

Нижегородский государственный технический
университет им. Р.Е. Алексеева

Орлов Артем Сергеевич

МЕТОДЫ И АЛГОРИТМЫ
РАСПОЗНАВАНИЯ МУЗЫКАЛЬНЫХ
ПРОИЗВЕДЕНИЙ

Научный руководитель:

кандидат технических наук, доцент
Гай Василий Евгеньевич

Нижний Новгород - 2016

Оглавление

Введение	3
Глава 1. Обзор известных методов и средств.....	9
Глава 2. Исследование и построение решения	16
2.1 Теория, выбранная для решения задачи	16
2.2 Исследование проблемной области поставленной задачи	20
2.2.1 Психоакустические и когнитивные исследования восприятия звука	20
2.2.2 Спектральные особенности музыкальных звуков	23
2.3 Построение необходимых математических моделей.....	32
2.4 Метод распознавания обертонового портрета	33
2.5 Методы компенсации атаки, неравномерности проигрывания и приведения полученной модели к классическому виду	42
2.6 Разработка программной системы	44
Глава 3. Экспериментальная часть	49
3.1 Планирование вычислительного эксперимента	49
3.2 Проведение вычислительного эксперимента и анализ его результатов	55
3.3 Сравнение полученных результатов с известными алгоритмами и методами	59
Заключение.....	61
Список литературы	63

ВВЕДЕНИЕ

В настоящее время одним из динамично развивающихся направлений в области искусственного интеллекта является область, связанная с задачей распознавания музыки (automatic music transcription), которая понимается как перевод музыкального звукового сигнала в символьное представление. Таким символьным представлением может быть как запись музыкального произведения в классической нотной нотации или же запись в формате MIDI [3].

Перспективность разработок в этой области подтверждается тем, что в данное время неизвестны системы, решающие данную проблему на уровне, который был бы аналогичен способности человека распознавать музыку [8,9].

В качестве практического применения технологии распознавания музыки, позволяющей транскрибировать запись музыкального звукового сигнала в символьное представление, необходимо отметить помощь музыкантам в создании нотной записи импровизированных выступлений с целью последующего воспроизведения, использования и распространения, а также анализ записей произведений традиционной народной музыки различных этносов (к примеру, полученной в ходе полевых исследований) для последующих культурологических исследований[10].

Отмечается, что подобная технология может иметь применение в поиске произведения по его фрагменту в специализированных музыкальных базах данных, а также в задачах выявления плагиата.[10]

Если рассмотреть подробнее данные области возможного применения, то становится возможным привести следующую подробную классификацию задач, в которых могут быть востребованы технологии распознавания музыкальных произведений:

- 1) осуществление поиска необходимого произведения по его фрагменту – в качестве такого фрагмента может выступать как

запись части произведения, полученное интересующимся в поиске лицом с помощью того или иного устройства звукозаписи, так и самостоятельное воспроизведение этого фрагмента по памяти. Программные системы, реализующие автоматическое распознавание музыки в таком случае могут использоваться для автоматического преобразования музыкальных произведений в сжатое символьное описание с последующим помещением в базу данных и последующим поиском в массиве этих произведений с помощью тех или иных поисковых алгоритмов. Преобразование в символьный вид, по сравнению с использованием исходных аудиопотоков даёт преимущество как в объёме необходимой памяти, так и в уменьшении вычислительных затрат на поиск в этой базе данных благодаря уменьшению её объёма.

- 2) поиск плагиата в музыкальной индустрии – аналогично предыдущему случаю, требуемые музыкальные композиции, которым может потребоваться защита от плагиата с целью сохранения авторских прав, распознаются автоматической программной системой и полученная в результате распознавания символьное представление, дающее сведения о структуре музыкального произведения, помещается в требуемую базу данных. Как и в предыдущем случае, исследуемое на плагиат произведение с помощью технологии автоматического распознавания музыки также преобразуется в символьный вид и с помощью специализированных поисковых алгоритмов служит образцом для осуществления поиска среди корпуса защищённых авторским правом музыкальных композиций.
- 3) проведение культурологических исследований – кроме случаев поиска целого произведения фрагмента также возможен и этот вариант использования: поскольку музыкальная база данных может хранить крайне большое количество музыкальных произведений с

дополнительными данными об этих произведениях (жанр, год выпуска, страна выпуска и т.д.), может возникнуть интерес, заключающийся в попытке тем или иным образом оценить существующие культурологические тенденции как в музыкальной индустрии, так и вообще в мировой музыке, классифицировать эти тенденции по эпохам, странам и провести иные действия.

- 4) использование в системах обучения начинающих музыкантов игре на различных музыкальных инструментах как в специализированных учебных заведениях, так и в частном порядке – как один из возможных вариантов использования, применение в качестве специализированного пользовательского ПО.

Таким образом, необходимость разработки программных систем, реализующих распознавание музыкальных произведений подтверждается как с точки зрения практического применения, так и с точки зрения сложности данной задачи и отсутствие общепринятых методов её решения.

В различных публикациях описаны разнообразные методы решения задачи распознавания музыкальных произведений, однако большинство исследователей сходится во мнении, что данная тематика не имеет удовлетворительного решения и поэтому исследования в этой области являются оправданными и перспективными.

Так, для обмена информацией между специалистами, работающими в различных областях обработки музыки с помощью вычислительных технологий в 1999-м году была создана организация International Society for Music Information Retrieval (ISMIR) – Международное общество извлечения музыкальной информации, направление работы которой посвящено проблемам компьютерной обработки музыки. Данная организация была призвана объединить всех исследователей, работающих в области компьютеризации музыки.

За прошедшие с её создания полтора десятилетия конференции, проводимые данной организацией, получили широкое признание во всём

мире.[19]. Это говорит о широкой заинтересованности в данных технологиях и идущем прогрессе в этой области.

Цель работы и задачи исследования

Целью исследования данной работы является разработка информационной модели системы распознавания музыкальных произведений с использованием положений теории активного восприятия (ТАВ) В.А. Утробина с последующей реализацией разработанных алгоритмов в программной системе, а также выполнение исследования эффективности разработанных алгоритмов и проведение сравнительного анализа полученных результатов с другими результатами и методами, полученными ранее. При этом распознавание музыкального произведения понимается как преобразование записи звукового сигнала в символьное представление, содержащее сведения о проигрываемых нотах и инструментах, которые это проигрывание осуществляют.

В качестве задач исследования выступают:

- 1) Изучение уже разработанных алгоритмов и методов, применяемых в данной проблемной области
- 2) Построение информационной модели системы, осуществляющей распознавание музыкальных произведений
- 3) Разработка и программная реализация разработанных алгоритмов в программной системе
- 4) Экспериментальное исследование полученной системы

Объект исследования

В качестве объекта исследования данной научной работы выступает сигнал, содержащий запись музыкального произведения.

Предмет исследования

Предметом исследования являются методы и алгоритмы распознавания структуры музыкального произведения.

Методы исследования

В качестве методов исследования использовались теория активного восприятия, аналитический метод и вычислительный эксперимент.

Положения, выносимые на защиту

На защиту выносятся следующие положения:

- информационная система распознавания музыкальных произведений с использованием теории активного восприятия.
- результаты вычислительного эксперимента

Научная новизна

Научная новизна работы состоит в разработке информационной модели системы распознавания музыкальных произведений на основе теории активного восприятия, а также разработанный на основе данной модели алгоритм.

Практическая ценность работы

Практическая ценность работы состоит в разработке программной системы, реализующей распознавание музыкальных произведений, которая может быть использована в различных практических приложениях, таких как создание специализированных музыкальных баз данных, содержащие сведения о различных музыкальных произведениях.

Реализация результатов работы

По результатам проведенных исследований были опубликованы тезисы на XXII международной научно-технической конференции «Информационные системы и технологии» ИСТ-2016 [25].

Апробация работы

Апробация этой работы была проведена в рамках публичного выступления на XXII международной научно-технической конференции.

Обоснованность и достоверность

Достоверность и обоснованность данной работы подтверждается широким тестированием разработанной программной системы в различных условиях использования.

Публикации

В процессе выполнения данной работы были опубликованы тезисы на XXI международной научно-технической конференции «Информационные системы и технологии» ИСТ-2015 [24], тезисы на XXII международной научно-технической конференции «Информационные системы и технологии» ИСТ-2016 [25].

Объем и структура работы. Работа состоит введения, трёх глав, заключения. Полный объём составляет 65 страниц, с 13 рисунками и 6 таблицами. Список литературы содержит 25 наименований.

ГЛАВА 1. ОБЗОР ИЗВЕСТНЫХ МЕТОДОВ И СРЕДСТВ

В настоящее время разработано множество техник для решения задачи автоматического распознавания музыки [8,9,10,11, и др.], которые описаны в различных научных публикациях.

Транскрибирование монофонической музыки считается наиболее легким вариантом общей задачи. В работе [12] для распознавания монофонической мелодии используется три параметра: высота основного тона, время начала ноты и длина её продолжительности. Для выявления нот в звуковом сигнале используется техника, названная авторами автокорреляционным отслеживанием основного тона (autocorrelation pitch tracking). Этот подход основан на использовании автокорреляционной функции, которая выявляет во входном сигнале скрытые периодичности. Автокорреляционная функция предназначена для определения корреляций отдельно взятого отрезка сигнала с последующими отрезками, благодаря чему и получила своё название, подчеркивающее её способность к обнаружению корреляций. Для каждой «периодичности» в исследуемом сигнале автокорреляционная функция демонстрирует отклик. Такой отклик является следствием существования в звуке ноты периодического изменения амплитуды. Таким образом, пик с наибольшим значением считается отображением частоты основного тона.

Поскольку музыкальное произведение, кроме идеальных случаев, содержит более чем одной ноты, то для выделения их границ используется сегментация сигнала на отдельные отрезки.

В качестве достоинств этого подхода называются простота и скорость работы реализованного алгоритма при хороших показываемых результатах.

Также в работе [12] описаны подходы к распознаванию простых полифонических композиций, состоящих из аккордов. Для решения этой задачи авторы используют преобразование входного сигнала в спектрограмму, которая в дальнейшем обрабатывается с помощью нейронной сети.

Работа[11] посвящена автоматическому транскрибированию музыки с использованием модели событий нот. В данном случае каждая нота представляется как событие, которому соответствует определённая скрытая марковская модель. В качестве основы авторы предлагают введение так называемой скрытой марковской модели тишины (silence model), с которой в дальнейшем происходит сравнение других скрытых марковских моделей, которую соответствуют тем или иным событиям ноты. Скрытая марковская модель – это такая марковская модель, в которой суждения о скрытых состояниях делаются по наблюдаемым следствиям этих состояний.

Для распознавания пианинной музыки авторы работы [13] также применяют скрытые марковские модели. Данный метод был опробован на сонатах Моцарта и показал результат в 39% ошибок – вставок и удалений нот по сравнению с оригинальной композицией.

В работе[14] для автоматического распознавания музыки используется подход, представляющий собой построение инвариантной к сдвигам модели скрытых переменных. Авторы этой работы основываются на теории вероятностного анализа скрытых компонентов (probabilistic latent component analysis, PLCA). Подобный анализ применялся другими авторами для задач распознавания музыкальных произведений.

Суть данного метода анализа состоит в том, чтобы аппроксимировать входные спектрограммы вероятностным распределением, зависящим от времени и частоты, и представить это распределение как разложение на сумму отдельных спектральных компонентов.

Авторы работы[14] ставят своей целью расширить возможности этого подхода в задачах распознавания музыки, для чего предлагают модель, поддерживающую множественные спектральные шаблоны, что позволяет достичь возможности распознавания отдельных музыкальных инструментов.

В конечном итоге, разработанная авторами система продемонстрировала хорошие результаты: так, по метрике корректности распознавания она превзошла аналогичные системы других авторов, основанные на других

подходах. Тем не менее, авторы признают несовершенство своей системы и в качестве следующих шагов называют цель для уменьшения потери нот сконцентрироваться на обработке амплитудных модуляций между различными звуковыми событиями, часто встречающихся в звуковых сигналах.

Авторы работы [15] также констатируют сложность задачи автоматического распознавания полифонической музыки, поэтому, как и многие другие авторы, фокусируются на распознавании полифонической музыки всего одного инструмента – пианино, и используют подход, названный ими коннекционистским. Коннекционизм в названии этого подхода означает, что модель музыкальной композиции представляется как набор простых элементов, связанных между собой сетевыми связями. Как и во многих других подходах, подход, использованный в работе [15] также в своей основе имеет переход от осциллограммы звукового сигнала, содержащего музыкальную композицию, к спектральному частотно-временному представлению с помощью модели слухового восприятия (auditory model), полученной на основе бионических исследований слухового восприятия у человека. Данная модель предполагает фильтрацию и частотное преобразование входного звукового сигнала. В дальнейшем происходит разложение. В качестве простых элементов, из которых складывается вся музыкальная композиция, выступают отдельные адаптивные осцилляторы (adaptive oscillators), которые представляют собой модели гармонических синусоидальных сигналов, которые подстраиваются (адаптируются) под входящий сигнал с помощью нейросетевых методов.

Таким образом, предлагаемая авторами система состоит из двух частей: подсистемы слухового восприятия, которая моделирует работу человеческого уха, выполняя спектральное преобразование осциллограммы сигнала в частотно-временное представление и нейросети адаптивных осцилляторов, которая выделяет на полученной на первом этапе спектрограмме пики, соответствующие частотам основного тона, а также

пики, соответствующие обертонам. С помощью адаптивных осцилляторов осуществляется попытка выявления в полученной спектрограмме периодических сигналов, которые соответствуют проигрываемым музыкальным событиям.

Каждый осциллятор обладает такими параметрами, как период колебаний (соответствующий частоте) и сдвиг фазы. На вход осциллятору с первого этапа обработки сигнала подаётся серия дискретных импульсов. После каждого цикла осцилляции осциллятор подстраивает свои параметры, стараясь подстроиться под входящий поток событий. Если входящий поток событий периодичен, то конечным эффектом будет подстройка осцилляторов под входящие события.

