.

### **8.2 Расчёт затрат на разработку ПО**

Данный модуль может быть использован в сервисе рекомендации, который позволит значительно улучшить качество рекомендаций. Также потенциальным пользователем, может быть любой человек, который хочет анализировать данные в музыкальной предметной области. Разработка программного средства производится по индивидуальному заказу организации-заказчика, однако в будущей перспективе может выйти для свободной реализации на рынке.

Для оценки экономической эффективности разработанного программного обеспечения проводится расчет затрат на разработку приложения, оценка прибыли от продажи одного такого приложения и расчет показателей эффективности инвестиций в его разработку.

Разрабатываемый программный продукт относится ко второй категории сложности, так как существует возможность переносимости программного обеспечения.



По степени новизны разрабатываемая система автоматизации относится к категории В с коэффициентом новизны  $K_n = 0,7$ , так как было выявлено несколько аналогов. Но стоит принять во внимание тот факт, что данный проект не подразумевает в себе использование принципиально нового типа электронно-вычислительных машин, операционных систем.

При разработке проекта используются существующие технологии и средства разработки, которые охватывают около 20 – 30% реализуемых функций, поэтому коэффициент использования стандартных модулей принимается равным 0,8.

При расчете сметы затрат будут использоваться данные, приведенные в таблице 8.1.

Таблица 8.1 – Исходные данные для расчета

Наименования показателей	Буквенные обозначения	Ед.измерения	Количество
1	2	3	4
Фонд социальной защиты населения (от заработной платы)	$H_{\text{соц}}$	%	34
Обязательное страхование (от несчастных случаев на производстве, от заработной платы)	$H_{\text{стр}}$	%	0,6
Налог на прибыль	$H_{\text{приб}}$	%	18
Налог на недвижимость (от стоимости зданий и сооружений)	$H_{\text{недв}}$	%	1
НДС (Налог на добавленную стоимость)	НДС	%	20
Норма дисконта	$E_n$	%	17
Тарифная ставка 1-го разряда	$T_{m1}$	руб.	31
Часовая тарифная ставка 1-го разряда	$T_{\text{ч}}$	руб.	0,19
Установленный фонд рабочего времени	$\Phi_{\text{рв}}$	часов	166
Продолжительность рабочего дня	$T_{\text{д}}$	часов	8
Тарифный коэффициент	$T_k$	-	2,03
Коэффициент премирования	$K_{\text{п}}$	единиц	1,5

Основой для расчёта сметы затрат является основная заработная плата разработчиков проекта. Затраты на основную заработную плату команды разработчиков определяются исходя из состава и численности команды, раз-

меров месячной заработной платы каждого из участников команды, а также общей трудоемкости разработки программного обеспечения.

Для осуществления упрощённого расчёта затрат на разработку ПО следует произвести расчёт следующих статей:

- затраты на основную заработную плату разработчиков;
- затраты на дополнительную заработную плату разработчиков;
- отчисления на социальные нужды;
- прочие затраты (амортизация оборудования, расходы на электроэнергию, командировочные расходы, накладные расходы и т.п.).

Расчёт затрат на основную заработную плату разработчиков осуществляется на основе численности и состава команды, размеров месячной заработной платы каждого из участников команды, а также общей трудоёмкости процесса разработки программного обеспечения.

В данном случае имеются четыре работника – два программиста I-й категории, руководитель проекта и специалист по анализу данных. Месячная тарифная ставка каждого исполнителя определяется по формуле:

$$T_M = T_{M1} \cdot T_K, \quad (8.1)$$

где  $T_{M1}$  — месячная тарифная ставка первого разряда, руб.;

$T_K$  — тарифный коэффициент.

Месячная тарифная ставка руководителя проекта составит:

$$T_{M.рук} = 19,35 \times 31 = 600 \text{ руб.} \quad (8.2)$$

Месячная тарифная ставка программиста составит:

$$T_{M.прог} = 16,13 \times 31 = 500 \text{ руб.} \quad (8.3)$$

Месячная тарифная ставка специалиста по анализу данных составит:

$$T_{M.спец} = 19,35 \times 31 = 600 \text{ руб.} \quad (8.4)$$

Исходя из месячной тарифной ставки рассчитывается часовая тарифная ставка:

$$T_ч = \frac{T_M}{\Phi_p}, \quad (8.5)$$

Часовая тарифная ставка руководителя проекта в соответствии с формулой составит:

