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

Харченко Владимира Алексеевича

РАЗРАБОТКА МОДЕЛИ ГОЛОСОВОГО УПРАВЛЕНИЯ АВТОМОБИЛЕМ НА ОСНОВЕ СУБПОЛОСНОГО АНАЛИЗА

магистерская диссертация

направления подготовки 11.04.02
Инфокоммуникационные технологии и системы связи

Научный руководитель
кандидат технических наук, доцент
кафедры ИТСиТ НИУ «БелГУ»
Балабанова Т.Н.

БЕЛГОРОД 2017

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Практически каждый день появляются новые возможности чтобы улучшить жизнь человека и внедрить в нее новые технологии и системы, что делает ее более удобной и комфортной. Распознавание голосовых команд является актуальным вопросом при разработке многих передовых технологий. Важным направлением автоматизированных систем, является создание систем голосового управления. Такие системы находят применение в: пунктах пропуска, в системе умный дом, в голосовом управлении мобильным телефоном, в повседневной жизни людей с ограниченными возможностями и т.д. Так, данное направление применяется даже в космонавтике, для управления механизмами в невесомости, когда космонавт находится в безопорном положении. К недостаткам существующих систем можно отнести высокую стоимость реализации голосового управления и недостаточно высокую степень распознавания голосовых команд. Автомобиль в современном обществе уже не роскошь и присутствует практически в каждой семье, а зачастую и не в одном экземпляре. Автомобили, управляемые вручную пытаются усовершенствовать, внедряя «умную» электронику, для упрощения управления до подачи голосовых команд. Однако, несмотря на то, что современная наука достигла очень высокого уровня, все еще довольно редко в повседневной жизни можно встретить транспортное средство, оснащенное системой распознавания голоса и управления с его помощью бортовыми командами. Ведущие автоконцерны мира, такие как Mercedes, BMW, Ford, стремятся к повышению комфорта и безопасности водителя, для этого используется качественная система управления бортовой электроникой при помощи голосовых команд («включи радио, включи следующую станцию, позвонить, номер»). Однако существующие системы распознавания голоса в автомобилях не всегда соответствуют желаемому водителем качеству. В существующих системах голосового управления зачастую пренебрегают

индивидуальными особенностями речи человека, поэтому, предлагается использовать метод, учитывающий распределение долей энергии речевого сигнала по частотным интервалам, и использовать эту особенность для верной идентификации сказанной фразы.

Целью данной научно-исследовательской работы является разработка системы голосового управления автомобилем на основе субполосного анализа и оценка ее работоспособности.

Основными задачами работы являются:

а) анализ существующих способов распознавания речи

б) анализ голосового управления автомобилем в существующих системах

в) реализация модели голосового управления в системе Matlab

г) проведение оценки эффективности разработанного метода

Объект исследования — распознавание речи.

Предметом исследования является речевой сигнал.

Методами исследования, использующимися при выполнении научно-исследовательской работы, являются:

1) анализ учебно-методической литературы, учебных пособий, электронных ресурсов;

2) изучение принципов составления и вычислительных алгоритмов;

3) проведение вычислительного эксперимента в соответствии с темой магистерской диссертации, представляющее собой компьютерное моделирование в программной среде MatLab.

Данная исследовательская работа состоит из двух основных частей: теоретической и практической.

В теоретической части научно-исследовательской работы описывается система голосового управления автомобилем, ее принципы и существующие разработки от ведущих мировых компаний, анализируются существующие методы распознавания речи, на которых может быть основано голосовое

управление автомобилем, рассматриваются основы субполосного анализа на котором и основан предлагаемый метод,

В практической части ведется разработка модели голосового управления автомобилем на основе субполосного анализа, оцениваются полученные результаты и принимается решение о качестве работы данного метода.

ГЛАВА 1 РАСПОЗНАВАНИЕ РЕЧИ И ЕГО ПРИМЕНЕНИЕ

В настоящее время все больший интерес приобретают системы распознавания речи, что обусловлено широким распространением мобильных устройств. Применение голосового управления к данным устройствам, ввиду ограниченности интерфейсов ручного ввода привело бы к расширению их возможностей и упрощению в использовании. Существующие технологии распознавания, как правило, не имеют широкого распространения ввиду ограничения их по одному или нескольким параметрам. Например, основная доля систем, позволяющих распознать большинство пользовательских запросов требуют подключения к сети Интернет (к удаленным серверам, на которых происходит обработка запросов), поскольку недостаточно их вычислительной мощности и ограниченности памяти, выделенной под словарь. По состоянию на 2012 год объем рынка систем распознавания речи оценивается в миллиард долларов США и планируется его рост. Большим спросом пользуется голосовая биометрия в судебно-медицинских и военных целях и является основным «драйвером» рынка.

К одним из первых программ для коммерческого использования можно отнести Voice Navigator, Dragon NaturallySpeaking и другие. Начало использования таких программ датируется началом девяностых годов. В основном, их используют люди с ограниченными возможностями, набор текста большого объема для которых является проблематичным. Преобразуя голосовое сообщение диктора в текст, данные программы позволяют не использовать руки при печатании. Качество перевода у данных программ довольно низкое, однако со временем оно существенно улучшается.

В результате усовершенствования производительности мобильных устройств стало возможным применение в приложениях к ним технологий по распознаванию речи. К такому приложению можно отнести Microsoft Voice Command. Так, данное приложение позволяет управлять многими функциями

при помощи голоса. К примеру, можно создать документ, набрать сообщение или воспроизвести музыкальный файл в плеере.

Можно найти следующее применение распознаванию речи в повседневной жизни, например, доктору в поликлинике вместо того чтобы заполнять вручную медицинскую карточку больного достаточно просто проговорить диагноз и данные о больном. В качестве другого примера можно привести систему «умный дом», где голосовые команды используются для управления электрикой в доме, к примеру, включением или выключением света, кондиционером, и т.д. Так, во многих приложениях на современных мобильных устройствах все больше применяются системы автоматического распознавания речи. Важно то, что данные применяемые системы, независимы от диктора и способны распознавать голос любого человека.

Следующим этапом во внедрении технологий распознавания отмечается применение интерфейсов безмолвного доступа (SSI). Данные системы основаны на обработке речи на начальной стадии артикулирования. Однако можно выделить значительный недостаток современных систем распознавания речи, это: слишком высокая чувствительность к воздействию шумов, т.е необходимо довольно разборчиво и четко произносить голосовые команды вовремя обращении к системе по распознаванию. Отметим что SSI основан на подходе заключающимся в использовании сенсоров не подверженных влиянию шумов в качестве дополнения к обработанным акустическим сигналам.

Можно выделить пять основных способов распознавания речи:

1. Распознавание отдельных команд – в данном способе требуется раздельно произносить слово\словосочетание, а затем проводится его распознавание. Качество распознавания данного способа ограничена размером имеющегося

2. Распознавание по грамматике – в данном способе проводится распознавание уже по фразам, которое соответствует определенному набору правил. Грамматика здесь задается путем использования стандартных XML-

языков, а обмен данными между системой распознавания и приложением осуществляется по протоколу MRCP.

3. Поиск ключевых слов в потоке слитной речи – в данном способе распознавание проводится по отдельным участкам речи. Здесь речь вполне может быть, как соответствующая набору неких правил, так и соответствующей определённым правилам. В данном методе нет необходимости переводить весь текст - в ней находятся лишь те участки, которые содержат заданные слова или словосочетания.

4. Распознавание слитной речи на большом словаре – в данном методе, сказанная фраза, дословно преобразуется в текст. При этом достоверность данного метода достаточно высока.

5. Распознавание речи с помощью нейронных систем – является довольно сложным методом, однако на основе нейронных сетей появляется возможность создания обучаемых и самообучающиеся систем, что является важной предпосылкой для их применения в системах распознавания (и синтеза) речи.

1.1 Классификация систем распознавания речи

При рассмотрении классификации систем распознавания речи следует отметить, что классификация может осуществляться по различным параметрам. По литературным источникам системы распознавания речи можно классифицировать следующим образом:

- В зависимости от размера словаря: системы распознавания речи с ограниченным набором слов; системы со словарем большого размера;
- В зависимости от привязки к диктору: системы, являющиеся дикторозависимыми и дикторонезависимыми;
- В зависимости от типа распознаваемой речи: системы, работающие со слитной речью или раздельной речью;

- В зависимости от назначения системы принято выделять системы диктовки и командные системы;
- В зависимости от алгоритма, используемого в системе распознавания выделяют: нейронные сети, скрытые Марковские модели, динамическое программирование;
- В зависимости от типа структурных единиц, используемых в системе (могут быть использованы слова, дифоны, фонемы, аллофоны, фразы);
- В зависимости от принципа, по которому выделяются структурные единицы системы распознавания можно разделить на системы, в которых осуществляется распознавание по шаблону и системы выделения лексических элементов;

1.2 Методы и алгоритмы распознавания речи

При рассмотрении систем распознавания речи следует отметить, что в существующих система используется, как правило два подхода, которые являются принципиально различными:

- распознавание голосовых меток;
- распознавание лексических элементов

Распознавание голосовых меток заключается в том, что распознавание фрагментов речи осуществляется по образцу, который записан заранее. Данный подход может быть использован только в простых системах распознавания речи, которые используются для выполнения изначально записанных речевых команд.

Распознавание лексических элементов является более сложным подходом. Данные методы предполагают выделение из потока речи отдельных лексических элементов, то есть фонем и аллофонов. В дальнейшем выделенные лексические элементы объединяются в слоги и морфемы. Таким

образом, в сложных системах распознавания речи используется именно этот подход.

Системы зависимые и независимые от диктора.

К классу систем, независимых от диктора относятся системы, которые работают вне зависимости от того, кто выступает в качестве диктора. Данные системы имеют возможность распознания речи любого диктора и не нуждаются в предварительном обучении.

К классу систем, зависящих от диктора относятся системы, которые требуют предварительного обучения и в его процессе настраиваются на определенного диктора. При смене диктора в таких системах возникает необходимость полной перенастройки.

Таким образом, для создания системы распознавания речи любого класса, используемая в промышленном масштабе необходим многолетний опыт в практическом применении различных речевых технологий.

Также отмечаем следующие методы распознавания речи:

а) Распознавание по образцу

Данный метод используется в стандартном мобильном телефоне. В данном случае система распознавания речи используется с целью ускоренного набора номера вызываемого абонента и списка контактов мобильного телефона.

Данная система работает следующим образом:

Если необходимо внести новый контакт в записную книжку, то система предполагает добавление голосовой метки, которая определяет этот контакт. Как правило это фамилия и имя абонента, которые необходимо произнести, возможно, для правильного внесения данную операцию приходится делать дважды или трижды.

Далее, при необходимости вызова данного абонента, активируется команда голосового набора путем нажатия определенной клавиши и произносится голосовая метка. По голосовой метке осуществляется выбор абонента и устанавливается с ним связь.

Следует отметить, что такое голосовое управление реализовано не только в мобильных телефонах, но и в других устройствах, таки, например, как компьютерная клавиатура. При организации голосового управления клавиатурой, в последнюю встраивается микрофон и назначаются голосовые метки для различных клавиш, либо комбинаций клавиш. Использование таких клавиатур является удобным для людей с ограниченными возможностями, а, так же, с помощью такой клавиатуры имеется возможность ускорить ввод информации. Однако, при наличии шумовой составляющей в помещении, качество распознавания голосовых меток значительно ухудшается.

Технология распознавания фрагментов по заранее записанным образцам применяется и во многих программах, позволяющих подключить голосовое управление к операционной системе Microsoft Windows и ее приложениям. При использовании этих программ Вы сможете запускать приложения, переключаться между ними, выбирать строки из меню и щелкать кнопки диалоговых окон, отдавая голосовые команды и не притрагиваясь руками к клавиатуре или мыши. Возможно, такие программы и ненамного ускорят работу с приложениями для обычных людей, но они отчасти помогут инвалидам, неспособным использовать стандартные средства общения с компьютером.

Эта технология работает достаточно хорошо, если телефоном пользуется только один человек, а общее количество голосовых меток не превышает десяток-другой. Если Вы «обучите» свой телефон (или клавиатуру с голосовым интерфейсом) реагировать на Ваш голос, то только Вы и сможете пользоваться речевыми метками. Таким образом, эти системы относятся к классу систем, зависимых от диктора. Впрочем, этот недостаток есть и у многих более совершенных систем распознавания речи, основанных на выделении из речи лексических элементов.

б) Предварительная обработка звуковых сигналов

Перед тем как предпринимать попытки распознавания речи, нужно выполнить предварительную обработку речевого сигнала. В ходе этой обработки следует

удалить шумы и посторонние сигналы, частотный спектр которых находится вне спектра человеческой речи. Такую обработку можно выполнить при помощи аналоговых или цифровых полосовых фильтров. Отфильтрованный звуковой сигнал нужно оцифровать, выполнив аналого-цифровое преобразование.

Всю предварительную обработку звукового сигнала можно сделать при помощи стандартного звукового адаптера, установленного в компьютере. Дополнительная цифровая обработка звукового сигнала (например, частотная фильтрация) может выполняться центральным процессором компьютера. Таким образом, при использовании современных персональных компьютеров системы распознавания речи не требуют для своей работы какого-либо специального аппаратного обеспечения.

Важным этапом предварительной обработки входного сигнала является нормализация уровня сигнала. Это позволяет уменьшить погрешности распознавания, связанные с тем, что диктор может произносить слова с разным уровнем громкости.

Заметим, однако, что если входной звуковой сигнал имеет слишком малый уровень громкости, то после нормализации может появиться шум. Поэтому для успешной работы системы распознавания речи необходимо отрегулировать оптимальным образом чувствительность микрофона. Чрезмерная чувствительность может привести к нелинейным искажениям сигнала и, как следствие, к увеличению погрешности распознавания речи.

в) Выделение информативных признаков речевого сигнала

В зависимости от различных обстоятельств форма огибающей речевого сигнала может меняться в широких пределах, что затрудняет задачу распознавания.

Для решения задачи распознавания необходимо выделить первичные признаки речи, которые будут использованы на последующих этапах процесса распознавания. Первичные признаки выделяются посредством анализа спектральных и динамических характеристик речевого сигнала.

Спектральное представление речи:

Для выделения информативных признаков речевого сигнала используется спектральное представление речи. При этом на первом этапе осуществляется получение частотного спектра речевого сигнала с помощью набора программных полосовых фильтров (выполняя так называемое дискретное преобразование Фурье).

