Федеральное государственное бюджетное образовательное
учреждение высшего образования
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ
УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ (ТУСУР)

На правах рукописи

Якимук Алексей Юрьевич

АЛГОРИТМЫ АНАЛИЗА ЧАСТОТЫ ОСНОВНОГО ТОНА ВОКАЛЬНОГО ИСПОЛНЕНИЯ

Специальность 05.13.17 – Теоретические основы информатики

Диссертация на соискание учёной степени кандидата технических наук

Научный руководитель:
доктор технических наук, профессор
Шелупанов А.А.

Томск 2019

Содержание

Введение

Актуальность темы. Популярность программных средств в задаче обучения конечного пользователя определенным навыкам растет с каждым днем. Сфера речевых технологий также относится к данному высказыванию. Применение специализированных программ способно помочь в обучении иностранным языкам или в выполнении упражнений для развития вокальных навыков. Общее число учебных заведений, реализующих программу в области музыкального искусства превышает отметку в 3000 [1]. Существующая форма обучения вокалу осуществляется в взаимодействии с репетитором. Чтобы обучение было максимально эффективным, необходимо проводить не менее 2 часов занятий в день [2], что является сложной задачей для большинства учеников. Самостоятельное выполнение упражнений редко способно развить музыкальный слух, а индивидуальные занятия с преподавателем ограничены его высокой загруженностью с другими учениками.

Самой распространенной методикой обучения вокалу является использование в качестве упражнения пение по нотам (сольфеджио). В связи с тем, что отсутствие развитого музыкального слуха не позволяет проведение оценки правильности исполнения ноты (в том числе степени отклонения от ее идеального звучания), самостоятельная практика сольфеджио может быть малоэффективной. Эту проблему можно решить с использованием специализированного программного средства, позволяющего в режиме реального времени предоставлять пользователю информацию о качестве исполнения выданного задания (количество правильно исполненных нот, точность исполнения нот с точки зрения высоты звучания и др.).

Основной тон содержит в себе информацию об интонационной структуре произнесения, индивидуальности голоса диктора и его эмоциональном состоянии, возрастных и патологических изменениях голосового аппарата. Системы, идентифицирующие исполненную ноту, основаны на алгоритмах вычисления частоты основного тона (ЧОТ). Алгоритмами вычисления ЧОТ речевого сигнала занимались такие ученые как А. Асеро, В.П. Бондаренко, А.А. Карпов, Л. Рабинер, А.Л. Ронжин, М.М. Сондхи, Г. Фант, М.В. Хитров, Л.А. Чистович, М. Шрёдер и

многие другие. Следует отметить, что существующие алгоритмы не позволяют вычислить значение фундаментальной частоты в вокальном исполнении с высокой точностью за счет наличия высокого процента грубых ошибок в них и ограничены узким спектром охватываемых частот. Большинство алгоритмов разрабатывались с целью анализа речевой информации, что накладывает ограничение в виде верхней границы определения ЧОТ, равной 400 Гц. Однако, во время пения частота звучания речевого сигнала может быть гораздо выше, что делает неприменимыми алгоритмы, ограниченные диапазоном для обработки речи. Также неприменимы алгоритмы, обладающие высоким процентом грубых ошибок, для идентификации звучащей ноты. Ошибка в частоте порядка 20% от ее значения может привести к промаху более чем на 3 ноты. Наличие таких ограничений делает неприменимыми существующие решения по идентификации нот в задаче обучения вокалу с помощью программных средств.

Как и в остальных задачах, решаемых исследованиями в области речевых технологий, ключевое место в данном исследовании занимает точность сегментации. Сегментация подразумевает выделение участков сигнала, соответствующих структурным единицам речевого сигнала. В случае, если за единицу принимать фонему, то сегментация будет определять переходы между фонемами. Таким образом, выбрав в качестве единицы спетую диктором ноту, можно применить сегментацию для определения их границ. Отметим, что в задаче идентификации нот сегментация необходима не только на этапе выделения вокализованных и невокализованных участков речевого сигнала, освещенной в таких работах, как [3-7]. Для решения задач обучения вокалу или получения партитуры на основании спетой последовательности нот сегментация также носит ключевой характер в вопросе определения длительности исполненной ноты. В некоторых упражнениях перед учениками ставится задание спеть ноты в определенном порядке или промежуток времени. В таком случае алгоритм сегментации может помочь в выставлении оценки для данных заданий. Особое внимание сегментации речевого сигнала в своих работах уделяли В.П. Бондаренко,

Т.К. Винцюк, Р.В. Шафер, Л.В. Златоустова, Р.К. Потапова, В.Н. Трунин-Донской, Л.В. Бондарко, Л.А. Вербицкая, Т.В. Шарий и многие другие.

Целью диссертационной работы является повышение качества распознавания звучащих нот в вокальном исполнении за счёт применения модели слуховой системы человека.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1) выполнить анализ текущего состояния предметной области: изучить существующие методы и алгоритмы распознавания нот, в том числе определения частоты основного тона сигнала;

2) модифицировать модель слуховой системы человека с точки зрения увеличения охватываемого диапазона определения частот основного тона;

3) разработать алгоритм сегментации и идентификации нот и определить способ оценки качества пения;

4) реализовать и апробировать программный комплекс по определению нот вокального исполнения.

Объектом исследования данной работы речевой сигнал вокального исполнения последовательности нот.

Предметом исследования является выделение последовательности спетых нот на основе частоты основного тона.

Методы исследования. Для решения задач, сформулированных в работе, использовались методы моделирования, системного анализа, цифровой обработки сигналов, математической статистики.

Достоверность результатов обеспечивается результатами проведенных численных экспериментов с использованием реальных данных, а также путём сопоставления результатов, полученных в диссертации, с результатами экспертной оценки.

Научная новизна результатов работы и проведенных исследований заключается в следующем:

1) Проведена модификация модели слуховой системы человека, позволившая расширить диапазон частот в 2 раза по сравнению с исходной моделью и отличающаяся возможностью произвольного указания границ определения тона.

2) Предложен алгоритм создания шаблонов для обнаружения частоты основного тона, отличающийся возможностью автоматической генерации наборов шаблонов с произвольным заданием граничных частот определения основного тона.

3) Разработан алгоритм распознавания нот, учитывающий минимальную длительность звучания нот и отличающийся учетом особенностей слуховой системы человека.

Теоретическая значимость работы заключается в развитии методов анализа частоты основного тона речевого сигнала. Модификация математической модели слуховой системы человека позволила расширить диапазон определения частот основного тона до 800 Гц. Улучшенный алгоритм идентификации частот основного тона речевого сигнала может быть также применен в исследовании параметров речевого сигнала.

Практическая значимость работы подтверждается использованием полученных в ней результатов для решения практических задач:

– автоматическое определения нот в вокальном исполнении;

– проведения оценки качества вокального исполнения заданного упражнения. Результаты внедрены в деятельность «Элекард-ЦТП» в рамках проекта по дистанционному обучению вокалу в формате видеоконференций.

Разработанные алгоритмы и методика использованы при выполнении проектной части государственного задания Министерства образования и науки Российской Федерации на 2017-2019 гг. № 2.3583.2017/4.6. Часть исследований проводилась при поддержке стипендии для акустиков – студентов и аспирантов из России, полученной от Американского акустического общества (Приложение В).

На защиту выносятся приведенные ниже положения.

1) Модифицированная модель слуховой системы человека, позволившая произвольно указывать границы определения тона и идентифицировать частоты

основного тона на диапазоне до 800 Гц с относительной погрешностью в указанном диапазоне не более 1%.

Соответствует пункту 5 паспорта специальности: Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях. разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений.

2) Алгоритм автоматизированного создания шаблонов для обнаружения частот основного тона, позволивший автоматически генерировать наборы шаблонов для произвольных диапазонов её поиска.

Соответствует пункту 5 паспорта специальности: Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях. разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений.

3) Алгоритм распознавания нот, позволивший определить не менее 95% спетых диктором нот.

Соответствует пункту 5 паспорта специальности: Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях. разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений.

Внедрение результатов диссертационного исследования. Результаты диссертационной работы внедрены в деятельность «Элекард-ЦТП» в рамках проекта по дистанционному обучению вокалу в формате видеоконференций.

Результаты диссертационной работы по исследованию слуховой системы человека используются в практических занятиях по дисциплинам «Моделирование автоматизированных информационных систем» и «Системный анализ» на факультете безопасности ТУСУР.

Апробация работы. Основные положения работы докладывались и обсуждались на следующих конференциях:

– XII Всероссийская научно-практическая конференция студентов, аспирантов и молодых ученых «Технологии Microsoft в теории и практике программирования» (ТПУ, г. Томск, 2015);

– Всероссийская научно-техническая конференция студентов, аспирантов и молодых ученых «Научная сессия ТУСУР» (ТУСУР, г. Томск, 2015, 2016);

– Международная научно-практическая конференция «Электронные средства и системы управления» (ТУСУР, г. Томск, 2015, 2016, 2018);

– XII Всероссийская научная конференция молодых ученых «Наука. Технологии. Инновации» (НГТУ, г. Новосибирск, 2018);

– III Всероссийская научно-практическая конференция «Информационные технологии в экономике и управлении» (ДГТУ, г. Махачкала, 2018);

– XI Всероссийская научно-практическая конференция «Проблемы управления качеством образования» (ПГАУ, г. Пенза, 2018);

– III Международная научно-практическая конференция «Проблемы и перспективы современного физико-математического, информационного и технологического образования» (Новокузнецкий институт КемГУ, г. Новокузнецк, 2019);

– XVI Международная конференция студентов, аспирантов и молодых ученых «Перспективы развития фундаментальных наук» (г. Томск, 2019);

– VII молодежная конференция «Математическое и программное обеспечение информационных, технических и экономических систем» (ТГУ, г. Томск, 2019)

– Томский IEEE семинар «Интеллектуальные системы моделирования, проектирования и управления».

Были получены 2 свидетельства о государственной регистрации программ для ЭВМ:

– Конев А.А., Якимук А.Ю., Осипов А.О. «Программный комплекс по определению нот вокального исполнения», свидетельство о государственной регистрации программы для ЭВМ №2017664232 от 19.12.2017;

– Конев А.А., Якимук А.Ю. «Программа для определения качества сегментации речевых сигналов», свидетельство о государственной регистрации программы для ЭВМ №2017664235 от 19.12.2017.

Публикации по теме диссертации. По результатам исследований опубликовано 19 работ, из них 3 статьи в журналах, входящих в перечень рекомендованных ВАК журналов, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, 14 публикаций в материалах международных и всероссийских научных конференций.

Личный вклад автора. Основные научные результаты получены лично автором. Автором был осуществлен анализ возможности модификации модели слуховой системы человека, разработка новых методов и алгоритмов, позволяющих получать результаты на большем диапазоне частот. Разработанные методы и алгоритмы были реализованы в виде комплекса программ также лично автором. Постановка задачи исследования осуществлялась научным руководителем д.т.н., профессором Шелупановым А.А.

Структура и объем работы. Диссертационная работа содержит введение, 4 главы, заключение, приложение и список источников из 157 наименований. Объем диссертационной работы 121 страницу, в том числе 12 таблиц и 53 рисунка.

Во введении обосновывается актуальность темы исследования, формулируется цель работы, излагаются полученные автором основные результаты проведенных исследований, показывается их научная новизна, теоретическая и практическая значимость, отражаются основные положения, выносимые на защиту.

В первой главе производится обзор проблемы исследования. Описываются алгоритмы анализа частоты основного тона, приводятся примеры применения алгоритмов вычисления частот основного тона сигнала к задачам, близким к анализу вокальных исполнений. Проводится обзор алгоритмов сегментации и их

роли в речевых технологиях. Приводятся показатели для рассмотренных алгоритмов с оценкой на пригодность к определению нот в пении. Также проводится обзор публикаций по теме обучения вокалу с точки зрения формирования в студентах способности к пению с помощью программных средств. Приводятся результаты обзора программ-аналогов с указанием их особенностей.

Во второй главе описывается формирование наборов шаблонов для определения частот основного тона в вокальном исполнении. Описывается модифицированная модель слуховой системы человека. Показаны результаты проведенного тестирования работы алгоритма идентификации частот основного тона на сгенерированных синусоидальных сигналах.

В третьей главе описывается разработанная методика распознавания нот вокального исполнения. Приводятся алгоритмы сегментации и идентификации нот. Для этапа определения нот обоснован выбор вычисления границ звучания ноты. Описаны стратегии, по которым собраны аудиозаписи с пением. Показаны результаты тестирования алгоритмов на аудиозаписях.

В четвертой главе содержится описание разработанного программного комплекса. Приводятся результаты тестирования работы комплекса на аудиозаписях с различными подходами к вокальному исполнению.

1 Обзор существующих методов, подходов и алгоритмов анализа частоты основного тона

В данной главе рассматриваются методы, подходы и алгоритмы, применяемые в задачах анализа частоты основного тона, сегментации речевого сигнала или обучения пению.

1.1 Применение алгоритмов анализа частоты основного тона для исследования вокальных исполнений

Знание значения частоты основного тона (ЧОТ) сигнала в конкретный момент времени имеет важное значение во многих сферах речевых технологий. В таких задачах как идентификация дикторов значение ЧОТ имеет не ключевую роль, поскольку особый интерес играют форманты, отражающие индивидуальные особенности человека [8-10]. С другой стороны, следует отметить, что определение основной частоты речевого сигнала с учетом особенностей формирования речи и восприятия речи, связанных с анатомией и физиологией человека, крайне важно в сфере реабилитации для онкологических больных после резекции гортани [11]. У таких больных часто возникают побочные эффекты, затрудняющие разговор и общение. Во время логопедической терапии широко используется подход перцептивной оценки качества голоса. Ряд исследований направлен на применение программных средств для оценки голоса в рамках реабилитации [12-14].

Кроме того, в сферах, направленных на обработку музыки или сигналов, подобных музыкальным, ситуация складывается обратным образом. Разнообразие исследований, касающихся вопросов обработки такого типа сигналов, достаточно обширно.

Пение может быть рассмотрено как особая форма речи, которая создается таким же образом, но при этом присутствует дополнительный контроль для создания музыкального аспекта. Естественная мелодия речи (просодия) отличается в разных языках и определяет контур высоты тона, вариации громкости, ритм и темп выражения эмоций. В пении же высоту, громкость и тембр определяет в первую очередь композиция за счет того, что для соответствия продолжительности ноты гласные звучат дольше, чем обычно.

Речевые гласные характеризуются особыми позициями формант. В пении положение формант может быть радикально изменено путем изменения длины и формы голосового тракта и положения артикуляторов. Идентичность голоса во многом определяется физическими характеристиками системы производства вокала. Форма голосового тракта определяет форманты, причем две форманты низшего порядка (F_1-F_2) являются наиболее важными для разборчивости речи, а форманты высшего порядка (F_3-F_5) способствуют идентификации говорящего. Опытные певцы могут точно контролировать частоты трех нижних формант, варьируя первую форманту с помощью степени раскрытия челюсти, вторую – управлением формой языка и третью положением кончика языка. Разные вокалисты настраивают свои частоты формант по-разному для каждого гласного, причем наиболее яркие различия заключаются в женских голосах сопрано, где высокие значения высоты звука (1000 Гц) по сравнению с обычным значением первой форманты (500 Гц) будут определять смещением этой форманты близко к ЧОТ. Это может привести к потере разборчивости, но при классическом пении интонация и музыкальные качества голоса являются наиболее важным аспектом, а разборчивость - вторым.

Следует отметить, что не во всех исследованиях, в рамках которых осуществляется выделение характеристик из музыкального сигнала, преследуются цели по получению информации об исполненных последовательностях нот. Кроме того, не всегда в подобных исследованиях используются записи с пением человека или игрой на музыкальных инструментах. В некоторых исследованиях осуществляется обработка записей пения или моментов общения, воспринимаемых человеком как таковое, представителей животного мира. В исследовании [15] уделяется внимание определению степени влияния городского шума на различие в пении птиц. Учеными было осуществлено сравнение записей пения воробьев, полученных в городской среде и за пределами города. Статья [16] в свою очередь направлена на обнаружение сходства в вокализации общения мышей. А в работе [17] учеными были использованы сведения, полученные из записей с пением одного из видов бесхвостых, как индикатор изменения климата в их регионе

обитания. В исследовании [18] предметом для изучения стало пение белогорлых воробьев. Учеными было определено, что самцы, исполняющие больший диапазон нот в своих песнях, дольше выживают, чем их собратья. Данное открытие позволило сделать вывод о возможности использовании характеристик пения для определения физической формы птиц.

Определение особенностей в частоте основного тона речевого сигнала является важной задачей и в сфере исследования особенностей языка. Для некоторых языков и акцентов характерно наличие появление восходящего или нисходящего тона, который по своим характеристикам может быть схож с вокальным исполнением. В работе [19] рассматривается возможность применения выявления в речи восходяще-нисходящего тона в качестве маркера для принятия решения о наличии валлийского акцента у диктора. Авторами оценивались интонации на основании повышения или понижения голоса. Полученный разброс ЧОТ был получен в диапазоне от 169 до 358 Гц. Исследование [20] затронуло вопрос определения особенностей в речи молодежи. На основании оценки ЧОТ записей речевых сигналов было определено, что для молодежи характерно намеренное растягивание гласных. Автором [21] проводилось исследование высотно-мелодического параметра речи, а именно влияние возрастных изменений на ритмические характеристики языка. Было определено, что с увеличением возраста диктора проявляется снижение максимальных и минимальных значений ЧОТ. Одной из особенностей публикации [22] является определение на основании идентификации в речи характеристик формы мелодии эмоционального состояния говорящего. Автор обратил внимание на такие характеристики как направление, характер и диапазон движения тона. В данном исследовании применялись в измерении ЧОТ величины, используемые при исследовании музыкальных мелодий. Это позволило наглядно представить особенности изменений в речи диктора при проявлении эмоций или намерений. Аналогичная задача была поставлена в работе [23], где по изменению разницы в акустических значения средней и максимальной ЧОТ фразы определялась оценочность в речи репортеров с целью передачи атмосферы происходящего.

Возвращаясь к теме обработки мелодии, основанной на пении человека и игре на музыкальных инструментах, отметим следующие направления научных работ. Ряд исследований, таких, как [24-30], преследуют цель создания систем поиска музыкальной информации (Music Information Retrieval – MIR). Учеными преследуются такие цели как разработка автоматизированной системы для определения плагиата между двумя отрывками музыкальных воспроизведений и создание средств для удобного поиска музыки по задаваемым критериям. Перед разрабатываемыми системами ставятся такие задачи как обнаружение пения в музыкальном произведении, классификация жанра, идентификация певца и многие другие. Такие работы, как [31], посвящены вопросу обнаружения иерархической структуры ритма в аудиосигнале с помощью алгоритма обнаружения изменений аккордов. Данный подход не рассматривает задачи по идентификации звучащей ноты. В [32] для оценки схожести гармонии композиций используется графическая вероятностная модель, содержащая информацию об аккорде и ладе момента времени звучания композиции. Авторы в зависимости от стиля определяли вероятность использования аккорда в контексте определенного лада. Преобразование композиций в набор векторов ими осуществлялся на основании принятой в европейской музыкальной традиции идентификации нот относительно абсолютной начальной частоты. Как указано в [33], при обработке музыки возникает сложность с наличием гармоник у некоторых инструментов. Помимо ЧОТ основной ноты звучат другие частоты, которые могут соответствовать другим нотам. Кроме того, в музыке спектр звукового сигнала чаще всего является суммой спектров отдельных инструментов, где каждым из инструментов воспроизводятся звуки в разных полосах частот.

Обзор исследований по обработке музыки показал, что с точки зрения обработки сигнала мелодия может быть представлена последовательностями ЧОТ, определяемыми в моменты звучания, то есть на участках, где активен инструмент, создающий мелодию. Основные алгоритмы транскрипции мелодии обычно содержат 2 этапа обработки. Во-первых, вычисляется представление, подчеркивающее наиболее вероятные значения ЧОТ во времени, например, в

форме матрицы выдачи, матрицы активации источника голоса или расширенной спектрограммы. Во-вторых, двоичная классификация выбранных ЧОТ между мелодическим и фоновым содержанием выполняется с использованием обнаружения мелодического контура и голоса. Например, в [34] подход к обнаружению мелодических и басовых линий в звуковом сигнале основан на оценке относительного доминирования каждой возможной ЧОТ в виде функции плотности вероятности ЧОТ. Проведенный учеными эксперимент показал, что используемая ими система, осуществляющая в реальном времени обработку аудиозаписи, смогла обнаружить требуемую информацию только на 80% участках от всей длительности протестированных записей.

Таким образом, в основе понимания того, какая нота была исполнена певцом или музыкальным инструментом, лежит знание того, на какой частоте основного тона происходили колебания в данный отрезок времени. Для каждой ноты существует дискретное значение частоты (таблица 1.1) [35-36].

Таблица 1.1 – Соответствие нот частотам основного тона

Нота	Октава								
	Субконтроктава	Контроктава	Большая октава	Малая октава	1 октава	2 октава	3 октава	4 октава	5 октава
До	–	32.70	65.41	130.82	261.63	523.25	1046.50	2093.00	4186.00
До-диез	–	34.65	69.30	138.59	277.18	554.36	1108.70	2217.40	4434.80
Ре	–	36.95	73.91	147.83	293.66	587.32	1174.60	2349.20	4698.40
Ре-диез	–	38.88	77.78	155.56	311.13	622.26	1244.50	2489.00	4978.00
Ми	20.61	41.21	82.41	164.81	329.63	659.26	1318.50	2637.00	5274.00
Фа	21.82	43.65	87.31	174.62	349.23	698.46	1396.90	2793.80	–
Фа-диез	23.12	46.25	92.50	185.00	369.99	739.98	1480.00	2960.00	–
Соль	24.50	49.00	98.00	196.00	392.00	784.00	1568.00	3136.00	–
Соль- диез	25.95	51.90	103.80	207.00	415.30	830.60	1661.20	3332.40	–
Ля	27.50	55.00	110.00	220.00	440.00	880.00	1720.00	3440.00	–
Си- бемоль	29.13	58.26	116.54	233.08	466.16	932.32	1864.60	3729.20	–
Си	30.87	61.74	123.48	246.96	493.88	987.75	1975.50	3951.00	–

Кроме того, из теории музыки [37] известно разделение певческих голосов в зависимости от исполняемых нот. Женские голоса включают в себя диапазон от ноты «фа малой октавы» (174.62 Гц), соответствующий самому низкому женскому певческому голосу – контральто, до ноты «соль-диез третьей октавы» (1661.20 Гц), соответствующий самому высокому женскому певческому голосу – колоратурному сопрано. Мужские голоса включают в себя диапазон от ноты «фа контроктавы» (43.65 Гц), соответствующей нижней границе баса-профундо – самого низкого мужского певческого голоса, до ноты «ми второй октавы» (659.26 Гц), соответствующей верхней границе самого высокого мужского певческого голоса – контратенора. В результате можно выделить диапазон частот, охватываемых в пении профессиональными оперными певцами – отрезок от 43.65 Гц до 1661.2 Гц. Однако, основываясь на данных о самых распространенных мужском (баритон) и женском (лирическое сопрано) [38] певческих голосах, исследуемый диапазон можно ограничить отрезком от 110 до 1318.5 Гц.

Следует отметить, что определенный диапазон для идентификации нот не охватывался учеными в рамках исследований ЧОТ. Большинство работ, посвященных алгоритмам вычисления значений ЧОТ сигнала в определенный момент времени, направлены на изучение речи. Примерами таких исследований могут служить [39-41]. При этом подавляющее большинство алгоритмов определения частоты основного тона используют вычисление периода основного тона по пикам речевого сигнала [42-44]. Основной интерес для [45] представляет разработка системы для коррекции слуха, что может объяснить ограничение применяемого в работе алгоритма отслеживания мгновенного значения ЧОТ (Instantaneous Robust Algorithm for Pitch Tracking – IRAPT) диапазоном от 50 до 500 Гц. В основе предложенного ими алгоритма лежит применение алгоритма RAPT [46]. В ходе эксперимента этой группе ученых удалось достичь для разработанного алгоритма процента грубых ошибок (gross pitch error - GPE) 1.625% для мужского голоса и 3.777% для женского, а также среднего процента мелких ошибок (mean fine pitch error - MFPE) 1.608% для мужского голоса и 0.977% для женского. В другой своей работе [47] эти же авторы предлагали другой алгоритм (оценка

мгновенной ЧОТ на основе многоскоростной обработки), где диапазон поиска частоты основного тона также был ограничен – от 100 до 450 Гц, что сказалось на проценте ошибок. В новой версии алгоритма GPE составил 0.743% для мужского голоса и 3.6% для женского, а MFPE – 1.268% для мужского голоса и 1.039% для женского. В исследовании [48] был предложен алгоритм PEFAC (Pitch Frequency Estimation Algorithm Robust to High Levels of Noise), позволяющий проводить оценку значения ЧОТ в условиях высокого шума с высокой точностью. Данный алгоритм также был изучен только на диапазоне до 400 Гц. С учетом специфики данного алгоритма, следует отметить его показатель вокализованности (Voicing Decision Error, VDE), который определяет количество правильно распознанных вокализованных фреймов, равный в среднем 46.45%, что превышает точность работы аналогов. При этом GPE для алгоритма при различных типах шумов составляет 40% при бульканье, 20% при шуме машин и 24% при белом шуме. Еще одним из алгоритмов, позволяющим выделить в сигнале значение частоты для основного тона, является SWIPE (Sawtooth Inspired Pitch Estimator) [49]. Как следует из названия алгоритма, анализ происходит по спектру пилообразного сигнала. Алгоритм является интегральным и показывает хорошие результаты в условиях низкого шума. Верхнее допустимое значение ЧОТ для алгоритма SWIPE – 500 Гц. Испытания алгоритма показали величины ошибок GPE 32.92 % и MFPE 4.45% при частотах близких к 500 Гц, что может быть вызвано особенностью реакции алгоритма на наличие субгармоник. Ошибки 1-го и 2-го рода при этом составляют 2.77% и 18.41% соответственно. Среди прочих стоит выделить алгоритм YIN [50], получивший свое название от одного из элементов философии – инь. Алгоритм основывается на использовании метода автокорреляции и, как заявлено авторами, не имеет верхней границы для определения ЧОТ сигнала и имеет процент грубых ошибок, равный 1.03%. По этой причине авторами алгоритм предлагается к использованию не только для речевых, но и для музыкальных сигналов. Однако, показанный в [47] сравнительный анализ перечисленных выше алгоритмов показал, что при обработке сигналов с частотами основного тона в

диапазоне от 100 до 350 Гц, GPE и MFPE у алгоритма YIN выше, чем у аналогов и составляют 3.96% и 1.389% соответственно.

В представленных выше исследованиях величина GPE показывает отношение количества анализируемых фреймов с отклонением полученной оценки ЧОТ более чем на ±20% от настоящего значения ЧОТ к общему числу вокализированных фреймов [51] и оценивается по формуле 1.1.

$$GPE(\%) = \frac{N_{GPE}}{N_v} \cdot 100, \tag{1.1}$$

где N_{GPE} – количество фреймов с отклонением полученной оценки более чем на ±20% от настоящего значения основного тона;

N_v – общее число вокализированных фреймов.

Как указывается в [52], погрешность в пределах 20% варьируется в пределах октавы, а разумность применения алгоритмов с таким количеством ошибок при оценивании ЧОТ определяется исключительно конечной целью работы. Однако, в вопросах касающихся идентификации нот данная ошибка приводит к смещению в результате не менее чем на 3 ноты.

Оценка MFPE, показывающая средний процент мелких ошибок, появляющихся при оценивании ЧОТ без учета грубых ошибок, вычисляется по формуле 1.2.

$$MFPE(\%) = \frac{1}{N_{FPE}} \cdot \sum_{n=1}^{N_{FPE}} \frac{|F0_{true}(n) - F0_{est}(n)|}{F0_{true}(n)} \cdot 100 \tag{1.2}$$

где N_{FPE} – число вокализированных фреймов без грубых ошибок;

$F0_{true}(n)$ – действительные значения ЧОТ;

$F0_{est}(n)$ – оценочные значения ЧОТ.

Если обратить внимание исключительно на пение певцов, то здесь также существует ряд направлений, связанных с особенностями исполнения. Существует широкий спектр областей оценки певческих голосов: определение типа, оценка дикции и орфоэпии, оценка тембровых качеств и др. [53]. Для работ [54-58] в качестве области исследований выбрано изучение вибрато и подобных ему особенностей пения. Авторами уделяется внимание частоте изменения ЧОТ в

секунду. По определению признаком присутствия в пении вибрато являются колебания в пределах полутона (соседние ноты) [54] с частотой колебаний от 5 до 7 в секунду [59]. Помимо вибрато, являющегося показателем профессионализма певца и придающего вокальному исполнению уникальность, которая позволяет выделить голос от фоновой музыки, наличие колебаний в исполняемой ноте может свидетельствовать о таких эффектах, как тремоляция и «качание звука» [60]. Считается, что к наличию вибрато в голосе склонны высокие голоса и наиболее распространен прием среди обладательниц сопрано [61-62]. Этот факт осложняет задачу определения звучащей ноты. С учетом выше изложенных ограничений в диапазоне существующих алгоритмов погрешность определения ЧОТ пропетой ноты может быть высока. В свою очередь, автоматическое определение частоты изменения ЧОТ в секунду с целью обнаружения тремоляции в пении также может осуществляться с большим числом ошибок 1-го и 2-го рода.

Особый интерес для анализа вокальных исполнений часто представляют записи, сделанные профессиональными оперными певцами. Примерами могут служить такие работы как [63-65]. В этих исследованиях изучается влияние выражаемых эмоций на речевой сигнал. Анализу эмоций в речи диктора уделяют свое внимание также и в таких работах, как [66-68], однако в них в качестве материала для исследования используется разговорная речь. Профессиональные оперные певцы выбираются по причине их способности выражать аутентичную эмоцию слушателям. В работе [69] в качестве объекта для изучения также были выбраны аудиозаписи вокального исполнения, в которых профессиональных оперных певцов попросили спеть последовательности нот, выражая определенные эмоции. Авторами исследования преследовалась цель обнаружения характеристик в записях, которые позволят однозначно идентифицировать эмоцию диктора. В предложенном наборе параметров, рекомендуемом для определения типа эмоции, присутствует частота основного тона. Анализ аудиозаписей показал, что, например, для ярости или гнева характерно низкое разнообразие частот основного тона, но высокая громкость речи. Несмотря на то, что фундаментальная частота оказалась включена в набор параметров, авторами было определено, что частота

основного тона играет второстепенную роль. При этом, авторы отмечают важность ЧОТ как маркера возбуждения и дали объяснение полученному результату. Певцы должны были придерживаться частот, предписанных партитурой, и не могли варьировать ее для выражения эмоций. Стоит отметить, что данное ограничение вызвано не требованиями исследования, а пением в целом. В нормальных условиях композитор будет использовать вариацию частот основного тона при написании мелодий. Однако, как показали результаты исследования, певцы могут вносить небольшую вариацию в частотах основного тона для поддержания эмоциональной интерпретации содержимого. Таким образом, наличие и отсутствие изменений в генерации частоты основного тона также способно служить маркером для определения эмоций. Анализ акустических параметров, связанных с воспринимаемыми эмоциями, показал, что гнев был связан с присутствием вибрато в пении, в то время как грусть характеризовалась отсутствием вибрато.

Проведенный анализ алгоритмов анализа частоты основного тона (таблица 1.2) показал, что существующие алгоритмы в текущем виде не могут быть применены в задаче обработки вокальных исполнений. В первую очередь это вызвано ограничением в диапазоне работы алгоритмов. Как было указано ранее, во время пения колебания ЧОТ в речевом сигнале могут происходить на частотах до 1400 Гц. Во-вторых, существующие алгоритмы имеют высокий показатель ошибок, что делает невозможным точное определение звучащей ноты. Помимо описанных выше алгоритмов в таблицу был включен алгоритм, основанный на математической модели слуховой системы человека [70]. Ограничение в частотном диапазоне для него определяется применением в рамках исследования параметров сигнала в слитной речи и при онкологических заболеваниях.

Таблица 1.2 – Показатели алгоритмов анализа частоты основного тона

Алгоритм	Показатели алгоритма
IRAPT	Процент грубых ошибок 1.62% для мужского и 3.77% для женского голоса. Диапазон работы алгоритма: от 50 до 500 Гц.
PEFAC	Диапазон работы алгоритма: до 400 Гц. Количество правильно распознанных фреймов равно 46.65%. Процент грубых ошибок не менее 24%.

Продолжение таблицы 1.2

SWIPE	Диапазон работы алгоритма: до 500 Гц. Процент грубых ошибок: 32.92% при частотах близких к 500 Гц.
YIN	Диапазон работы алгоритма: от 100 до 350 Гц. Процент грубых ошибок: 3.96%.
Основанный на математической модели слуховой системы человека	Диапазон работы алгоритма: от 70 до 400 Гц. Погрешность определения ЧОТ составляет не более 0.6%.

1.2 Применение алгоритмов сегментации при исследовании вокальных исполнений

Под сегментацией понимается действие по разбиению чего-то целого на набор сегментов. В сфере автоматической обработки речи сегментация может применяться в контексте синтеза или распознавания речи [71-73] с целью изучения динамики сигнала [74] или для определения ключевых слов. Например, в [75] проводится оценка разборчивости слов в рамках решения задач реабилитации речи после комплексного лечения онкологических заболеваний. Таким образом, сегментация является ключевым инструментом в области обработки речевой информации. Многие исследования в ней требуют наличия аудиоинформации и синхронизированной с ней фонетической транскрипции. Существует два подхода к сегментации речевого сигнала.

Первый подход заключается в анализе разрывов сигналов без учета лингвистических знаний о содержимом сообщения (например, орфографическая или фонетическая транскрипция). Такой подход называется неявной сегментацией и не зависит от диктора. По причине отсутствия дополнительной информации, первая фаза при использовании такого подхода заключается в анализе акустических характеристик, присутствующих в сигнале. Данная группа алгоритмов является активно развивающимся направлением в области сегментации и в некоторых работах носит название обработки голосовых сигналов в условиях нулевых ресурсов [76]. Ранние работы в этом направлении акцентировали внимание на идентификации изолированных повторяющихся структур в речевом корпусе, в то время как более современные исследования

полного покрытия преследуют цель полной сегментации и кластеризации звука в словоподобные участки. В работе [77] с этой целью применяется байесовская модель с сегментными представлениями слов: каждый сегмент слова представлен в виде фиксированного акустического вложения, полученного путем сопоставления последовательности характерных кадров с одним вектором вложения.

Второй подход, соответственно, учитывает языковые особенности в речевом сигнале и называется явной сегментацией. При таком подходе полученные сегменты определяются фонетической транскрипцией. Определению фонетических особенностей речевого сигнала посвящены исследования [78-82]. Обычно в алгоритмах данной категории применяют скрытые марковские модели [83]. Примерами использования скрытых марковских моделей в задаче сегментации речи могут служить такие работы как [84, 85]. В [84] предлагается использование моделей для автоматической фонетической сегментации в условиях недостаточного количества данных. Авторами рассматривается эффективность искусственного увеличения начального количества материала, обработанного вручную для тренировки.

Заинтересованность исследователей в разработке системы, способной автоматически сегментировать речевой сигнал с результатами, близкими к ручной сегментации, объясняется высоким расходом ресурсов на выполнение последней. В работе [86] упоминается, что обработка аудиозаписи длительностью в 30 секунд может потребовать не менее часа сосредоточенной работы эксперта. Зачастую в системах, направленных на автоматическую сегментацию сигнала, используются предварительно обученные модели, независимые от диктора. Существенным недостатком является тот факт, что они охватывают очень ограниченное количество языков и могут не работать должным образом для разных стилей речи.

Большинство алгоритмов автоматической сегментации речевого сигнала достигают схожих показателей с точки зрения правильного определения границ в процентах [87]. Результаты экспериментов для них существенно ниже, чем полученные экспертами вручную эталонные данные по транскрипции.

