СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 4](#_Toc513065561)

[**0.1** История задачи 4](#_Toc513065562)

[**0.2** Существующие подходы и их слабые стороны 6](#_Toc513065563)

[**0.3** Цель и задачи работы 11](#_Toc513065564)

[**0.4** Используемые методы 12](#_Toc513065565)

[**1** Обзор литературы 12](#_Toc513065566)

[**1.1** Звук 12](#_Toc513065567)

[**1.3** Некоторые понятия из теории музыки 14](#_Toc513065568)

[**1.4** Цифровой звук 15](#_Toc513065569)

[**1.5** Свойства музыкальных звукозаписей 16](#_Toc513065570)

[**1.6** Формализация задачи 16](#_Toc513065571)

[**1.6.1** Частотно-временное представление 17](#_Toc513065572)

[**1.6.2** Классификация 17](#_Toc513065573)

[**1.7** Предварительная обработка 17](#_Toc513065574)

[**1.8** Спектрограмма 18](#_Toc513065575)

[**1.9** Векторы признаков 20](#_Toc513065576)

[СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ 21](#_Toc513065577)

# ВВЕДЕНИЕ

## История задачи

Музыкальная транскрипция – это процесс анализа акустического музыкального сигнала и запись музыкальных параметров звуков, возникающих в нем. Традиционно музыка записывается в виде нотной нотации, которая позволяет указать высоту тона, время начала и продолжительность каждого воспроизводимого звука.

Музыкальную транскрипцию также можно представить в репрезентационном смысле как преобразование акустического сигнала в символическое представление. С другой стороны, музыкальная нотация - это скорее инструкция для исполнителя, чем представление музыкальной записи.

Она описывает музыку на языке, понимаемом музыкантом и служит для воспроизводства музыкальных звуков. С этой точки зрения, музыкальная транскрипция может быть рассмотрена как обнаружение «рецепта» или реверс-инжиниринг «исходного кода» музыкального сигнала. Прикладная нотация не обязательно должна быть традиционной музыкальной нотацией, скорее любым символическим представлением, которое дает необходимую информацию для исполнения записи используя доступные музыкальные инструменты. В случае электронного синтезатора, используемого для повторного синтеза, MIDI файл является примером такого музыкального представления записи.

Музыкальная нотация не только позволяет воспроизводить музыкальную запись, но также и модифицировать, поправлять и обрабатывать музыку на более высоком уровне абстракции. Другое важное применение музыкальной транскрипции – это структурированное аудио кодирование. MIDI-подобное представление невероятно компактно, но в то же время оно сохраняет идентифицируемость и характеристики музыкального произведения в значительной степени. Другое применение музыкальной транскрипции включает *поиск информации* на основе мелодии фрагмента, *музыкальный анализ* импровизируемой этнической музыки и *интерактивные музыкальные системы,* которые генерируют аккомпанементпоющему или играющему солисту.

Автоматическая транскрипция полифонической музыки является субъектом все возрастающих исследований на протяжении последних 30 лет. Проблема во многом аналогична автоматическому распознаванию речи, но не получила сопоставимых академического или коммерческого применения.

Несмотря на большое кол-во попыток решить проблему, практически применимая, универсальная система музыкальной транскрипции не существует и по сей день. Существующие системы на порядок хуже специалистов в точности и гибкости. Самые новые системы, однако, достигли определенной степени точности в транскрибировании полифонической музыки ограниченной сложности. Типичные ограничения целевого сигнала таковы:

* число одновременных нот ограниченно (или фиксировано);
* интерференция ударных и перкуссионных инструментов обычно запрещено;

Как ни странно, даже транскрипция одноголосого произведения является еще нерешенной проблемой.

## **0.2** Существующие подходы и их слабые стороны

Мурер [1977] впервые представил ограниченную систему для транскрипции дуэта. С тех пор ряд акустических моделей для полифонической транскрипции были представлены как в частотной области [Rossi et al., 1997, Sterian, 1999, Dixon, 2000], так и во временной [Bello et al., 2002].

Однако эти методы основаны на анализе, который предполагает, что сигнал будет иметь определенную структуру. Музыкальный тон создается с периодичностью на определенной основной частоте в звуковом сигнале.

Предположение, что высота ноты возникает из гармонических составляющих сильно основана на музыкальной акустике, но для транскрипции это необязательно. Во многих полях (таких как автоматическое распознавание речи) классификаторы для определенных событий построенных с использованием минимальных предварительных знаний о том, как они представлены в признаках. Marolt [2004] представил классификационный подход с использованием нейронных сетей, но блок фильтров адаптивных осцилляторов был необходимых для уменьшения ошибочных вставок нот. Байесовские модели также были предложены для транскрипции музыки, однако, такое их трактование слишком основано на физическом представлении модели генерации музыкального звука [1].

К сожалению, взаимодействие гармонических рядов, происходящее в полифонической музыке значительно запутывает автоматическую транскрипцию.

Poliner, Ellis (2006):

* метод опорных векторов и обучение на спектральных свойствах для классификации нот во фрейме;
* выход классификатора пропускается через скрытые марковские модели;
* система используется для транскрипции как синтезированных, так и живых записей фортепиано [2].

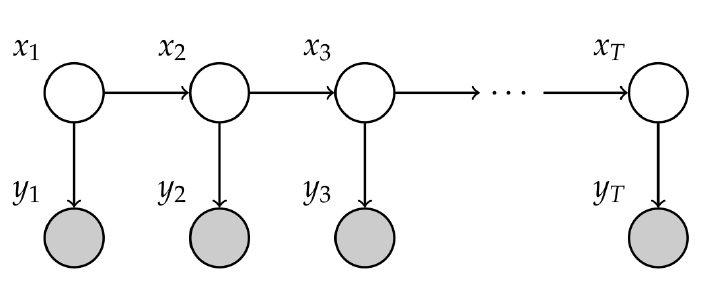


Рисунок 0.1 – Пост-процессинг скрытыми марковскими моделями

Ryyn ̈anen, Klapuri (2005):

* первичная обработка — multiple-F0 estimator;
* HMM для нот, модель тишины и музыкологическая модель;
* музыкологическая модель отвечает за поиск переходов между HMM и моделью тишины;

События нажатия нот описываются скрытой марковской моделью. Модель использует три акустические характеристики с оценкой множественной основной частоты (F0) для вычисления вероятности различных нот и выполняет временную сегментацию нот. Переходы между нотами контролируются с помощью музыкальной модели с оценкой музыкального ключа и bigram моделей. Окончательная транскрипция получается путем поиска нескольких путей через модели нот [3].