На втором этапе выполняются преобразования полученного спектра речевого сигнала:

- логарифмическое изменение масштаба в пространстве амплитуд и частот;
- сглаживание спектра с целью выделения его огибающей;
- кепстральный анализ, т.е. обратное преобразование Фурье от логарифма прямого преобразования.

Перечисленные выше преобразования позволяют учитывать такие особенности речевого сигнала, как понижение информативности высокочастотных участков спектра, логарифмическую чувствительность человеческого уха, и т.д.

Учет динамики речи:

Помимо спектральных характеристик, необходимо учитывать и динамические особенности речи. Для этого используют дельта-параметры, представляющие собой производные по времени от основных параметров.

При этом мы можем отслеживать не только изменение параметров речи, но и скорость их изменения.

Выделение фонем и аллофонов:

Для выделения фонем и аллофонов применяются нейронные сети и метод формирования нейронных ансамблей.

При этом обучение выделению примитивов речи (фонем и аллофонов) может заключаться в формировании нейронных ансамблей, ядра которых соответствуют наиболее частой форме каждого примитива.

Формирование нейронных ансамблей представляет собой процесс обучения нейронной сети без учителя, при котором происходит статистическая обработка всех сигналов, поступающих на вход нейронной сети. При этом формируются ансамбли, соответствующие наиболее часто встречающимся сигналам. Запоминание редких сигналов происходит позже и требует подключения механизма внимания или иного контроля высшего уровня.

г) Уровни распознавания слитной речи

Распознавание слитной речи представляет собой многоуровневый процесс. После предварительной обработки речевого сигнала и выделения из него информативных признаков выполняется выделение лексических элементов речи. Это первый уровень распознавания.

На втором уровне выделяются слоги и морфемы, на третьем — слова, предложения и сообщения.

На каждом уровне сигнал кодируется представителями предыдущих уровней. То есть слоги и морфемы составляются из фонем и аллофонов, слова — из слогов и морфем, предложения и сообщения — из слов.

Рисунок 1 - Три уровня распознавания слитной речи

При переходе с уровня на уровень помимо представителей сигналов передаются и некоторые дополнительные признаки, временные зависимости и отношения между сигналами. Собирая сигналы с предыдущих уровней, высшие уровни располагают большим объемом информации (или её другим представлением), и могут осуществлять управление процессами на низших уровнях, например, с привлечением механизма внимания.

Механизм внимания используется при обучении нейронной сети. В случае использования такого механизма при появлении образца, неизвестного нейронной сети, скорость обучения многократно возрастает. При этом редко встречающийся образец запоминается в нейронной сети.

д) Применение нейронных сетей для распознавания речи

При обучении сети с учителем можно научить сеть распознавать объекты, принадлежащие заранее определенному набору классов. Если же сеть обучается без учителя, то она может группировать объекты по классам в соответствии с их цифровыми параметрами.

Таким образом, на базе нейронных сетей можно создавать обучаемые и самообучающиеся системы. Также выдвигаются некоторые требования к самообучающимся системам:

- Разработка системы заключается только в построении архитектуры системы

В процессе создания системы разработчик создает только функциональную часть, но не наполняет (или наполняет в минимальных объемах) систему информацией. Основную часть информации система получает в процессе обучения.

- Возможность контроля своих действий с последующей коррекцией

Этот принцип говорит о необходимости обратной связи Действие-Результат-Коррекция в системе. Такие цепочки очень широко распространены в сложных биологических организмах и используются на всех уровнях — от контроля мышечных сокращений на самом низком уровне до управления сложными механизмами поведения.

· Возможность накопления знаний об объектах рабочей области

Знание об объекте — это способность манипулировать его образом в памяти.

Количество знаний об объекте определяется не только набором его свойств, но ещё и информацией о его взаимодействии с другими объектами, поведении при различных воздействиях, нахождении в разных состояниях, и т.д., т.е. его поведении во внешнем окружении.

Например, знание о геометрическом объекте предполагает возможность предсказать вид его перспективной проекции при любом повороте и освещении. Это свойство наделяет систему возможностью абстрагирования от реальных объектов, т.е. возможностью анализировать объект при его отсутствии, открывая тем самым новые возможности в обучении.

· Автономность системы

При интеграции комплекса действий, которые система способна совершать, с комплексом датчиков, позволяющих контролировать свои действия и внешнюю среду, наделенная вышеприведенными свойствами система будет способна взаимодействовать с внешним миром на довольно сложном уровне.

При этом она будет адекватно реагировать на изменение внешнего окружения (естественно, если это будет заложено в систему на этапе обучения). Способность корректировать свое поведение в зависимости от внешних условий позволит частично или полностью устранить необходимость контроля извне, т.е. система станет автономной.

Возможность создания на базе искусственных нейронных сетей самообучающихся систем является важной предпосылкой для их применения в системах распознавания (и синтеза) речи.

Представление речи в виде набора числовых параметров

После выделения информативных признаков речевого сигнала можно представить эти признаки в виде некоторого набора числовых параметров (т.е. в виде вектора в некотором числовом пространстве). Далее задача

распознавания примитивов речи (фонем и аллофонов) сводится к их классификации при помощи обучаемой нейронной сети.

Нейронные сети можно использовать и более высоких уровнях распознавания слитной речи для выделения слогов, морфем и слов.

е) Нейронные ансамбли

Отмечается, что в качестве модели нейронной сети, пригодной для распознавания речи и обучаемой без учителя можно выбрать самоорганизующуюся карту признаков Кохонена. В ней для множества входных сигналов формируется нейронные ансамбли, представляющие эти сигналы. Этот алгоритм обладает способностью к статистическому усреднению, что позволяет решить проблему изменчивости речи.

По сравнению с классическим программированием, когда алгоритм решения той или иной задачи задан жестко, нейронные сети позволяют динамически изменять алгоритм простым изменением архитектуры сети.

ж) Генетические алгоритмы

Возможность изменения алгоритма работы нейронной сети простым изменением ее архитектуры позволяют решать задачи совершенно новым способом, с помощью так называемых генетических алгоритмов.

При использовании генетических алгоритмов создаются правила отбора, позволяющие определить, лучше или хуже справляется новая нейронная сеть с решением задачи. Кроме того, определяются правила модификации нейронной сети.

Изменяя достаточно долго архитектуру нейронной сети и отбирая те архитектуры, которые позволяют решить задачу наилучшим образом, рано или поздно можно получить верное решение задачи.

Генетические алгоритмы обязаны своим появлением эволюционной теории (отсюда и характерные термины: популяция, гены, родители-потомки, скрещивание, мутация). Таким образом, существует возможность создания таких нейронных сетей, которые ранее не изучались исследователями (или не поддаются аналитическому изучению).

з) Реализация уровня ввода и вывода в системе SAS

Эта система, выполненная с использованием технологии нейронных сетей, предназначена не только для распознавания, но и для синтеза речи

Блок-схема системы SAS, соответствующая уровню ввода/вывода, показана на рисунке 2.

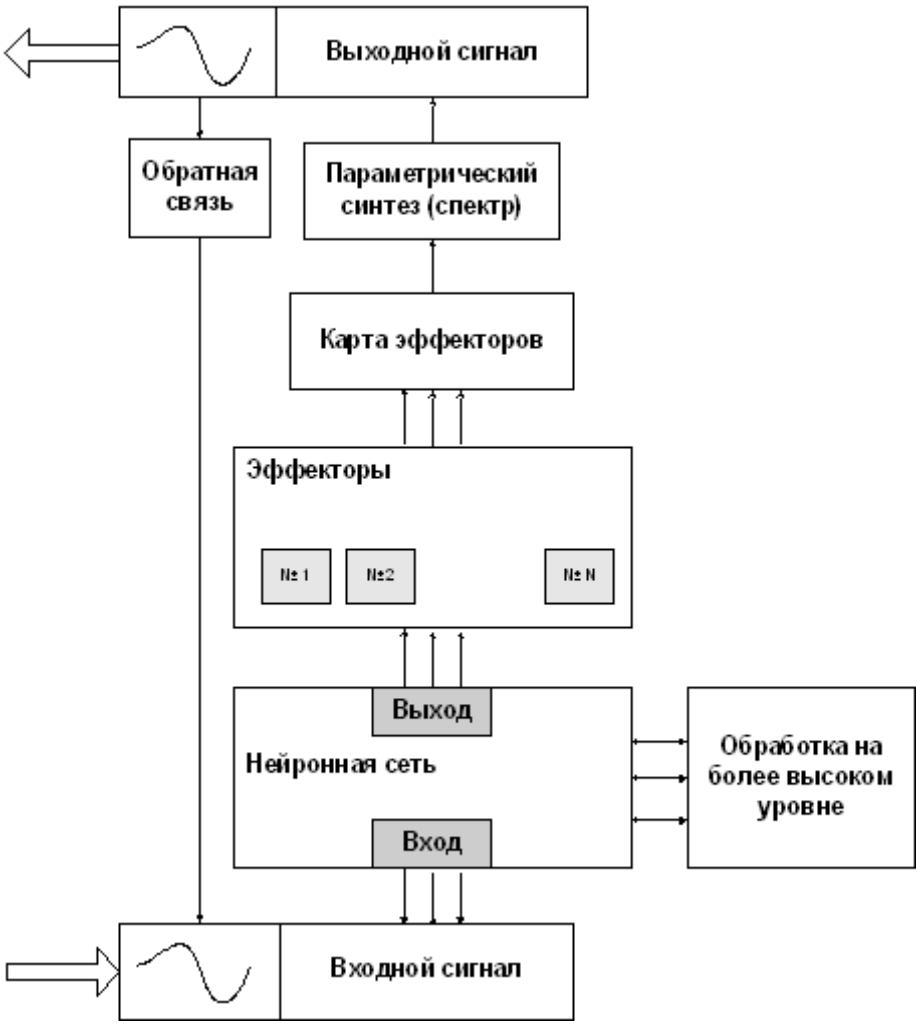

Рисунок 2-Блок-схема уровня ввода/вывода

При распознавании речи система SAS осуществляет ввод звуковой информации, предварительную обработку, получение энергетического спектра и выделение примитивов речи.

При синтезе речи осуществляется выделение из нейронной сети запомненного примитива, синтез спектра (частотный параметрический синтез) и преобразование спектра в звуковой сигнал. При обучении последовательным

повторением двух вышеописанных процедур осуществляется запоминание примитивов речи в нейронной сети.

Процесс ввода звука:

На рисунке 3 изображен процесс ввода звука в системе SAS.

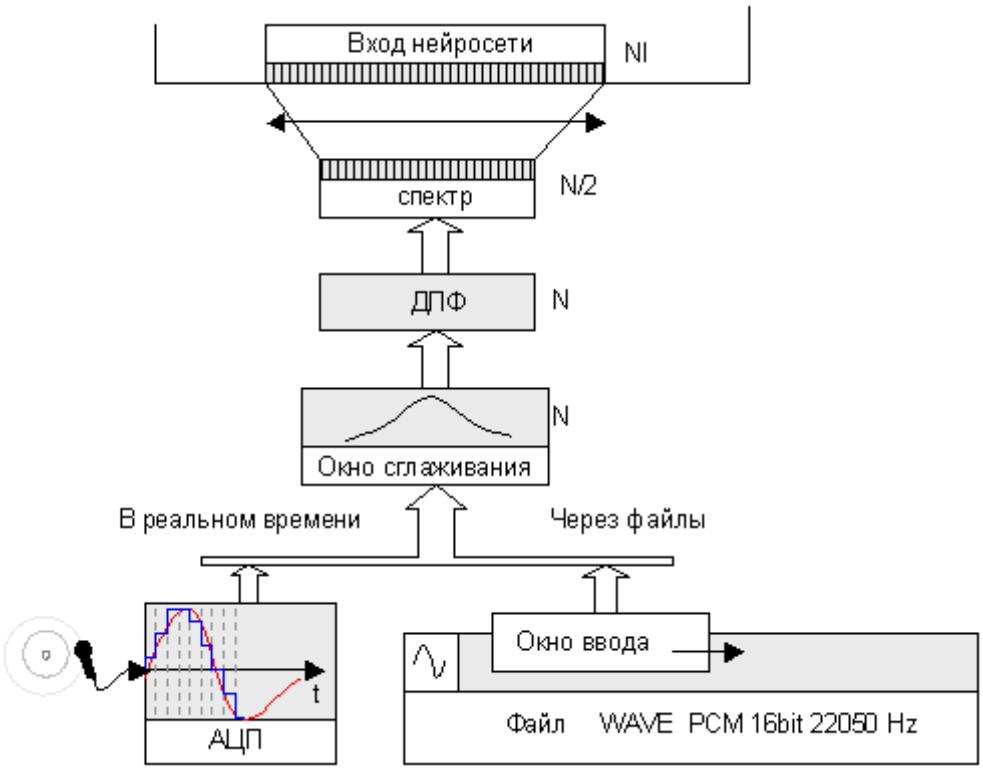

Рисунок 3 - Процесс ввода звука в системе SAS

Ввод звука осуществляется в реальном времени через звуковую карту или через файлы формата WAV в кодировке PCM (разрядность 16 бит, частота дискретизации 22 050 Гц). Работа с файлами была предусмотрена, чтобы облегчить многократное повторение обработки нейронной сети, что особенно важно при обучении.

Предварительная обработка звука:

Согласно рисунку 3, звуковые сигналы, полученные в реальном времени или введенные из файлов формата WAV, подвергаются в системе SAS-предварительной обработке.

При обработке файла по нему перемещается окно ввода, размер которого Равен N элементов – размеру окна дискретного преобразования Фурье (ДПФ).

Смещение окна относительно предыдущего положения можно регулировать. В каждом положении окна оно заполняется 16-разрядными данными (система работает только с такими звуковыми данными, в которых каждый отсчет кодируется 16 битами).

После ввода данных в окно перед вычислением ДПФ на него накладывается окно сглаживания Хэмминга:

$$X'[i] = X[i](0.54 - 0.46\cos\left(\frac{2\pi i}{N-1}\right)) \quad (1)$$

Здесь X[i] — исходный массив данных, X’[i] — массив данных, полученный после наложения окна сглаживания, N — размер ДПФ.

Наложение окна Хэмминга немного понижает контрастность спектра, но позволяет убрать боковые лепестки резких частот, при этом особенно хорошо проявляется гармонический состав речи. Сказанное иллюстрирует рисунок 4.