Статистический анализ показал, что в основном ошибки связаны с неизвестностью длительности сегмента и наличием последовательностей схожих сегментов. Авторами была проверена гипотеза, показавшая, что акустические изменения являются довольно хорошим индикатором границ сегментов: более двух третей предполагаемых границ совпадают с границами сегментов. По сравнению с предварительно обученными моделями адаптированные модели более точно характеризуют акустические свойства обрабатываемых сигналов. Это означает, что может быть достигнута более высокая точность сегментации. Авторами [88] была рассмотрена возможность адаптации модели сегментации музыки за счет использования меры доверия, получаемой на основе вероятностей распознавания и используемой для выбора достоверных данных. Авторами [89] также отмечалась острая необходимость практического подхода к анализу структуры. В своем исследовании они предложили подход к автоматической сегментации музыки для извлечения фраз и предложений музыкальной структуры, при котором для сегментации музыки используются как ритмические особенности, так и концептуальная мелодическая форма. Обычно другие музыкальные характеристики, такие, как плавность и гармонизация, игнорируются. Однако, в этом исследовании применили концепцию мелодической формы для повышения эффективности музыкальной сегментации. При этом отмечается, что предложенный подход к сегментации на основе фраз может перегружать длинные фразы. Проблема чрезмерной сегментации фраз не очень значимо снижает точность извлечения предложения в связи с тем, что применяемый шаг извлечения предложения объединяет последовательные подфразы.

В большинстве систем обнаружение разговорного элемента происходит в два этапа: сначала речь сегментируется, а далее происходит верификация в рамках сегментов. Анализ речевой информации с учетом знания особенностей речи накладывает на исследователей обязательство по предварительной обработке данных. В частности, в работе [90] авторам потребовалось осуществить устранение неоднозначности длительных сигналов для проведения сегментации. Как показывает практика, в ударном и в последнем слогах слов гласные могут

удлиняться. Данный факт может существенно упростить задачу сегментации слова из непрерывной речи. Однако, в таких языках, как английский, где преобладают ударения на первый слог, увеличенный по длительности сигнал может осложнить определение границ слов. Проведенный анализ большого корпуса английской речи показал, что говорящие предоставляют информацию о распределении, достаточную для того, чтобы потенциально позволить слушателям определить, связано ли удлинение гласного с лексическим ударением или окончательностью слова. Однако, авторами работы [91] утверждается, что преобладание ударных слогов в начале слова в английском языке существенно облегчает сегментацию в зашумленных условиях прослушивания. Таким преимуществом не обладают языки с различной ритмической структурой, к которой относится, в частности, испанский язык. В работе [92] используются акустические особенности речевого потока для обнаружения границ слов. Авторами было предложено применение спектральных характеристик сигнала для сегментации слов в арабской речи. Примером исследования разговорной русской речи может служить [93].

Обзор существующих решений по распознаванию слов показывает, что начало слова имеет более высокую информативность для сегментации, чем окончание. Ассиметричное распределение информации внутри слов происходит из-за коммуникативного давления, позволяющего распознавать слова в речи как можно раньше. Посредством анализа энтропии в работе [94] было показано, что музыкальные сегменты также имеют более высокое информационное содержание (то есть более высокую энтропию) в начале сегмента, чем в конце. Тем не менее, данный эффект не столь выразителен, как в речи, а наибольшая информация наблюдалась в середине выделенных сегментов. Авторы предполагают, что причиной этому может то, что первые и последние ноты музыкального примера имеют тенденцию быть тонально стабильными, с большей гибкостью в первой ноте для обеспечения исходного контекста.

Авторами работы [95] проводилось исследование восприятия акустической информации в частотной области выше 5 кГц. Проведенная ранее серия экспериментов показала, что слушатели способны идентифицировать то, о чем

говорят или поют дикторы на основании высокочастотной энергии в условиях присутствия маскирующего шума на низких частотах. Поэтому авторами была поставлена цель по определению способности слушателей транскрибировать короткие семантически непредсказуемые, но синтаксически правильно сформированные разговорные фразы, которые были отфильтрованы верхними частотами на частоте 5,6 кГц. Особый интерес вызвала способность некоторых слушателей правильно определять расположение границ слов даже без наличия низкочастотной информации.

В [96] также была рассмотрена проблема обработки аудиозаписей, содержащих сложных с точки зрения спектральной составляющих данных, к которым относят речь и музыку. Авторами был разработан метод классификации, основанный на синусоидальных траекториях для речи и музыки, и метод обнаружения, применяемый для речи с фоновой музыкой.

Важную роль играет сегментация и в исследованиях, посвященных системам поиска музыкальной информации. Знание структурной информации аудиосигнала позволит облегчить поиск и просмотр музыкальных коллекций, визуализировать музыкальную структуру исполнения, определить текст или классифицировать стиль музыки. Наиболее распространенный подход в этих системах направлен на обнаружение границ с помощью оценки новизны, которая описана в [97-98]. Методы, основанные на этой идее, ограничиваются композициями, которые следуют определенным правилам и принципам, так как требуют наличия предметной области. Авторами [97] предложено использование генетических алгоритмов для сегментации музыки. Предложенный ими метод не основан на его априорных знаниях, что позволяет применить алгоритм к более широкому музыкальному спектру. Музыка анализируется на основе ее самоподобия, а повторы используются для обнаружения сегментов. Затем алгоритм группирует схожие сегменты для создания групп сегментов (то есть наборов нескольких непересекающихся сегментов, которые удовлетворяют заданному условию подобия). Как замечено в [85], автоматическое распознавание текстов песен по-прежнему является сложной задачей, поскольку вариаций певческого голоса

намного больше, чем у говорящего, и отсутствует большая база данных с образцами пения. Авторами было решено использовать скрытую марковскую модель для распознавания слогов в записях с пением «а капелла». В исследовании [99] разрабатывают систему автоматического выравнивания текстовой лирики с музыкальным звуком. В ней для входного аудиосигнала сначала выполняется структурная сегментация, и аналогичным сегментам присваивается метка путем вычисления расстояния между парами сегментов.

Одним из вариантов подхода к сегментации музыкальных произведений [100] заключается в анализе их структуры. При этом в первую очередь извлекается функция тембра из акустического сигнала и строится матрица самоподобия, с помощью которой определяется сходство между функциями в музыке. Следующим шагом определяются границы предполагаемых сегментов на основании данных об отклонениях в матрице. Подобные сегменты, такие как повторы в музыкальном клипе, группируются и объединяются. Таким образом, каждое музыкальное произведение может быть представлено последовательностью состояний, где каждое состояние представляет музыкальный сегмент с похожей характеристикой.

На этапе классификации большинство систем основаны на моделях гауссовых смесей (GMM) или скрытых марковских моделях (HMM). Тем не менее, некоторые системы используют другие классификаторы речи или музыки. Встречаются работы, использующие для этой цели многоуровневый персептрон, максимальный апостериорный классификатор, k-ближайших соседи и разные гибридные системы.

Примером сегментации с применением классификатора может служить исследование [101], где предлагается использовать декомпозицию аудиосигнала на основе вейвлетов, которая позволяет провести хороший анализ нестационарного сигнала, такого как речь или музыка. На начальном этапе вычисляются различные типы энергии в каждой полосе частот, полученные в результате вейвлет-разложения. Для этого используются два классификатора: первый определяет наличие/отсутствие речи, а второй – наличие/отсутствие музыки. Проведенные исследования показали, что данный подход дает более высокие результаты, чем

основанный на мел-кепстральных коэффициентах. Схожий подход выбран в работе [102]. Здесь авторами используется система классификации, основанная на скрытых марковских моделях. Сигнал на основании векторов признаков делится на четыре класса: речевой, неречевой, музыкальный и не музыкальный.

Использование апостериорных функций, основанных на вероятности, для сегментирования аудиосигнала, содержащего речь или музыку, является не менее широко используемым методом. В исследовании [103] акустические модели на основе скрытой марковской модели также используются для вычисления апостериорных вероятностей. Акустические модели включают состояния контекстно-независимых источников в качестве модуля моделирования. Энтропия и динамизм обнаруживаются с помощью апостериорных вероятностей, и эти значения используются как функция для распознавания речи и музыки. Реализованный авторами классификатор на основе скрытой марковской модели и использующий декодирование Витерби с использованием различительных признаков аудиосигналов позволяет сегментировать как речь, так и музыку.

Отдельный интерес представляет метод сегментирования музыкального звука с иерархической тембровой моделью [104]. До определенного момента в традиционных музыковедческих подходах к извлечению структуры придавалось мало значения тембру как структурному измерению. Было замечено, что короткие последовательности соседних кадров могут быть назначены одному и тому же типу тембра, а общий тембр часто изменяется в течение любого отрезка значительной длины. Несмотря на то, что тембр меняется от кадра к кадру, распределение типов тембров остается довольно непротиворечивым в течение структурных сегментов и может использоваться для характеристики типов сегментов.

Авторами [105] был предложен метод оценки преобладающей тональности в музыкальном отрывке. Пространство тонов моделируется скрытой марковской моделью из 24 состояний, где каждое состояние представляет один из 24 основных и вспомогательных тонов, а каждое наблюдение представляет переход аккордов или пару последовательных аккордов. Использование переходов аккордов в качестве наблюдений моделирует большую временную зависимость между

последовательными аккордами, чем наблюдения отдельных аккордов. Ключевые переходы и вероятности эмиссии аккордов инициализируются с использованием результатов перцептивных тестов, чтобы отразить человеческое ожидание гармонических отношений. Параметры скрытой марковской модели затем обучаются на основе каждой песни с использованием аннотированных вручную символов аккордов, прежде чем модель для каждой песни будет декодирована, чтобы дать вероятность каждого ключа в каждом временном кадре.

Позднее в исследовании [106] были применены перечисленные выше алгоритмы для сегментации аудиозаписей с песнями с целью определения порядка рангов по точности работы. Кроме того, набор алгоритмов был дополнен их модифицированными версиями, включающими комбинирование функций. Тембральная сегментация в модифицированном алгоритме реализуется с использованием более широких возможностей для обучения HMM. Второй комбинированный алгоритм объединяет два метода непосредственно перед этапом кластеризации. Алгоритм распознавания аккордов, описанный в [107], используется первым для создания наблюдений для гармонической модели алгоритма. Скрытая марковская модель обучается для сегментации только гармонии, затем вероятности апостериорного состояния вычисляются для определения вероятности каждого тона в каждый период времени. Стадия тембрального извлечения выполняется так же, как в классическом алгоритме, но с использованием скрытой марковской модели с 24 состояниями для получения гистограмм состояния с тем же числом измерений, что и их вероятности. Вероятность последнего гармонического состояния добавляется к гистограмме признаков после нормализации для придания обоим признакам одинакового веса, а затем алгоритм кластеризации применяется к большим векторам признаков. Результаты показали, что протестированные алгоритмы не могут быть ранжированы, так как ни один алгоритм не выделил сегменты в каждом испытании лучше, чем другие. В одних случаях алгоритм кластеризации с комбинацией функций показывал лучшую точность границ по сравнению с другими методами, поскольку он мог использовать улучшенную точность гармонических

характеристик, но алгоритм кластеризации не допускал короткие сегменты, созданные классическим гармоническим алгоритмом. В других случаях лучше работал алгоритм, основанный на скрытой марковской модели с комбинированием функций, в то время как простой тембральный алгоритм находил лишние сегменты. В целом результаты показали, что отдельно гармоническая сегментация работает хуже других, но в сочетании с тембральной техникой она может улучшить точность, даже если дает плохую сегментацию, поскольку она менее чувствительна к изменениям инструментовки, которые часто происходят в середине сегмента.

Распространенным подходом к определению границ сегментов в музыкальном произведении является акцентирование внимания на начало и конец фраз. В музыке понятие «фраза» обычно отождествляется с четырьмя тактами. Такой длины придерживаются в простых произведениях, к которым относятся, например, народные песни. Авторами [108] был предложен метод сегментации музыки, основанный на фразах, при котором музыка разбивается на множество четырехмерных фрагментов, оцениваемых впоследствии на предмет значимости. С учетом того, что не все произведения используют четыре такта в качестве фразы, авторы обратили внимание на метод обнаружения локальных границ [109-110]. Данный метод позволяет обнаруживать локальные границы в мелодической поверхности и может использоваться для сегментации музыки. В нем используются длительность, высота тона и ноты паузы для разделения музыки на короткие фразы. Отмечая недостаток модели, заключающийся в нехватке точности восприятия фрагментов, авторы предложили [111] эвристический метод, учитывающий продолжительность и ноту паузы фрагментов в методе обнаружения локальных границ, чтобы получить более длинные предложения. Предложенный алгоритм позволил правильно определить мелодию в 77.9% аудиозаписях с одним каналом и 53.4 % в аудиозаписях с мелодией, распределенной по нескольким каналам.

Результаты анализа алгоритмов сегментации отражены в таблице 1.3. Исследование показало, что автоматическая сегментация показывает хорошие результаты в тех алгоритмах, где используется предварительно обученная модель

для конкретного языка. Однако данные алгоритмы не могут должным образом обработать разные стили речи. Следовательно, при анализе различных стилей вокального исполнения также возможны ошибки. Основные причины ошибок в данных алгоритмах обуславливаются неизвестной длительностью сегментов и наличием последовательности одинаковых сегментов. Помимо описанных выше алгоритмов, в таблицу был включен алгоритм сегментации на вокализованные и невокализованные участки, разработанный в рамках работы над моделью слуховой системы человека [70].

Таблица 1.3 – Показатели алгоритмов сегментации

Алгоритм	Показатели алгоритма
Байесовская модель с сегментными представлениями слов	Кластеризация на словоподобные участки. Точность работы в английском языке при известном словаре около 84% [77].
С применением скрытой марковской модели	Алгоритм, учитывающий особенности индийских языков показал результат на уровне 86% правильно распознанных границ слов [83].
С использованием меры доверия	Сегментация музыки с применением данной модели осуществлена не менее чем с 14% ошибками[88].
Сегментации на вокализованные и невокализованные участки	Проведение автоматической сегментации с надежностью 0.89-0.93. Относительное количество пропущенных границ: 0-0.03. Относительное количество лишних границ: 0.04-0.11.

1.3 Исследование параметров вокальных исполнений

В сфере музыкального обучения сложно выделить господствующую методику развития вокальных данных. Сборники указаний начинающим певцам и музыкантам получили название сольфеджио. Исследования истории развития подобных методик [112-113] сходятся на том, что, начиная с конца XIX века, в содержании руководств произошли значительные изменения. Наибольшее влияние оказал Н.А. Римский-Корсаков, полагавший невозможным развитие музыкального слуха без понимания музыкальных средств. Базовыми элементами музыкального языка являются гармония и ритм, формирующие соответственно гармонический и ритмический музыкальный слух. Гармонический слух подразделяется на слух

строя, помогающий отличать музыкальные словосочетания из комбинаций обертонов сложных звуков от немузыкальных, и слух лада, фиксирующий отношения звуков по высоте и тональной функции [114]. В методических руководствах по развитию слуха строя им предлагалось в том числе пение под настроенное фортепиано и пение в хоре.

В дальнейшем развитие сольфеджио претерпевало множество изменений, но самыми значительными стали те, которые были вызваны работами ученых с достоверными сведениями о природе и функциях музыкального слуха. В частности, акустико-психологические исследования таких ученых, как Ревеш [115] и Володин [116] показали избирательный характер восприятия музыки. Согласно полученным учеными результатам, восприятие высоты звука складывается из тембровой насыщенности и фонической сбалансированности звучания. Музыкальный звук обрабатывается слухом по разным психологическим критериям, что стало основой для составления рекомендаций по развитию звуковысотного восприятия, в котором для тембровой составляющей предлагается глиссандо, а для фонического – упражнения с интервальными комбинациями нот.

Примерами проявления внимания к существующим техническим средствам с точки зрения применения в обучении вокальному мастерству могут служить такие работы, как [117-121]. В работе [117] уделяется внимание акустическому исследованию частоты основного тона, интенсивности и длительности в компьютерной программе Praat [122]. В исследовании [118] отмечено, что развитие вокального образования на современном этапе может осуществляться комплексно – развитие технологий позволило сформировать ресурсы с образцами записей исполнителей музыки вокального жанра, изучение которых способствует накоплению музыкального опыта и знаний. Кроме того, существует множество специализированных компьютерных программ, с помощью которых можно овладеть новыми навыками или развить имеющиеся. Автором был произведен анализ существующих программ, которые могут применяться при обучении пению. В рамках эксперимента были сделаны эталонные записи с вокальным исполнением последовательности нот, а затем обработаны с помощью

исследуемых тренажеров. В таблице приведены основные особенности и результаты проведенных испытаний. В исследовании [119] также проводится анализ тренажерных и тестовых программ для занятий сольфеджио. Однако основное внимание в данном обзоре уделяется инструментам для развития музыкального слуха.

Таблица 1.4 – Показатели для программ-тренажеров пения

Название программы	Показатели программы
Melodyne[123]	Программа работает только с аудиозаписями. Для каждой ноты отображает длительность звучания. Отсутствуют упражнения. В рамках эксперимента удалось распознать правильно 72% спетых нот.
Sing and See[124]	Позволяет работать в режиме реального времени. Сведений об алгоритмах обработки в открытом доступе нет, но визуальная оценка позволяет сделать вывод о пиковом методе анализа ЧОТ. Об этом свидетельствуют зависания программы при резких изменениях в звучании. Не содержит в себе упражнений. В рамках эксперимента удалось распознать правильно 54% спетых нот.
Singing Coach[125]	Наличие возможности детальной настройки относительно типа голоса исполнителя позволяет указывать некорректные данные для оценки качества пения. В рамках эксперимента удалось распознать правильно 63% спетых нот.
EarMaster[126]	Алгоритм распознавания нот не всегда воспринимает исполнение диктора и не отображает результат пока пользователь не исполнит нужную ноту. Данное обстоятельство не позволяет скорректировать действия пользователя при выполнении упражнения. В рамках эксперимента удалось распознать правильно 58% спетых нот.
VocTeacher[127]	Интерактивное приложение-тренажер, позволяющее произвести оценку качества исполнения самому пользователем на основании выводов о схожести целевых нот и полученных областей красных точек, указывающих на место распознанной певческой активности. При анализе записей непрерывного звучания с нотой программа выдает отрезки с паузами, которых в исполнении не было. В рамках эксперимента удалось распознать правильно 47% спетых нот.

Как утверждается в [120], самостоятельная работа над правильным исполнением нот может быть затруднена наличием разницы в восприятии собственного голоса от того, как его воспринимают окружающие. Эту проблему способно частично решить использование технических средств, чье функционирование по качеству работы достаточно близко к экспертной оценке. Более того, внедрение информационных технологий в процесс обучения вокалу позволяет применять разнообразные методы, приемы, формы и средства, проводить контрольно-оценочную деятельность с использованием современных способов оценивания [121], что также способно повысить интерес обучающихся. Как свидетельствует работа [53], современное состояние исследований в области речевых технологий делает возможным выделение в голосах певцов характеристик, которые сложно определить на слух, но имеющих большое значение для оценки вокальной одаренности испытуемого. Автором предлагается использование комплексного метода для определения типа певческого голоса, его силы, динамического диапазона и других показателей.

В работе [128] приводится классификация существующих методов обучения вокалу. Наибольшее влияние, по мнению автора публикации, на формирование умений и навыков певца оказывают практические методы. К ним относятся методы упражнений и практических работ. Упражнения являются наиболее распространенной и эффективной формой закрепления знаний, певческих умений и навыков. Практические работы, в свою очередь, направлены на углубление приобретенных навыков.

С целью оптимизации упражнений в исследовании [129] обращено внимание на необходимость ограничения числа элементов в задании количеством от 5 до 9. Это ограничение обусловлено исследованием американского психолога Дж.А. Миллера [130], в котором было определено, что память человека в состоянии воспринимать именно такое число смысловых единиц.

1.4 Выводы по главе

В настоящей главе были рассмотрены алгоритмы анализа частоты основного тона речевого сигнала и алгоритмы сегментации, а также их применимость для

обработки вокальных исполнений. В результате проведенного анализа был сделан вывод, что в сфере речевых технологий отсутствуют алгоритмы, направленные на точную идентификацию спетой диктором ноты. Исследование вокальных исполнений осуществляется только в направлениях, посвященных изучению проявления эмоций в голосе. Кроме того, существующие алгоритмы анализа основного тона сигнала обладают высоким процентом грубых ошибок, что делает их неприменимыми для задач обучения пению. Ограничение в охватываемом диапазоне обработки сигналов также не позволяет применить их в сфере распознавания нот.

Было определено, что наименьшим процентом ошибок при идентификации частот основного тона обладает алгоритм, основанный на математической модели слуховой системы человека. Ограничение в диапазоне обработки данного алгоритма определяется сферой использования. Поскольку применяемая в алгоритме модель была построена под речевой диапазон, математическая модель должна быть доработана до диапазона звучания вокального исполнения.

В качестве основы для алгоритма сегментации нот может быть использован подход алгоритма сегментации на вокализованные и невокализованные участки, также основанный на применении математической модели слуховой системы человека. Данный алгоритм обладает высокой надежностью при автоматической сегментации и низкой долями пропущенных и лишних границ.

2 Формирование набора шаблонов для определения частоты основного тона

2.1 Математическая модель слуховой системы человека

Для определения ЧОТ в исследуемом алгоритме используется эффект одновременной маскировки, заключающийся в следующем. Каждой точке вдоль основной мембраны внутреннего уха, которая преобразует механические колебания в нервные импульсы, ставится в соответствие частота звука, вызывающая максимальный отклик в данной точке. Чем больше расстояние от этой точки, тем ниже амплитуда отклика. Восприятие сигналов сложной формы (в том числе речевого сигнала) характеризуется тем, что отклик будет происходить на все частотные компоненты сигнала. Если амплитуда отклика на компоненту с собственной частотой окажется ниже, чем на другие, то данная компонента слуховой системой восприниматься не будет.

Описанный выше принцип применяется для определения значения частоты основного тона в речевом сигнале. Основная мембрана может быть рассмотрена как набор частотных резонансных фильтров. Данный подход к задаче определения особенностей в структуре речевого сигнала представлен в [70].

При создании фильтров используются такие характеристики, как частота дискретизации, требуемое количество каналов фильтрации, значения верхней и нижней границ частот проводимого анализа, а также коэффициент точности для создаваемых фильтров. На этапе вычисления весовых коэффициентов одновременной маскировки и генерации набора шаблонов необходимо также определить несколько переменных.

Шаг квантования по времени в секундах определяется по формуле 2.1.

$$T_k = \frac{1}{F_k}, \tag{2.1}$$

где F_k– частота квантования по времени.

При дальнейших вычислениях используются следующие параметры:

$\alpha = 0.1$ – погрешность задания количества весовых коэффициентов системы фильтров;

F_{0n} – нижняя частота основного тона для формирования масок;

F_{0v}– верхняя частота основного тона для формирования масок.

На этапе расчета параметров системы фильтра для анализа частот основного тона сигнала в первую очередь требуется вычислить коэффициенты для расчета шкал резонансных частот (формула 2.2) и добротностей (формула 2.3).

$$dx = \frac{1}{K} \cdot \ln\left(\frac{F_0}{F_n}\right), \tag{2.2}$$

F_0 –верхняя частота системы фильтров в Гц;

F_n – нижняя частота системы фильтров в Гц.

$$c = \frac{1}{K} \cdot \ln\left(\frac{Q_0}{Q_n}\right), \tag{2.3}$$

Q_0 – добротность на верхней частоте системы фильтров;

Q_n – добротность на нижней частоте системы фильтров.

Следующим шагом для анализируемого диапазона частот основного тона речевого сигнала по формуле 2.4 вычисляется шкала частот, а по формуле 2.5 шкала добротностей.

$$F_k = \frac{F_0}{\exp(k \cdot dx)}, \tag{2.4}$$

где k – текущий канал анализа из диапазона от 0 до K

$$Q_k = Q_0 \cdot \exp(-c \cdot k) \tag{2.5}$$

На основании полученных данных определяется число коэффициентов системы фильтров для диапазона частот основного тона по формуле 2.6.

$$I_0(k) = \left\lceil \frac{2.4 \cdot Q_k}{\omega_0 \cdot \exp(-k \cdot dx) \cdot T_k} \cdot \ln\left(\frac{1}{\alpha}\right) \right\rceil, \tag{2.6}$$

где ω_0 –верхняя частота системы фильтров в рад/с.

Далее по сформированной шкале частот с учетом значений коэффициентов системы фильтров эмпирическим путем был сформирован набор шаблонов, использовавшийся в рамках исследования параметров сигнала в слитной речи и при онкологических заболеваниях.

2.2 Модификация математической модели

Модификация заключается в реализации возможности указания алгоритму граничных частот определения ЧОТ. Данная модификация была осуществлена с

целью автоматизации генерации наборов шаблонов для исследуемых частот. Как было указано ранее, в первоначальной версии модели набор шаблонов для алгоритма идентификации частот основного тона был получен эмпирическим путем и может применяться только к диапазону ЧОТ от 70 до 400 Гц. В случае применения шаблонов к сигналам с частотами выше 400 Гц алгоритм идентифицировал значения частот только для посторонних шумов. Применение эмпирического подхода к определению нового набора шаблонов осложняется увеличением числа каналов определения частот основного тона.

Каждый из шаблонов в наборе содержит в себе последовательности из 0 и 1, используемых при маскировании сигнала. Графически применение шаблонов можно представить следующим образом. Шаблон (рисунок 2.1) представляет из себя последовательность из 2 «черных» и 3 «белых» полос, где черные полосы соответствуют участкам с вокализацией, а белые – невокализованным. На рисунке P_0 обозначен результат применения маскировки.

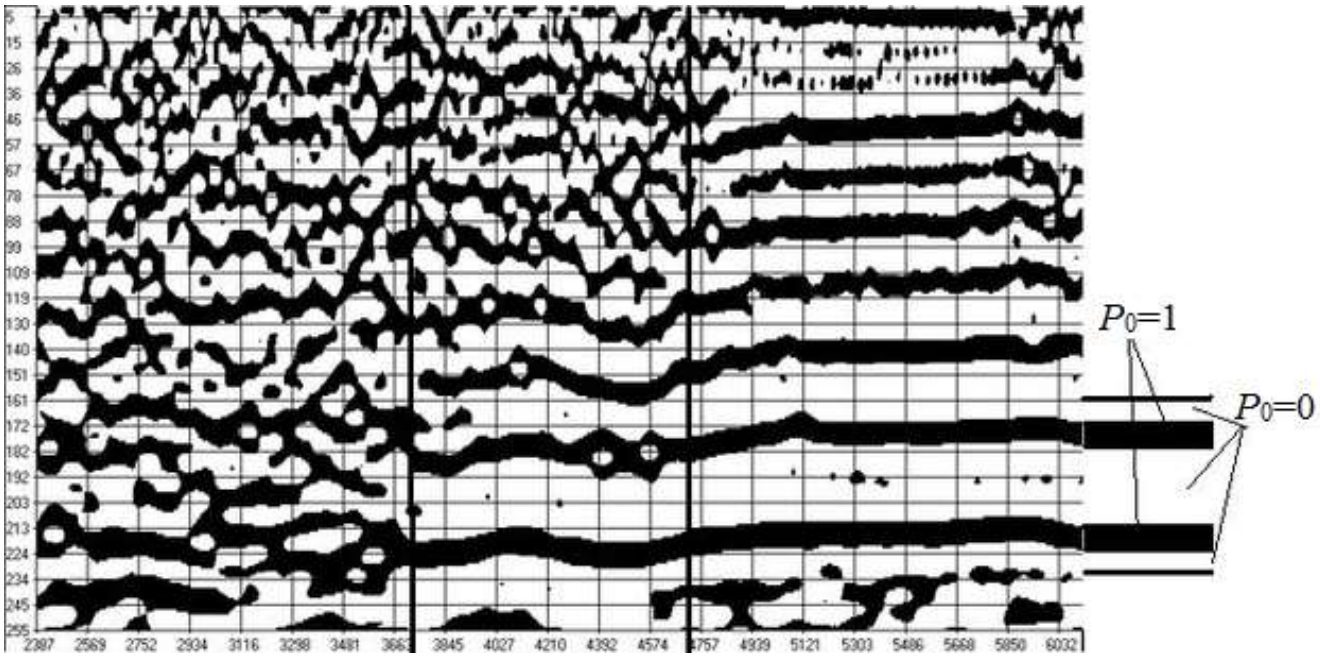


Рисунок 2.1 – Применение шаблона при идентификации ЧОТ

На основании математической модели слуховой системы человека был определен порядок операций, которые необходимо выполнить при генерации набора шаблонов. Работа алгоритма генерации шаблонов [131] заключается в выполнении следующих шагов:

1) вычисление граничных номеров каналов ЧОТ (k_{0n} – нижняя и k_{0v} – верхняя);

2) формирование тестовых сигналов для создания шаблонов;

3) одновременная маскировка (вычисление массива результата маскировки $P_0[k_t,k]$);

4) вычисление массива номеров первых каналов свертки $N_1[k_t]$, массива количества каналов в шаблоне $N_k[k_t]$, массива набора шаблонов $Tpl[k_t,k]$.

Таким образом, генерация шаблонов происходит по упрощенной схеме, представленной на рисунке 2.2.

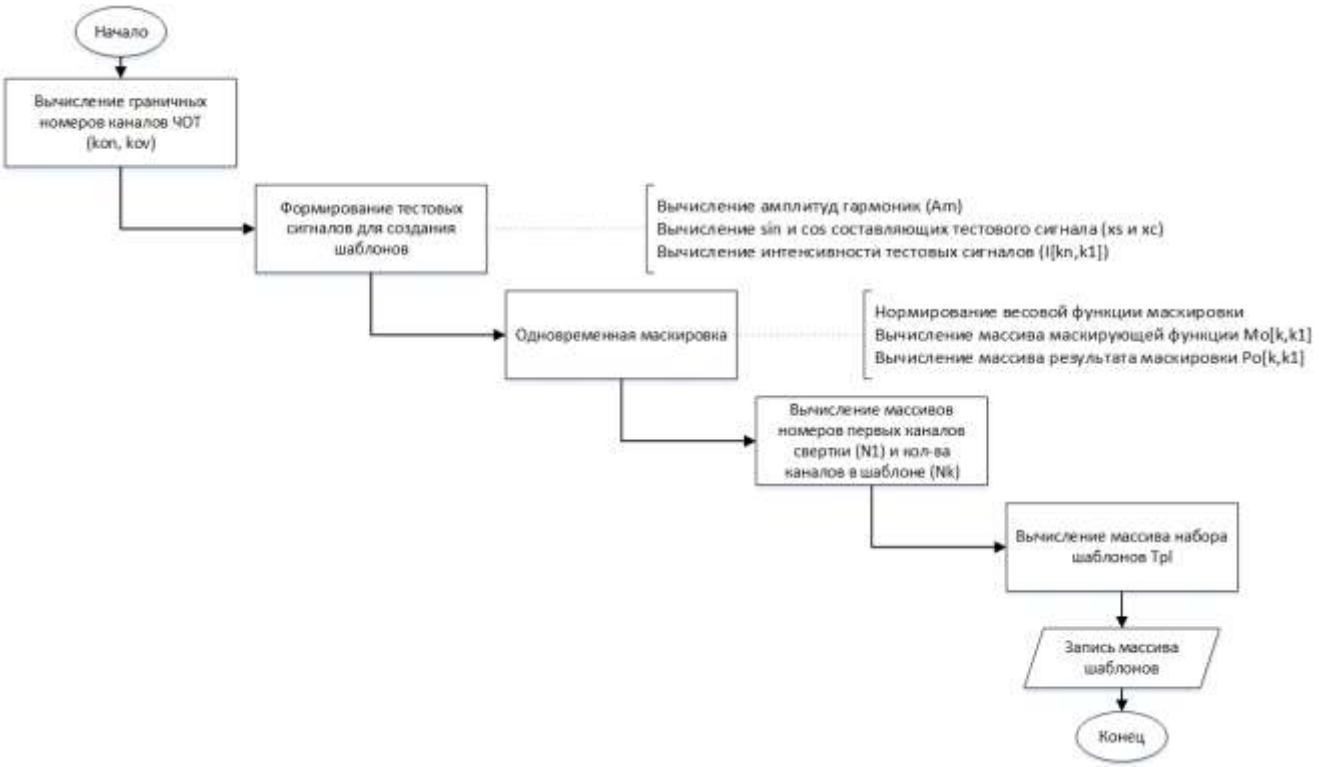


Рисунок 2.2 – Блок-схема алгоритма генерации шаблонов

Отталкиваясь от формул, полученных в математической модели слуховой системы человека, было определено, что вычисление значений номеров каналов, соответствующих нижней и верхней границам определения частоты основного тона, k_{0n} и k_{0v} осуществляется по формулам 2.7 и 2.8.

$$k_{0n} = \left\lfloor \frac{1}{dx} \cdot \ln\left(\frac{F_0}{F_{0n}}\right) \right\rfloor \quad (2.7)$$

$$k_{0v} = \left\lfloor \frac{1}{dx} \cdot \ln\left(\frac{F_0}{F_{0v}}\right) \right\rfloor \quad (2.8)$$

На основании данных о граничных значениях номеров каналов становится возможным определение количества масок необходимого для работы алгоритма генерации шаблонов для проведения анализа (формула 2.9), что в свою очередь позволит получить шкалу частот для исследуемого диапазона (формула 2.10).

$$k_m = R \cdot (k_{0n} - k_{0v}), \tag{2.9}$$

где R – коэффициент умножения (по умолчанию $R = 1$).

$$F_{k_1} = \frac{F_0}{exp\left(\frac{k_1 \cdot dx}{R} + k_{0v} \cdot dx\right)}, \tag{2.10}$$

где k_1 – номер маски в диапазоне от 0 до k_m.

С этого момента наступает этап формирования тестовых сигналов. Полученные значения каналов используются при обработке тестовых сигналов сверткой с фильтрами, что в результате позволяет определить амплитуду сигнала. Амплитуды выбранных гармоник основного тона вычисляются по формуле 2.11.

$$A_m = \frac{1}{1 + a \cdot m}, \tag{2.11}$$

где (a – коэффициент (по умолчанию $a = 0.25$);

m – учитываемая гармоника основного тона (от 0-й до 3-й).

Частоты выбранных гармоник основного тона определяются по формуле 2.12 в рад/с и по формуле 2.13 в Гц.

$$\omega_{m,\ k_1} = 2 \cdot \pi \cdot (m + 1) \cdot F_{k_1} \tag{2.12}$$

$$Fx_{m,\ k_1} = F_{k1n} \cdot (m + 1) \tag{2.13}$$

Следующим шагом необходимо определить реакцию системы фильтров на сформированные тестовые сигналы. Для этого воспользуемся формулами 2.14 и 2.15, чтобы получить их sin и cos составляющие.

$$xs(k, k_1) = \sum_m A_m \cdot e^{-1.44 \cdot (Q_k)^2 \left(1 - \frac{Fx_{m,\ k_1}}{F_{n0}} \cdot e^{k \cdot dx}\right)^2} \cdot \sin\left(\omega_{m,\ k_1 n} \cdot T_k \cdot I\right), \tag{2.14}$$

где I – момент времени формирования реакции системы фильтров на тестовые сигналы в тактах времени T_k.

$$xc(k,k_1) = \sum_m A_m \cdot e^{-1.44\cdot(Q_k)^2\left(1-\frac{Fx_{m,k_1}}{F_{n0}}\cdot e^{k\cdot dx}\right)^2} \cdot \cos\left(\omega_{m,\ k_1}\cdot T_k \cdot I\right) \quad (2.15)$$

В таком случае, реакция системы фильтров на тестовые сигналы будет определяться формулой 2.16.

$$I_{k,\ k_1} = (xs(k,k_1))^2 + (xc(k,k_1))^2 \quad (2.16)$$

После этого осуществляется одновременная маскировка. В первую очередь необходимо определить форму весовой функции маскировки по формуле 2.17.

$$H_0(k,n) = e^{-2.88\cdot\delta n\cdot(Q_k)^2\left(1-e^{(k-n)\cdot dx}\right)^2}, \quad (2.17)$$

где δn – коэффициент, определяющий ширину весовой функции маскировки;

n – номер канала в диапазоне от 0 до K.

