На правах рукописи

Якимук Алексей Юрьевич

АЛГОРИТМЫ АНАЛИЗА ЧАСТОТЫ ОСНОВНОГО ТОНА ВОКАЛЬНОГО ИСПОЛНЕНИЯ

Специальность 05.13.17 –
«Теоретические основы информатики»

Автореферат
диссертации на соискание учёной степени
кандидата технических наук

Томск – 2019

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего образования «Томский государственный университет систем управления и радиоэлектроники»

Научный руководитель – доктор технических наук, профессор
Шелупанов Александр Александрович

Официальные оппоненты: **Лобанов Борис Мефодьевич**,
доктор технических наук, главный научный сотрудник лаборатории распознавания и синтеза речи Государственного научного учреждения «Объединённый институт проблем информатики Национальной академии наук Беларуси»

Фадеев Александр Сергеевич,
кандидат технических наук, проректор по цифровизации, директор центра цифровых образовательных технологий, доцент отделения информационных технологий Томского политехнического университета

Ведущая организация – Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук

Защита состоится «26» декабря 2019 г. в 16:00 часов на заседании диссертационного совета Д 212.268.05 при Томском государственном университете систем управления и радиоэлектроники (ТУСУР) по адресу: 634050, г. Томск, пр. Ленина 40, ауд. 201.

С диссертацией можно ознакомиться в библиотеке ТУСУР по адресу: 634045, г. Томск, ул. Красноармейская 146, а также на сайте ТУСУР: https://postgraduate.tusur.ru/urls/j3udrfi2

Автореферат разослан «___» _____ 2019 г.

Ученый секретарь
диссертационного совета  Костюченко Евгений Юрьевич

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования.

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. Существующая форма обучения вокалу осуществляется в взаимодействии с репетитором. Чтобы обучение было максимально эффективным, необходимо проводить не менее 2 часов занятий в день, 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. 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, позволяющему 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.).

Как правило, системы идентифицирующие исполненную ноту основаны на алгоритмах вычисления частоты основного тона (ЧОТ). Алгоритмами нахождения 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. 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 также занимает точность сегментации. В задаче идентификации нот сегментация необходима не только на этапе выделения вокализованных и невокализованных участков. 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 исследования является повышение качества распознавания звучащих нот в вокальном исполнении за счёт применения модели слуховой системы человека.

Для достижения поставленной цели исследования необходимо 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 является речевой сигнал вокального исполнения последовательности нот.

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

Основные **методы исследования**, примененные в диссертационной работе – это методы системного анализа, методы моделирования, цифровой обработки сигналов и математической статистики.

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 результатов исследования заключается в развитии технологии анализа частоты основного тона речевого сигнала. Осуществленная модификация математической модели слуховой системы человека позволила расширить диапазон оценки частот 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.

Положения, выносимые на защиту:

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 05.13.17, заключающемся в следующем: 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 05.13.17, заключающемся в следующем: 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 05.13.17, заключающемся в следующем: 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.

Обоснованность и достоверность результатов диссертационной работы определяется полученными результатами осуществленных численных экспериментов на реальных данных, а также возможностью сопоставления результатов, полученных в диссертации, с результатами экспертной оценки.

Внедрение результатов. Результаты диссертационной работы внедрены в деятельность «Элекард-ЦТП» в рамках проекта по дистанционному обучению вокалу в формате видеоконференций, а также в учебный процесс факультета безопасности Томского государственного университета систем управления и радиоэлектроники.

Личный вклад. 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.

Апробация работы. Основные и промежуточные результаты исследования 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».

Работа выполнена в рамках 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.

Публикации по теме диссертации. По материалам исследования опубликовано 19 работ, в том числе 3 работы в изданиях, рекомендованных ВАК РФ.

Структура и объем работы. Диссертация содержит введение, четыре главы, заключение, 3 приложения и список источников из 157 наименований. Объем работы – 121 страница.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность темы исследования, проводимой в научно-квалификационной работе, формулируется цель и задачи, излагаются основные полученные автором 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. Проведенное исследование показало, что существующие алгоритмы анализа частоты основного тона сигнала обладают следующими недостатками:

- высокий процент грубых ошибок;
- ограничение в диапазоне обработки.