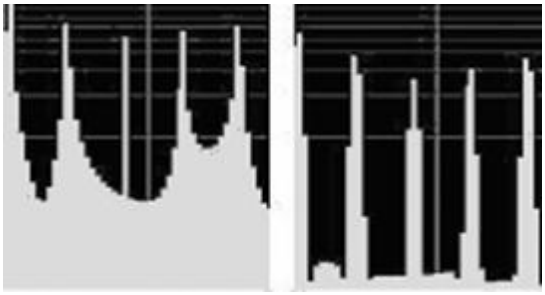

Рисунок 4 - Действие окна сглаживания Хэмминга (логарифмический масштаб)

Выполнение дискретного преобразования Фурье:

Результат сглаживания Хэмминга подвергается в системе SAS дискретному преобразованию Фурье по алгоритму быстрого преобразования Фурье. В результате этого преобразования получается амплитудный спектр и информация о фазе сигнала (в реальных и мнимых коэффициентах).

Информация о фазе сигнала отбрасывается и вычисляется энергетический спектр:

$$E[i] = \sqrt{\mathrm{Re}\, C[i] \cdot \mathrm{Im}\, C[i]},$$
$$i = 0 \ldots NS - 1, NS = N / 2 \quad (2)$$

Здесь E[i] – энергии частот.

Так как звуковые данные не содержат мнимой части, то по свойству ДПФ результат получается симметричным, т.е. E[i] = E[N-i]. Таким образом, размер информативной части спектра NS равен N/2.

и) Нормирование частотного спектра

Так как все вычисления в нейронных сетях производятся над числами с плавающей точкой, значения параметров объектов, классифицируемых с помощью нейронных сетей, ограничены диапазоном [0.0,1.0].

Для выполнения обработки спектра нейронной сетью в системе SAS полученный спектр нормируется на 1.0. Для этого каждый компонент вектора делится на его длину:

$$E'[i] = E[i]/(|E|) \quad (3)$$

$$|E| = \sum_{i=0}^{NS-1} E[i] \quad (4)$$

к) Логарифмическое сжатие спектра

Исследования показали, что информативность различных частей спектра неодинакова: в низкочастотной области спектра содержится больше информации, чем в высокочастотной области спектра.

Поэтому для более экономного использования входов нейронной сети и увеличения необходимо уменьшить число элементов, получающих информацию из высокочастотной области спектра. Это и означает сжатие высокочастотной области спектра в пространстве частот.

В системе SAS применен наиболее распространенный и простой метод — логарифмическое сжатие, или mel-сжатие.

Вот формула, по которой выполняется логарифмическое сжатие спектра:

$$m = 1125 \cdot \log(0.0016 f + 1) \quad (5)$$

Здесь f —частота в спектре, Гц, m — частота в новом сжатом частотном пространстве.

Рисунок 5 иллюстрирует процесс логарифмического сжатия частотного спектра.

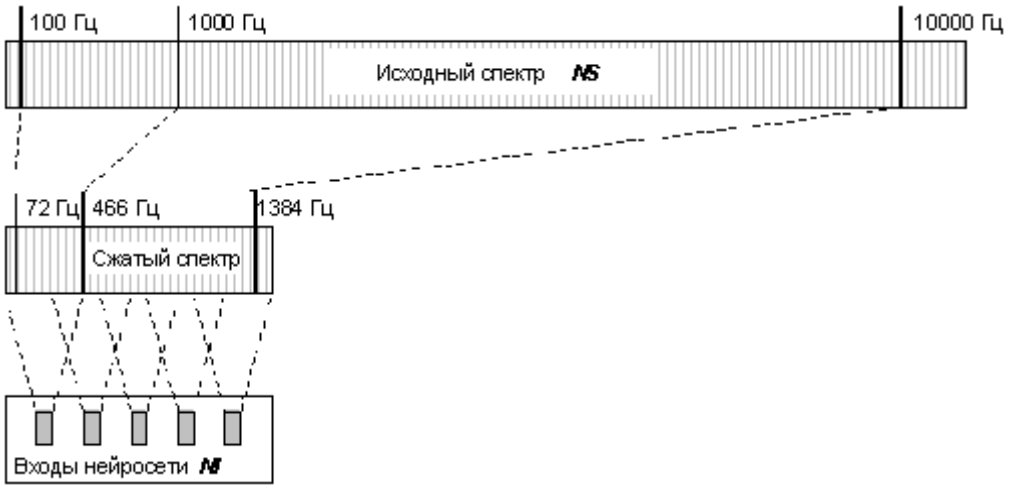

Рисунок 5 - Нелинейное преобразование спектра в пространстве частот

л) Применение вейвлет-преобразований

В только что описанной системе SAS для выделения из речи синтаксических элементов применялось быстрое преобразование Фурье.

Однако, как отмечают исследователи, анализ Фурье обладает целым рядом недостатков, в результате которых происходит потеря информации о временных характеристиках обрабатываемых сигналов. Этот анализ подразумевает использование искусственных приемов, с помощью которых осуществляется частотно-временная локализация, например, окон данных (на рисунке 3 это окно обозначено как Окно ввода).

В современных технологиях обработки и распознавания сигналов применяются так называемые вейвлет-преобразования и вейвлет-анализ.

Термин вейвлет (wavelets) можно перевести как «маленькая волна». Вейвлеты представляют собой новый инструмент решения различных задач прикладной математики. Вейвлет-анализ, детальное знакомство с которым требует определенных познаний в математике, лишен недостатков анализа

Фурье. Он позволяет достичь неплохих результатов при использовании в системах распознавания речи.

Отличие анализа Фурье от вейвлет-анализа заключается в следующем:

Фурье-анализ предполагает разложение исходной периодической функции в ряд, в результате чего исходная функция может быть представлена в виде суперпозиции синусоидальных волн различной частоты. Такая суперпозиция и есть спектр сигнала.

Что же касается вейвлет-анализа, то здесь входной сигнал раскладывается в базис функций, характеризующих как частоту, так и время. Поэтому с помощью вейвлетов можно анализировать свойства сигнала одновременно и в физическом пространстве (время, координата), и в частотном пространстве. Чтобы подчеркнуть такое обстоятельство, в зарубежной литературе Фурье-анализ называют single spectrum, а спектры, полученные на основе вейвлет-преобразований — itime-scale spectrum, или wavelet spectrum.

Функции-базисы для вейвлетных преобразований конструируются на основе производных функций Гаусса.

На рисунке 6, показаны наиболее часто используемые вейвлеты.

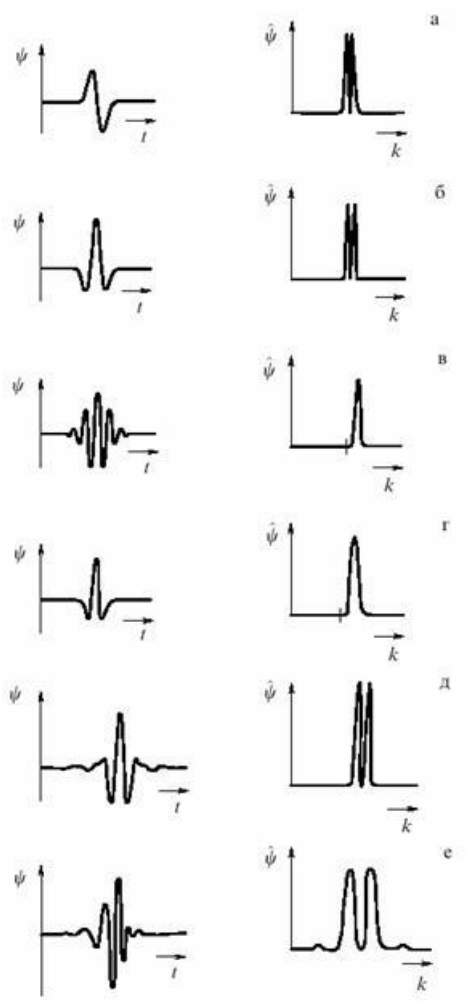

Рисунок 6 - Часто используемые вейвлеты

Эти функции имеют свои названия, которые приведены в таблице 1

Таблица 1 - Часто используемые вейвлеты

Обозначение на рисунке 6	Название
а	WAWE-вейвлет
б	МНАТ-вейвлет. Получил свое название от «мексиканская шляпа, сомбреро» (Mexican Hat)
в	Morlet
г	Paul
д	LMB
е	Daubeshies

При использовании вейвлет-преобразований для распознавания речи разработчик должен выбрать нужную функцию. От правильного выбора зависит успешность распознавания.

Анализ существующих методов.

Существуют два подхода к распознаванию речи. Первый подход реализует распознавание элементов речи по образцу и применяется в различные рода системах голосового управления. Второй подход основан на выделении в речи лексических элементов — фонем, аллофонов, морфем и т.д. Этот подход пригоден для создания систем диктовки текста, рассмотренных нами в следующей главе.

Выделяются системы распознавания речи, требующие обучения и зависящие от диктора, а также системы, способные работать без предварительного обучения и, следовательно, не зависимые от диктора.

Перед тем как приступить к выделению из речи лексических элементов, необходимо выполнить предварительную обработку речевого сигнала. В ходе этой обработки из сигнала удаляются шумы, выполняется частотная фильтрация и оцифровка, а также нормализация уровня сигнала.

Были рассмотрены две методики выделения из речи лексических элементов.

Первая методика предполагает использование дискретного преобразования Фурье. Непосредственно лексические элементы выделяются из оцифрованной речи при помощи нейронной сети, способной к обучению. При этом речь представляется в виде некоторого набора числовых параметров, так как нейронные сети работают именно с наборами таких параметров.

Вторая методика выделения лексических элементов речи, основана на применении вейвлет-преобразований. В отличие от дискретного преобразования Фурье, этот метод исключает потерю информации о временных характеристиках обрабатываемых сигналов. Мы отметили, что при использовании вейвлет-преобразований входной сигнал раскладывается не в

базисе периодических функций (как в дискретном преобразовании Фурье), а в базисе функций, характеризующих как частоту, так и время.

1.3 Голосовое управление автомобилем: значение, возможности, применение

Современные автомобили «умнеют» с каждым годом, оснащаясь новыми системами и решениями, призванными сделать вождение максимально безопасным и комфортным. Обилие кнопок и переключателей для управления огромным количеством функций и параметров, делает водительское место похожим на пилотскую кабину. Система голосового управления, установленная в автомобиле, позволяет водителю не отвлекаться от дорожной ситуации на манипуляции с кнопками.

Первая система, осуществляющая голосовое управление, была настолько далека от совершенства, что удобство ее использования вызывало сомнения. Распознавать слова она еще не могла, требуя проговаривать команду по буквам, вдобавок невысокое качество микрофона и несовершенство алгоритмов шумоочистки не всегда гарантировали правильное определение произносимого, голосовое управление оставляло желать лучшего.

Отсутствие бортового компьютера, интерфейса взаимодействия у многих систем автомобиля, не позволяло управлять ими голосом. Единственный способ что-то им «сообщить» – протянуть руку и нажать/повернуть.

Современная система значительно расширила свой потенциал. Разработчики стремятся максимально приблизить формат взаимодействия к естественному диалогу. Разговорный язык, диалекты, посторонние шумы, быстрота произношения, особенности дикции человека и даже

нестандартные формулировки команд все меньше становятся препятствием для качественного функционирования.

Перечень функций, «подвластных голосу», постоянно растет. Сегодня таким способом можно управлять:

- климатом;
- мультимедийной системой. Помимо уже привычного перелистывания музыкальных треков, радиостанций и TV-каналов, можно слушать чтение текстовых книг, что поможет скрасить времяпровождение в пробках;
- перемещением по спискам меню;
- телефоном. К поиску контактов добавилась возможность прослушивать и создавать сообщения электронной почты и SMS. Надо только синхронизировать телефон с автомобилем посредством Bluetooth или кабеля USB;
- параметрами бортового компьютера, в том числе прослушивать его сообщения;
- навигацией;
- электронной почтой;
- санкционированным доступом к вождению автомобиля по распознаванию голоса человека, его индивидуальной биометрике.

Этот список можно долго продолжать.

Стремление к естественности диалогового общения прослеживается на каждом уровне. Помимо прочего, это:

1. Правильная интерпретация. Система способна расшифровывать сокращения, «знает» многие аббревиатуры, форматы даты и чисел, и так далее, …даже ненормативную лексику;
2. Естественность воспроизводимой речи. Текст, произносимый модулем голосового управления, построен корректно, с правильными ударениями и звучит естественно;

3. Удобство взаимодействия. После успешного распознавания команды подается соответствующий сигнал;
4. Режим постоянного отслеживания команд. Обычно перед подачей команды требуется нажать на специальную кнопку, расположенную на руле, но в некоторых новых реализациях, голосовое управление имеет режим непрерывного прослушивания, и может выделять команды без нажатия на кнопку.

Сюда же можно добавить, что постепенно увеличивается и количество понимаемых языков и диалектов.

ГЛАВА 2 ТЕОРЕТИЧЕСКИЙ ОБЗОР ВОЗМОЖНОСТИ РЕАЛИЗАЦИИ СИСТЕМЫ ГОЛОСОВОГО УПРАВЛЕНИЯ НА ОСНОВЕ СУБПОЛОСНОГО АНАЛИЗА

Существующие системы понимания речи пока еще значительно уступают речевым способностям человека, что свидетельствует об их недостаточной адекватности и ограничивает применение речевых технологий в промышленности и быту. Из имеющихся программных продуктов рынка систем распознавания речи лишь немногие поддерживают русский язык.

В настоящее время актуальным является разработка методов автоматического анализа устной речи на компьютерах, одним из направлений которых является совершенствование человеко-машинных интерфейсов. Существующие системы голосового управления как правило основываются на нейронных сетях, и уже доказали свою эффективность, однако при этом, они не учитывая распределение долей энергии по частотным интервалам. Распределение долей энергии является важным параметром в обработке речевых данных, по их изменению от интервала к интервалу можно определить границу звука, или же найти особенности, присущие лишь данному диктору.

2.1 Основы субполосного анализа речевых сигналов

Предполагается, что РС представлены эквидистантными отсчетами

$$x_i = x(i\Delta t), i = 1,2,...,$$

с частотой дискретизации:

$$f_Д = 1/\Delta t \geq 8000 Гц.$$

Известно, что все звуки русской речи обладают свойствами концентрации энергии в пределах малой доли частоты дискретизации. Поэтому адекватным подходом к обработке РС является применение субполосного анализа, когда их характеристики соотносятся некоторым разбиением области нормированных частот

$$-\pi \le \omega \le \pi \,, \tag{6}$$

на частотные интервалы:

$$\Omega_r = [-\Omega_{2r}, -\Omega_{1r}] \vee [\Omega_{1r}, \Omega_{2r}], \tag{7}$$

где $r = 1,..., R;$;

$$\Omega_{2r} > \Omega_{1r}; \Omega_{2r} \le \pi; \tag{8}$$

В соответствии с конечной длительностью звуков речи анализу должны подвергаться конечные наборы отсчетов РС (векторы)

$$\vec{x} = (x_1,..., x_N)^T \tag{9}$$

где Т – символ транспонирования.