Далее осуществляется нормирование весовой функции маскировки с помощью формул 2.18 и 2.19.

$$B(n) = \sum_k H_0(k,n) \quad (2.18)$$

$$W_{k,\ n} = \frac{H_0(k,n)}{B(n)} \quad (2.19)$$

В результате, функция одновременной маскировки принимает вид, представленный формулой 2.20. Результат маскирования получается с применением формулы 2.21.

$$M_{0_{k,\ k_1}} = \sum_{n=0}^{k} I_{n,\ k_1}\cdot W_{k,\ n} \quad (2.20)$$

$$P_{0_{k,\ k_1}} = \begin{cases} 1, \text{если } I_{k,\ k_1} - M_{0_{k,\ k_1}} > 0 \\ 0, \text{если } I_{k,k_1} - M_{0_{k,\ k_1}} \le 0 \end{cases} \quad (2.21)$$

На основании полученных формул был разработан алгоритм генерации набора шаблонов (рисунок 2.3), включающий в себя вычисление массива номеров первых каналов свертки $N_1[k_t]$, при $0 \le k_t \le k_{0n}-k_{0v}$ и массива количества каналов в шаблоне $N_k[k_t]$, при $0 \le k_t \le k_{0n}-k_{0v}$ с последующим формированием массива набора

шаблонов при $0 \le k_t \le k_{0n}-k_{ov}$; $0 \le k \le N_k[k_t]-1$ по формуле 2.22. После этого осуществляется запись массива набора шаблонов Tpl.

$$Tpl = P_{0_{k,\ N_1[k]}} \tag{2.22}$$

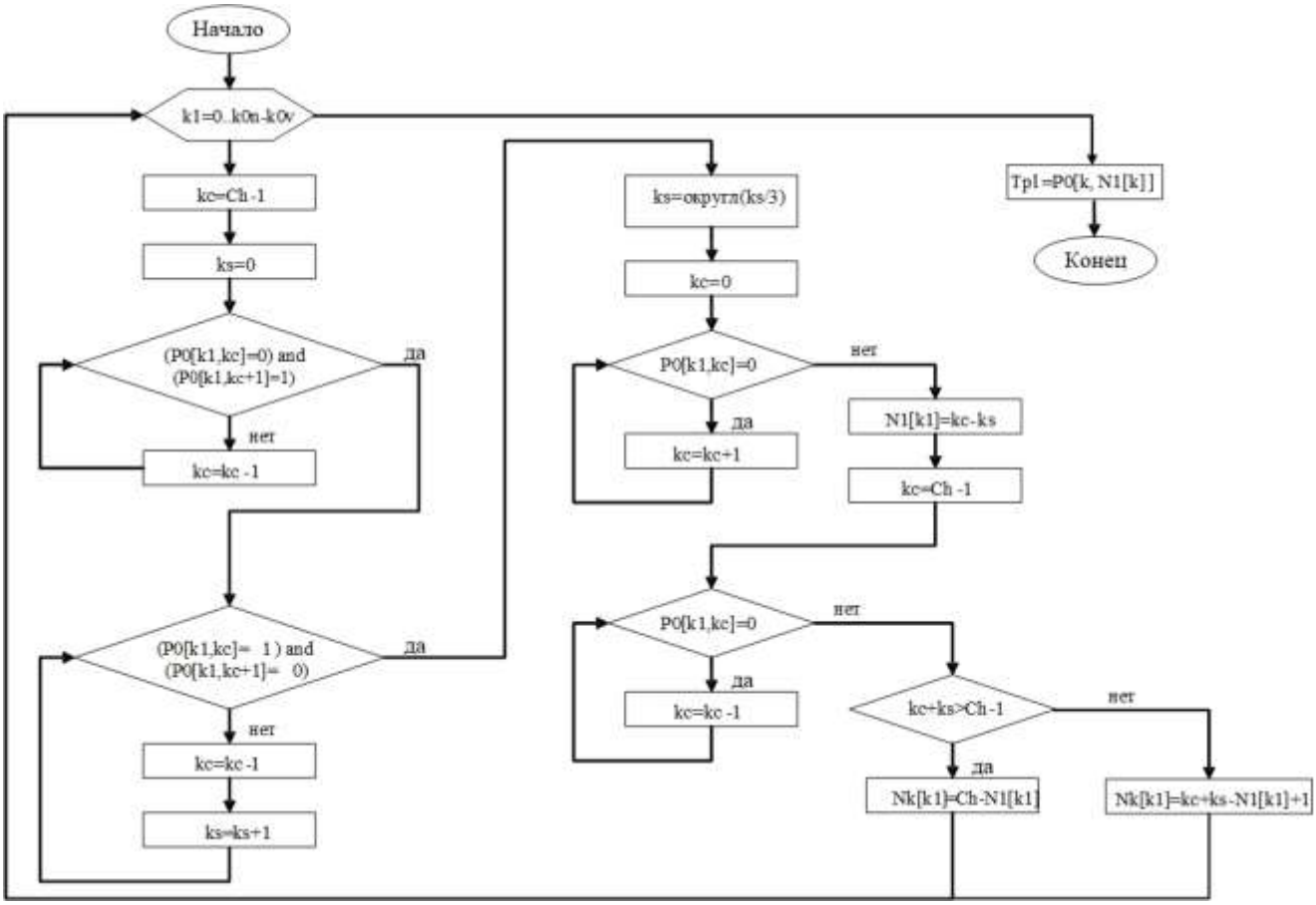


Рисунок 2.3 – Алгоритм генерации набора шаблонов

В результате, задавая значения граничных частот для определения ЧОТ, мы указываем алгоритму генерации шаблонов диапазон каналов, для которых необходимо сформировать набор шаблонов.

2.3 Эксперименты по определению частоты основного тона на синусоидальных сигналах

На основании полученных шаблонов происходит определение номера канала ЧОТ и определение значения ЧОТ в Гц по алгоритму, представленному на рисунке 2.4. Номер канала ЧОТ (kff) и минимальная мера различия (dmin) приравниваются к текущему количеству каналов фильтрации (Ch). Цикл сравнения массива шаблонов с результатами маскировки осуществляется при $0 \le k_t \le k_{0n}-k_{0v}$. В случае обнаружения совпадения происходит выход из цикла, а полученный результат

счетчика d сравнивается с минимальной мерой различия. Если минимальная мера различия оказывается больше или равен d, то к номеру канала ЧОТ прибавляется номер верхней границы. Дальше минимальная мера различия сравнивается с порогом вокализации.

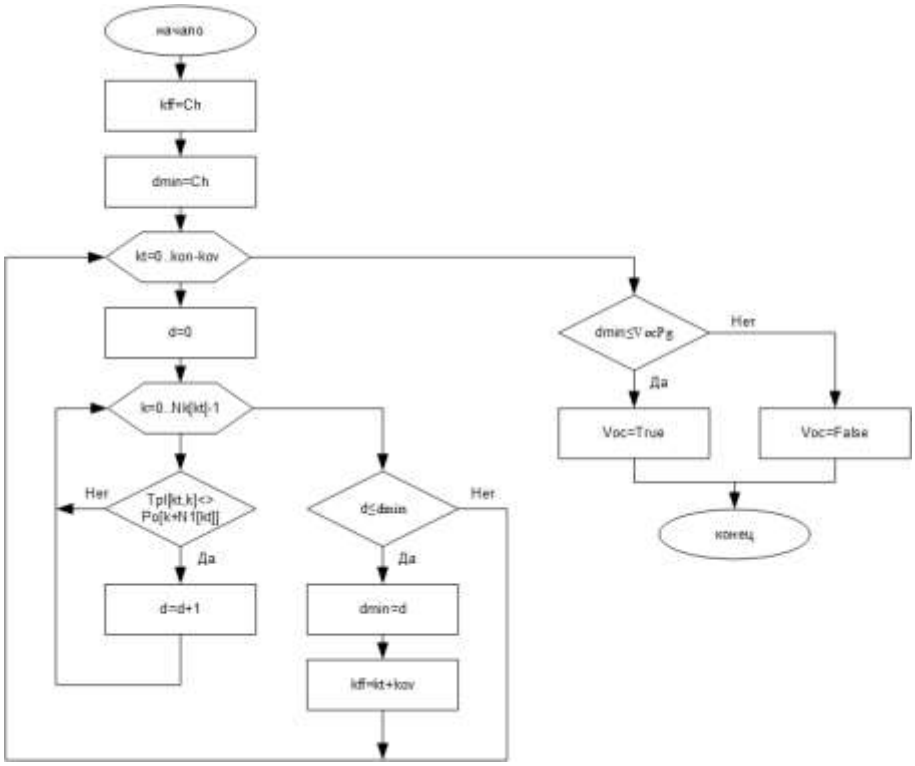


Рисунок 2.4 – Алгоритм определения номера канала ЧОТ

Для проведения экспериментальной оценки корректности работы алгоритма идентификации частот основного тона были сгенерированы синусоидальные сигналы с заданными значениями ЧОТ.

Шаг квантования по времени в секундах вычисляется по формуле 2.23.

$$T_k = \frac{1}{F_v}, \tag{2.23}$$

где $F_v = 8000$ – частота выборки в Гц.

По формуле 2.24 рассчитывается

$$x_i = \sum_n (A_n \cdot \sin(2 \cdot \pi \cdot n \cdot F0 \cdot T_k \cdot i)), \tag{2.24}$$

где A_n – амплитуды гармоник в диапазоне от 0.6 до 1;

n – число учитываемых гармоник в диапазоне от 1 до 8;

$F0$ – частота основной гармоники в герцах;

i – момент времени в диапазоне от 0 до I.

Генерация тестового сигнала осуществляется по формуле 2.25.

$$X = \left\lfloor (2^{15} - 1) \cdot \frac{x - X_{min}}{X_{max} - X_{min}} \right\rfloor, \quad (2.25)$$

где X_{min} – минимальное значение x;

X_{max} – максимальное значение x.

Полученный тестовый сигнал с частотой выборки F_v и разрядностью 16 бит записывается в файл. Для оценки точности работы алгоритма были сгенерированы синусоидальные сигналы с известной частотой в диапазоне от 70 до 800 Гц. На рисунке 2.5 показан результат работы алгоритма для синусоидального сигнала с частотой основного тона 200 Гц. Графики для сигналов с частотами до 600 Гц включительно выглядят аналогичным образом. По оси абсцисс – время в миллисекундах, по оси ординат – значение частоты основного тона в Гц.

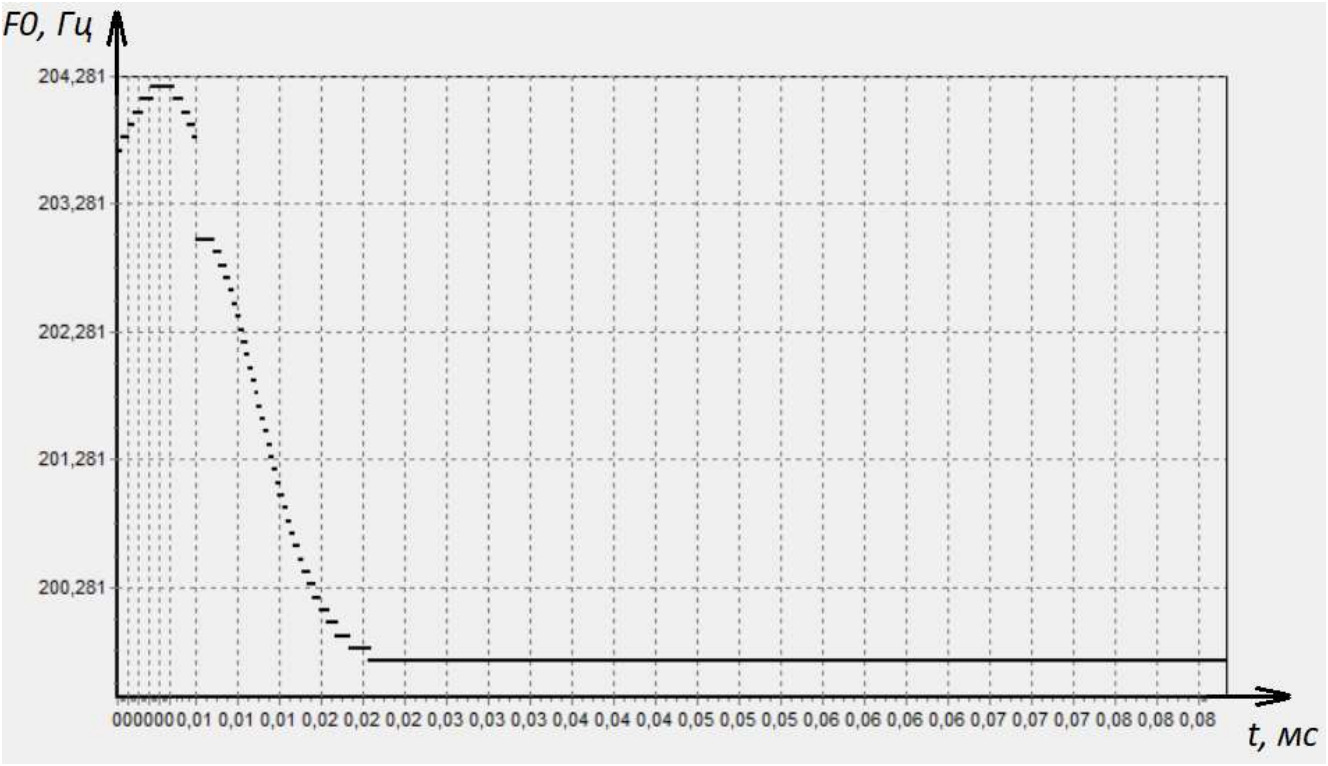


Рисунок 2.5 – График определения частоты основного тона синусоидального сигнала на 200 Гц

Результаты работы алгоритма для синусоидальных сигналов с разными частотами были проанализированы на предмет ошибок. В таблице 2.1 представлены значения относительных погрешностей определения частоты основного тона. Видно, что относительная погрешность возрастает с увеличением частоты сигнала.

Таблица 2.1 – Относительная ошибка определения частоты основного тона

Частота основного тона синусоидального сигнала, Гц	Относительная погрешность, %	Частота основного тона синусоидального сигнала, Гц	Относительная погрешность, %
70	<0.01	400	0.50
90	<0.01	450	0.44
120	<0.01	500	0.60
150	<0.01	550	0.73
200	<0.01	600	0.83
250	0.40	650	50.00
300	0.33	700	48.14
350	0.29		

Что касается сигналов с частотами 650 и 700 Гц, погрешности для них рассчитаны исходя из определенного номера канала, так как частота на каждом временном отсчете определяется как разная. Номер канала для 650 Гц (рисунок 2.6) приблизительно соответствует номеру канала для сигнала 325 Гц, а номер канала для 700 Гц (рисунок 2.7) приблизительно соответствует номеру канала для сигнала 363 Гц. Стоит добавить, что анализу подлежат лишь вокализованные участки (имеющие периодическую структуру). Отсюда следует то, что наличие «пунктира» на графиках объясняется тем, что алгоритм определяет некоторые участки сигнала как невокализованные.

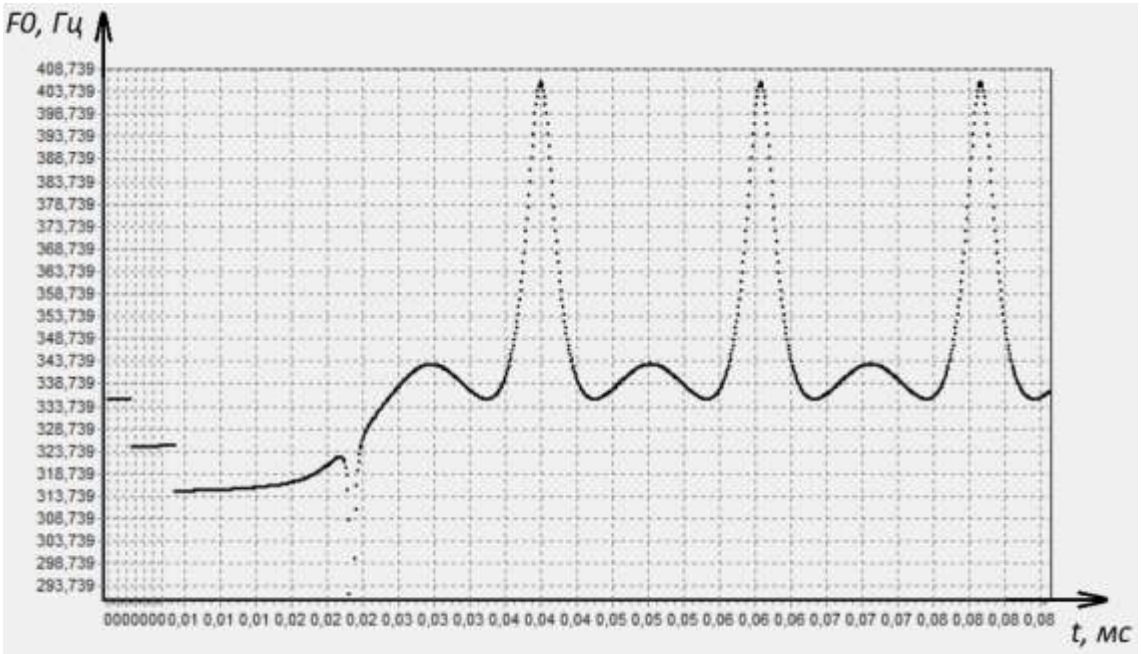


Рисунок 2.6 – График определения частоты основного тона синусоидального сигнала на 650 Гц

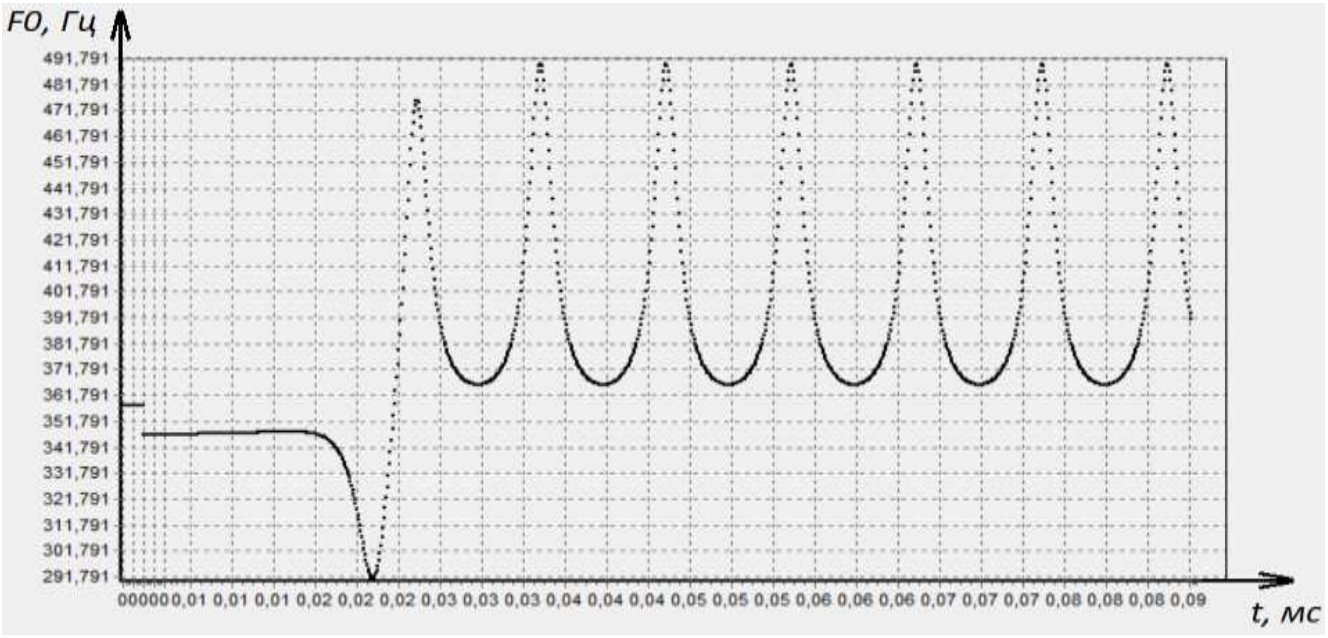


Рисунок 2.7 – График определения частоты основного тона синусоидального сигнала на 700 Гц

Экспериментально была определена граница частот, начиная с которой алгоритм идентификации ЧОТ вычисляет неправильное значение, равная 620 Гц. С целью устранения ошибок в работе алгоритма был исследован порог вокализации, учитываемый в алгоритме. Было определено, что при использовании значения порога вокализации равным 4 частоты синусоидальных сигналов выше 600 Гц определяются корректно (рисунки 2.8 - 2.10).

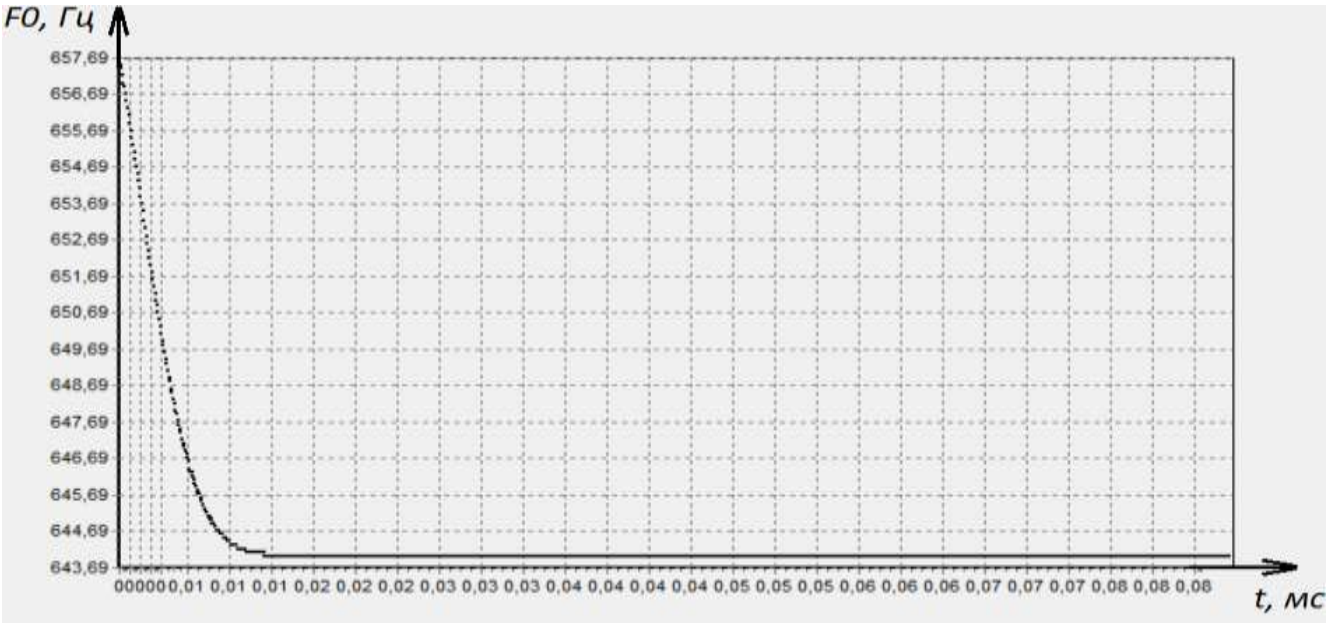


Рисунок 2.8 – График зависимости частоты основного тона для синусоидального сигнала с ЧОТ 650 Гц при заданном пороге вокализации

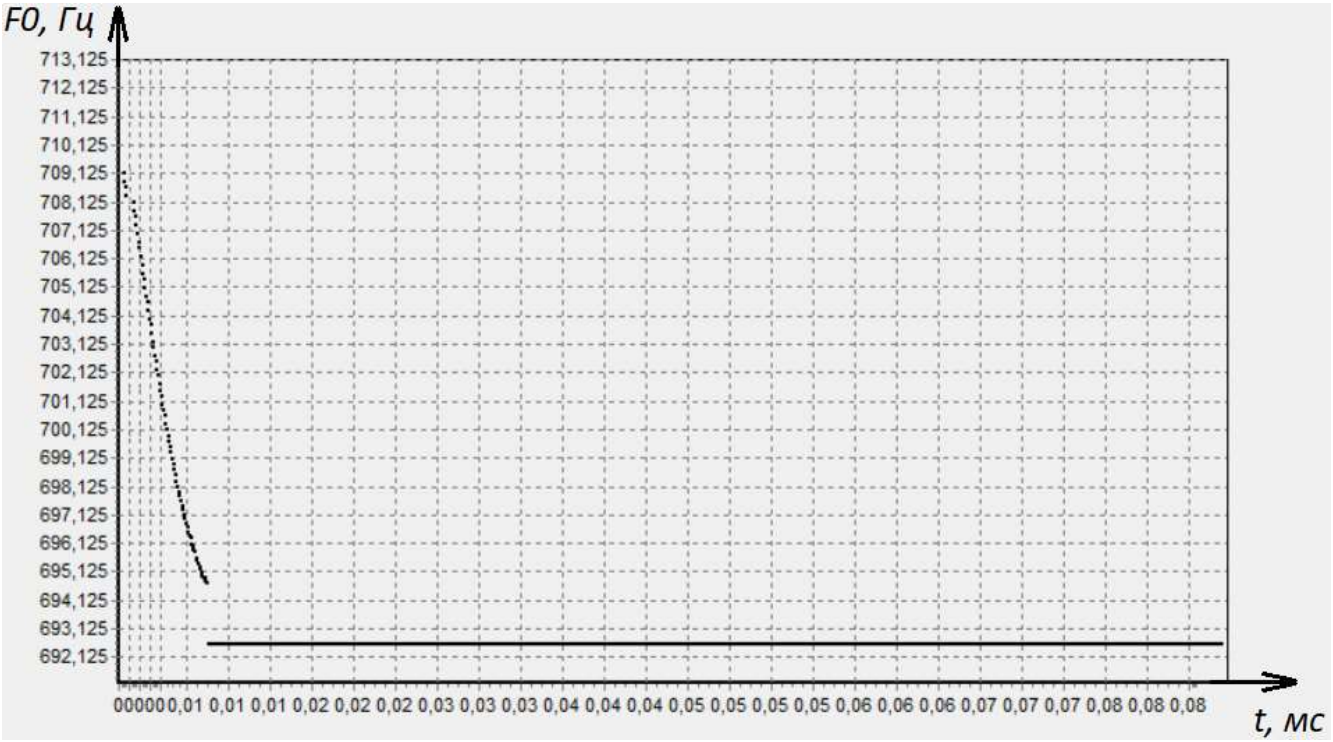


Рисунок 2.9 – График зависимости частоты основного тона для синусоидального сигнала с ЧОТ 700 Гц при заданном пороге вокализации

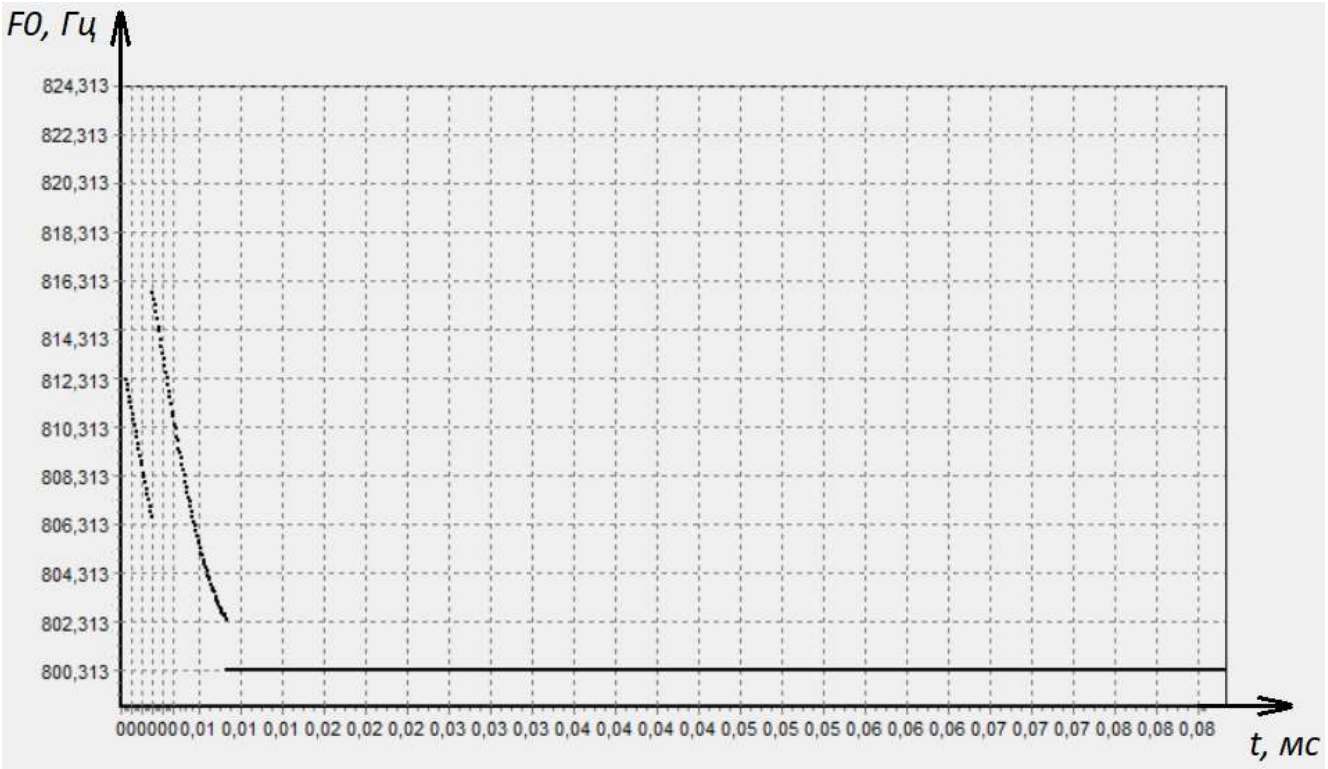


Рисунок 2.10 – График зависимости частоты основного тона для синусоидального сигнала с ЧОТ 800 Гц при заданном пороге вокализации

В таблице 2.2 представлены значения относительных погрешностей определения частоты основного тона после произведенных изменений порога

вокализации для частот от 600 до 800 Гц. Относительная погрешность для частот до 600 Гц осталась на прежнем уровне.

Таблица 2.2 – Относительная ошибка определения частоты основного тона

Частота основного тона синусоидального сигнала, Гц	Относительная погрешность, %	Частота основного тона синусоидального сигнала, Гц	Относительная погрешность, %
600	0.83	700	0.85
630	0.79	750	0.88
660	0.91	800	0.94

Кроме того, при помощи алгоритма были исследованы записи для частот выше 800 Гц. Полученные результаты показали, что границы применимости модифицированной модели при относительной погрешности менее 1% составляют диапазон от 70 до 800 Гц.

Далее работа алгоритма с выбранными значениями порога вокализации была проверена на записях женского голоса, а именно вокальных исполнений. Вокальные исполнения удобны для оценки работы алгоритма, так как частоты звучания нот заранее известны (таблица 1.1).

В качестве примера, на рисунке 2.11 представлен результат одновременной маскировки, полученный для ноты «ми второй октавы». Результат идентификации частот в данной ноте (рисунок 2.12) позволяет определить, что в момент исполнения ноты, ЧОТ сигнала колебались в диапазоне от 619 до 679 Гц. При этом основная часть частот принадлежит уровню в 659 Гц, что соответствует частоте звучания анализируемой ноты.

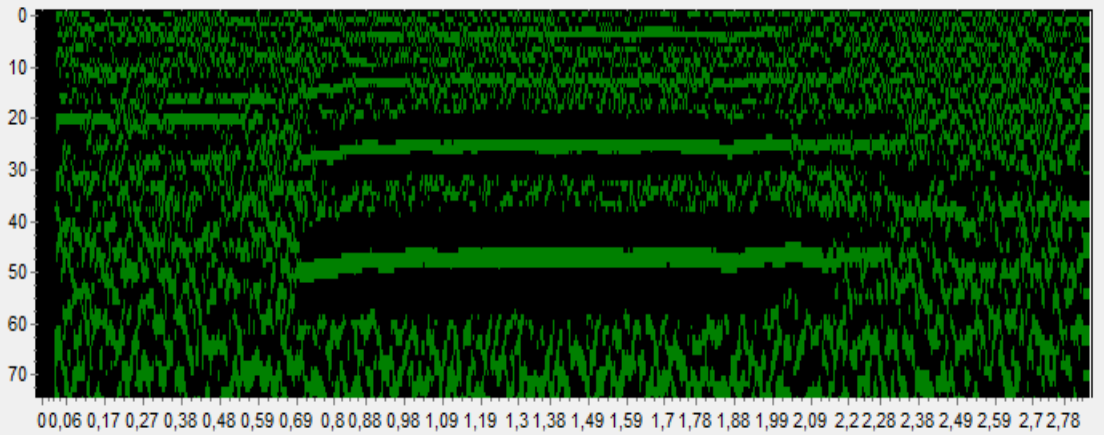


Рисунок 2.11 – Результат одновременной маскировки для спетой во 2 октаве ноты «ми» (частота звучания 659.26 Гц)

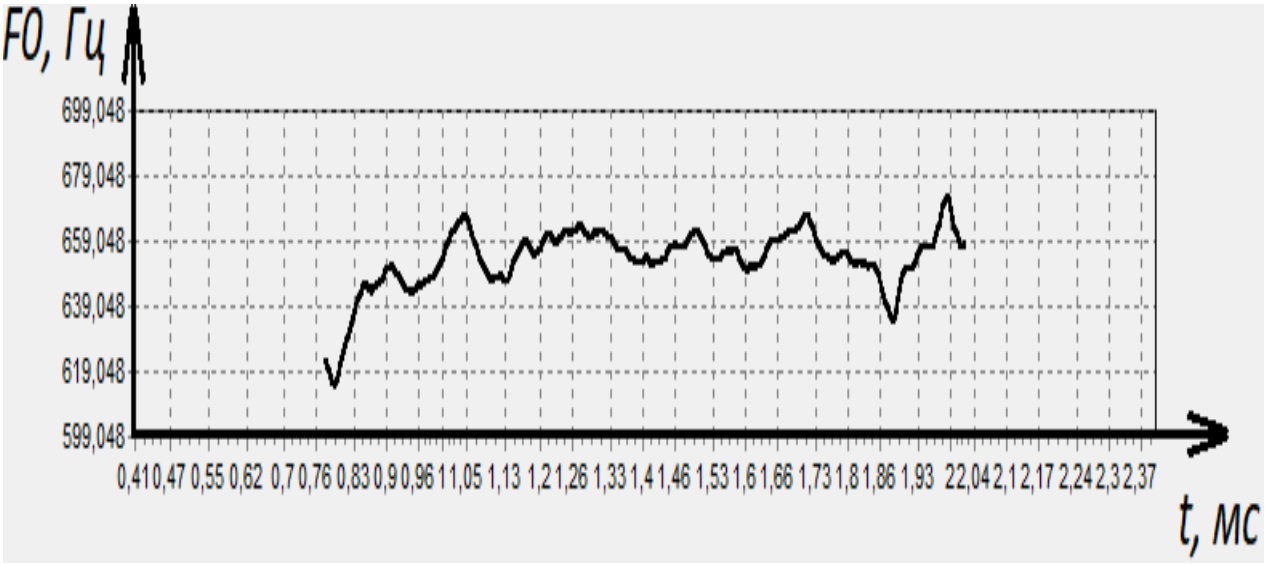


Рисунок 2.12 – График зависимости частот основного тона от времени для спетой во 2 октаве ноты «ми» (частота звучания 659.26 Гц)

Как можно увидеть из полученных графиков, пение ноты в записи происходило с 0.8 до 2.2 секунды. Участки от 0 до 0.8 секунды и от 2.2 до 2.37 секунды соответствуют тишине, воспринимаемой алгоритмом как невокализованные сегменты.

С целью проверки корректности работы алгоритма автоматической генерации набора шаблонов для произвольных границ определения частоты основного тона был сгенерирован набор для диапазона от 70 до 400 Гц. Полученный набор был сравнен с набором, который был эмпирически сформирован в исходной модели. Сравнение показало полное совпадения наборов для указанного диапазона. Также были исследованы шаблоны, генерируемые для отдельных интервалов определения частоты основного тона. В качестве эксперимента были сгенерированы шаблоны для интервалов частот: от 70 до 200 Гц, от 200 до 400 Гц, от 400 до 600 Гц и от 600 до 800 Гц. Для каждого диапазона была сгенерирована синусоида с ЧОТ, лежащей в указанном диапазоне: 150 Гц, 300 Гц, 450 Гц, 600 Гц и 750 Гц.

На рисунках 2.13-2.17 приводятся результаты обработки каждой синусоиды с помощью набора шаблонов, сгенерированного для указанного диапазона частот. Полученные графики свидетельствуют о корректном определении частот основного тона на сгенерированных наборах шаблонов.