$$T_{ч.рук} = \frac{600}{166} = 3,61 \text{ руб.} \quad (8.6)$$

Часовая тарифная ставка программиста первой категории в соответствии с формулой составит:

$$T_{ч.прог} = \frac{500}{166} = 3,01 \text{ руб.} \quad (8.7)$$

Часовая тарифная ставка специалиста по анализу данных в соответствии с формулой составит:

$$T_{\text{ч.спец}} = \frac{600}{166} = 3,61 \text{ руб.} \quad (8.8)$$

Основная заработная плата исполнителей на конкретное программное средство является суммой заработных плат каждого из исполнителей в отдельности и определяется по формуле:

$$Z_o = \sum_{i=1}^n T_{\text{чи}} \cdot t_i, \quad (8.9)$$

где  $n$  — количество исполнителей, занятых разработкой конкретного ПО;

$T_{\text{чи}}$  — часовая тарифная ставка  $i$ -го исполнителя, руб;

$t_i$  — трудоемкость работ, выполняемых  $i$ -го исполнителем, час.

Для руководителя, при заработной плате равной 600 рублей, часовая заработная плата равна 3,61 рубля. Для разработчика, при заработной плате равно 500 рублей, часовая заработная плата равна 3,01 рубля. Для специалиста по анализу данных, часовая заработная плата равна 3,61. Трудоемкость определяется исходя из сложности разработки программного продукта и объема выполняемых им функций. В нашем случае она составляет 90 дней или 720 часов. Тогда основная зарплата исполнителей равна:

$$Z_o = (3,01 + 3,61 + 3,01 + 3,61) \times 720 = 9532,8 \text{ руб.} \quad (8.10)$$

Затраты на дополнительную заработную плату команды разработчиков включает выплаты, предусмотренные законодательством о труде (оплата отпусков, льготных часов, времени выполнения государственных обязанностей и других выплат, не связанных с основной деятельностью исполнителей), и определяется по формуле:

$$Z_d = \frac{Z_o \cdot H_d}{100}, \quad (8.11)$$

где  $Z_o$  — затраты на основную заработную плату с учетом премии, руб;

$H_d$  — норматив дополнительной заработной платы, 15 %.

В результате подстановки получим:

$$Z_d = \frac{14\,299,2 \times 15}{100} = 2144,88 \text{ руб.} \quad (8.12)$$

Отчисления на социальные нужды включают в предусмотренные законодательством отчисления в фонд социальной защиты (34%) и фонд обязательного страхования (0,6%) в процентах от основной и дополнительной

Таблица 8.2 – Расчет затрат на основную заработную плату команды

№	Участник команды	Выполняемые работы	Месячная заработная плата, р	Часовая заработная плата, р.	Трудоемкость работ, ч.	Основная заработная плата, р.
1	Руководитель проекта	Контроль, помощь	600	3,61	720	2599,2
2	Программист 1-й категории	Разработка	500	3,01	720	2167,2
3	Программист 1-й категории	Разработка	500	3,01	720	2167,2
4	Специалиста по анализу данных	Разработка	600	3,61	720	2599,2
ПРЕМИЯ (50%)						4766,4
Итого затраты на основную заработную плату разработчиков						14 299,2

заработной платы и вычисляются по формуле:

$$З_{\text{соц}} = \frac{(З_o + З_d) \cdot Н_{\text{соц}}}{100}, \quad (8.13)$$

где  $Н_{\text{соц}}$  — норматив отчислений на социальные нужды 34% и норматив отчислений в фонд социального страхования, 0,6 % .

$$З_{\text{соц}} = \frac{(14\,299,2 + 2144,88) \times 34,6}{100} = 5689,65 \text{ руб.} \quad (8.14)$$

Расходы по статье «Прочие затраты» включают затраты на приобретение и подготовку специальной научно-технической информации и специальной литературы. Определяются по формуле:

$$З_{\text{пз}} = \frac{З_o \cdot Н_{\text{пз}}}{100}, \quad (8.15)$$

где  $Н_{\text{пз}}$  — норматив прочих затрат в целом по организации, 100 % .

$$З_{\text{пз}} = \frac{14\,299,2 \times 100}{100} = 14\,299,2 \text{ руб.} \quad (8.16)$$

Общая сумма расходов по смете определяется:

$$З_{\text{п}} = З_o + З_d + З_{\text{соц}} + Р_{\text{пр}}, \quad (8.17)$$

Подставив рассчитанные ранее значения в формулу, получим:

$$З_{\pi} = 14\,299,2 + 2144,88 + 5689,65 + 14\,299,2 = 36\,432,93 \text{ руб.} \quad (8.18)$$

Полная сумма затрат на разработку программного обеспечения находится путем суммирования всех рассчитанных статей затрат (таблица 8.3).