Указанные недостатки делают алгоритмы неприменимыми для задач распознавания нот в вокальном исполнении и обучения пению. Обзор алгоритмов сегментации выявил, что наиболее подходящим для определения границ звучащей ноты является сегментация на вокализованные и невокализованные участки за счет высокой надежности при автоматической сегментации и низкой доли пропущенных и лишних границ.

В результате проведенного анализа существующих методов, подходов и алгоритмов, применяемых в задачах вычисления частоты основного тона, сегментации речевого сигнала или обучения пению, были определены следующие требования к программному комплексу по определению нот вокального исполнения:

– распознавание нот в вокальном исполнении с частотами основного тона, звучащими выше 400 Гц;

– сегментация нот с учетом минимальной длительности звучания ноты;

– оценка качества пения.

Во второй главе описывается модификация математической модели слуховой системы, позволившая автоматическое формирование наборов шаблонов для определения частот основного тона в вокальном исполнении. Описывается исходная модель слуховой системы человека и проведенная модификация. Показаны результаты проведенного тестирования работы алгоритма идентификации частот основного тона на сгенерированных синусоидальных сигналах.

Идея исходной математической модели слуховой системы человека состоит 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. В исходной модели эмпирическим путем был сформирован набор шаблонов для вычисления частоты основного тона. Для исследуемого диапазона от 70 до 400 Гц был получен список каналов определения частоты основного тона, по которым были синтезированы синусоидальные сигналы с заданной частотой звучания. Шаблон, включаемый в набор для определения заданного канала, определялся эмпирически на основании минимальной меры различия.

Модифицированная модель слуховой системы человека отличается учетом возможности задания границ определения частот основного тона. Данная модификация была осуществлена с целью автоматизации генерации наборов шаблонов для исследуемых частот. 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.

Основываясь на 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 (k0n – нижняя и k0v – верхняя);

2) формирование тестовых сигналов для создания шаблонов;

3) одновременная маскировка (вычисление массива результата маскировки P0[kt,k]);

4) вычисление массива номеров первых каналов свертки N1[kt], массива количества каналов в шаблоне Nk[kt], массива набора шаблонов Tpl[kt,k].

Таким образом, генерация шаблонов происходит по упрощенной схеме, представленной на рисунке 1.

