

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
**«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»**
(Н И У « Б е л Г У »)

ИНСТИТУТ ИНЖЕНЕРНЫХ ТЕХНОЛОГИЙ И ЕСТЕСТВЕННЫХ НАУК
КАФЕДРА ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫХ
СИСТЕМ И ТЕХНОЛОГИЙ

**ИССЛЕДОВАНИЕ МЕТОДОВ РАСПОЗНАВАНИЯ ЗВУКОВ РУССКОЙ
РЕЧИ**

Магистерская диссертация
обучающегося по направлению подготовки 11.04.02
Инфокоммуникационные технологии и системы связи,
магистерская программа «Системы и устройства радиотехники и связи»
очной формы обучения, группы 07001532
Нагибина Евгения Юрьевича

Научный руководитель
канд. техн. наук,
доцент кафедры
Информационно-
телекоммуникационных
систем и технологий
НИУ «БелГУ» Прохоренко Е.И.

Рецензент
Ведущий инженер электросвязи
участка систем коммутации № 1
г. Белгорода Белгородского
филиала ПАО «Ростелеком»
Уманец С.В.

БЕЛГОРОД 2017

ОГЛАВЛЕНИЕ	
ВВЕДЕНИЕ	3
ГЛАВА 1 ОБЗОР МЕТОДОВ РАСПОЗНАВАНИЯ РЕЧИ	6
1.1 Основные аспекты обработки речевых сигналов	6
1.2 Классификация методов распознавания речи	10
1.3 Анализ существующих методов распознавания речи	12
1.4 Исследование существующих систем распознавания речи	16
ГЛАВА 2 ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК РЕЧЕВЫХ СИГНАЛОВ ПОРОЖДЕННЫХ РАЗЛИЧНЫМИ ЗВУКАМИ РУССКОЙ РЕЧИ	19
2.1 Исследование временных характеристик речевых сигналов	19
2.2 Исследование мел-кепстральных характеристик	34
2.3 Исследование частотных (субполосных) характеристик	42
2.4 Анализ возможности применения полученных параметров в системах распознавания	58
ЗАКЛЮЧЕНИЕ	67
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	69

ВВЕДЕНИЕ

Актуальность диссертационного исследования. В настоящее время для людей речь является важнейшим средством связи. Для выполнения эффективной передачи на большие расстояния, речь должна быть преобразована в непрерывный электрический сигнал, т.е. сигнал, у которого можно отсчитать бесконечно большое число значений в заданном интервале времени. Сегодня научные сообщества вкладывают огромные средства в развитие новшеств и научно-исследовательские разработки для разрешения проблемы, касающихся понимания речи и ее распознавания. Стимулом для этого служат практические требования, которые связаны с созданием военной системы и системы коммерческого назначения. Также усложняют ситуацию и такие факторы, как индивидуальные особенности пользователя. Потому что даже для четко произнесенных и аккуратно сформулированных слов и выражений вероятность правильного распознавания может быть невелика, и на выходе конечное распознавание получается с большим количеством словесных ошибок, при этом устройство коррекции ошибок во многих системах слабо отлажено. Спонтанная речь может сопровождаться грамматическими ошибками и так называемым речевым «бардаком». Влияние также могут оказывать наличие искажений и акустических помех, которые имеют свойство изменяться.

На данном этапе для упрощения понимания употребления малых многозначных слов проводится интенсивный поиск возможности решения задачи распознавания речи, так как существующие технологии пока не имеют достаточных возможностей для их повсеместного использования. Распознавание речи в настоящее время нашло настоящее применение в жизни, наверное, только в тех случаях, когда используемый словарь сокращен до 10 знаков, например в компьютерных системах, предназначенных для обработки номеров и других кодов доступа, обрабатывающих передаваемые данные по телефону. Именно поэтому

актуальная задача – распознавание хотя бы 20 тысяч слов естественного языка – остается пока недостижимой. Такие возможности пока недоступны для обширного коммерческого использования. Но, несмотря на это, некоторые компании пытаются применять использовать уже существующие в данной области науки знания. Сегодня повсеместному распространению голосового управления мешает недостаточная вычислительная мощность процессоров и проблема наличия посторонних (внешних) шумов.

Целью данной магистерской работы является разработка рекомендаций по использованию существующих методов обработки речевых данных в задачах распознавания звуков русской речи.

Для достижения поставленной цели необходимо решить следующие задачи:

- провести анализ и исследование существующих методов распознавания речи;
- провести исследование временных характеристик речевых сигналов;
- провести исследование мел-кепстральных характеристик речевого сигнала;
- провести исследование субполосных характеристик речевых сигналов;
- провести анализ возможности применения полученных параметров в системах распознавания;

Объектом исследования являются методы идентификации диктора по голосу, а **предметом исследования** – речевые сигналы, соответствующие звукам русской речи.

В процессе работы над научно-исследовательской работой были использованы методы:

субполосного анализа/синтеза
цифровой обработки речевых данных

Личный вклад соискателя. Все изложенные в диссертации результаты исследования получены либо соискателем лично, либо при его непосредственном участии.

Публикации. По теме диссертационного исследования опубликована 1 печатная работа.

Объем и структура работы. Диссертация состоит из Введения, двух глав и заключения. Работа изложена на 73 страницах машинописного текста, включая 35 рисунков, 22 таблицы и список литературных источников из наименований.

ГЛАВА 1 ОБЗОР МЕТОДОВ РАСПОЗНАВАНИЯ РЕЧИ

1.1 Основные аспекты обработки речевых сигналов

Речь – форма общения людей, сложившаяся в ходе материальной преобразующей деятельности. Речь - это язык в действии. В ходе речевого общения происходит непрерывное кодирование и декодирование информации. Речь может служить как для целей регулирования личной деятельности (внутренняя речь), так и для процессов восприятия сообщений, предназначенных непосредственно для общения.

Основные виды речи человека:

• устная; • письменная; • диалогическая; • монологическая; • внешняя; • внутренняя.

При устной речи люди непосредственно общаются друг с другом, произносят различные звуки, которые относятся к тем или иным явлениям или предметам. Наборы звуков передаются и воспринимаются на слух через соответствующие колебания давления воздуха.

Письменная речь - речь, изложенная на бумаге или на какой-нибудь другой поверхности: пергаменте, баннере, полотне и т. п.) с помощью специальных графических знаков (знаков письменности). Использование П. р. позволяет более целостно структурировать свою речь, постепенно выстраивая и дополняя, а также исправляя ее в случае необходимости, что способствует в конечном счете выработке и применению более сложных синтаксических конструкций, чем это свойственно устной речи.

Диалогический тип речи может быть осуществлен при участии как минимум двух людей. Каждая фраза человека адресована одному или нескольким людям, а их собственные высказывания являются ответом на реплику данного человека. Монологическая – речь, которая произнесена от начала и до конца только одним человеком. Диалогу свойственно прерывания речи одного человека репликами других людей. Монолог же не

предполагает какой-то речевой реакции и должен быть понятным сам по себе. И диалог и монолог могут быть как, устными, так и письменными.

Внешняя речь – система письменных знаков и символов для передачи информации, а также совокупность звуковых сигналов, используемых человеком, процесс материализации мысли. Внешней речи могут быть свойственны жаргон и интонация. Жаргон - стилистические особенности (лексические, фразеологические) языка узкой социальной или профессиональной группы людей. Интонация - совокупность элементов речи (мелодика, ритм, темп, интенсивность, акцентный строй, тембр и др.), которые организуют речь фонетически и являющихся средством выражения различных значений, их эмоции.

Распознавание речи – процесс преобразования речевого сигнала в цифровую информацию (например, текстовые данные). Обратной задачей является синтез речи. Сейчас можно выделить 4 сравнительно изолированных направления в области развития речевых технологий:

1. Распознавание речи – т.е. преобразование речевого акустического сигнала в цепочку символов, слов. Эти системы могут быть оценены по некоторым характеристикам. Первое, это объём словаря: малые объёмы до 20 слов, большие – тысячи и десятки тысяч. Количество дикторов: от одного до произвольного. Стил произнесения: от изолированных команд до слитной речи и от чтения до спонтанной речи. Оценкой качества работы систем распознавания речи обычно определяется надёжностью распознавания слов, или же, другими словами, процентом ошибок.

2. Определение индивидуальности говорящего. Эти системы можно поделить на два класса: верификация говорящего (т.е. подтверждение его личности) и идентификация говорящего (т.е. определение его личности из заранее ограниченного числа людей). Оба эти класса далее могут быть разделены на тексто - зависимые и тексто - независимые.

3. Синтез речи. Практически существует два класса:

а) воспроизведение записанного в той или иной форме ограниченного числа сообщений;

б) синтез речи по тексту. Синтезаторы характеризуются по следующим параметрам: разборчивость (словесная или слоговая), естественность звучания, помехоустойчивость.

4. Компрессия речи. Основной (и единственный) оценочно-показательный признак этих систем, это степень сжатия: от низкой (16-32 кбит/сек) до высокой (1200-2400 кбит/сек и ниже). Разборчивость сжатой (компрессированной) речи и есть оценка качества работы систем компрессии. В некоторых приложениях очень важными параметрами являются узнаваемость голоса говорящего и возможность определения стрессового уровня говорящего.

Важно также дать определения понятиям квантования, дискретизации и интерполяции.

Квантование по уровню заключается в замене непрерывного множества значений сигнала $s(t)$ множеством дискретных значений, а дискретизация по времени заключается в замене непрерывного сигнала $s(t)$ дискретным по времени сигналом, значения которого для фиксированного момента времени совпадают соответственно с мгновенными значениями непрерывного сигнала. На рисунке 1.1 представлен пример квантования сигнала по времени и по уровню, где $x_0(t)$ – исходный сигнал; $x(t)$ – результат квантования;

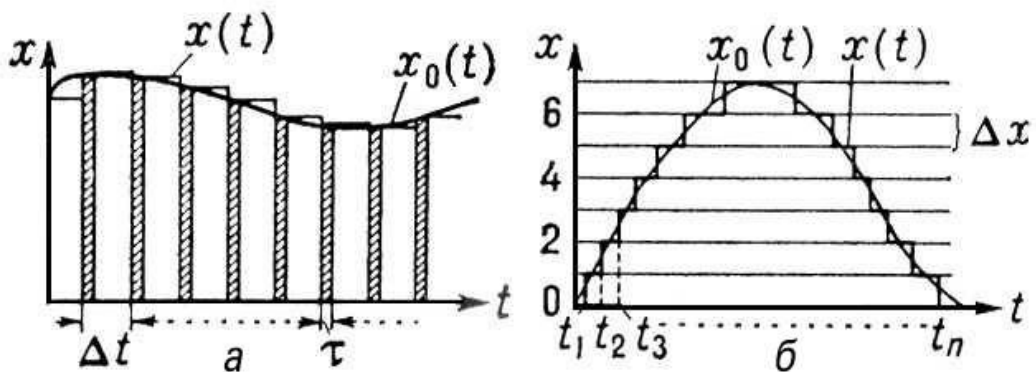


Рисунок 1.1 – Квантование сигнала: а) по времени; б) по уровню

Некоторые критерии выбора интервала временной дискретизации

- 1) Частотный критерий
- 2) Корреляционный

Интерполяция – восстановление функции по точкам, заданным в каких-то значениях.

Предварительная обработка речевого сигнала включает в себя следующие этапы: процесс ввода речевого сигнала; выделение границы речевого сигнала; цифровая фильтрация; нарезка речевого сигнала перекрывающимися кадрами; обработка сигнала в окне; спектральное преобразование; нормирование частотного спектра. Рассмотрим некоторые этапы.

Процесс ввода речевого сигнала.

Ввод звука осуществляется в реальном времени через файлы формата WAV или звуковую карту компьютера. В системах передачи информации частота дискретизации 8кГц и квантование 16 бит являются типовыми параметрами хранения и обработки речевой информации.

Выделение границы речевого сигнала

Для вычленения из входного сигнала участков, содержащих только речь, используются следующие характеристики речевого сигнала: кратковременная энергия речевого сигнала; число нулей интенсивности (мгновенная частота); плотность распределения значения отчетов паузы.

Цифровая фильтрация

Вместе с полезным сигналом обычно попадают различные шумы. Шум, несомненно, оказывает отрицательное влияние на качество работы систем распознавания речи, поэтому с ним необходимо бороться. Применение пропускающего полосового фильтра и предварительного фильтра позволяет снизить уровень шума в системе.

Нормирование частотного спектра.

Все вычисления производятся над числами с плавающей точкой. Поэтому значения параметров объектов, классифицируемых с помощью

нейронных сетей, ограничены диапазоном $[0.0, 1.0]$. Для выполнения обработки спектра нейронной сетью полученный спектр нормируется на 1.0. Для этого каждый компонент вектора делится на его максимальный компонент.

Основной идеей в задаче распознавания речи, является не полное восстановление лингвистического содержания высказывания, а определение смысла этого высказывания по так называемым ключевым словам из относительно малого словаря. Характерными параметрами для данного подхода являются:

- Конечный алфавит речевых сообщений.
- Распознавание высказывания, приводимое к выбору канонической фразы, при котором коэффициент ветвления равен 200.
- База распознавания – ограниченный набор ключевых слов.

1.2 Классификация методов распознавания речи

Системы распознавания речи классифицируются:

- по размеру словаря (ограниченный набор слов, словарь большого размера) – принцип в том, что чем больше размер словаря, который заложен в систему распознавания, тем больше частота ошибок при распознавании слов системой. Например, словарь из 10 цифр может быть распознан практически безошибочно, тогда как частота ошибок при распознавании словаря в 100000 слов может достигать 45%. С другой стороны, даже распознавание небольшого словаря может давать большое количество ошибок распознавания, если слова в этом словаре очень похожи друг на друга.

- по зависимости от диктора (дикторозависимые и дикторонезависимые системы) - по умолчанию, дикторозависимая система предназначена для использования одним пользователем, в то время как

дикторонезависимая система предназначена для работы с любым диктором. Дикторонезависимость – труднодостижимая цель, так как при обучении системы, она настраивается на параметры того диктора, на примере которого обучается. Частота ошибок распознавания таких систем обычно в 3-5 раз больше, чем частота ошибок дикторозависимых систем.

- по типу речи (слитная или раздельная речь) - если в речи каждое слово разделяется от другого участком тишины, то говорят, что эта речь – раздельная. Слитная речь – это естественно произнесенные предложения. Распознавание слитной речи намного труднее в связи с тем, что границы отдельных слов не четко определены и их произношение сильно искажено смазыванием произносимых звуков.

- по назначению (системы диктовки, командные системы) - назначение системы определяет требуемый уровень абстракции, на котором будет происходить распознавание произнесенной речи. В командной системе (например, голосовой набор в сотовом телефоне) скорее всего, распознавание слова или фразы будет происходить как распознавание единого речевого элемента. А система диктовки текста потребует большей точности распознавания и, скорее всего, при интерпретации произнесенной фразы будет полагаться не только на то, что было произнесено в текущий момент, но и на то, как оно соотносится с тем, что было произнесено до этого. Также, в системе должен быть встроен набор грамматических правил, которым должен удовлетворять произносимый и распознаваемый текст. Чем строже эти правила, тем проще реализовать систему распознавания и тем ограниченной будет набор предложений, которые она сможет распознать.

- по используемому алгоритму (нейронные сети, скрытые Марковские модели, динамическое программирование);

- по типу структурной единицы (фразы, слова, фонемы, дифоны, аллофоны);

- по принципу выделения структурных единиц (распознавание по шаблону, выделение лексических элементов) - в качестве акустической

модели выступает каким-либо образом сохраненный пример распознаваемой структурной единицы (слова, команды). Вариативность распознавания такой моделью достигается путем сохранения различных вариантов произношения одного и того же элемента (множество дикторов много раз повторяют одну и ту же команду). Используется, в основном, для распознавания слов как единого целого (командные системы).

1.3 Анализ существующих методов распознавания речи

Распознавание и порождение (синтез) речи компьютером является важной проблемой. В исследованиях важной задачей была находка упрощенного способа общения человека с компьютером. Вопрос не решен до сих пор, хотя в последнее время были достигнуты значительные успехи: уже многие годы голосовые команды являются одной из возможных опций программного обеспечения персональных компьютеров, появление функций распознавания речи уже обычное дело в ряде текстовых процессоров, системы распознавания речи работают там, где требуется оказание справочных услуг и в системах безопасности.

Методы цифровой обработки сигналов обычно осуществляют преобразование и очистку звукового сигнала в цифровой формат данных и другие представления, которые могут непосредственно обрабатываться системой распознавания речи. Эти задачи включают также фильтрацию шумовых сигналов, которые примешиваются к звуку при передаче акустических сигналов от воспринимающих устройств (микрофонов) или по сети. Методы же распознавания образов используют при выделении и распознавании отдельных слов или предложений речевого потока или в некоторых случаях для идентификации говорящего.

Кроме того, системы распознавания и синтеза речи затрагивают вопросы лингвистики, в которой заложены фундаментальные концепции и принципы распознавания речи и понимания языка.