Таблица 8.3 – Затраты на разработку программного обеспечения

Статья затрат	Сумма, руб.
Основная заработная плата команды разработчиков	14 299,2
Дополнительная заработная плата команды разработчиков	2144,88
Отчисления на социальные нужды	5689,65
Прочие затраты	14 299,2
Общая сумма затрат на разработку	36 432,93

### 8.3 Оценка результата (эффекта) от продажи ПО

Экономический эффект представляет собой прирост чистой прибыли, полученный организацией в результате использования разработанного ПО. Как правило, он может быть достигнут за счет:

- уменьшения (экономии) затрат на заработную плату за счет замены «ручных» операций и бизнес-процессов информационной системой;
- ускорения скорости обслуживания клиентов и рост возможности обслуживания большего их количества в единицу времени, т.е. рост производительности труда;
- появления нового канала сбыта продукции или получения заказов (как в случае внедрения интернет-магазина);
- и т.п.

Экономический эффект организации-разработчика программного обеспечения в данном случае заключается в получении прибыли от его продажи множеству потребителей. Прибыль от реализации в данном случае напрямую зависит от объемов продаж, цены реализации и затрат на разработку данного ПО.

При свободной реализации на рынке IT экономический эффект заключается в получении прибыли от его продажи множеству потребителей. Прибыль от реализации в данном случае напрямую зависит от объемов продаж, цены реализации и затрат на разработку данного проекта. Организация, осуществляющая финансово-хозяйственную деятельность, заинтересована не только в наибольшей величине прибыли, но и в отдаче вложенных средств. Отдача или эффективность вложенных средств характеризуется размером

прибыли, получаемой предприятием. Показатели рентабельности характеризуют эффективность работы организации в целом, а также доходность различных направлений деятельности.

Далее следует определить цену на одну копию (лицензию) ПО. Цена формируется на основе затрат на разработку и реализацию ПО. Тогда расчет прибыли от продажи одной копии (лицензии) ПО осуществляется по формуле:

$$Ц = П_{ед} + НДС + \frac{З_p}{N}, \quad (8.19)$$

- где Ц — цена реализации одной копии (лицензии) ПО, руб.;
- $З_p$  — сумма расходов на разработку и реализацию программного обеспечения, руб. В дипломном проекте расходы на реализацию приняты равным 5%;
- N — количество копий (лицензий) ПО, которое будет куплено клиентами за год, руб.;
- $П_{ед}$  — прибыль, получаемая организацией-разработчиком от реализации одной копии программного продукта, руб.;
- НДС — сумма налога на добавленную стоимость, руб.;

Прибыль от продажи одной копии (лицензии) ПО осуществляется по формуле:

$$П_{ед} = \frac{C_{п} \cdot Y_p}{100 \times N}, \quad (8.20)$$

- где  $C_{п}$  — себестоимость ПО, руб.;
- $Y_p$  — запланированный уровень рентабельности, 35 %.

$$П_{ед} = \frac{36\,432,93 \times 35}{100 \times 20} = 637,6 \text{ руб.} \quad (8.21)$$

Цена одной копии (лицензии) программного обеспечения данного направления составляет 637,6 руб.

Следовательно годовая прибыль составит:

$$П = 637,6 \times 20 = 12\,751,5 \text{ руб.} \quad (8.22)$$

Подставив рассчитанные ранее значения в формулу, получим:

$$Ц = 637,6 + \frac{36\,432,93 \times 1,05}{20} = 2550,32 \text{ руб.} \quad (8.23)$$

Рентабельность затрат рассчитаем по формуле. Проект будет экономически эффективным, если рентабельность затрат на разработку программного средства будет не меньше средней процентной ставки по банковским депозитным вкладам.

$$P = 100 \cdot \frac{П}{З_p}, \quad (8.24)$$

где  $P$  — годовая прибыль, руб.;

$Z_p$  — сумма расходов на разработку и реализацию, руб.

$$P = 100 \times \frac{12\,751,5}{36\,432,93} = 35\%. \quad (8.25)$$

Средняя процентная ставка по депозиту за март 2017 года составляет 7,9% для юридических лиц, что меньше рентабельности данного проекта. Соответственно проект является экономически эффективным. Учитывая налог на прибыль, можно рассчитать итоговую сумму, которая останется разработчику и будет являться его экономическим эффектом:

$$\text{ЧП} = P - \frac{P \cdot N_{\text{приб}}}{100}, \quad (8.26)$$

где  $N_{\text{приб}}$  — ставка налога на прибыль, 18 % ;

$P$  — прогнозируемая прибыль, руб..

$$\text{ЧП} = 12\,751,5 - \frac{12\,751,5 \cdot 18}{100} = 10\,456,23 \text{ руб.} \quad (8.27)$$

Чистая прибыль от реализации ПО (ЧП = 10456,23 рублей) остается организации-разработчику и представляет собой экономический эффект.