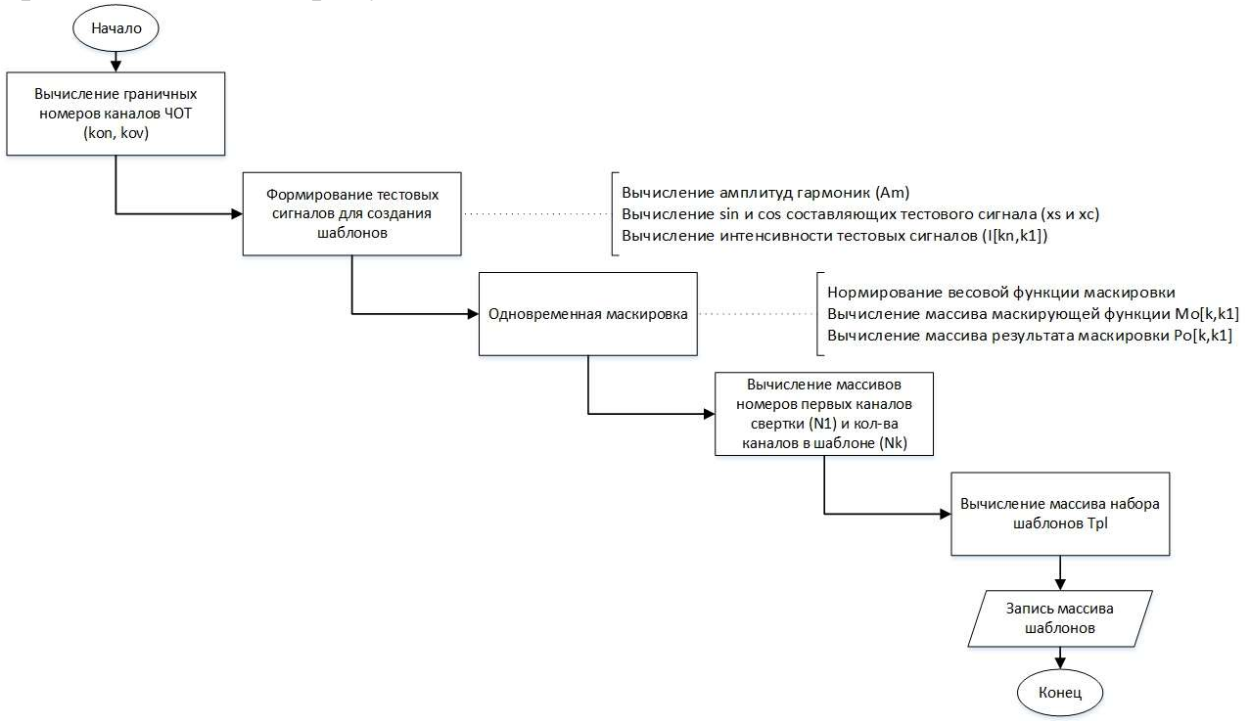

Рисунок 1 – Блок-схема алгоритма генерации шаблонов

Шаг квантования по времени в секундах определяется по формуле 1.

$$T_k = \frac{1}{F_k}, \quad (1)$$

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 ЧОТ для формирования масок;

F_{0v}– верхняя ЧОТ для формирования масок.

Для расчета параметров системы фильтра для анализа частот основного тона сигнала в первую очередь требуется вычислить коэффициенты для расчета шкал резонансных частот (формула 2) и добротностей (формула 3).

$$dx = \frac{1}{K} \cdot \ln\left(\frac{F_0}{F_n}\right), \quad (2)$$

F_0 –верхняя частота системы фильтров в Гц;

F_n – нижняя частота системы фильтров в Гц.

$$c = \frac{1}{K} \cdot \ln\left(\frac{Q_0}{Q_n}\right), \quad (3)$$

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 4 вычисляется шкала частот, а по формуле 5 шкала добротностей.

$$F_k = \frac{F_0}{\exp(k \cdot dx)}, \quad (4)$$

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

$$Q_k = Q_0 \cdot \exp(-c \cdot k) \quad (5)$$

На основании полученных данных определяется число коэффициентов системы фильтров для диапазона частот основного тона по формуле 6.

$$I_0(k) = \left\lceil \frac{2.4 \cdot Q_k}{\omega_0 \cdot \exp(-k \cdot dx) \cdot T_k} \cdot \ln\left(\frac{1}{\alpha}\right) \right\rceil, \quad (6)$$

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

Исходная математическая модель ограничивалась формулами 1-6. Отталкиваясь от формул выше, было определено, что вычисление значений номеров каналов, соответствующих нижней и верхней границам определения частоты основного тона, k_{0n} и k_{0v} осуществляется по формулам 7 и 8.

$$k_{0n} = \left\lceil \frac{1}{dx} \cdot \ln\left(\frac{F_0}{F_{0n}}\right) \right\rceil \quad (7)$$

$$k_{0v} = \left\lfloor \frac{1}{dx} \cdot \ln\left(\frac{F_0}{F_{0v}}\right) \right\rfloor \quad (8)$$

Таким образов, возможность указания граничных частот анализа позволяет исследовать модель на частотах, выходящих за пределы, определенные для речевых сигналов. Анализ результатов работы алгоритма идентификации частот при разных значениях номеров каналов основного тона позволяет определить оптимальные параметры для создания шаблонов.

Алгоритм создания шаблонов позволяет на основании указанных границ определения частот основного тона получить набор шаблонов для заданного диапазона. Для вычислений требуется определить количество масок для проведения анализа (формула 9) и получить шкалу частот для исследуемого диапазона (формула 10).

$$k_m = R \cdot (k_{0n} - k_{0v}), \quad (9)$$

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

$$F_{k1} = \frac{F_0}{exp\left(\frac{k1 \cdot dx}{R} + k_{0v} \cdot dx\right)}, \quad (10)$$

где $k1$ – 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 11.

$$A_m = \frac{1}{1 + a \cdot m}, \quad (11)$$

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 12 в рад/с и по формуле 13 в Гц.