Предлагаемая авторами система состоит из 88 сетей осцилляторов по числу 88 нот, которых может проигрывать пианино (ноты в диапазоне A0 – C8). Начальная частота каждого осциллятора соответствует высоте каждой ноты, для отслеживания которой он предназначен.

Также необходимо отметить существование на рынке прикладных приложений программной системы WIDISoft разработки группы российских разработчиков. Данная система является автоматизированной средой распознавания, то есть в ней для успешного функционирования необходимо ручное задание параметров распознавания для различных музыкальных композиций. Авторами данной системы была опубликована пара работ, посвященные функционированию разработанного ими алгоритма [16,17].

Работа [16] посвящена выявлению во входящем звуковом сигнале начала нот (онсетов). Как и подходы многих других авторов, описанный в данной работе алгоритм основан на использовании сонограммы звуковой записи, которая в дальнейшем исследуется с помощью нейросетевых методов. Для решения своей задачи авторы предлагают использование двухуровневой нейронной сети, первый уровень которой состоит из отдельных нейросетей, определяющих появление онсетов на каждой из полос частот и обучаемых на эталонных произведениях, для которых вручную

описаны начало каждой ноты. На второй уровень передаются результаты работы всех нейросетей первого уровня. Нейросеть второго уровня предназначена для окончательного принятия решения о появлении в том или ином месте произведения той или иной ноты. Благодаря двухуровневой структуре верхний уровень позволяет корректировать результаты работы нейросетей нижнего уровня и таким образом получать более качественные результаты распознавания. Нейросети обоих уровней обучаются методом обратного распространения ошибки по эталонным записям. В качестве достоинств изложенного двухуровневого подхода отмечается возможность бороться с эффектом биений, при котором возможно периодическое пропадание основного тона звучащей ноты.

Работа [17] посвящена определению темпа распознаваемого музыкального произведения. Для решения этой задачи автор предлагает метод конкурирующих гипотез. Сущность этого алгоритма заключается в последовательном анализе ряда, содержащего соответствующие началам нот отсчёты времени. В начале работы алгоритма строятся локальные гипотезы, содержащие время начала нот, период, с которым ноты следуют, а также параметр, называемый штрафом. Данный параметр необходим для того, чтобы в последующем отбирать из всего множества сгенерированных гипотез такие, которые наилучшим образом соответствуют темпу на том или ином отрезке музыкального произведения. В дальнейшем алгоритм на основе первоначально полученных гипотез строит следующее поколение дочерних гипотез, отбирая из первоначальной популяции те гипотезы, которые обладают наименьшим штрафом.

Таким образом осуществляется построение цепочки гипотез с помощью механизма штрафов. Штрафы каждой из гипотез назначаются по принципу наименьшей разницы между тем положением ноты, которое предсказывается гипотезой, и реальным расположением этой ноты: чем меньше расстояние между предсказанием и действительным положением, тем меньше штраф. В качестве критерия верности гипотезы автор предъявляет требование, чтобы

временной промежуток между предсказанием гипотезы и действительным расположением ноты был меньше достаточно малого порога различения одновременного звучания, то есть такого порога, при котором две ноты, чье начало начинается со сдвигом меньше этого значения, слышатся одновременно звучащими. В качестве такого порога предлагается величина в 20 мс. Штрафы гипотезе назначаются в том случае, если расхождение между предсказанием и действительным положением ноты больше порога различения. Также возможна ситуация, при которой гипотеза предсказывает не следующую ноту, а ноту, следующую за ней. В таком случае за пропуск ноты назначается дополнительная величина штрафа.

Таким образом, в механизме назначения штрафа присутствует всего четыре возможных компонента: штраф за несовпадение, штраф за изменение периода, штраф за пропуск ноты, а также штраф, унаследованный от родительской гипотезы.

Отмечается, что представленный алгоритм демонстрирует хорошую устойчивость в случае изменения темпа и в случае сложных ритмических картин.

Таким образом, всё возможное количество используемых методов можно классифицировать по признаку принадлежности к нейросетевым и непринадлежности к ним. Нейросети широко используются в настоящее время в различных областях распознавания образов, поскольку зарекомендовали себя как инструмент, пригодный для осуществления задач распознавания требуемых образов по массиву предъявляемых им образцов. Поскольку реальные музыкальные сигналы могут быть, во-первых, крайне зашумлены, а, во-вторых, подвергаться различным влияниям вариаций проигрывания, то использование нейросетевых методов для их распознавания выглядит приемлемым методом для их выявления с целью последующего символического преобразования. В этом случае происходит обучение используемой нейросети на различных обучающих выборках.

Для улучшения качества распознавания нейросетевыми методами, как следует из приведенного обзора, различные авторы применяют методы, варьирующие как по типу входным данным, так и по методу получения этих данных для обработки с помощью нейросети.

Подводя итог для вышеприведённого обзора литературы, можно привести следующую классификацию рассмотренных методов используемых в системах распознавания:

1. Нейросетевые
 - а. одноуровневые
 - б. многоуровневые
2. Классифицирующие
 - а. Скрытые марковские модели
 - б. Вероятностный анализ скрытых компонентов (PLCA)
3. Комбинированные
 - а. Адаптивные осцилляторы и нейросети

Все эти методы имеют те или иные недостатки, существование которых является подтверждением как сложности проблемы, так и констатацией того факта, что данная область является перспективной с точки зрения новых исследований.

ГЛАВА 2. ИССЛЕДОВАНИЕ И ПОСТРОЕНИЕ РЕШЕНИЯ

2.1 Теория, выбранная для решения задачи

В качестве теории, выбранной для решения поставленной задачи, используется теория активного восприятия (ТАВ) В.А.Утробина. [1,2,3]

Данная теория фокусируется на проблематике раскрытия априорной неопределенности изображения как объекта исследования. Математической базой этой теории является алгебра изображения.[1,2]

В ТАВ постулируется, что основная задача распознавания (в «узком» смысле) заключается в построении правила, позволяющего отнести наблюдаемый объект, заданный на системе признаков (параметров), к одному из классов, удовлетворяющих требованиям сходства максимумам однородности внутриклассовых объектов при максимуме различия на межклассовых объектах. Любая система восприятия, по определению, решает задачу отражения предмета - объекта окружающего мира, в совокупность его свойств с учетом его объективной целостности. Результатом отражения является перцептивный образ, а сам процесс отражения – результат взаимодействия замкнутой системы: окружающий мир и наблюдатель, погруженный в этот мир и неискажающий его. [3]

Объект исследования в данной теории определяется как функция наблюдателя (в случае наблюдателя-человека - возникающая на сетчатке глазного яблока). Постулируется, что данная функция удовлетворяет следующим естественным ограничениям:

1) в каждый фиксированный момент времени t каждый элемент множества изображения M есть неотрицательная действительная функция действительных аргументов $M_i = m(x,y)$ если (x,y) принадлежат области G и $M_i = 0$, если (x,y) не принадлежат этой области, при этом G является замкнутой двумерной области евклидова пространства. Данная функция является суммируемой и квадратично интегрируемой на этом множестве (что является свойством измеримости), которое наделено свойствами

упорядоченности, структурированности и дискретности в пространстве и времени.[3]

Таким образом, постулируется, что изображение – это двумерный сигнал, представленный в фиксированный момент времени функцией $\mu(x,y)$ либо множеством точек-пикселей, в котором каждая точка координатно упорядочена и ей соответствует значение этой функции.[2]

Из свойства измеримости и наблюдаемости следует конечность множества M , которое, таким образом, имеет мощность $k_{\max}=P^{N \times M}$, где $N \times M$ – области G в силу свойства пространственной дискретности, а P – число градаций яркости. Таким образом, всё множество возможных изображений представляется конечным.[2,3]

Благодаря конечности объекта исследования и его принадлежности к множеству положительных действительных чисел, множество его точек сходится в среднем квадратичном, то есть всё изображение в целом и любая его подобласть определения допускают преобразование проектирования.

$$m(G_i) = \iint_{(x,y) \in G_i} \mu(x,y) dx dy \quad (1)$$

Преобразование (1) определено как Q -преобразование в силу следующих свойств:

1. Преобразование применимо к любому объекту исследования любой размерности и в любом диапазоне частот, удовлетворяющим относительно слабым ограничениям определения на изображение
2. Оно реализует отображение в действительное абсолютно гладкое пространство.
3. Данное преобразование фундаментально в силу фундаментальности последовательности Коши, существующей в M , достаточно и реализуемо.

Q -преобразование реализует отображение изображения, как объекта исследования, в абсолютно (бесконечно) гладкое многообразие [3]

Так как (1) – преобразование интегрирования, то существует обратное ему преобразование внешнего дифференцирования. Вместе они образуют композицию $U = d \circ \int$, раскрывающую неопределенность объекта исследования – изображения M в поле зрения G .

Данная композиция была названа U-преобразованием. [2,3]

Теория активного восприятия для обработки изображения предполагает использование алгебры изображения. Данная алгебра строится благодаря постановке в соответствии каждому фильтру U-преобразования соответствующего оператора, который имеет ту же конструкцию, что и фильтр, однако элементы его матрицы соответственно заменены на $+1 \rightarrow 1$ и $-1 \rightarrow 0$. Иными словами, операторы образуются путём принятия для множества фильтров отображения $(+1 \rightarrow 1)$ и $(-1 \rightarrow 0)$. [3]

Благодаря этому отображению на основе данных операторов, соответствующих вышеописанным фильтрам, можно ввести алгебру на теоретико-множественных операциях объединения (сложения) и пересечения (умножения). Такая алгебра является алгеброй синтеза изображений на планигоне, а синтезированный объект соответственно является образом (зрительным образом) наблюдаемого изображения. Отмечается, что семейства этих групп допускают своё использование на этапах принятия решений и понимания анализируемого изображения. [3]

В данной алгебре A_V существуют следующие алгебраические группы:

1) P_{ni} – группы на трёх элементах (названы полными), образованы на тройках операторов (V_i, V_j, V_k) , для которых справедливы соотношения: $V_i + V_j + V_k \equiv e_1$ – единица. В полной группе допускается два возможных типа описания: первое строится на операции пересечения (умножения) $V_i V_j V_k$, при этом число инверсий чётно и операции объединения (сложения) $\bar{V}_i + V_j + V_k$ при этом число инверсий – нечётно. Количество допустимых образов полных групп, с учётом инверсий операторов, равняется четырём. Образ полной группы является компактом из четырех связанных элементов.

2) P_{si} – группы на четырёх элементах (названы замкнутыми), образованы на четвёрке операторов (V_i, V_j, V_n, V_m) , где $(V_i, V_j, V_k) \in P_{ni}$, $(V_n, V_m, V_k) \in P_{nj}$, с описанием $V_i V_j + V_n \bar{V}_m$ (где необходимое число инверсий операторов нечетно) и единицей $- V_i + V_j + V_n + \bar{V}_m \equiv e_1$. Число возможных образов замкнутых групп, с учётом инверсий операторов, равно восьми. Образ полной группы – компакт из восьми связанных элементов.

Множества $\{P_{ni}\}$, $\{P_{si}\}$ конечны и имеют мощность 35 и 105 соответственно. Множества $\{V_i\}$, $\{P_{ni}\}$, $\{P_{si}\}$ – множества эталонов, используемые при решении задачи узнавания объекта исследования в пространстве эталонов. Обозначение полной группы на операции сложения – P_{nia} , на операции умножения – P_{nim} . [2,3,4,5]

Введение данных групп позволяет определить процесс формирования врожденных эталонов [4]. Использование эталонов как «образцов» классов распознаваемых объектов широко используется в теории распознавания образов, при этом для формирования данных выработанных эталонов обычно выступает процесс обучения.

Пространство классов есть пространство эталонных изображений, представленных в пятнадцатимерном признаковом пространстве, которое построено на базисе пятнадцати независимых градиентных преобразований d/dx^i , где $i = 1, \dots, 15$, образующих евклидово пространство E^{15} . В этом пространстве подмножества эталонных изображений образуют классы близких эталонных изображений - точек подмножества, для которых близость необходимо должна определяться некоторым евклидовым расстоянием r . Отмечается, что такой подход является естественным, правомерным и представляет собой фактически стандарт в теории распознавания образов, однако требует знания значения r и априорного знания, чем является эталон класса. В качестве последнего должно выступать некоторое обобщенное изображение на подмножестве изображений, позволяющее успешно (и желательно достоверно) решить проблему

классификации. Такое обобщенное изображение является выработанным, обобщенным эталоном класса эталонных изображений.[3]

Таким образом, а теории активного восприятия пространством врожденных эталонов являются:[3,4]

1. пространство операторов
2. пространство полных групп, имеющее мощность 35 или, если учитывать нулевой оператор, 36
3. пространство замкнутых групп, имеющее мощность 105.

В работе [6], посвящённой оценке частоты основного тона в звуковом сигнале в условиях помех, данная алгебра используется для формирования описания звукового сигнала [6], для чего с помощью замкнутых и полных групп выполняется спектрально-корреляционный анализ. При этом полные группы позволяют выявить корреляционные связи между операторами, замкнутые группы позволяют выявить корреляционные связи между полными группами.

В качестве аналогии между языком и алгеброй групп можно поставить в соответствие оператору алфавит языка, а в соответствие полной группе – слово, замкнутой группе – словосочетание.[6]

В дальнейшем в работе [6] методология теории активного восприятия используется для определения частоты основного тона путём разбиения входного сигнала на ключевые сегменты с дальнейшим построением для этих групп ключевых сегментов.

2.2 Исследование проблемной области поставленной задачи

2.2.1 Психоакустические и когнитивные исследования восприятия звука

Для апробации выбранной теории к задаче распознавания музыкальных произведений необходимо исследование проблемной области, связанной с особенностями восприятия звуков человеком.