Положим

$$X_N(\omega) = \sum_{i=1}^{N} x_i e^{-j\omega(i-1)} \tag{10}$$

Имеет место формула обращения

$$x_k = \int_{-\pi}^{\pi} X_N(\omega) e^{i\omega(k-1)} d\omega / 2\pi, \tag{11}$$

и справедлива формула Парсеваля

$$g_{xy} = (\vec{x}_N \vec{y}_N) = \sum_{i=1}^{M} x_i y_i = \int_{-\pi}^{\pi} X_N(\omega) Y_N^*(\omega) d\omega / 2\pi, \tag{12}$$

где звездочка означает комплексное сопряжение, $\vec{y}_N = (\vec{y}_1,..., \vec{y}_N)^T$ а .

В частности, имея в виду частотные интервалы (7), соотношение (12) можно переписать в виде суммы:

$$(\vec{x}_N, \vec{y}_N) = \sum_{i=1}^{R} G_r(\vec{x}_N, \vec{y}_N), \tag{13}$$

слагаемые которой

$$G_r(\vec{x}_N, \vec{y}_N) = \int_{\omega \in \Omega r} X_N(\omega) Y_N^*(\omega) d\omega / 2\pi, \tag{14}$$

естественно называть субполосными корреляциями.

Кроме того, можно ввести понятие частей энергии, попадающих в частотные интервалы

$$\left\|\vec{x}_N\right\|^2 = \sum_{r=1}^{R} P_r(\vec{x}_N); \qquad \left\|\vec{y}_N\right\|^2 = \sum_{r=1}^{R} P_r(\vec{y}_N); \tag{15}$$

где

$$P_r(\vec{z}_N) = \int_{\omega \in \Omega r} \left|Z_N(\omega)\right|^2 d\omega / 2\pi, \tag{16}$$

где $\vec{Z}_N(\omega)$ – трансформанта Фурье вектора $\vec{z}_N$, и понятие субполосных нормированных корреляций

$$\rho_r(\vec{x}_N, \vec{y}_N) = \frac{G_r(\vec{x}_N, \vec{y}_N)}{\sqrt{P_r(\vec{x}_N) P_r(\vec{y}_N)}}, \tag{17}$$

Они, очевидно, удовлетворяют неравенству

$$\left|\rho_r(\vec{x}_N, \vec{y}_N)\right| \le 0. \tag{18}$$

Примечательно, что для вычисления характеристик (16) и (17) нет необходимости переходить в частотную область, т.к. подстановка в представление (14) и (16) определений вида (10) позволяет получить реализуемые непосредственно во временной области билинейные и квадратичные формы

$$G_r(\vec{x}_N, \vec{y}_N) = \vec{x}_N^T A_r \vec{y}_N, \tag{19}$$

$$P_r(\vec{z}_N) = \vec{z}_N^T A_r \vec{z}_N, \tag{20}$$

где Ar субполосная матрица с элементами.

$$a_{ik}^{r} = \{\sin(\Omega_{2r}(i-k)) - \sin(\Omega_{1r}(i-k))\} / \pi(i-k), \tag{21}$$

Отметим, что соотношения (19) и (20) позволяют вычислить точные значения частей энергии отрезков сигналов, приходящихся на заданный частотный интервал, и соответствующих субполосных корреляций.

2.1.1 Селекция пауз между звуками речи

Известно, что энергия порождаемых звуками отрезков речевых сигналов распределены в совокупности достаточно узких частотных интервалов, верхняя граница которых не превышает 7 килогерц, а суммарная ширина не превышает половины этого частотного диапазона. В свою очередь, посторонние шумы при отсутствии речи имеют иное частотное распределение энергии, что позволяет на этой основе построить решающие процедуры селекции пауз.

Исходная (нулевая) гипотеза формируется следующим образом.

H0 : отрезок $\vec{x}_N$ сигнала зарегистрирован в паузе речи так что

$$\vec{x}_N = \vec{u}_N \tag{22}$$

где – $\vec{u}_N = (u_1,..., u_N)^T$ вектор отрезков шумов.

Альтернатива имеет следующую формулировку.

H1 : хотя бы часть отчетов зарегистрирована в присутствии речевого воздействия, которые аддитивно взаимодействуют с шумом, то есть

$$\vec{x}_N = \vec{z}_N + \vec{u}_N, \tag{23}$$

где - $\vec{z}_N = (\vec{z}_1,..., \vec{z}_N)^T$ вектор отчетов возбуждаемых речью, часть из которых может быть равна нулю.

В качестве решающей функции предлагается использовать:

$$F_r(\vec{x}) = \frac{P_r(\vec{x}_N)}{E[P_r(\vec{u})]}, \tag{24}$$

где E – символ математического ожидания.

Гипотеза H0 отвергается при выполнении следующего неравенства

$$\max F_r(\vec{x}) > h_\alpha \tag{25}$$

где максимум определяется для всех частотных интервалов, а h_α – некоторый порог.

Предполагается, что имеется возможность предварительного обучения, на этапе которого при заведомом отсутствии речи можно определить оценки $\overline{P}_r(\vec{u}_N)$ математических ожиданий частей энергий шумов и оценку величины порога в (2.4), удовлетворяющую условию:

$$PR\left\{P_{r\max}(\vec{u}_N) / \overline{P}_r(\vec{u}_N) \geq h_\alpha\right\} \leq \vec{\alpha}, \tag{26}$$

где PR – оценка вероятности, α – желаемый уровень вероятности ошибок первого рода, а $\vec{\alpha}$ его оценка при использовании оценок математических ожиданий.

Оценивание математических ожиданий и порога можно осуществить по одному и тому же достаточно большому количеству отчетов шумов, при отсутствии речи, например по 10 000 отрезкам необходимой длительности (порядка 1,5 секунд).

Предполагая для простоты, что шумы в паузах, являются гауссовыми с независимыми отсчетами, причем

$$E[u_t] = 0, \tag{27}$$

$$\sigma_0^2 = E[u_t^2], \tag{28}$$

можно показать справедливость следующих соотношений

$$m_r = E[P_r(\vec{u})] = \sigma_u^2 N(\Delta\Omega_r / \pi)^2, \tag{29}$$

$$\sigma_r^2 = E\left[(P_r(\vec{u}) - m_r)^2\right] = 2N\sigma_u^4(\Delta\Omega_r / \pi)^2, \tag{30}$$

где $\Delta\Omega_r = \Omega_{2r} - \Omega_{1r}$

Таким образом, имеет место

$$\gamma_r = \sigma_r / m_r = 2/(N)^{1/2}, \tag{31}$$

То есть в отсутствии сигнала дисперсия решающей функции (24) обратно пропорциональна длительности обрабатываемого отрезка, а ее математическое ожидание равно единице. В виду нестационарности речевых воздействий исследовать мощность критерия (25) (левая часть) не представляется возможным. Отметим только, что использование максимального значения решающей функции, по крайней мере, в случае белого шума, позволяет в среднем эффективно отреагировать на появления дополнительной энергии, которая сосредоточена в малой доле частотной полосы.

2.1.2 Селекция вокализованных звуков русской речи и оценка периода основного тона

В речевом сигнале можно выделить два основных типа структур: периодическая (вокализованные звуки) и шум (невокализованные). Периодическая структура возникает при использовании в артикуляции источника колебаний – голосовых связок. Шумовые источники подразделяются на турбулентные и импульсные. Кроме периодического и шумового типа звуков, существуют звуки, в образовании которых принимают участие шумовой источник и голосовые связки. Одним из наиболее значимых параметров периодического сигнала является частота основного тона. Оценка частоты основного тона выполняется по вокализованной компоненте.

Селекция вокализованных звуков русской речи и оценка периода основного тона. Вокализованные звуки являются периодическими, что может быть положено в основу их селекции. При этом необходимо оценивать период основного тона, если отвергается основная гипотеза:

$$\vec{x} = (x_1, \ldots, x_N)^T$$

H0 - анализируемый отрезок порожден невокализованным звуком речи.

В настоящее время существует два основных подхода к определению периода основного тона (ОТ): на основе анализа спектров и корреляционном анализе.

Спектральный метод заключается в определении в диапазоне априори допустимых значений частоты ОТ такой из них, в которой наблюдается максимум модуля спектра Фурье. Такая частота при выполнении некоторых условий принимается за частоту основного тона.

Основной недостаток спектрального оценивания заключается в следующем. Пусть последовательльность отсчетов сигнала $(x_1, x_2, ..)$ имеет периодический характер так что

$$x_{i+km} = x_i, k = 0,1... \tag{32}$$

Тогда соответствующая трансформанта Фурье (спектр)

$$X(\omega) = \sum_{i=1}^{CM} x_i e^{-j\omega(i-1)};$$

может быть представлена в виде

$$X(\omega) = \sum_{k=1}^{C} e^{-j\omega M(k-1)} \sum_{i=1}^{M} x_i e^{-j\omega(i-1)}; \tag{33}$$

Таким образом

$$\left|X(\omega)\right|^2 = \left|X_p(\omega)\right|^2 * \sin^2(CM\omega/2)/\sin^2(M\omega/2), \tag{34}$$

$X_p(\omega)$ -внутриперiодный спектр сигнала (вторая сумма в (33)). Легко понять, что первый сомножитель будет достигать максимального значения в следующих точках оси частот:

$$\omega_m = m2\pi / M, m = 1,2,..... \tag{35}$$

причем именно значение $2\pi / M, (m = 1)$ соответствует частоте основного тона. Однако влияние $\left|X_p(\omega)\right|^2$ может проявляться в том, что максимум правой части будет соответствовать другому значению m. Именно

это не позволяет методически надежно определять период основного тона по спектру анализируемого отрезка сигнала. В основе корреляционного метода определения периода основного тона используется характеристика

$$\rho_{\tau,N} = \sum_{i=1}^{N} x_i x_{i+\tau} / \sqrt{\sum_{i=1}^{N} x_i^2 \sum_{k=1}^{N} x_{k+\tau}^2}, \tag{36}$$

которая является оценкой нормированного коэффициента корреляции. Гипотеза H0 отвергается при выполнении неравенства

$$\max_{0\le\tau\le L} \rho_{\tau,N} = h \in (0{,}7 \div 1), \tag{37}$$

где L определяется отношением частоты дискретизации к минимально возможной частоте основного тона.

Тогда в качестве оценки периода основного тона принимается

$$M = \arg\max \rho_{\tau,N}, 0 \le \tau \le L. \tag{38}$$

Иными словами, максимальное значение характеристики (39) должно превышать некоторый порог, что принимается за признак почти периодического поведения отрезков сигнала.

Одним из недостатков такого подхода является присутствие искажающих шумов, что маскирует наличие периодичности в сигнале.

Кроме того, концентрация спектра $|X(\omega)|^2$ вблизи частоты, не совпадающей $2\pi / M$, с приводит к тому, что максимальное значение (36) будет достигаться при меньшем, чем длина интервала между возбуждающими гортань воздействиями.

На рисунке 7 представлены фрагмент РС, порожденного звуком «и», и значения нормированных коэффициентов корреляции при различных значения смещения

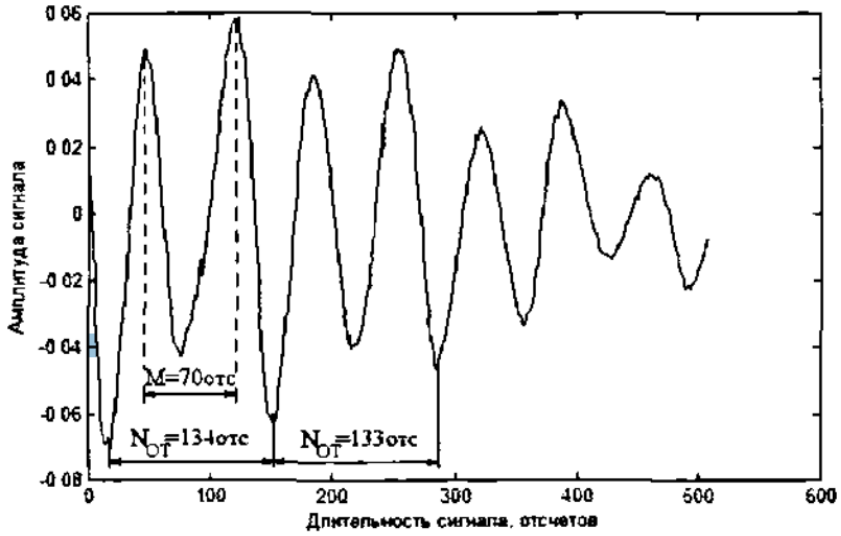

Рисунок 7 - Фрагмент речевого сигнала, порожденного звуком «и» (f_d =16кГц)

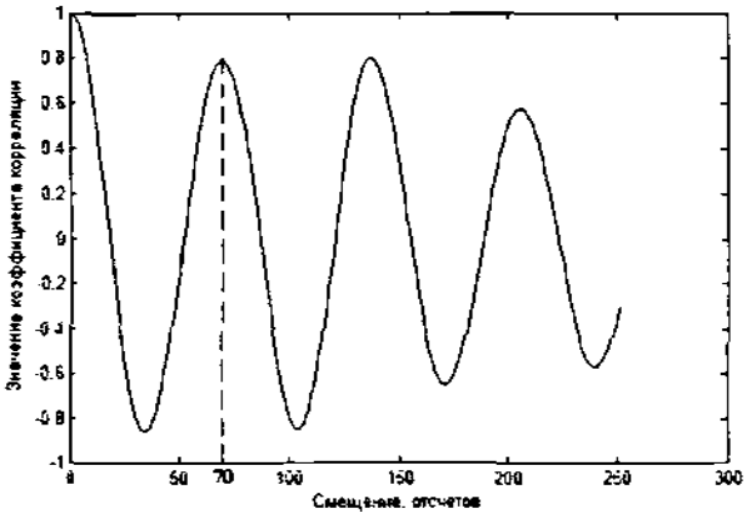

Рисунок 8 - Значения нормированных коэффициентов корреляции (N=256, f_d =16кГц)

На рисунке 7 N_{OT} - «истинный» период основного тона РС, а М - его оценка при использовании корреляционного метода. На рисунке 8 пунктиром отмечен максимум характеристики (36). Видно, что максимальное значение характеристики (36) достигается при значении в 2 раза меньшем, чем величина периода основного тона.