Cистема Klapuri вычисляет множественные фундаментальные частоты из спектральных пиков, используя вычислительную модель слуховой периферии человека. Затем дискретные скрытые марковские модели (HMM) итеративно применяются для извлечения последовательности мелодий из фундаментальных оценок частоты [Ryyn ̈anen and Klapuri, 2005].

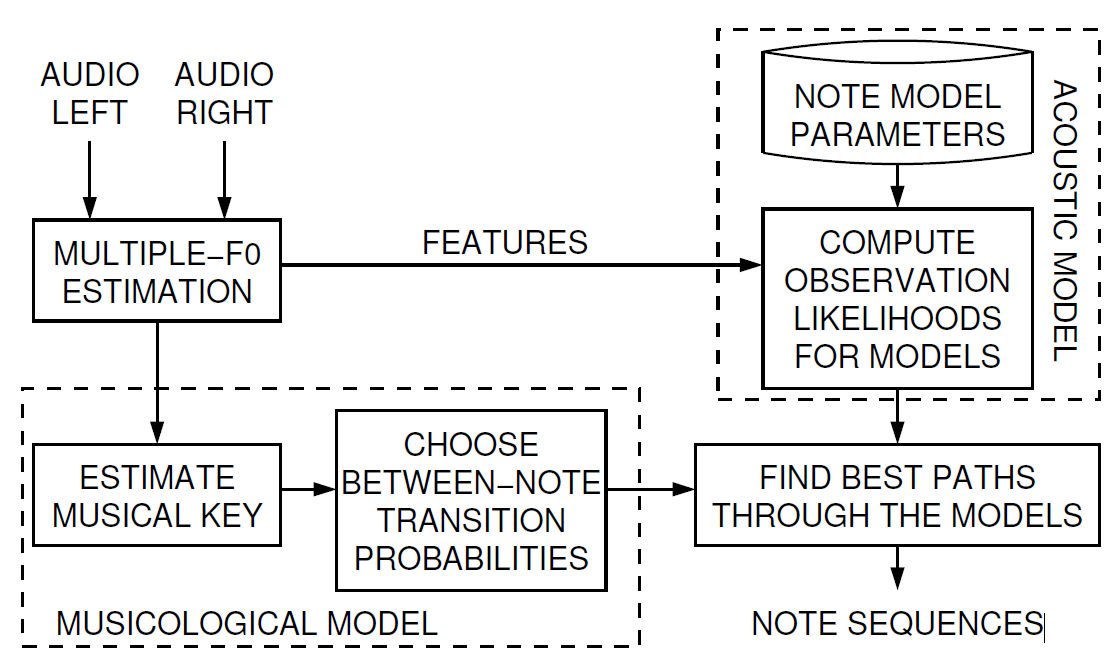


Рисунок 0.2 – Блок-диаграмма метода Ryyn ̈anen, Klapuri (2005)

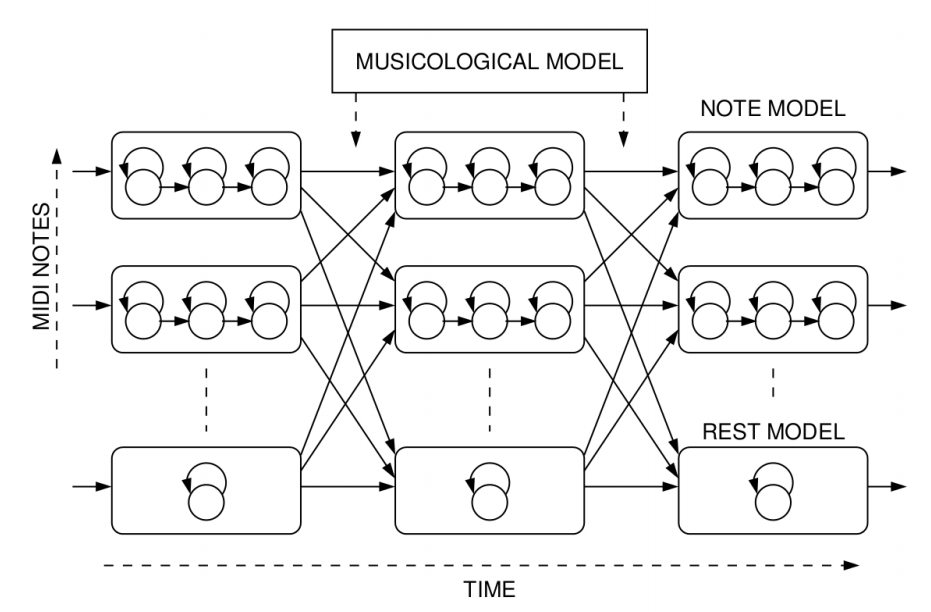


Рисунок 0.3 – Сеть модели ноты и модели тишины

Marolt (2004):

* используются нейросети и их обобщения;
* вводятся «адаптивные осцилляторы» для определения частот звуков;
* с помощью особой синхронизации осцилляторов удается учитывать обертоны в музыке;
* объединяя такие осцилляторы в сети, получем систему для отслеживания целых групп обертонов, т.е. фактически получаем необходимую транскрипцию;

На рисунке 0.4 - модель отслеживания частот, нейронная сеть для учета временных задержек. Система также содержит детектор появления ноты, модуль для детектирования повторяющихся нот и простой алгоритм для оценки длины и громкости [4].