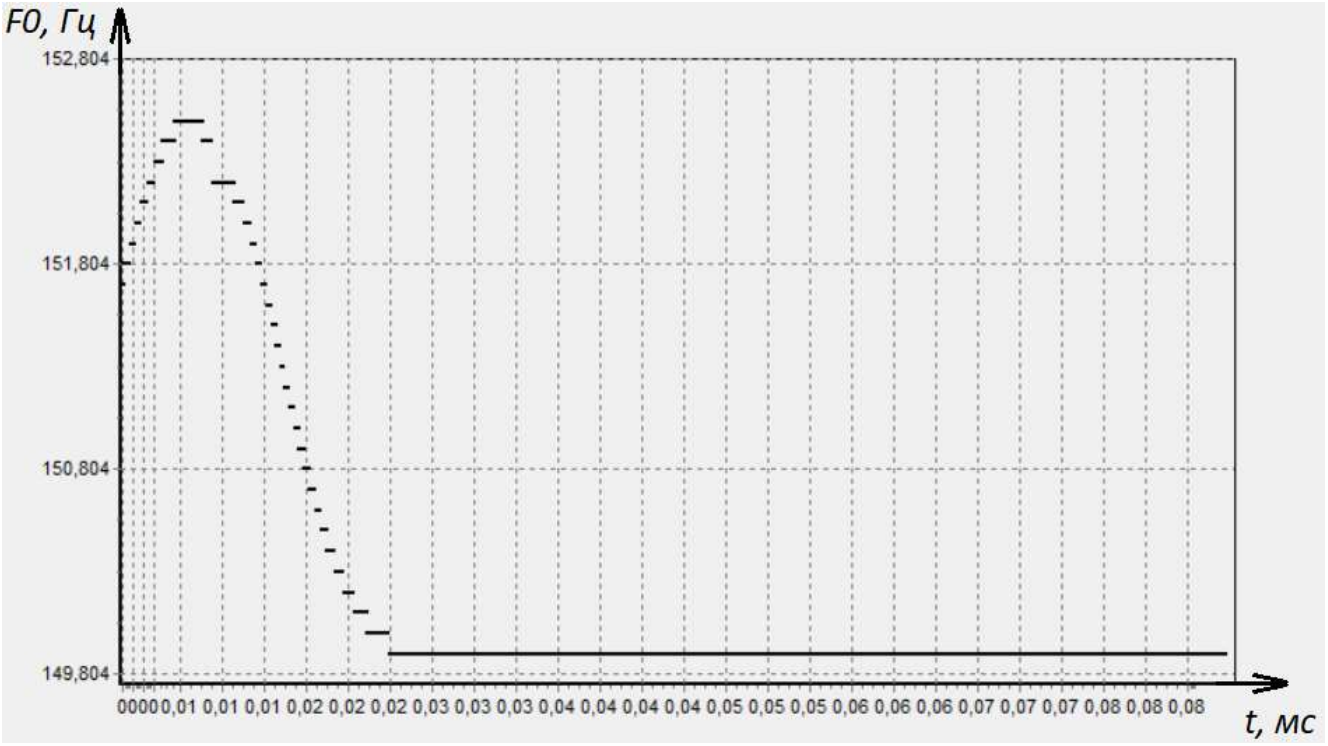


Рисунок 2.13 – График зависимости частоты основного тона для синусоидального сигнала с ЧОТ 150 Гц на наборе шаблонов 70-200 Гц

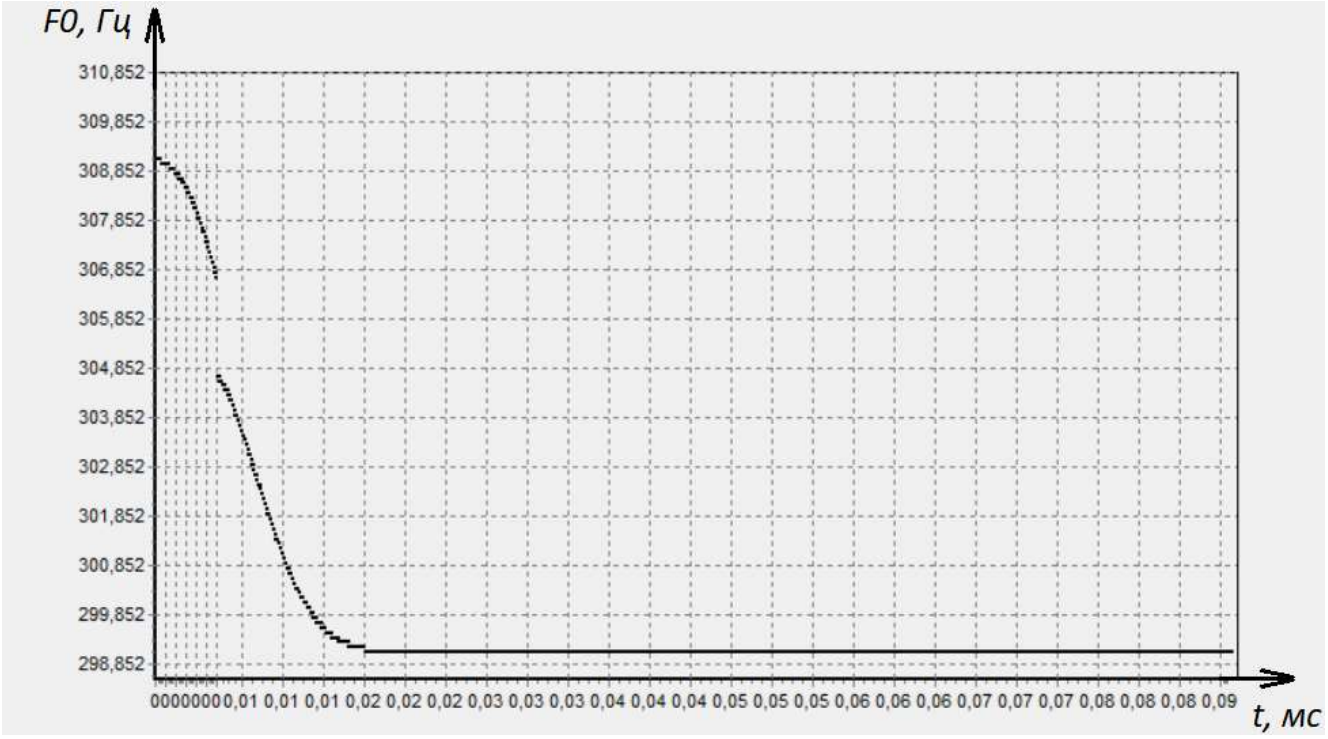


Рисунок 2.14 – График зависимости частоты основного тона для синусоидального сигнала с ЧОТ 300 Гц на наборе шаблонов 200-400 Гц

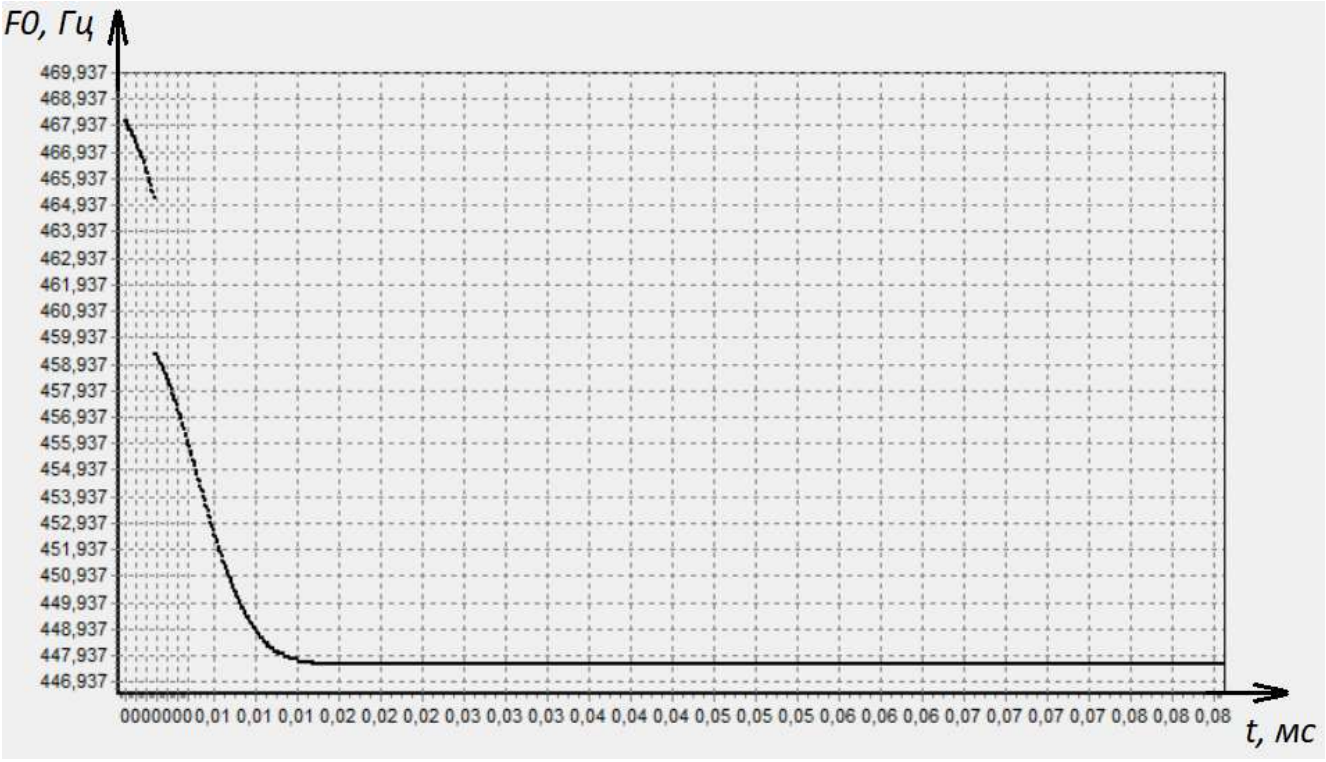


Рисунок 2.15 – График зависимости частоты основного тона для синусоидального сигнала с ЧОТ 450 Гц на наборе шаблонов 400-600 Гц

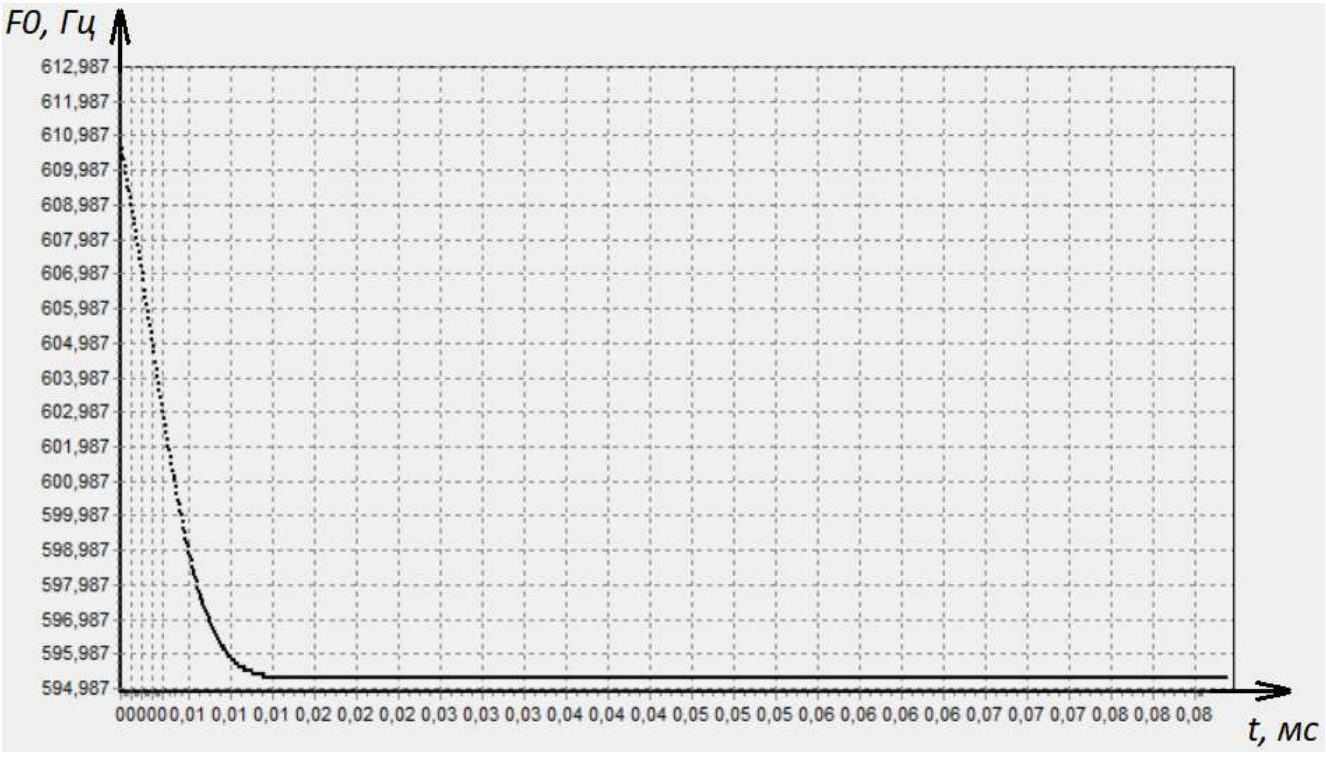


Рисунок 2.16 – График зависимости частоты основного тона для синусоидального сигнала с ЧОТ 600 Гц на наборе шаблонов 600-800 Гц

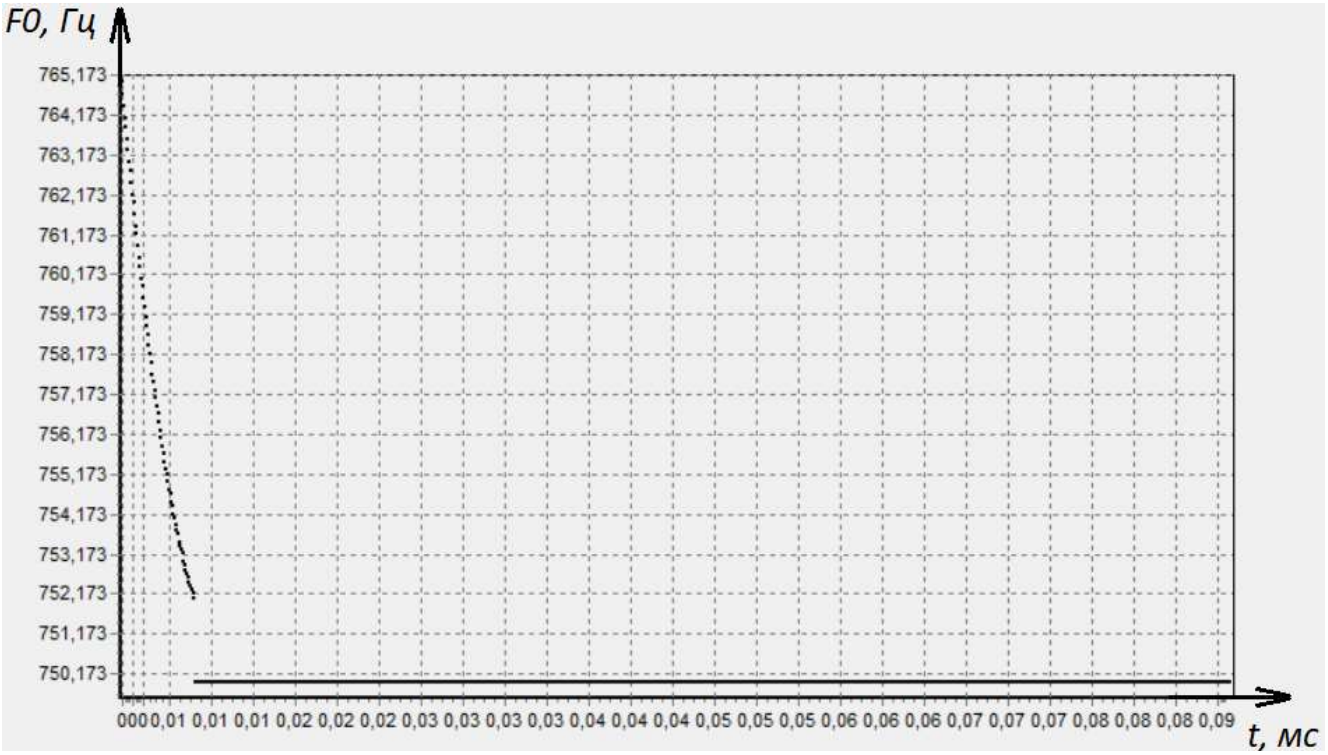


Рисунок 2.17 – График зависимости частоты основного тона для синусоидального сигнала с ЧОТ 750 Гц на наборе шаблонов 600-800 Гц

Результаты анализа синусоидальных сигналов с помощью наборов шаблонов для отдельных диапазонов частот были сопоставлены с результатами анализа при общем наборе шаблонов, сгенерированном для диапазона от 70 до 800 Гц. Было определено, что точность определения частот основного тона не зависит от размера исследуемого диапазона. Таким образом, при обработке сигналов с частотой основного тона, лежащей в диапазоне от 70 до 800 Гц, не требуется предварительная обработка с целью определения диапазона вычисления точного значения ЧОТ.

2.4 Выводы по главе

Была произведена модификация математической модели слуховой системы за счет внесения возможности автоматического учета границ определения частоты основного тона сигнала. Модель была протестирована на сгенерированных синусоидальных сигналах. Полученные результаты по идентификации частот основного тона показали высокую точность в диапазоне от 70 до 800 Гц включительно. Относительная ошибка алгоритма идентификации ЧОТ составила менее 1%, что позволяет применить модифицированную математическую модель

слуховой системы человека не только для анализа параметров речевого сигнала [132], но и для идентификации нот. Таким образом, становится возможной разработка алгоритма распознавания нот. Данный алгоритм должен включать в себя этапы сегментации и идентификации нот.

Алгоритм автоматической генерации шаблонов для указанного диапазона частот был проверен на диапазоне, обрабатываемом исходной математической моделью. Полученный автоматически набор шаблонов для частот основного тона в диапазоне от 70 до 400 Гц соответствует набору, полученному эмпирически в [70]. Кроме того, алгоритм автоматической генерации шаблонов был исследован на предмет влияния размера указываемого диапазона на точность вычисления частоты основного тона. Были получены наборы шаблонов для отдельных отрезков частот в диапазоне от 70 до 800 Гц, которые далее были протестированы на синусоидальных сигналах. Было определено, что точность определения частот основного тона не зависит от размера исследуемого диапазона.

3 Алгоритм распознавания нот вокального исполнения

Предлагаемая идея распознавания спетых нот заключается в использовании частот основного тона, идентифицированных для вокального исполнения с применением шаблонов, соответствующих модели слуховой системы человека.

Распознавание нот в вокальном исполнении включает в себя следующие шаги:

1) вычисление частот основного тона для сигнала в каждый момент времени;

2) сегментация и идентификация нот на основании полученного массива частот в каждый момент времени;

3) корректировка сегментированных нот с учетом минимальной длительности звучания.

3.1 Алгоритмы сегментации, автоматизации оценки качества сегментации и идентификации нот в вокальном исполнении

Как было сказано в главе 1, каждой ноте в теории музыки соответствует конкретное значение частоты основного тона. В связи с тем, что пение ноты на конкретном значении является невыполнимой задачей, было принято решение перейти от дискретных значений к интервалам частот. На основании данных из таблицы 1.1 было получено множество отрезков для каждой ноты, где нижняя границы ноты и верхняя границы ноты определялись по формулам 3.1 и 3.2 соответственно.

$$f_{i\text{н}} = \frac{f_{i-1}+f_i}{2} \tag{3.1}$$

$$f_{i\text{в}} = \frac{f_{i+1}+f_i}{2} \tag{3.2}$$

где f_i – частота i-й ноты;

$f_{i\text{н}}$ и $f_{i\text{в}}$ – значения частот для нижней и верхней границ, соответственно.

Отдельный этап исследования вокального исполнения был посвящен определению корректности применения среднего арифметического при вычислении граничных значений. В работах [133-134] было проведено сравнение результатов идентификации нот при различных подходах к усреднению. К применяемому методу были добавлены: средняя гармоническая, средняя геометрическая, средняя арифметическая, средняя квадратическая и средняя

кубическая. Поскольку при определении значения частоты следующей ноты ее значение умножается на 2 в степени 1/12, было решено применить данных подход для определения граничных значений нот. В полученной методике усреднения, названной логарифмической, применялись формулы 3.3 и 3.4 для вычисления граничных значений.

$$f_{i\text{н}} = f_{i-1} \cdot 2^{\frac{1}{24}} \tag{3.3}$$

$$f_{i\text{в}} = f_i \cdot 2^{\frac{1}{24}} \tag{3.4}$$

где f_i – частота i-й ноты;

$f_{i\text{н}}$ и $f_{i\text{в}}$ – значения частот для нижней и верхней границ, соответственно.

В таблице 3.1 представлен фрагмент полученного набора шкал с интервалами в диапазоне от ноты «ля 1-й октавы» до ноты «соль-диез 2-й октавы».

Таблица 3.1 – Шкалы идентификации нот (фрагмент)

Частота звучания ноты	Среднее											
	Гармоническое		Логарифмическое		Геометрическое		Арифметическое		Квадратическое		Кубическое	
	Ниж.гр.	Верх.гр.	Ниж.гр.	Верх.гр.	Ниж.гр.	Верх.гр.	Ниж.гр.	Верх.гр.	Ниж.гр.	Верх.гр.	Ниж.гр.	Верх.гр.
698,456	678,290	718,623	678,571	718,921	678,572	718,923	678,856	719,223	679,138	719,522	679,421	719,822
739,989	718,623	761,355	718,921	761,671	718,923	761,672	719,223	761,990	719,522	762,308	719,822	762,625
783,991	761,355	806,627	761,671	806,962	761,672	806,963	761,990	807,300	762,308	807,636	762,625	807,972
830,609	806,627	854,591	806,962	854,946	806,963	854,948	807,300	855,305	807,636	855,661	807,972	856,017

Сравнение полученных шкал показало, что отклонение от эталонного значения ноты для каждого из методов усреднения данных составляет от 2.72 % у кубического метода до 2.88% у гармонического для нижней границы и от 2.88% у гармонического до 3.05 у кубического для верхней границы. В среднем ширина интервала относительно значения исследуемой ноты составляет: для гармонического метода – 5.774%, для логарифмического и геометрического – 5.777%, для арифметического – 5.779%, для квадратического – 5.782%, а для кубического – 5.784%. В ходе вычислительного эксперимента было замечено, что с помощью программы были получены идентичные результаты вне зависимости от выбора шкалы. Причиной для подобного результата может служить то, что среди записей самой высокой исполненной нотой является «фа-диез второй октавы», звучащая на высоте 739.98 Гц. На данном уровне частот разница между границами

интервалов разных шкал достаточно мала: разница между средним гармоническим и средним кубическим не превышает 1.5 Гц, что составляет 0.17% от значения частоты, на которой прозвучала исследуемая нота.

Основанием для сегментации аудиосигнала с вокальным исполнением на вокализованные и невокализованные участки служит применение понятия минимальной длительности звучания ноты (рисунок 3.1).

Рисунок 3.1 – Сегментация вокализованных и невокализованных участков с учетом минимальной длительности звучания ноты

В одной секунде аудиозаписи содержится количество значений мгновенной частоты основного тона равное частоте дискретизации записи. Следует учесть, что разброс между соседними значениями может быть существенно больше границ ноты. Данные разрывы могут быть расценены как индикаторы границ вокализованного участка, так и случайным всплеском в рамках одного участка. На рисунке 3.1 схематично представлен процесс сегментации. В верхней части схемы представлена минимальная длительность ноты, в центральной части участки различной величины с множеством значений частот основного тона, принадлежащих одному классу (ноте), а в нижней части рисунка – результат определения вокализованности участка. В левой части представлена ситуация, когда в множестве частот одной ноты возникают участки, отнесенные к другим нотам. В данном случае алгоритм воспринимает вокализованный участок целостным, если его общая длительность с учетом присутствия посторонних нот, выше минимальной длительности звучания ноты. В дальнейшем такие ноты дополнительно оцениваются алгоритмом на предмет чистоты звучания (процент посторонних частот). В средней части представлена ситуация, где в полученном

множестве частот общая длительность меньше минимальной длительности звучания ноты. В таком случае участок относится к невокализованным и воспринимается как шум. В правой части представлена ситуация, где множество частот, отнесенных к ноте, составляют минимальную длительность звучания. Такая ситуация может быть расценена как чисто спетая нота.

С учетом данных об интервалах звучания нот массив частот основного тона переводится в массив звучащих нот для каждого момента времени. Последовательности идентичных нот в данном массиве преобразуются в набор сегментов различной длительности, все частоты в которых относятся к вычисленным диапазонам звучания нот. В рамках алгоритма идентификации нот этот процесс можно представить в виде обобщенной блок-схемы сегментации нот по длительности звучания [135]. В представленной схеме (рисунок 3.2) используются следующие обозначения:

A – начало ноты;

B – конец ноты;

MinDuration – минимальная длительность ноты;

det – диапазон ноты / 2;

Note( ) – функция определения принадлежности частоты к ноте;

notes – массив найденных нот;

note – ноты из массива notes;

NOTESN – массив нот после проверки на минимальную длительность.

На начальной стадии у алгоритма обнуляются данные о начале и конце обрабатываемой ноты, задаются минимальная длительность звучания ноты и диапазон, в пределах которого определяется принадлежность к ноте. В случае, если в данный момент нет вокализованного участка, оцениваемого на соответствие установленным критериям принятия решения о найденной ноте, алгоритм закрепляет первый из необработанных участков как эталон. Затем для каждого сегмента звучания проверяется факт отсутствия шумов или тишины. Если данный сегмент содержит в себе ноту, то она сравнивается с закрепленным эталоном. В случае совпадения – длительность отрезка прибавляется к длительности

эталонного. Если в сегменте была другая нота, шум или тишина, то данный отрезок добавляется к длительности тишины. Пока длительность тишины меньше минимальной длительности звучания ноты, данные действия повторяются. Если в итоге длительность звучания ноты была меньше минимальной длительности звучания ноты, то эталонный сегмент приравнивается к тишине, а следующий за ним принимается как эталон для анализа.

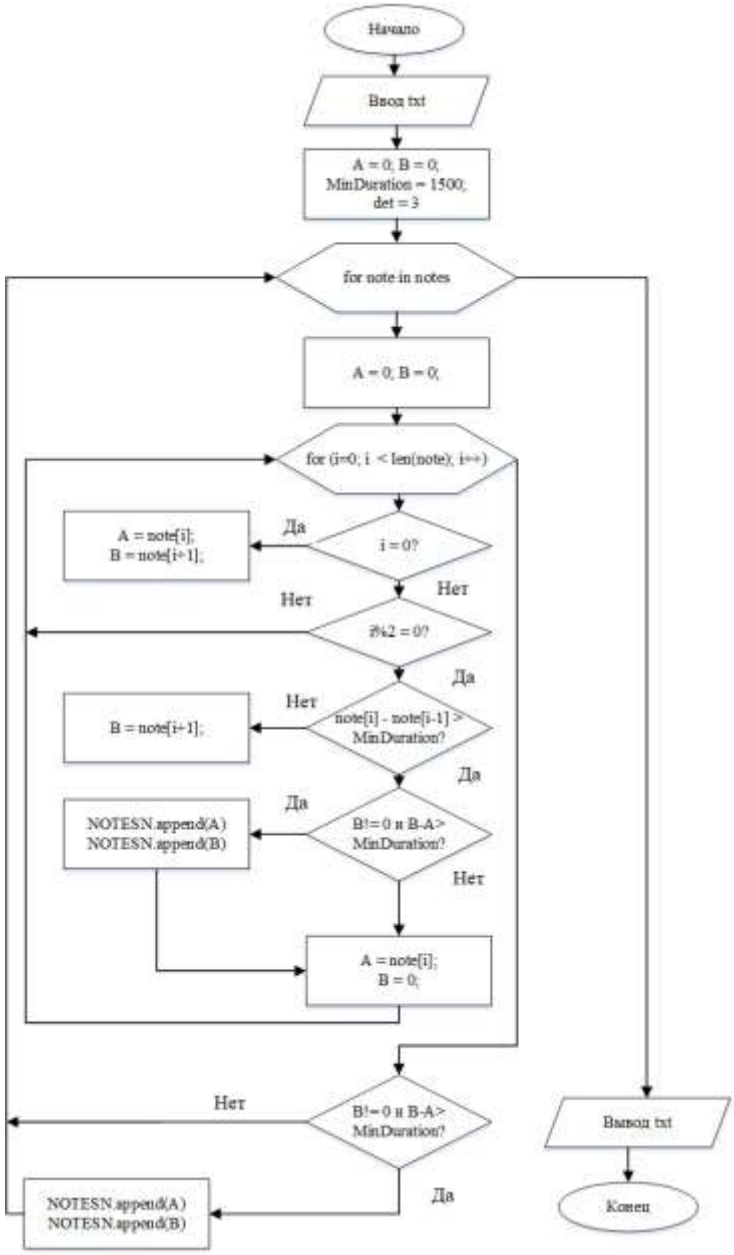


Рисунок 3.2 – Алгоритм сегментации нот по длительности звучания

Позже данный подход был исследован на речевых сигналах в [136], где сегментация осуществлялась на вокализованные и невокализованные участки на основании минимальной меры различия. При этом вокализованные участки

включают в себя звонкие согласные и сонорные звуки, а невокализованные – глухие согласные и «тишину».

Алгоритм основан на анализе в каждый дискретный момент времени частотной области, включающей две гармоники речевого сигнала (два непрерывных интервала единиц определенной длины, разделенных интервалом нолей). Для этого создается набор шаблонов, с которыми сравнивается структура сигнала в текущий момент времени. Шаблоны включают в себя первую и вторую гармоники основного тона. После прохождения сигнала через систему фильтров [137-138] на каждом временном отрезке производится его свёртка с частотной маской.

Алгоритм сегментации состоит из двух этапов:

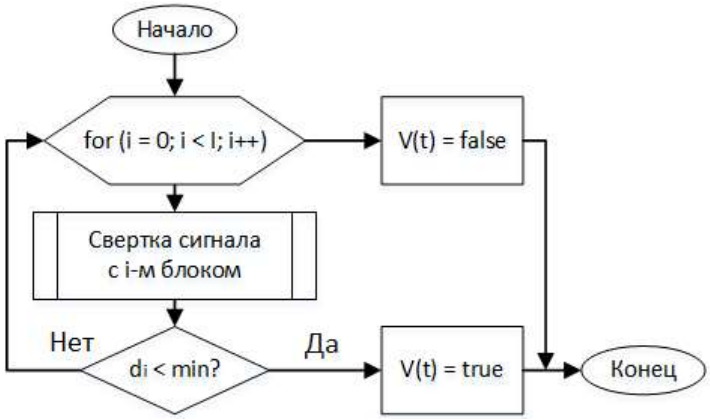
1) определение вокализованности текущего временного отсчета;

2) сегментация речевого сигнала на вокализованные и невокализованные участки.

На первом этапе используются три способа вычисления меры различия. Для первого способа вычисляется количество отличающихся каналов в шаблоне. Для второго – считается количество отличающихся каналов в шаблоне и делится на общее количество каналов, которое было в самом шаблоне. Для третьего способа вместо отличающихся, суммируются совпавшие каналы в шаблоне.

Минимумы меры различия находятся в точках совпадения с маской, имеющей такую же частоту основного тона. Величина минимума различается для каждой маски, но остаётся постоянной для любого речевого сигнала. Если в какой-то момент времени эта величина больше заданного порога *min*, то данный участок признаётся невокализованным. На рисунке 3.3 представлен алгоритм определения вокализованности текущего временного отсчета.

Для каждого момента времени осуществляется свертка всего перечня шаблонов с обрабатываемым сигналом. В случае, если хоть у одного из шаблонов мера различия с сигналом будет меньше порога вокализации, система получает сигнал о вокализации текущего момента времени. В противном случае момент времени считается невокализованным.



Обозначения:
$V(t)$ – признак вокализации текущего момента времени;
d_i – мера различия i-го шаблона и сигнала;
min – порог вокализации;
i – порядковый номер шаблона;
I – количество шаблонов.

Рисунок 3.3 – Базовый алгоритм определения вокализованности временного отсчета

При исследовании, является ли отсчет вокализованным или нет, учитывается наличие определенного значения порога. Выброс в динамике минимальной меры различия на вокализованных участках часто превышает этот порог. При определенных уровнях порога участок с выбросом определится как невокализованный, что в итоге может привести к дополнительным ошибкам в сегментации. На рисунке 3.4 представлен усовершенствованный алгоритм сегментации речевого сигнала на вокализованные и невокализованные участки на основе сглаживания значений минимальной меры различия. Данный алгоритм отличается добавлением этапа с сглаживанием значений мер различия, осуществляющимся по алгоритму на рисунке 3.5. Перед тем как перейти к сглаживанию, в алгоритме осуществляется поиск наличия тех моментов времени, в которых результат маскировки несопоставим ни с одним из элементов набора шаблонов. В случае обнаружения таких моментов текущее значение меры различия, отнесенное к общему количеству каналов в шаблоне, сравнивается с минимальной мерой различия. Если оно окажется меньше, то минимальная мера различия приравнивается к текущей мере различия.

При сглаживании происходит определение минимальной меры различия для каждого из значений в выбранном окне. Первым шагом осуществляется сдвиг на 1 позицию в массиве значений мер различия. После этого определяется сумма всех значений в обрабатываемом диапазоне без учета текущего минимального значения меры различия. Полученная сумма делится на общее количество временных

отсчетов в выбранном окне и возвращается в общий алгоритм, где сравнивается с порогом при определении периодической структуры сигнала.

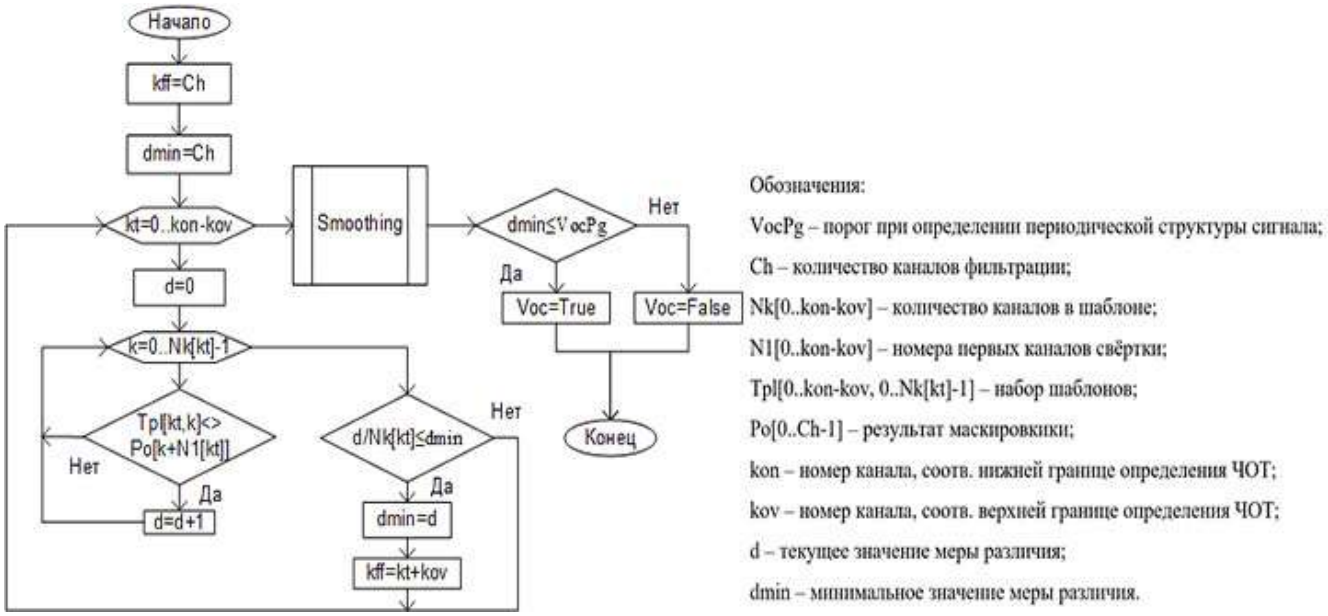


Рисунок 3.4 – Модифицированный алгоритм определения вокализованности временного отсчета с учетом получения минимального значения меры различия

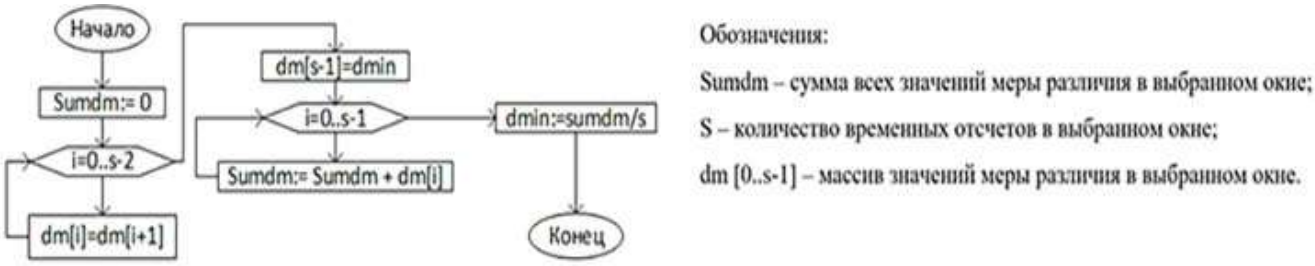


Рисунок 3.5 – Блок-схема подпрограммы,
отвечающей за сглаживание значений

Проведенное исследование по сбору статистических данных о качестве сегментации на вокализованные и невокализованные участки выявило, что без сглаживания результаты сегментации получаются гораздо хуже, чем со сглаживанием. Было определено сочетание входных параметров [136], при котором результаты автоматической сегментации наиболее приближены к результатам ручной сегментации сигнала. Тестирование сегментации речевого сигнала на вокализованные и невокализованные участки осуществлялось для аудиозаписей с произнесением слогов [139]. Полученные результаты были учтены в алгоритме сегментации и идентификации нот.