Таким образом, необходимо использовать иные методы определения периода основного тона, устойчивые как к воздействию шумов, так и к

влиянию периодичности сигнала между двумя последовательными возбуждающими гортань воздействиями.

Представляется естественным ориентироваться на поиск наименьшей частоты из набора (35).

Для этого в работе введено понятие нормированной субполосной корреляции (НСК)

$$\varphi_t^n = \Phi_t^n / \sqrt{P_n(\vec{x}_1) P_n(\vec{x}_\tau)}, \tag{39}$$

где

$$\Phi_t^n = \int\limits_{\omega \in V_n} X_1(\omega) X_\tau^*(\omega) d\omega / 2\pi, \tag{40}$$

$$X_1(\omega) = \sum_{i=1}^{N} x_i e^{-j\omega(i-1)}, X_\tau(\omega) = \sum_{i=1}^{N} x_{i+\tau} e^{-j\omega(i-1)}, \tag{41}$$

$$\vec{x}_1 = (x_1, .. x_2 .., x_N)^T, \vec{x}_r = (x_{1+r}, .. x_{2+r} .., x_{N+r})^T, \tag{42}$$

$$V_n = [-V_{2n}, -V_{1n}) \cup [V_{1n}, V_{2n}), \tag{43}$$

$$P_n(\vec{y}) = \int\limits_{\omega \in V_n} |Y(\omega)|^2 d\omega / 2\pi, \tag{44}$$

$$Y(\omega) = \sum_{i=1}^{N} y_i e^{-j\omega(i-1)}, \tag{45}$$

Наличие в (46) знаменателя позволяет обеспечить чувствительность к частотным полосам с малым уровнем энергии. Подставляя в соотношение (39) и (40) определения (44) и (45), нетрудно получить представления для НСК непосредственно во временной области

$$\varphi_t^n = \vec{x}_1^T A_n \vec{x}_\tau / \sqrt{\vec{x}_1^T A_n \vec{x}_1 \vec{x}_\tau^T A_n \vec{x}_\tau}, \tag{46}$$

где - $A_n = \{a_{ik}^n\}$ субполосные матрицы.

Очевидно, что

$$\varphi_t^n \le 1, \tag{47}$$

причем правая часть достигается только при выполнении условия пропорциональности;

$$\vec{x}_\tau \le c\vec{x}_1, \tag{48}$$

Таким образом, определение периода основного тона сводится к вычислению характеристик (39) при разных значениях τ и ω_n.

Если условие

$$\max_{0\le\tau\le L} \varphi_\tau^n = h_\varphi \tag{49}$$

для некоторого частотного интервала выполняется, то принимается решение, что анализируемый отрезок РС относится к вокализованному звуку. Рекомендуется использовать частотный интервал n=1.

Для установления h_φ порога проведены вычислительные эксперименты, которые показали, что наименьшая вероятность ошибок в селекции вокализованных звуков русской речи обеспечивается при величине порогового значения

$$h_\varphi = 0{,}8. \tag{50}$$

Использование z-преобразования Фишера вида

$$z_r^n = 1/2\ln((1+\varphi_\tau^n)/(1-\varphi_\tau^n)), \tag{51}$$

позволяет повысить чувствительность НСК в области значений φ_τ^n близких к максимальному, что способствует более точному оцениванию периода ОТ. При этом рекомендуется использовать соотношение

$$z_r = \sum_{n=0}^{N/2-1} k_n * z_\tau^n; \tag{52}$$

где $k_n = Pd_n(\vec{x}_1)$ - весовой коэффициент, представляющий собой долю энергии.

Тогда в качестве оценки периода основного тона принимается значение

$$M = \arg\max z_r, 0 \le \tau \le L. \tag{53}$$

если выполнено условие

$$z_M \geq h_z = 1{,}1. \tag{54}$$

Указанное h_z значение получено при подстановке (50) в (51) с учетом свойства

$$\sum_{n=0}^{N/2-1} k_n = 1.$$

Вычисление значения числителя НСК вида (46) ускоряется за счет использования представления вида

$$\Phi_t^n(\vec{x}_1, \vec{x}_\tau) = \sum_{k=1}^{J/2} \lambda_k^r (\alpha_{2kr}(\vec{x}_1)\alpha_{2kr}(\vec{x}_\tau) + (\alpha_{(2k-1)r}(\vec{x}_1)\alpha_{(2k-1)r}(\vec{x}_\tau)), \tag{55}$$

где - $\alpha_{kr}(\vec{x}_1), \alpha_{kr}(\vec{x}_\tau)$ скалярные произведения (проекции) вида.

На рисунках (9) и (10) представлен фрагмент РС, порожденного звуком «и» и результат оценки на основе соотношений (52), (53) для этого фрагмента.

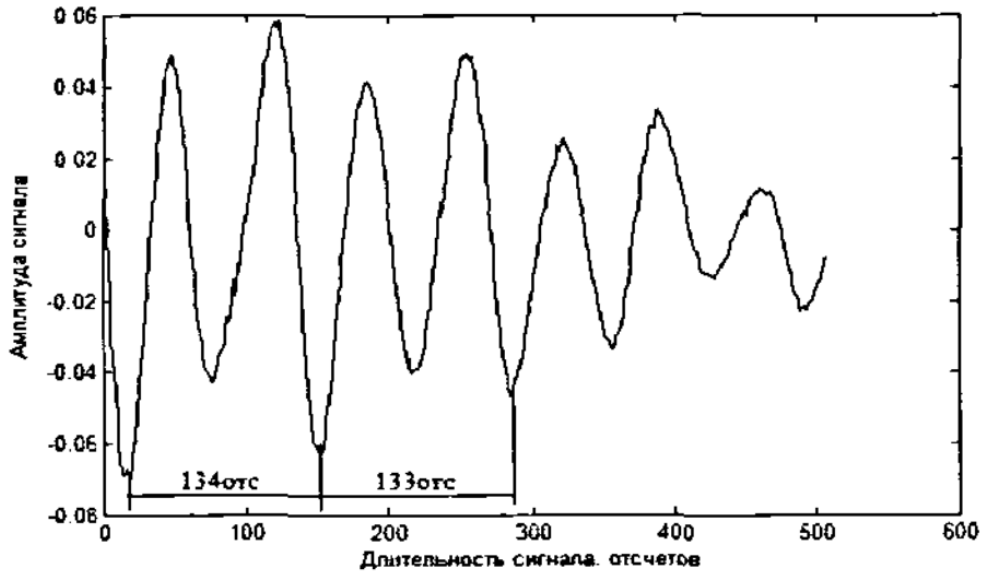

Рисунок 9 - Фрагмент речевого сигнала, порожденного звуком «и» (f_d =16кГц)

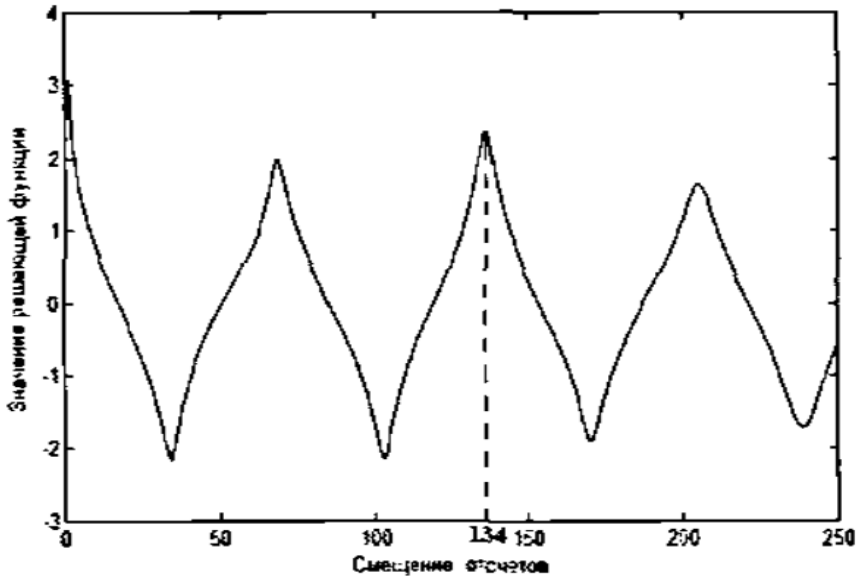

Рисунок 10 – Значения РФ НСК (N=256, f_d =16кГц)

Из рисунка 10 видно, что максимум РФ достигается при сдвиге равном периоду основного тона, в отличие от корреляционного метода оценивания периода основного тона.

Для оценки эффективности предлагаемой решающей процедуры были оценены вероятности ошибок первого и второго рода:

$$P_{1ош} = 1 - N_{авт.невокал} / N_{невокал} \tag{56}$$

$$P_{2ош} = 1 - N_{авт.вокал} / N_{вокал} \tag{57}$$

а также погрешность оценки периода основного тона:

$$\sigma^{zv} = \sqrt{\sum_{i=1}^{N_{zv}} (M_i - N_{OT})^2 / \sum_{i=1}^{N_{zv}} N_{OT}^2} \tag{58}$$

Здесь $N_{авт.невокал}$ - количество правильно определенных невокализованных отрезков; $N_{авт.вокал}$ - количество правильно определенных вокализованных отрезков; $N_{вокал}$ - количество отрезков, соответствующих вокализованных звукам речи, использованных для анализа (66000 отрезков); - количество

$N_{невокал}$ отрезков, соответствующих невокализованных звукам речи, использованных для анализа (45000 отрезков); N_{zv} - количество отрезков анализа одного звука; M_i - значение периода основного тона, определенное автоматически; N_{OT} - значение периода основного тона, определенное в ручную.

2.2 Голосовое управление в существующих системах

Главный модуль голосового управления распознает речевые команды, преобразует их в соответствующие сигналы, которые передаются системам автомобиля, выполняющим требуемое действие. Как правило, микрофон встроен в салонное зеркало заднего вида.

Рисунок 11 – Система голосового управления в автомобиле

Однако конкретная реализация зависит от установки: штатная это система или доустановленная, ведь речевое управление встречается во многих опционах. Ярким примером здесь выступает охранная система Pandora DXL5000. Ее функции по охране дополнены взаимодействием водителя и автомобиля на любом расстоянии с помощью мобильного телефона. В память сохраняются образцы всех речевых команд. При поступлении устного

сообщения (звонок с телефона) ищется 100% совпадение. Если оно не найдено, срабатывает охранная блокировка.

Также можно приобрести модуль голосового управления, не предусмотренный изготовителем вашего автомобиля. В этом случае перечень устройств невелик, как правило, это стеклоподъемники, дворники, салонное освещение и наружная оптика, люк, центральный замок и тому подобное. Подключение таких систем не всегда требует в автомобиле наличия CAN или другой какой-либо шины, подсоединяясь к управляемому механизму напрямую.

Владельцы айфонов не остались без внимания. Компания Apple предоставляет им такую возможность благодаря продукту Siri Eyes Free. Теперь можно голосовыми командами пользоваться популярными приложениями iPhone: звонки, сообщения, навигация, медиа, а также информация о погодных условиях, спортивных событиях, курсе валют и прочее. Как уже было сказано, надо лишь подключить смартфон по USB и перед произнесением команды или вопроса нажать специальную кнопку в автомобиле.

Использование Siri Eyes Free в своих авто одобрили BMW, Audi, Toyota, Land Rover, Jaguar, Honda, Mercedes-Benz и другие. В выпускниках конвейеров Ford, Hyundai и Kia голосовые возможности предоставляет компания Microsoft.

Не все производители используют разработку голосового управления сторонних компаний. У многих есть собственные достижения в этой области. Среди последних упомянутых, наиболее развитыми можно назвать Ford Sync, CUE для Cadillac (мультимедийная система собственной разработки с голосовыми возможностями), SDS для Audi, Linguatronic для Mercedes-Benz.

Все они имеют свои преимущества и недостатки. Так, система Ford Sync «владеет» 19-ю языками, включая разновидности французского и английского. CUE распознает произвольные фразы, а не только заранее предусмотренные. Название города или улицы, произнесенное для некоторых версий SDS, должно быть озвучено по буквам.

Многие минусы, выявленные водителями, не столь значительны, чтобы система, понимающая устные команды, оставалась без участия. Безусловно, это удобно, а на этапе знакомства кому-то даже интересно. Привыкнув со временем разговаривать со своим автомобилем, некоторые сразу и не скажут, где находится, например, блок управления климатом.

ГЛАВА 3 СЕГМЕНТАЦИЯ РЕЧЕВЫХ СИГНАЛОВ

Предварительным этапом распознавания речевого сигнала является сегментация речевого сигнала на звуки речи. В результате сегментации речевой сигнал разбивается на участки, каждый их которых соответствует определенному звуку речи. В данной работе будет рассмотрено 3 метода сегментации. Результаты их сегментации будут объединены и прорежены для большей эффективности метода.

3.1 Субполосный метод сегментации (1)

Основным свойством данного метода сегментации является то, что энергия отрезков речевых сигналов сосредоточена в достаточно узких частотных полосах, расположение и состав которых определяется произносимым звуком речи.

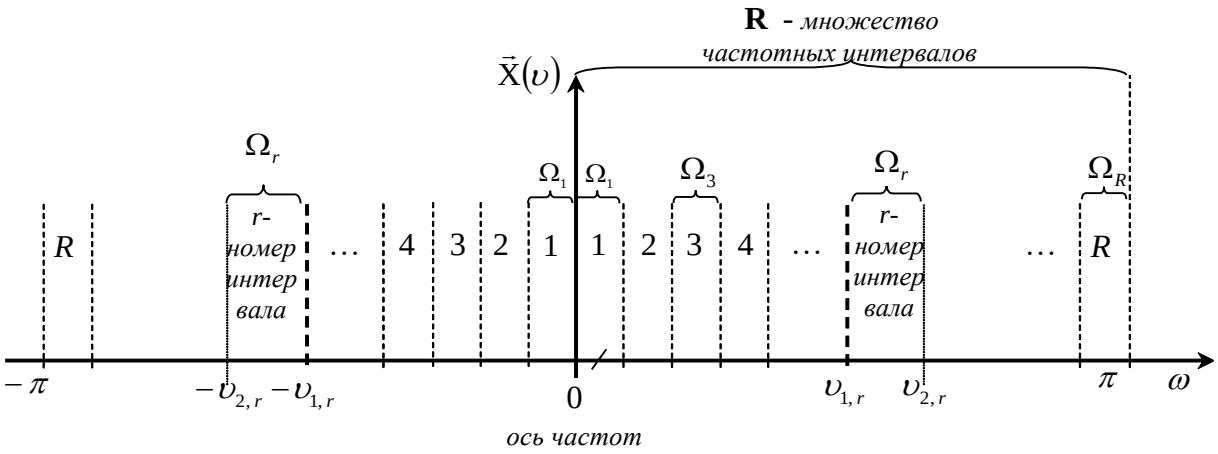

Рисунок 12 – Пример разбиения частотной полосы

Основная суть метода заключается в следующем.