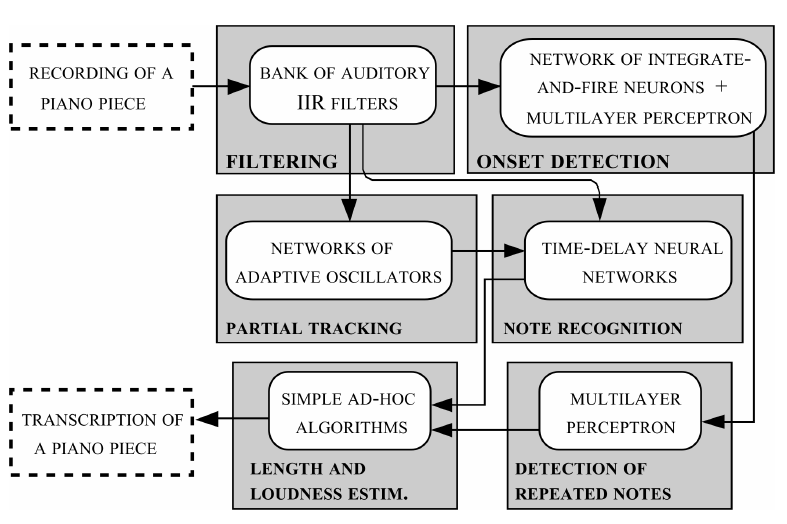


Рисунок 0.4 – Структура системы Marolt (2004)

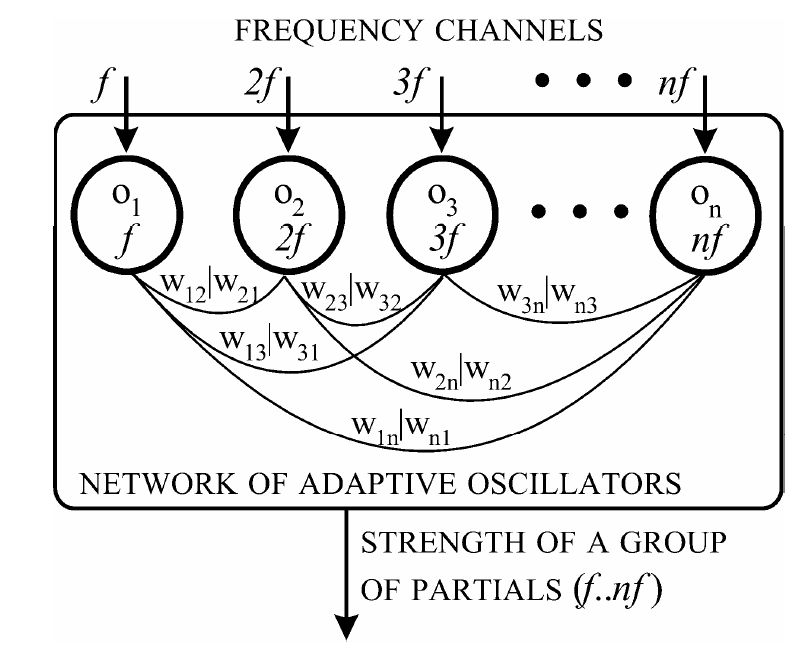


Рисунок 0.5 – Сеть адаптивных осцилляторов

Сеть осцилляторов может доходить до десяти взаимосвязанных осцилляторов. Их начальные частоты устанавливаются как целые кратные частоте первого осциллятора. Система использует 88 сетей осцилляторов для отслеживания групп, соответствующих всем 88 тонам фортепиано (A0-C8). Начальная частота первого осциллятора в каждой сети установлена на высоту одного из 88 тонов фортепиано (см рисунок 0.5).

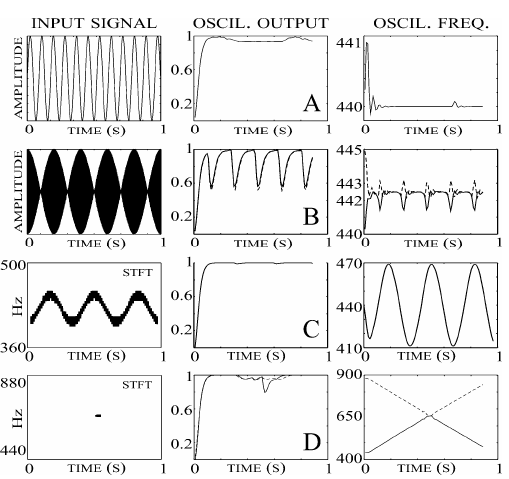


Рисунок 0.6 – Отслеживание частот адаптивными осцилляторами

Zalani, Mittal (2014):

* RNN-RBM для обучения признаков;
* STFT тоже учитывается как признаки;
* 88 SVM-классификаторов One-vs-All;
* пост-процессинг скрытыми марковскими моделями;

В этом подходе обучают 88 бинарных классификаторов, которые служат для транскрибирования нот в полифонической записи. Каждый классификатор определяет присутствие одной ноты в записи в каждый отдельный отсчет времени. Обучение без учителя RNN-RBM (Рекурсивные нейронные сети и ограниченная машина Больцмана). Convolutional Deep Belief Network (CDBN) применены. Классификация методом опорных векторов (SVM) «один против всех» one-vs-all. HMM сглаживание для улучшения результатов [5].

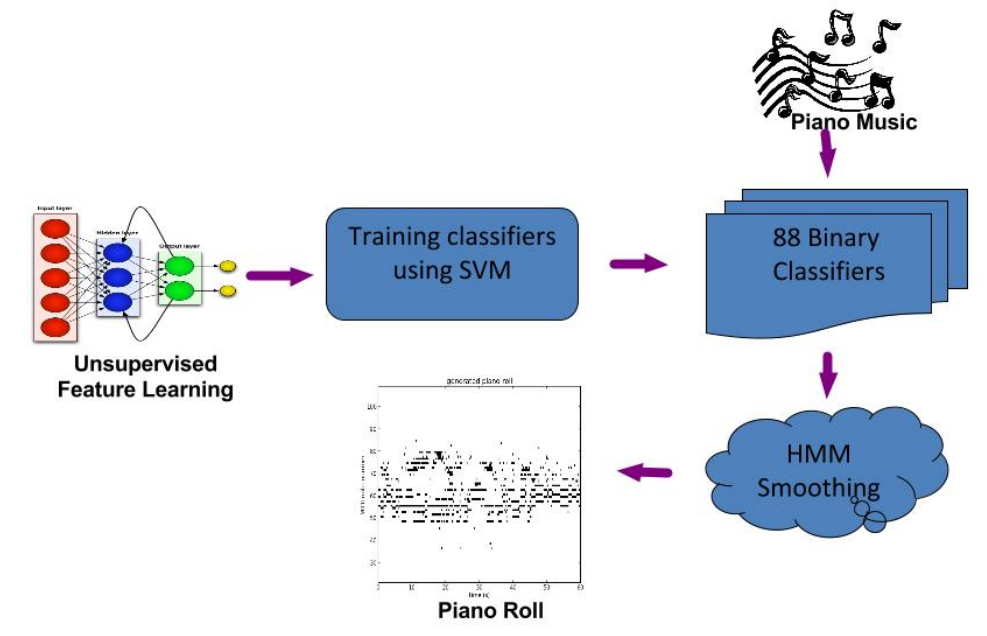


Рисунок 0.7 – Подход Zalani, Mittal (2014)

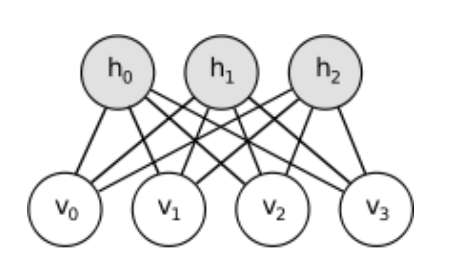


Рисунок 0.8 – Ограниченная машина Больцмана

представляет видимые блоки, – скрытые (см рисунок 0.8).