Эта необходимость вызвана тем, что музыка является продуктом человеческой культуры, поэтому для её исследования требуются факты,

накопленные массивом наук о человеке, которые призваны дать представление о тех аспектах проблемы, на которых необходимо фокусироваться при построении решения.

Прежде всего, необходимо рассмотреть вопрос об устройстве слуховой системы и выделить аспекты, которые могут быть важны для решения поставленной в рамках данной работы задачи.

Эта система состоит из следующих компонентов: наружного уха, среднего уха и внутреннего уха.

Наружное ухо состоит из ушной раковины и ушного прохода, связанного с барабанной перепонкой, и осуществляет улавливание звуковых волн и передачу их через барабанную перепонку системе среднего уха.

Среднее ухо состоит из трёх косточек (имеющих названия молоточек, наковальня и стремечко), и его функция заключается в передаче звуковых колебаний от барабанной перепонки во внутреннее ухо.

Внутреннее ухо состоит из улитки (cochlea), представляющей собой орган, имеющий вид, похожий на улитку (в связи с чем и было дано его название), заполненный жидкостью. Поверхность улитки покрыта слуховыми рецепторами, которые при возбуждении звуковыми колебаниями передают нервные импульсы в мозг.

Наиболее интересным является рассмотрение улитки. Согласно резонансной теории слуха, выдвинутой Г. Гельмгольцем, улитка функционирует как набор резонаторов. При прохождении звуковой волны от начала до конца улитки некоторые области её сенсорной поверхности подвергаются большему возбуждению на соответствующих им резонансных частотах, чем другие. Этот эффект вызван разной шириной канала улитки и различной жёсткостью её поверхности при движении от базального конца к апикальному. Конечным же итогом прохождения звукового пакета через объём улитки является разложение сложных звуков на парциальные тоны. [19,20].

Преобразованием, аналогичным тому, которое осуществляется улиткой уха, можно назвать преобразование Фурье, которое заключается в разложении одной функции вещественной переменной на составляющие, представляющие собой гармонические колебания.

Таким образом, можно считать, что человек преобразует звуки в вид, аналогичный тому, который осуществляется в преобразовании входного сигнала в спектрограмму. Это позволяет разложить сигнал с быстроменяющейся амплитудой звуковой волны на более медленно изменяющееся спектральное представление поступающего звука.

Именно сходство принципа работы улитки внутреннего уха с преобразованием Фурье и определило столь широкое использование спектральных преобразований как в приложениях распознавания музыкальных произведений, так и в более широкой области задач компьютерной обработки звука, таких, как, например, обработка речи.

Таким образом, стандартом де-факто во множестве работ, посвященных автоматическому распознаванию музыки, используется подход, основанный на формировании спектрограмм на первоначальном этапе информационных преобразований. Как правило, эти спектрограммы формируются с помощью оконного преобразования Фурье или вейвлет-преобразования, которое заключается в использовании фрейма определенной длины, перемещающегося по массиву отсчётов сигнала с определенным шагом. В дальнейшем все информационные преобразования осуществляются с полученной спектрограммой. Необходимо отметить, что в большинстве работ используются различные вариации нейросетевых методов. [8]

В конечном итоге, также представляется необходимым в качестве первого этапа информационных преобразований в разрабатываемой программной системе использовать построение спектрограммы входного звукового сигнала, содержащего аудиозапись музыкального произведения. Последующие же этапы информационных преобразований предлагается

решать с помощью методов, основанных на адаптации теории активного восприятия к условиям задачи распознавания музыкальных произведений.

Следующим важным аспектом, который должен быть рассмотрен в рамках построения решения задачи данной работы, является феномен, получивший название «эффекта вечеринки» («cocktail party phenomenon»).

Данный эффект заключается в способности человека выделять в сложной акустической обстановке отдельные, требуемые ему звуки. Примером проявления этого эффекта, по которому ему и было дано полученное название, является возможность человеческого мозга выделять на шумной вечеринке отдельный голос знакомого приятеля. [

Ещё одним проявлением этого эффекта является способность человека обучаться новым, никогда ранее не слышанным звукам, и впоследствии выделять их из массы других играемых звуков. [20, 21]

На основании этого предполагается, что этот эффект вызван активным использованием памяти в процессе акустического восприятия [21]. *В конечном итоге всё вышеизложенное приводит к выводу, что в распознавании полифонических музыкальных произведений ведущую роль играет память о звуках, которые были услышаны ранее, и применительно к данному исследованию это означает необходимость создания специальной библиотеки образцов звучания музыкальных инструментов для реализации как функциональности распознавания полифонии, так и для распознавания типа инструмента, которым проигрывается тот или иной звук.*

2.2.2 Спектральные особенности музыкальных звуков

Следующим важным аспектом исследуемой предметной области является вопрос о спектральной структуре музыкальных звуков.

Прежде всего, каждый музыкальный звук содержит в своём составе основной тон, представляющий собой гармонику наибольшей амплитуды, которая и определяет проигрываемую ноту.

Кроме основного тона музыкальный звук также содержит сторонние гармоники, носящие название обертонов. Существование сторонних гармоник порождает окрас звука и формирует обертоновый портрет, который различен у каждого инструмента. Благодаря этому на основе образца играемого звука становится возможным определение типа инструмента.

Для наглядной демонстрации этих различий можно привести обертоновые портреты, полученные для скрипки, пианино и гитары, полученные, и представленные на рис. 1, 2, 3.

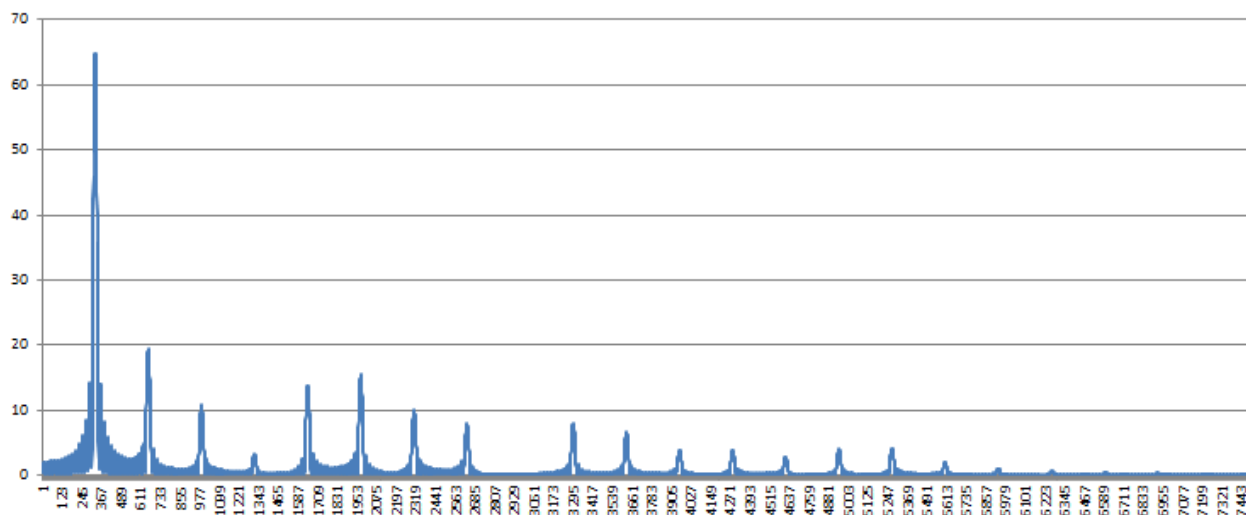


Рисунок 1. Спектр звука, проигрываемого скрипкой

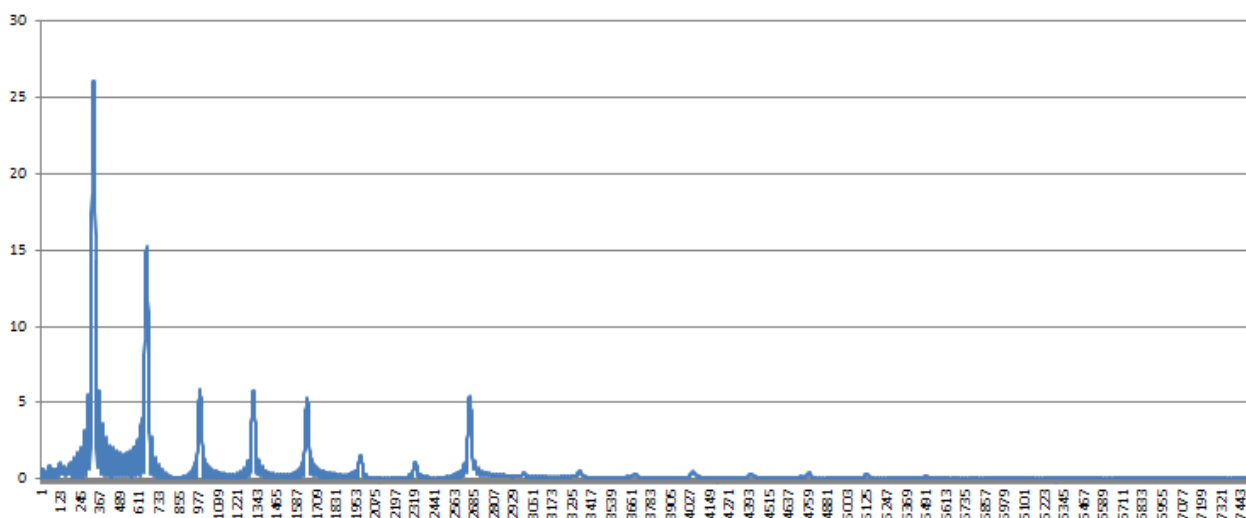


Рис. 2 Спектр звука, проигрываемого пианино

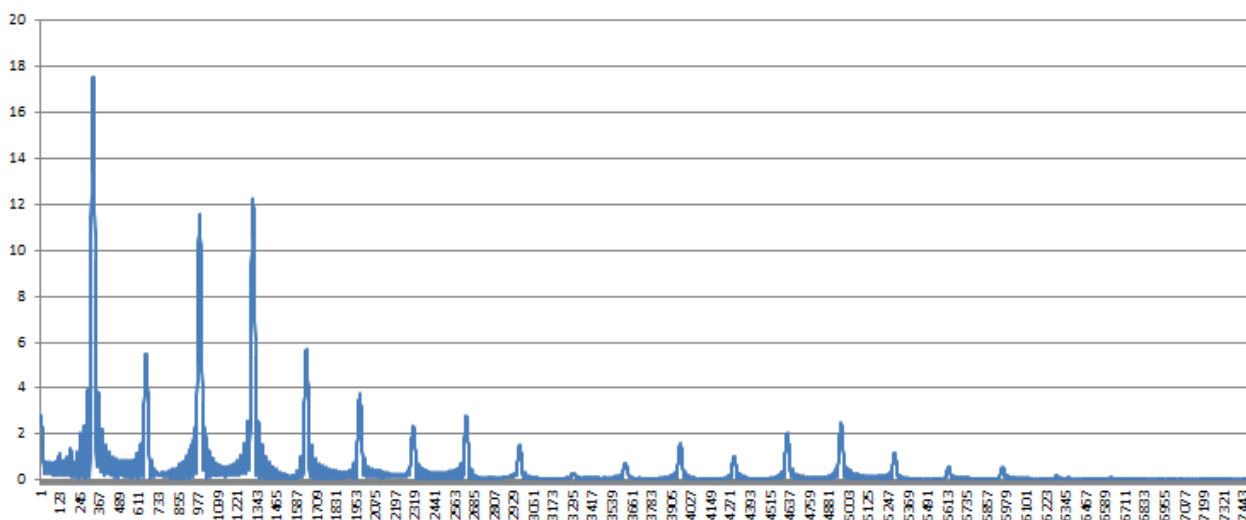


Рис. 3 Спектр звука, проигрываемого гитарой

Все звуки, представленные на этих рисунках, имеют одну и ту же частоту основного тона, которая представлена как пик с наибольшей амплитудой, однако относительная конфигурация других пиков, и являющихся обертонами гармониками, отлична для каждого случая. Как можно видеть, преобладающие обертоны для скрипки являются гармоники высоких частот, для пианино более характерно излучение энергии в нижней части спектра, гитара же занимает некоторое промежуточное положение между скрипкой и пианино.

Также необходимо отметить, что исследование спектра музыкальных звуков, в том числе представленных на вышеприведенных рисунках, показывает, что, как правило, обертоны следуют на частотах, кратных частоте основного тона.

В связи с этим интересно отметить, что благодаря тому, что сейчас в музыке стандартом де-факто является равномерно темперированный строй, в котором ноты частоты нот выводятся как степени двойки, эффект следования обертонов на кратных основной частоте частотах приводит к тому, что

обертоны также оказываются на частотах, соответствующих нотам равномерно темперированного строя.

Для демонстрации проблемы распознавания полифонических мелодий можно произведений можно привести спектры, представленные на рис. 4 и рис. 5.

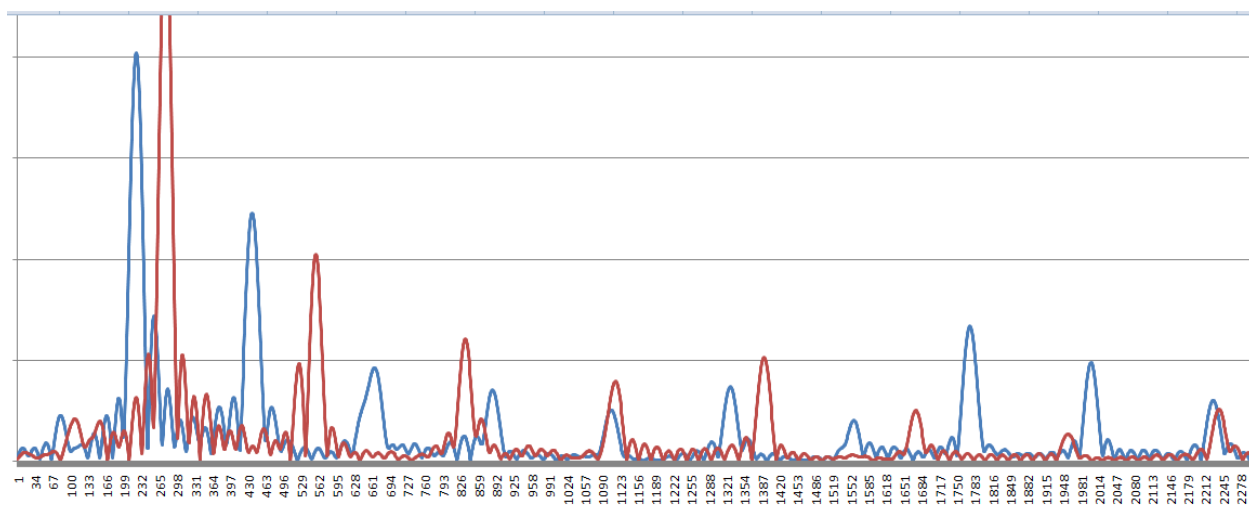


Рис.4 Спектры двух различных музыкальных звуков, представленные на одном графике.