$$\omega_{m,\ k1} = 2 \cdot \pi \cdot (m + 1) \cdot F_{k1} \quad (12)$$

$$Fx_{m,\ k1} = F_{k1n} \cdot (m + 1) \quad (13)$$

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.

$$xs(k, k1) = \sum_m A_m \cdot e^{-1.44 \cdot (Q_k)^2 \left(1 - \frac{Fx_{m,\ k1}}{F_{n0}} \cdot e^{k \cdot dx}\right)^2} \cdot \sin\left(\omega_{m,\ k1n} \cdot T_k \cdot I\right), \quad (14)$$

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

$$xc(k, k1) = \sum_m A_m \cdot e^{-1.44 \cdot (Q_k)^2 \left(1 - \frac{Fx_{m,\ k1}}{F_{n0}} \cdot e^{k \cdot dx}\right)^2} \cdot \cos\left(\omega_{m,\ k1} \cdot T_k \cdot I\right) \quad (15)$$

В таком случае, реакция системы фильтров на тестовые сигналы будет определяться формулой 16.

$$I_{kn,\ k1} = (xs(k, k1))^2 + (xc(k, k1))^2 \quad (16)$$

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

$$H_0(k, n) = e^{-2.88 \cdot \delta n \cdot (Q_k)^2 \left(1 - e^{(k-n) \cdot dx}\right)^2}, \quad (17)$$

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 18 и 19.

$$B(n) = \sum_{kn} H_0(k,n) \tag{18}$$

$$W_{k,\ n} = \frac{H_0(k,n)}{B(n)} \tag{19}$$

В результате, функция одновременной маскировки принимает вид, представленный формулой 20. Результат маскирования получается с применением формулы 21.

$$M_{0_{k,\ k1}} = \sum_{n=0}^{k} I_{n,\ k1} \cdot W_{k,\ n} \tag{20}$$

$$P_{0_{k,\ k1}} = \begin{cases} 1, \text{если } I_{k,\ k1} - M_{0_{k,\ k1}} > 0 \\ 0, \text{если } I_{k,\ k1} - M_{0_{k,\ k1}} \le 0 \end{cases} \tag{21}$$

Дальше, для получения набора шаблонов требуется провести вычисление массивов номеров первых каналов свертки $N_1[k_t]$, при $0 \le k_t \le k_{0n}-k_{0v}$ и количества каналов в шаблоне $N_k[k_t]$, при $0 \le k_t \le k_{0n}-k_{0v}$ по схеме на рисунке 2.

Вычисление массива набора шаблонов осуществляется при $0 \le k_t \le k_{0n}-k_{ov}$; $0 \le k \le N_k[k_t]-1$ по формуле 22. После этого осуществляется запись массива набора шаблонов Tpl.