Кроме того, в программе были учтены такие параметры, как темп исполнения, что позволило повысить точность идентификации нот за счет возможности определения минимальной длительности звучания нот по оценке характеристик аудиосигнала. С целью определения качества исполнения диктором ноты была добавлена оценка относительно максимальной длительности ошибок подряд.

Схожий подход был применен при анализе шепотной речи [140]. Основное отличие шепотной речи от вокализованной состоит в том, что шепотная речь происходит от турбулентного шума, создаваемого трением воздуха в гортани и над ней. Голосовые связки при этом не вибрируют. Таким образом, в шепотной речи отсутствует частота основного тона и гармоническая структура [141]. Одна из задач, которая при этом преследовалась, заключалась в определении возможности нахождения вокализованных участков и нот в условиях отсутствия голосового сигнала. Исследование показало, что алгоритм работает исправно и не определяет лишних участков. Вокализованные участки не будут определяться, если в сигнале присутствует шепот без голоса.

Для каждого момента времени определено значение частоты основного тона, поэтому возможно определить, какая нота «прозвучала» в данный отсчет. Последовательно обрабатывается каждый момент, что в результате позволяет получить массив отрезков звучания той или иной ноты. Полученный массив оценивается аналогично алгоритму, представленному на рисунке 3.2. В случае если в пределах минимальной длительности ноты обнаруживается сегмент, принадлежащий той же ноте, что и эталонная для данного этапа оценки, то этот сегмент с всеми сегментами между ним и эталонным складываются в общую длительность ноты, к чистому звучанию ноты добавляется длительность звучания текущей ноты. К длительности неправильных нот добавляются длительности элементов, встреченных с момента последней правильной ноты до текущей оцениваемой. Данные длительности в итоге оцениваются при определении чистоты исполнения текущей ноты и вынесении решения при выставлении оценки студенту. В случаях, если в пределах минимальной длительности ноты не

обнаружен сегмент той же ноты, что и эталонная, то за эталонную принимается следующая нота, а текущая нота приравнивается к тишине.

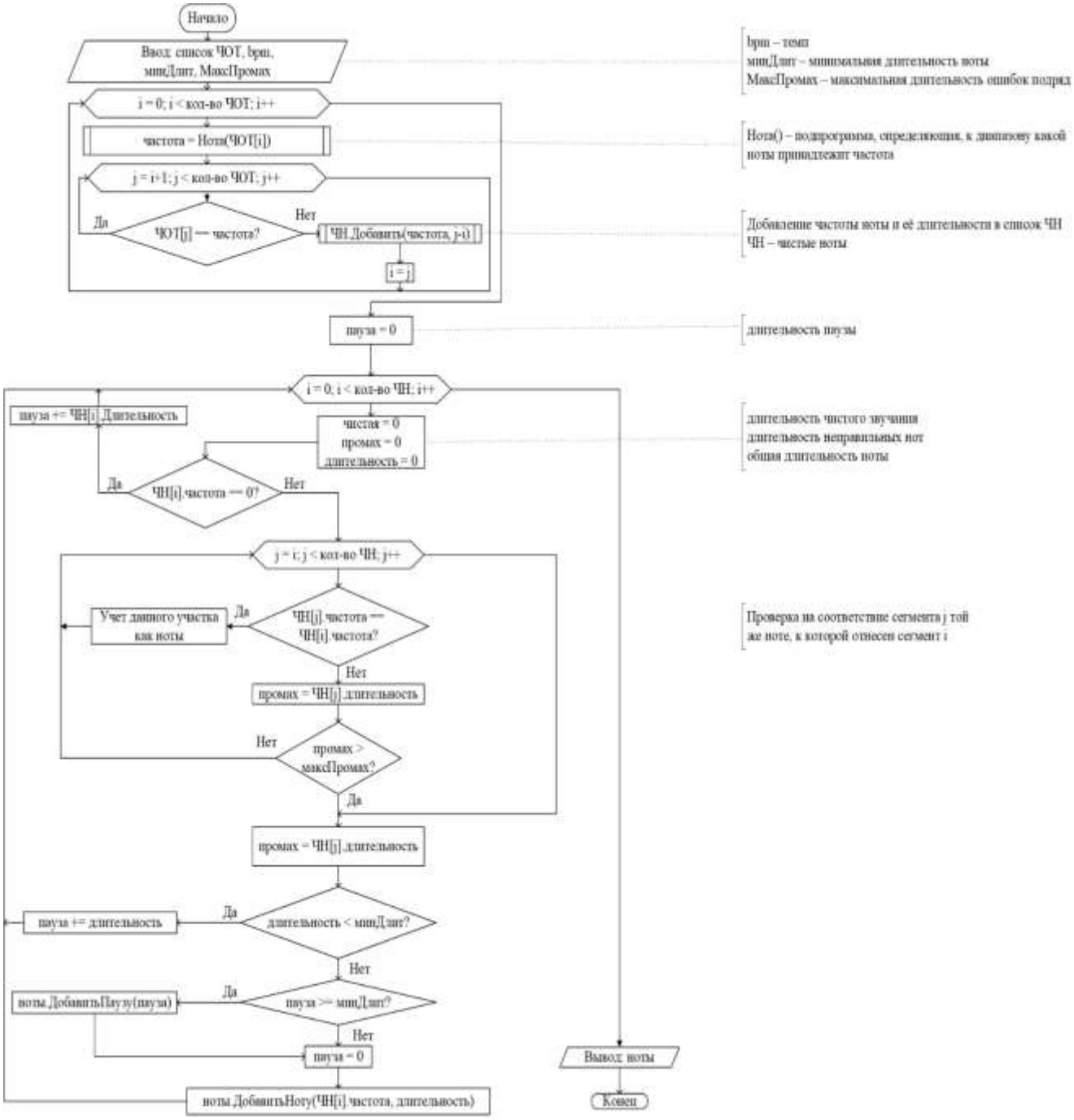


Рисунок 3.6 – Алгоритм сегментации и идентификации нот

Оценка качества пения осуществляется по двум критериям: точность попадания в тон P и точность попадания в ритм R. Как было сказано выше, оценка нот производится на основе сопоставления найденных и эталонных нот. Эталонным нотам ставятся в соответствие найденные ноты по принципу: если N%

длительности звучания ноты по времени соответствует эталонной ноте, то найденная нота ставится ей в соответствие.

Для определения точности попадания в тон определяется средняя ЧОТ - *M* нот, соответствующих эталонной. Значение *M* вычисляется по формуле 3.5.

$$M_k = \frac{\sum_{j}^{i} ЧОТ}{j - i}, \tag{3.5}$$

где k – порядковый номер эталонной ноты,

j – порядковый номер первой и последней ЧОТ, соответствующей эталонной ноте.

Отклонение в центах (сотая часть полутона) *C* определяется по следующей формуле 3.6.

$$C_k = \begin{cases} \dfrac{M_k - F_l}{F_l - F_{l-1}}, \ если \ M_k < F_l \\ \dfrac{M_k - F_l}{F_{l+1} - F_l}, \ если \ M_k > F_l \end{cases}, \tag{3.6}$$

где k – это порядковый номер эталонной ноты,

F_l – частота эталонной ноты,

F_{l-1} – частота ноты на полутон ниже эталонной,

F_{l+1} – частота на полутон выше эталонной.

Точность попадания в тон P вычисляется по формуле 3.7.

$$P_k = 1 - |C_k| * 100\% \tag{3.7}$$

На рисунке 3.7 представлено отображение программным комплексом информации о полученных отклонениях между эталонными и вычисленными значениями. Кружком выделен участок, на котором при пении колебания осуществлялись в двух нотах вместо одной. В результате, при оценке исполнения текущей ноты будет сообщено, что длительность исполнения была меньше заданного (на длину лишней ноты).

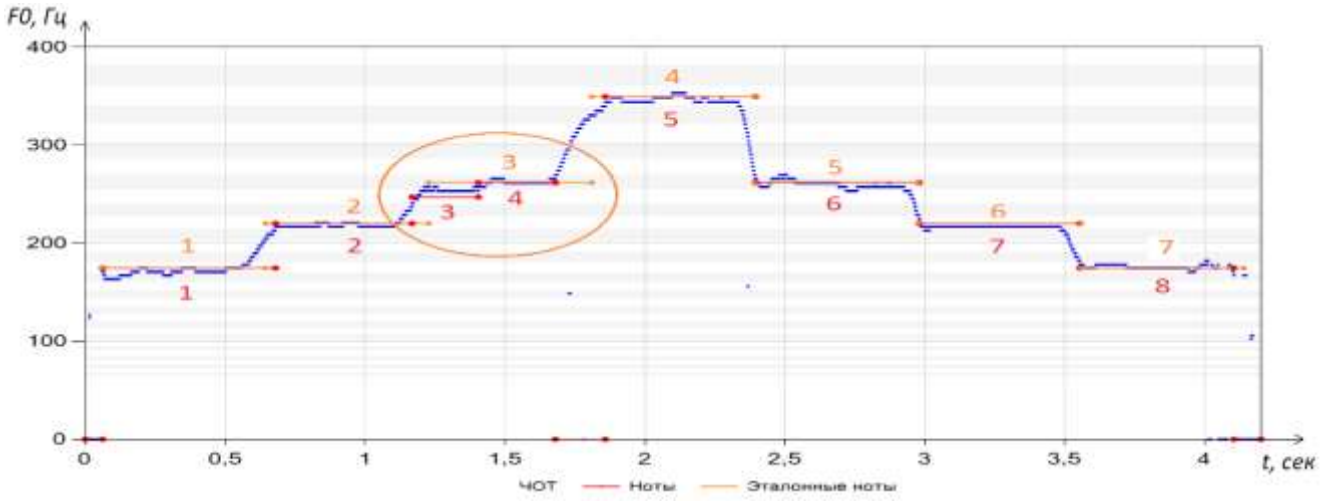


Рисунок 3.7 – Оценка качества пения

Для вычисления точности попадания в ритм вычисляется разница D_1 между началом эталонной ноты T_{R1} и началом первой найденной ноты T_{N1}, соответствующей эталонной, а также разница D_2 между концом эталонной T_{R2} ноты и концом последней найденной ноты T_{N2} (рис. 3.8) применяется формула 3.8.

$$D_1 = T_{R1} - T_{N1}; D_2 = T_{R2} - T_{N2} \quad (3.8)$$

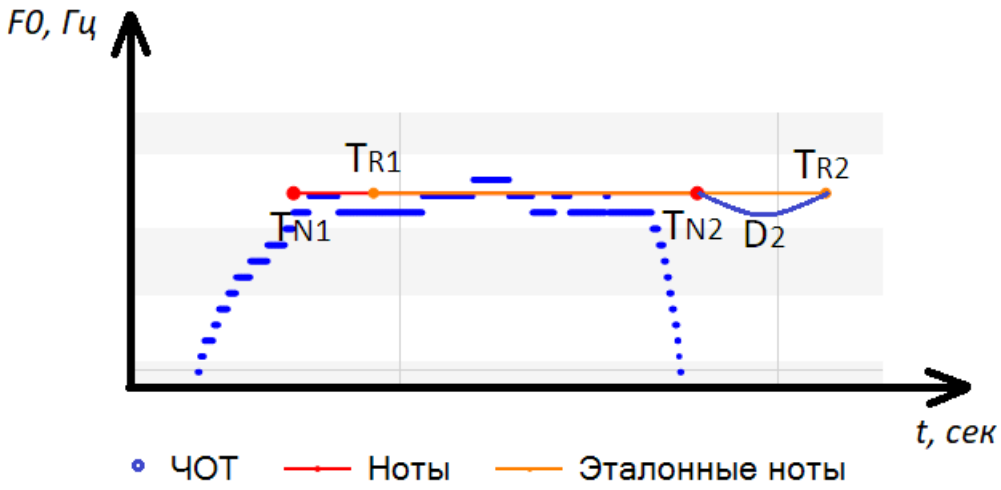


Рисунок 3.8 – Определение точности попадания в ритм

Точность попадания в ритм вычисляется по формуле 3.9.

$$R_k = \left(1 - \frac{|D_1| + |D_2|}{D_k}\right) \cdot 100\% \quad (3.9)$$

где D_k – длительность эталонной ноты.

Для оценивания всей спетой вокальной партии вычисляются средние значения точности попадания в тон и в ритм по формулам 3.10 и 3.11, соответственно.

$$R = \frac{\sum_{k=1}^{n} R_k}{n}; \tag{3.10}$$

$$P = \frac{\sum_{k=1}^{n} P_k}{n} \tag{3.11}$$

3.2 Проведение экспериментов на нотах

При исследовании надежности детектора частот основного тона [142] были определены несколько стратегий, на основе которых были сделаны 16 записей (8 женским голосом и 8 мужским голосом):

− тестирование коротких звуков (стаккато),

− тестирование звуков, находящихся на минимальном интервале,

− тестирование звуков, находящихся в среднем интервале,

− тестирование звуков с произношением слов,

− тестирование мелодии с произношением слов.

8 записей женского голос включают в себя:

№1. Локация: первая октава. Ноты: до, ре, ми, фа, соль, фа, ми, ре, до. Стаккато (отрывисто), без произношения нот (только звуки).

№2. Локация: первая октава. Ноты: до, ре, ми, фа, соль, фа, ми, ре, до. Легато (связанно), без произношения нот (только звуки).

№3. Локация: первая октава. Ноты: си, ре-диез, фа-диез, ре-диез, си. Стаккато, без произношения нот.

№4. Локация: малая октава, первая октава. Ноты: си, ре-диез, фа-диез, ре-диез, си. Легато, без произношения нот.

№5. Локация: малая октава, первая октава. Ноты: ля, ля-диез, си, до, до-диез, до, си, ля-диез, ля. Стаккато, без произношения нот.

№6. Локация: малая октава, первая октава. Ноты: ля, ля-диез, си, до, до-диез, до, си, ля-диез, ля. Легато, без произношения нот.

№7. Локация: первая октава. Ноты: до, ми, ре, фа, ми, соль. Стаккато, с произношением нот.

№8. Локация: первая октава. Ноты: до, ми, ре, фа, ми, соль. Легато, с произношением нот.

8 записей мужского голоса включают в себя:

№1. Локация: малая октава. Ноты: до, ре, ми, фа, соль, фа, ми, ре, до. Стаккато (отрывисто), без произношения нот (только звуки).

№2. Локация: малая октава. Ноты: до, ре, ми, фа, соль, фа, ми, ре, до. Легато (связанно), без произношения нот (только звуки).

№3. Локация: малая октава. Ноты: си, ре-диез, фа-диез, ре-диез, си. Стаккато, без произношения нот.

№4. Локация: малая октава. Ноты: си, ре-диез, фа-диез, ре-диез, си. Легато, без произношения нот.

№5. Локация: малая октава. Ноты: ля, ля-диез, си, до, до-диез, до, си. Стаккато, без произношения нот.

№6. Локация: малая октава. Ноты: ля, ля-диез, си, до, до-диез, до, си, ля-диез, ля. Легато, без произношения нот.

№7. Локация: малая октава. Ноты: до, ми, ре, фа, ми, соль. Стаккато, с произношением нот.

№8. Локация: малая октава. Ноты: до, ми, ре, фа, ми, соль. Легато, с произношением нот.

Отличие мужского набора записей заключалось в отсутствии в записи № 5 Стаккато двух последних нот «ля-диез» и «ля» ввиду проблем их воспроизведения отрывисто неподготовленным мужским голосом.

В рамках работы по повышению качества идентификации нот в автоматизированной системе распознавания вокала [143] были сопоставлены результаты обработки аудиозаписей разработанным комплексом и программами-аналогами.

Ниже представлены результаты тестирования для 6-й локации, спетой женским голосом. В аудиозаписи диктором легато исполнены следующие ноты

малой и первой октавы: ля, ля-диез, си, до, до-диез, до, си, ля-диез, ля. Пение осуществлялось без произношения нот.

На рисунке 3.9 представлен график частот основного тона для анализируемой аудиозаписи, полученный в разрабатываемом комплексе. В качестве сравнения на рисунке 3.10 показан аналогичный график, полученный в Praat. Как можно видеть, визуально данные графики похожи.

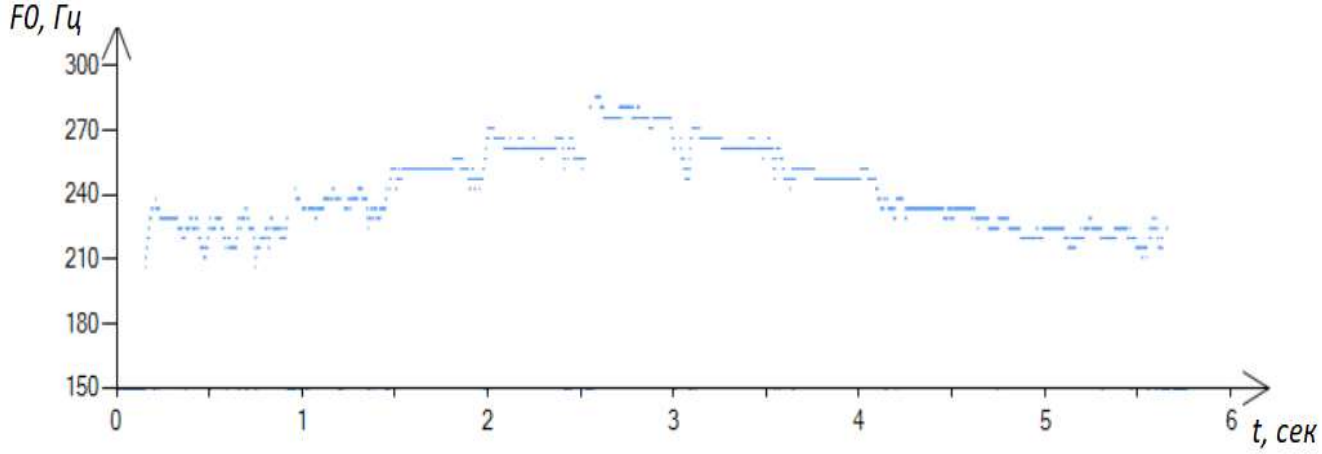


Рисунок 3.9 – График частот основного тона в первой локации

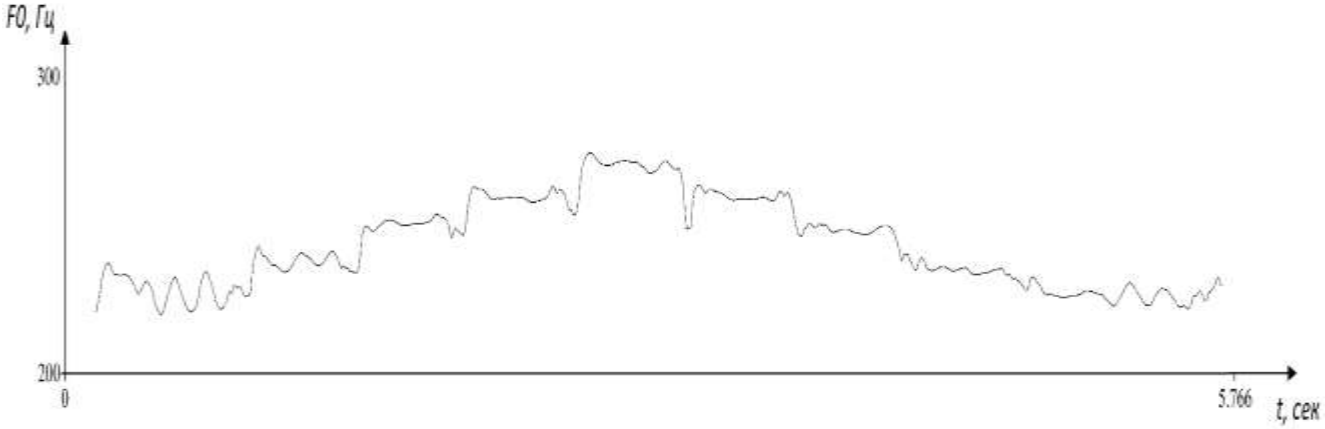


Рисунок 3.10 – График частот основного тона в Praat

В отличие от Praat, разработанный комплекс способен производить сегментацию на основании полученных частот (рисунок 3.11) спетых диктором нот. В Praat для получения нот необходимо осуществить ручную сегментацию (рисунок 3.12), записывая границы исполнения нот в ходе параллельного прослушивания записи. Для каждого выделенного сегмента определялись моменты времени, в которое звучала максимальная и минимальная частоты основного тона сигнала.

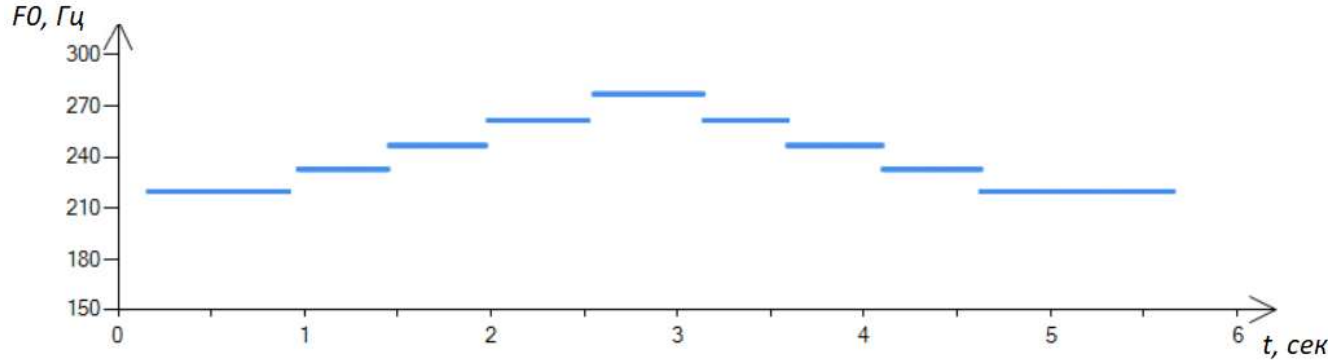


Рисунок 3.11 – График сегментированных нот в первой локации

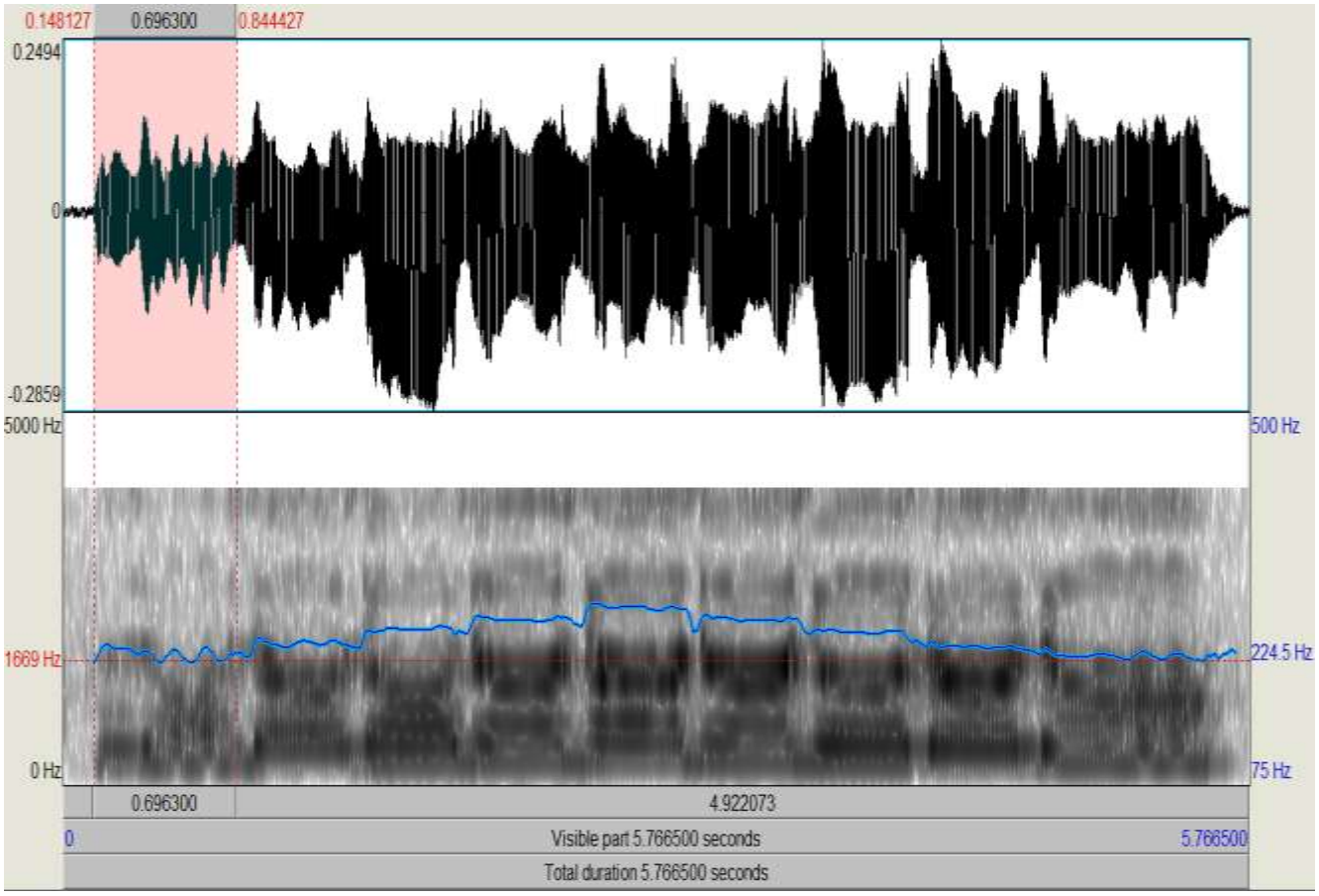


Рисунок 3.12 – Процесс ручной сегментации в Praat

На рисунке 3.13 представлен результат ручной сегментации, перенесенный на график частот. Поскольку полученные интервалы оценивались также и по высоте звучания, были получены границы для определения звучания нот (таблица 3.1). Как можно видеть в результате сравнения реально прозвучавших нот и результатов из таблицы, при ручной сегментации возможно получить значения реально прозвучавших нот при условии, если частоты основного тона идентифицированы верно. Однако, данный процесс отнимает время и человеческие ресурсы.

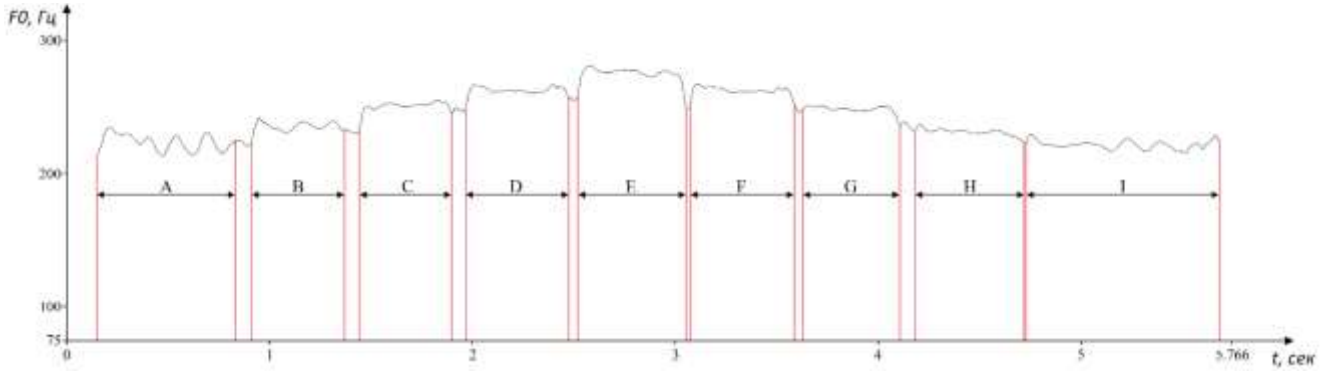


Рисунок 3.13 – Результат идентификации ЧОТ в Praat с ручной сегментацией нот

Таблица 3.2 – Результаты ручной сегментации для локации

Участок	Диапазон частот	Время звучания	Предполагаемая нота
A	216 – 228 Гц	0.69 с	Ля малой октавы
B	232 – 242 Гц	0.47 с	Си-бемоль малой октавы
C	245 – 254 Гц	0.43 с	Си малой октавы
D	256 – 264 Гц	0.51 с	До первой октавы
E	271 – 283 Гц	0.52 с	До-диез первой октавы
F	258 – 264 Гц	0.51 с	До первой октавы
G	246 – 252 Гц	0.43 с	Си малой октавы
H	232 – 237 Гц	0.51 с	Си-бемоль малой октавы
I	219 – 229 Гц	0.95 с	Ля малой октавы

На рисунке 3.14 представлен результат оценки аудиозаписи с помощью программы Melodyne. Распознанные программой ноты: A#3 (Ля-диез малой октавы), A3 (Ля малой октавы), A#3 (Ля-диез малой октавы), B3 (Си малой октавы), C4 (До первой октавы), C#4 (До-диез первой октавы), C4 (До первой октавы), B3 (Си малой октавы), A#3 (Ля-диез малой октавы), A3 (Ля малой октавы). Как видно из рисунка и результатов, алгоритм программы обнаружил небольшой всплеск в начале записи на уровне «ля-диез малой октавы», воспринятый в качестве реальной ноты. По экспертной оценке, данной ноты в аудиозаписи не прозвучало. Из этого следует, что данный участок является примером ошибки, когда алгоритм распознает ноту в момент времени, в котором ноты нет. Длительность звучания данного всплеска составила менее 0.2 с, что не было воспринято экспертом как отдельная нота и было отнесено к ноте «ля малой октавы» исполненной в начале

записи. Таким образом, часть ноты была распознана неправильно, что может быть интерпретировано как ошибка 2-го рода (ошибочно распознанная нота).

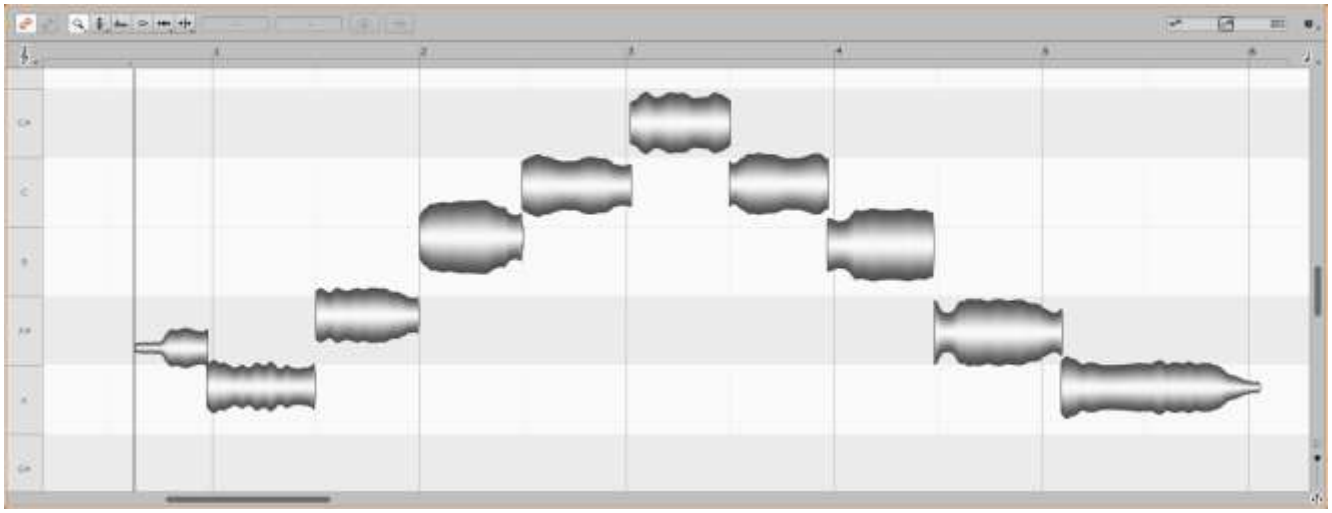


Рисунок 3.14 – Идентификация нот в Melodyne в локации

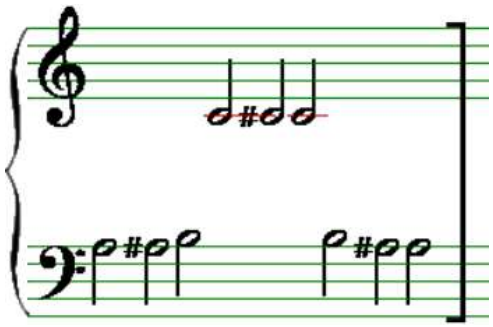


Рисунок 3.15 – Результаты идентификации нот разработанной программой

На рисунке 3.15 представлены результаты идентификации нот после сегментации, произведенной программой на рисунке 3.11. Распознанные ноты: «ля», «ля-диез», «си» – исполненные в малой октаве; «до», «до-диез», «до» – исполненные в первой октаве; «си», «ля-диез», «ля» – исполненные в малой октаве.

Исследование аудиозаписей показало схожие результаты обработки для последовательностей нот, спетых стаккато. Однако, с локациями, исполненными легато, аналоги справились хуже. Из-за специфики исполнения легато переходы между нотами, не являющимися соседними по отношению друг к другу, могут быть ошибочно восприняты за какую-либо ноту, расположенную между ними.

В общей сложности, в 16 тестах были проанализированы 114 нот, 58 из которых были спеты женским голос, а 56 – мужским. Для женского голоса были получены результаты, представленные в таблице 3.3. Для мужского голоса, аналогично представлены в таблице 3.4.

Таблица 3.3 – Результат тестирования разработанной программы и аналогов

	Praat	Melodyne	Разработанная программа
Верно распознано нот	45	58	58
Ошибочно нераспознанных (ошибки 1 рода)	13	0	0
Ошибочно распознанных (ошибки 2 рода)	0	4	0

Таблица 3.4 – Результат тестирования для мужского голоса

	Praat	Melodyne	Разработанная программа
Верно распознано нот	41	53	55
Ошибочно нераспознанных (ошибки 1 рода)	15	3	1
Ошибочно распознанных (ошибки 2 рода)	0	2	0

Таким образом, разработанная программа из 114 нот правильно идентифицировала 113 нот, 1 ноту с ошибкой. За счет применения меры минимальной длительности звучания не было определено лишних нот в местах всплесков в звучании. В изученных аналогах доля ошибок первого и второго рода существенно выше. Для программы Praat не было определено лишних нот за счет ручной обработки сигналов, а ошибки при идентификации нот на сегментах можно объяснить сложностью ручного определения звучания ноты в рамках диапазонов, охватывающих несколько нот. В программе Melodyne были получены значения нот для отклонений от звучания чистой ноты, возникающие при переходах от одной ноты к другой. Также данные переходы повлияли на границы звучания ноты, что привело к идентификации соседних нот вместо спетой диктором.

3.3 Проверка корректности экспертных оценок

С целью определения корректности оценки экспертами эталонных значений нот был собран набор из 29 нот в диапазоне от ре малой октавы до фа-диез второй октавы. В рамках исследования применялись оценки 3 экспертов, главным критерием к которым являлось с обязательным наличием музыкального образования по направлению «вокал». Экспертам были выданы пронумерованные

аудиозаписи и таблицы для записи определенных ими нот. Результаты представлены в таблице 3.5.