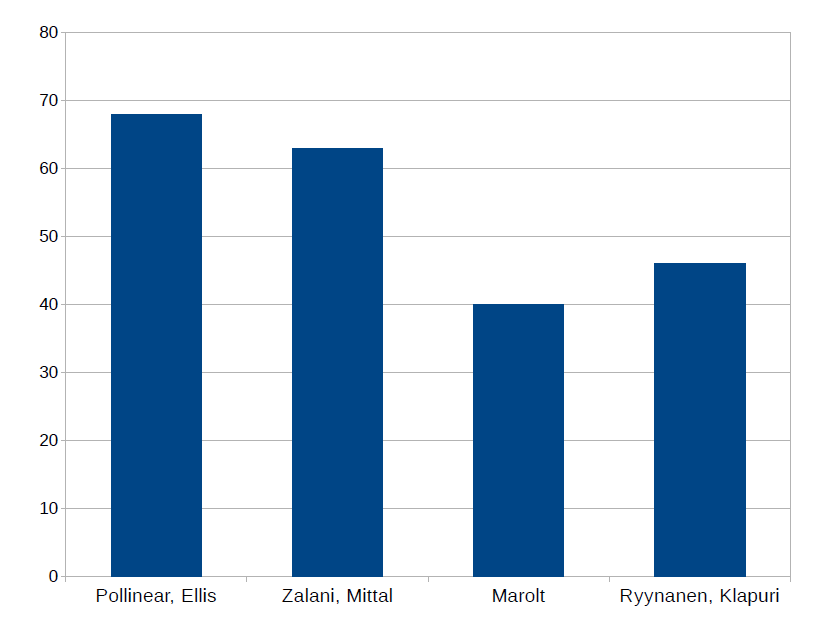


Рисунок 0.9 – Сравнение точности распознавания аналогов

## **0.3** Цель и задачи работы

*Цель работы*: состоит в разработке метода автоматической транскрипции, который по возможности учитывает опыт и недостатки аналогов для получения лучшего результата распознавания.

С учётом описанных выше недостатков существующих подходов, для достижения поставленной цели перед автором данной работы были поставлены следующие *задачи*:

1. Разработать метод для более точного выделения в спектре звука информации, с целью получения более аккуратного отображения фрагментов звукозаписи в пространство признаков.
2. Выделить новые наиболее информативные признаки, желательно небольшой размерности, например, использовать мел-кепстральные коэффициенты (MFCC) в качестве признаков.
3. Реализовать описанные алгоритмы в виде программного продукта, позволяющего распознавать последовательность нот в поданном на вход звуковом файле.

## Используемые методы

1. Непрерывное вейвлет-преобразование (CWT) или преобразование постоянного качества (constant-q) для преобразования звукозаписи в частотно-временное представление вместо STFT.
2. Получение векторов признаков, как признаки учитывать также MFCC.
3. RNN-RBM для обучения признаков.
4. 88 SVM-классификаторов One-vs-All.
5. Пост-процессинг скрытыми марковскими моделями.

Взаимодействие методов показано на рисунке:

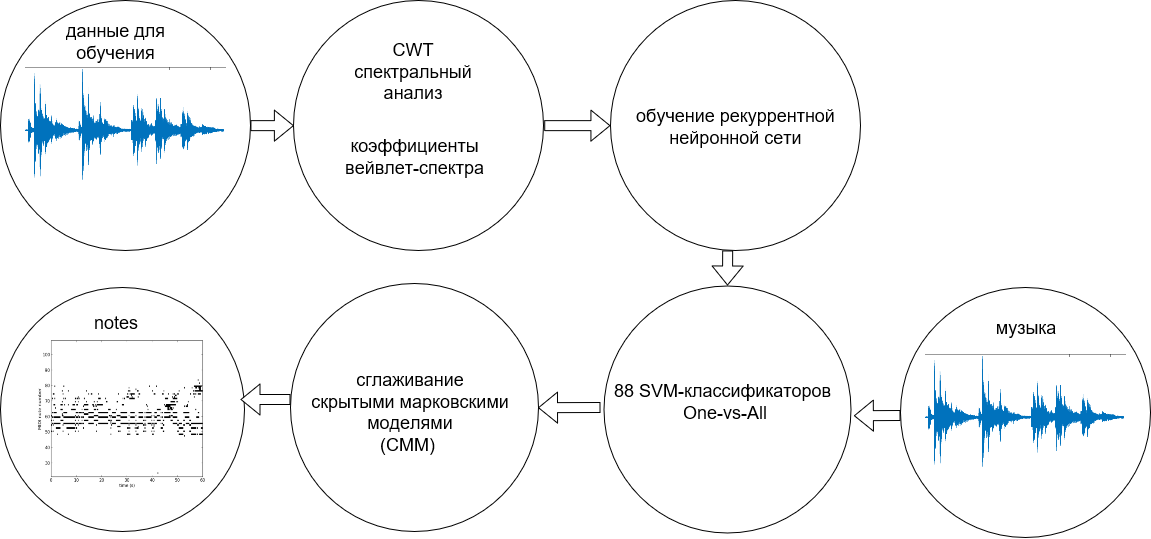


Рисунок 0.10 – Предложенный подход

# **1** Обзор литературы

## Звук

Звук – объективно существующее в природе физическое явление, вызываемое механическими колебаниями какого-либо упругого тела, вследствие чего образуются звуковые волны, воспринимаемые ухом и преобразуемые в нём в нервные импульсы. Звуковыми волнами называются периодически чередующиеся сгущения и разрежения в какой-либо упругой (т.е. звукопроводящей) среде; звуковые волны воспринимаются слуховыми органами человека и животных и при помощи центростремительных нервов передаются в большие полушария головного мозга, где и осознаются как конкретные звуки.