$$Tpl = P_{0_{k,\ N_1[k]}} \tag{22}$$

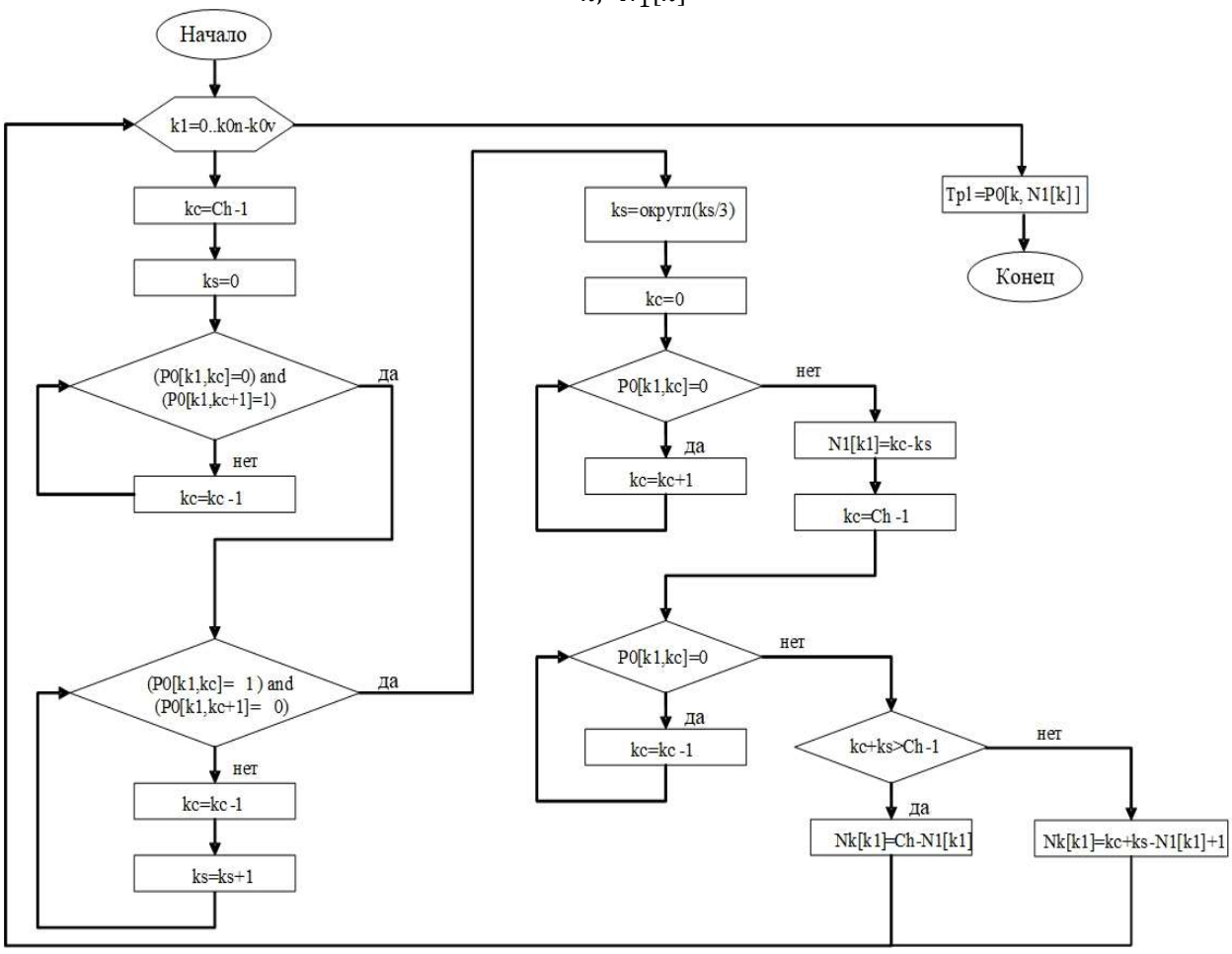

Рисунок 2 – Алгоритм создания шаблонов

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

Полученные шаблоны в рамках алгоритма вычисления частоты основного тона были протестированы на сгенерированных синусоидальных сигналах. Полученные результаты по идентификации частот показали высокую точность в диапазоне от 70 до 800 Гц включительно. Относительная ошибка алгоритма идентификации ЧОТ составила менее 1% (таблица 1), что позволяет применять модифицированную математическую модель слуховой системы человека не только для анализа параметров речевого сигнала, но и для идентификации нот.

Таблица 1 – 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

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 базовой модели. 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.

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.

Основная идея ***алгоритма распознавания 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.

Для определения корректных границ звучания нот на основании дискретных значений из теории музыки было проведено сравнение методов усреднения данных. В качестве метода усреднения был выбран метод, приближенный к определению значения частоты, на которой звучит нота. Кроме применяемого метода были рассмотрены: 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 результате нижняя и верхняя границы ноты определялись по формулам 22 и 23 соответственно.

$$f_{i\text{н}} = \frac{f_{i-1}+f_i}{2} \tag{22}$$

$$f_{i\text{в}} = \frac{f_{i+1}+f_i}{2} \tag{23}$$

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

Таблица 2 – 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%. Были получены

идентичные результаты распознавания нот вне зависимости от выбора шкалы. Причиной для подобного результата может служить то, что среди протестированных 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

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

Работа алгоритма на этапе идентификации частот основного тона осуществляется по следующей стратегии: для каждого дискретного момента времени, анализируется частотная область, определенная двумя гармониками речевого сигнала, интерпретированными в виде непрерывных интервалов единиц некоторой длины, между которыми находится интервал с нолями. Генерируется набор шаблонов, содержащих в себе первую и вторую гармоники основного тона, с которым будет сравниваться структура сигнала. После фильтрации сигнал проходит через свёртку с частотной маской.