Таблица 3.5 – Результаты оценки записей с нотами экспертами

Номер записи	Соответствующая нота		
	Эксперт 1	Эксперт 2	Эксперт 3
001	До-диез первой октавы	До-диез первой октавы	До-диез первой октавы
002	Ре-диез малой октавы	Ре-диез малой октавы	Ре-диез малой октавы
003	Си первой октавы	Си первой октавы	Си первой октавы
004	Ля-диез первой октавы	Ля-диез первой октавы	Ля первой октавы
005	Ре первой октавы	Ре первой октавы	Ре первой октавы
006	Ре-диез второй октавы	Соль-диез первой октавы	Соль первой октавы
007	Ми первой октавы	Ми первой октавы	Ми первой октавы
008	До-диез второй октавы	До-диез второй октавы	До-диез второй октавы
009	Фа-диез второй октавы	Фа-диез второй октавы	Фа-диез второй октавы
010	Си малой октавы	Си малой октавы	Си малой октавы
011	Ля-диез малой октавы	Ля-диез малой октавы	Ля-диез малой октавы
012	Ми малой октавы	Ми малой октавы	Ми малой октавы
013	Соль-диез первой октавы	Соль-диез малой октавы	Соль-диез первой октавы
014	Фа-диез малой октавы	Фа-диез малой октавы	Фа-диез малой октавы
015	Ми второй октавы	Ми второй октавы	Ми второй октавы
016	Ля первой октавы	Ля первой октавы	Ля первой октавы
017	Фа второй октавы	До второй октавы	До-диез второй октавы
018	Ля-диез первой октавы	Ре-диез первой октавы	Ре-диез первой октавы
019	Фа второй октавы	Фа второй октавы	Фа-диез второй октавы
020	Фа большой октавы	Фа малой октавы	Фа малой октавы
021	Соль первой октавы	Соль первой октавы	Соль первой октавы
022	До-диез второй октавы	Фа-диез первой октавы	До второй октавы
023	До первой октавы	До первой октавы	До первой октавы
024	Ля малой октавы	Ля малой октавы	Ля малой октавы
025	Ре-диез второй октавы	Ре-диез второй октавы	Ре-диез второй октавы
026	Си большой октавы	Соль малой октавы	Соль малой октавы
027	Фа первой октавы	Фа первой октавы	Фа первой октавы
028	До-диез большой октавы	Ре малой октавы	Ре малой октавы
029	Ре второй октавы	Ре второй октавы	Ре-диез второй октавы

Анализ полученных результатов показал, что в 62% все эксперты и в 28% случаев 2 эксперта из 3 давали одинаковую оценку для нот. Далее на основе данных анкетного опроса была составлена сводная матрица рангов (таблица 3.6).

Таблица 3.6 – Матрица рангов для экспертов

№	Эксперт 1	Эксперт 2	Эксперт 3	Сумма рангов
001	11	12	11	34
002	4	2	2	8
003	20	22	21	63
004	18,5	21	19,5	59
005	12	13	12	37
006	24,5	19	16,5	60
007	13	15	14	42
008	21,5	24	23,5	69
009	29	29	28,5	86,5
010	9	10	9	28
011	8	9	8	25
012	5	3	3	11
013	16	7	18	41
014	6	5	5	16
015	26	27	27	80
016	17	20	19,5	56,5
017	27,5	23	23,5	74
018	18,5	14	13	45,5
019	27,5	28	28,5	84
020	2	4	4	10
021	15	18	16,5	49,5
022	21,5	17	22	60,5
023	10	11	10	31
024	7	8	7	22
025	24,5	26	25,5	76
026	3	6	6	15
027	14	16	15	45
028	1	1	1	3
029	23	25	25,5	73.5
∑	435	435	435	1305

К полученным результатам были применены формулы для расчета коэффициента конкордации Кендалла:

$$W = \frac{12S}{m^2(n^3-n)} \tag{3.12}$$

$$S = \sum_{i=1}^{n}(\sum_{j=1}^{m} R_{ij})^2 - \frac{(\sum_{i=1}^{n}\sum_{j=1}^{m} R_{ij})^2}{n} \tag{3.13}$$

где m - число экспертов в группе;

n - число факторов;

S - сумма квадратов разностей рангов (отклонений от среднего).

По формуле (3.13) получаем:

$$S = 76400{,}5 - \frac{1305^2}{29} = 17675{,}5$$

Далее вычисляется коэффициент конкордации Кендалла по формуле (3.12):

$$W = \frac{12 \cdot 17675{,}5}{9 \cdot (24389 - 29)} = 0{,}967$$

Полученное значение W = 0,967 говорит о наличии высокой степени согласованности мнений экспертов. Далее необходимо провести оценку значимости коэффициента конкордации по формуле 3.12.

$$X^2 = W \cdot m \cdot (n - 1) \qquad (3.12)$$

Табличное значение X^2 для 28 степеней свободы составляет 48,3. Получим по формуле 3.12:

$$X^2 = 0{,}967 \cdot 3 \cdot (29 - 1) = 81{,}228.$$

Полученное значение X^2 больше, чем табличное значение, что свидетельствует о значимости полученном коэффициенте конкордации. Согласованность экспертов удовлетворительная.

3.4 Выводы по главе

На основе модифицированной модели слуховой системы человека был разработан алгоритм распознавания нот в вокальном исполнении, включающий в себя вычисление ЧОТ, на основании которых происходит сегментация и идентификация нот. Алгоритм сегментации и идентификации нот состоит из этапов идентификации нот в каждый момент времени и последующей сегментацией на основании значения минимальной длительности звучания ноты. Для нот был определен подход к вычислению границ звучания с обоснованием корректности выбранных границ. В качестве минимальной меры различия в алгоритме был использован учет минимальной длительности звучания ноты. Данный подход был также исследован при анализе шепотной речи, где в качестве меры различия была использована интенсивность. Были сформированы стратегии

записи вокального исполнения подобные музыкальным упражнениям объемом от 5 до 9 нот.

Работа алгоритмов была протестирована на собранных аудиозаписях. В результате проведенного эксперимента было распознано 113 нот из 114 прозвучавших, что составляет 99%. Результаты были сравнены с данными, полученными в приложениях, показавшим наилучший результат на этапе обзора аналогов (раздел 1.3) и с результатами ручной обработки записей. Оценка коэффициента конкордации показала удовлетворительную согласованность экспертов.

На основании разработанных алгоритмов может быть спроектирован программный комплекс. В рамках дальнейшего исследования необходимо собрать аудиозаписи, содержащие различные варианты исполнения нот, и оценить точность работы алгоритмов на собранных данных.

4 Разработка программного комплекса исследования вокализованной речи

4.1 Структура программного комплекса

В основу программного комплекса исследования вокализованной речи [144-145] были положены алгоритмы слуховой системы человека, которые описаны в [146]. Разработка программного комплекса по исследованию речевых сигналов выполнялась с учетом результатов, представленных в [147]. Было решено выделить в отдельную библиотеку алгоритмы выделения частоты основного тона с учетом особенностей слуховой системы человека, воспринимающей звук. Применение сформированной библиотеки позволило достичь более точного определения значения частоты основного тона по сравнению с аналогами, в частности с пиковыми методами [148]. Структура разработанного программного комплекса на уровне блоков представлена на рисунке 4.1.

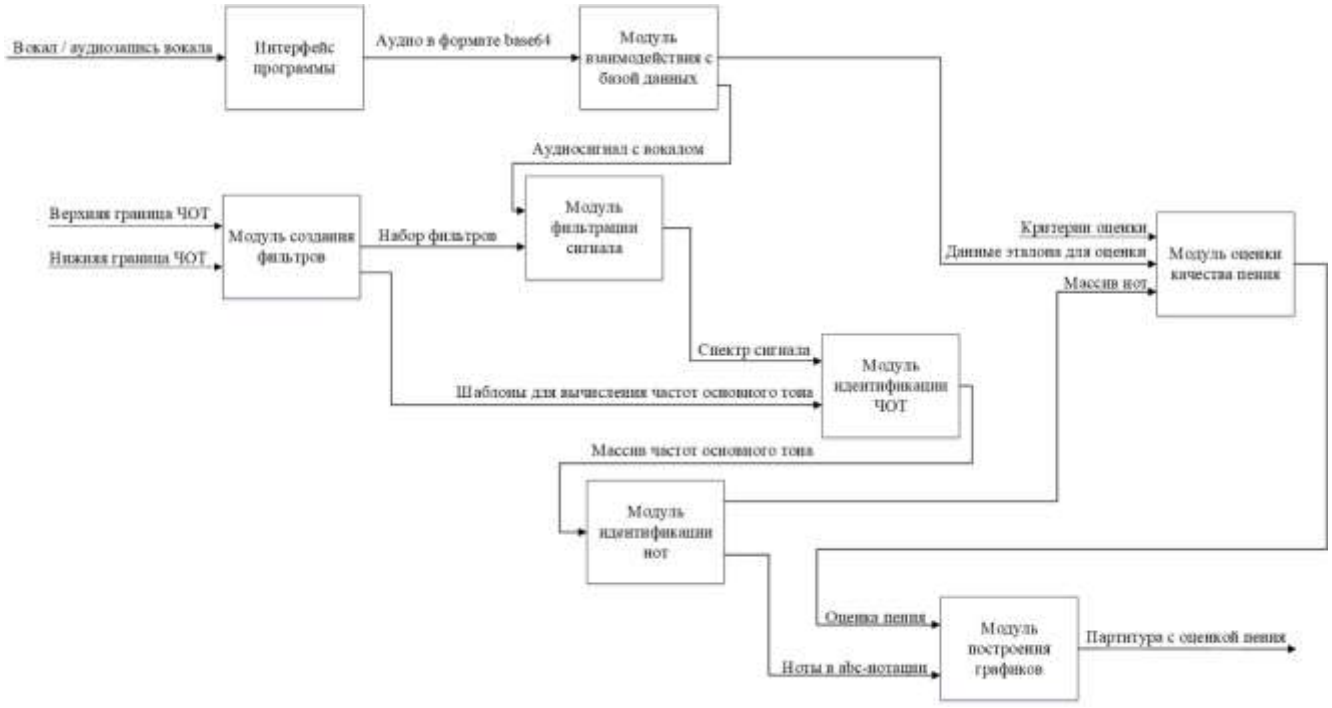


Рисунок 4.1 – Структура программного комплекса

Программный комплекс включает в себя базу данных, содержащую информацию об эталонных значениях границ сегментов, речевые сигналы и их описания. Данный принцип был выбран для формирования упражнений, которые необходимо выполнить студенту. В базу заносят сигнал, который будет использоваться в качестве эталона при одном из упражнений. Данный сигнал обрабатывается вручную экспертом с целью определения точных границ звучания

нот. Также сегментацию сигнала можно осуществить автоматически, определив параллельно корректность найденных границ по алгоритму, представленному в предыдущей главе. Модуль, ответственный за оценку качества сегментации, также осуществляет взаимодействие с остальной частью программного комплекса по исследованию речевых сигналов. Если осуществляется выполнение одного из упражнений, внесенного в базу данных, система сопоставляет полученные при исполнении нот границы звучания с эталонными, сообщая пользователю о степени корректности исполнения задания.

Модуль ручной сегментации был реализован до того, как программный комплекс был дополнен алгоритмом автоматической сегментации. Впоследствии данный модуль использовался для проверки корректности работы алгоритма путем сравнения результатов ручной обработки экспертом с автоматически полученными результатами.

На этапе (рисунок 4.2) создания фильтров программа производит генерацию исходных инструментов для заявленной области обработки. Для этого в границах от нижней частоты $F_{0н}$ до верхней $F_{0в}$ по формулам создается набор фильтров, необходимый для преобразования речевого сигнала в его спектрограмму. Помимо этого, создается набор шаблонов, который сопоставляется с полученным после фильтрации сигнала спектром прослушанного сигнала. На основе поступивших данных алгоритмом определения ЧОТ программой выдается массив мгновенных значений частот F_0. По разработанному в ходе исследования алгоритму получения нот на основе данного массива программой выдается заключение о том, какая последовательность нот была зафиксирована в речевом сигнале [149].

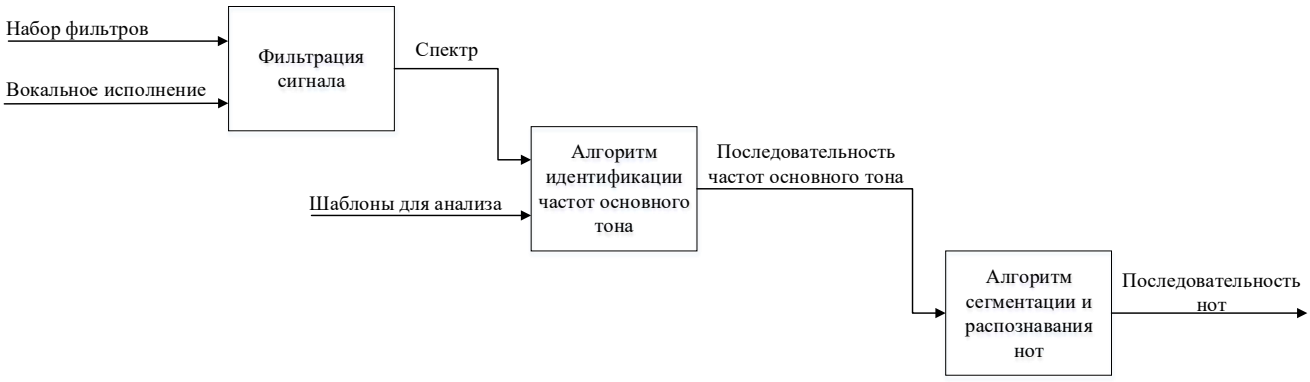


Рисунок 4.2 – Этапы распознавания нот в вокальном исполнении

Одной из особенностей программного комплекса, реализованной в рамках [150], является использование клиент-серверной модели. По мере того, как пользователь записывает аудиофайл в клиентской части приложения, происходит процесс преобразования аудиофайла, представленный на рисунке 4.3.

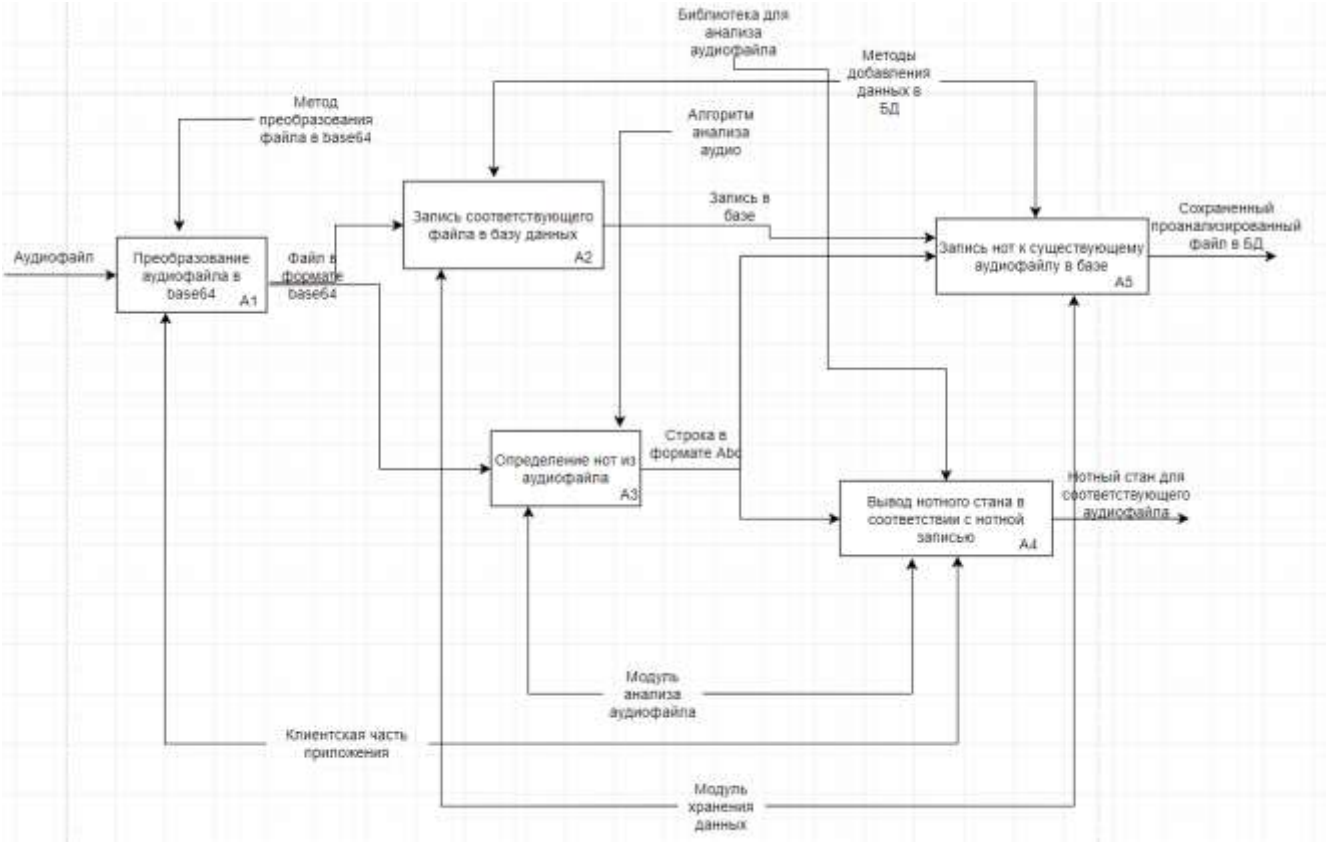


Рисунок 4.3 – Обработка аудиофайла

Записанные файлы отправляются на сервер для последующей обработки алгоритмами сегментации и идентификации нот. После чего на клиентскую машину отправляется сообщение в abc-нотации с результатами. Преобразование аудиофайла в формат base64 для последующего хранения в базе данных производится при помощи встроенного метода JavaScript – btoa(). Для того, чтобы декодировать base64 строку на сервере, необходимо аналогично использовать стандартный метод atob() применительно к необходимой строке. Для определения нот в аудиофайле применяется интегрированная в проект библиотека, которая реализует фильтрацию, определяет вокализованные и невокализованные участки, определяет частоту основного тона и сопоставляет полученные частоты нотам, после чего выдает на выходе строку в формате Abc, если ноты были найдены. Для

отображения нот используется библиотека Abc.js, которая принимает на вход строку, содержащую названия нот.

Добавление в базу данных состоит из следующих этапов:

1) Сервер принимает файл в случае его соответствия требованиям с клиентской части посредством HTTP-запроса;

2) Выполняется скрипт, записывающий полученную информацию в базу данных.

При разработке проекта к front-end части был применён компонентный подход. Благодаря данному подходу в решении появляется читаемость, возможность быстрого и верного рефакторинга с отсутствием возможности утери работоспособности приложения.

Взаимодействие с сервером со стороны front-end было реализовано с помощью технологии HTTP Requests. Данный вид работы с сервером обеспечивается как встроенными методами javascript, так и при помощи различных библиотек, предназначенных для данных целей, одной из которых является библиотека axios. Для поддержки масштабируемости и расширяемости, которую обеспечивает компонентный подход, работа с данными на front-end производится в специальном хранилище – Vuex storage. Обрабатываются как и локальные данные, так и получаемые с сервера, а также производится обработка запросов на сервер.

Как было отмечено, в [151], применение программных средств для обучения вокальному мастерству не должно расцениваться как замена преподавателя. Это объясняется тем, что при непосредственном общении с педагогом студент приучается аналитически вникать в музыкальный материал [152]. В задачи преподавателя входит подбор учебного материала таким образом, чтобы каждая новая задача ставила перед учеником новые, но всегда преодолимые трудности, соответствующие его общему музыкальному и вокально-техническому развитию [153].

В качестве экспериментального набора упражнений было сформировано несколько заданий. Согласно [154], все упражнения, четкое и систематическое

выполнение которых формируют фундамент певческих навыков, можно разделить на 2 группы: интонируемые упражнения и упражнения по слуховому анализу. Упражнения по слуховому анализу осуществляются в формате, не подразумевающем пение ученика. В качестве интонируемых упражнений были выбраны:

– пение гамм в октавном диапазоне вверх и вниз;

– построение голосом аккордов, а именно одноголосное пение мелодически связанных аккордовых последовательностей;

– пение по заданному (на фортепьяно или камертоном) звуку мелодических мотивов, настраивающих слух в соответствующей ладо-тональности.

В свою очередь, в [2] говорится о том, что на начальном этапе обучения упражнения не должны ставить перед учеником сложных задач. Данные упражнения должны быть направлены на формирование качественного певческого звука за счет применения основных видов вокальных движений: стаккато, легато и т.п.

Для первых двух типов выбранных упражнений были использованы локации, описанные в разделе 3.2: в качестве примеров пения гамм в октавном диапазоне вверх и вниз могут служить локации 1-6, а локации 7-8 – являются примером мелодически связанных аккордовых последовательностей.

Для пения по заданному звуку было решено применить аудиозапись с пением человека, получившим музыкальное образование по вокалу. Запись данного человека выступала в роли эталонной. В рамках эксперимента, описанного в [155], 7 человек без музыкального образования прослушивали аудиозапись и старались спеть ноты на том же уровне, что и преподаватель (эталонная запись). На рисунке 4.4 представлен пример распределения ЧОТ для спетой стаккато локации, содержащей следующую последовательность нот первой октавы: до, ре, ми, фа, соль, фа, ми, ре, до.

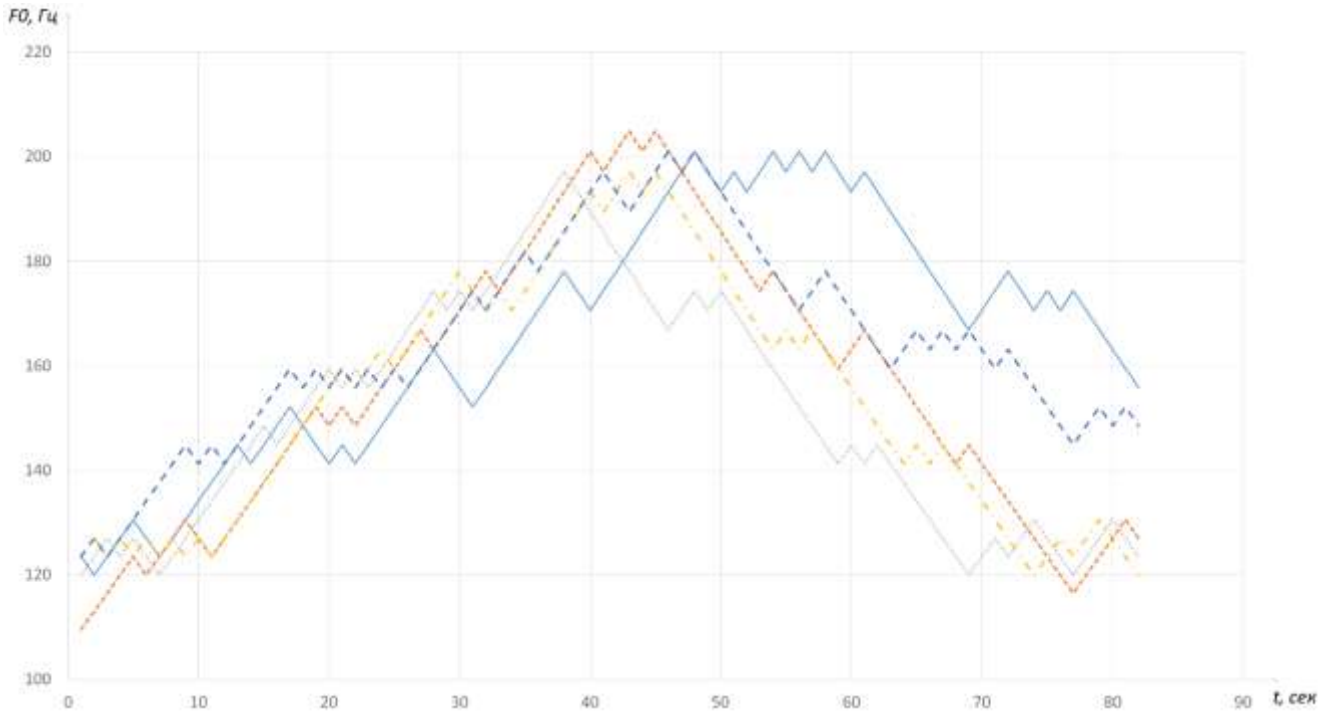


Рисунок 4.4 – Распределение ЧОТ обработанных аудиозаписей для локации

Для определения сходства вокальных исполнений был применен метод выделения синхронности, описанный в [156]. Степень сходства выше 90% была получена для тех исполнений, где пение учеников было достаточно близким к пению в эталонной аудиозаписи. Экспертная оценка аудиозаписей, получивших степень родства в 50%, показала, что данные локации были исполнены учениками хуже. Следовательно, алгоритм оценки сходства вокальных исполнений справился со своей задачей и может быть применен в рамках упражнения по повторению за преподавателем мелодического мотива.

4.2 Описание собранной базы

В качестве материалов для тестирования помимо синтезированных тестовых синусоидальных сигналов были собраны:

– 60 аудиозаписей для каждой отдельной ноты от большой до третьей октавы, сыгранных на фортепиано;

– 9 аудиозаписей для нот в диапазоне от 392 Гц до 739.98 Гц (первая-вторая октава), сыгранных на гитаре;

– 8 аудиозаписей, содержащих 58 нот, спетых стаккато и легато в диапазоне первая-вторая октава женским голосом;

– 8 аудиозаписей, содержащих 58 нот, спетых стаккато и легато в диапазоне малая-первая октава женским голосом;

– 8 аудиозаписей, содержащих 56 нот, спетых стаккато и легато в диапазоне малой октаве мужским голосом;

– 24 аудиозаписи, спетые альтом и баритоном и содержащие 196 нот, исполненных с такими приемами, как крещендо, декрещендо, арпеджио, восходящим и нисходящим глиссандо, в диапазоне большая-первая октава;

– 29 аудиозаписей, спетых женским голосом, содержащие 29 нот в диапазоне от ре малой октавы до фа-диез второй октавы.

4.3 Проведение экспериментов по распознаванию нот в вокальном исполнении на заданных частотах основного тона

В первую очередь в программе были протестированы аудиозаписи с сигналами наиболее близкими к использованным ранее. Для определения корректности исполнения чистой ноты были собраны аудиозаписи с нотами, сыгранными на настроенном фортепиано, где частота звучания каждой ноты точно соответствовала частоте, представленной в таблице 1.1. Отсутствие посторонних нот и изменений в частоте колебания должно было исключить посторонние воздействия на работу алгоритмов. В ходе эксперимента были протестированы аудиозаписи от нот до-диез большой октавы (69.3 Гц) до ноты си второй октавы (987.75 Гц). Все ноты до си-бемоль второй октавы (932.32 Гц) включительно были идентифицированы без ошибок.

Как можно видеть на рисунке 4.5, в результате идентификации частот в аудиосигнале было получено много шумов на уровне до 200 Гц. Аналогичные результаты получены для нот выше 932 Гц. Обилие шумов не позволяет алгоритму четко идентифицировать конкретную ноту. Таким образом, было показано, что на частотах выше 800 Гц наблюдается низкая надежность определения частоты основного тона сигнала. Однако успешность сегментации и идентификации нот, сыгранных на фортепиано в диапазоне до 932 Гц включительно, позволяет судить о корректности работы алгоритмов программного комплекса для одиночных нот.

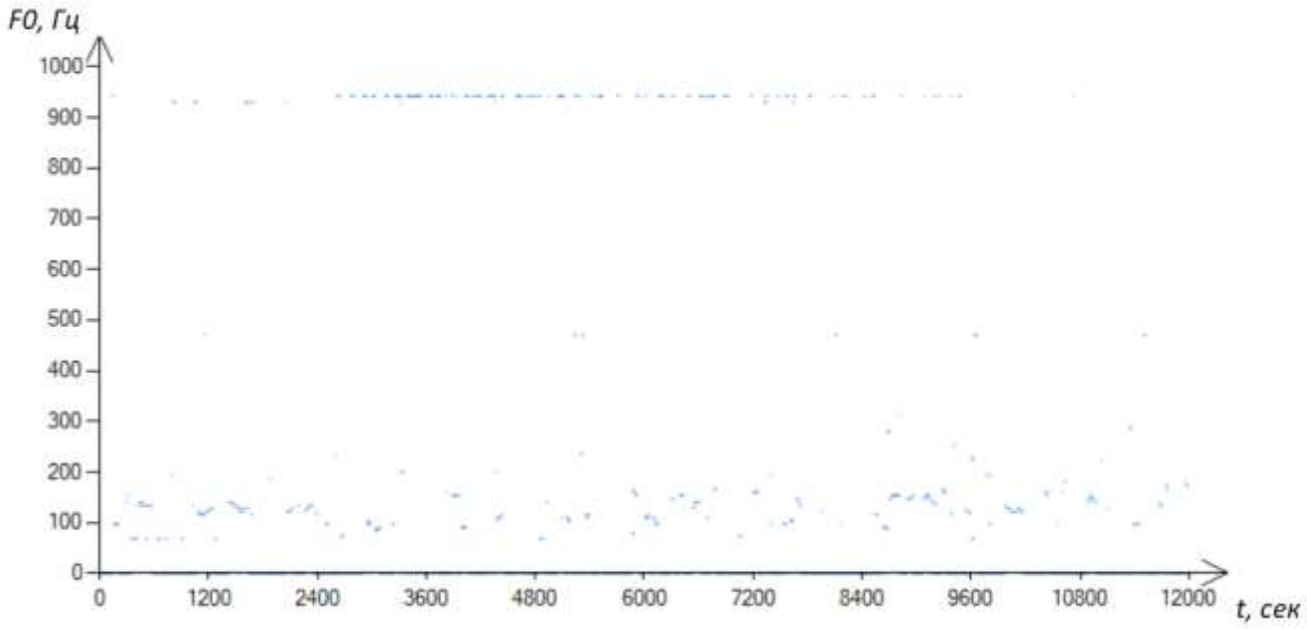


Рисунок 4.5 – Результат идентификации частот основного тона для ноты си-бемоль второй октавы (932.32 Гц)

Кроме того, было решено протестировать работу алгоритмов также и на аудиозаписях с нотами, сыгранными на гитаре. В качестве эксперимента было решено тестировать ноты, находящиеся в диапазоне от 400 до 800 Гц. В каждой аудиозаписи содержалась последовательность из 4 нот, сыгранная с паузами между нотами. Эксперимент был проведен для анализа ситуаций, когда колебания накладываются друг на друга: колебания предыдущей ноты еще не затихли, а новой уже прозвучали. С каждой следующей нотой сложность определения текущей прозвучавшей ноты возрастает. Данная ситуация будет справедлива также и для записей, в которых последовательно сыграны несколько нот на фортепиано, поет несколько людей одновременно или пение происходит под аккомпанемент музыкального инструмента.

Кроме аудиозаписей с пением стаккато и легато, тестирование которых было приведено в главе 3.2, были рассмотрены такие приемы, как крещендо и декрещендо, арпеджио и глиссандо (восходящее и нисходящее). Было решено проверить наличие влияния различных способов исполнения нот на качество их идентификации.

Как видно из рисунка 4.6, при пении арпеджио отсутствует влияние на качество работы алгоритмов в программном комплексе. Под термином понимается

последовательное исполнение нот без наложения друг на друга. Обычно данный термин применяется к игре на музыкальных инструментах. Спецификой пения арпеджио является сохранение однородности тембра голоса на всем протяжении диапазона пения. Отсутствие влияния на качество идентификации нот заключается в том, что в алгоритме при анализе учитывается только частота основного тона.

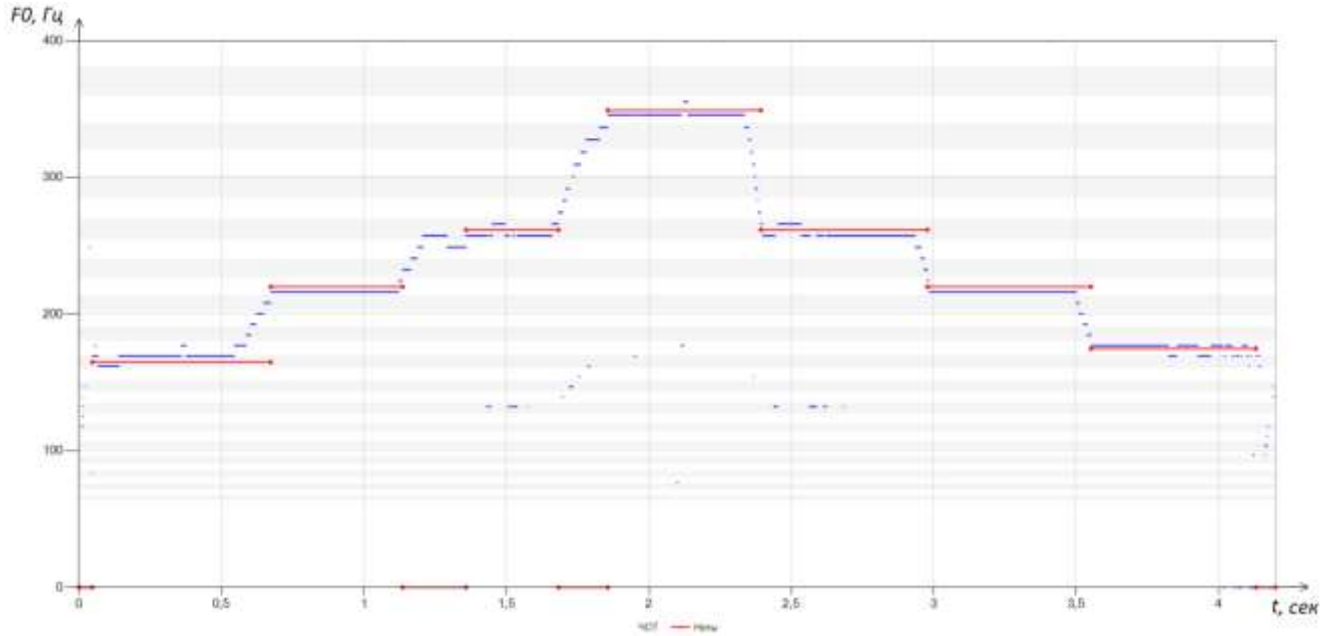


Рисунок 4.6 – Результат обработки аудиозаписи с арпеджио

Крещендо и декрещендо – это исполнение ноты с постепенным увеличением или уменьшением силы звука. Как можно видеть по результатам анализа аудиозаписей (рисунки 4.7-4.8), изменения в силе звука также не оказывают влияние на качество распознавания спетых нот.