Линейное предсказание является одним из наиболее лучших методов анализа речевых сигналов. Этот метод эффективен при оценке параметров основного тона, форманты и спектра. Также полезен для хранения и передачи сокращенной речи. Важность метода объясняется высокой точностью получаемых оценок и относительной простотой вычисления.

Основной принцип метода линейного предсказания состоит в том, что текущий отсчет речевого сигнала можно аппроксимировать линейной комбинацией предшествующих отсчетов. Коэффициент предсказания при этом определяется однозначно минимизацией среднего квадрата разности между отсчетами речевого сигнала и их предсказанными значениями (на конечном интервале). Коэффициенты предсказания - это весовые коэффициенты, используемые в линейной комбинации. Метод линейного предсказания можно применять для сокращения объема цифрового речевого сигнала.

Основной целью обработки речевых сигналов является получение наиболее удобного и компактного представления содержащейся в них информации. Точность представления определяется той информацией, которую необходимо сохранить или выделить. Например, цифровая обработка может применяться для выяснения, является ли данное колебание речевым сигналом. Сходная, но несколько более сложная задача состоит в том, чтобы классифицировать колебания на вокализованную речь, невокализованную речь и паузу (шум).

Одним из наиболее известных методов анализа речи во временной области можно назвать метод, предложенный Л.Рабинером и Р.Шафером. Он основан на измерении кратковременного среднего значения сигнала и кратковременной функции среднего числа переходов через нуль. Как отмечалось выше, амплитуда речевого сигнала существенно изменяется во

времени. Подобные изменения амплитуды хорошо описываются с помощью функции кратковременной энергии сигнала. В общем случае определить функцию энергии можно как

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 \quad (1.1)$$

Это выражение может быть переписано в виде

$$E_n = \sum_{m=-\infty}^{\infty} x^2(m)h(n-m) \quad , \quad (1.2)$$

где $h(n) = w^2(n)$

Выбор импульсной характеристики $h(n)$ или окна составляет основу описания сигнала с помощью функции энергии.

Скрытой Марковской моделью (СММ) называется модель, состоящая из N состояний, в каждом из которых некоторая система может принимать одно из M значений какого-либо параметра. Вероятности переходов между состояниями задается матрицей вероятностей $A=\{a_{ij}\}$, где a_{ij} – вероятность перехода из i -го в j -е состояние. Вероятности выпадения каждого из M значений параметра в каждом из N состояний задается вектором $B=\{b_j(k)\}$, где $b_j(k)$ – вероятность выпадения k -го значения параметра в j -м состоянии. Вероятность наступления начального состояния задается вектором $\pi=\{\pi_i\}$, где π_i – вероятность того, что в начальный момент система окажется в i -м состоянии.

Таким образом, скрытой Марковской моделью называется тройка $\lambda=\{A,B,\pi\}$. Использование скрытых Марковских моделей для распознавания речи основано на двух приближениях:

1) Речь может быть разбита на фрагменты, соответствующие состояниям в СММ, параметры речи в пределах каждого фрагмента считаются постоянными.

2) Вероятность каждого фрагмента зависит только от текущего состояния системы и не зависит от предыдущих состояний.

Для осуществления распознавания на основе скрытых моделей Маркова необходимо построить кодовую книгу, содержащую множество эталонных наборов для характерных признаков речи (например, коэффициентов линейного предсказания, распределения энергии по частотам и т.д.). Для этого записываются эталонные речевые фрагменты, разбиваются на элементарные составляющие (отрезки речи, в течение которых можно считать параметры речевого сигнала постоянными) и для каждого из них вычисляются значения характерных признаков. Одной элементарной составляющей будет соответствовать один набор признаков из множества наборов признаков словаря.

Фрагмент речи разбивается на отрезки, в течение которых параметры речи можно считать постоянными. Для каждого отрезка вычисляются характерные признаки и подбирается запись кодовой книги с наиболее подходящими характеристиками. Номера этих записей и образуют последовательность наблюдений $O=\{o_1, o_2, \dots, o_i\}$ для модели Маркова. Каждому слову словаря соответствует одна такая последовательность. Далее A – матрица вероятностей переходов из одного минимального отрезка речи (номера записи кодовой книги) в другой минимальный отрезок речи (номер записи кодовой книги). B – вероятности выпадения в каждом состоянии конкретного номера кодовой книги рисунок 1.2.

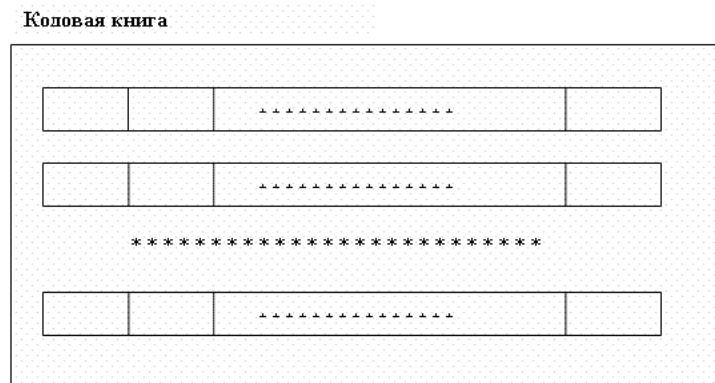


Рисунок 1.2 –Кодовая книга

1.4 Исследование существующих систем распознавания речи

В настоящее время речевое распознавание находит все новые и новые области применения, начиная от приложений, осуществляющих преобразование речевой информации в текст и заканчивая бортовыми устройствами управления автомобилем. Все системы распознавания речи можно разделить на два класса:

1) Системы, зависимые от диктора - настраиваются на речь диктора в процессе обучения. Для работы с другим диктором такие системы требуют полной перенастройки.

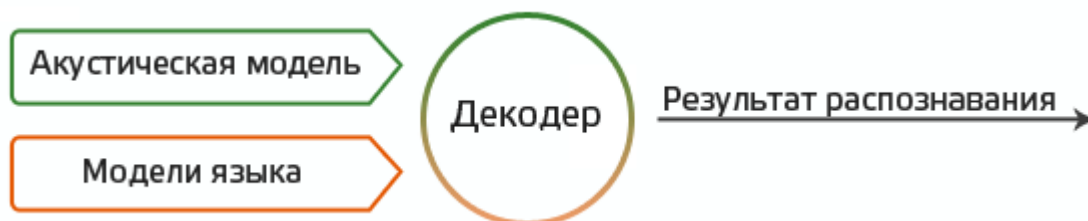


Рисунок 1.3 – Распознавание речи

2) Системы, не зависимые от диктора - работа которых не зависит от диктора. Такие системы не требуют предварительного обучения и способны распознавать речь любого диктора.

Для того чтобы понимать слитную речь, необходимо было перейти к словарям гораздо больших размеров, от нескольких десятков до сотен тысяч слов. Методы, использовавшиеся в системах первого вида, не подходили для решения этой задачи, так как просто невозможно создать эталоны для такого количества слов. Все многообразие существующих систем распознавания речи можно условно разделить на следующие группы:

1. Программные ядра для аппаратных реализаций систем распознавания речи.

2. Наборы библиотек, утилит для разработки приложений, использующих речевое распознавание.

3. Независимые пользовательские приложения, осуществляющие речевое управление и/или преобразование речи в текст.
4. Специализированные приложения, использующие распознавание речи.
5. Устройства, выполняющие распознавание на аппаратном уровне.
6. Теоретические исследования и разработки.

Ниже каждая из этих групп рассмотрена более подробно.

1. Программные ядра для аппаратных реализаций

Ядро программы – набор данных и правил, по которым осуществляется обработка данных. Ядро TTS (Text-to-Speech) engine предоставляет возможность синтеза речи по тексту, а ASR (Automatic Speech Recognition) engine – распознавания речи.

2. Наборы библиотек для разработки приложений

Siri (Speech Interpretation and Recognition Interface) – персональный помощник и вопросно-ответная система, разработанная для iOS. Данное приложение использует обработку естественной речи, чтобы отвечать на вопросы и давать рекомендации.

RealSpeaker – распознаёт 11 языков. Можно голосом печатать сообщения в социальных сетях, текстовых редакторов, word, Evernote, Skype, twitter. Программка в триал версии бесплатная. Про версия стоит 1000 руб. Хорошее решение. Нужен микрофон, видео камера и интернет.

3. Независимые пользовательские приложения

Dragon NaturallySpeaking Preferred фирмы Dragon Systems – единственная программа, приблизившаяся к тому, чтобы соответствовать заявленным характеристикам. В целом он очень близко подходит к достижению заявленной безошибочности распознавания - 95%. Обладает способностью с первого раза правильно записывать произнесенные слова.

4. Специализированные приложения

Распознавание речи может применяться для подразделений МВД, МЧС, МО, служб экстренной помощи, центров обработки вызовов и для других пользователей, в деятельности которых особое значение придается регистрации и обработке речевой информации.

5. Устройства, выполняющие распознавание на аппаратном уровне

Для использования функций речевого распознавания в различных устройствах, роботах, игрушках, разрабатываются аппаратные методы решения данной проблемы.

6. Теоретические исследования и разработки

Разработкой теоретической базы в области речевых технологий занимаются множество исследовательских групп по всему миру. В первую очередь это такие крупные корпорации как IBM, Intel, Microsoft, AT&T. Эти компании занимаются теорией распознавания уже не один десяток лет и являются законодателями в этой области.

ГЛАВА 2 ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК РЕЧЕВЫХ СИГНАЛОВ ПОРОЖДЕННЫЕ РАЗЛИЧНЫМИ ЗВУКАМИ РУССКОЙ РЕЧИ

2.1 Исследование временных характеристик речевых сигналов

Речевой сигнал имеет двойственную природу – с одной стороны, это обычный акустический сигнал, который представляет собой процесс распространения энергии акустических колебаний в упругой среде. Как любой акустический сигнал, он может быть представлен в виде звуковых волн, представляющих собой распространение процессов сжатия и разряжения частиц среды, формы фронтов которых зависят от свойств источника и условий распространения. Поэтому, как и другие акустические сигналы, речь характеризуется определенным набором объективных характеристик: зависимостью звукового давления от времени (временной структурой звуковой волны), длительностью звучания, спектральным составом, местом расположения источника в пространстве и пр.

Длительность - это продолжительность звука со свойственным ему количеством колебаний в единицу времени. Обычно время колебания измеряется в миллисекундах.

Различные звуки русской речи имеют различную длительность, что связано, не только со скоростью произнесения, но и типом звука. Таким образом, интерес представляет исследование длительностей фрагментов речевых сигналов, порожденных различными звуками русской речи. В рамках данной работы проведено исследование зависимости средней длительности фрагментов речевых сигналов от типа звука русской речи. Словесное описание вычислительного эксперимента представлено ниже.

Вычислительный эксперимент № 1

Цель: выявить зависимость длительности речевых сигналов (РС) от типа звуков русской речи (РР).

Исходные данные: речевой сигнал, соответствующий звукам русской речи, записанные с частотой дискретизации 16 кГц, разрядностью 16 бит. Речевой материал записан в одних и тех же акустических условиях при использовании одного и того же оборудования. Звуки РР выделены «на слух» из слитно записанных фраз.

План

1. Загрузить речевой сигнал, соответствующий определенному звуку РР.
2. Для каждого РС определить длительность фрагмента N .
3. Для каждого типа звука РР определить минимальное значение длительности РС N_{min} .
4. Для каждого типа звука РР определить максимальное значение длительности РС N_{max} .
5. Для каждого типа звука РР определить среднее значение длительности РС N_s .
6. Для каждого типа звука РР определить величину среднеквадратического отклонения (СКО) значений длительности РС от средней длительности РС, соответствующего данному звуку РР.

$$\sigma = \sqrt{M[X - M(X)]^2}$$

7. Для каждого звука РР построить гистограммы частотей длительностей РС.
8. Проанализировать полученные результаты
9. В соответствии с полученными результатами разбить все звуки РР на несколько подгрупп.

Результаты эксперимента

Максимальные, минимальные, средние значения длительности, а также значения СКО представлены в таблице 2.1. Подгруппы были сформированы согласно средним значениям длительности. Единица измерения – м/с.

Таблица 2.1 – Значения длительности

№ группы	№ подгруппы	Тип зв. РР	Min значение	Max значение	Среднее	СКО
1	1	п	12,56	79,75	35,62	528,66
	2	й	24,38	60,50	39,08	174,52
2	1	к	9,00	164,19	46,77	973,47
	2	р	18,31	129,50	50,91	575,18
	3	т	20,25	104,06	53,28	550,56
	4	и	22,69	132,38	57,24	612,86
	5	г	20,31	99,13	57,98	572,29
	6	э	22,56	128,56	58,45	522,44
	7	в	23,63	102,75	60,55	422,22
	8	л	13,06	226,00	65,25	1372,28
	9	у	22,25	117,19	68,86	559,91
3	1	а	26,50	157,06	71,13	752,90
	2	ж	31,44	139,31	73,14	739,13
	3	н	21,94	169,81	73,18	785,24
	4	ы	21,38	143,38	73,80	1146,68
	5	д	27,31	159,56	74,00	874,51
	6	е	33,63	142,94	74,04	760,98
	7	м	34,19	149,31	77,00	690,16
	8	б	24,75	145,00	79,19	1066,68
4	1	ю	74,31	95,69	83,23	82,42
	2	ф	42,75	122,06	86,53	633,54
	3	з	54,56	132,88	89,08	529,26
	4	х	65,56	110,56	90,01	168,90
	5	о	39,69	193,31	91,61	1383,41
	6	ё	74,75	134,75	95,29	778,92
	7	ч	56,19	133,06	96,09	539,09
5	1	ц	49,75	234,44	106,69	3088,26
	2	с	52,69	222,88	110,80	1562,61
	3	я	46,44	190,88	112,33	1772,20
	4	ш	90,50	162,56	122,94	460,19
	5	щ	72,56	176,81	137,96	967,36

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 5 подгрупп, в зависимости от значения параметра «длительность». Звуки «п», «й» имеют самые малые средние значения. Звуки «ц», «с», «я», «ш», «щ» входят в 5 подгруппу с самым большим числом средних значений.

Вычислительный эксперимент № 2

При анализе сигналов в дискретном времени часто исследуют количество переходов сигнала через ноль. Частота переходов сигнала через ноль может служить простейшей характеристикой спектральных свойств сигнала, хотя сама обработка производится во временной области. Это наиболее справедливо для узкополосных сигналов. Речевой сигнал является широкополосным и функция среднего числа переходов через ноль может быть грубой оценкой спектральных свойств, особенно на фоне шума. Кратковременная функция среднего числа переходов через ноль или нулевых пересечений основана на сравнении знаков соседних отсчетов и определяется.

$$Z_s = \frac{1}{L} \sum_{m=-\infty}^{\infty} \frac{|\operatorname{sgn}(S(m)) - \operatorname{sgn}(S(m-1))|}{2} w(n-m) \quad (2.1)$$

Где $S_{gn}(x)$ - знаковая функция, $S_{gn}(x) = 1$, если $(x) \geq 0$ и $S_{gn}(x) = -1$, если $(x) < 0$

Цель: выявить зависимость значений числа переходов через ноль через фиксированный интервал времени от типа звуков РР.

Исходные данные: речевой сигнал, соответствующий звукам русской речи, записанные с частотой дискретизации 16 кГц, разрядностью 16 бит, длительность отрезка анализа для определения числа переходов через ноль $T=16$ мс. ($N=256$). Речевой материал записан в одних и тех же акустических

условиях при использовании одного и того же оборудования. Звуки РР выделены «на слух» из слитно записанных фраз.

План

1. Загрузить речевой сигнал, соответствующий определенному звуку РР.
2. Каждый РС разбить на отрезки равной длительности.
3. Для каждого отрезка определить количество переходов через ноль.

$$Z_s = \frac{1}{L} \sum_{m=-\infty}^{\infty} \frac{|\operatorname{sgn}(S(m)) - \operatorname{sgn}(S(m-1))|}{2} w(n-m)$$

4. Для каждого типа звука РР определить минимальное значение числа переходов через ноль.
5. Для каждого типа звука РР определить максимальное значение числа переходов через ноль.
6. Для каждого типа звука РР определить среднее значение числа переходов через ноль.
7. Для каждого типа звука РР определить СКО значений числа переходов через ноль от среднего числа РС, соответствующий данному типу звуков РР.

$$\sigma = \sqrt{M[X - M(X)]^2}$$

8. Для каждого звука РР построить гистограммы частоты числа переходов через ноль.
9. Проанализировать полученные результаты.
10. В соответствии с полученными результатами разбить все звуки РР на несколько подгрупп.

Результаты эксперимента

Максимальные, минимальные, средние значения переходов через 0, а также значения СКО представлены в таблице 2.2. Подгруппы были сформированы согласно средним значениям. Единица измерения – м/с.