Алгоритм сегментации включает в себя два этапов:

1) определение вокализованности текущего временного отсчета;
2) сегментация речевого сигнала на вокализованные и невокализованные участки.

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.

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.

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 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.

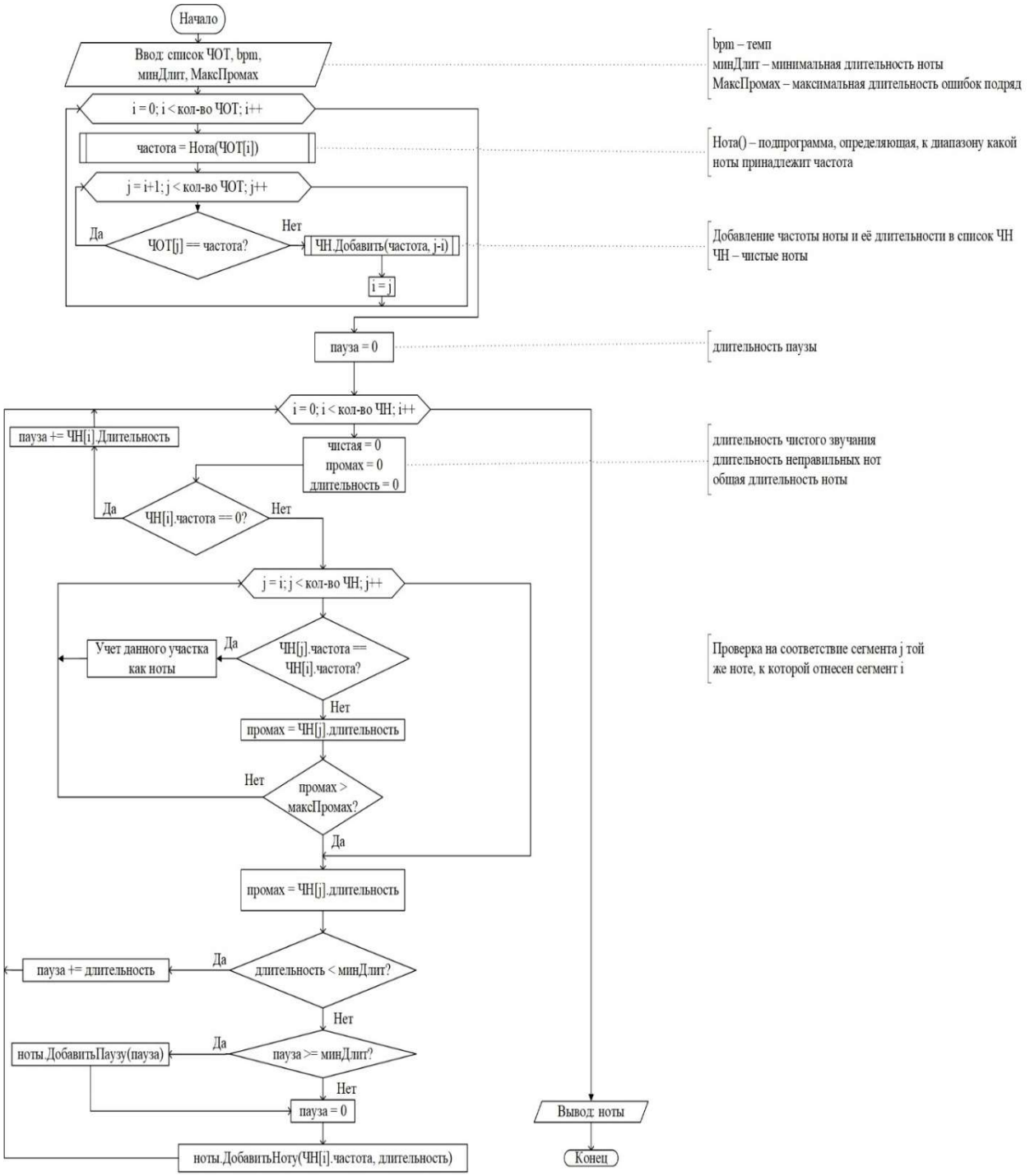

Рисунок 3 – Алгоритм сегментации и идентификации нот

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 и с результатами ручной обработки записей. Оценка коэффициента конкордации показала удовлетворительную согласованность экспертов.