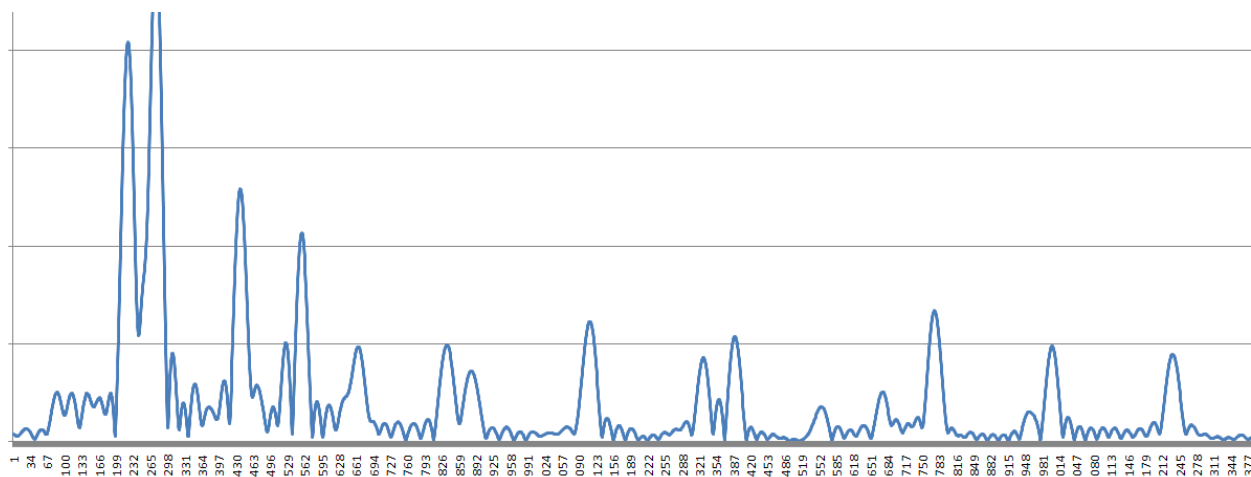


Рис.5 Спектр двух музыкальных звуков, проигрываемых одновременно

На рис. 4 показаны два спектра двух различных музыкальных звуков, представленные на одном графике. На рис. 5 показан спектр сигнала, получающийся при одновременном проигрывании двух этих звуков. Как

можно видеть, спектр результирующей суммы практически является суммой спектров отдельных спектров отдельных его слагаемых, что объяснимо соответствующим свойством преобразования Фурье.

Кроме обертонового портрета также существует другой аспект, вызванный существованием атаки. Атаку можно интерпретировать как «первоначальный импульс звукоизвлечения» [23].

«В аспекте сугубо физическом, тот или иной характер атаки звука обусловлен крутизной и длительностью переднего фронта спектрального состава частотных колебаний первоначального звукового импульса, образующегося в момент извлечения звука при игре на каком-либо музыкальном инструменте или при исполнении вокальных партий.»[23]

С точки зрения физического процесса проигрывания инструментом звука, существование атаки обусловлено неравномерным излучением звуковой энергии в окружающую среду с течением времени.

Как и обертоновый портрет, атака различается у разных инструментов: если скрипка может выдавать одинаковый спектральный портрет на всём протяжении ноты, то в начале гитарного музыкального события следует столь мощная атака, что основной тон начального фрагмента может не соответствовать основному тону самой ноты, и переход к основному тону происходит лишь с течением времени по мере угасания атаки.

В качестве примера атаки можно привести увеличенный фрагмент спектрограммы музыкального звука, представленный на рис. 6.

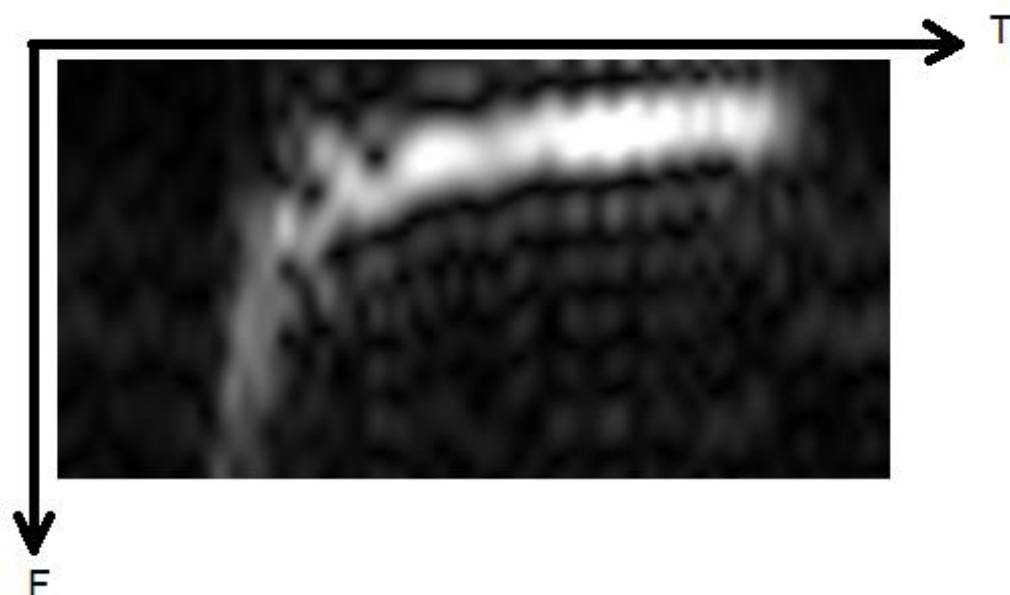


Рис. 6. Увеличенный фрагмент спектрограммы, демонстрирующий развитие атаки во времени

Из-за того, что в начале музыкального звука следует мощный энергетический импульс, начало проигрываемой ноты сдвинуто в область верхних частот и переходит на частоту основного тона только на протяжении некоторого времени.

В конечном итоге, рассматривая атакующий и обертоновый портреты одновременно, можно дать следующее схематичное представление музыкального звука, представленное на рис. 7.

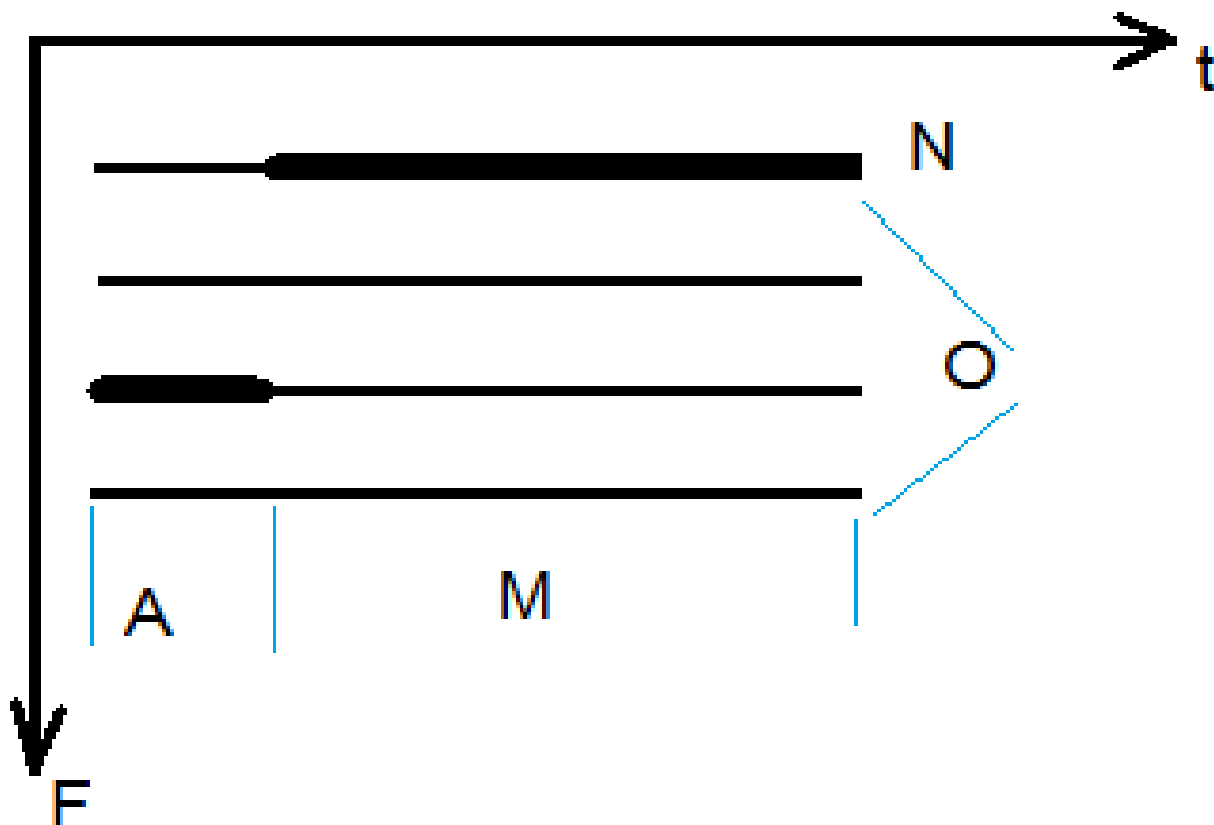


Рис. 7. Схематичное представление спектрограммной структуры музыкального события

На рисунке 7 обозначены следующие компоненты музыкального события:

- A – атака
- M – основная часть звукового события
- O – обертоны
- N – основной тон

Жирной линией выделена гармоника с наибольшей амплитудой. Как можно видеть, спектрографический портрет в начале музыкального звука отличается от такового при последующем его развитии.

Как можно видеть, что вначале на спектрограмме следует атака, при которой основная энергия может излучаться на гармонике, не соответствующей основному тону (гармоника с наибольшей амплитудой показана жирной выделенной линией), в дальнейшем же спектральный

портрет нормализуется и наибольшая гармоника начинает соответствовать требуемой ноте.

В связи с распознаванием полифонических произведений необходимо также дать представление о том, как будет выглядеть спектральная картина при проигрывании полифонического произведения (рис. 8)

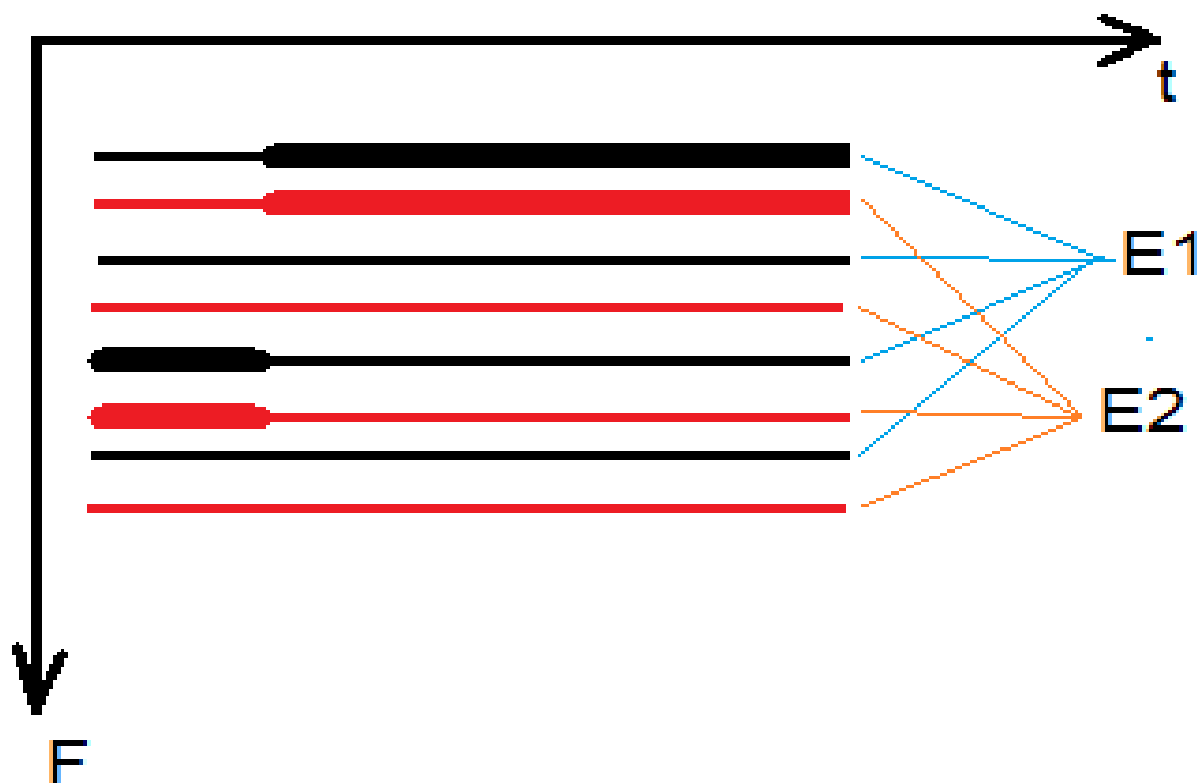


Рис. 8 Схематическое изображение спектрограммы одновременного проигрывания двух музыкальных событий.