Таблица 2.2 – значения переходов через 0

№ группы	№ подгруппы	Тип зв. РР	Min значение	Max значение	Среднее	СКО
1	1	у	2,00	17,00	6,16	12,75
	2	ю	3,00	15,00	7,07	5,14
	3	б	3,00	15,00	8,49	22,50
	4	л	3,00	36,00	8,82	120,19
	5	в	5,00	38,00	9,35	126,67
	6	ы	5,00	18,00	9,58	25,81
	7	н	4,00	24,00	9,82	30,27
	8	о	3,00	26,00	10,63	92,55
	9	м	5,00	24,00	11,76	34,57
	10	э	3,00	23,00	11,81	107,52
	11	д	3,00	21,00	11,83	24,41
	12	ё	5,00	30,00	12,58	43,74
	13	р	6,00	29,00	12,88	82,23
	14	и	3,00	40,00	13,90	172,11
2	1	г	10,00	24,00	15,27	5,75
	2	а	6,00	33,00	16,10	133,25
	3	я	5,00	32,00	16,90	32,88
	4	е	8,00	38,00	18,31	419,54
	5	й	11,00	37,00	19,10	101,12
	6	к	4,00	103,00	21,23	3829,95
	7	п	8,00	38,00	22,28	35,67
	9	х	12,00	37,00	22,71	20,67
3	1	з	5,00	169,00	28,93	1585,59
	2	ф	15,00	53,00	37,84	86,09
4	1	ш	14,00	98,00	66,82	395,85
5	1	ч	43,00	113,00	87,43	945,30
	2	ж	17,00	131,00	90,54	5169,07
	3	щ	43,00	131,00	95,22	2136,38
6	1	т	43,00	164,00	107,49	1043,26
	2	ц	37,00	182,00	110,29	10150,04
	3	с	51,00	172,00	112,25	6577,12

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 6 подгрупп, в зависимости от значения

параметра «переходов через ноль» Звук «ш» (среднее значение 66.82 м/с) отнесен в отдельную подгруппу, так как разница по сравнению с ближайшим звуком составляет 28.98 м/с.

Вычислительный эксперимент № 3

Цель: выявить зависимость изменения энергии отрезков РС «внутри» звуков от типов РР.

Исходные данные: речевой сигнал, соответствующий звукам русской речи, записанные с частотой дискретизации 16 кГц, разрядностью 16 бит. Речевой материал записан в одних и тех же акустических условиях при использовании одного и того же оборудования. Звуки РР выделены «на слух» из слитно записанных фраз.

План

1. Загрузить речевой сигнал, соответствующий определенному звуку РР. (\vec{X})
2. Для каждого РС определить значение энергии ($E = \|\vec{X}\|^2$) и длительность фрагмента (N).
3. Для каждого РС определить долю энергии, сосредоточенной в первой $\frac{E_1}{E}$ и второй $\frac{E_2}{E}$ половине фрагмента.
4. Для каждого типа звука РР определить максимальное значение долей энергий в 1-ой и 2-ой половинах фрагментов и соотношение между м/ж их энергиями.
5. Для каждого типа звука РР определить среднее значение долей энергий в 1-ой и 2-ой половинах фрагментов и соотношение между м/ж их энергиями.
6. Для каждого типа звука РР определить СКО от соответствующих средних значений.

$$\sigma = \sqrt{M[X - M(X)]^2}$$

7. Для каждого РС определить значение энергии в начале (для отрезка(X_n)), середине (для отрезка(X_c)) и конце (для отрезка(X_k)) звука.
8. Для каждого РС определить долю энергии, сосредоточенной в начале $\frac{E_n}{E}$, середине $\frac{E_c}{E}$ и конце $\frac{E_k}{E}$ звука.
9. Для каждого РС определить соотношение м/ж энергией в начале, середине и конце звука $(\frac{E_c}{E_n}, \frac{E_n}{E_k}, \frac{E_c}{E_k})$
10. Для каждого типа звука РР определить минимальное значение долей энергии в начале, середине и конце звука, а также соотношение м/ж этими энергиями.
11. Для каждого типа звука РР определить максимальное значение долей энергии в начале, середине и конце звука, а также соотношение м/ж этими энергиями.
12. Для каждого типа звука РР определить среднее значение долей энергии в начале, середине и конце звука, а также соотношение м/ж этими энергиями.
13. Для каждого типа звука РР определить СКО долей энергии в начале, середине и конце звука от соответствующих средних значений

$$\sigma = \sqrt{M[X - M(X)]^2}$$

14. Проанализировать полученные результаты
15. В соответствии с полученными результатами разбить все звуки РР на несколько подгрупп.

Результаты эксперимента

В таблице 2.3 представлены значения отношения энергии первой половины звука ко второй. Подгруппы были сформированы согласно максимальным значениям. Единица измерения м/с.

Таблица 2.3 – Отношение энергий

№ группы	№ подгруппы	Тип зв. РР	Min значение	Max значение	Среднее	СКО
1	1	й	0,56	1,27	0,85	0,07
	2	ё	0,28	2,01	0,89	0,63
	3	щ	0,41	2,60	1,29	0,53
	4	н	0,25	2,70	0,96	0,37
	5	ю	1,45	2,92	2,27	0,38
2	1	е	0,33	3,49	1,17	0,28
	2	б	0,29	3,62	1,21	0,82
	3	ф	0,02	3,77	1,44	1,72
	4	я	0,20	3,80	1,08	0,62
	5	ы	0,68	3,87	1,63	0,78
	6	э	0,71	3,88	1,39	0,44
	7	ч	0,12	4,06	1,52	1,03
	8	х	1,35	4,10	2,43	0,93
	9	т	0,09	4,35	1,75	1,93
	10	ж	0,41	4,61	1,22	1,02
	11	п	0,02	4,98	1,20	1,57
3	1	у	0,25	5,41	1,48	0,99
	2	о	0,52	6,05	1,55	1,51
	3	з	0,36	6,26	1,52	1,89
	4	ш	0,79	6,64	2,60	3,44
	5	ц	0,64	7,03	2,87	4,19
	6	с	0,24	7,64	2,37	2,58
4	1	к	0,44	8,10	2,47	3,25
	2	м	0,18	8,64	1,09	2,01
	3	р	0,19	8,87	1,68	3,32
	4	г	0,56	10,05	2,57	6,01
5	1	а	0,54	16,07	1,62	3,91
	2	и	0,35	17,30	1,97	6,46
	3	л	0,24	17,47	2,26	7,89
	4	д	0,08	18,08	2,58	12,00
6	1	в	0,15	50,02	2,72	67,93

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 6 подгрупп, в зависимости от значения

параметра «отношение энергий». Звук «в» (максимальное значение 50.02 м/с) отнесен в отдельную подгруппу, так как разница по сравнению с ближайшим звуком составляет 31.94 м/с.

Таблица 2.4 показывает значения отношения энергии начала к концу. Подгруппы были сформированы согласно максимальным значениям. Единица измерения – м/с.

Таблица 2.4 - отношение начала к концу

№ группы	№ подгруппы	Тип зв. РР	Min значение	Max значение	Среднее	СКО
1	1	й	0,57	2,29	1,13	0,46
	2	ё	0,19	2,68	1,03	1,35
	3	щ	0,36	4,06	1,70	1,57
	4	б	0,16	4,20	1,19	1,09
	5	ю	1,54	4,52	3,08	1,49
	6	н	0,14	4,76	1,11	1,16
	7	я	0,12	5,01	1,24	1,40
	8	ж	0,34	5,37	1,28	1,56
	9	п	0,01	5,54	1,18	2,01
	10	е	0,26	5,67	1,30	0,82
2	1	т	0,02	6,20	1,91	3,34
	2	ф	0,02	7,32	2,18	6,47
	3	э	0,64	7,87	1,81	2,47
	4	х	1,58	7,87	3,57	3,70
	5	ч	0,05	7,90	1,95	4,12
	6	ы	0,52	8,03	2,30	4,91
	7	ц	0,46	10,21	4,24	10,73
	8	к	0,22	10,75	3,23	7,55
3	1	с	0,15	12,58	3,64	7,95
	2	о	0,30	12,69	2,17	7,62
	3	у	0,05	12,76	1,95	5,44
	4	м	0,11	12,99	1,29	4,86
	5	з	0,25	14,66	2,23	11,96
	6	г	0,52	14,91	3,68	15,20
	7	ш	0,80	15,02	4,64	23,75
4	1	р	0,16	21,11	2,42	14,74
	2	д	0,02	28,16	3,77	34,13
5	1	а	0,33	38,38	2,58	25,00
	2	л	0,12	39,55	3,72	38,87
	3	и	0,37	44,30	3,25	50,40
6	1	в	0,05	70,95	3,44	137,81

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 6 подгрупп, в зависимости от значения параметра «отношение начала к концу». Звук «в» (максимальное значение 70.95 м/с) отнесен в отдельную подгруппу, так как разница по сравнению с ближайшим звуком составляет 26.65 м/с.

Таблица 2.5 показывает значения отношения энергии середины к концу. Подгруппы были сформированы согласно максимальным значениям. Единица измерения – м/с.

Таблица 2.5 - отношение энергии середины к концу

№ группы	№ подгруппы	Тип зв. РР	Min значение	Max значение	Среднее	СКО
1	1	ж	0,52	1,45	0,88	0,08
	2	б	0,24	1,53	0,70	0,10
	3	ё	0,68	2,03	1,19	0,36
	4	й	0,72	2,69	1,31	0,65
	5	п	0,02	3,04	0,71	0,56
	6	ю	1,97	3,08	2,40	0,24
	7	е	0,56	3,31	1,47	0,38
	8	х	1,13	3,36	2,22	0,41
	9	я	0,64	3,72	1,50	0,65
2	1	н	0,45	4,07	1,11	0,47
	2	в	0,37	4,19	1,16	0,57
	3	з	0,22	4,30	1,15	1,14
	4	ф	0,01	4,52	1,31	1,49
	5	м	0,26	4,67	1,16	0,85
	6	щ	0,90	4,88	2,24	1,38
	7	г	0,48	5,56	1,59	1,58
	8	э	0,86	5,97	1,91	1,57
	9	у	0,53	6,14	1,88	1,50
3	1	р	0,13	7,55	1,29	2,07
	2	ч	0,69	7,71	3,79	4,13
	3	к	0,17	7,80	1,93	2,97
	4	д	0,14	8,17	1,40	2,19
	5	т	0,04	8,45	1,57	2,41
	6	о	0,86	8,60	2,09	2,73
	7	ц	0,82	8,60	2,92	5,15
	8	ш	1,05	8,93	3,32	4,24

Окончание таблицы 2.5

4	1	л	0,20	11,59	2,02	4,32
	2	ы	1,00	12,11	2,50	7,51
	3	и	0,12	13,92	2,41	4,69
	4	с	0,53	15,07	2,49	6,10
	5	а	0,27	17,08	2,23	5,62

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 4 подгруппы, в зависимости от значения параметра «отношение энергии середины к концу».

Таблица 2.6 показывает значения отношения энергии середины к началу. Подгруппы были сформированы согласно максимальным значениям. Единица измерения – м/с.

Таблица 2.6 - отношение энергии середины к началу

№ группы	№ подгруппы	Тип зв. РР	Min значение	Max значение	Среднее	СКО
1	1	з	0,29	1,26	0,76	0,10
	2	г	0,11	1,32	0,65	0,10
	3	х	0,25	1,39	0,76	0,11
	4	ю	0,62	1,40	0,90	0,13
	5	й	0,88	1,56	1,18	0,06
	6	ы	0,42	2,14	1,23	0,21
	7	ш	0,26	2,24	1,18	0,38
	8	е	0,58	2,27	1,29	0,17
	9	э	0,62	2,31	1,23	0,21
	10	б	0,31	2,84	1,02	0,68
	11	ж	0,12	2,85	1,05	0,38
	12	и	0,24	3,40	1,30	0,42
	13	ё	0,76	3,81	2,69	1,88
	14	щ	0,67	3,88	1,74	0,84
	15	а	0,16	3,93	1,29	0,34
	16	р	0,12	3,98	0,87	0,60
2	1	н	0,45	4,43	1,51	0,77
	2	ц	0,19	4,63	1,23	1,53

Окончание таблицы 2.6

2	3	с	0,12	4,64	1,06	1,08	
	4	м	0,36	4,81	1,57	0,99	
	5	л	0,15	4,90	1,03	0,63	
	6	я	0,74	5,97	2,06	2,29	
	7	о	0,46	6,24	1,42	0,86	
	8	т	0,23	6,49	1,54	2,29	
	9	к	0,15	7,37	1,06	1,76	
	10	ф	0,27	8,37	1,67	5,52	
	3	1	у	0,48	10,15	1,71	3,41
		2	в	0,04	10,35	1,57	3,43
3		ч	0,42	16,46	4,50	24,25	
4		п	0,11	21,88	2,47	22,39	
5		д	0,08	28,01	1,98	26,94	

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 3 подгруппы, в зависимости от значения параметра «отношение энергии середины к началу».

Таблица 2.7 показывает значения доли энергии в начале звука. Подгруппы были сформированы согласно максимальным значениям. Единица измерения – м/с.

Таблица 2.7 – доля начала

№ группы	№ подгруппы	Тип зв. РР	Min значение	Max значение	Среднее	СКО
1	1	й	0,23	0,38	0,31	0,00
	2	ё	0,10	0,47	0,23	0,03
2	1	я	0,06	0,52	0,28	0,02
	2	щ	0,15	0,52	0,31	0,01
	3	ю	0,33	0,53	0,46	0,01
	4	э	0,22	0,53	0,35	0,01
	5	н	0,09	0,57	0,29	0,02
	6	е	0,14	0,57	0,32	0,01
	7	б	0,10	0,62	0,35	0,03
	8	ы	0,20	0,64	0,36	0,01
	9	у	0,03	0,64	0,34	0,02
	10	ч	0,03	0,65	0,25	0,03
	11	о	0,09	0,65	0,34	0,01

Окончание таблицы 2.7

	12	т	0,02	0,65	0,34	0,04
	13	ф	0,02	0,66	0,33	0,06
	14	х	0,34	0,68	0,50	0,01
3	1	м	0,08	0,70	0,28	0,02
	2	з	0,17	0,73	0,40	0,02
	3	ш	0,25	0,75	0,43	0,02
	4	ц	0,13	0,75	0,47	0,04
	5	ж	0,16	0,77	0,35	0,02
	6	п	0,01	0,77	0,30	0,05
	7	к	0,08	0,78	0,45	0,04
	8	и	0,15	0,79	0,36	0,02
	9	р	0,10	0,82	0,39	0,03
	10	с	0,09	0,83	0,47	0,03
	11	л	0,07	0,84	0,41	0,03
	12	а	0,16	0,85	0,34	0,01
	13	г	0,24	0,85	0,49	0,03
	14	д	0,01	0,90	0,42	0,05
	15	в	0,03	0,95	0,34	0,03

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 3 подгруппы, в зависимости от значения параметра «доля начала».

Таблица 2.8 показывает значения доли энергии в середине звука. Подгруппы были сформированы согласно максимальным значениям. Единица измерения – м/с.

Таблица 2.8 - доля середины

№ группы	№ подгруппы	Тип зв. РР	Min значение	Max значение	Среднее	СКО
1	1	з	0,15	0,37	0,26	0,00
	2	б	0,15	0,37	0,25	0,01
	3	г	0,10	0,40	0,27	0,01
	4	ё	0,36	0,41	0,38	0,00
	5	д	0,07	0,44	0,26	0,01
	6	ж	0,09	0,44	0,30	0,01
	7	й	0,28	0,45	0,36	0,00
	8	л	0,13	0,46	0,32	0,00
	9	ю	0,32	0,46	0,38	0,00

Окончание таблицы 2.8

	10	х	0,17	0,47	0,34	0,01
	11	н	0,23	0,48	0,34	0,00
	12	ф	0,01	0,48	0,26	0,02
	13	в	0,04	0,48	0,32	0,01
2	1	е	0,29	0,50	0,38	0,00
	2	э	0,26	0,52	0,39	0,00
	3	я	0,31	0,54	0,40	0,00
	4	ш	0,19	0,56	0,41	0,01
	5	м	0,18	0,57	0,33	0,01
	6	ы	0,27	0,57	0,40	0,01
	7	у	0,25	0,57	0,39	0,01
	8	щ	0,31	0,58	0,44	0,01
	9	р	0,08	0,61	0,26	0,02
	10	и	0,08	0,62	0,40	0,01
	11	п	0,02	0,63	0,21	0,02
	12	ц	0,14	0,66	0,37	0,02
	13	о	0,30	0,66	0,40	0,01
	14	к	0,10	0,67	0,30	0,02
3	1	а	0,13	0,70	0,39	0,01
	2	с	0,10	0,76	0,34	0,02
	3	т	0,03	0,79	0,32	0,03
	4	ч	0,27	0,83	0,55	0,02

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 3 подгруппы, в зависимости от значения параметра «доля середины».

Таблица 2.9 показывает долю энергии в конце звука. Подгруппы были сформированы согласно максимальным значениям. Единица измерения – м/с.