Все звуки вокруг нас распадаются на 2 типа: с определенный высотой (музыкальные звуки) и с неопределённой высотой (шумовые звуки). Музыкальные звуки составляют звуковой фонд музыки, в то время как шумовые звуки применяются лишь эпизодически. Музыкальный звук имеет 4 основных свойства: высота, длительность, громкость, тембр [6].

* 1. Свойства звука

*Высота*. Высота звука обусловлена частотой колебаний вибратора и находится от неё в прямой зависимости. Частота колебаний находится в обратной зависимости от величины (длины и толщины) звучащего тела и в прямой – от упругости.

Слух человека воспринимает звуки в диапазоне частот от 16 до 20000 Гц, в раннем детстве до 22000 Гц, в старости до 14000-15000Гц. Наиболее точно и ясно человек воспринимает звуки в пределах 16-4200-4500Гц, этот диапазон и используется в музыке. Зависимость между частотой колебаний и высотой звука – геометрическая прогрессия. При увеличении частоты на 110 Гц (это приблизительно соответствует укорачиванию струны в два раза) от A (110Гц) образуются интервалы: ч.8, ч.5, ч.4, б. 3, м.3, м.3, несколько б.2, несколько м.2. Дальше образуются интервалы меньше полутона. Этот звуковой ряд соответствует натуральному ряду чисел и называется натуральным звукорядом. Его можно получить при делении струны на 2, 3, 4, 5, 6 и т.д. частей, чем пользуются при исполнении на струнных инструментах флажолетов. Эталон высоты звука – 440 Гц (а первой октавы).

Акустическая единица измерения звуковысотных расстояний – цент = 1/100 темперированного полутона. Порог различения изменения высоты звука в среднем регистре – 5 центов.

*Длительность.* Длительность звука – выраженное в ритмических единицах время, в течение которого совершаются колебательные движения вибратора. Прямая зависимость. Длительность музыкального звука колеблется от 0,015-0,02 с до нескольких минут (педальные звуки органа). В тактовой нотации (с 17 в.) ноты указывают лишь относительную длительность звука, реальное значение которой зависит от темпа.

*Громкость*. Громкость звука – отражение в восприятии силы звука, обусловленной амплитудой колебаний. Применяемые в музыкальной практике обозначения динамических оттенков показывают не абсолютные значения громкости звука, а соотношения между их градациями.

Колебания бывают 2 видов: затухающие (т.е. с постепенно уменьшающейся за счёт сопротивления воздуха и внутреннего торможения амплитудой – рояль, арфа, струнно-щипковые) и незатухающие (с постоянной или произвольно меняющейся амплитудой – орган, скрипка при игре смычком). При затухающих колебаниях громкость звука постепенно уменьшается до полного затихания (высота остаётся практически неизменной). При незатухающих колебаниях громкость можно варьировать в зависимости от художественных целей.

Интенсивность (сила) звука – отношение падающей на поверхность звуковой мощности к площади этой поверхности, измеряется в Вт/м2. При росте силы звука в геометрической прогрессии громкость возрастает лишь в арифметической.

## Некоторые понятия из теории музыки

Человек воспринимает звуки с частотами и (до 5000 Гц) как очень похожие и тесно связанные друг с другом. Расстояние между такими звуками называется октавой. Даже некоторые животные – например, обезьяны и кошки, – воспринимают звуки, отличающиеся на октаву, как похожи.

Звукоряд делится на октавы на основе октавного сходства его звуков и отражающей это сходство повторности их названий. В свою очередь, каждая октава имеет своё название: субконтроктава, контроктава, большая октава, малая октава и октавы с первой по пятую. Началом октавы принято считать звук ступени до.

Темперированным называется строй, который делит каждую октаву звукоряда на равные части. С начала XVIII века в европейской музыке принята двенадцатизвуковая (двенадцатиступенная) темперация, делящая октаву на 12 равных друг другу частей, называемых полутонами.

Полутон является наименьшим расстоянием по высоте, возможным в двенадцатизвуковом темперированном строе. Он образуется между звуками любых двух соседних ступеней звукоряда.

Расстояние, образованное двумя полутонами, называется целым тоном. Расстояние между двумя соседними основными ступенями звукоряда (соответствующими белым клавишам фортепиано) может быть равно полутону (например, ми–фа) или целому тону (например, фа–соль).

Частоту каждой ступени звукоряда можно вычислить по формуле

= ·

где 𝑓0 – частота настройки музыкальных инструментов. В 1939 году на международной конференции в Лондоне был принят стандарт для частоты настройки 𝑓0 = 440 Гц. Эту частоту фиксируют для звука ля первой октавы.

Клавиатура фортепиано охватывает 88 ступеней: от ля субконтроктавы до ступени до пятой октавы. Частота, соответствующая 𝑘-й слева клавише фортепиано (отсчитывается с нуля), может быть вычислена по формуле

= 27.5 ·

Широко используемый в настоящее время стандарт MIDI (Musical Instrument Digital Interface), задающий формат обмена данными между электронными музыкальными инструментами, определяет 128 возможных значений для частоты звука. Частота, соответствующая ступени с номером 𝑘, 0 ≤ 𝑘 ≤ 127, может быть получена по формуле.

**

И наоборот, номер ступени может быть получен из частоты по формуле



Приведенные выше формулы справедливы для стандартного значения частоты настройки = 440 Гц. В рамках стандарта MIDI звук ля первой октавы соответствует 69-й ступени [7].

## Цифровой звук

Звуковой сигнал 𝑥(𝑡) может быть представлен в цифровом виде при помощи операций дискретизации и квантования. Для этого с некоторой частотой 𝜈 раз в секунду измеряется амплитуда функции 𝑥(𝑡) (дискретизация), после чего каждое полученное значение заменяется на ближайшее из заданного множества возможных значений амплитуды (квантование). Как правило, это множество содержит 28, 216 или 224 элементов, чтобы каждое значение можно было представить целым числом байт. Частота 𝜈 часто выбирается равной 44100 Гц (по историческим причинам). При этом 𝜈 называют частотой дискретизации, а значения – отсчётами исходного сигнала 𝑥(𝑡)). В соответствии с классической теоремой Котельникова, если спектр сигнала 𝑥(𝑡) ограничен сверху частотой 𝜈/2 (т.е. 𝑎𝑘 = 0 для > 𝜈/2), то исходный сигнал может быть восстановлен однозначно и без потерь по измеренным значениям . При квантовании эти значения заменяются на , поэтому исходный сигнал может быть восстановлен из оцифрованного только с некоторой ошибкой, которая тем меньше, чем больше возможных значений амплитуды использовалось при квантовании. Для большинства звукозаписей эта ошибка незаметна на слух. Отметим ещё раз, что спектр любых оцифрованных звуковых сигналов ограничен [7].