В данном случае можно наблюдать одновременное проигрывание двух событий E_1 и E_2 . Предполагается, что оба события играютя одновременно, то есть фактически они представляют собой аккорд.

Кроме факторов, связанных с развитием спектрального портрета во времени также необходимо учитывать аспекты, связанные с особенностями «живого» проигрывания. Так, например, скрипач управляет процессом звукоизвлечения благодаря тонкостям управления смычком и тонкостям управления грифом. Всё это порождает очень мощный инструмент для того,

чтобы осуществлять очень тонкое управление различными оттенками звука, что благоприятно сказывается на восприятии звука слушателем.

Практика музыкального искусства наглядно показывает, что наиболее нравящимися слушателям являются не произведения, в которых каждый звук одной и той же ноты одного и того же инструмента играет без малейших изменений (что при живом исполнении вообще недостижимо физически), а такие, в которых исполнитель при проигрывании своей партитуры добавляет различные оттенки, изменения в звучание. С точки зрения формального музыковедения такие произведения являются идентичными, однако с точки зрения слушателя «живое» исполнение является более предпочтительным, чем «механическое».

Этот эффект также приводит к тому, что в процессе живого исполнения музыкального произведения также имеет место быть различие в проигрывании каждой ноты в партии при переходе от одной ноты к другой. Однако это различие никак не проявляется в стандартной нотной записи, что говорит о необходимости разработки метода для компенсации этого различия для целей преобразования проигранной мелодии в готовое представление классической музыкальной нотации. Иными словами, необходима разработка методов компенсации неравномерности проигрывания.

Таким образом, становится возможным выделить следующие подзадачи, на которых необходимо фокусироваться при разработке требуемых методов и алгоритмов:

1. Распознавание обертонового портрета
2. Обработка атакного портрета
3. Обработка вариаций воспроизведения для преодоления особенностей игры при реальном воспроизведении
4. Обработка характеристик ноты, необходимых для преобразования к классическому виду.

Определение данных подзадач необходимо провести для того, чтобы в дальнейшем определить структуру и вид этих подсистем для реализации программной системы.

2.3 Построение необходимых математических моделей

Прежде всего, необходимо ввести модель музыкального события, которая могла бы осуществлять представление проигрываемых музыкальных событий. Его можно определить как кортеж $E = \langle N, t_n, t_d, I \rangle$, где N – код ноты, t_n – время начала проигрывания ноты, t_d – длительность проигрывания, I – тип инструмента.

Символьное представление музыкального произведения – это множество событий $M = \{E_1, E_2, E_3 \dots E_n\}$, где E_i – события, составляющие это музыкальное произведение.

В качестве представления, пригодного для преобразования в классический вид можно представить модель, аналогичную преобразованиям в классический вид, однако включающий в себя обозначение высоты ноты и её длительность в долях такта.

Таким образом, классическую ноту можно представить как кортеж кортеж $N = \langle N, t_n, t_d \rangle$, исключив из рассмотрения тип инструмента, так как обычно тип инструмента определяется для всей партии. Таким образом, также целесообразно ввести следующую модель для партий. Партия будет представлять из себя также кортеж $P = \{N_1, N_2, \dots N_n\}$, а всё музыкальное произведение будет иметь вид множества $S = \{P_1, P_2, \dots P_n\}$, где P_i – это отдельная партия.

Также представляется необходимым формирование модели входного сигнала. В качестве входных данных разрабатываемая программная система использует аудиофайлы формата WAV, аудиоданные в которых представляют собой оцифрованный звуковой сигнал в импульсно-кодовой модуляции. Таким образом, входной сигнал S возможно представить как упорядоченный массив $S = \{s_1, s_2 \dots s_n\}$, где s_i – значение i -го отсчета.

В качестве модели спектрограммы возможно выбрать модель матрицы, в которой номер столбца соответствует номеру фрейма, а номер строки – номеру гармоники. Таким образом, если спектрограмма содержит n отсчётов, и m гармоник, то её модель можно представить следующим образом:

$$G = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ g_{21} & g_{22} & \cdots & g_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nm} \end{pmatrix}$$

где g_{ij} – величина гармоники с номером i в фрейме j .

2.4 Метод распознавания обертонального портрета

Теория активного восприятия, будучи применяемой для вопросов обработки изображения предполагает использование множество масок, покрывающих область определения изображения G , определено на системе базисных функций [2].

Аналогично случаю обработки изображения, в работе [6] вводятся аналогичные фильтры, также в количестве 16-ти, что позволяет проводить более тонкий анализ сигнала. Вид этих фильтров представлен на рис. 9.

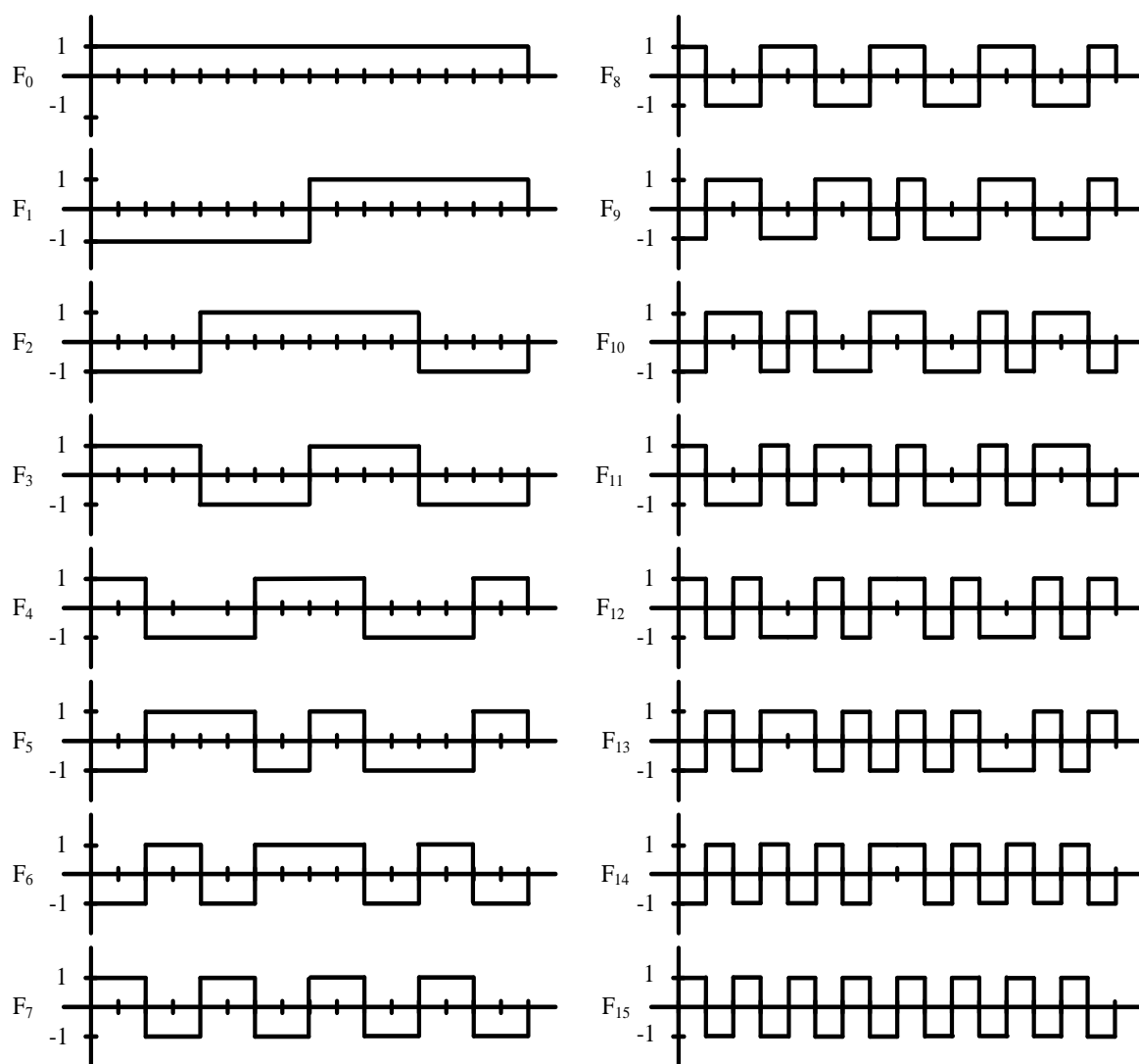


Рис. 9. Фильтры, используемые для анализа сигнала

С позиции алгебры описания изображений, как уже упоминалось, пространством врожденных эталонов является:

1) пространство операторов $V = \{V_i\}$ с базисом e_i , где орт e_i определен координатным базисом $\{d/dx^i\}$. Данное множество имеет мощность, равную пятнадцати, таким образом, для полного представления необходимо евклидово пространство E^{15}

2) пространство полных групп $P_n = \{P_{ni}\}$, имеющее мощность 35

3) пространство замкнутых групп, имеющее мощность 105.

Таким образом, теорией активного восприятия предоставляется три типа пространств врожденных эталонов, которые могут быть использованы

для решения задач распознавания. Поскольку, как уже было указано, для реализации распознавания полифонических музыкальных произведений необходимо создание специализированной библиотеки звуковых музыкальных событий (необходимость данной библиотеки, как также указывалось, диктуется и исследованиями в психоакустике, что выражено в так называемом «эффекте вечеринки»), то вышеизложенный математический аппарат, основанный на создании врожденных эталонов и создании пространств данных эталонов представляется пригодным для решения задач, поставленных в рамках данной работы.

С учетом изложенных концепций ТАВ в качестве этого метода может выступить применение фильтров к полученной спектрограмме и построение с помощью теории групп его образа.

Исследование обертонального портрета показывает, что в большинстве случаев необходимые для распознавания типа инструментов обертоны находятся в частотной области от 0 до 4000 Гц. Таким образом, для решения задачи о распознавании обертонального портрета будет целесообразно прежде всего выбрать эту область.

В дальнейшем предлагается осуществить применение фильтров У-преобразования данной частотной области спектра фрейма таким образом, как показано на рис. 10.

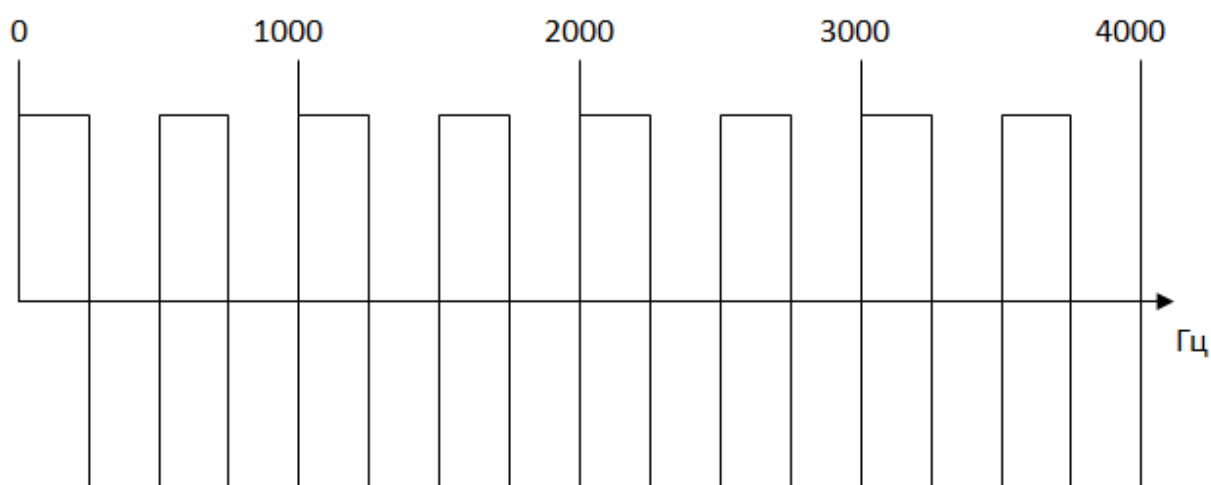


Рис.10 Применение фильтра F15 к частотному диапазону 0 – 4000 Гц

После применения всех 16 фильтров (включая нулевой) к данной частотной области получается набор из 16 коэффициентов, каждому из которых соответствует свой оператор.

Благодаря применению фильтров к разным экземплярам образцов становится возможным формирование пространства врождённых эталонов, соответствующая пространству операторов. Величина же каждого коэффициента, даваемая фильтром, соответствует отклику оператора, соответствующего этому фильтру. В дальнейшем становится возможным сформировать следующие пространства врождённых эталонов на полных и замкнутых группах.

В работе [5] такой набор коэффициентов, полученных в результате применения фильтров U-преобразования к сигналу, получил наименование «вектор-описание».

Проблематика распознавания полифонического произведения, как уже было рассмотрено ранее, состоит в том, что одновременно может проигрываться несколько музыкальных звуков, которые накладываются друг на друга. Таким образом, для распознавания полифонического произведения необходимо провести обратную операцию, то есть разложить условную сумму этих звуковых событий на отдельные составляющие. Эта сумма может быть только условной в виду того, что кроме самих исходных амплитуд сигналов могут свой вклад в конечный результат также неизбежно вносят шумы, интерференция звуковых волн двух проигрываемых музыкальных звуков, а также различные воздействия на вид результирующей звуковой волны из-за возникновения эхо и других факторов.

Всё это приводит к тому, что задача о распознавании полифонической музыки считается в настоящее время крайне сложной и не имеющей удовлетворительного решения.

Рассмотрим, какими методами может располагать теория активного восприятия для решения этой задачи.

Рассмотрим результат последовательного применения U-преобразования к двум спектрам исходных музыкальных событий, а потом к одновременному их проигрыванию. Можно легко показать, что U-преобразование является линейным оператором, поэтому в идеальном случае вектор-описание суммы спектров будет являться суммой отдельных векторов-описаний спектров.