Таблица 2.9 - доля конца

№ группы	№ подгруппы	Тип зв. РР	Min значение	Max значение	Среднее	СКО
1	1	ю	0,12	0,21	0,16	0,00
	2	х	0,08	0,23	0,16	0,00
	3	ц	0,06	0,30	0,17	0,01
2	1	ш	0,04	0,35	0,16	0,01
	2	ы	0,05	0,38	0,24	0,01
	3	э	0,07	0,38	0,26	0,01

Окончание таблицы 2.9

	4	й	0,17	0,41	0,33	0,01
	5	о	0,05	0,43	0,26	0,01
	6	щ	0,10	0,44	0,25	0,01
	7	г	0,06	0,45	0,25	0,02
3	1	ё	0,18	0,53	0,40	0,02
	2	ж	0,14	0,54	0,35	0,01
	3	е	0,10	0,55	0,30	0,01
	4	я	0,10	0,55	0,33	0,02
	5	ч	0,08	0,58	0,20	0,02
	6	с	0,05	0,60	0,19	0,02
	7	а	0,02	0,63	0,27	0,01
	8	н	0,10	0,63	0,37	0,02
	9	в	0,01	0,63	0,35	0,02
	10	у	0,05	0,63	0,27	0,02
	11	б	0,15	0,64	0,40	0,02
	12	к	0,06	0,64	0,24	0,02
	13	и	0,02	0,67	0,24	0,01
	14	з	0,05	0,68	0,34	0,02
	15	л	0,02	0,68	0,28	0,02
	16	м	0,05	0,71	0,39	0,03
	17	р	0,04	0,73	0,35	0,03
4	1	д	0,03	0,80	0,31	0,04
	2	т	0,09	0,90	0,34	0,05
	3	п	0,14	0,94	0,49	0,07
	4	ф	0,08	0,97	0,41	0,08

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 3 подгруппы, в зависимости от значения параметра «доля конца».

2.2 Исследование мел-кепстральных характеристик

Кепстр - энергетический спектр функции $\ln|S(\omega)|^2$, определяемый выражением

$$C_s(q) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \ln|S(\omega)|^2 e^{i\omega q} d\omega \quad (2.2)$$

где $S(\omega)$ - амплитудный спектр непрерывного сигнала. Поскольку $S^2(\omega)$ имеет смысл спектральной плотности энергии сигнала $s(t)$, то C_s

истолковывается как энергетический спектр функции $\ln [S(w)]^2$. Но из (2.2) очевидно, что аргумент q этого «спектра» имеет размерность времени, а не частоты. Этим и объясняется распространение термина «кепстр», который образован перестановкой букв в термине «спектр».

Другими словами, кепстр определяет последовательность коэффициентов разложения функции $\lg [\Phi (z)]$ в степенной ряд.

Логарифм спектра мощности колебания, содержащего отраженный сигнал, имеет аддитивную периодическую компоненту, созданную этим сигналом, и поэтому преобразование Фурье от логарифма спектра мощности имеет пик на месте, соответствующем задержке отраженного сигнала.

Вычислительный эксперимент 1

Понятия локального максимума и локального минимума объединяются общим термином локальный экстремум. Слово "локальный" для краткости часто опускают и говорят просто о максимумах и минимумах функции. Таким образом экстремум - максимальное или минимальное значение функции на заданном множестве.

Цель: Выявить зависимость количества локальных максимумов в кепстре от типа звуков русской речи.

Исходные данные: речевой сигнал, соответствующий звукам русской речи, записанные с частотой дискретизации 16 кГц, разрядностью 16 бит, длительность отрезка анализа для определения числа переходов через ноль $T=16$ мс. ($N=256$). Количество интервалов, на которые разбивается ось частот ($R=40$).

План

1. Загрузить речевой сигнал, соответствующий определенным звукам $PP(\vec{X})$
2. Каждый речевой сигнал разбить на отрезки одинаковой длительности $N(\vec{X}_N)$

3. Для каждого отрезка РС оценить распределение энергии в шкале мел, используя субполосный анализ по R интервалам (\vec{P}_k)
4. Для каждого отрезка РС оценить значения кепстральных коэффициентов, используя ДПФ. (\vec{K}_k)
5. Для каждого отрезка РС оценить число локальных максимумов в кепстре
6. Для каждого звука РР определить минимальное количество локальных максимумов.
7. Для каждого звука РР определить максимальное количество локальных максимумов.
8. Для каждого звука РР определить среднее количество локальных максимумов.
9. Для каждого звука РР определить СКО локальных максимумов.

$$\sigma = \sqrt{M[X - M(X)]^2}$$

10. Проанализировать полученные результаты
11. В соответствии с полученными результатами разбить все звуки РР на несколько подгрупп

Результаты исследований

В таблице 2.10 представлены максимальные, минимальные, средние значения, а также значения СКО количества максимумов. Подгруппы были сформированы согласно значениям СКО. Единица измерения – м/с.

Таблица 2.10 – количества максимумов

№ группы	№ подгруппы	Тип зв. РР	Min значение	Max значение	Среднее	СКО
1	1	я	0,00	13,00	1,30	15,21
	2	э	0,00	14,00	1,40	17,64
2	1	д	0,00	15,00	1,50	20,25
	2	с	0,00	15,00	1,50	20,25
	3	у	0,00	15,00	1,50	20,25
	4	ф	0,00	15,00	1,50	20,25
	5	ш	0,00	15,00	1,50	20,25
	6	а	0,00	16,00	1,60	23,04
	7	в	0,00	16,00	1,60	23,04
	8	е	0,00	16,00	1,60	23,04
	9	м	0,00	16,00	1,60	23,04
	10	н	0,00	16,00	1,60	23,04
	11	ц	0,00	16,00	1,60	23,04
	12	ч	0,00	16,00	1,60	23,04
	13	щ	0,00	16,00	1,60	23,04
3	1	п	0,00	16,00	1,78	25,28
	2	г	0,00	17,00	1,70	26,01
	3	и	0,00	17,00	1,70	26,01
	4	к	0,00	17,00	1,70	26,01
	5	л	0,00	17,00	1,70	26,01
	6	о	0,00	17,00	1,70	26,01
	7	х	0,00	17,00	1,70	26,01
	8	й	0,00	12,00	3,00	27,00
	9	б	0,00	18,00	1,80	29,16
	10	ж	0,00	18,00	1,80	29,16
	11	з	0,00	18,00	1,80	29,16
	12	т	0,00	18,00	1,80	29,16
	13	ы	0,00	18,00	1,80	29,16
	14	р	0,00	19,00	1,90	32,49
4	1	ё	0,00	16,00	5,33	56,89
	2	ю	0,00	16,00	5,33	56,89

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 4 подгруппы, в зависимости от значения параметра «количество максимумов». Звуки «ё», «ю» имеют наибольшее значение СКО.

Вычислительный эксперимент № 2

Цель: выявить номера максимальных элементов кепстра от типа звуков РР.

Исходные данные: речевой сигнал, соответствующий звукам русской речи, записанные с частотой дискретизации 16 кГц, разрядностью 16 бит, длительность отрезка анализа для определения числа переходов через ноль $T=16$ мс. ($N=256$). Количество интервалов, на которые разбивается ось частот ($R=40$).

План

1. Загрузить речевой сигнал, соответствующий определенным звукам РР (\vec{X})
2. Каждый речевой сигнал разбить на отрезки одинаковой длительности N (\vec{X}_N)
3. Для каждого отрезка РС оценить распределение энергии в шкале мел, используя субполосный анализ по R интервалам (\vec{P}_k)
4. Для каждого отрезка РС оценить значения кепстральных коэффициентов, используя ДПФ. (\vec{K}_k)
5. Для каждого отрезка РС оценить номер с максимальным значением кепстра.
6. Для каждого отрезка РС определить минимальные значения номера с максимальным значением кепстра.
7. Для каждого отрезка РС определить максимальные значения номера с максимальным значением кепстра.
8. Для каждого отрезка РС определить средние значения номера с максимальным значением кепстра.
9. Для каждого отрезка РС определить СКО номера с максимальным значением кепстра.

$$\sigma = \sqrt{M[X - M(X)]^2}$$

10. Проанализировать полученные результаты.

11. В соответствии с полученными результатами разбить все звуки РР на несколько подгрупп.

Результаты исследований

В таблице 2.11 представлены максимальные, минимальные, средние значения, а также значения СКО номера максимальных элементов в кепстре. Подгруппы были сформированы согласно значениям СКО.

Таблица 2.11 – Номера максимальных элементов

№ группы	№ подгруппы	Тип зв. РР	Min значение	Max значение	Среднее	СКО
1	1	а	0,00	1,00	0,10	0,09
	2	б	0,00	1,00	0,10	0,09
	3	в	0,00	1,00	0,10	0,09
	4	д	0,00	1,00	0,10	0,09
	5	е	0,00	1,00	0,10	0,09
	6	ж	0,00	1,00	0,10	0,09
	7	з	0,00	1,00	0,10	0,09
	8	и	0,00	1,00	0,10	0,09
	9	й	0,00	1,00	0,10	0,09
	10	к	0,00	1,00	0,10	0,09
	11	л	0,00	1,00	0,10	0,09
	12	н	0,00	1,00	0,10	0,09
	13	о	0,00	1,00	0,10	0,09
	14	п	0,00	1,00	0,10	0,09
	15	р	0,00	1,00	0,10	0,09
	16	с	0,00	1,00	0,10	0,09
	17	т	0,00	1,00	0,10	0,09
	18	у	0,00	1,00	0,10	0,09
	19	ф	0,00	1,00	0,10	0,09
	20	х	0,00	1,00	0,10	0,09
	21	ц	0,00	1,00	0,10	0,09
	22	ч	0,00	1,00	0,10	0,09
	23	щ	0,00	1,00	0,10	0,09
	24	ы	0,00	1,00	0,10	0,09
	25	э	0,00	1,00	0,10	0,09

Окончание таблицы 2.11

	26	ю	0,00	1,00	0,10	0,09
	27	я	0,00	1,00	0,10	0,09
2	28	г	0,00	1,00	0,11	0,10
	1	м	0,00	1,00	0,25	0,19
	2	ё	0,00	1,00	0,33	0,22
	3	ш	0,00	1,00	0,33	0,22

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 2 подгруппы, в зависимости от значения параметра «номера максимальных элементов». Практически все значения идентичны, за исключением звуков «м» «ё» «ш», которые и определены в отдельную подгруппу.

Вычислительный эксперимент № 3

Цель: Выявить зависимость доли максимальных значений кепстра от типа звуков РР.

Исходные данные: речевой сигнал, соответствующий звукам русской речи, записанные с частотой дискретизации 16 кГц, разрядностью 16 бит, длительность отрезка анализа для определения числа переходов через ноль $T=16$ мс. ($N=256$). Количество интервалов, на которые разбивается ось частот ($R=40$).

План

1. Загрузить речевой сигнал, соответствующий определенным звукам РР (\vec{X})
2. Каждый речевой сигнал разбить на отрезки одинаковой длительности N (\vec{X}_N)
3. Для каждого отрезка РС оценить распределение энергии в шкале мел, используя субполосный анализ по R интервалам (\vec{P}_K)

4. Для каждого отрезка РС оценить значения кепстральных коэффициентов, используя ДПФ. (\vec{K}_k).
5. Для каждого отрезка РС оценить долю максимальных значений в кепстре.
6. Для каждого звука РР определить минимальные значения доли максимальных элементов кепстра.
7. Для каждого звука РР определить максимальные значения доли максимальных элементов кепстра.
8. Для каждого звука РР определить среднее значения доли максимальных элементов кепстра.
9. Для каждого звука РР определить СКО доли максимальных элементов кепстра.

$$\sigma = \sqrt{M[X - M(X)]^2}$$

10. Проанализировать полученные результаты.
11. В соответствии с полученными результатами разбить все звуки РР на несколько подгрупп.

Результаты исследований

В таблице 2.12 представлены максимальные, минимальные, средние значения, а также значения СКО долей максимальных элементов в кепстре. Подгруппы были сформированы согласно значениям СКО.

Таблица 2.12 – доли максимальных элементов

№ группы	№ подгруппы	Тип зв. РР	Min значение	Мах значение	Среднее	СКО
1	1	э	0,00	0,38	0,04	0,01
	2	я	0,00	0,42	0,04	0,02
	3	в	0,00	0,54	0,05	0,03
	4	ы	0,00	0,56	0,06	0,03
	5	а	0,00	0,56	0,06	0,03

Окончание таблицы 2.12

	6	м	0,00	0,57	0,06	0,03	
	7	г	0,00	0,58	0,06	0,03	
	8	н	0,00	0,58	0,06	0,03	
	9	у	0,00	0,59	0,06	0,03	
	10	о	0,00	0,59	0,06	0,03	
	11	щ	0,00	0,61	0,06	0,03	
	12	л	0,00	0,61	0,06	0,03	
	13	д	0,00	0,61	0,06	0,03	
	14	ф	0,00	0,61	0,06	0,03	
	15	ч	0,00	0,61	0,06	0,03	
	16	е	0,00	0,62	0,06	0,03	
	17	ш	0,00	0,62	0,06	0,03	
	2	1	б	0,00	0,63	0,06	0,04
		2	ц	0,00	0,63	0,06	0,04
		3	и	0,00	0,63	0,06	0,04
		4	з	0,00	0,64	0,06	0,04
		5	с	0,00	0,64	0,06	0,04
6		п	0,00	0,61	0,07	0,04	
7		к	0,00	0,64	0,06	0,04	
8		х	0,00	0,64	0,06	0,04	
9		т	0,00	0,64	0,06	0,04	
10		р	0,00	0,64	0,06	0,04	
11		й	0,00	0,45	0,11	0,04	
12		ж	0,00	0,65	0,06	0,04	
3	1	ё	0,00	0,57	0,19	0,07	
	2	ю	0,00	0,60	0,20	0,08	

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 3 подгруппы, в зависимости от значения параметра «доли максимальных элементов». Разница в полученных значениях незначительна.

2.3 Исследование частотных (субполосных) характеристик

В некотором смысле субполосный анализ – это задача описания свойств объекта с позиций разбиения области пространственных частот на подобласти. При этом основное внимание уделяется задаче вычисления

точных значений долей энергии в заданных подобластях пространственных частот и оптимальному выделению субполосных компонент.

Вычислительный эксперимент № 1

Цель: выявить зависимость величины частотной концентрации от типа звуков РР.

Исходные данные: речевой сигнал, соответствующий звукам русской речи, записанные с частотой дискретизации 16 кГц, разрядностью 16 бит, длительность отрезка анализа для определения числа переходов через ноль $T=16$ мс. ($N=256$). Количество частотных интервалов, на которые разбивается ось частот $R=128$. Значение доли энергии $m=0.7; 0.8; 0.9$

План

1. Загрузить речевой сигнал, соответствующий определенным звукам РР (\vec{X})
2. Каждый речевой сигнал разбить на отрезки одинаковой длительности N (\vec{X}_N)
3. Для каждого отрезка РС оценить значение частотной концентрации при заданном значении m по формуле:

$$W_{NR} = f_{NR}^m / R$$

где R – количество равновеликих интервалов, на которые разбивается ось частот; f_{NR}^m – минимальное количество частотных интервалов (частотная концентрация), в которых сосредоточена заданная доля энергии m звукового отрезка, т.е.:

$$f_{NR}^m = \min d_{NR}^m$$

4. Для каждого звука РР вычислить минимальные значения частотной концентрации.
5. Для каждого звука РР вычислить максимальные значения частотной концентрации.

6. Для каждого звука РР вычислить средние значения частотной концентрации.
7. Для каждого звука РР вычислить СКО частотной концентрации от среднего значения для каждого звука РР.
8. Проанализировать полученные результаты.
9. В соответствии с полученными результатами разбить все звуки РР на несколько подгрупп.

Результаты исследований

В таблице 2.13 представлены максимальные, минимальные, средние значения, а также значения СКО частотной концентрации. Подгруппы были сформированы согласно средним значениям.

Таблица 2.13 – частотная концентрация

№ группы	№ подгруппы	Тип зв. РР	Min значение	Max значение	Среднее	СКО
1	1	б	3,00	10,00	3,95	0,67
	2	н	3,00	10,00	4,26	0,44
	3	в	3,00	10,00	4,34	1,40
	4	м	3,00	10,00	4,81	1,94
	5	у	3,00	9,00	4,96	1,17
	6	ю	3,00	15,00	5,06	3,19
	7	д	3,00	15,00	5,10	4,37
2	1	г	3,00	29,00	6,07	9,30
	2	ё	3,00	14,00	6,94	3,85
	3	е	3,00	15,00	6,95	3,85
	4	л	4,00	27,00	7,06	8,66
	5	э	3,00	17,00	7,16	7,19
	6	ы	3,00	14,00	7,21	5,25
	7	и	3,00	17,00	7,22	5,35
	8	о	5,00	13,00	7,52	1,44
	9	й	3,00	16,00	8,26	5,57
	10	а	5,00	19,00	8,85	8,56
	11	я	4,00	18,00	9,18	3,76
	12	х	3,00	26,00	9,63	14,45
	13	р	4,00	20,00	9,75	10,45
3	1	ж	3,00	37,00	10,31	51,29
	2	п	3,00	30,00	11,09	21,62

Окончание таблицы 2.13

3	3	з	3,00	40,00	12,62	96,96
	4	к	4,00	33,00	13,67	53,57
4	1	т	4,00	30,00	18,99	33,04
	2	щ	6,00	33,00	21,35	17,17
	3	ц	4,00	39,00	21,41	83,29
	4	ф	4,00	44,00	21,45	123,33
	5	ш	3,00	42,00	21,80	83,85
	6	ч	6,00	36,00	21,87	43,70
	7	с	3,00	37,00	24,23	35,27

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 4 подгруппы, в зависимости от значения параметра «частотная концентрация».