В четвертой главе содержится описание разработанного программного комплекса. Структура разработанного программного комплекса на уровне блоков представлена на рисунке 4.

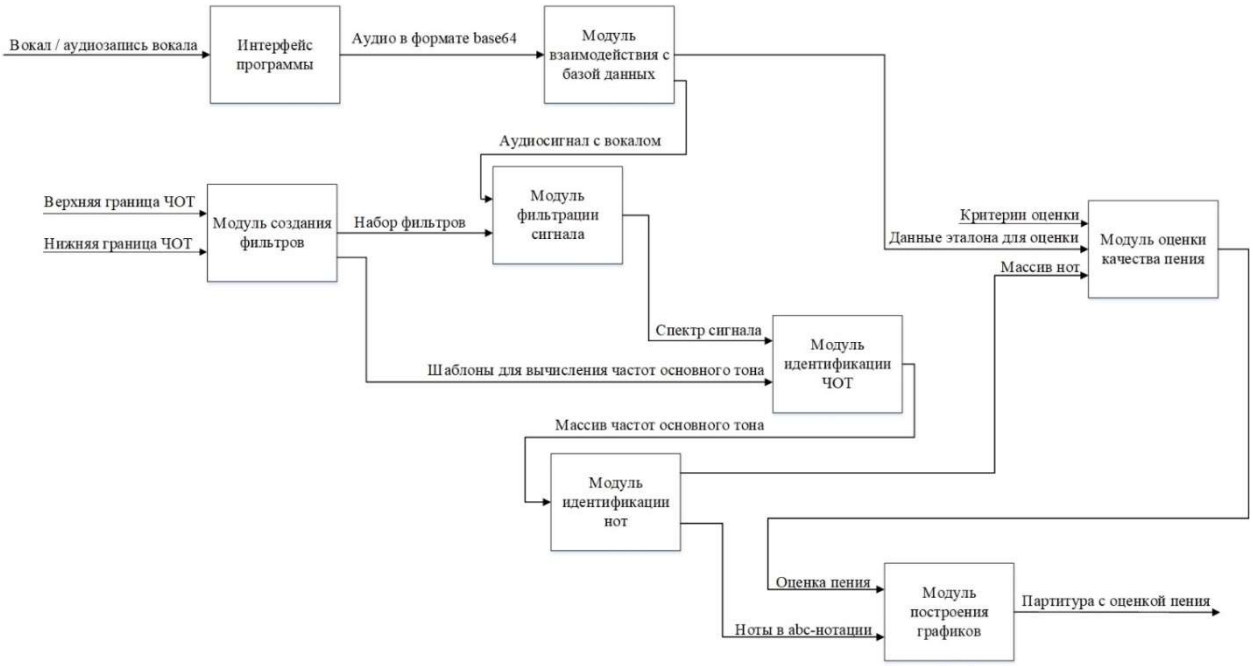

Рисунок 4 – Структура программного комплекса

По мере того, как пользователь записывает аудиофайл в клиентской части приложения, происходит процесс преобразования 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.

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.

Для пения по заданному звуку было решено применить аудиозапись с пением человека, получившим музыкальное образование по вокалу. Запись данного человека выступала в роли эталонной. С данной записью сравнивались записи учеников, выполнявших аналогичное задание. Для определения сходства вокальных исполнений был применен метод выделения синхронности.

Приводятся результаты тестирования работы комплекса на аудиозаписях с различными подходами к вокальному исполнению. 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.