Однако в реальном случае эта сумма из-за описанных выше факторов (помехи, эхо и т.д.) будет отличаться от идеальной суммы, однако по-прежнему будет близка к ней ввиду того, что основные звуковые сигналы, несущие информацию (обертоны музыкальных звуков) будут преобладать над шумом. Это будет выполняться до тех пор, пока громкость шума будет достаточно низка относительно полезного сигнала.

Тем не менее, в условиях, когда шумы достаточно низки, подобное свойство U-преобразования делает возможным предложить следующий вариант алгоритма распознавания двух звуковых событий, играющих одновременно:

- 1) На первом этапе, этапе построения библиотеки строятся векторы-описания для всех образцов музыкальных звуков для различных инструментов.
- 2) На следующем этапе осуществляется сложение всех возможных пар векторов по принципу «каждый с каждым»: если количество образцов в библиотеке равняется N , то при сложении всех возможных пар векторов будет получено $N \times N = N^2$ новых векторов-описаний.
- 3) В дальнейшем, при проигрывании пары музыкальных звуков, имеющих в библиотеке, вычислить вектор-описание спектра одновременного проигрывания.
- 4) Среди всех возможных векторов-описаний, полученных на шаге 2, выбрать те, которые близки к вектору-описанию входного сигнала, полученному на шаге 3. Согласно свойству линейности U-

преобразования данный вектор может являться вектором-описанием тех сигналов, которые образуют проигранную комбинацию. Поскольку на шаге 2 была известна та пара образцов, вектор-описание которых было сложено, то таким образом из найденного по принципу большего сходства вектору становится возможным получить представление о тех образцах, одновременным проигрыванием которых и был получен поступивший на обработку сигнал.

Однако в этом случае крайне вероятна ситуация, когда две различные пары образцов при получении суммы их векторов-описаний дадут один и тот же результат, что при поиске составляющих даст большую вероятность ошибки.

Рассмотрим, какие методы может ещё предоставить теория групп для решения данной проблемы.

Теория групп предполагает построение на пространстве операторов двух классов групп: полных групп и замкнутых групп.

Полная группа, согласно определению – это такая группа из трёх операторов V_i, V_j, V_k , для которых выполняется условие $V_i + V_j + V_k = e_1$ ($e_1 = V_0, i \neq j \neq k \neq 0$). полные группы позволяют выявить корреляционные связи между операторами [6]. Полные группы позволяют реализовывать Данный класс групп два описания – на операции пересечения $V_i V_j V_k$ (операция умножения, число инверсий – чётно) и операции объединения $\bar{V}_i + V_j + V_k$ (операции сложения, число инверсий – нечётно).

Для каждой полной группы возможно привести четыре образа с учётом инверсий.

Замкнутые группы строятся следующим образом:

- 1) Из набора полных групп выбирается одна полная группа $V_i V_j V_k$
- 2) Для выбранной группы формируется четыре образа на операции умножения: $V_i V_j V_k, \bar{V}_i \bar{V}_j V_k, \bar{V}_i V_j \bar{V}_k, V_i \bar{V}_j \bar{V}_k$;

- 3) Полученные образы анализируются следующим образом: если в состав образа не входит оператор V_0 , тогда входящие в него операторы, образуют замкнутую группу.

После просмотра всех возможных групп становится возможным создание набора замкнутых групп.

Таким образом, иерархия операторов, полных и замкнутых групп имеет вид, представленный на рис. 11

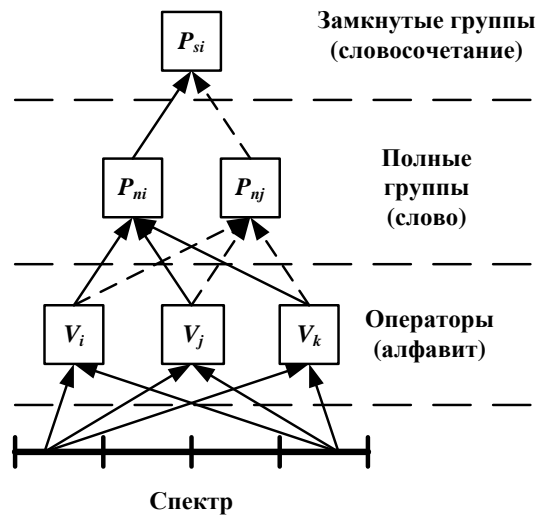


Рис. 11. Иерархия операторов, полных групп и замкнутых групп

Каждому оператору, согласно принятым в ТАВ определению, соответствует фильтр, получаемый заменой $-1 \rightarrow 0$ и $1 \rightarrow 1$; то есть, таким образом, каждому оператору соответствует числовой коэффициент, полученный в результате применения соответствующего фильтра к сигналу. Таким образом, каждой группе можно поставить в соответствие три коэффициента, соответствующие входящим в эту группу операторам.

ТАВ также предполагает использование такого понятия, как масса группы. Масса группы – это сумма отсчётов сигнала, которая находится под образом этой группы. [6]

Если дан сигнал f , то для вычисления массы группы используется следующий алгоритм:

$$\begin{aligned} \forall k \in \overline{1, L} \\ \text{if } B_{grp}[k] = 1 \\ M_{grp} = M_{grp} + f[k] \end{aligned}$$

где $f[k]$ – k -ый отсчёт сигнала f , B_{grp} – образ группы, M_{grp} – масса группы.

Таким образом, построение полных групп даёт дополнительные возможности для различения сумм фильтров, соответствующих этим операторам, благодаря тому, что сам механизм построения полных групп учитывает также пространственное расположение областей «1» и «0», что позволяет осуществлять дополнительное различение пространства входных сигналов, что позволяет осуществлять дальнейшее различение. Кроме того, ещё одним способом различения двух входных событий также может служить возможность построения образа полной группы и вычисление массы, находящимся под этим образом.

В работе [6] для определения частоты основного тона для каждого ключевого сегмента осуществляется построение набора замкнутых групп, для каждой из которых впоследствии вычисляется масса, а затем осуществляется сортировка групп по убыванию массы. В дальнейшем для определения частоты основного тона производится сравнение наборов групп, полученных по i -му и j -му сегментам и в том случае, если эти наборы совпадают не менее чем на некоторое пороговое значение M %, то ключевые сегменты, которым соответствуют эти наборы, признаются подобными с некоторой точностью, а расстояние между ними заносится в массив H . Для вычисления частоты основного тона определяется номер максимального элемента гистограммы H , который соответствует наиболее часто встречающемуся расстоянию между сегментами и таким образом соответствующим частоте основного тона. Сама же частота основного тона вычисляется на как отношение частоты дискретизации к этому расстоянию, измеряющемуся в отсчётах.

Таким образом, построение описания эталона с использованием замкнутых групп позволяет осуществить эффективное разделение пространства элементов в пространстве классов.

Для решения задачи распознавания суммы обертональных портретов использование замкнутых групп может позволить разрешить проблему одинаковых сумм применением фильтров U-преобразования благодаря тому, что позволяет производить более тонкое пространственное разделение входных спектральных сигналов.

Таким образом, становится возможным предложить следующий алгоритм для построения библиотеки обертональных портретов отдельных музыкальных событий:

- 1) Формируется библиотека образцов музыкальных звуков
- 2) Для каждого звука строится спектрограмма и в этой спектрограмме выбирается фрейм, соответствующий основной части музыкального звука (то есть такой части, которая следует за атакой и где спектральные характеристики звука уже нормализованы)
- 3) К данному фрейму применяются фильтры U-преобразования так, как показано на рис. 10. Таким образом осуществляется построение вектора-описания для данного фрейма.
- 4) Для данного фрейма также строится описание на массиве замкнутых групп и формируется массив масс замкнутых групп.
- 5) Для библиотеки нот строится массив пар сумм векторов-описаний (для распознавания двух инструментов), троек (для распознавания трёх инструментов) и так далее в зависимости от того, сколько может быть одновременно играемых инструментов в музыкальном произведении

Для распознавания входного сигнала предлагается использовать следующий алгоритм:

- 1) Для обрабатываемого произведения строится спектрограмма

- 2) К каждому фрейму применяются фильтры U-преобразования так, как показано на рис. 10
- 3) Для этого вектора-описания ищутся вектора-описания среди построенных на этапе формирования библиотеки для пар, троек и кортежей более высокого порядка (при их наличии), такие, чтобы евклидово расстояние между ними было минимально. Для этого получаемые расстояния динамически сортируются.
- 4) Для полученного массива осуществляется проверка, с помощью построения описаний на основе теории групп следующим образом: для всех образцов векторов-описаний вычисляется описание на основе масс замкнутых групп, в дальнейшем все массы аналогичных групп для кортежа складываются. Аналогичный набор формируется и для анализируемого фрейма. Затем происходит поиск наиболее похожего набора групп среди анализируемого сигнала аналогично тому, каков используется в шаге 3 для векторов-описаний. Набор с наименьшим евклидовым расстоянием считается присутствующим в сигнале и таким образом для анализируемого фрейма формируется кортеж, содержащий сведения о том, какого типа события присутствуют в данном фрейме.

Подобный подход позволяет производить в начале менее вычислительно трудоёмкий поиск среди векторов-описаний, а не использовать поиск по замкнутым группам для всех кортежей. Для следующего этапа информационных преобразований, которые заключаются в обработке атаки, необходимо приведение евклидова расстояния между образом исследуемого фрейма и образом суммы библиотечного фрейма.

2.5 Методы компенсации атаки, неравномерности проигрывания и приведения полученной модели к классическому виду

На предыдущем этапе информационных преобразований входящий сигнал, преобразованный в спектрограмму, преобразуется в массив

кортежей, каждый из которых представляет собой номера тех музыкальных звуков из библиотеки, которые при анализе на предыдущем этапе признаются наличествующими в этом фрейме. Для анализа атаки необходимо использование евклидова расстояния между каждым из фреймов и тем кортежом образов, которые признаются наличествующими в этом фрейме.

Следуя анализу структуры музыкального события, приведенному на рис. 6, с течением времени должно осуществляться приближение спектральных характеристик атаки к спектральным характеристикам основной ноты.

Таким образом, выделение атаки становится возможным совместить с определением границ музыкальных событий при помощи порогового фильтра.

После применения этого порогового фильтра, выделяющего границы музыкальных событий, формируется массив звуковых событий.

В результате проведения предыдущих информационных преобразований входной сигнал распознаётся как массив звуковых событий. Такой массив событий может использоваться для преобразования входного сигнала в формат MIDI, при этом не имеет значения, насколько точной будет являться длина музыкального события, так как слуховая система человека будет воспринимать множество событий со слегка варьирующей длиной как множество одного и того же размера (до некоторого субъективно определяемого предела). Однако в различных приложениях необходимо, чтобы полученный сигнал был приведён к более стандартизированному виду, такому, какой даётся стандартной нотной нотацией. Для решения этой проблемы необходимо провести постобработку полученной модели музыкального произведения с целью перехода от модели звуковых событий, в которой наступление и длина звуковых событий различна, к более точной модели классической нотной записи с фиксированной шириной темпа и размером нот.

Для решения данной проблемы целесообразно принять величины длительностей музыкальных событий и времена их наступления как некоторые случайные величины, а в качестве истинных значений длительностей и величины темпа принять значения, представляющие собой математическое ожидание этих значений.

Таким образом осуществляется нормализация характеристик звукового события и формирование из него записи, приближенной к классической.

В конечном итоге, схема информационных преобразований имеет вид, представленный на рис.12

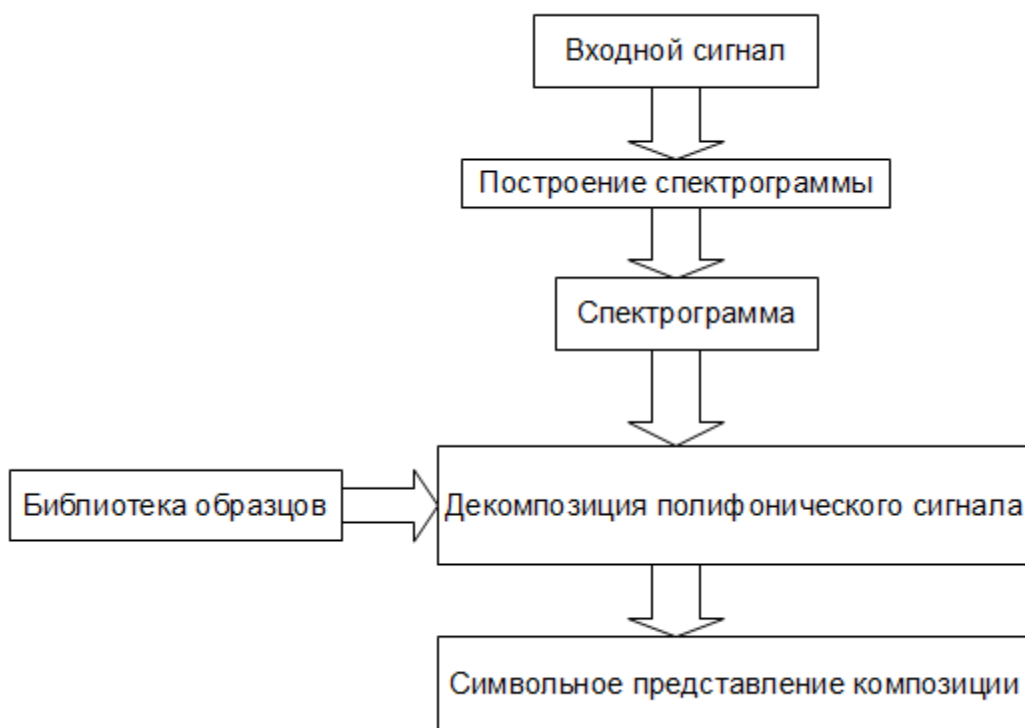


Рис.12. Схема информационных преобразований

2.6 Разработка программной системы

Для разработки программной системы был выбран язык программирования C#.