Вычислительный эксперимент № 2

Цель: выявить зависимость номера частотного интервала с максимальной энергией от типа звуков русской речи.

Исходные данные: речевой сигнал, соответствующий различным звукам русской речи, записанный с частотой дискретизации 16 кГц и разрядностью 16 бит. Длительность отрезка анализа $T=16$ мс ($N=256$). Количество частотных интервалов, на которые разбивается ось частот $R=128$.

План

1. Загрузить речевой сигнал, соответствующий звукам русской речи.
2. Разбить речевой сигнал на отрезки одинаковой длительности (\vec{X}_N).
3. Для каждого отрезка РС оценить распределение энергии в частотной области, используя субполосный анализ.

$$P_r(\vec{x}_N) = \int_{\omega \in V_r} |X(\omega)|^2 d\omega / 2\pi$$

$$\Phi_N^r = \int_{\omega \in V_r} X(\omega)Y^*(\omega)d\omega / 2\pi$$

Соотношение $P_r(\bar{x}_N)$ определяет часть энергии, а Φ_N^r - субполосную корреляцию спектров различных отрезков.

4. Для каждого отрезка РС определить номер частотного интервала с максимальной энергией.
5. Для каждого отрезка РС определить максимальную долю энергии, сосредоточенную в одном интервале.
6. Для каждого звука РР определить минимальное значение номера частотного интервала с максимальной энергией.
7. Для каждого звука РР определить максимальное значение номера частотного интервала с максимальной энергией.
8. Для каждого звука РР определить среднее значение номера частотного интервала с максимальной энергией.
9. Для каждого звука РР определить СКО номера частотного интервала с максимальной энергией.
10. Для каждого звука РР определить минимальное значение максимальной доли энергии, сосредоточенной в одном интервале.
11. Для каждого звука РР определить максимальное значение максимальной доли энергии, сосредоточенной в одном интервале.
12. Для каждого звука РР определить среднее значение максимальной доли энергии, сосредоточенной в одном интервале.
13. Для каждого звука РР определить СКО максимальной доли энергии, сосредоточенной в одном интервале.
14. Проанализировать полученные результаты
15. В соответствии с полученными результатами разбить звуки РР на подгруппы.

Результаты исследований

В таблицах 2.14 – 2.15 представлены номера максимальных интервалов и доли максимальных интервалов (их максимальные, минимальные, средние

значения, а также значения СКО). Подгруппы были сформированы согласно средним значениям. Единица измерения – м/с.

Таблица 2.14 – номера максимальных интервалов

№ группы	№ подгруппы	Тип зв. РР	Min значение	Max значение	Среднее	СКО
1	1	ю	1,00	4,00	2,24	1,25
	2	б	1,00	4,00	2,34	0,63
	3	в	1,00	4,00	2,39	0,96
	4	м	1,00	3,00	2,40	0,49
	5	д	1,00	6,00	2,44	1,15
	6	н	1,00	4,00	2,50	0,55
	7	г	1,00	6,00	2,53	0,62
	8	э	1,00	6,00	2,98	1,46
2	1	ж	1,00	20,00	3,05	7,37
	2	и	1,00	4,00	3,21	0,31
	3	ы	1,00	5,00	3,24	0,95
	4	у	1,00	5,00	3,49	0,30
	5	й	3,00	5,00	3,59	0,25
	6	ё	1,00	5,00	3,68	0,64
	7	р	1,00	6,00	3,74	1,03
	8	е	1,00	5,00	3,81	0,39
	9	о	1,00	12,00	3,84	1,89
	10	л	2,00	7,00	3,92	1,32
	11	я	1,00	15,00	3,96	3,21
3	1	п	1,00	19,00	4,72	8,77
	2	х	1,00	15,00	5,04	17,99
	3	з	1,00	49,00	5,28	111,97
	4	а	2,00	15,00	5,42	3,22
	5	ф	1,00	46,00	7,87	84,97
4	1	к	1,00	32,00	9,60	115,79
	2	т	1,00	39,00	10,74	147,05
	3	ш	1,00	49,00	14,48	165,66
	4	ч	1,00	41,00	16,47	111,99
5	1	ц	1,00	57,00	19,51	387,02
	2	щ	1,00	49,00	23,01	43,93
	3	с	1,00	50,00	27,42	189,29

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 5 подгрупп, в зависимости от значения параметра «номера максимальных интервалов». Наибольшее количество гласных звуков находится во второй подгруппе.

Таблица 2.15 – доли максимальных интервалов

№ группы	№ подгруппы	Тип зв. РР	Min значение	Мах значение	Среднее	СКО
1	1	с	0,06	0,42	0,15	0,00
	2	щ	0,06	0,38	0,17	0,00
	3	т	0,06	0,46	0,18	0,01
	4	ц	0,05	0,42	0,18	0,01
	5	ч	0,06	0,40	0,18	0,01
	6	ф	0,05	0,42	0,18	0,01
	7	ш	0,04	0,44	0,20	0,01
2	1	к	0,08	0,40	0,22	0,01
	2	п	0,06	0,40	0,22	0,00
	3	я	0,1	0,41	0,23	0,00
	4	о	0,1	0,39	0,23	0,00
	5	х	0,1	0,50	0,25	0,00
	6	р	0,1	0,42	0,25	0,00
	7	а	0,1	0,41	0,26	0,00
	8	з	0,07	0,50	0,27	0,01
	9	ж	0,09	0,53	0,29	0,01
	10	ё	0,1	0,44	0,29	0,00
	11	э	0,1	0,46	0,30	0,00
	12	л	0,1	0,44	0,30	0,00
	13	й	0,1	0,43	0,30	0,00
	14	ы	0,1	0,48	0,30	0,00
3	1	е	0,1	0,46	0,31	0,00
	2	и	0,1	0,45	0,31	0,00
	3	г	0,09	0,49	0,32	0,00
	4	д	0,1	0,49	0,32	0,01
	5	ю	0,1	0,54	0,33	0,00
	6	у	0,1	0,45	0,33	0,00
	7	м	0,1	0,46	0,33	0,00
	8	н	0,1	0,48	0,33	0,00
	9	в	0,1	0,51	0,36	0,00
	10	б	0,1	0,49	0,36	0,00

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 3 подгруппы, в зависимости от значения параметра «доли максимальных интервалов».

Вычислительный эксперимент № 3

Форманта — термин фонетики, обозначающий акустическую характеристику звуков речи (прежде всего гласных), связанную с уровнем частоты голосового тона и образующую тембр звука.

Тембр нашего голоса – не бесконечное количество звуковых волн и не сумма основного звука, а довольно ограниченная по своему спектру сложная звуковая волна. Эта звуковая волна состоит из нескольких областей – групп частот, усиленных резонаторной системой вокального тракта, тогда как другие, которые ротоглоточной полостью не отфильтрованы, не оказывают влияния на тембр голоса вообще. Эти области усиленных частот и называются формантами.

Самую низкую по частоте форманту называют Основной, в ней содержится основной тон звука, сама нота. Следующую форманту называют Первой речевой. Эта форманта в речи возникает, благодаря глоточной полости. Для распознавания гласного звука важны первые две форманты – первая и следующая за ней вторая речевая. «Ответственной» за вторую речевую форманту является ротовая полость. Именно размер и взаимное расположение в спектре первой и второй формант определяют то, какую гласную мы услышим.

В обычной разговорной речи этих трех формант (не забыли про Основную) вполне достаточно для создания всего спектра гласных звуков языка. При пении вокальный аппарат человека организован несколько иначе, и это приводит к иному взаимодействию человеческих резонаторов

Цель: выявить зависимость количества и положения формантных областей от типа звука РР.

Исходные данные: РС, соответствующий различному звуку РР, записанный с частотой дискретизации 16 кГц и разрядностью 16 бит. Длительность отрезка анализа $T=16$ мс ($N=256$). Количество частотных интервалов, на которые разбивается ось частот $R=128$. Доля энергии сигнала $m=0.9$.

План

1. Загрузить речевой сигнал, соответствующий звукам русской речи.
2. Разбить речевой сигнал на отрезки одинаковой длительности (\vec{X}_N).

3. Для каждого отрезка РС оценить распределение энергии в частотной области, используя субполосный анализ.
4. Для каждого отрезка РС определить информационные частотные интервалы, как интервалы, в которых сосредоточены основная доля энергии m .
5. Для каждого отрезка РС определить количество формантных областей.
6. Для каждого отрезка РС определить ширину каждой формантной области.
7. Для каждого звука РР определить центральную частоту каждой формантной области.
8. Для каждого звука РР определить минимальное значение числа формантных областей.
9. Для каждого звука РР определить максимальное значение числа формантных областей.
10. Для каждого звука РР определить среднее значение числа формантных областей.
11. Для каждого звука РР определить СКО числа формантных областей.
12. Проанализировать полученные результаты.
13. В соответствии с полученными результатами все звуки РР разбить на подгруппы.

Результаты исследований

В таблице 2.16 представлены значения количества формант (их максимальные, минимальные, средние значения, а также значения СКО). Подгруппы были сформированы согласно средним значениям. Единица измерения – м/с.

Таблица 2.16 – количество формант

№ группы	№ подгруппы	Тип зв. РР	Min значение	Max значение	Среднее	СКО
1	1	б	1,00	3,00	1,01	0,01
	2	н	1,00	3,00	1,01	0,01
	3	в	1,00	3,00	1,08	0,08
	4	у	1,00	3,00	1,13	0,12
	5	м	1,00	3,00	1,15	0,14
	6	ю	1,00	3,00	1,17	0,15
	7	д	1,00	4,00	1,17	0,19
	8	г	1,00	8,00	1,39	0,84
	9	о	1,00	4,00	1,49	0,30
	10	л	1,00	8,00	1,53	0,75
	11	ё	1,00	3,00	1,67	0,40
	12	э	1,00	4,00	1,71	0,61
	13	е	1,00	4,00	1,87	0,50
	14	ы	1,00	4,00	1,90	0,47
	15	и	1,00	4,00	1,90	0,45
2	1	а	1,00	6,00	2,01	0,89
	2	я	1,00	5,00	2,09	0,34
	3	х	1,00	7,00	2,16	0,85
	4	п	1,00	6,00	2,20	0,96
	5	й	1,00	5,00	2,24	0,55
	6	р	1,00	6,00	2,26	0,73
	7	ж	1,00	10,00	2,55	3,49
	8	з	1,00	10,00	2,74	4,69
	9	к	1,00	7,00	2,88	2,64
	10	т	1,00	8,00	3,53	1,53
3	1	щ	1,00	11,00	4,10	1,43
	2	ф	1,00	12,00	4,11	5,35
	3	ц	1,00	11,00	4,13	3,57
	4	ч	1,00	9,00	4,37	1,78
	5	ш	1,00	10,00	4,57	3,46
	6	с	1,00	9,00	4,77	2,32

Полученные результаты проведенных экспериментов позволили разбить все звуки русской речи на 3 подгруппы, в зависимости от значения параметра «количество формант». Наибольшее количество гласных звуков находится в первой подгруппе, согласных – в третьей.

Для всех звуков русской речи были построены гистограммы временных, мел-кепстральных и частотных характеристик. На рисунках 2.1 –

2.32 представлены гистограммы на примере двух звуков – гласной «О» и согласной «Л».

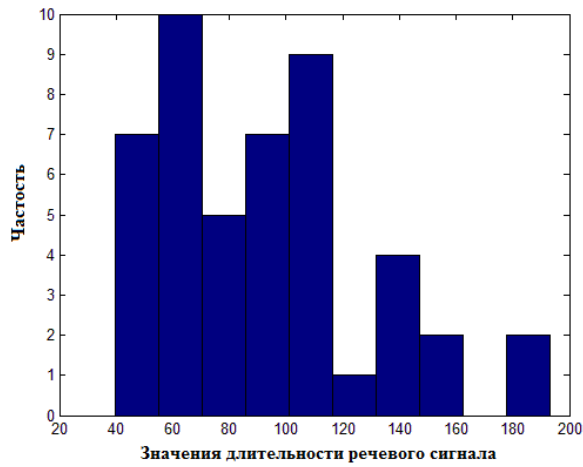


Рисунок 2.1 – Гистограмма длительности звука «О»

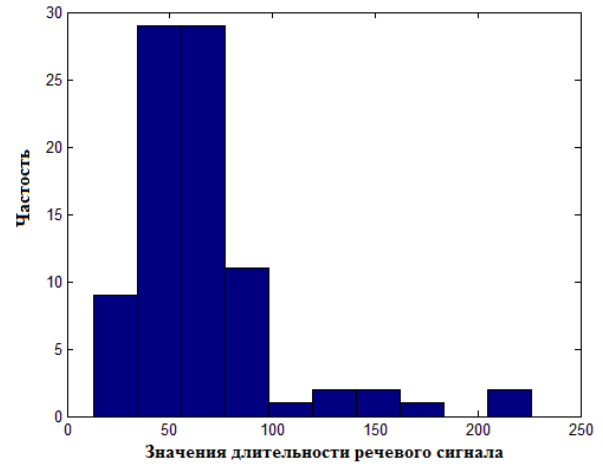


Рисунок 2.2 – Гистограмма длительности звука «Л»

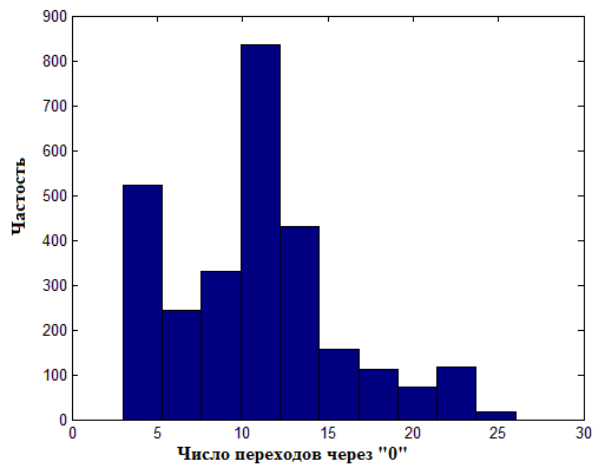


Рисунок 2.3 – Гистограмма числа переходов через «0» звука «О»

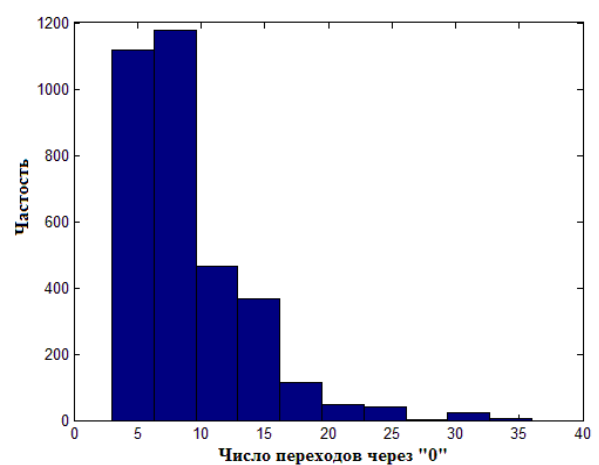


Рисунок 2.4 – Гистограмма числа переходов через «0» звука «Л»

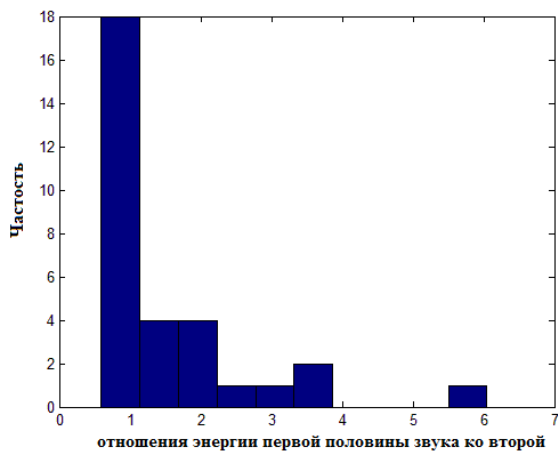


Рисунок 2.5 – Гистограмма отношения энергий звука «О»

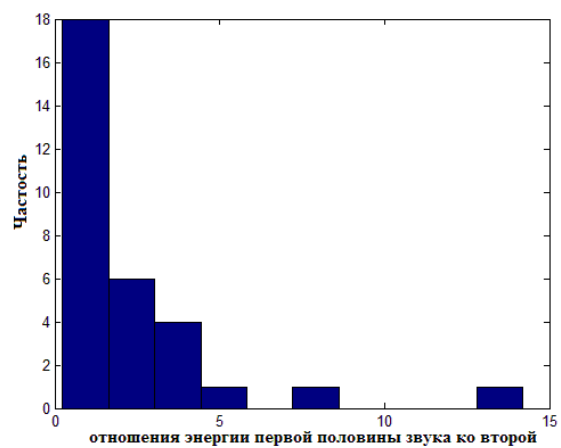


Рисунок 2.6 – Гистограмма отношения энергий звука «Л»