Пусть компоненты вектора представляют собой значения некоторого сигнала (функции времени)

$$\vec{x} = (x_1, \ldots, x_N)^T. \tag{59}$$

Трансформанта Фурье отрезка отсчетов сигнала (вектора), в качестве области определения которой естественно рассматривать (нормированная частота)

$$X(\upsilon) = \sum_{k=1}^{N} x_k e^{-j(k-1)\upsilon}, \tag{60}$$

Из равенства Парсеваля нетрудно получить долю энергии отрезка сигнала, соответствующую частотному интервалу $V = [-\upsilon_2, -\upsilon_1) \cup [\upsilon_1, \upsilon_2)$

$$P_V(\vec{x}) = \frac{1}{2\pi} \int_{\upsilon \in V} |X(\upsilon)|^2 d\upsilon \tag{61}$$

Если в правую часть соотношения (59) подставить определение (60), то в результате преобразований получим

$$P_V(\vec{x}) = \vec{x}^T A_V \vec{x}, \tag{62}$$

где $A_V = \{a_{ik}\}$ – симметричная матрица, элементы которой определяются

$$a_{ik} = \begin{cases} \dfrac{\sin[\upsilon_2(i-k)] - \sin[\upsilon_1(i-k)]}{\pi(i-k)}, & i \neq k \\ \dfrac{\upsilon_2 - \upsilon_1}{\pi}, & i = k \end{cases} \tag{63}$$

Таким образом, долю энергий отрезка сигнала в любом частотном интервале можно вычислить на основе представления (56), не вычисляя при этом соответствующую трансформанту Фурье. Матрицу вида $A_V = \{a_{ik}\}$ естественно называть субполосной матрицей.

В данном методе границы звуков вычисляются следующим образом:

$$\frac{\sum_{r \in R_1} \left(\sqrt{P_r(\vec{x}_1)} - \sqrt{P_r(\vec{x}_2)} \right)^2}{\|\vec{x}\|^2} \leq \frac{B\left(\dfrac{\|\vec{x}\|^2 \cdot \Delta\omega}{\pi} \right)}{\sum_{r \in R_1} P_r(\vec{x})}, \tag{64}$$

г

д

е

EMBED Equation.3 $\vec{x}_1 = (x_{11}, ..., x_{1i}, ..., x_{1N})^T, \vec{x}_2 = (x_{21}, ..., x_{2k}, ..., x_{2N})^T, \vec{x} = [\vec{x}_1{}^T, \vec{x}_2{}^T]^T$ - сигналы

в

$$P_r(\vec{x}) \geq \frac{\|\vec{x}\|^2 \cdot \Delta\omega}{\pi}, \quad (65)$$

B – количество частотных интервалов, удовлетворяющих (65); $\Delta\omega$ - ширина частотного интервала.

Согласно неравенству,(65) определяются частотные интервалы, являющиеся информационными, затем согласно неравенству (64) вычисляются границы звуков.

3.2 Субполосный метод сегментации (2)

В основе данного метода лежит тот же принцип что и в п.3.1, однако границы между звуками определяются следующим образом:

Основная гипотеза H_0: отрезки сигнала $\vec{x}_N = (x_1,..., x_N)^T$ и $\vec{x}_N^N = (x_{N+1},..., x_{2N})^T$ порождены одним и тем же звуком речи .

Положим

$$d_n(\vec{z}_N) = P_r(\vec{z}_N) / \|\vec{z}_N\|^2 \quad (66)$$

И введем на основе этих долей энергии субполосного расстояния

$$V_n = (\sum_{r=1}^{R}((d_r(\vec{x}_N))^{1/2} - (d_r(\vec{x}_N^N))^{1/2})^2 / 2)^{1/2} = (1 - \sum_{r=1}^{R}(d_r(\vec{x}_N) d_r(\vec{x}_N^N))^{1/2})^{1/2}, \quad (67)$$

Для проверки исходной гипотезы предлагается использовать решающую функцию:

$$W_N = s_1^2 / s_2^2 * V_N, \quad (68)$$

где

$$s_1^2 = \max\left\{\|\vec{x}_N\|^2, \|\vec{x}_N^N\|^2\right\}, \quad (69)$$

$$s_2^2 = \min\left\{\|\vec{x}_N\|^2, \|\vec{x}_N^N\|^2\right\}, \quad (70)$$

Гипотеза отвергается при выполнении неравенства

$$W_N > \vartheta_\alpha, \quad (71)$$

где ϑ_α -порог, который соответствует некоторой желаемой вероятности ошибок первого рода

3.3 Особенность локальных минимумов распределения кратковременной энергии речевого сигнала

Особенностью данного метода является чувствительность к быстрому переходу между звуками, что может быть полезным при высоком темпе речи.

На первом этапе сигнал разбивается на фреймы, для которых вычисляется кратковременная энергия согласно формуле (72)

$$E_i = \frac{\sum_{n=1}^{N} x_i(n)}{N}, \quad (72)$$

На втором этапе применяется сглаживание распределения энергии с помощью сглаживающего фильтра Савицкого—Голея

На третьем этапе производится сопоставление локальных минимумов энергий границам звуков.

ГЛАВА 4 РЕАЛИЗАЦИЯ МОДЕЛИ ГОЛОСОВОГО УПРАВЛЕНИЯ В СИСТЕМЕ MATLAB

На первом этапе практических исследований в качестве слов «запускающих» одно из направлений голосового управления (управление звонками, управление музыкой в магнитоле, навигация) используем следующие слова: 1) позвонить 2) музыка 3) навигация. Слова записаны с частотой дискретизации 8000 Гц, каналом моно и разрядностью 16 бит. После распознавания сказанного слова система выдает уточняющий запрос. Данный запрос служит для сокращения словаря, что в свою очередь повышает вероятность правильного распознавания.

Блок-схема принципа запросов указана на рисунке 13

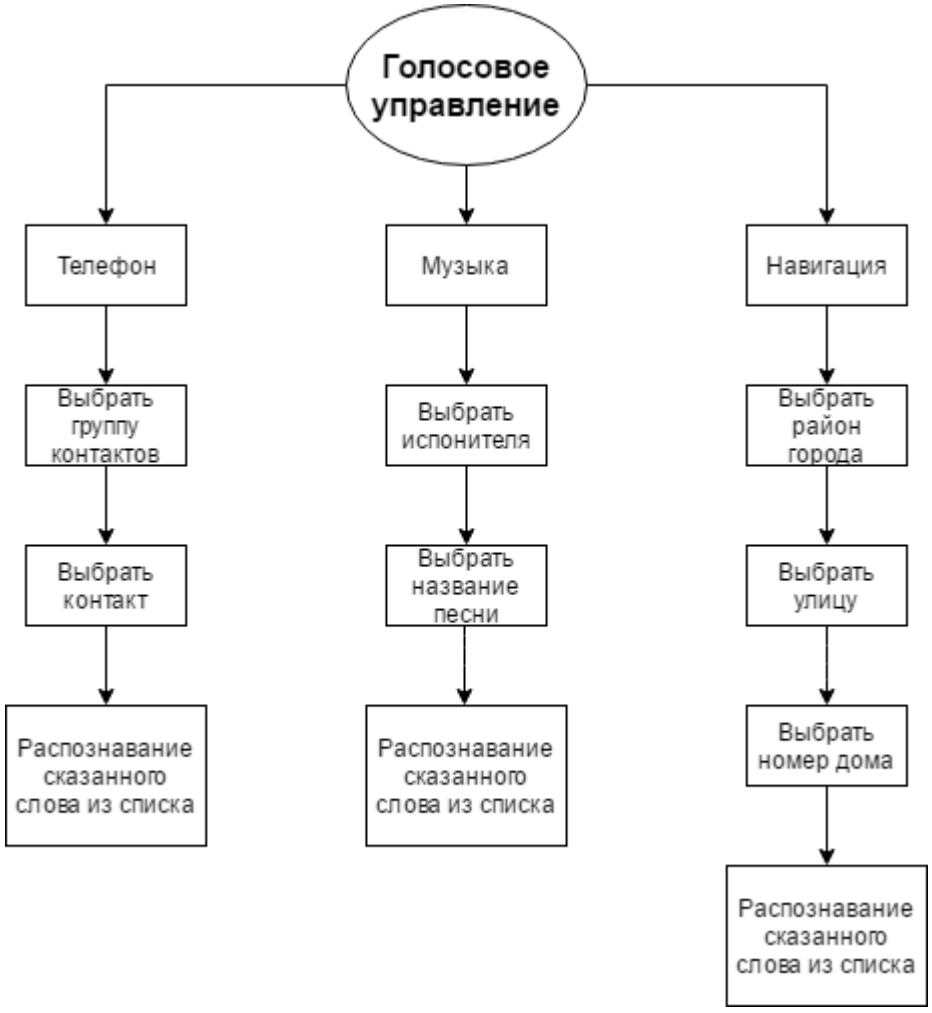

Рисунок 13 – Блок-схема организации запрос-ответа

После распознавания сказанного, система выдает встречный запрос о указании группы контактов\исполнителя\района города.

На рисунке 14 указан полный список исследуемых слов

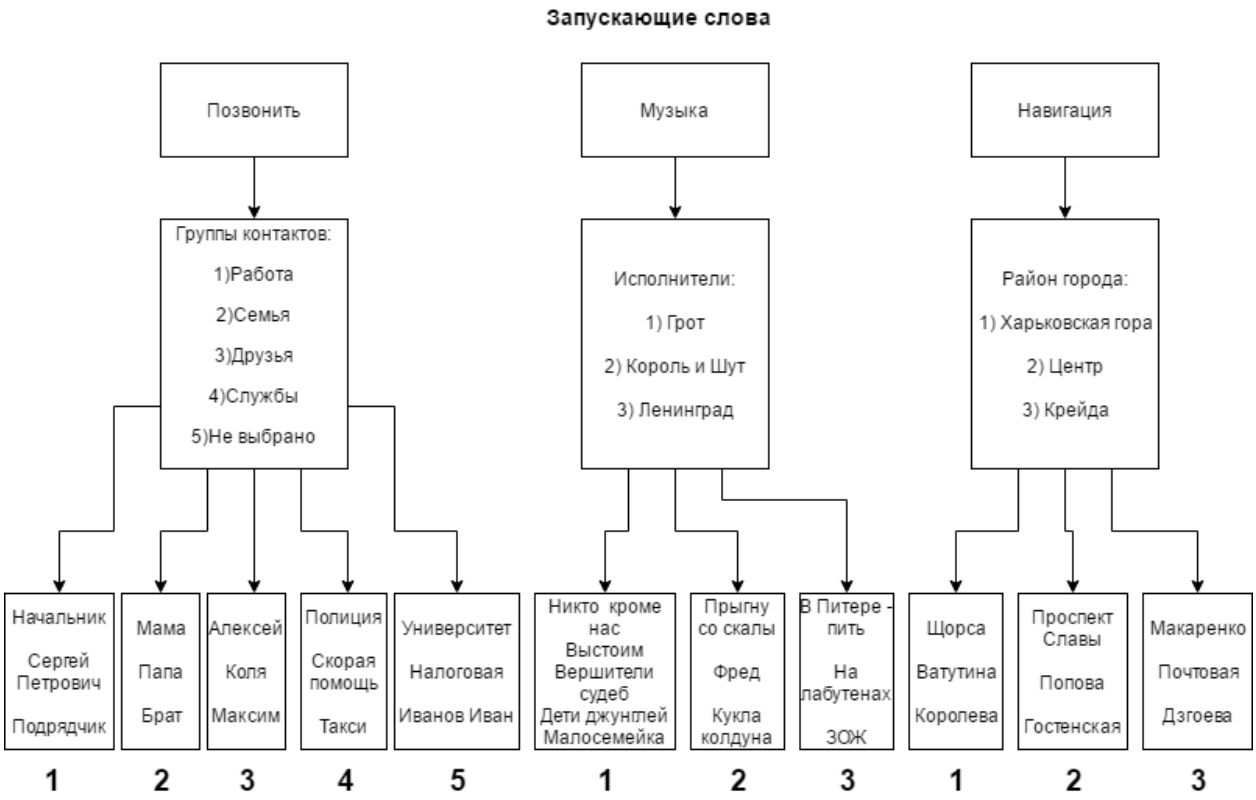

Рисунок 14 – Список исследуемых групп и относящихся к ним слов

Произведение распознавания речевого сигнала производится по следующему принципу (рисунок 15):

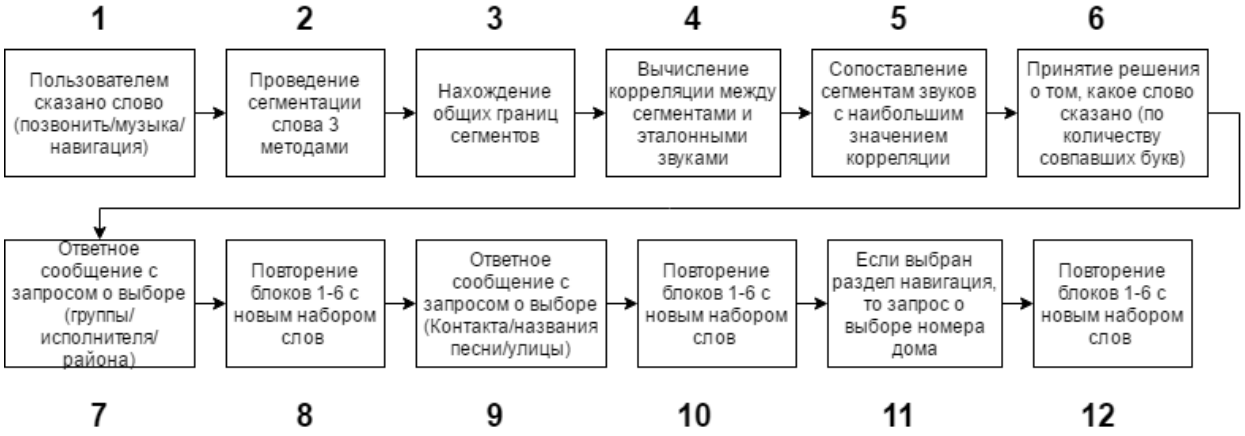

Рисунок 15 – Принцип предлагаемого метода распознавания

Проиллюстрируем работу исследуемого метода на примере «запускающих» слов (позвонить\музыка\навигация). На рисунках 16-18 представлены графики речевых сигналов, соответствующие «запускающим» словам.