В качестве среды разработки используется IDE Microsoft Visual Studio 2010, полученная по студенческой лицензии для образовательных целей.

В качестве парадигмы разработки используется объектно-ориентированный подход (ООП).

Для реализации поставленных заданием целей необходимо определить структуры, соответствующие разработанным математическим моделям и создать алгоритмы, реализующие соответствующие методы.

В качестве модели музыкального события был создан класс *SEvent*, содержащий, согласно требуемой математической модели, следующие параметры:

- `start` – тип `double` – время начала данного музыкального события, сек.
- `duration` – тип `double` – длительность данного музыкального события, в секундах
- `note` – тип `int` – код MIDI, соответствующий ноте, проигрываемой во время данного события
- `instr` – тип `int` – код MIDI, соответствующий инструменту, проигрывающему ноту в данном музыкальном событии.

Класс *SEvent* используется классом *MusicComposition*, который является представлением музыкального произведения.

Для реализации библиотеки был создан отдельный класс *SoundsLibrary*.

Данный класс реализует специальные методы, необходимые как для пополнения библиотеки образцами звучания музыкальных инструментов, так и для выполнения обработки загруженных образцов звучаний с целью получения эталонов.

Полученные в результате выполнения методов класса *MusicLibrary* данные могут использоваться в дальнейшем в следующих методах.

В качестве входных данных используются файлы формата WAV(Waveform Audio File Format), использующийся для хранения несжатого звука в импульсно-кодовой модуляции. Для обработки входных

файлов необходимо создание отдельного класса, который будет реализовывать загрузку содержимого этих файлов и представление этого содержимого в удобном для обработки виде.

Для реализации данной функциональности был реализован отдельный класс, получивший наименование *WAVLoader*. В качестве методов данного класса были определены метод *LoadFile* следующего вида, принимающий

Для реализации загрузки содержимого WAV-файла необходимо чтение заголовка WAV-файла, что и реализуется методом *LoadFile*. Данные об амплитуде звукового сигнала в формате WAV могут иметь тип знакового или беззнакового целого, поэтому чтение заголовка файла должно дать представление об используемом типе кодирования величин (знаковое или беззнаковое), что необходимо для того, чтобы преобразовать входное содержимое файла в массив *double*. После того, как звуковая волна была преобразована в массив *double*, данный массив может использоваться в следующих этапах обработки в других подсистемах разработанной программной системы.

В соответствии с разработанным методом необходимо провести построение спектрограммы как для образцов звучания в библиотеке, так и для всех распознаваемых композиций.

Для реализации процесса построения спектрограммы для дальнейших информационных преобразований необходимо построение отдельного класса, получившего наименование *SpectrogramBuilder*.

Данный класс реализует метод *buildSpectrogram*. В качестве аргументов данного метода используются полученный класс *WAVLoader* массив типа *double*, представляющий данные о величине амплитуды входного звукового сигнала, шаг фрейма, то есть величину, на которую должен смещаться фрейм при построении спектрограммы, и длину этого фрейма.

Для представления полученной спектрограммы используется класс *Spectrogram*.

Данный класс содержит информацию о ширине фрейма `frame_width`, об области частот, для которой строится спектрограмма (при использовании построения от нуля до максимальной используется значение `double max_F`), `b` в качестве основного массива данных выступает матрица, ставящая в соответствие каждой гармонике для данного фрейма значение её амплитуды.

Для реализации алгоритмов распознавания используется класс *RecognitionCore*.

Данный класс содержит метод *Recognize*, принимающий на вход данные, загруженный из файла WAV с помощью класса *WAVLoader*. В соответствии с разработанным алгоритмом этот метод во время своей работы использует такие структуры, как фильтры и группы. Поэтому кроме всего прочего необходима реализация классов, представляющих данную функциональность.

Для предоставления реализации фильтров был создан класс *Filters*. Данный класс содержит конструктор, определяющий длину требуемых фильтров и методы *getAll*, возвращающий весь массив фильтров, и метод *getFunction*, представляющий передачу функции с заданным номером.

Для представления групп используется класс *Groups*, служащий для построения сначала массива полных групп, а на их основе – массива замкнутых групп. Основным методом, который представляет данный класс, является метод *getClosedGroups()*, который возвращает в качестве двумерного массива `int` набор замкнутых групп, при этом для обозначения операторов, входящих в группу, используются их номера.

Для реализации библиотеки используется класс *Library*, предоставляющий для класса *RecognitionCore* массив образцов фреймов обертональных портретов для осуществления основного алгоритма распознавания

Основной метод *Recognize* класса *RecognitionCore* реализует алгоритмы, описанные в пп. 2.4 – 2.5 распознавания.

После осуществления процесса распознавания результат распознавания выводится в окно «Результат распознавания».

В конечном итоге, разработанная программная система имеет следующий пользовательский вид, представленный на рис. 13.

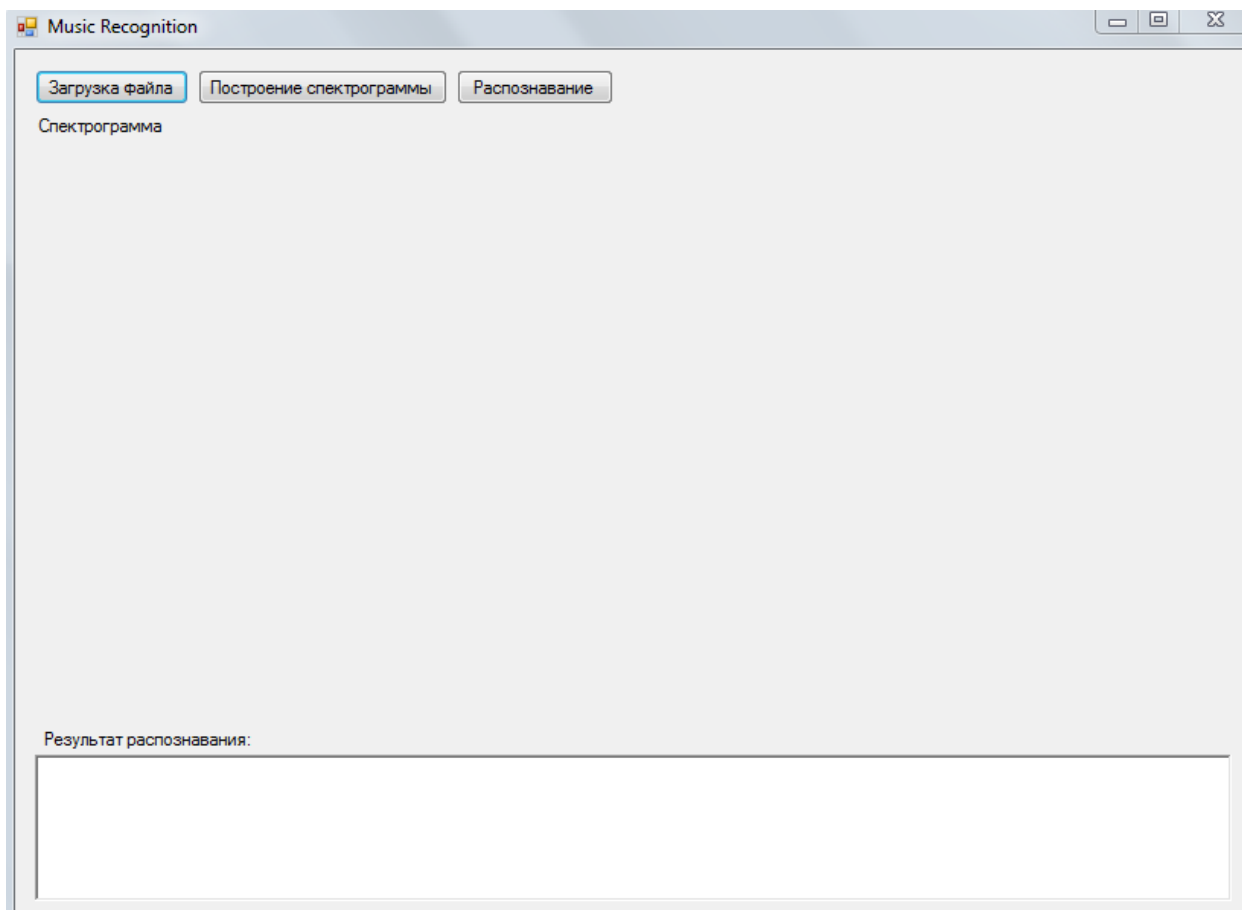


Рис. 13. Вид пользовательского интерфейса разработанной программной системы

ГЛАВА 3. ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

3.1 Планирование вычислительного эксперимента

В рамках данной работы необходимо проведение экспериментального исследования разработанной программной системы. Для реализации этой задачи прежде всего необходимо провести планирование данного вычислительного эксперимента с целью выявления всех потенциальных возможностей и недостатков разработанной системы.

Испытание разработанной программной системы в различных условиях использования необходимо для выявления всех возможных аспектов процесса распознавания, а также для выявления возможных ограничений и возможностей разработанных методов и алгоритмов. Это, в свою очередь, должно дать представление о том, как разработанная программная система будет вести себя в условиях реального мира, а также о том, какие возможные улучшения в рамках дальнейшего развития данного научного направления могут быть достигнуты на основе результатов, полученных в данной работе. Также, в частности, тестирование также должно дать представление о дальнейшем развитии разработанной программной системы.

В качестве таких условий могут выступать добавление различных шумовых компонентов, выступающих в роли аддитивного шума.

Разрабатываемая система предназначена для работы в различных условиях реального мира. В связи с этим необходимо провести анализ возможных зашумлений, с которым возможно столкновение в условиях реального мира.

Прежде всего, возможен вариант с реализацией зашумления различными белыми шумами. Белый шум имеет место быть во множестве физических процессов, в том числе и тех, которые используются в различных приложениях звукозаписи и звукоизвлечения.

В дальнейшем необходимо изучить воздействие, оказываемое импульсным шумом на результаты распознавания. В реальных приложениях

шумом такого рода можно назвать, например, помехи в звуковом сигнале радиоприёмника из-за, например, дробового шума, вызванного эффектами дискретности заряда электронов при протекании в электрической цепи, содержащей каскады усиления слабого сигнала из транзисторов или радиоламп, или шум, возникающий при проигрывании грампластинки из-за повреждений на её поверхности или наличия на ней пыли.

Важность исследования влияния шума такого рода на результаты распознавания обусловлена тем, что одними из важных перспективных приложений технологий автоматического распознавания музыки является выявление плагиата в музыкальной индустрии. При таком варианте использования вполне представима ситуация, когда возникает необходимость получить символическое представление музыкальной композиции, проигранной в радиоэфире.

Также возможна ситуация, когда необходимо получить композицию из старых грамзаписей. Особенно остро эта проблема может стоять в случае записей джаз-исполнителей первой половины и середины XX века, что обусловлено не только использованием несовершенных технических средств звукозаписи в то время, но и тем, что зачастую для джаз-произведений вообще отсутствует какие-либо нотные записи о структуре проигрываемых в них композиции ввиду того, что для джаза характерны импровизированные выступления, когда музыканты сочиняют играемые мелодии «на ходу».

В качестве тестирования разработанной программной системы в различных условиях распознавания также необходимо обратить пристальное внимание на структуру и музыкальной композиции, распознавание которой осуществляется в рамках вычислительного эксперимента, осуществляемого в результате решения задачи по тестированию разработанных методов и алгоритмов.

Целесообразнее всего осуществить метод «от простого к сложному», переходя от простых монофонических композиций (поскольку монофоническая композиция также является частным случаем более общего

класса всех возможных музыкальных произведений), и в дальнейшем через различные усложнения тестовых примеров перейти ко всё более и более сложным случаям, тестируя таким образом разработанную программную систему во всё более и более сложных случаях.

С точки зрения рассмотрения структуры прежде всего целесообразно перейти от случая с простой монофонической композиции с нотами, играемыми одним инструментом с наличием промежутков между ними. Наличие данных промежутков должно привести к тому, что такая композиция станет идеальным случаем при использовании в проигрываемой композиции тех же образцов звучания, которые используются в библиотеке разработанной программной системы распознавания. Фактор наличия пауз между играемыми событиями приведет к тому, что между нотами будут отсутствовать различные взаимодействия, которые могут оказать то или иное шумовое влияние друг на друга. В свою очередь, применение в проигрываемом сигнале тех же образцов звучания, которые используются в библиотеке программной системы гарантирует отсутствие как различных помех, вызванных зашумлением при звукозаписи или звукоизвлечения, так и вариаций воспроизведения, неизбежно возникающих при живом исполнении.

В дальнейшем необходимо усложнить поставленную задачу, реализовав вычислительный эксперимент с распознаванием модельной композицией, аналогичной в предыдущем случае, однако с отсутствием пауз между играемыми нотами. Отсутствие пауз между играемыми нотами может привести к наличию различных шумовых возмущений, вызванных наложением их друг на друга, что, в свою очередь, выявит реакцию разработанной программной системы на это возмущение. Однако данный случай также должен быть лёгок для распознающей системы.

Два вышеописанных эксперимента необходимо осуществить с различными инструментами, чтобы исследовать возможности разработанной программной системы по распознаванию различных типов инструментов.

В дальнейшем необходимо перейти к другому случаю, который предполагает распознавание музыкальной композиции, содержащей аккорды, то есть последовательность пар нот. Данный случай можно считать простым примером простой полифонической музыкальной композиции.

Данный вариант также необходимо сделать с использованием различных инструментов, однако по-прежнему следует применить использование тех же образцов звучания, которые используются в библиотеке программной системы. Использование «идеальных» музыкальных событий в данной композиции позволит исследовать возникающие между нотами возмущения в отсутствие помех, вызванных процессами звукоизвлечения или вариациями проигрывания.