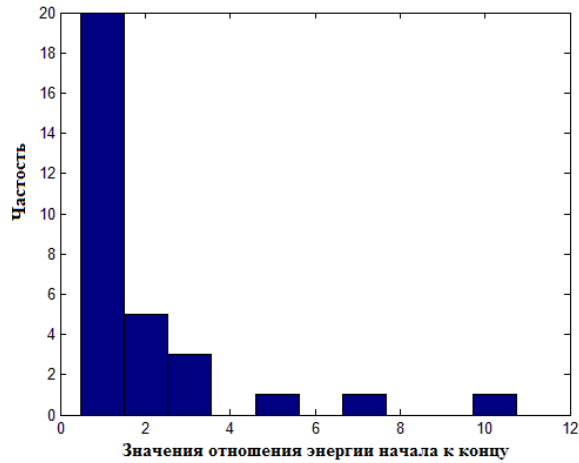


Рисунок 2.7 – Гистограмма отношения начала к концу звука «О»

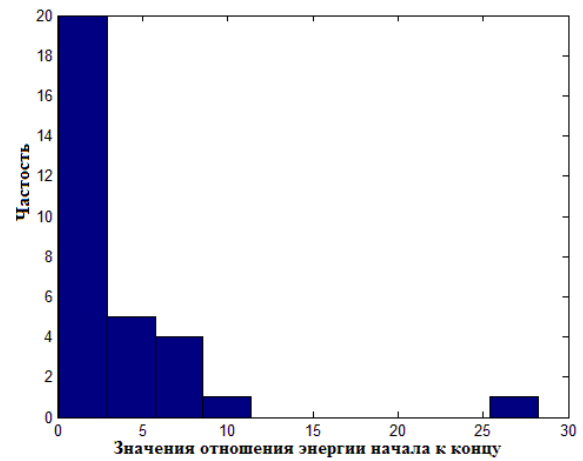


Рисунок 2.8 – Гистограмма отношения начала к концу звука «Л»

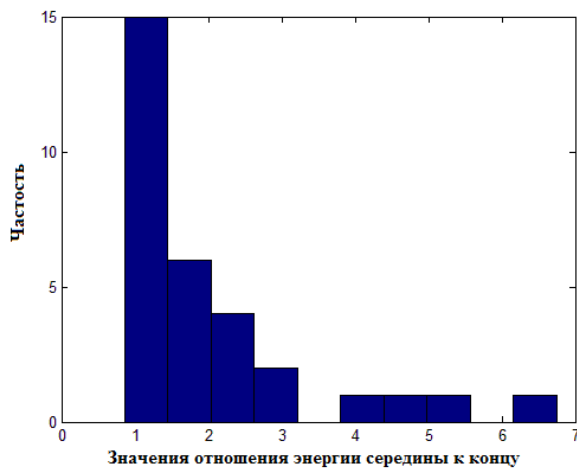


Рисунок 2.9 – Гистограмма отношения середины к концу звука «О»

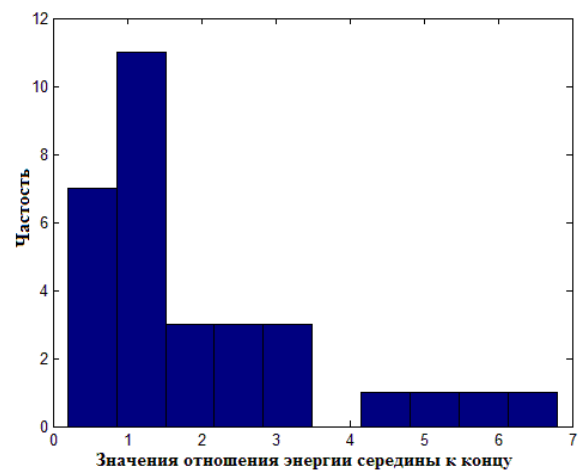


Рисунок 2.10 – Гистограмма отношения середины к концу звука «Л»

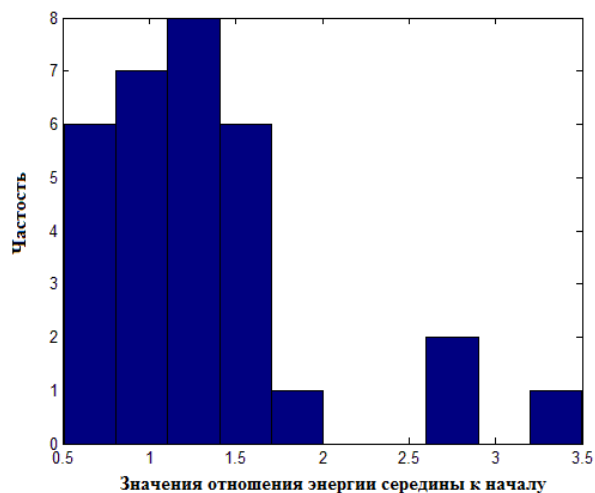


Рисунок 2.11 – Гистограмма отношения середины к началу звука «О»

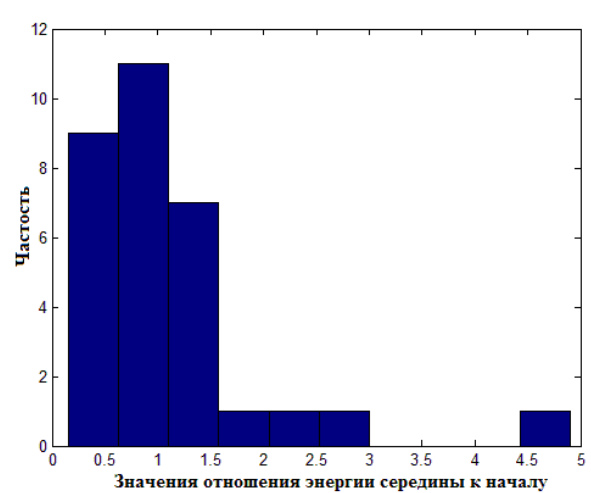


Рисунок 2.12 – Гистограмма отношения середины к началу звука «Л»

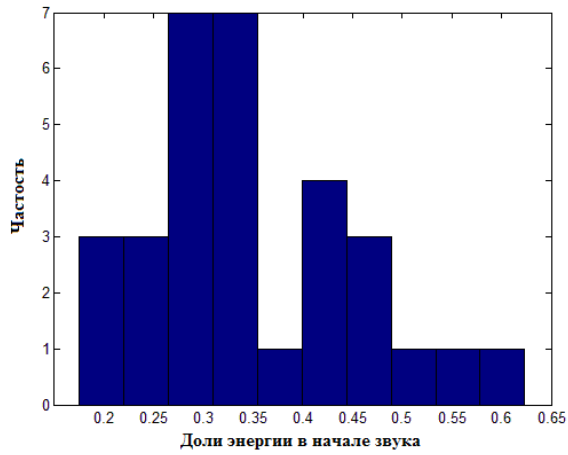


Рисунок 2.13 – Гистограмма доли энергии в начале звука «О»

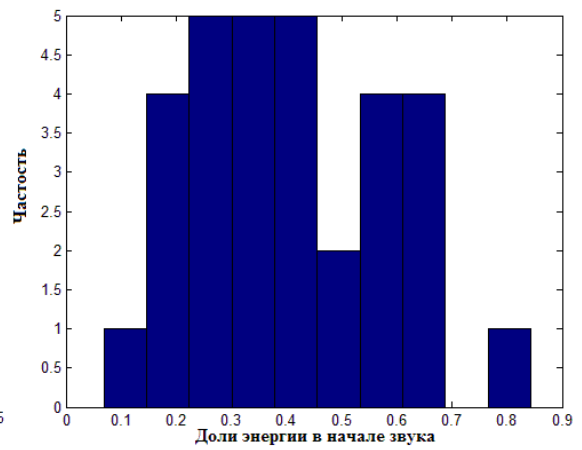


Рисунок 2.14 – Гистограмма доли энергии в начале звука «Л»

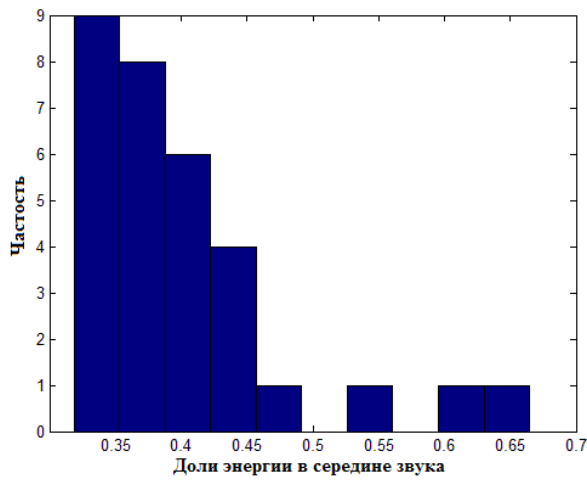


Рисунок 2.15 – Гистограмма доли энергии в середине звука «О»

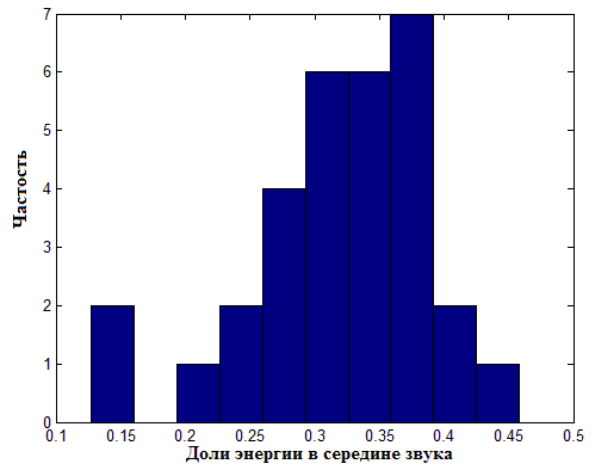


Рисунок 2.16 – Гистограмма доли энергии в середине звука «Л»

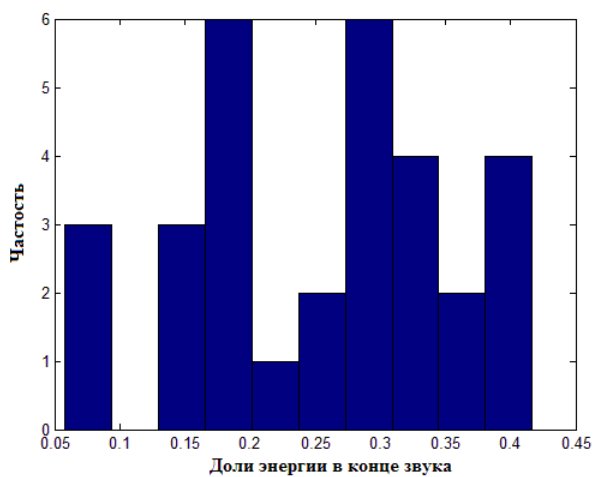


Рисунок 2.17 – Гистограмма доли энергии в конце звука «О»

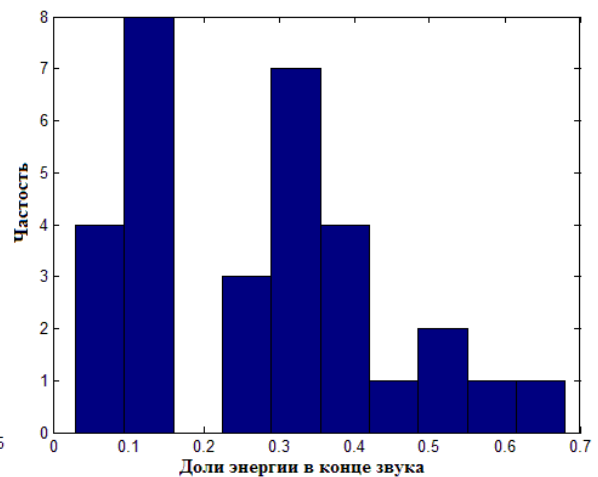


Рисунок 2.18 – Гистограмма доли энергии в конце звука «Л»

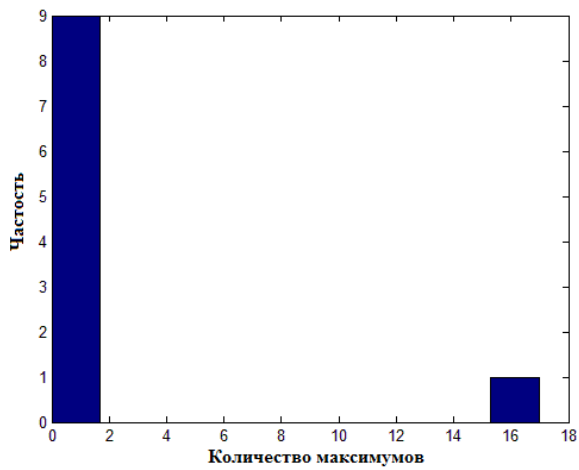


Рисунок 2.19 – Гистограмма количества максимумов звука «О»

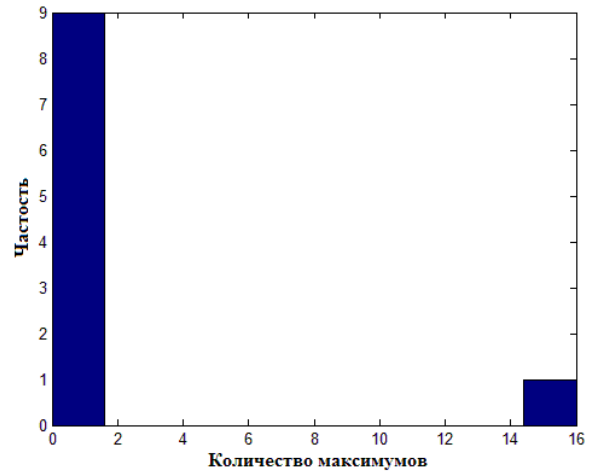


Рисунок 2.20 – Гистограмма количества максимумов звука «Л»

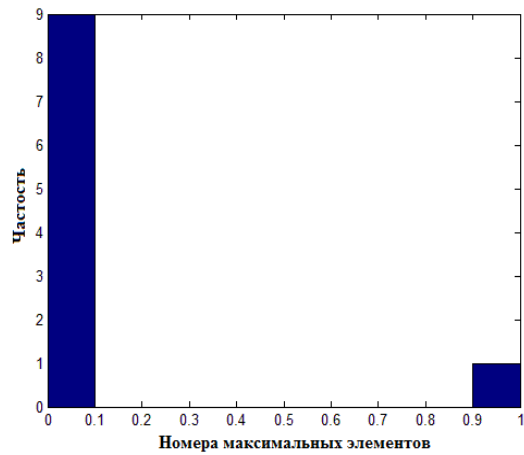


Рисунок 2.21 – Гистограмма номеров максимальных элементов звука «О»

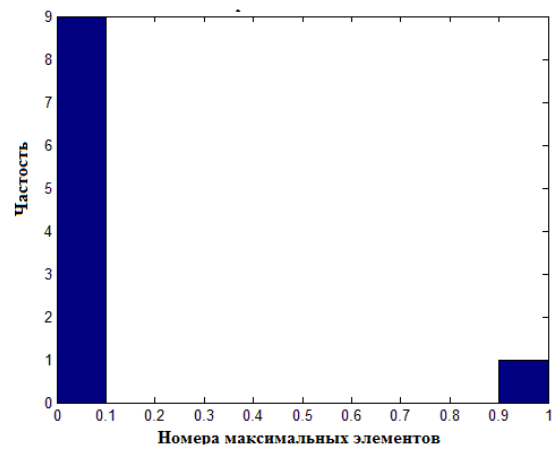


Рисунок 2.22 – Гистограмма номеров максимальных элементов звука «Л»

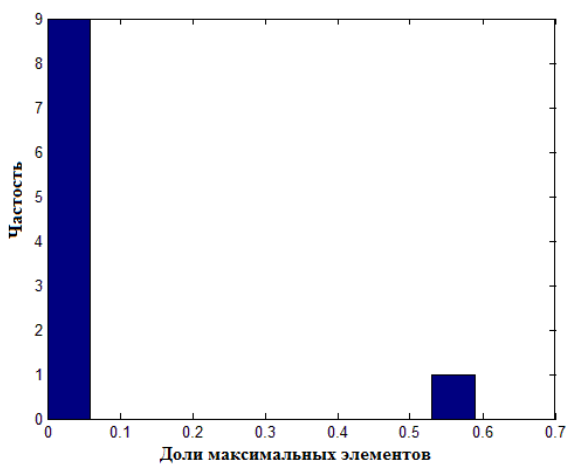


Рисунок 2.23 – Гистограмма долей максимальных элементов звука «О»

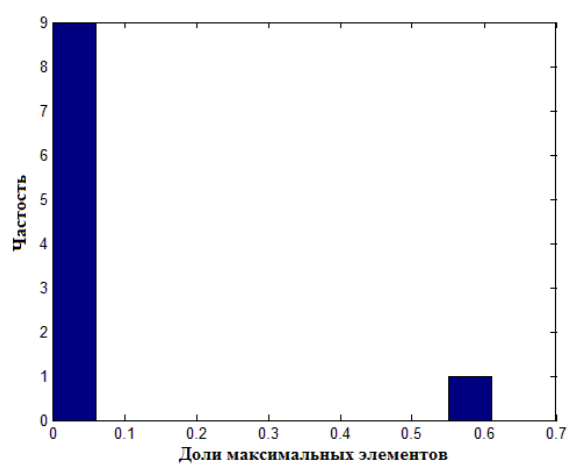


Рисунок 2.24 – Гистограмма долей максимальных элементов звука «Л»