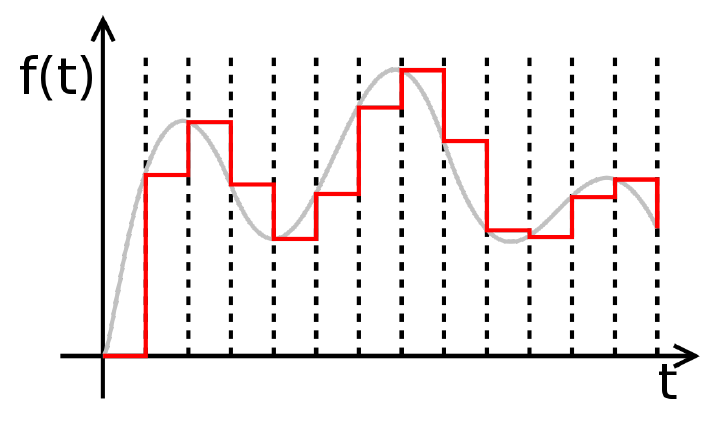


Рисунок 1.1 – Дискретизация и квантование

## Свойства музыкальных звукозаписей

Формат файла определяет структуру и особенности представления звуковых данных при хранении на компьютере. Для устранения избыточности данных используются аудиокодеки, при помощи которых производится сжатие аудиоданных. Выделяют три группы звуковых форматов аудиофайлов:

* форматы без сжатия (WAV, AIFF);
* форматы со сжатием без потерь (APE, FLAC);
* форматы с применением сжатия с потерями (mp3, ogg);

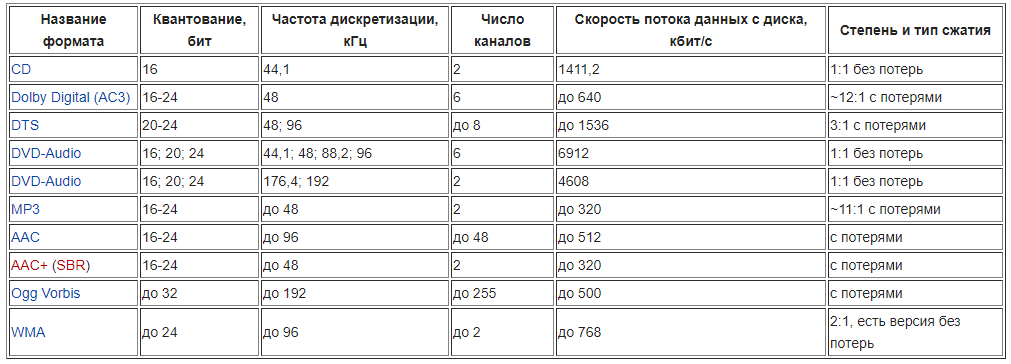


Рисунок 1.2 – Некоторые виды цифровых аудиформатов в сравнении

Музыкальные звукозаписи в целом обладают рядом свойств, которые нужно учитывать при транскрибировании:

1. Наличие гармоник у музыкальных инструментов с ясно выраженной высотой звучания. В звучании таких инструментов можно выделить отдельную ноту. При этом наряду с частотой, соответствующей этой основной ноте, звучат другие частоты. Их звучание менее выражено, но они могут соответствовать другим ступеням музыкальной системы. Математически это означает, что если таково, что – наибольшая по абсолютному значению компонента спектра звучащей ноты, то существует по меньшей мере одно значение 𝑘 > такое, что существенно отлична от 0. Соотношения между парами (, 𝑘) для разных 𝑘 и разных (соответствующих разным нотам) во многом задают тембр музыкального инструмента.
2. Наличие периодичностей (повторений). Одна и та же музыкальная фраза, последовательность аккордов и даже целый фрагмент композиции могут повториться в точности или с небольшими изменениями [7].

## Формализация задачи

Теперь, с использованием введённых понятий и обозначений, можно дать формальную постановку основной задачи, решаемой в данной работе: пусть заданы звуковой сигнал 𝑥(𝑡), 𝑡 ∈ [, ] и множество возможных названий аккордов 𝑌 . Необходимо для каждого момента времени 𝑡 ∈ [, ] указать аккорд 𝑦 ∈ 𝑌 , звучащий в этот момент. Эту общую задачу можно разделить на отдельные шаги, которые соответствуют конкретным задачам, перечисленным во введении [7].

### **1.6.1** Частотно-временное представление

Представление звука в виде последовательности отсчётов амплитуды не является удобным для обработки: неясно, как сопоставить аккорду последовательность отсчётов и наоборот. Поэтому естественным первым шагом является часто используемый при обработке звука переход к частотно-временному представлению звукозаписи или получение её спектрограммы 𝐶𝑁×𝑀. Она представляет из себя матрицу, каждая из 𝑁 строк которой соответствует определённой частоте, а каждый из *M* столбцов представляет из себя спектр фрагмента звукозаписи в пределах короткого промежутка. Элементами этой матрицы являются значения интенсивности данной частоты в пределах данного промежутка времени.

Частотно-временное представление может быть получено с помощью разных техник Short-time Fourier transform (STFT), Continuous wavelet transform (CWT) или Constant-Q transform (CQT).

### **1.6.2** Классификация

Имеем множество всех возможных векторов-столбцов спектрограммы. Для каждого вектора из данной последовательности необходимо указать множество нот, соответствующий этому вектору. Принимая во внимание известные из равномерно темперированного строя частоты нот, можно по спектру звука в данном фрагменте делать предположения относительно звучащих нот и аккорда. Возможно использование не только текущего, но и предыдущих векторов (а также последующих, если от алгоритма не требуется выдавать результат в реальном времени). Также возможно использование результатов классификации других векторов. На этом этапе главными вопросами являются выбор метода классификации и отыскание такого набора преобразований множества, который позволит упростить классификацию с использованием выбранного метода. Результатом преобразования будет последовательность векторов признаков, отличная от исходной последовательности векторов-столбцов спектрограммы.