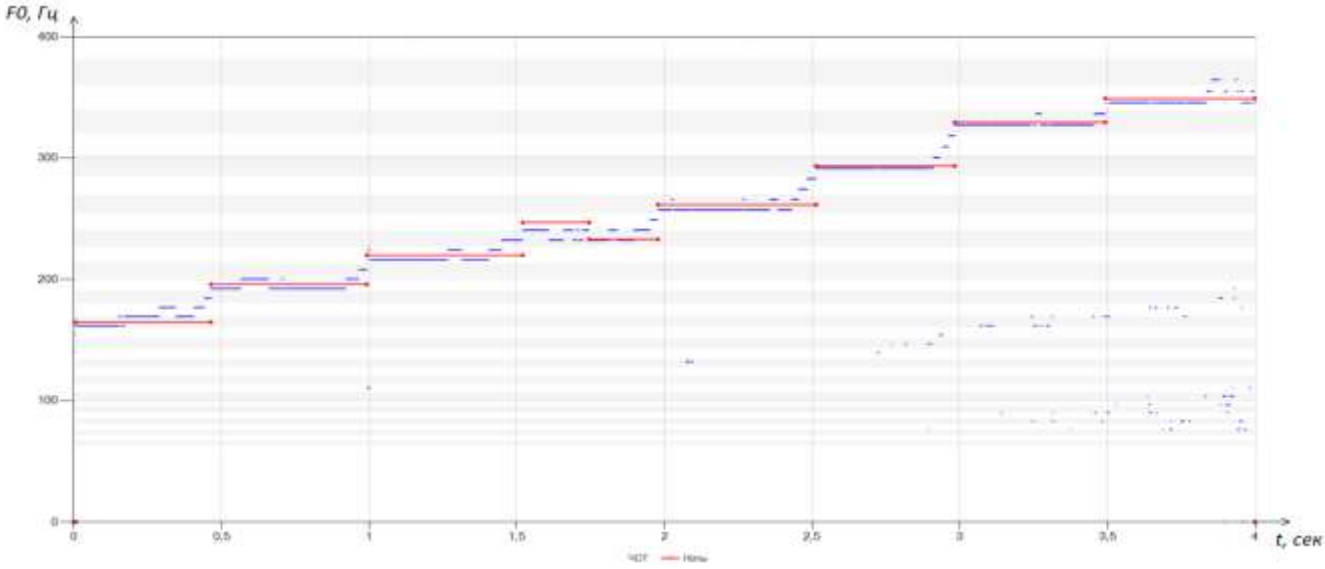


Рисунок 4.7 – Результат обработки аудиозаписи с крещендо

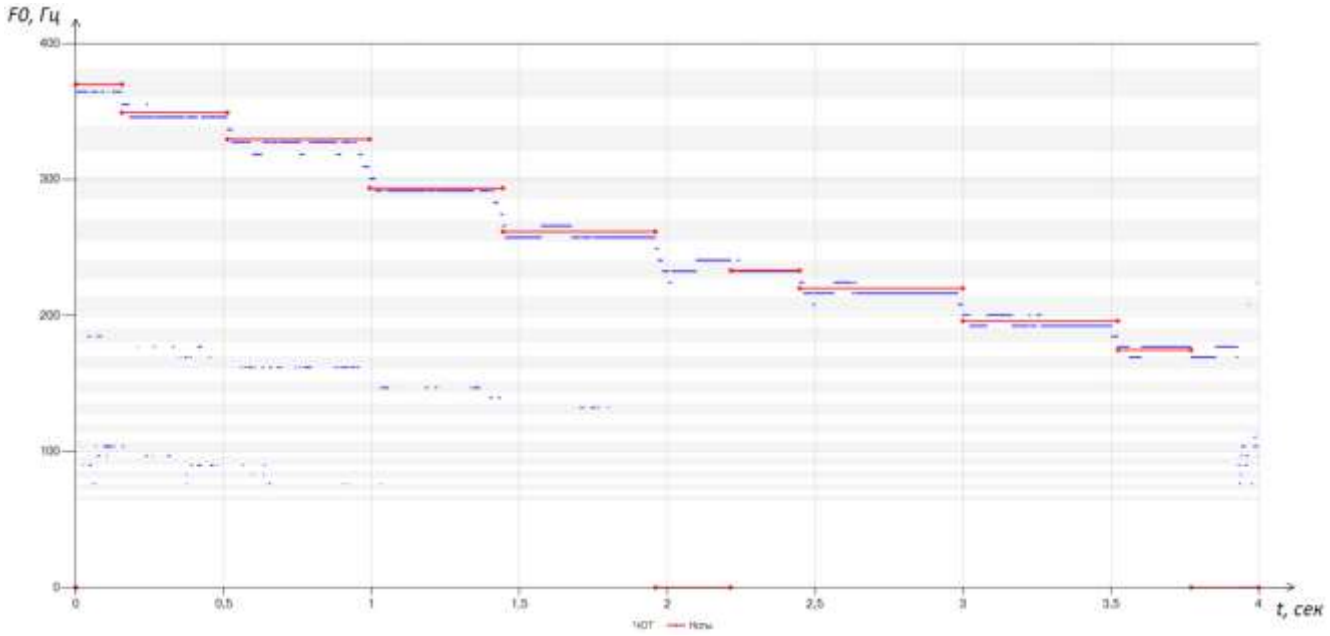


Рисунок 4.8 – Результат обработки аудиозаписи с декрещендо

При исследовании аудиозаписей с вокальным исполнением нот, у которых частота основного тона находится в диапазоне от 400 до 600 Гц, был обнаружен эффект вибрато, оказывающий воздействие на точность распознавания нот в указанном интервале частот. Спецификой работы одного из этапов алгоритма идентификации нот, отвечающего за определение принадлежности вокализованного участка к ноте, заключается в ориентированности на чистое исполнение.

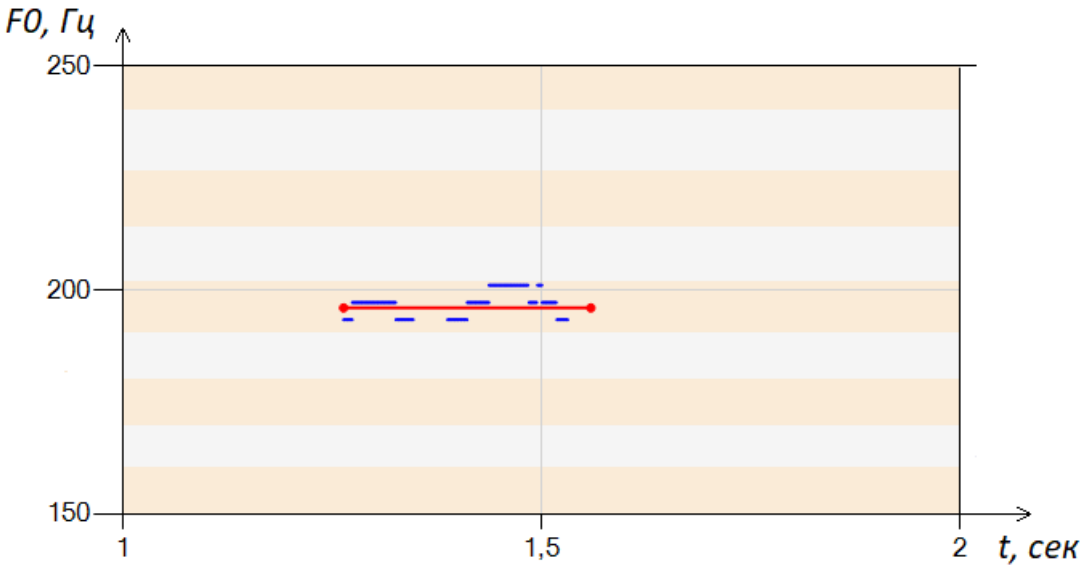


Рисунок 4.9 – Участок графика с нотой, спетой диктором без вибрато в голосе

Исполнить ноту так, чтобы все частоты в момент пения находились в пределах полосы, отведенной под нее, достаточно сложно. По этой причине учитываются соседние ноты выше и ниже исполняемой. На рисунках 4.9 и 4.10 спектр частот разделен на участки, соответствующие границам нот. Каждый участок выделен горизонтальной полосой на фоне графика частот. По оси X – время, по оси Y – частоты.

Как можно видеть на рисунке 4.9, сегмент, спетый в пределах одной ноты, был распознан программой. Количество моментов, попавших в соседние ноты, незначительно и не оказало влияние на результат идентификации. Преобладающее число распознанных частот основного тона оказалось в диапазоне спетой ноты. С учетом прописанных в алгоритме требований по точности исполнения, программа смогла сделать вывод о спетой ноте. В случае, если количество фрагментов в каждом из 3 участков оказалось приблизительно одинаковым, алгоритм воспринимает весь вокализованный сегмент как шум.

Эту ситуацию можно увидеть на рисунке 4.10. Спетая диктором нота была исполнена с вибрато в голосе. Как можно заметить на исследуемом участке присутствуют колебания в 4 соседних нотах. Кроме того, оценка вклада в каждую из нот у спетого сегмента меньше требуемого показателя. Считается, что к наличию вибрато в голосе склонны высокие голоса, и наиболее распространен прием среди обладательниц сопрано.

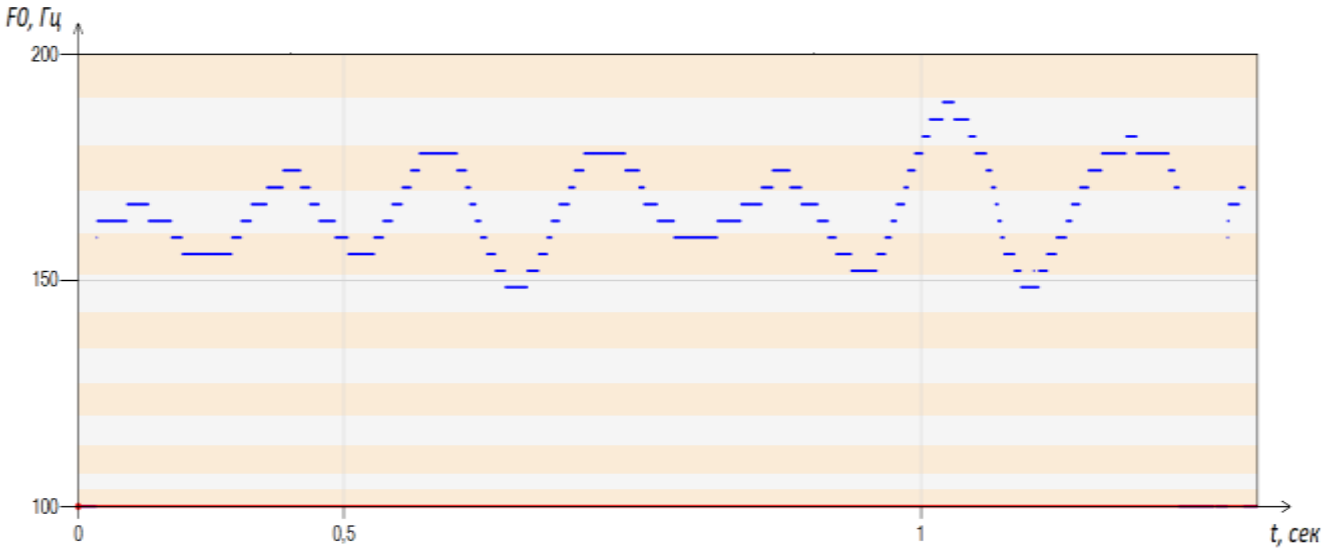


Рисунок 4.10 – Участок графика с нотой, спетой диктором с вибрато в голосе

Другим примером внесения колористического эффекта в пение может служить такой прием как глиссандо. Как можно увидеть на рисунке 4.11, при таком приеме в пении происходит плавное скольжение от одного звука к другому.

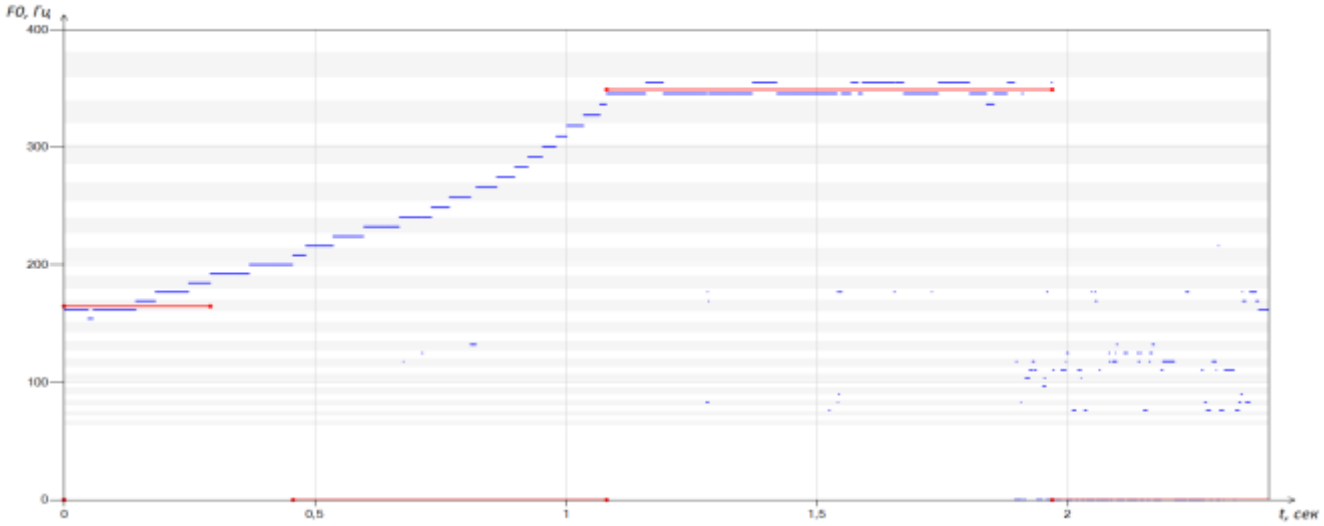


Рисунок 4.11 – Восходящее глиссандо с ноты «соль малой октавы» до ноты «фа 1 октавы»

При таком пении переход происходит слишком быстро для того, чтобы суметь идентифицировать каждую отдельную ноту в момент скольжения, поскольку на каждый сегмент из охваченных нот приходится отводится менее 0,1 секунды.

В общей сложности из 196 нот, спетых с различными подходами к вокальному исполнению, было распознано 187 нот, определенных экспертами, что составляет 95.4%.

4.4 Внедрение в дистанционное обучение вокалу

В рамках подготовки разработанных алгоритмов к внедрению в дистанционное обучение вокалу с помощью приложения, разрабатываемого «Элекард-ЦТП», была изменена модель взаимодействия преподавателя с обучающимися. В первоначальной версии результаты выполнения упражнений студентом записывались в формате wav и передавались по сети преподавателю. Применение алгоритма распознавания нот в вокальном исполнении позволило снизить объем трафика, передаваемого по сети, более чем на 90% за счет перехода

к отправлению преподавателю текстового сообщения с abc-нотацией распознанных нот.

Собранная база аудиозаписей с вокальным исполнением была дополнена записями, полученными от учеников музыкальной школы. Был сформирован набор эталонных записей с пением последовательностей нот с различным стилем исполнения. В данный набор вошли следующие упражнения:

1. исполнение гаммы со следующими нотами: «до», «ре», «ми», «фа», «соль», «фа», «ми», «ре», «до» 1 октавы;

2. исполнение гаммы со следующими нотами: «соль», «ми», «до», «до», «ми», «соль» 1 октавы;

3. исполнение гаммы со следующими нотами: «фа», «соль», «ля», «си», «ля-диез», «ля», «соль», «ля», «си» малой октавы; «до» 1 октавы; «ля-диез», «ля», «соль», «фа» малой октавы;

4. исполнение гаммы стаккато со следующими нотами: «ля-диез» малой октавы; «до», «фа», «соль», «фа-диез», «ре-диез», «до» 1 октавы; «си» малой октавы; «до» 1 октавы;

5. пение последовательности нот: «соль», «ми», «фа», «ре», «ми», «до», «ре» 1 октавы; «си» малой октавы; «соль», «ми», «фа», «ре», «ми», «до», «ре» 1 октавы; «си» малой октавы; «до» 1 октавы; «си» малой октавы; «до» малой октавы;

6. восходящее глиссандо от ноты «си» малой октавы до ноты «фа-диез» 1 октавы, «ми», «до» 1 октавы; восходящее глиссандо от ноты «си» малой октавы до ноты «соль» 1 октавы, «ми», «до» 1 октавы; восходящее глиссандо от ноты «до» до ноты «соль», «ми», «до» 1 октавы;

7. пение в быстром темпе переходов между нотами: «соль-диез» малой октавы; «до» 1 октавы; «соль-диез» малой октавы; «ре-диез» 1 октавы; «соль-диез» малой октавы; «соль-диез», «ре-диез», «до» 1 октавы; «соль-диез» малой октавы.

Кроме того, в качестве эталонных упражнений выступили мотивы из следующих песен:

1. отрывок мелодии из песни «Конь», содержащая 45 нот, спетых без произношения слов;

2. отрывок мелодии из песни «В лунном сиянии», содержащий последовательность из 21 ноты, спетую без произношения слов;

3. отрывок мелодии из песни «Луч солнца золотого», содержащий последовательность из 14 нот, спетую без произношения слов;

4. отрывок песни с произношением слов: «Горна река так быстротечна и чиста, войду в воду твою с головой».

Полученный набор был дополнен упражнениями из сборника одноголосных музыкальных диктантов [157] упражнениями из разделов: ритмические группы, мажорный ряд, движение мелодии по звукам тонического трезвучия, секвенции.

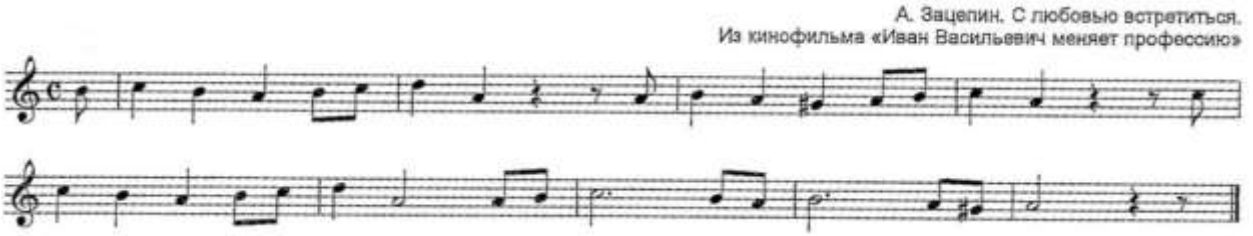


Рисунок 4.12 – Ноты для музыкального диктанта по песне «С любовью встретиться»

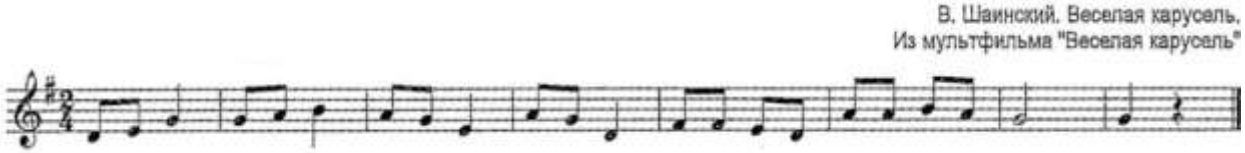


Рисунок 4.13 – Ноты для музыкального диктанта по песне «Веселая карусель»

Рисунок 4.14 – Ноты для музыкального диктанта по песне «Первоклашка»

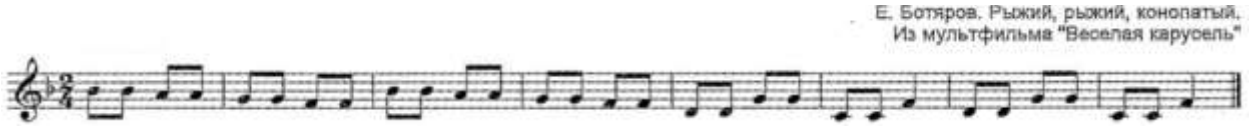


Рисунок 4.15 – Ноты для музыкального диктанта по песне «Рыжий, рыжий, конопатый»

Рисунок 4.16 – Ноты для музыкального диктанта по песне «Ой, цветет калина»

Рисунок 4.17 – Ноты для музыкального диктанта по песне «Песенка Деда Мороза»

Каждая эталонная запись была проанализирована экспертом на предмет звучания нот. Результаты экспертной оценки были сопоставлены с результатами, полученными с помощью программного комплекса. Оценка показала корректность работы алгоритмов сегментации и идентификации нот на представленном наборе записей.

В рамках эксперимента по составлению упражнений с пением по заданному звуку мелодических мотивов, настраивающих слух в соответствующей ладо-тональности, был проведен сбор записей с учеников музыкальной школы. Каждому ученику давалось задание прослушать эталонную запись и сделать 5 записей с пением прослушанного задания. В общей сложности, было собрано 17 эталонных записей и 740 записей учеников, содержащих в совокупности 13078 спетых нот.

С целью определения частоты корректной работы программы была проведена экспертная оценка набор записей для 1-го упражнения. Всего было оценено 53 записи, содержащие 477 нот. В таблице приведена статистика по результатам распознавания нот разработанным программным комплексом и программой Melodyne.

В общей сложности, разработанный программный комплекс распознал 469 нот из 477 прозвучавших, а программа Melodyne смогла выделить 392 ноты.

Таблица 4.1 – Результаты исследования записей с помощью программ

Показатель	Разработанный программный комплекс	Melodyne
Распознаны все ноты	47 записей	26 записей
Не распознана 1 нота	4 записей	9 записей
Не распознана 2 ноты	2 записей	6 записей
Не распознана 3 и более нот	0 записей	12 записей

В качестве примера далее приводятся несколько результатов обработки записей разных дикторов.

В записи №7 экспертом были выделены следующие ноты: «до», «ре», «ми», «фа», «соль», «фа», «ре-диез», «до-диез» 1 октавы и «си» малой октавы. Программа Melodyne (рисунок 4.18) смогла распознать следующие ноты: «си» малой октавы; «до», «ре-диез», «фа-диез», «фа», «ре-диез», «до-диез» 1 октавы и «си» малой октавы, что содержит 5 из 9 выделенных экспертом значений. Разработанный программный комплекс (рисунок 4.19) распознал следующие ноты: «си» малой октавы; «до», «ре», «ми», «соль», «фа», «ре-диез», «до-диез» 1 октавы; «си» малой октавы, что на 8/9 соответствует экспертной оценки. Нота «фа» на участке от 1.75 до 1.9 секунды не была распознана программой из-за перехода на соседнюю ноту.

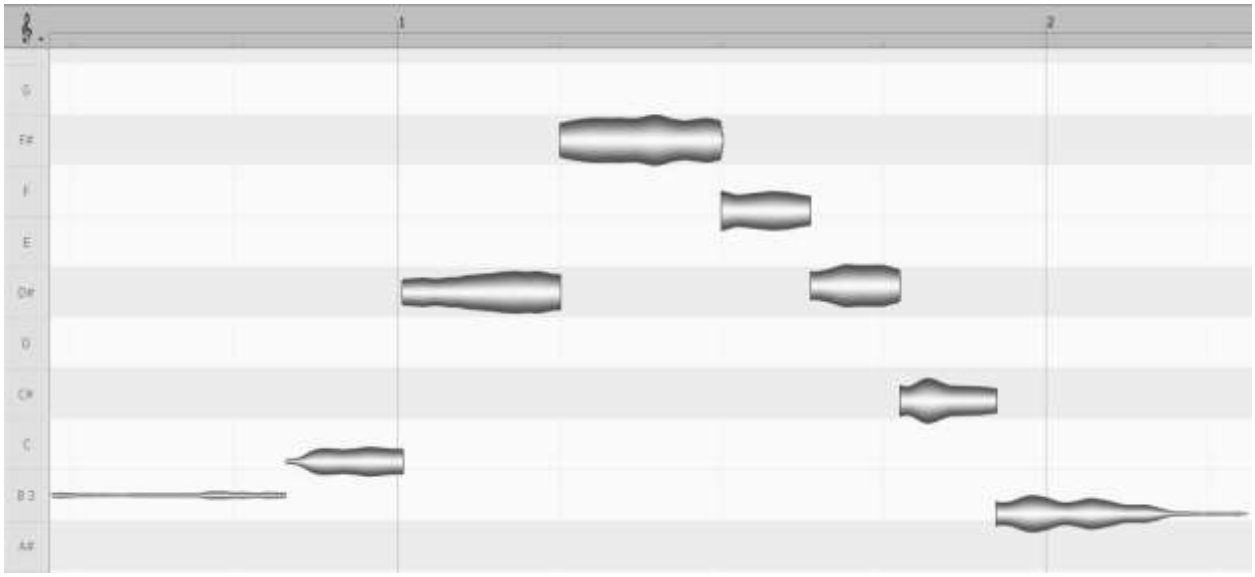


Рисунок 4.18 – Результат обработки записи №7 программой Melodyne

Рисунок 4.19 – Результат обработки записи №7
разработанным программным комплексом

В записи №28 экспертом были выделены следующие ноты: «си» малой октавы; «до-диез», «ре-диез», «фа», «фа-диез», «ми», «ре» первой октавы; «си», «ля-диез» малой октавы. Программа Melodyne (рисунок 4.20) смогла распознать следующие ноты: «си», «ре-диез», «фа», «фа-диез», «ре» и «ля-диез», что содержит 6 из 9 выделенных экспертом значений. Разработанный программный комплекс (рисунок 4.21) распознал следующие ноты: «си» малой октавы; «до-диез», «ре-диез», «фа», «фа-диез», «ми», «ре» первой октавы; «си», «ля-диез» малой октавы, что полностью соответствует экспертной оценке.

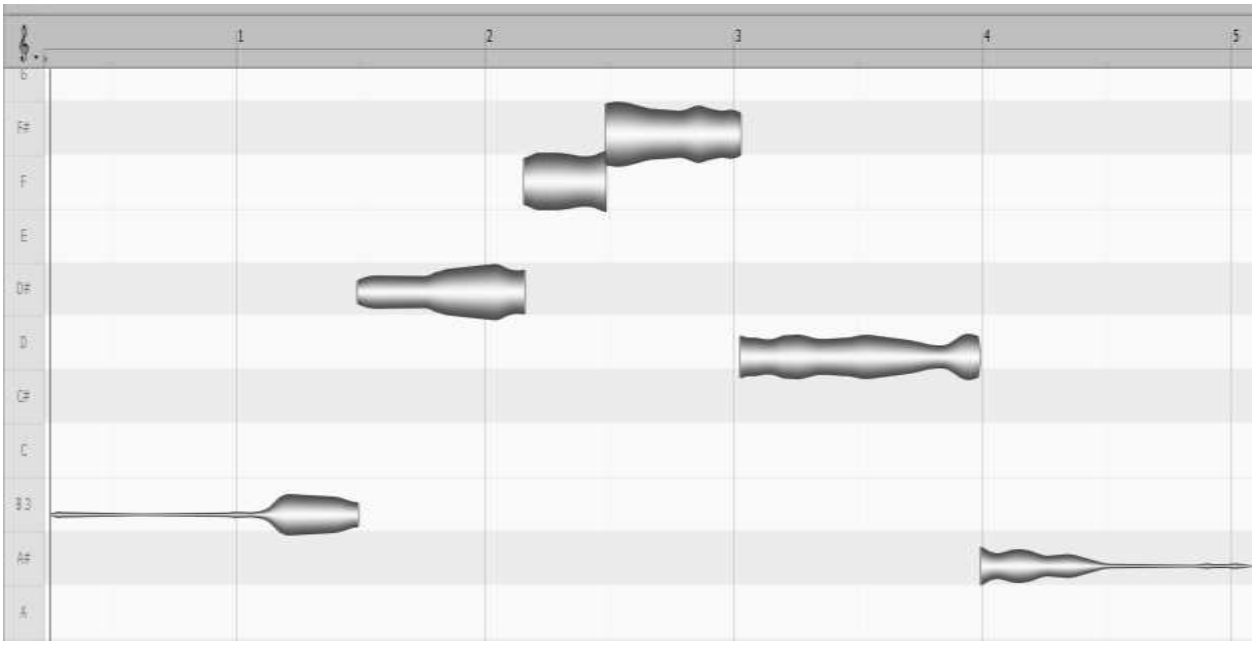


Рисунок 4.20 – Результат обработки записи №28 программой Melodyne

Рисунок 4.21 – Результат обработки записи №28
разработанным программным комплексом

Нижняя и верхняя границы доверительного интервала для неизвестной частоты наступления события определяется по формулам 4.1 и 4.2.

$$P_{\text{н}} = \frac{n}{t^2+n}\left[\omega + \frac{t^2}{2n} - t\sqrt{\frac{\omega(1-\omega)}{n} + \left(\frac{t}{2n}\right)^2}\right], \quad (4.1)$$

$$P_{\text{в}} = \frac{n}{t^2+n}\left[\omega + \frac{t^2}{2n} + t\sqrt{\frac{\omega(1-\omega)}{n} + \left(\frac{t}{2n}\right)^2}\right] \quad (4.2)$$

где n – общее количество измерений;

ω – относительная частота;

t – значение обратной функции Лапласа (для вероятности 0,95 составляет 1,96).

Таким образом, при $n = 477$ и $\omega = 8/477$ вычисления границ по формулам 4.1 и 4.2 принимают вид:

$$P_{\text{н}} = \frac{477}{1{,}96^2+477}\left[\omega + \frac{1{,}96^2}{2\cdot 477} - 1{,}96\cdot\sqrt{\frac{0{,}02\cdot(1-0{,}02)}{477} + \left(\frac{1{,}96}{2\cdot 477}\right)^2}\right] = 0{,}009$$

$$P_{\text{в}} = \frac{477}{1{,}96^2+477}\left[\omega + \frac{1{,}96^2}{2\cdot 477} + 1{,}96\cdot\sqrt{\frac{0{,}02\cdot(1-0{,}02)}{477} + \left(\frac{1{,}96}{2\cdot 477}\right)^2}\right] = 0{,}033$$

В результате частота ошибок в работе программы с вероятностью 0,95 не превышает 3.3%.

4.5 Выводы по главе

На основе предложенной методики распознавании нот в вокальном исполнении был разработан программный комплекс. Разработанный программный комплекс способен работать с аудиозаписями вокальных исполнений, загруженных из файлов формата wav, а также с записями, сделанными через интерфейс программы. В рамках сбора данных для тестирования работы комплекса были собраны записи с игрой на музыкальных инструментах, с пением в различных диапазонах и с применением разных способов пения. Комплекс был протестирован на аудиозаписях с различными подходами к вокальному исполнению (такими как стаккато, легато, арпеджио, крещендо, декрещендо, восходящее и нисходящее

глиссандо). Результаты эксперимента показали, что для записей музыкальных инструментов, содержащих отдельную сыгранную ноту, алгоритм работает без ошибок до частоты 932.32 Гц, что соответствует ноте «си-бемоль второй октавы».

Было определено, что при пении арпеджио, крещендо и декрещендо отсутствует влияние на качество работы алгоритмов в программном комплексе. Отсутствие влияния на качество идентификации нот заключается в том, что в алгоритме при анализе учитывается только частота основного тона. В общей сложности из 196 нот, спетых с различными подходами к вокальному исполнению, было распознано 187 нот, что составляет 95.4%.

Разработанные алгоритмы были внедрены в деятельность «Элекард-ЦТП» в рамках проекта по дистанционному обучению вокалу в формате видеоконференций, что позволило снизить объем трафика, передаваемого по сети, более чем на 90% за счет перехода от передачи аудиозаписей с пением в формате wav к отправлению преподавателю текстового сообщения с abc-нотацией распознанных нот.

Программный комплекс был оценен на предмет частоты ошибок в работе программы. С вероятностью 0.95 частота возникновения ошибок не превышает 3.3%.

Заключение

В диссертационной работе решена задача повышения качества распознавания звучащих нот в вокальном исполнении за счёт применения модели слуховой системы человека.

Основные результаты диссертационной работы:

1. Произведен обзор существующих методов и алгоритмов распознавания нот, в том числе определения частот основного тона. Был сделан вывод, что в сфере речевых технологий отсутствуют алгоритмы, направленные на точную идентификацию спетой диктором ноты. Кроме того, было определено, что существующие алгоритмы анализа частоты основного тона неприменимы к вокальным исполнениям по 2 причинам: высокий процент грубых ошибок и ограничение полосы исследования диапазоном до 400 Гц.

2. Проведена модификация модели слуховой системы человека на предмет увеличения диапазона анализа частот основного тона. Для этого была добавлена возможность автоматического учета границ определения частот основного тона сигнала. Модель была протестирована на сгенерированных синусоидальных сигналах. Полученные результаты по идентификации частот основного тона показали высокую точность в диапазоне от 70 до 800 Гц включительно. Относительная ошибка алгоритма идентификации ЧОТ составила менее 1%, что позволяет применить модифицированную математическую модель слуховой системы человека не только для анализа параметров речевого сигнала, но и для идентификации нот.

3. Описан алгоритм сегментации и идентификации нот, состоящий из этапа идентификации нот в каждый момент времени с их последующей сегментацией на основании значения минимальной длительности звучания ноты. Для нот был определен подход к вычислению границ звучания с обоснованием корректности выбранных границ. В качестве минимальной меры различия в алгоритме был использован учет минимальной длительности звучания ноты. Работа алгоритмов была протестирована на собранных аудиозаписях. В результате проведенного эксперимента было распознано 113 нот из 114 прозвучавших, что составляет 99%.

Результаты были сравнены с данными, полученными в приложениях, показавшим наилучший результат на этапе обзора аналогов. Оценка коэффициента конкордации показала удовлетворительную согласованность экспертов.

4. Разработан программный комплекс, способный работать с аудиозаписями вокальных исполнений, загруженных из файлов формата wav, а также с записями, сделанными через интерфейс программы. Комплекс был протестирован на аудиозаписях с различными подходами к вокальному исполнению (такими как стаккато, легато, арпеджио, крещендо, декрещендо, восходящее и нисходящее глиссандо, вибрато). Результаты эксперимента показали, что при анализе аудиозаписей:

– вокального исполнения, содержащих исполнения с применением стаккато, легато, арпеджио, крещендо и декрещендо, алгоритм распознал безошибочно не менее 95% нот;

– вокального исполнения, содержащего исполнения с применением таких техник, как глиссандо и вибрато, алгоритм правильно указывает диапазоны, в которых происходит изменение звучания ноты.

5. Разработанные алгоритмы были внедрены в деятельность «Элекард-ЦТП» в рамках дистанционного обучения вокалу в формате видеоконференций, что позволило снизить объем трафика, передаваемого по сети, более чем на 90% за счет перехода от передачи аудиозаписей с пением в формате wav к отправлению преподавателю текстового сообщения с abc-нотацией распознанных нот. Программный комплекс был оценен на предмет частоты ошибок в работе. С вероятностью 0,95 частота возникновения ошибок не превышает 3.3%.

Список использованных источников

1 Минкультуры представило программу развития музыкального образования на ближайшую пятилетку [Электронный ресурс]. – Режим доступа: http://www.mkrf.ru/press/news/minkultury-predstavilo-programmu-razvitiya-muzykalnogo-obrazovaniya-na-blizhaysh20171006172610.

2 Дмитриев Л.Б. Основы вокальной методики. Москва: Министерство культуры СССР. – 1966. – С. 611-613.

3 Tashev I. Dual stage probabilistic voice activity detector / Tashev I., Lovitt A., Acero A. // The Journal of the Acoustical Society of America. – Vol. 127. – Issue 3. – 2010. – P. 1816-1816.

4 Fujihara H. F0 Estimation Method for Singing Voice in Polyphonic Audio Signal Based on Statistical Vocal Model and Viterbi Search / Fujihara H., Kitahara T., Masataka G., Komatani K., Ogata T., Okuno H. // Conference: Acoustics, Speech and Signal Processing. – Vol. 5. – 2006. – P. 253-256.

5 Peiszer E. Automatic Audio Segmentation: Segment Boundary and Structure Detection in Popular Music. / Peiszer E., Lidy T., Rauber A. // Proc. of LSAS. – Paris, France. – 2008. – 114 p.

6 Li F. An Automatic Segmentation Method of Popular Music Based on SVM and Self-similarity / Li F.,You Y., Lu Y., Pan Y. // Lecture notes in computer science, human centered computing. – Vol. 8944. – 2015. – P.15–25.

7 Maddage N. Automatic Structure Detection for Popular Music // IEEE MultiMedia. – Vol. 13. – 2006. – P. 65-77.

8 Гураков И.А. Статистические распределения формант различных дикторов при проведении фоноскопических экспертиз методом формантного выравнивания / Гураков И.А., Костюченко Е.Ю., Новохрестова Д.И., Шелупанов А.А.// Информационные технологии в управлении (ИТУ-2018) материалы конференции. – 2018. – С. 501-509.

9 Шелупанов А.А. Математическое и программно-алгоритмическое обеспечение в задачах идентификации и распознавания речи / Шелупанов А.А., Мещеряков Р.В., Конев А.А., Бондаренко В.П.// Вестник Сибирского

государственного аэрокосмического университета им. академика М.Ф. Решетнева. – 2006. – № S. – С. 11-15.

10 Рахманенко И.А. Верификация диктора по произвольной фразе с помощью сверточной глубокой сети доверия и Гаусовой смеси /Рахманенко И.А., Мещеряков Р.В.// Безопасные информационные технологии Сборник трудов Восьмой всероссийской научно-технической конференции. НУК «Информатика и системы управления». Под. ред. М.А.Басараба. – 2017. – С. 364-367.

11 Kharchenko S.S. Fundamental frequency evaluation subsystem for natural speech rehabilitation software calculation module for cancer patients after larynx resection /Kharchenko S.S., Mescheryakov R.V., Volf D.A., Balatskaya L.N., Choinzonov E.L.// Proceedings - 2015 International Conference on Biomedical Engineering and Computational Technologies, SIBIRCON 2015. – 2015. – P. 197-200.

12 Балацкая Л.Н. Речевая реабилитация и качество жизни после хирургического лечения больных раком гортани//Сибирский онкологический журнал. – 2003. – № 2. – С. 54-57.

13 Kostuchenko E., Assessment of Syllable Intelligibility Based on Convolutional Neural Networks for Speech Rehabilitation After Speech Organs Surgical Interventions /Kostuchenko E., Novokhrestova D., Pekarskikh S., Shelupanov A., Nemirovich-Danchenko M., Choynzonov E., Balatskaya L.// SPECOM 2019: Speech and Computer. – 2019. – P. 359-369.

14 Dietz J.H. Adaptive rehabilitation in cancer: A program to improve quality of survival // Postrad. Med. –1980. – Vol. 68. – P. 145-163.

15 Job J. R. Song adjustments by an open habitat bird to anthropogenic noise, urban structure, and vegetation /Job J. R., Kohler S. L., Gill S. A.// Behavioral Ecology. – Vol. 27. – Issue 6. – 2016. – P. 1734–1744.

16 Portfors C. V., Perkel, D. J. The role of ultrasonic vocalizations in mouse communication. Current opinion in neurobiology. – 2014. – Vol. 28. – P. 115-20.

17 Luque A., Romero-Lemos J., Carrasco A., Barbancho J. Non-sequential automatic classification of anuran sounds for the estimation of climate-change indicators // Expert Systems with Applications. – Vol. 95. – 2018. – P. 248-260.