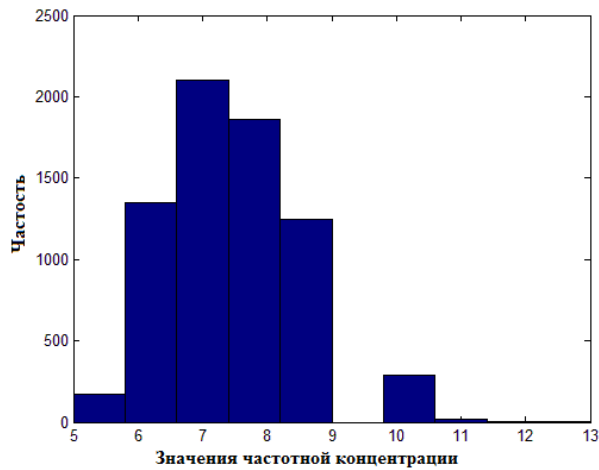


Рисунок 2.25 – Гистограмма значений частотной концентрации звука «О»

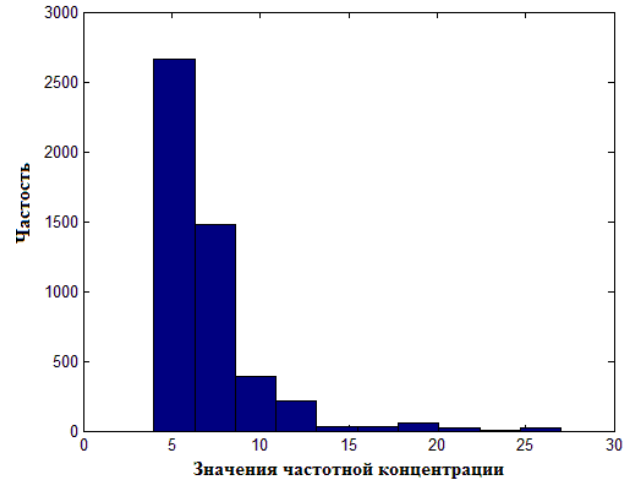


Рисунок 2.25 – Гистограмма значений частотной концентрации звука «Л»

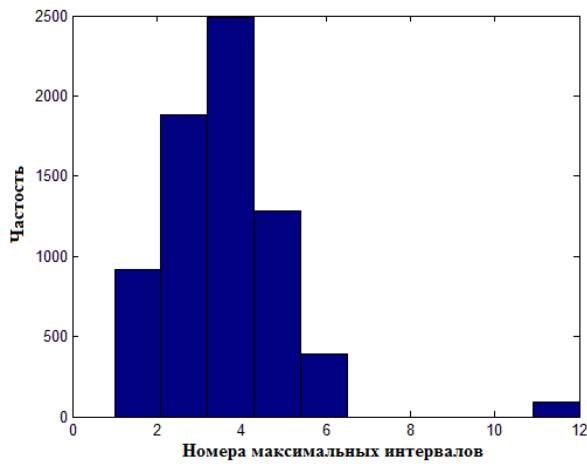


Рисунок 2.27 – Гистограмма номеров максимальных интервалов звука «О»

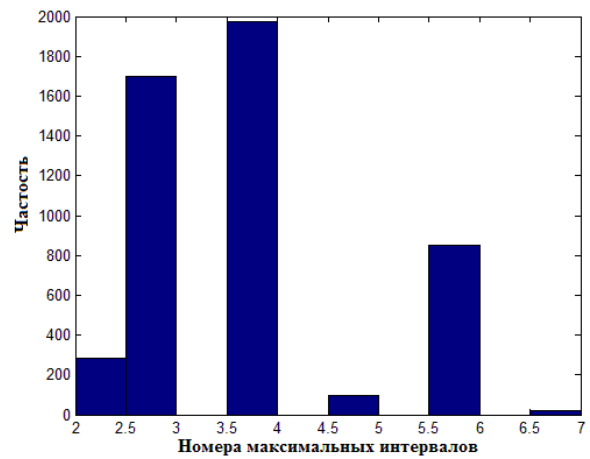


Рисунок 2.28 – Гистограмма номеров максимальных интервалов звука «Л»

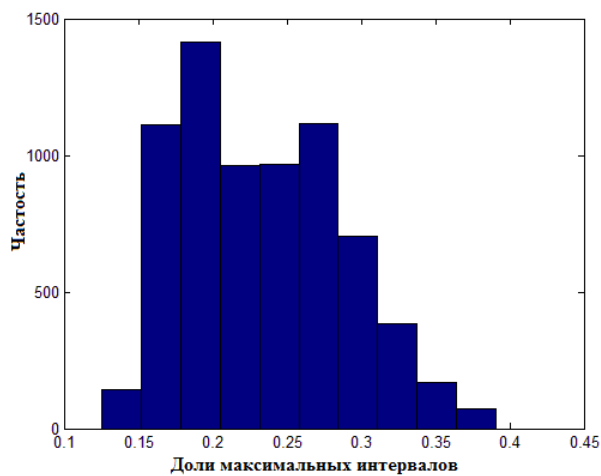


Рисунок 2.29 – Гистограмма долей максимальных интервалов звука «О»

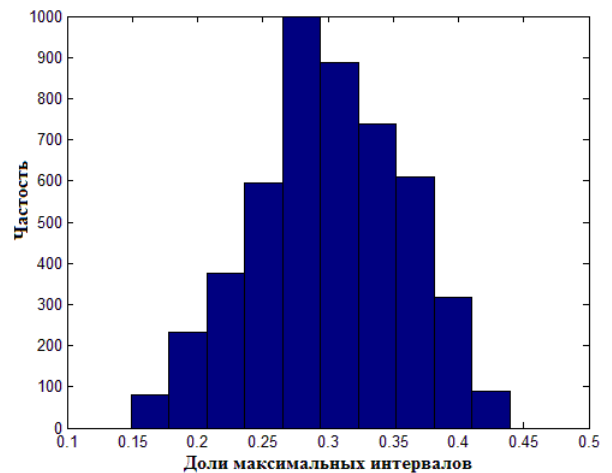


Рисунок 2.30 – Гистограмма долей максимальных интервалов звука «Л»

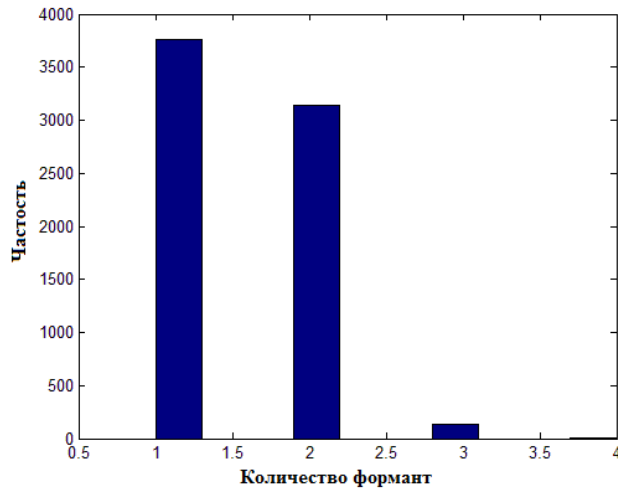


Рисунок 2.31 – Гистограмма количества формант звука «О»

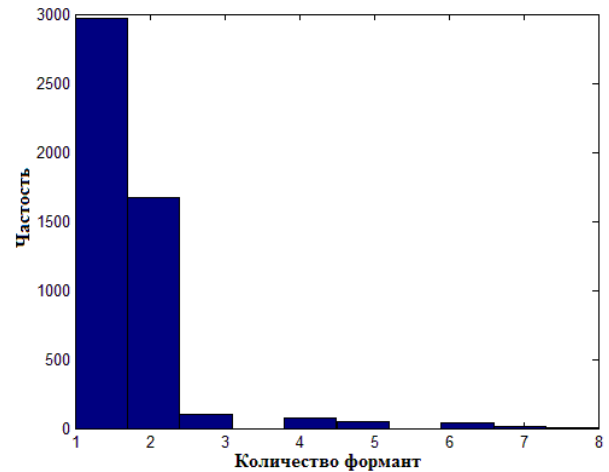


Рисунок 2.32 – Гистограмма количества формант звука «Л»

Данное исследование показывает, что различные звуки могут иметь сходные значения некоторых параметров, такие, как параметры – «длительность», «отношение энергий», «отношение начала к концу», «отношение середины к концу», «доли максимальных коэффициентов кепстра», «количество максимумов кепстра», «номера максимальных коэффициентов кепстра» представленные на рисунках 2.1 – 2.2, 2.7 – 2.10, 2.13 – 2.14, а также 2.19 – 2.24 и 2.27 – 2.28, но большинство параметров для двух различных звуков будут иметь разные значения. Это такие параметры, как «переходы через ноль», «отношение энергий первой половине ко второй», «отношение середины к концу», «доли энергии в середине звука», «доли энергии в конце звука», «частотная концентрация», «доли максимальных элементов». Таким образом, сравнение звуков по одному параметру не имеет смысла, а сравнение по всему набору может быть использовано как один из подходов в системах распознавания.

2.4 Анализ возможности применения полученных параметров в системах распознавания

Как было сказано выше, сравнение звуков по одному параметру не имеет смысла, а сравнение по всему набору может быть использовано как один из подходов в системах распознавания. Данное утверждение подтверждается исследованием набора параметров для отдельных звуков по сравнению с другими звуками.

В таблицах 2.17 - 2.22 приведен сравнительный анализ схожести значений всех групп характеристик на примере гласных звуков «А» «И» «О» и согласных «К» «М» «Ф».

Таблица 2.17 - результат исследования звука А

	Временные характеристики	Кепстральные характеристики	Частотные характеристики	Всего
	Количество наборов характеристик соответствующих звуку А			
А	9	3	4	16
	Количество наборов характеристик соответствующих звуку А и другому указанному звуку			
е	4	3	1	8
и	6	1	1	8
л	5	2	1	8
р	3	1	3	7
х	2	1	4	7
ч	3	3	1	7
д	3	3	0	6
ж	4	1	1	6
з	3	1	2	6
к	3	1	2	6
н	3	3	0	6
п	2	1	3	6
с	4	2	0	6
в	2	3	0	5
й	2	1	2	5
м	3	2	0	5
я	2	1	2	5
б	3	1	0	4
ф	0	3	1	4
ш	2	2	0	4
щ	1	3	0	4

Окончание таблицы 2.17

ы	2	1	1	4
о	0	2	1	3
у	1	2	0	3
ц	1	2	0	3
э	1	1	1	3
г	0	1	1	2
ё	1	0	1	2
т	0	1	1	2
ю	1	0	0	1

В таблице 2.17 определено, что из 16 наборов характеристик, соответствующих звуку «А», в 8 случаях характеристики совпадают с характеристиками звуков «Е» «И» «Л». В 7 случаях совпадают характеристики звуков «Р» «Х» «Ч». При этом, временные характеристики звука «Е» совпадают с временными характеристиками звука «А» в 4 случаях, а звука «И» - в 6 случаях, кепстральные характеристики звука «Е» совпадают с «А» в 3 случаях, а у звука «И» в 1 случае, и частотные характеристики звуков «Е» «И» совпадают со звуком «А» в одном случае.

Таблица 2.18 - результат исследования звука «И»

	Временные характеристики	Кепстральные характеристики	Частотные характеристики	Всего
	Количество наборов характеристик соответствующих звуку И			
И	9	3	4	16
	Количество наборов характеристик соответствующих звуку И и другому указанному звуку			
л	7	2	3	12
р	6	3	2	11
б	4	3	2	9
у	5	1	3	9
а	6	1	1	8
е	3	1	4	8
в	4	1	2	7
ж	3	3	1	7
к	4	3	0	7
о	2	2	3	7
ы	4	1	2	7
г	2	1	3	6
ё	3	0	3	6
й	1	3	2	6

Окончание таблицы 2.18

м	4	0	2	6
э	4	0	2	6
д	2	1	2	5
з	3	2	0	5
н	2	1	2	5
п	2	3	0	5
х	1	3	1	5
с	2	2	0	4
т	1	3	0	4
ц	2	2	0	4
щ	3	1	0	4
ю	2	0	2	4
я	2	0	2	4
ч	1	1	0	2
ш	2	0	0	2
ф	0	1	0	1

В таблице 2.18 определено, что из 16 наборов характеристик, соответствующих звуку «И», в 12 случаях характеристики совпадают с характеристиками звука «Л». В 11 случаях характеристики совпадают с характеристиками звука «Р». В 9 случаях совпадают характеристики звуков «Б» и «У». В 8 случаях совпадают характеристики звуков «Е» «А». При этом, временные характеристики звука «Л» совпадают с временными характеристиками звука «И» в 7 случаях, звука «Р» - в 6 случаях, а звуком «Б» и «У» - в 3 и 5 случаях соответственно. Мел-кепстральные характеристики звука «Л» совпадают с характеристиками звука «И» в 2 случаях, характеристики звука «Р» - в 3 случаях. Звук «Б» в 3 случаях, звук «У» в одном случае. Частотные характеристики звуков «Л» совпадают в трех случаях, звука «Р» в двух случаях.

Таблица 2.19 - результат исследования звука «О»

	Временные характеристики	Кепстральные характеристики	Частотные характеристики	Всего
	Количество наборов характеристик соответствующих звуку О			
О	9	3	4	16
	Количество наборов характеристик соответствующих звуку И и другому указанному звуку			

Окончание таблицы 2.19

л	1	3	3	7
ы	3	1	3	7
э	4	1	2	7
я	3	1	3	7
г	2	2	2	6
и	1	2	3	6
к	3	2	1	6
м	4	1	1	6
н	3	2	1	6
у	4	1	1	6
х	2	2	2	6
ш	5	1	0	6
е	2	1	2	5
ё	2	0	3	5
з	3	2	0	5
й	1	2	2	5
р	3	1	1	5
ц	4	1	0	5
ч	3	2	0	5
щ	3	2	0	5
а	0	2	2	4
б	2	1	1	4
в	1	2	1	4
д	1	2	1	4
с	3	1	0	4
т	3	1	0	4
ф	3	1	0	4
ж	0	2	1	3
п	1	1	1	3
ю	2	0	1	3
л	1	3	3	7

В таблице 2.19 определено, что из 16 наборов характеристик, соответствующих звуку «О», в 7 случаях характеристики совпадают с характеристиками звуков «Я» «Э» «И» «Л». При этом, временные характеристики звука «Л» совпадают с временными характеристиками звука «О» в 1 случае, звука «Ы» - в 3 случаях, звука «Э» в 4 случаях, звука «Я» в 1 случае. Кепстральные характеристики звука «Л» совпадают с характеристиками звука «О» в 3 случаях, в одном случае совпадают характеристики звуков «Ы» «Э» «Я». Частотные характеристики звуков «Л» «Ы» «Я» совпадают со звуком «И» в трех случаях.

Таблица 2.20 - результат исследования звука «К»

	Временные характеристики	Кепстральные характеристики	Частотные характеристики	Всего
	Количество наборов характеристик соответствующих звуку К			
К	9	3	4	16
	Количество наборов характеристик соответствующих звуку И и другому указанному звуку			
р	6	2	2	10
п	3	3	3	9
т	4	2	2	8
ж	2	3	2	7
з	2	3	2	7
х	2	3	2	7
а	3	1	2	6
и	4	2	0	6
л	4	2	0	6
о	3	2	1	6
ц	5	1	0	6
й	1	3	1	5
м	5	0	0	5
с	3	2	0	5
ч	3	1	1	5
я	4	0	1	5
б	2	2	0	4
в	3	1	0	4
г	3	1	0	4
у	3	1	0	4
ш	3	0	1	4
д	2	1	0	3
е	2	1	0	3
ф	2	1	0	3
э	3	0	0	3
н	2	0	0	2
щ	1	1	0	2
ы	2	0	0	2
ё	1	0	0	1
ю	0	0	0	0

В таблице 2.20 определено, что из 16 наборов характеристик, соответствующих звуку «К», в 10 случаях характеристики совпадают с характеристиками звука «Р». В 9 случаях совпадают характеристики звука «П». В 8 случаях совпадают характеристики звуков «Т». При этом, временные характеристики звука «Р» совпадают с временными характеристиками звука «К» в 6 случаях, звука «П» - в 3 случаях, звука «Т» в 4 случаях. Кепстральные характеристики звука «Р» совпадают с

характеристиками звука «К» в 2 случаях, звука «П» в 3 случаях, звука «Т» в 2 случаях. Частотные характеристики звуков «Р» «Т» совпадают с характеристиками звука «К» в 2 случаях. Частотные характеристики звука «П» совпадают со звуком «К» в 3 случаях.