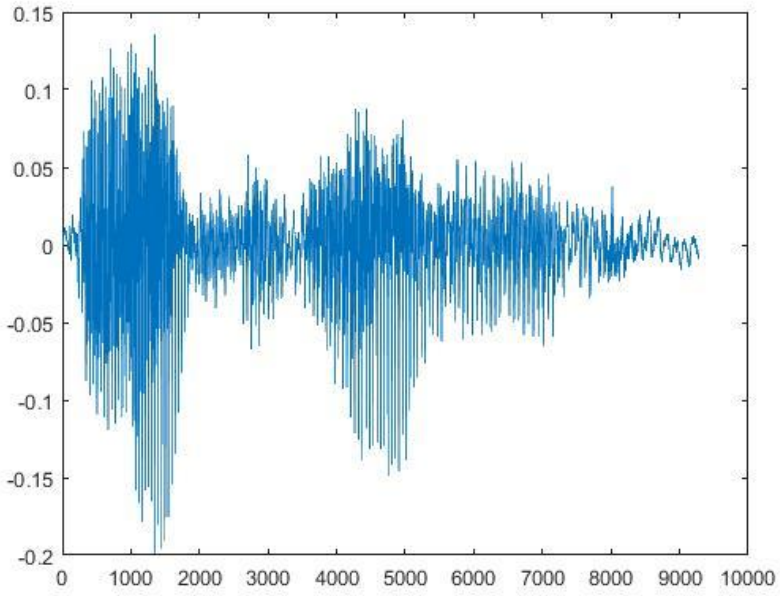

Рисунок 16 – График слова «позвонить»

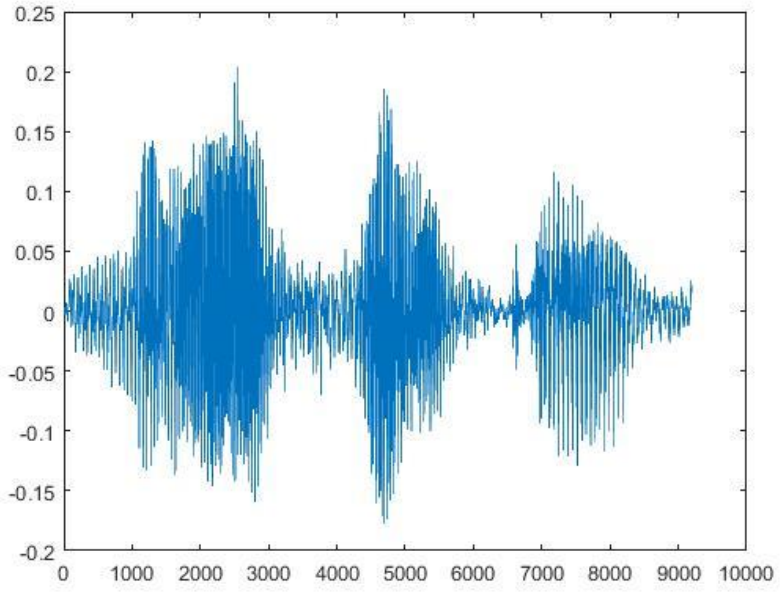

Рисунок 17 – График слова «музыка»

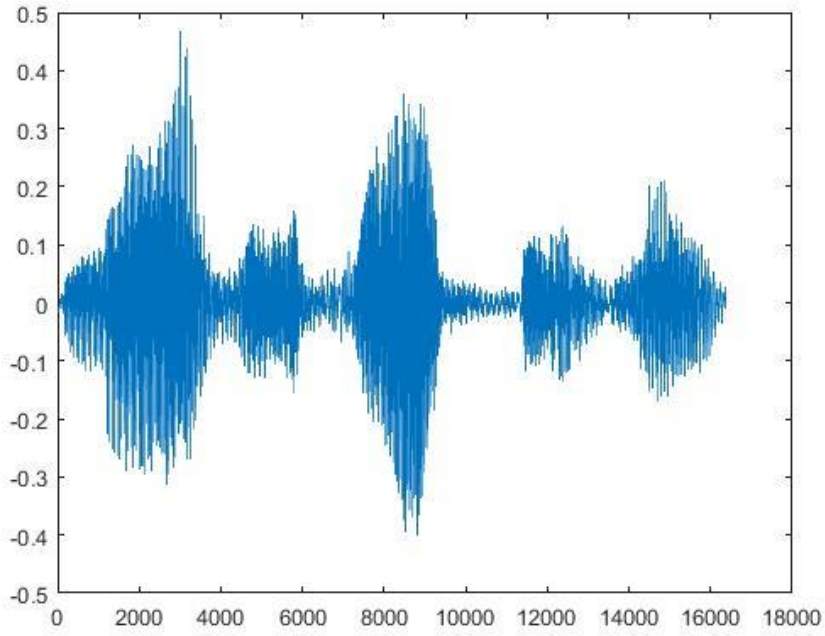

Рисунок 18 – График слова «навигация»

Так, из графиков можно заметить, что сигналы разные по форме и длительности, с разным количеством тихих периодов. Обнаружив и подсчитав данные периоды можно сделать предположение относительно того, к какому из слов принадлежит сказанное. Так, в русском языке, количество слогов в слове равно количеству гласных букв. Предположим, что переход от слога к слогу сопровождается перепадом энергии. И т.к слова разные по количеству слогов, проведя сегментацию и вычислив число тихих переходов можно с большой долей вероятности говорить о том, какое из слов было сказано.

Для наглядности качества работы методов проведем сегментацию данных слов на слух, и вычислим границы звуков. Графики слов (позвонить\ музыка\ навигация) изображены на рисунках 19-21.

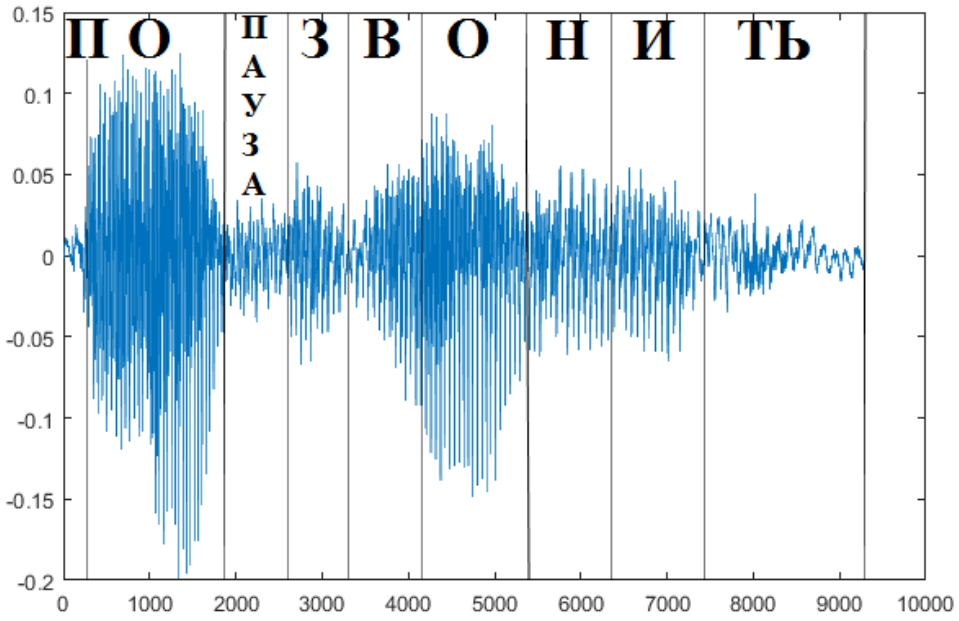

Рисунок 19 – График слова «позвонить» отсегментированного вручную

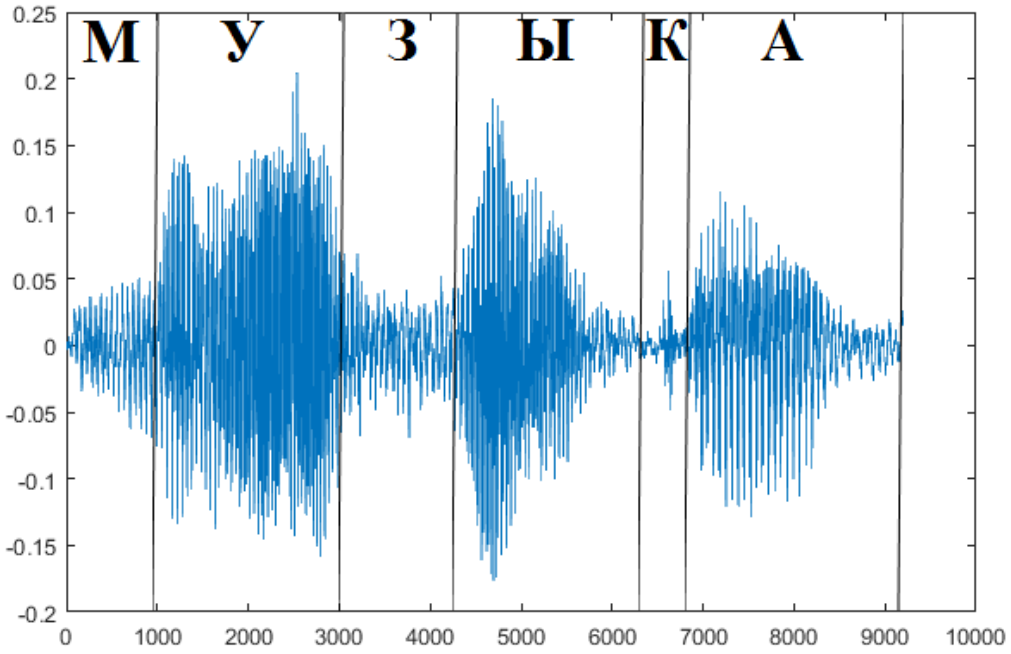

Рисунок 20 – График слова «музыка» отсегментированного вручную

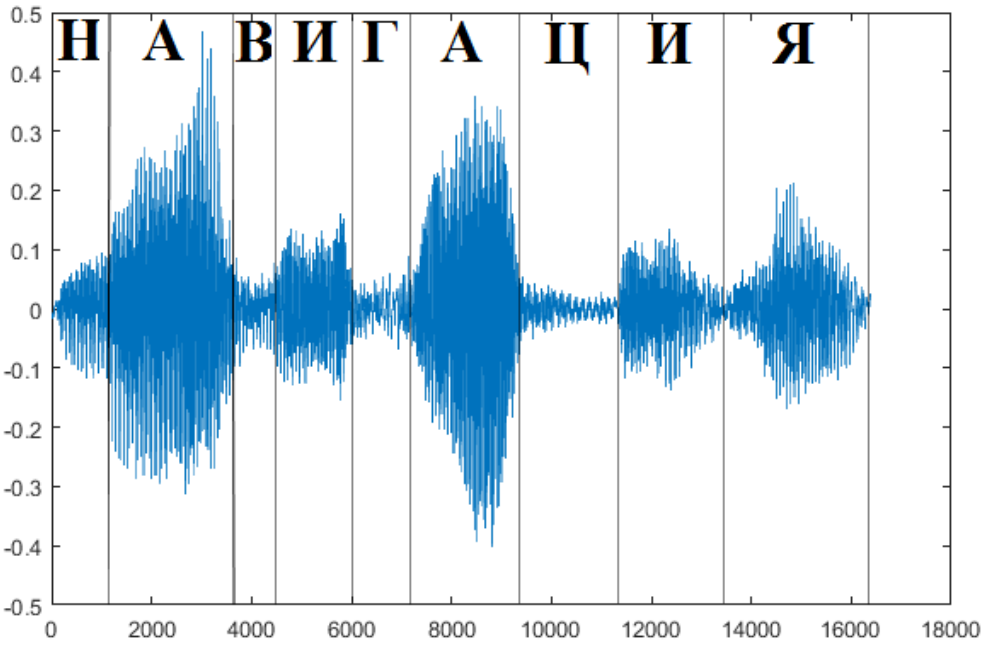

Рисунок 21 – График слова «навигация» отсегментированного вручную

Принятие решение относительно присутствия границы сегмента в первом субполосном методе сегментации происходит тогда, когда решающая функция принимает значение больше порога. Решающая функция (оранжевый) и порог (фиолетовый) представлены на рисунке 22.