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 программного комплекса была проведена экспертная оценка набор записей для 1-го упражнения. Всего было оценено 53 записи, содержащие 477 нот. Верхняя граница доверительного интервала для неизвестной частоты наступления события определяется по формуле 24.

$$P_{\text{в}} = \frac{n}{t^2+n}\left[\omega + \frac{t^2}{2n} + t\sqrt{\frac{\omega(1-\omega)}{n} + \left(\frac{t}{2n}\right)^2}\right] \quad (24)$$

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 формуле 24 принимает вид:

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

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

В заключении приведены основные результаты и выводы по проделанной работе.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

В результате выполненного исследования в области разработки моделей классификации угроз нарушения информационной безопасности была решена научно-техническая задача повышения качества распознавания звучащих нот в вокальном исполнении за счёт применения модели слуховой системы человека. Поставленная в начале исследования цель достигнута. Получены следующие основные результаты:

1) выполнен анализ текущего состояния предметной области: методов и алгоритмов распознавания нот, в том числе определения частот основного тона. Было определено, что существующие алгоритмы анализа частоты основного тона неприменимы к вокальным исполнениям по 2 причинам: высокий процент грубых ошибок и узкая полоса исследования данных;

2) проведена модификация модели слуховой системы человека на предмет увеличения диапазона анализа частот основного тона. 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 нот составила менее 1%;

3) сформирована методика распознавания нот в вокальном исполнении, включающая в себя вычисление ЧОТ, на основании которых происходит сегментация и идентификация нот;

4) 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. 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;

5) разработан программный комплекс, способный работать с аудиозаписями вокальных исполнений. Комплекс был протестирован на аудиозаписях с различными подходами к вокальному исполнению (такими как стаккато, легато, арпеджио, крещендо, декрещендо, восходящее и нисходящее

глиссандо, вибрато). В ходе эксперимента была определена точность распознавания нот – более 95%.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

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 ПУБЛИКАЦИЙ ПО ТЕМЕ РАБОТЫ

Публикации в ведущих рецензируемых журналах, рекомендованных ВАК для публикации результатов кандидатских и докторских диссертационных работ по специальности 05.13.17 – «Теоретические основы информатики»:

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

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

3. Катаева Е.С., **Якимук А.Ю.** Применение метода выделения синхронности при оценке сходства вокальных исполнений // Доклады Томского государственного университета систем управления и радиоэлектроники. – 2019. – Т. 22. – №3.

Публикации в ведущих рецензируемых журналах, рекомендованных ВАК для публикации результатов кандидатских и докторских диссертационных работ по другим специальностям:

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

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

Публикации в научных изданиях, индексируемых Scopus:

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

7. **Yakimuk A.Yu.**, Konev A.A., Andreeva Yu.V., Nemirovich-Danchenko M.M. Applying the principle of distribution in the program complex for vocal recognition // IOP Conf. Series: Materials Science and Engineering. – 2019. – Vol. 597. – No. 1. – P. 012072. – DOI:10.1088/1757-899X/597/1/012072

Публикации в тезисах и материалах научных конференций:

8. **Якимук А.Ю.**, Конев А.А. Программное обеспечение для автоматического распознавания мелодии // 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.

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

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

11. **Якимук А.Ю.** Алгоритмы анализа частоты основного тона вокального исполнения // 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.

12. **Якимук А.Ю.**, Егошин Н.С., Осипов А.О., Боков И.М. Повышение качества идентификации нот в автоматизированной системе распознавания вокала // Электронные средства и системы управления: Материалы докладов XII Международной научно-практической конференции – Томск: В-Спектр, 2016. – С. 29-32.

13. **Якимук А.Ю.** Генерация фильтров для одновременной маскировки // 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.

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

15. **Якимук А.Ю.** Этапы работы программного комплекса, определяющего ноты вокального исполнения // 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.

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

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

18. **Якимук А.Ю.**, Холопов М.Д. Распределенный программный комплекс по распознаванию нот // 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.

19. **Якимук А.Ю.** Влияние вибрато на качество распознавания нот в вокальном исполнении // VII молодежная научная конференция "Математическое и программное обеспечение информационных, технических и экономических систем"

Свидетельства о государственной регистрации программы для ЭВМ:

20. 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 г.

21. 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 г.

Якимук Алексей Юрьевич

Алгоритмы анализа частоты основного тона вокального исполнения

Автореф. дис. на соискание ученой степени канд. техн. наук

Тираж 100 экз. Заказ ___.

Томский государственный университет
систем управления и радиоэлектроники

634050, г. Томск, пр. Ленина, 40.
Тел. (3822) 53-30-18.