Таблица 2.21 - результат исследования звука «М»

	Временные характеристики	Кепстральные характеристики	Частотные характеристики	Всего
	Количество наборов характеристик соответствующих звуку М			
М	9	3	4	16
	Количество наборов характеристик соответствующих звуку М и другому указанному звуку			
н	5	2	4	11
б	5	1	4	10
в	4	2	4	10
у	5	2	3	10
д	3	2	4	9
е	3	2	2	7
ы	3	1	3	7
а	3	2	1	6
л	4	1	1	6
о	4	1	1	6
ш	3	3	0	6
э	3	1	2	6
г	2	0	3	5
и	4	0	1	5
к	5	0	0	5
р	5	0	0	5
с	4	1	0	5
ю	1	0	4	5
ё	2	1	1	4
з	4	0	0	4
ф	2	2	0	4
ц	3	1	0	4
щ	2	2	0	4
я	3	1	0	4
ж	3	0	0	3
ч	1	2	0	3
п	2	0	0	2
т	1	0	0	1
й	0	0	0	0
х	0	0	0	0

В таблице 2.21 определено, что из 16 наборов характеристик, соответствующих звуку «М», в 11 случаях характеристики совпадают с характеристиками звука «Н». В 10 случаях совпадают характеристики звуков «Б» «В» «У». В 9 случаях совпадают характеристики звука «Д». При этом, временные характеристики звуков «Н» «Б» «У» совпадают с временными характеристиками звука «М» в 5 случаях, звука «В» - в 4 случаях. Кепстральные характеристики звука «Б» совпадают с характеристиками звука «М» в 1 случае. Кепстральные характеристики звуков «Н» «В» «У» в 1 случае. Частотные характеристики звуков «Р» «Т» совпадают с характеристиками звука «К» в 2 случаях. Частотные характеристики звуков «Н» «Б» «В» совпадают с характеристиками звука «М» в 4 случаях.

Таблица 2.22 - результат исследования звука «Ф»

	Временные характеристики	Кепстральные характеристики	Частотные характеристики	Всего
	Количество наборов характеристик соответствующих звуку Ф			
Ф	9	3	4	16
	Количество наборов характеристик соответствующих звуку Ф и другому указанному звуку			
ч	5	3	3	11
х	6	1	1	8
ц	3	2	3	8
щ	2	3	3	8
н	4	3	0	7
т	4	1	2	7
з	4	1	1	6
с	1	2	3	6
в	2	3	0	5
е	2	3	0	5
о	3	2	0	5
у	2	3	0	5
э	4	1	0	5
а	1	3	0	4
б	3	1	0	4
д	1	3	0	4
л	2	2	0	4
м	2	2	0	4

Окончание таблицы 2.22

ш	0	1	3	4
я	3	1	0	4
ж	2	1	0	3
к	2	1	0	3
п	1	1	1	3
ы	2	1	0	3
ю	3	0	0	3
г	1	1	0	2
ё	2	0	0	2
й	1	1	0	2
и	0	1	0	1
р	0	1	0	1

В таблице 2.22 определено, что из 16 наборов характеристик, соответствующих звуку «М», в 11 случаях характеристики совпадают с характеристиками звука «Ч». В 8 случаях совпадают характеристики звуков «Х» «Ц» «Щ». При этом, временные характеристики звука «Ч» совпадают с временными характеристиками звука «Ф» в 5 случаях, звука «Х» - в 6 случаях, звука «Ц» - в 3 случаях, звука «Щ» - в 2 случаях. Кепстральные характеристики звука «Х» совпадают с характеристиками звука «Ф» в 1 случае. Кепстральные характеристики звуков «Ч» «Щ» в 3 случаях, звука «Ц» - в 2 случаях. Частотные характеристики звуков «Р» «Т» совпадают с характеристиками звука «К» в 2 случаях. Частотные характеристики звуков «Ц» «Ч» «Щ» совпадают с характеристиками звука «Ф» в 3 случаях. У звука «Х» совпадение частотных характеристик со звуком «Ф» происходит лишь в одном случае.

Результаты перекрестного сравнения, приведенные в таблице, показывают, что вероятность распознавания отдельного звука с использованием подобного набора колеблется в интервале от 20 до 60 процентов в зависимости от звука и от таких особенностей, как особенности голоса диктора, разного типа аппаратуры, условий записи звука, воздействие внешних шумов. Таким образом, построить самостоятельную систему

распознавания, основанную лишь на параметрическом подходе весьма затруднительно.

ЗАКЛЮЧЕНИЕ

1. Анализ существующих методов распознавания речи позволяет сказать, что основной целью обработки речевых сигналов является получение наиболее удобного и компактного представления содержащейся в них информации. Точность представления определяется той информацией, которую необходимо сохранить или выделить.

2. Временные характеристики речевых сигналов на примере длительности имеют изменяются в зависимости от скорости воспроизведения, а также зависят непосредственно от типа звука. Исследование параметров значений длительности происходит путем изучения длин фрагментов речевых сигналов. При анализе переходов сигнала через ноль характерно проявление квазипериодичности. Особенно ярко это проявляется для звуков, которые принято относить к классу вокализованных: гласные и сонорные согласные. Проявление квазипериодичности связано с работой голосовых связок.

3. При исследовании мел-кепстральных характеристик были найдены различия, которые есть между частотными составляющими в традиционном спектре и частотными составляющими в спектре, называемом кепстром. В первом случае любая частотная составляющая имеет физический смысл, сигнал с такой частотой и амплитудой действительно присутствует в исходном сигнале во временной области. Во втором же случае, в кепстре, присутствие гармоник может совсем не означать, что в исходном спектре существуют соответствующие периодичности, а только то, в какой мере вид дискретных составляющих в исходном спектре более или менее островершинный.

3. При анализе частотного распределения энергии фрагментов РС, порождаемых различными звуками русской речи, интерес представляет не

только ширина диапазона, в котором сосредоточена подавляющая доля энергии, но и местоположение этого диапазона в частотной области.

4. Результаты перекрестного сравнения, приведенные в таблице, показывают, что вероятность распознавания отдельного звука с использованием подобного набора колеблется в интервале от 20 до 60 процентов в зависимости от звука и от таких особенностей, как особенности голоса диктора, разного типа аппаратуры, условий записи звука, воздействие внешних шумов. Таким образом, построить самостоятельную систему распознавания, основанную лишь на параметрическом подходе весьма затруднительно. Такой подход может быть полезен как элемент системы распознавания, с помощью которого можно ограничить количество вероятных типов звука для дальнейшего исследования.

Цель исследования достигнута, все задачи выполнены.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Аграновский, А.В. Теоретические аспекты алгоритмов обработки и классификации речевых сигналов [Текст]/ А.В. Аграновский, Д.А. Леднов – М.: Радио и связь, 2004. – 164 с.
2. Алдошина, И.А. Основы психоакустики. Слух и речь. Часть 1 [Текст] / И.А. Алдошина // Информационно-технический журнал «Звукорежиссер». - 2002. - №1. - С.38-44.
3. Алдошина, И.А. Слуховые модели восприятия линейных и нелинейных искажений в музыке и речи. Часть 1 [Текст] / И.А. Алдошина // Информационно-технический журнал «Звукорежиссер». - 2006. - №3. – С.38-44.
4. Бабкин, В.В. Помехоустойчивый выделитель основного тона речи [Текст] /В.В. Бабкин// Труды 7-й Международной Конференции и Выставки Цифровая Обработка Сигналов и ее Применение (DSPA-2005) – Москва, 2005. – С.175-178.
5. Бабкина Л.Н., Молчанов А.П. «Способ адаптивной фильтрации речевых сигналов в слуховых аппаратах», патент RU -2047946 от 10.11.95
6. Бабкина, Л.Н. Особенности отображения частотной структуры сигналов в периферическом отделе слухового анализатора в норме и патологии [Текст]/ Л.Н. Бабкина, А.П. Молчанов // Вестник оториноларингологии. – 2000. – №3. – С.28-30
7. Баронин, С.П. Спектральный анализ и проблема сокращенного описания речевых сигналов [Текст] - В кн.: Спектральный анализ звуков речи и интонации звуков речи и интонации. М. 1969, с. 13-30.
8. Баронин, С.П. Автокорреляционный метод выделения основного тона речи. Пятьдесят лет спустя [Текст] /С.П. Баронин// Речевые технологии, 2/2008. – 2008. – С.3-12.

9. Бахвалов, Н.С. Численные методы [Текст] / Н.С. Бахвалов, Н.П. Жидков, Г.М. Кобельников. – 3-е изд., перераб. – М.: БИНОМ. Лаборатория знаний, 2003. – 632 с.
10. Белошапкова, В.А. Современный русский язык: Учеб. для филол. спец. ун-тов [Текст] / В.А. Белошапкова, Е.А. Брызгунова, Е.А. Земская под ред. В.А. Белошапковой – М.: Высш. шк., 1989. – 800 с.
11. Буланин, Л.Л. Фонетика современного русского языка [Текст]. – М.: «Высшая школа», 1970. – 206 с.
12. Васильев, Е.М. Система распознавания фонетических образов на основе нейросетевой модели восприятия речи [Текст] / Е.М. Васильев, В.В. Меренков // Вестник Воронежского государственного технического университета. – 2009. – Т.5. № 10. – С.130-134.
13. Винцюк, Т.К., Анализ, распознавание и интерпретация речевых сигналов [Текст] / Винцюк Т.К. - Киев: Наук.думка, 1987. - 264с.
14. Воеводин, В.В. Матрицы и вычисления [Текст] / В.В. Воеводин, Ю.А. Кузнецов. – М.: Наука, 1984. – 318 с.
15. Герасимов, А.В. Применение метода модифицированного линейного предсказания к задачам выделения акустических признаков речевых сигналов [Текст] / А.В.Герасимов, О.А. Морозов, В.Р. Фидельман // Радиотехника и Электроника. – 2005. – том 50. №10. – С. 1287-1292.
16. Гольдштейн, В.С. IP-Телефония [Текст] / В.С. Гольдштейн, А.В. Пинчук, А.Л.Суховицкий. – М.: Радио и связь, 2001. – 336с.: ил.
17. Гудонавичюс, Р.В. Распознавание речевых сигналов по их структурным свойствам [Текст] / Р.В. Гудонавичюс, П.П. Кемешис, А.Б. Читавичюс – Л.: «Энергия», 1977. – 64 с.
18. Деркач, М.Ф. Динамические спектры речевых сигналов [Текст] / М.Ф. Деркач, Р.Я. Гумецкий, Б.М. Гура, М.Е. Чабан – Львов: Виша школа. Изд-во при Львов. ун-те, 1983. – 168 с.

19. Дорохин, О.А. Сегментация речевого сигнала [Текст] / О.А. Дорохин, Д.Г. Старушко, Е.Е. Федоров, В.Ю. Шелепов // Институт проблем искусственного интеллекта. – 2000. – 3/2000. – С.450-458.

20. Коваль С.Л., Лабутин П.В., Раев А.Н. Патент РФ 2230375 от 10.06.2004 «Метод распознавания диктора и устройство для его осуществления».

21. Колерс, П.А. Распознавание образов. Исследование живых и автоматических распознающих систем [Текст]/ П.А. Колерс, Е.Д. Мюрей, пер. Л.И. Титомира – М.: «Мир», 1970. – 288 с.

22. Коротаев, Г.А. Системы анализа и синтеза речевого сигнала с линейным предсказанием [Текст] / Г.А. Коротаев // Зарубежная радиоэлектроника. – 1976. – № 10. – С.3-14.

23. Коротаев, Г.А. Некоторые аспекты линейного предсказания при анализе речевого сигнала [Текст] / Г.А. Коротаев // Зарубежная радиоэлектроника. – 1991. – № 7. – С.13-31.

24. Лабутин, В. К. Модели механизмов слуха [Текст] / В.К. Лабутин, А. П. Молчанов. – М.: Энергия, 1973. – 200 с.

25. Ле, Н.В. Распознавание речи на основе искусственных нейронных сетей [Текст] / Н.В. Ле, Д.П. Панченко // Технические науки в России и за рубежом: материалы междунар. заоч. науч. конф.– Москва. – 2011. – С.8-11.

26. Леонович, А.А. Современные технологии распознавания речи [Текст] /А.А. Леонович // Материалы конференции «Диалог: Компьютерная лингвистика и интеллектуальные технологии». – Звенигород. – 2005.

27. Ли, У.А. Методы автоматического распознавания речи. [Текст] В 2-х книгах. Кн.1. / Пер. с англ./Под ред. У.Ли. – М.; Мир, 1983. –328 с.

28. Ли, У.А., Методы автоматического распознавания речи. [Текст] В 2-х книгах. Кн.2. /Пер. с англ. Под ред. У.Ли. – М.; Мир, 1983. – 392 с.

29. Лузин, Д.А. Разработка и исследование системы автоматического выделения основного тона речи [Текст]: автореф. дис. канд. техн. наук /Д.А. Лузин. – Ижевск, 2009. – 26с.

30. Мазуренко, И.Л. Компьютерные системы распознавания речи [Текст] / И.Л. Мазуренко // Интеллектуальные системы. – Москва. – 1998. – т.3. вып. 1-2. – С.117-134.
31. Мазуренко, И.Л. Одна модель распознавания речи [Текст] / И.Л. Мазуренко // Компьютерные аспекты в научных исследованиях и учебном процессе. – Москва – 1996 – С.107-112.
32. Маркел, Дж.Д. Линейное предсказание речи [Текст] / Дж.Д. Маркел, А.Х. Грэй. Пер. с англ. Под ред. Ю.Н. Прохорова – М.: Связь, 1980. – 308с.
33. Матвеев, Ю.Н. Система идентификации дикторов по голосу для конкурса NIST SRE 2010/ Ю.Н. Матвеев, К.К. Симончик // 20я Международная Конференция по Компьютерной Графике и Зрению. – Санкт-Петербург. – 2010. – С.35-39.
34. Ниценко, А.В. Алгоритмы пофонемного распознавания слов наперед заданного словаря [Текст] / А.В. Ниценко, В.Ю. Шелепов // Искусственный интеллект. – 2004. – С.633-639.
35. Новиков, Л.А. Современный русский язык [Текст] / Л.А. Новиков, Л.Г. Зубкова, В.В. Иванов и др. под общей ред. Л.А. Новикова – СПб.: «Лань», 1999. – 865 с.
36. Питмен, Э. Основы теории статистических выводов [Текст] / Э. Питмен // Пер. с англ. – М.: Мир, 1986. – 104с.
37. Прохоров, Ю.Н. Рекуррентное оценивание параметров речевых сигналов [Текст] / Ю.Н. Прохоров – М.: Наука, 1977. – С.67-80.
38. Рабинер, Л. Теория и применение цифровой обработки сигналов /Л.Рабинер, Б.Гоулд – М.: Мир, 1978. – 848с.
39. Рабинер, Л.Р. Цифровая обработка речевых сигналов [Текст]/ Л.Р. Рабинер, Р.Ф. Шафер – М.: Радио и связь, 1981. – 496 с.
40. Савченко, В.В. Различение случайных сигналов в частотной области [Текст] /В.В. Савченко // Радиотехника и электроника. – 1997. – Т.42, №4. – С.426-429.

41. Сапожков, М.А. Вокодерная связь [Текст]/ М.А. Сапожков, В.Г. Михайлов – М.: «Радио и связь», 1983. – 248 с.
42. Сорокин, В.Н. Артикуляторно-ориентированная система распознавания речи [текст] / В.Н. Сорокин, А.Н. Ижнин, А.И. Цыплихин, Д.Н. Чепелев // Труды Международного семинара «Диалог - 2003». – 2003. С.657-662.
43. Сорокин, В.Н. Модель многослойного первичного анализа речевых сигналов [Текст] / В.Н. Сорокин // Труды 13-й сессии Российского акустического общества. – 2003. – С.11-16.
44. Сорокин, В.Н. Первичный анализ речевых сигналов [Текст] / В.Н. Сорокин, Д.Н. Чепелев // Акустический ж. – 2005. – Т.51, №4. – С.536-542.
45. Сорокин, В.Н. Сегментация и распознавание гласных [Текст] / В.Н. Сорокин, А.И. Цыплихин // Информационные процессы. – 2004. – Т.4, №2. – С. 202-220.
46. Сорокин, В.Н. Сегментация речи на кардинальные элементы [Текст] / В.Н. Сорокин, А.И. Цыплихин // Информационные процессы. – 2006. – Т.6, №3. – С.177-207.
47. Сорокин, В.Н. Синтез речи [Текст] / В.Н.Сорокин. – М.: Наука, 1992. – 392 с.
48. Сорокин, В.Н. Теория речеобразования [Текст] / В.Н. Сорокин – М.: Радио и связь, 1985. – 312 с.
49. Фант, Г. Акустическая теория речеобразования [Текст] / Г. Фант – М.:Наука, 1964. – 304 с.
50. Фирсова, А.А. Исследование решающей функции максимальной чувствительности к изменению частей энергии в частотных интервалах [Текст] / С.П. Белов, А.А. Фирсова // Научные ведомости Белгородского государственного университета. Сер. История. Политология. Экономика. Информатика. – 2012. – №13(132), выпуск 23/1. – С.227-231.