#### 8.4 Расчёт показателей эффективности инвестиций в разработку ПО

Так как сумма инвестиций в разработку ПО больше суммы годового экономического эффекта, то экономическая целесообразность инвестиций в разработку и использование программного средства осуществляется на основе расчёта и оценки следующих показателей:

- чистый дисконтированный доход (ЧДД);
- срок окупаемости инвестиций;
- рентабельность инвестиций.

Метод чистой дисконтированной доходности основан на сопоставлении дисконтированной стоимости денежных поступлений (инвестиций), генерируемых предприятием в течение прогнозируемого периода. Целью данного метода является выявление реального размера прибыли, который может быть получен организацией вследствие реализации данного инвестиционного проекта.

Коэффициент дисконтирования соответствующего года  $t$  определяется по формуле:

$$a_t = \frac{1}{(1 + E_n)^t}, \quad (8.28)$$

где  $E_n$  — норма дисконта, равная или больше средней процентной ставки по банковским депозитам, действующей на момент проведения расчётов;

$t$  — порядковый номер года периода реализации инвестиционного проекта (1 – 2017, 2 – 2018, 3 – 2019, 4 – 2020).

Подставляя значения, получим значения коэффициентов дисконтирования для 2017 – 2020 годов:

$$a_1 = \frac{1}{(1 + 0,17)^1} = 0,85, \quad (8.29)$$

$$a_2 = \frac{1}{(1 + 0,17)^2} = 0,73, \quad (8.30)$$

$$a_3 = \frac{1}{(1 + 0,17)^3} = 0,62, \quad (8.31)$$

$$a_4 = \frac{1}{(1 + 0,17)^4} = 0,53. \quad (8.32)$$

Чистый дисконтированный доход рассчитывается по следующей формуле:

$$\text{ЧДД} = \sum_{t=1}^n (P_t \cdot a_t - Z_t \cdot a_t), \quad (8.33)$$

где  $n$  — расчётный период, лет;

$P_t$  — результат (экономический эффект), полученный в году  $t$ , руб.;

$Z_t$  — затраты (инвестиции в разработку ПО) в году  $t$ , руб.

$$\text{ЧДД} = -20\,129,21 + 9308,59 + 7905,93 + 6758,29 = 3843,6 \text{ руб.} \quad (8.34)$$

Рентабельность инвестиций рассчитывается по формуле:

$$P_{\text{и}} = \frac{\sum_{t=0}^n P_t \cdot a_t}{\sum_{t=0}^n Z_t \cdot a_t} \cdot 100\%, \quad (8.35)$$

По формуле рентабельность инвестиций равна:

$$P_{\text{и}} = \frac{10\,838,77 + 9308,59 + 7905,93 + 6758,29}{30\,967,99} \cdot 100 = 112,3\%. \quad (8.36)$$

На основании проведенной анализа экономической эффективности разработки системы можно сделать следующие выводы. Из оценки затрат на создание и поддержание программного продукта и прибыли, полученной за продажу копии программы, мы подсчитали рентабельность затрат. Она оказалась выше средней процентной ставки по депозиту, что показывает выгоду от реализации данной продукции. Показатели эффективности



Таблица 8.4 – Расчет эффективности инвестиционного проекта по разработке программного обеспечения

Показатели	2017	2018	2019	2020
<b>РЕЗУЛЬТАТ</b>				
Экономический эффект	12 751,5	12 751,5	12 751,5	12 751,5
Дисконтированный результат	10 838,77	9308,59	7905,93	6758,29
<b>ЗАТРАТЫ</b>				
Инвестиции в разработку программного средства	36 432,93	0	0	0
Дисконтированные инвестиции	30 967,99	0	0	0
Чистый дисконтированный доход по годам	–20 129,21	9308,59	7905,93	6758,29
Дисконтированный результат	–20 129,21	–10 819,6	–2913,69	3843,6
Коэффициент дисконтирования	0,85	0,73	0,62	0,53

инвестиций на создание программного обеспечения в результате подсчетов показали, что разработка данного программного продукта является экономически целесообразной.