## Предварительная обработка

Предварительная обработка может содержать следующие шаги:

* преобразование стерофонических записей в монофонические;
* возможно понижение частоты дискретизации с 44100 Гц до 11025 Гц, при этом теряются частоты выше 5.5 кГц, что не является критичным, так как в музыке большая часть информации находится на частотах до 5 кГц;
* преобразование сигнала из временного представления в частотно-временное;

## Спектрограмма

Переход к частотно-временному представлению звукозаписи в виде спектрограммы является ключевым, поскольку даёт возможность работать с отдельными частотными компонентами звука. Для этого звукозапись делится на короткие, возможно, пересекающиеся фрагменты, на каждом из которых вычисляется спектр звука.

Могут быть применены следующие алгоритмы преобразования звукозаписи к частотно-временному представлению:

STFT (Short-time Fourier transform), получают спектрограмму звукозаписи. Это позволяет работать с отдельными частотными компонентами звука. Ось x спектрограммы – время, y – частота, цвет – интенсивность определенной частоты в данный момент времени. Для получения спектрограммы сигнал разделяется на временные сегменты с перекрытием, выполняется на каждом сегменте FFT и далее совмещаются все FFT сегментов на одном графике. Недостаток STFT в сложности нахождения компромисса между временным и частотным разрешениями. Вместе с тем, при использовании алгоритмов быстрого преобразования Фурье невозможно произвольным образом выбирать частоты его компонент. Это создает неудобства при дальнейшей обработке, поскольку невозможно точно определить количество звуковой энергии, приходящейся на частоты, соответствующие ступеням звукоряда. При фиксированном размере окна FFT, информация, используемая для расчета FFT ограничена. Если выбрать большое окно, то получим очень хорошее частотное разрешение, но это приведет к перекрытию нот. С другой стороны, выбирая малое окно, достигаем хорошего разделения нот, но частотное представление нот хуже, что, возможно, уменьшит способность разделять ноты.

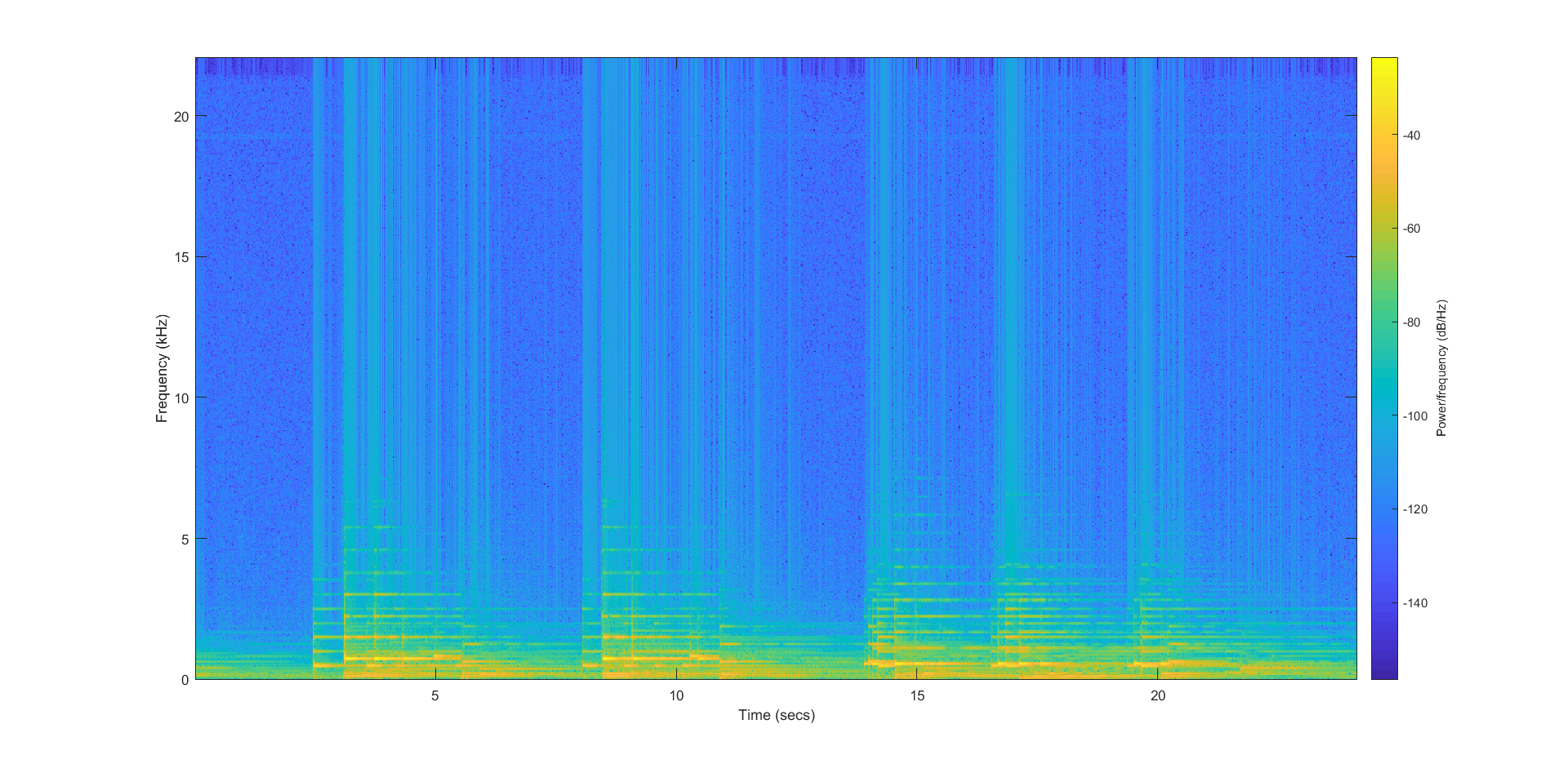


Рисунок 2.1 – STFT участка звукозаписи

В преобразовании постоянного качества (constant-Q преобразование) в отличие от преобразования Фурье, размер анализируемого фрагмента и размер оконной функции зависят от номера соответствующей частотной компоненты. Достоинством этого преобразования является легкость дальнейшей работы со спектром, поскольку его компоненты напрямую соответствуют ступеням звукоряда. Недостатками являются большая сложность вычислений и зависимость от правильного определения частоты настройки.