18 Grunst M. L., Grunst A.S., Formica V. A., Gonser R.A., Tuttle E.M. Multiple signaling functions of song in a polymorphic species with alternative reproductive strategies // Ecology and evolution. – 2017. – Vol. 8(2). – P. 1369-1383.

19 Федотова М.В. Мелодическая структура восходяще-нисходящего тона как маркер валлийского акцента в английском языке // Вестник Московского государственного лингвистического университета. Гуманитарные науки. – 2011. – № 607. – С. 233-244.

20 Жаровская Е.В. Просодические особенности речи молодежи // Филологические науки. Вопросы теории и практики. – 2018. – № 8-1 (86). – С. 95-99.

21 Сокорева Т.В. Роль высотно-мелодического компонента в сохранении и развитии ритмических тенденций // Вестник Московского государственного лингвистического университета. Гуманитарные науки. – 2017. – № 771. – С. 105-117.

22 Жаровская Е.В. Характеристика элементов мелодического рисунка речи // Филологические науки. Вопросы теории и практики. – 2017. – № 7-3 (73). – С. 112-114.

23 Шук С.В. Акустические признаки позитивной и негативной оценки в британском радиорепортаже // Вестник Полоцкого государственного университета. Серия А: Гуманитарные науки. – 2011. – № 10. – С. 78-82.

24 Murthy Y.V.S. , Koolagudi S.G. Classification of Vocal and Non-vocal segments in Audio Clips using Genetic Algorithm based Feature Selection (GAFS) // Expert Systems with Applications. – 2018. – Vol. 106. – P. 77-91.

25 Finley, Michael & Razi, Abolfazl. Musical Key Estimation with Unsupervised Pattern Recognition. // 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). – 2019. – P. 401-408.

26 Bader R. Computational Music Archiving as Physical Culture Theory // Computational Phonogram Archiving. Current Research in Systematic Musicology. – Vol 5. – Springer, Cham. – 2019. – P. 3-34.

27 McFee B., Wook K.J., Cartwright M., Salamon J. M., Bittner R., Pablo B. J. Open-Source Practices for Music Signal Processing Research: Recommendations for Transparent, Sustainable, and Reproducible Audio Research // IEEE Signal Processing Magazine. – 2019. – Vol. 36. – P. 128-137.

28 Li H., You H., Fei X., Yang M., Chao K., He C. Automatic Note Recognition and Generation of MDL and MML using FFT // 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE), Xi'an. – 2018. – P. 195-200.

29 Kiska T., Galaz Z., Zvoncak V., Mucha J., Mekyska J., Smekal Z. Music Information Retrieval Techniques for Determining the Place of Origin of a Music Interpretation // 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). – Moscow, Russia. – 2018. – P. 1-5.

30 Murthy, Y. V. Srinivasa et al. Vocal and Non-vocal Segmentation based on the Analysis of Formant Structure // 2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR). – 2017. – P. 1-6.

31 Masataka Goto, Yoichi Muraoka. Real-time beat tracking for drumless audio signals: Chord change detection for musical decisions // Speech Communication. – Vol. 27. – Issues 3–4. – 1999. – P. 311-335.

32 Балтийский И.А., Николенко С.И. Обзор графических вероятностных моделей гармонии для анализа музыкальных произведений // Труды СПИИРАН. – 2011. – № 2 (17). – С. 174-196.

33 Глазырин Н.Ю. О задаче распознавания аккордов в цифровых звукозаписях // Известия Иркутского государственного университета. Серия: Математика. – 2013. – Т. 6. – № 2. – С. 2-17.

34 Masataka Goto. A real-time music-scene-description system: predominant-F0 estimation for detecting melody and bass lines in real-world audio signals // Speech Communication. – Vol. 43. – Issue 4. – 2004. – P. 311-329.

35 Способин И.В. Элементарная теория музыки. — М.: Музыка, 1968. – 204 с.

36 Тюлин Ю.Н. Краткий теоретический курс гармонии. — М.: Музыка, 1978. – 212 с.

37 Искусство пения: Учебное пособие. — 4-е изд., стер. — СПб.: Издательство «Лань»; Издательство «ПЛАНЕТА МУЗЫКИ», 2019. — 212 с.: ил., ноты. — (Учебники для вузов. Специальная литература).

38 Aronson, Arnold Elvin; Bless, Diane M. Clinical Voice Disorders (4th ed.). New York, NY: Thieme Medical Publishers. – 2009. – 278 p.

39 Фант Г. Анализ и синтез речи / Г. Фант. – Новосибирск: Наука. – 1970. – 306 с.

40 Шарий, Т. В. О проблеме параметризации речевого сигнала в современных системах распознавания речи / Т. В. Шарий // Вісник Донецького національного університету. – Сер. А: Природничі науки. – Вип. 2. – 2008. – С. 536–541.

41 Винцюк Т. К. Алгоритмы распознавания слов и слитных фраз и результаты их моделирования / Т. К. Винцюк, О. Н. Гаврилюк, Н. Г. Пучкова. Тезисы докладов VIII Всесоюзного семинара АРСО. – Львов, 1974. – Ч.3 – С. 33-37

42 Рабинер Р. Л. Цифровая обработка речевых сигналов / Р. Л. Рабинер, Р. В. Шафер – М.: Радио и связь. – 1981. – 496 с

43 Матвеев Ю.Н., Симончик К.К., Тропченко А.Ю., Хитров М.В. Цифровая обработка сигналов // Учебное пособие: СПбНИУ ИТМО. – 2013. – 166 с.

44 M.R. Schroeder. Number theory in music, speech, and acoustics // The Journal of the Acoustical Society of America. – 1984. – Vol. 76. – Issue S1. –S58-S58.

45 Azarov E. Instantaneous pitch estimation based on RAPT framework / E. Azarov, M. Vashkevich, A. Petrovsky // Proceedings of EUSIPCO'12 — European Signal Processing Conference. – Bucharest, Romania – August 27-31, 2012. — P. 2787-2791.

46 Talkin D. A robust algorithm for pitch tracking (RAPT) // Speech Coding and Synthesis, W. B. Kleijn and K. Paliwal, Eds. New York: Elseviever. – 1995. — P. 495-518.

47 Вашкевич М.И., Азаров И.С., Петровский А.А. Оценка мгновенной частоты основного тона речевого сигнала на основе многоскоростной обработки // Речевые технологии. – 2018. – № 1-2. – С. 12-24.

48 Gonzalez S. PEFAC — A Pitch Estimation Algorithm Robust to High Levels of Noise / S. Gonzalez, M. Brookes // IEEE/ACM Transactions on Audio, Speech, and Language Processing. – 2014. – Vol. 22. – No.2. – P. 518-530.

49 Гитлин В.Б. Выделение основного тона речи методом SWIPE из сигнала, ограниченного полосой телефонного канала / В.Б. Гитлин, Д.Ю. Вашурин // Речевые технологии. – 2014. – №1. – С. 57-74.

50 De Cheveigné, A., Kawahara, H. YIN, a fundamental frequency estimator for speech and music // The Journal of the Acoustical Society of America. 2002. Vol. 111. [Электронный ресурс]. – Режим доступа: http://asa.scitation.org/doi/abs/10.1121/1.1458024.

51 Camacho A. SWIPE: A sawtooth waveform inspired pitch estimator for speech and music // Ph.D. dissertation. – Florida: Univ. of Florida, 2007. – 116 p.

52 Вольф Д.А. Модель, численная и программная реализация оценивания частоты основного тона речевого сигнала с помощью сингулярного спектрального анализа: дис. … канд. техн. наук. – Томск, 2015. – 149 с.

53 Морозов В.П. Компьютерная диагностика вокальной одаренности // Голос и речь. – 2010. – № 1. – С. 81-93.

54 Leydon C., Bauer J.J., Larson C.R. The role of auditory feedback in sustaining vocal vibrato // Acoustical Society of America. – Vol. 114(3). – 2003. – P. 1575-1581.

55 Prame E. Vibrato extent and intonation in professional Western lyric singing // Acoustical Society of America. – Vol. 102(1). – 1997. – P. 616–621.

56 Reddy A., Subramanian U. Singers’ and nonsingers’ perception of vocal vibrato // J. Voice. – Vol. 29(5). – 2015. – P. 603–610

57 Агин М.С. Основные недостатки певческого голоса и речи и пути их преодоления // Голос и речь. – 2011. – № 3. – С. 79-90.

58 Zhang M., Bocko M., Beauchamp J. Measurement and analysis of musical vibrato parameters // Journal of the Acoustical Society of America. – Vol. 137. – 2015. – P. 2404-2404.

59 Морозов В.П., Морозов П.В. Вибрато голоса мастеров вокального искусства // Компьютерные исследования. Вопросы вокального образования методические рекомендации Совета по вокальному искусству для преподавателей вузов и средних спец. учебных заведений. – Санкт-Петербург. – 2007. – С. 33-45.

60 Абдуллин Э.Б., Чжан И. Анализ причин возникновения вокальной тремоляции и способы её устранения // Вестник кафедры ЮНЕСКО Музыкальное искусство и образование. – 2017. – № 4 (20). – С. 125-131.

61 Michel C., Ruiz M. (2017). The physics of singing vibrato // Physics Education. – Vol. 52 (4). – 2017. – P. 1-6.

62 Frič M., Pavlechová A. Listening evaluation and classification of female singing voice categories // Logopedics Phoniatrics Vocology. – 2019. – P. 1-13.

63 Jansens S., Bloothooft G., De Krom G. Perception and acoustics of emotions in singing // Proceedings of the Fifth European Conference on Speech Communication and Technology, Rhodes, Greece. – Vol. 4. – 1997. – P. 2155–2158.

64 Sundberg J. Acoustic and psychoacoustic aspects of vocal vibrato // Dejonckere PH, Hirano M, Sundberg J, eds. Vibrato. San Diego, Calif: Singular Publishing Group Inc. –1995. – P.35-62.

65 Kotlyar G. M., Morozov V. P. Acoustical correlates of the emotional content of vocalized speech // Sov. Phys. Acoust. – Vol. 22. – 1976. – P. 208–211.

66 Гай В.Е., Утробин В.А., Родионов П.А., Дербасов М.О. Оценка эмоционального состояния человека по голосу с позиций теории активного восприятия // Системы управления и информационные технологии. – 2015. – Т. 59. – № 1-1. – С. 118-122.

67 Nwe T. L., Foo S. W., De Silva L. C. Speech emotion recognition using hidden Markov models // Speech communication. – 2003. – Vol. 41. – No. 4. – P. 603-623.

68 El Ayadi M., Kamel M. S., Karray F. Survey on speech emotion recognition: Features, classification schemes, and databases // Pattern Recognition. – 2011. –Vol. 44. – No. 3. – P. 572-587.

69 Scherer K., Sundberg J., Fantini B., Trznadel S., Eyben F. The expression of emotion in the singing voice: Acoustic patterns in vocal performance // The Journal of the Acoustical Society of America. – Vol. 142. – 2017. – P. 1805-1815.

70 Конев А. А. Модель и алгоритмы анализа и сегментации речевого сигнала // Диссертация на соискание ученой степени кандидата технических наук. – Томск: ТУСУР, 2007. – 150 с.

71 Bouafif L., Ouni K. A Speech Tool Software for Signal Processing Applications // 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT). – 2012. – P. 788-791.

72 Cherouat S., Marir F. Pitch detection and Voicing/Unvoicing decision of Arabic speech signal by HOS-Polycesptre // 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT). – 2012. – P. 768-771.

73 Кипяткова, И. С. Автоматическая обработка разговорной русской речи: монография / И. С. Кипяткова, А. Л. Ронжин, А. А. Карпов. – СПИИРАН. – СПб.: ГУАП, 2013. – 314 с.

74 Elmir Y. Score Level Fusion Based Multimodal Biometric Identification (Fingerprint & Voice) / Y. Elmir, A. Elberrichi, R.Adjoudj // Proceedings of 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), 2012. – P. 146-150.

75 The Evaluation Process Automation of Phrase and Word Intelligibility Using Speech Recognition Systems / Kostuchenko E. et. // SPECOM 2019: Speech and Computer – P. 237-246.

76 Гай В.Е. Обзор методов распознавания голосовых сигналов в условиях нулевых ресурсов / В.Е. Гай, Филяков А.А., Лукьянчикова А.В. // Фундаментальные проблемы радиоэлектронного приборостроения: материалы Международной научно-технической конференции «INTERMATIC–2015». 2015. С. 215-218.

77 Kamper H., Jansen A., Goldwater S. A segmental framework for fully-unsupervised large-vocabulary speech recognition // Computer Speech and Language. – 2017. – Vol. 46. – P. 154–174.

78 Общая и прикладная фонетика: учеб. пособие / Л. В. Златоустова, Р. К. Потапова, В. В. Потапов, В. Н. Трунин-Донской. – 2-е изд., перераб. и доп. – М.: Изд-во МГУ, 1997. – 416 с.

79 Бондарко Л.В. Основы общей фонетики / Л.В. Бондарко, Л.А. Вербицкая, М.В Гордина. – СПб., 2004. – 160 с.

80 Schroeter J. Speaker adaptation in articulatory speech analysis by synthesis / J. Schroeter, J. N. Larar, M. M. Sondhi. // The Journal of the Acoustical Society of America. – 1987. – Vol. 82. – P. 54. doi 10.1121/1.2024868.

81 Физиология речи. Восприятие речи человеком / Л. А. Чистович и др. – Л.: Наука, 1976. – 388 с.

82 Златоустова Л.В. Фонетические единицы русской речи / Л.В. Златоустова // М.: Изд-во Московского института, 1981. – 105 с.

83 Sangeetha J. Robust Automatic Continuous Speech Segmentation for Indian Languages to Improve Speech to Speech Translation / J. Sangeetha, S. Jothilakshmi // International Journal of Computer Applications. – Vol. 53 (15). – 2012. – P.13-16

84 A Phonetic Segmentation Procedure Based on Hidden Markov Models / E. Pakoci et. // Lecture Notes in Computer Science. – 2016. – Vol. 9811. – P. 67–74.

85 Awata S. Vowel duration dependent hidden Markov model for automatic lyrics recognition / S. Awata, S. Sako, T. Kitamura // Journal of the Acoustical Society of America. – 2016. – Vol. 140. – P. 3427. Doi 10.1121/1.4971035.

86 Train&Align: a new online tool for automatic phonetic alignment / S. Brognaux, S. Roekhaut, T. Drugman, R Beaufort // Proceedings of IEEE Signal Processing Society. Spoken Language Technology Workshop (SLT), 2012. – P. 416-421.

87 Scharenborg O. Unsupervised speech segmentation: An analysis of the hypothesized phone boundaries / Scharenborg O., Wan V., Ernestus M. // The Journal of the Acoustical Society of America. – 2010. – Vol. 127, № 2. – P. 1084-1095.

88 Zheng N.-H Music segmentation based on model adaptation and smoothing processing / N.-H. Zheng, , Y.-L. Zhang, X. Li. – 2011. – Vol. 28. – P. 271-275.

89 Chen H.-C. Music segmentation by rhythmic features and melodic shapes / H.-C. Chen , C.-H. Lin , A Chen. // Proceedings of 2004 IEEE International Conference on Multimedia and Expo (ICME). – Vol.3. – P. 1643-1646. Doi 10.1109/ICME.2004.1394566.

90 Monaghan P. Disambiguating durational cues for speech segmentation / P. Monaghan, L. White, M. Merkx // The Journal of the Acoustical Society of America – 2013. – Vol. 134(1). – P. 45–57.

91 The role of stress and word size in Spanish speech segmentation / A. Lacross et. // The Journal of the Acoustical Society of America. – 2016 – Vol. 140. – P. 484-490. Doi 10.1121/1.4971227.

92 Benati N. Spoken term detection based on acoustic speech segmentation / N. Benati, H. Bahi // Proceedings of 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications. SETIT 2016. – P. 267–271.

93 Кипяткова И.С. Автоматическая обработка разговорной русской речи / И.С. Кипяткова, А.Л. Ронжин, А.А. Карпов. – Санкт-Петербург, 2013. –314 с.

94 Chan A. Information Distribution Within Musical Segments. / A. Chan, J. Hsiao // Music Perception: An Interdisciplinary Journal. – 2016. – Vol. 34. – P. 218-242. 10.1525/mp.2016.34.2.218.

95 Vitela A Lexical segmentation of speech from energy above 5 kHz / A Vitela, B. Monson, A. Lotto // The Journal of the Acoustical Society of America. – 2013. – Vol. 134 (5). – P. 4072. Doi 10.1121/1.4830868.

96 Taniguchi T. Detection of speech and music based on spectral tracking / T. Taniguchi, M. Tohyama, K. Shirai // Speech Communication. – 2008. – Vol. 50, iss. 7. – P. 547-563.

97 Music Segmentation With Genetic Algorithms / B. Rafael, S. Oertl, M. Affenzeller, S. Wagner // Proceedings of International Workshop on Database and Expert Systems Applications, DEXA, 2009. – P. 256-260. Doi 10.1109/DEXA.2009.16.

98 Peiszer E. Automatic audio segmentation: Segment boundary and structure detection in popular music: master's thesis / E. Peiszer //Vienna: Vienna University of Technology, 2007. – 114 pp.

99 Lee K. Segmentation-based lyrics-audio alignment using dynamic programming / Lee K., Cremer M. // Proceedings of the 9th International Conference on Music Information Retrieval, 2008. – P. 395–400.

100 Jun, S. Music segmentation and summarization based on self-similarity matrix / S. Jun, E. Hwang // Proceedings of the 7th International Conference on Ubiquitous Information Management and Communication (ICUIMC), 2013. – doi 10.1145/2448556.2448638.

101 A Wavelet-Based Parameterization for Speech/Music Segmentation / E. Didiot, I. Illina, O. Mella, D. Fohr // Music Discrimination. Computer Speech and Language. – 2010. – Vol. 24(2). – P 341-357.

102 Didiot E. Speech/music segmentation for automatic transcription of continuous speech / E. Didiot. – Nancy: Henri Poincaré University, 2007 – 160 pp.

103 Demir C. Speech-music segmentation system for speech recognition / C. Demir, U. D. Mehmet // Proceedings of IEEE 17th Signal Processing and Communications Applications (SIU), 2009. – P. 624-627. Doi 10.1109/SIU.2009.5136473.

104 Levy M. Extraction of high-level musical structure from audio data and its application to thumbnail generation / M. Levy, M. Sandler, M. Casey // Proceedings of the 2006 IEEE Conference on Acoustics, Speech and Signal Processing, 2006.

105 Noland K. Key estimation using a hidden Markov model / K. Noland, M. Sandler // Proceedings of the 7th International Conference on Music Information Retrieval, Victoria, Canada, 2006.

106 Levy M. A Comparison of Timbral and Harmonic Music Segmentation Algorithms / M. Levy, K. Noland, M. Sandler // Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07, Honolulu, HI, 2007. – P. 1433-1436.

107 Harte C. Automatic chord identification using a quantised chromagram / C. Harte, M. Sandler // Proceedings of AES 118th Convention, Barcelona, 2005. – P. 1-6.

108 Ke W.J. Representative Music Fragment Extraction by Using Segmentation Techniques / W.-J. Ke, C.-W. Chang, H.C. Jiau // Proceedings of the International Computer Symposium, 2004. – P. 1156-1161.

109 Cambouropoulos E. A Formal Theory for the Discovery of Local Boundaries in a Melodic Surface / E. Cambouropoulos // Proceedings of the III Journees d' Informatique Musicale, 1996. – P. 1-10.

110 Cambouropoulos E. The Local Boundary Detection Model (LBDM) and its Application in the Study of Expressive Timing / E. Cambouropoulos // Proceedings of the International Computer Music Conference, 2001. – P. 17–22.

111 Chang C.-W. A Heuristic Approach for Music Segmentation / C.-W. Chang, W.-J. Ke, H.C. Jiau // Innovative Computing, Informatio and Control: Proceedings of international Conference, 2007. – P. 228. doi 10.1109/ICICIC.2007.30.

112 Логинова Л.Н. Актуальные проблемы современного сольфеджио / Л.Н. Логинова // Южно-Российский музыкальный альманах. – 2012. – № 2 (11). – С. 34-42.

113 Воронцова И.В. История сольфеджио: от научения к учению / И.В. Воронцова // Музыкальная академия. – 2011. – № 3. – С. 171-172.

114 Римский-Корсаков Н.А. О музыкальном образовании // Н.А. Римский-Корсаков. – Литературное наследие и переписка. – М.: Музыка, 1963.– Т. 2.

115 Володин, А. А. Психологические аспекты восприятия музыкальных звуков: дис. … докт. психологич. наук / Володин Андрей Александрович. – МГУ им. Ломоносова, 1969. – 508 С.

116 Révész G. Introduction to the Psychogy of music / G. Révész. – N. Y.: Dover Publications, 2001. – 288 pp.

117 Белов Е.В. Компьютерное обучение вокалу / Е.В Белов. // Педагогическая информатика. – 2011. – №2. – С. 63-66.

118 Сороколетова Н.Ю. Функциональные возможности компьютерного приложения Praat / Н.Ю. Сороколетова // Иностранные языки: лингвистические и методические аспекты. – 2014. – № 27. – С. 127-131.

119 Дадиомов А.Е. Обзор компьютерных обучающих программ по сольфеджио / А.Е. Дадиомов // Вопросы музыкознания: Теория. История. Методика: сборник научных статей. – Москва, 2011. – С. 134-141.

120 Пыхов В.В. Использование технических средств в процессе обучения вокалу / В.В. Пыхов // Традиции и инновации в педагогическом образовании: сборник научных трудов III Международного круглого стола. – 2017. – С. 238-240.

121 Моисеев Е.О. Технологические аспекты использования электронных образовательных ресурсов в процессе обучения подростков эстрадному вокалу / Е.О. Моисеев // Электронное обучение в непрерывном образовании. – 2018. – С. 382-388.

122 Praat: doing Phonetics by Computer [Электронный ресурс]. – Режим доступа: http://www.fon.hum.uva.nl/praat, свободный (дата обращения: 13.08.2019 г.).

123 Melodyne [Электронный ресурс]. – Режим доступа: https://www.celemony.com/en/start, свободный (дата обращения: 23.05.2019 г.).

124 Singing Software | Voice Training | Vocal Training | SING&SEE [Электронный ресурс]. – Режим доступа: https://www.singandsee.com, свободный (дата обращения: 10.05.2019 г.).

125 Singing Coach [Электронный ресурс]. – Режим доступа: http://www.listening-singing-teacher.com, свободный (дата обращения: 13.01.2019 г.).

126 EarMaster [Электронный ресурс]. – Режим доступа: https://www.earmaster.com, свободный (дата обращения: 17.03.2019 г.).

127 VocTeacher [Электронный ресурс]. – Режим доступа: http://vocteacher.mazaycom.ru/ru/index.php, свободный (дата обращения: 02.04.2019 г.).

128 Мациевская С.В. Методы обучения вокалу / С.В Мациевская. // Вестник Полоцкого государственного университета. Серия Е: Педагогические науки. – 2011. – № 7. – С. 52-55.

129 Родина Т.Б. Музыкальный диктант на занятиях сольфеджио / Т.Б. Родина // Aktualni pedagogika. – 2017. – № 1. – С. 86-93.

130 Миллер Дж. А. Магическое число семь плюс или минус два. О некоторых пределах нашей способности перерабатывать информацию / Дж. А. Миллер // Инженерная психология: сборник статей / под ред. Д. Ю. Панова и В. П. Зинченко; перевод с английского. – М.: Издательство «Прогресс», 1964. – С. 564–580.

131 Якимук А.Ю. Генерация фильтров для одновременной маскировки / А.Ю. Якимук // Электронные средства и системы управления: Материалы докладов XIV Международной научно-практической конференции (28–30 ноября 2018 г.): в 2 ч. – Ч. 2. – Томск: В-Спектр, 2018. – С. 29-31.

132 Егошин Н.С. Идентификация параметров речевого сигнала / Н.С. Егошин, А.А. Конев, А.Ю. Якимук // Электронные средства и системы управления. – 2015. – № 1-2. – С. 147-150.

133 Томская К.М. Определение метода шкалирования для идентификации нот с помощью частот основного тона / К.М. Томская // Российская наука в современном мире: сборник статей XVI международной научно-практической конференции. – 2018. – С. 88-91.

134 Якимук А.Ю. Исследование работы алгоритма идентификации нот для выбора метода определения границ ноты / А.Ю. Якимук, К.М. Томская // Наука. Технологии. Инновации Сборник научных трудов. В 9-ти частях. Под ред. А.В. Гадюкиной. – 2018. – С. 215-219.

135 Якимук А.Ю. Программное обеспечение для автоматического распознавания мелодии / А.Ю. Якимук, А.А. Конев // Технологии Microsoft в теории и практике программирования: сборник трудов XII Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых – Томск: Изд-во Томского политехнического университета, 2015. – С. 247-248.

136 Якимук А.Ю. Алгоритм сегментации речевого сигнала на основе значений минимальной меры различия / А.Ю. Якимук, А.А. Конев // Информатика и системы управления. – 2018. – № 2 (56). – С. 108-121.

137 Бондаренко В.П. Адаптивный анализ голосового сигнала / В.П. Бондаренко, В.П. Коцубинский, Р.В. Мещеряков // Интеллектуальные системы в управлении, конструировании и образовании – Томск: STT, 2004. – 216 с. – С.58-61.

138 Бондаренко В.П. Модель одновременной маскировки / В.П. Бондаренко, А.А. Пономарев, Е.А. Рогозинская // Интеллектуальные системы в управлении, конструировании и образовании – Томск: STT, 2004. – 216 с. – С. 167-174.

139 Якимук А.Ю. Этапы работы программного комплекса, определяющего ноты вокального исполнения / А.Ю. Якимук // Информационные технологии в экономике и управлении: материалы III Всероссийской научно-практической конференции, г. Махачкала, 29-30 ноября 2018 г.: Дагестанский государственный технический университет. – Махачкала, 2018. – С. 154-157.

140 Якимук А.Ю. Алгоритмическое обеспечение системы анализа шепотной речи / А.Ю. Якимук, А.А. Конев, Ю.А. Терещенко // «Вестник Брянского государственного технического университета». – №10 (71). – 2018. – С. 62-71.

141 Cirillo J. Communication by unvoiced speech: the role of whispering / J. Cirillo // Annals of the Brazilian Academy of Sciences. – 2004. – Vol. 76 №. 2. – P. 413-423.

142 Якимук А.Ю. Исследование надёжности детектора частоты основного тона голосового сигнала / А.Ю. Якимук // Научная сессия ТУСУР – 2015: Материалы Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых – Томск: В-Спектр, 2015. – С. 194-196.

143 Якимук А.Ю. Повышение качества идентификации нот в автоматизированной системе распознавания вокала / А.Ю. Якимук, Н.С. Егошин, А.О. Осипов, И.М. Боков // Электронные средства и системы управления:

Материалы докладов XII Международной научно-практической конференции – Томск: В-Спектр, 2016. – С. 29-32.

144 Якимук А.Ю. Программный комплекс для автоматизации моделирования сегментации речевых сигналов и вокальных исполнений / А.Ю. Якимук, А.А. Конев, А.О. Осипов // Вестник Иркутского государственного технического университета. – 2017. – Т. 21. № 10 (129). – С. 53-64.

145 Konev, A. The program complex for vocal recognition / A. Konev, E. Kostyuchenko, A. Yakimuk // Journal of Physics: Conference Series. – 2017. – Vol. 803. – Issue 1. – P. 012077. – DOI: 10.1088/1742-6596/803/1/012077.

146 Бондаренко В.П. Обработка речевых сигналов в задачах идентификации / В.П. Бондаренко, А.А. Конев, Р.В. Мещеряков // Известия высших учебных заведений. Физика. 2006. – Т. 49. № 9. – С. 207–210.

147 Конев А.А. Автоматическое распознавание музыкальных нот / А.А. Конев, А.А. Онищенко, Е.Ю. Костюченко, А.Ю. Якимук // Научный вестник Новосибирского государственного технического университета. – 2015. – № 3 (60). – С. 32-47.

148 Yakimuk A.Yu. Applying the principle of distribution in the program complex for vocal recognition / A.Yu. Yakimuk, A.A. Konev, Yu.V. Andreeva, M.M. Nemirovich-Danchenko // IOP Conf. Series: Materials Science and Engineering. – 2019. – Vol. 597 – P. 012072. doi:10.1088/1757-899X/597/1/012072.

149 Якимук А.Ю. Алгоритмы анализа частоты основного тона вокального исполнения / А.Ю. Якимук // Научная сессия ТУСУР–2016: материалы Международной научно-технической конференции студентов, аспирантов и молодых ученых, Томск, 25–27 мая 2016 г. – Томск: В-Спектр, 2016. – С. 245-248.

150 Якимук А.Ю. Распределенный программный комплекс по распознаванию нот / А.Ю. Якимук, М.Д. Холопов // Перспективы развития фундаментальных наук: сборник трудов XVI Международной конференции студентов, аспирантов и молодых ученых (Томск, 23–26 апреля 2019 г.) в 7 томах. Том 7. IT-технологии и электроника / под ред. И.А. Кузиной, Г.А. Вороновой. – Томск: Изд-во Томского политехнического университета, 2019. – С. 125-127.

151 Якимук А.Ю. Применение программных средств при обучении вокалу / А.Ю. Якимук // Проблемы управления качеством образования: сборник статей XI Всероссийской научно-практической конференции / МНИЦ ПГАУ. – Пенза: РИО ПГАУ, 2018. – С. 141-144.

152 Иванова А.З. Роль педагога в формировании личности студента-певца / А.З. Иванова // Современное общество: проблемы, идеи, перспективы движения в социокультурном пространстве: сборник научных статей по итогам III Международной очно-заочной научно-практической конференции. – 2017. – С. 133-135.

153 Курлапов Н.И. Значение вокального педагога в формирования личности ученика / Н.И. Курлапов // Обучение и воспитание: методики и практика 2016/2017 учебного года: сборник материалов XXXV Международной научно-практической конференции. – 2017. – С. 90-97.

154 Островский А.Л. Учебник сольфеджио. Вып. 1. 2-е издание / А.Л. Островский. — М.: Музыка, 1966. — 228 с.

155 Катаева Е.С. Применение выделения синхронности для оценки сходства вокальных исполнений / Е.С. Катаева, Ю.Р. Свешникова, А.Ю. Якимук // Информационно-коммуникационные технологии в педагогическом образовании. 2019. № 4 (61). С. 54-58.

156 Катаева Е. С. Применение алгоритма выделения синхронности для метеорологических временных рядов/ Е. С. Катаева, Г.М. Кошкин // Известия вузов. Физика. — Т.56, № 9/2. — С.229–231.

157 Мелодии кино и мультфильмов: сборник одноголосных музыкальных диктантов / состав.: Жанна Борисевич, Маргарита Кочарян. — Винница: Нова Книга, 2012. — 88 с.: ноты.

Приложение А

Свидетельства о государственной регистрации программы для ЭВМ

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2017664232

Программный комплекс по определению нот вокального исполнения

Правообладатель: ***Федеральное государственное бюджетное образовательное учреждение высшего образования «Томский государственный университет систем управления и радиоэлектроники» (ТУСУР) (RU)***

Авторы: ***Конев Антон Александрович (RU), Якимук Алексей Юрьевич (RU), Осипов Андрей Олегович (RU)***

Заявка № **2017660858**
Дата поступления **26 октября 2017 г.**
Дата государственной регистрации
в Реестре программ для ЭВМ ***19 декабря 2017 г.***

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2017664235

Программа для определения качества сегментации речевых сигналов

Правообладатель: ***Федеральное государственное бюджетное образовательное учреждение высшего образования «Томский государственный университет систем управления и радиоэлектроники» (ТУСУР) (RU)***

Авторы: ***Конев Антон Александрович (RU),***
Якимук Алексей Юрьевич (RU)

Заявка № **2017660865**
Дата поступления **26 октября 2017 г.**
Дата государственной регистрации
в Реестре программ для ЭВМ ***19 декабря 2017 г.***

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев

Приложение Б

Акты внедрения

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ» (ТУСУР)

УТВЕРЖДАЮ
Проректор по учебной работе ТУСУР
_____ П.В. Сенченко
«20» 09 2019 г.

АКТ
о внедрении результатов диссертационной работы
Якимука Алексея Юрьевича в учебный процесс

Комиссия в составе:

Давыдова Е.М., к.т.н., декан факультета безопасности ТУСУР – председатель комиссии;

Конев А.А., к.т.н., доцент кафедры КИБЭВС ТУСУР;

Кручинин Д.В., к.ф.-м.н., доцент кафедры КИБЭВС ТУСУР;

Рахманенко И.А., к.т.н., доцент кафедры БИС ТУСУР

составила настоящий акт о нижеследующем.

Результаты диссертационной работы Якимука А.Ю., используются в учебном процессе на факультете безопасности ТУСУР при проведении практических занятий по дисциплинам «Системный анализ» и «Моделирование автоматизированных информационных систем» в рамках подготовки студентов, обучающихся по специальностям «10.05.02 – Информационная безопасность телекоммуникационных систем», «10.05.03 – Информационная безопасность автоматизированных систем» и «10.05.04 – Информационно-аналитические системы безопасности».

Результаты Якимука А.Ю. по исследованию слуховой системы человека используются в практических занятиях по дисциплинам «Моделирование автоматизированных информационных систем» и «Системный анализ», что

позволяет студентам ознакомиться с примером создания математической модели по реальной системе.

Кроме того, студенты факультета безопасности имеют возможность ознакомиться с результатами диссертационного исследования в ходе выполнения групповых проектов, научно-исследовательских и дипломных работ и использовать их в практических работах по исследованию параметров речевых сигналов.

Освоение студентами предложенного Якимуком А.Ю. подхода позволяет сформировать навыки построения математических моделей и моделирования в решении прикладных задач

Настоящий акт составлен в 3 (трех) экземплярах.

Давыдова Е.М.
к.т.н., декан факультета
безопасности ТУСУР

«20» 09 2019 г.

Конев А.А.
к.т.н., доцент кафедры
КИБЭВС ТУСУР

«20» 09 2019 г.

Кручинин Д.В.
к.ф.-м.н., доцент кафедры
КИБЭВС ТУСУР

«20» 09 2019 г.

Рахманенко И.А.
к.т.н., доцент кафедры
БИС ТУСУР

«20» 09 2019 г.

Общество с ограниченной ответственностью «Элекард-ЦТП»

ИНН 7017257102 КПП 701701001
634055, Томская обл., г. Томск,
Пр-кт Развития, д 3
Тел/факс (8-3822) 509-892
16.10.2019 № 11
На №_____от_____

УТВЕРЖДАЮ
Директор
ООО "Элекард-ЦТП"
_____ Шум А.А.
«16» октября 2019 г.

АКТ

внедрения результатов диссертационной работы

Якимука Алексея Юрьевича

Комиссия в составе председателя: директора Шума А.Л., членов комиссии:

- исполнительного директор Ширшина В.А.
- ведущего инженера Левикина В.А.
- аналитика Оленевой А.Е.

составили настоящий акт о том, что результаты диссертационной работы Якимука А.Ю. «Алгоритмы анализа частоты основного тона вокального исполнения», представленной на соискание ученой степени кандидата технических наук, внедрены в деятельность «Элекард-ЦТП» в процессе работы над приложением для визуального контроля качества видео потоков и уровня аудио сигнала в режиме реального времени «MediaQAnalitic module».

На базе приложения был реализован метод дистанционного обучения пению в формате видеоконференций. При таком формате, распознавание нот в вокальном исполнении осуществляется на стороне обучающегося, а к преподавателю поступает результат выполнения задания в формате нотной записи в abc-нотации.

Переработка напеваемых мелодий в ноты по алгоритму, разработанному Якимуком А.Ю., позволила сократить объем трафика, передаваемого по сети. Сокращение объема трафика происходит за счет перехода от передачи

аудиозаписей в формате wav к передаче текстового сообщения с abc-нотацией распознанных нот. Это позволяет снизить общий объем трафика более чем на 90%. За счет этого появилась возможность использовать данные подходы в интернете вещей, где классически существуют проблемы, связанные с пропускной способностью.

Председатель комиссии Шум А.Л.

Члены комиссии Ширшин В.А.

Левикин В.А

Оленева А.Е.

Приложение В

Сертификат гранта Американского Акустического Общества

Acoustical Society of America

ASA
ACOUSTICAL SOCIETY OF AMERICA

The Acoustical Society of America recognizes

Alexey Yakimuk

Tomsk State University of Control Systems and Radioelectronics

as recipient of the

ASA International Student Grant

to assist the research of promising graduate students in acoustics

9 November 2018

ASA President

Chair, Committee on International Research and Education