## ЗАКЛЮЧЕНИЕ

В результате преддипломной практики были проведены исследования в области выделения информационных образов и сервисов рекомендации. Было проведено исследование существующих способов выделения информационных признаков при решении задач: распознавания неперкуSSIONных инструментов, распознавания заимствований и классификации жанров. Были рассмотрены существующие сервисы и приложения рекомендации музыки на основе акустического анализа, рассмотрены методы и работы, а также их положительные и отрицательные стороны. Также была разработана архитектура модуля выделения информационных образов из музыкальных произведений, и описана его структура.

В процессе реализации данного проекта в рамках дипломного проектирования будет создан конечный продукт реализующий данную архитектуру.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Yandex N. V. Десять миллионов треков на Яндекс.Музыке [Электронный ресурс]. — Электронные данные. — Режим доступа: <https://yandex.ru/blog/company/69072>. — Дата доступа: 05.02.2017.
- 2 Kevin Murnane. The US Music Industry Crossed A Threshold In 2016 [Электронный ресурс]. — Электронные данные. — Режим доступа: <https://www.forbes.com/sites/kevinmurnane/2017/01/18/the-us-music-industry-passed-a-milestone-in-2016/#312060b15a90>. — Дата доступа: 05.02.2017.
- 3 Yandex N. V. Как это работает? Рекомендации в Яндекс.Музыке [Электронный ресурс]. — Электронные данные. — Режим доступа: <https://yandex.ru/blog/company/92883>. — Дата доступа: 05.02.2017.
- 4 M. Jones. Introduction to approaches and algorithms [Электронный ресурс]. — Электронные данные. — Режим доступа: [https://www.ibm.com/developerworks/opensource/library/os-recommender1/index.html?S\\_TACT=105AGX99&S\\_CMP=CP](https://www.ibm.com/developerworks/opensource/library/os-recommender1/index.html?S_TACT=105AGX99&S_CMP=CP). — Дата доступа: 05.02.2017.
- 5 George Tzanetakis Georg Essl, Perry Cook. Automatic Musical Genre Classification Of Audio Signals. — 2001.
- 6 Bashi, Jamil George. Music Similarity Measures for Interpolative Playlist Generation. — 2008.
- 7 Balen, Jan Van. Automatic Recognition of Samples in Musical Audio. — 2011.
- 8 Martin, Keith Dana. Sound-Source Recognition: A Theory and Computational Model. — 1999.
- 9 Brown, Judith C. Calculation of a constant Q spectral transform. — 1991.
- 10 Ian Glover, Peter Grant. Digital Communications. — 1998.
- 11 Mohit Rajani, Luke Ekkizogloy. Supervised Learning in Genre Classification. — 2009.
- 12 Logan, Beth. Mel Frequency Cepstral Coefficient for Music Modeling. — 2000.
- 13 HOLO — Система анализа музыки [Электронный ресурс]. — Электронные данные. — Режим доступа: <https://habrahabr.ru/post/194724/>. — Дата доступа: 05.02.2017.
- 14 Формирование музыкальных предпочтений у нейронной сети — эксперимент по созданию умного плеера [Электронный ресурс]. — Электронные данные. — Режим доступа: <https://habrahabr.ru/post/263811/>. — Дата доступа: 05.02.2017.
- 15 About Pandora [Электронный ресурс]. — Электронные данные. — Режим доступа: <http://www.pandora.com/about>. — Дата доступа: 05.02.2017.
- 16 The Best Music Services Compared [Электронный ресурс]. — Электронные данные. — Режим доступа:

<http://www.techlicious.com/guide/best-music-service-best-for-you/>. — Дата доступа: 05.02.2017.

17 Consumer item matching method and system US 7003515 B1 [Электронный ресурс]. — Электронные данные. — Режим доступа: <http://www.google.com/patents/US7003515?dq=7,003,515>. — Дата доступа: 05.02.2017.

18 Tim Westergren (Music Genome Project Founder) [Электронный ресурс]. — Электронные данные. — Режим доступа: <http://www.tinymixtapes.com/features/tim-westergren-music-genome-project-founder>. — Дата доступа: 05.02.2017.

ПРИЛОЖЕНИЕ А  
(обязательное)

Вводный плакат

ПРИЛОЖЕНИЕ Б  
(обязательное)

Схема структурная