Гребёнка фильтров (filter bank). В цифровой обработке сигналов любое преобразование сигнала называют фильтром. В данном случае под фильтром понимается полосовой фильтр – преобразование, сохраняющее в звуке только частоты, находящиеся в некоторой полосе частот. Быстрое преобразование Фурье эквивалентно вполне определенной гребёнке достаточно грубых фильтров. Вместо них можно использовать любые другие фильтры, у каждого из которых центр полосы пропускания соответствует частоте одной из ступеней звукоряда, а ширина полосы пропускания достаточно мала, чтобы не охватывать частоты соседних ступеней. Эти фильтры можно подобрать так, чтобы они были менее грубыми, то есть более точно определяли количество звуковой энергии, приходящейся на их полосы пропускания. Недостатком данного метода является большая вычислительная сложность в сравнении с алгоритмом быстрого преобразования Фурье.

Вейвлет-преобразование Вейвлет-преобразование (WT) – еще одна техника для перевода сигнала из временного представления в частотно-временное. WT является альтернативой STFT и позволяет преодолеть проблемы связанные с неопределенностью частотного и временного разрешений. STFT представляет одинаковое временное разрешение для всех частот, в то время как CWT предоставляет высокое временное разрешение и низкое частотное для высоких частот и высокое частотное разрешение и низкое временное для низких частот. Человеческое ухо демонстрирует аналогичные частотно-временные характеристики разрешения.

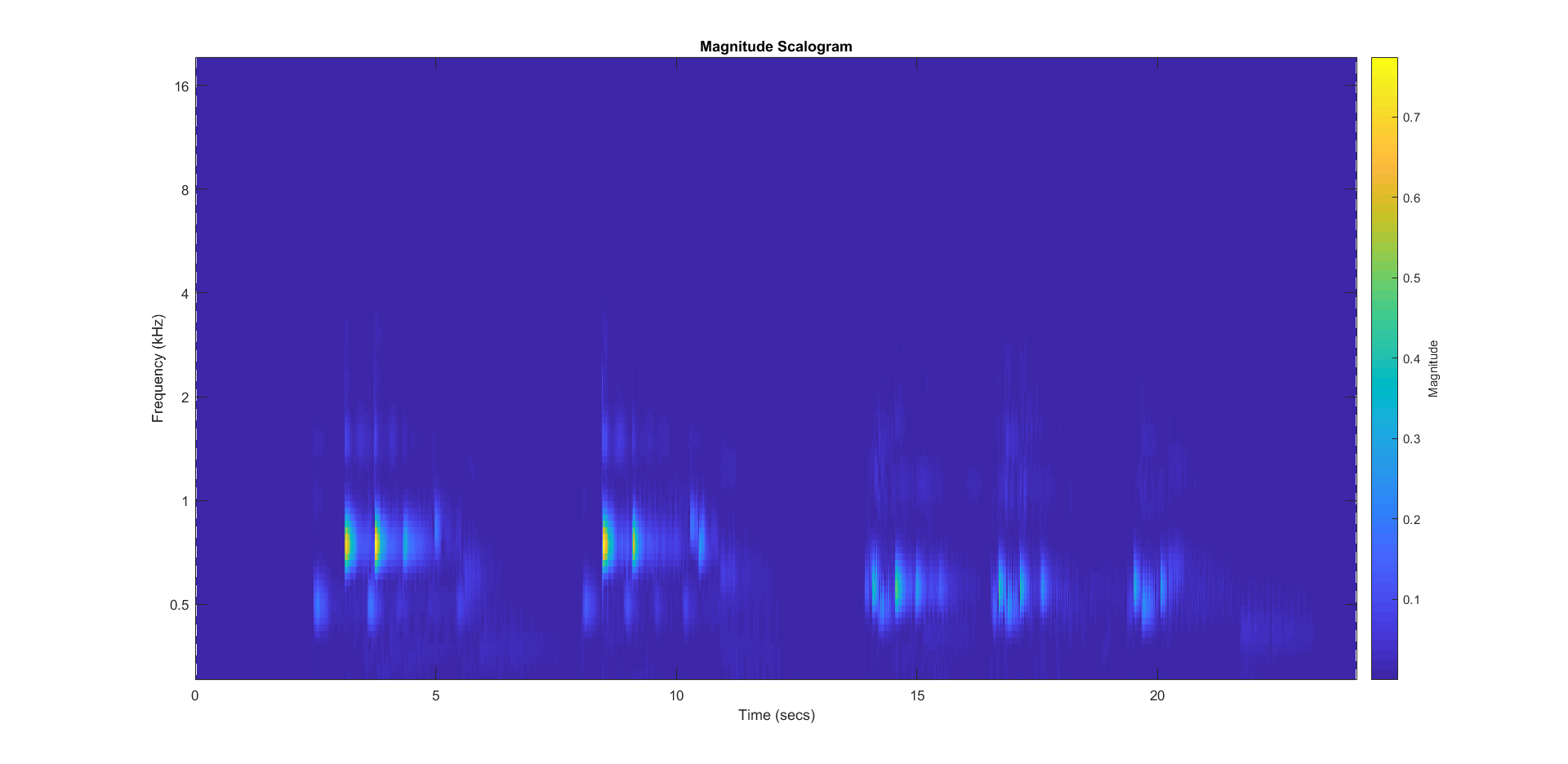


Рисунок 2.2 – CWT участка звукозаписи

## Векторы признаков

В качестве признаков могут быть использованы:

* коэффициенты MFCC;
* коэффициенты вейвлет-преобразования;
* столбцы спектрограммы;
* хроматические вектора - векторы признаков, полученные путём объединения спектральной информации по всем октавам, каждый хроматический вектор имеет всего 12 компонент;

# СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

[1] E. Poliner, P. W. Ellis — A Discriminative Model for Polyphonic

Piano Transcription, 2006.

[2] Нижибицкий Е. — Музыкальная транскрипция при помощи методов машинного обучения, 2014.

[3] M. Ryyn ̈anen, A. Klapuri — Polyphonic Music Transcription Using Note Event Modeling for MIREX 2008.

[4] M. Marolt — A Connectionist Approach to Automatic Transcription of Polyphonic Piano Music, 2004.

[5] A. Zalani, A. Mittal — Polyphonic Music Transcription: A Deep Learning Approach, 2014.

[6] Звук и его свойства [Электронный ресурс]. – Электронные данные. – Режим доступа: http://theory.solfa.ru/20-вопросов-по-теории-музыки-классическ/1-звук-и-его-свойства/.

[7] Глазырин Н. — Алгоритмическое распознавание аккордов в цифровом звуке, 2014.