В дальнейшем представляется необходимым исследовать реакцию разработанной программной системы на простую полифоническую музыкальную композицию, также проигрываемую эталонными нотами, используемыми в библиотеке, состоящую из мелодии, проигрываемой на фоне аккордов. Подобный вариант будет близок к реальному использованию системы на настоящих музыкальных композициях, поскольку будет реализовывать достаточно сложный вариант, близкий к тем композициям, которые на самом деле выпускаются музыкальной индустрией. Таким образом станет возможным исследовать реакцию разработанной программной системы в условиях, которые с одной стороны достаточно сильно приближены к реальным, а с другой – являются достаточно модельными для выявления реакций системы на них.

После того, как были получены представления о реакциях системы в идеальных случаях, необходимо перейти к исследованиям реакции системы в данных случаях, с добавлением дополнительных шумовых компонент к исследуемым музыкальным композициям. Подобное добавление сделает возможным исследовать реакцию разработанной программной системы на воздействия, аналогичные встречающимся в реальном мире. Данное исследование может дать представление о том, какой вклад воздействие

зашумления может оказать на результаты распознавания в случаях реальных музыкальных композиций в реальных условиях использования.

Получив сведения о реакциях системы на музыкальные композиции, состоящие из эталонных музыкальных событий в дальнейшем очевидно станет целесообразным переход к композициям, которые, с одной стороны, были получены в результате настоящего проигрывания настоящими музыкантами с одной стороны, а с другой – в результате такого живого исполнения были зашумлены в результате реальных процессов записи звука, а также реальных процессов его воспроизведения. Такими процессами звукозаписи могут быть неизбежно возникающие в результате шума, а процессами передачи звука, например, помехи, добавляемые в результате передачи через радиоэфир или телеэфир.

Кроме того, вариации исполнения также оказывают значительное влияние.

Практика музыкального искусства наглядно показывает, что наиболее нравящимися слушателям являются не произведения, в которых каждый звук одной и той же ноты одного и того же инструмента играет без малейших изменений (что при живом исполнении вообще недостижимо физически), а такие, в которых исполнитель при проигрывании своей партитуры добавляет различные оттенки, изменения в звучание. С точки зрения формального музыковедения такие произведения являются идентичными, однако с точки зрения слушателя «живое» исполнение является более предпочтительным, чем «механическое».

Кроме того, если касаться применения системы распознавания в качестве одного из компонентов обучающей системы для обучения игры на музыкальных инструментах, то возможен вариант нарушения темпа неопытным музыкантом, что также создаёт трудности для процесса распознавания с целью перевода проигранной композиции в стандартную музыкальную нотацию.

Для этого было выбрано произведение С. Шнурова «привет Морриконе», реально полученное в результате проигрывания её исполнителями. Особенностью данной композиции является наличие простой аккордовой мелодии, играемой на низких частотах с помощью скрипки, в то время как на высоких частотах с помощью пианино исполняется отдельная партия. Данный случай представляется наиболее легким в виду того, что количество инструментов, задействованных в исполнении данной мелодии довольно мало, с другой стороны, это реальное произведение, а композиций, подобных этой, существует довольно большое количество. Кроме того, она, как полученная в результате живого исполнения содержит на высоком уровне.

3.2 Проведение вычислительного эксперимента и анализ его результатов

После того, как был вычислительный эксперимент был спланирован, необходимо осуществить переход к его выполнению.

Для его проведения использовалась компьютерная система с конфигурацией Intel Core i5 2.67 ГГц 3.86 Гб ОЗУ.

В соответствии с планом эксперимента сначала было осуществлено распознавание мелодии с паузами между сигналами. В качестве модельной модели была выбрана простая мелодия «до-ре-ми-фа-соль-ля-си»

Результаты эксперимента представлены в таблице 1.

Таблица 1

№ испытания	Масимальная амплитуда шумовой компоненты, % от амплитуды	Результат распознавания, % верно распознанных нот от оригинальной композиции
1	0	100
2	1	91
3	2	83
4	5	83
5	10	41
6	30	33

Как можно видеть из представленной таблицы, разработанный алгоритм показывает хорошие результаты даже для шумов, величиной в 10% зашумления от максимальной амплитуды сигнала.

В дальнейшем было исследовано воздействие импульсного шума на результаты распознавания той же монофонической мелодии с паузами.

Таблица 2

№ испытания	Масимальная плотность шумовой компоненты, % единицы времени исходного сигнала	Результат распознавания, % верно распознанных нот от оригинальной композиции
1	0	100
2	1	91
3	2	91
4	5	91
5	10	50
6	30	25

Таким образом была исследована реакция системы на самый простой из всех возможных случаев музыкальной композиции – на простую монофоническую мелодию, проигрываемую с паузами между нотами.

В дальнейшем было осуществлено исследование воздействия шумов на результаты распознавания музыкальной композиции, содержащий простую монофоническую мелодию без пауз между ними. Согласно плану исследования, разработанная система была протестирована на воздействие белого шума.

Результаты вычислительного эксперимента представлены в таблице 3.

Таблица 3

№ испытания	Масимальная амплитуда шумовой компоненты, % от максимальной амплитуды исходного сигнала	Результат распознавания, % верно распознанных нот от оригинальной композиции
1	0	100
2	1	91
3	2	83
4	5	83
5	10	41
6	30	33

Результаты исследования воздействия импульсного шума представлены в таблице 4.

Таблица 4

№ испытания	Масимальная плотность шумовой компоненты, %	Результат распознавания, % верно распознанных нот от оригинальной композиции
1	0	100
2	1	91
3	2	91
4	5	91
5	10	50

Как можно видеть из представленной таблицы, качество распознавания не ухудшилось, что говорит о том, что исключение пауз между играемыми звуковыми событиями не оказывает сильного влияния на результат распознавания и свидетельствует о качестве разработанной системы.

В дальнейшем было исследовано воздействие белого и импульсного шума на идеальную аккордовую композицию. В качестве данной композиции были применена композиция «до-ре-ми-фа-соль», играемая образцами из библиотеки музыкальных событий с помощью скрипки на двух разных октавах: на первой и на второй. Эксперимент был проведён пять раз, результаты его представлены в таблице

Результаты воздействия белого шума представлены в таблице 4.

Таблица 5

№ испытания	Максимальная амплитуда шумовой компоненты, % от максимальной амплитуды исходного сигнала	Результат распознавания, % верно распознанных нот от оригинальной композиции
1	0	95
2	0.1	83
3	1	43
4	2	21

Таким образом, выяснено, что с распознавание аккордовых композиций система справляется хуже, чем с распознаванием монофонической мелодии, однако при низких шуме результат находится примерно на уровне разработок других авторов.

В дальнейшем была исследована реальная аккордовая композиция, которая показала результат распознавания в 68% (вставок и удалений нот по сравнению с оригинальной композицией).

В дальнейшем было проведено исследование композиции С. Шнурова «Привет Морриконе», в результате был получен результат в 61% соответствия.

3.3 Сравнение полученных результатов с известными алгоритмами и методами

Для сравнения полученных результатов было выбрано три описанных в литературе подхода, предложенные следующими авторами:

- подход скрытых марковских моделей, описанный в работе [13] (Raphael C.)
- нейросетевой подход, описанный в работе [12] (Bello J. P. et al.)
- модель событий со скрытой марковской моделью [11] (Eronen A., Klapuri A.)

Также

Приведём следующую таблицу сравнительного анализа разработанной программной системы и других методов и алгоритмов.

Результаты представлены в таблице 6.

Таблица 6.- Сравнительный анализ разработанной системы

Критерий сравнения	Степень соответствия оригинальной композиции, %				
	Разработанная система	C. Raphael	Bello J. P.	Eronen A., Klapuri A.	WIDI
Распознавание монофонической мелодии с па	100	100	-	-	100
Распознавание простой монофонической мелодии без пауз	100	80	-	-	100
Распознавание простой аккордовой мелодии	64	44	71	78	71
Распознавание реальной мелодии (2 скрипки + 1 пианино)	61	-	-	-	69

ЗАКЛЮЧЕНИЕ

В результате выполнения задач, поставленных в рамках данной магистерской выпускной квалификационной работы были разработаны и реализованы в программной системе методы и алгоритмы, пригодные для успешного распознавания музыкальных произведений.

Было подтверждено, что теория активного восприятия пригодна для построения распознающих систем, предназначенных для автоматического распознавания музыкальных произведений.

В качестве разработанного метода был выбран метод построения библиотеки, позволивший получить удовлетворительные результаты распознавания для различных музыкальных инструментов.

Библиотека содержит экземпляры образцов проигрывания, что позволяет осуществлять, во-первых, распознавание звуков, а во-вторых, приводит к улучшенному качеству распознавания полифонических композиций.

Полученная программная система позволила распознавать полифонические композиции с возможностью распознавания типа инструментов, которым проигрывается та или иная партия.

Также был реализован механизм, который позволяет осуществлять приведение музыкального произведения к классическому виду, что даёт возможность представления распознанного музыкального произведения в виде традиционной нотной записи.

Разработанная библиотека обладает возможностями расширения, что позволяет добавлять новые образцы распознавания и таким образом обучать разработанную программную систему распознаванию музыкальных композиций, которые проигрываются различным самым широким типом новых инструментов. Распознавание этих типов осуществляется с помощью выделения таких характеристик музыкальных инструментов, как обертоновый портрет и темпоральный портрет атаки.

Было произведено исследование полученных результатов, в частности, разработанные алгоритмы были исследованы на предмет устойчивости к различным зашумлениям и другим различным воздействиям. Было показано наличие устойчивости разработанных алгоритмов к данным аддитивным шумам, что свидетельствует о хорошей практической перспективности разработанной программной системы.

В качестве потенциальных возможностей для развития разработанной программной системы можно назвать перспективу, заключающуюся в увеличении библиотеки музыкальных звуков, а также возможность распознавания песенных композиций.

Список литературы

1. Утробин, В. А. Компьютерная обработка изображений. Анализ и синтез: учеб. пособие /В.А. Утробин; НГТУ. –Нижний Новгород, 2003. – 228с
2. Утробин В.А. Физические интерпретации элементов алгебры изображения // Успехи физических наук, Т. 174, №10, 2004, С.1089–1104.
3. Утробин В. А. Элементы теории активного восприятия изображений //Труды Нижегородского государственного технического университета им. Р.Е. Алексеева. – 2010. – Т. 81. – №. 2. – С. 61-69.
4. Десятников И. Е. Поиск изображений по визуальному содержанию в графических базах данных и сети Интернет //ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ. – 2013. – С. 89.
5. ДЕСЯТНИКОВ И. Е., УТРОБИН В. А. Алгоритмы поиска изображений в базах видеоданных //Компьютерная оптика. – 2011. – Т. 35. – №. 3.
6. Гай, В. Е. Метод оценки частоты основного тона в условиях помех // Цифровая обработка сигналов. 2013. № 4. С. 72–76.
7. Гай В. Е., Утробин В. А. Методика выделения в звуковом сигнале участков речевой активности//Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – №. 4.
8. Klapuri A. P. Automatic music transcription as we know it today //Journal of New Music Research. – 2004. – Т. 33. – №. 3. – С. 269-282.
9. Benetos E. et al. Automatic Music Transcription: Breaking the Glass Ceiling //ISMIR. – 2012. – С. 379-384.
10. Benetos E. et al. Automatic music transcription: challenges and future directions //Journal of Intelligent Information Systems. – 2013. – Т. 41. – №. 3. – С. 407-434.

11. Ryynänen M. P., Klapuri A. Polyphonic music transcription using note event modeling //Applications of Signal Processing to Audio and Acoustics, 2005. IEEE Workshop on. – IEEE, 2005. – С. 319-322.
12. Bello J. P. et al. Techniques for Automatic Music Transcription //ISMIR. – 2000.
13. Raphael C. Automatic Transcription of Piano Music //ISMIR. – 2002.
14. Benetos E., Dixon S. A shift-invariant latent variable model for automatic music transcription //Computer Music Journal. – 2012. – Т. 36. – №. 4. – С. 81-94.
15. Marolt M. A connectionist approach to automatic transcription of polyphonic piano music //Multimedia, IEEE Transactions on. – 2004. – Т. 6. – №. 3. – С. 439-449.
16. Андреев С. А. Определение моментов начала нот (онсетов) при анализе музыкальных произведений. // Сборник докладов Всероссийской конференции Математические методы распознавания образов-13 (ММРО-13). — Макс Пресс Москва, 2007. — С. 277–279.
17. Курганский Д. А. Определение темпа музыкального произведения методом конкурирующих гипотез. // Сборник докладов Всероссийской конференции Математические методы распознавания образов-13 (ММРО-13). — Макс Пресс Москва, 2007. — С. 343–345.
18. Downie J. S., Byrd D., Crawford T. Ten Years of ISMIR: Reflections on Challenges and Opportunities //ISMIR. – 2009. – С. 13-18.c
19. Рубинштейн С. Л. Основы общей психологии. – Издательский дом "Питер", 1999 ; Schnupp J., Nelken I., King A. Auditory neuroscience: Making sense of sound. – MIT Press, 2011.
20. Bronkhorst, Adelbert W. "The Cocktail Party Phenomenon: A Review on Speech Intelligibility in Multiple-Talker Conditions" . *Acta Acustica united with Acustica* 86 - 2000 – С. 117–128.
21. Shinn-Cunningham, Barbara G. "Object-based auditory and visual attention". *Trends in Cognitive Sciences* - 2008 – С. 182–186.

22. Conway A. R. A., Cowan N., Bunting M. F. The cocktail party phenomenon revisited: The importance of working memory capacity // Psychonomic bulletin & review. – 2001. – Т. 8. – №. 2. – С. 331-335.
23. Алдошина И.А. Основы психоакустики. Оборонгиз., Москва, 2000. - С. 154.
24. Орлов А.С. Алгоритм распознавания музыкальных мелодий в условиях априорной неопределенности // Информационные системы и технологии. - НГТУ. 2015.
25. Орлов А.С., Гай В.Е. Система распознавания музыкальных произведений // Информационные системы и технологии. - НГТУ. 2016.