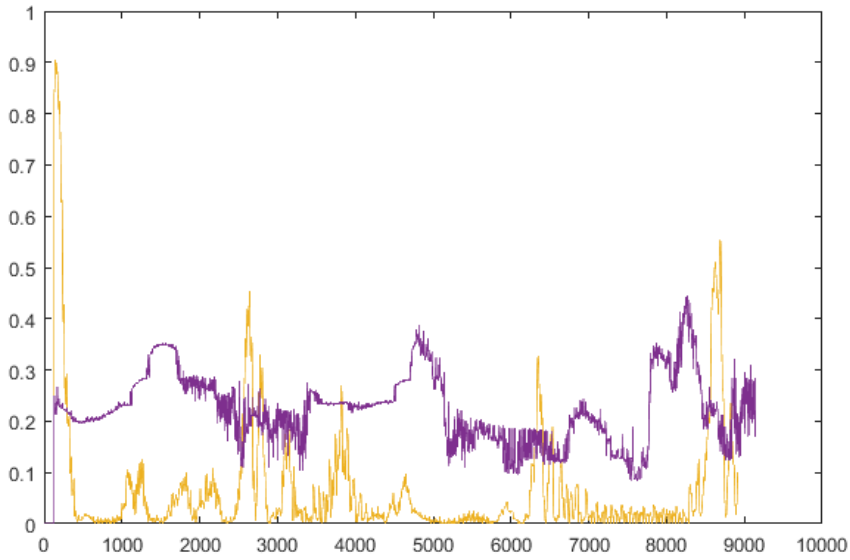

Рисунок 22 – Решающая функция (оранжевый) и порог (фиолетовый) для слова «позвонить»

Из рисунка 22 следуют следующие границы:

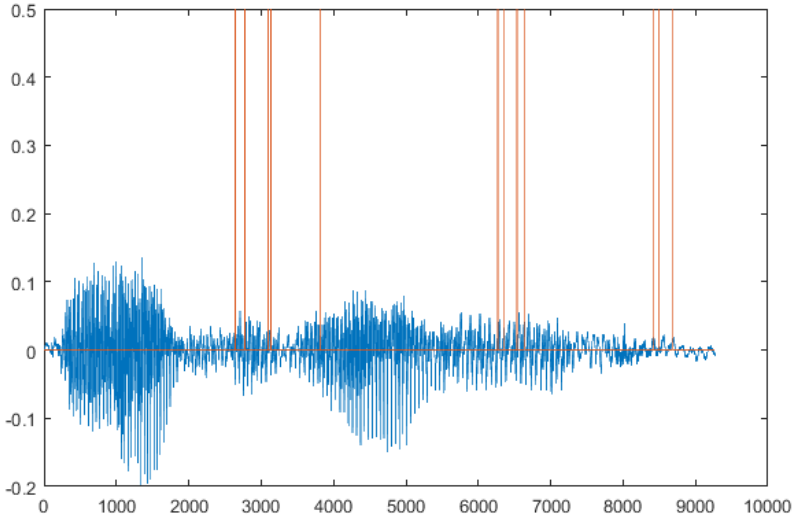

Рисунок 23 – Определение границ сегментов (красный) субполосным методом сегментации 1 для слова «позвонить»

Субполосный метод сегментации 2 зависит от порога, задаваемого пользователем. Данный порог соответствует некоторой желаемой вероятности ошибок первого рода. В данной работе используем порог равный 0,4.

Отсегментированный субполосным методом 2 сигнал указан на рисунке 24

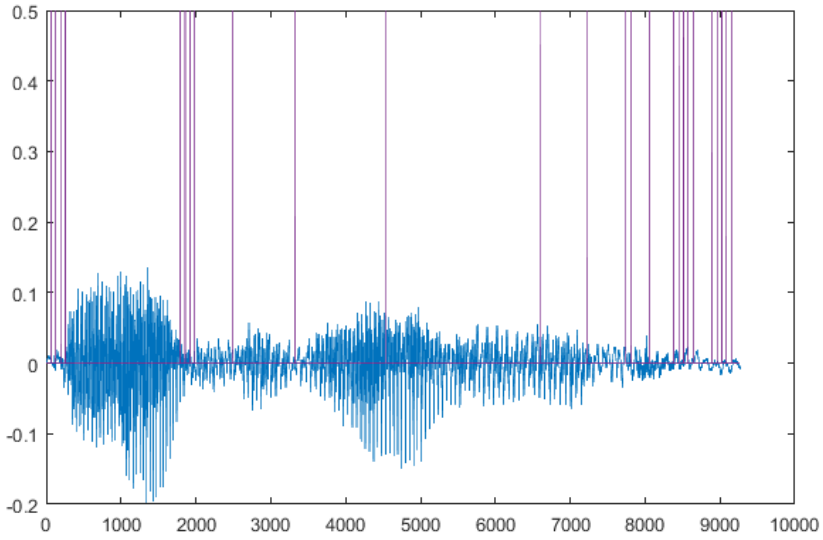

Рисунок 24 – Определение границ сегментов (фиолетовый) субполосным методом сегментации 2 для слова «позвонить»

Метод локальных минимумов основан на нахождении минимумов распределения локальной энергии соответствующих границам сигналов

Отсегментированный методом локальных минимумов сигнал указан на рисунке 25

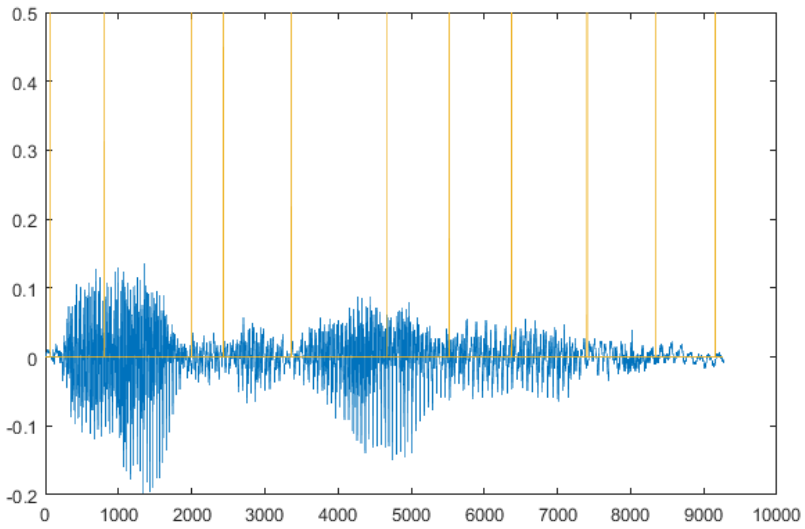

Рисунок 25 – Определение границ сегментов (оранжевый) методом локальных минимумов для слова «позвонить»

Совмещая все границы звуков получаем:

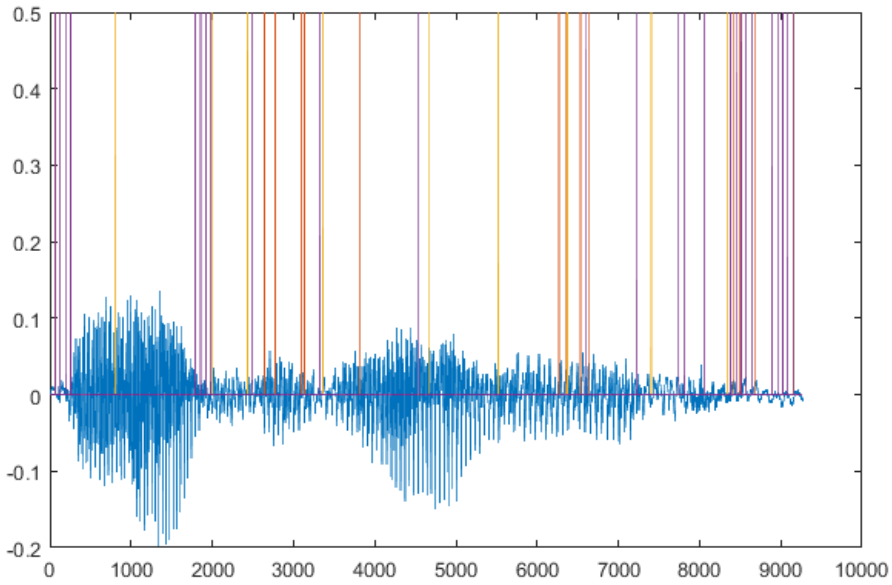

Рисунок 26 – Совмещенные границы звуков для 3-х методов слова «позвонить»

Для удаления лишних границ звуков используем следующее правило, предполагая, что минимальная длительность звука не превышает 300 отсчетов, оставляем только те границы, которые совпадают с разницей не более чем в 150 отчетов как минимум для 2 методов. Граница принимает среднее значение между двумя предполагаемыми границами, полученными в результате сегментации. Например, если в результате сегментации получено 2 границы звуков на 950-м отсчете одним методом сегментации и на 1050-м отсчете другим методом сегментации, то принимаем решение, что граница сигнала находится на 1000-м отсчете.

После применения данного правила к отсегментированному сигналу получаем следующие границы:

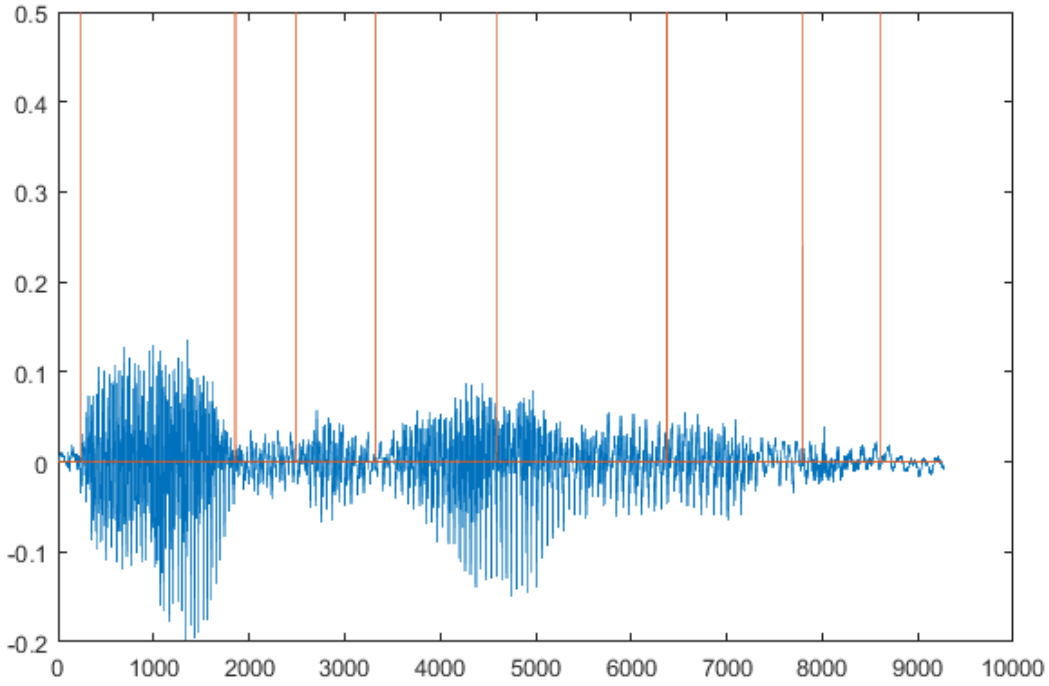

Рисунок 27 – Сегментация слова «позвонить» после применения к совмещенным границам (рисунок 26) вышеуказанных правил

Проведя операцию по сокращению границ сегментов из рисунка 27 можем сделать вывод, что данный метод с высокой степенью точности обнаруживает границы звуков и не оставляет при этом множества лишних границ звуков.

После проведения сегментации сравниваем каждый из получившихся сегментов с записанными эталонными звуками. Распознавание звука достигается путем вычисления корреляции сегмента с каждым из эталонных звуков, звук с наибольшей корреляцией принимаем за сказанный диктором

Корреляция вычислялась по формуле:

$$R(i,j) = \frac{C(i,j)}{\sqrt{C(i,i) \cdot C(j,j)}} \quad (15)$$

где C = cov(X) - матрица ковариаций.

В программной среде Matlab для реализации корреляции использующей данную формулу ссуществует встроенная функция corrcoef (x,y), с помощью которой и вычислялась корреляция в данной работе.

Так, сегмент сравнивался со всеми возможными звуками и сочетанием пар звуков присутствующих в словаре. В результате получаем следующие данные

Таблица 2 – Значения корреляций между сегментами и эталонными звуками для слова «позвонить»

	П	О	З	В	О	Н	И	ТЬ
1	2	3	4	5	6	7	8	9
В	0,012	0,064	0,001	0,04	0,01	0,038	0,005	0,007
ВО	0,006	0,057	0,0001	0,0045	0,1	0,002	0,007	0,04
З	0,079	0,006	0,053	0,15	0,05	0,04	0,02	0,056
ЗВ	0,039	0,055	0,002	0,021	0,029	0,016	0,04	0,015
И	0,04	0,088	0,025	0,09	0,004	0,15	0,097	0,002
ИТЬ	0,019	0,055	0,005	0,01	0,013	0,006	0,014	0,03
Н	0,012	0,013	0,037	0,017	0,04	0,16	0,02	0,03
НИ	0,013	0,01	0,004	0,014	0,0155	0,0002	0,0035	0,005
О	0,01	0,11	0,01	0,003	0,087	0,04	0,04	0,003

Продолжение таблицы 2

1	2	3	4	5	6	7	8	9
ОЗ	0,005	0,04	0,026	0,027	0,016	0,035	0,0044	0,01
ОН	0,035	0,046	0,034	0,027	0,027	0,093	0,023	0,01
П	0,085	0,02	0,001	0,03	0,07	0,014	0,03	0,009
ПО	0,023	0,017	0,005	0,01	0,006	0,006	0,05	0,0015
ТЬ	0,064	0,013	0,046	0,04	0,007	0,028	0,01	0,11
МУ	0,005	0,0175	0,02	0,002	0,024	0,05	0,04	0,005
А	0,038	0,01	0,038	0,195	0,013	0,085	0,017	0,0026
К	0,17	0,0065	0,008	0,012	0,015	0,053	0,018	0,003
ИЯ	0,0006	0,007	0,015	0,05	0,0017	0,024	0,028	0,015
ЦИ	0,013	0,0037	0,0036	0,018	0,012	0,047	0,08	0,026
АЦ	0,0015	0,0018	0,001	0,01	0,029	0,01	0,0066	0,014
ГА	0,002	0,014	0,014	0,014	0,0055	0,084	0,0035	0,017

Продолжение таблицы 2

1	2	3	4	5	6	7	8	9
ВИ	0,02	0,06	0,002	0,0095	0,06	0,045	0,02	0,002
АВ	0,03	0,0005	0,0075	0,0038	0,01	0,0034	0,0065	0,02
НА	0,009	0,039	0,009	0,018	0,004	0,00009	0,0009	0,011
Я	0,008	0,023	0,009	0,048	0,069	0,062	0,14	0,005
Ц	0,09	0,02	0,03	0,16	0,02	0,078	0,02	0,0047
Г	0,011	0,056	0,013	0,012	0,011	0,014	0,043	0,016
УЗ	0,01	0,002	0,026	0,03	0,0114	0,004	0,023	0,01
Ы	0,092	0,009	0,065	-0,089	0,03	0,043	0,038	0,033
У	0,01	0,02	0,024	0,097	-0,0078	0,039	0,073	0,012
М	0,015	0,102	0,006	0,05	0,012	0,006	0,076	0,014
КА	0,0007	0,076	0,0056	0,003	0,0188	0,036	0,032	0,012
ЫК	0,018	0,108	0,048	0,004	0,024	0,009	0,037	0,004
ЗЫ	0,0017	0,037	0,01	0,01	0,016	0,011	0,005	0,007

Согласно таблице 2 сопоставляя сегменты эталонным звукам с максимальным значением корреляции получаем слово «КОЫАОНИТЬ», где 5 из 8 звуков распознано верно (62,5%)

Проделаем аналогичные операции для слов «музыка» и «навигация»:

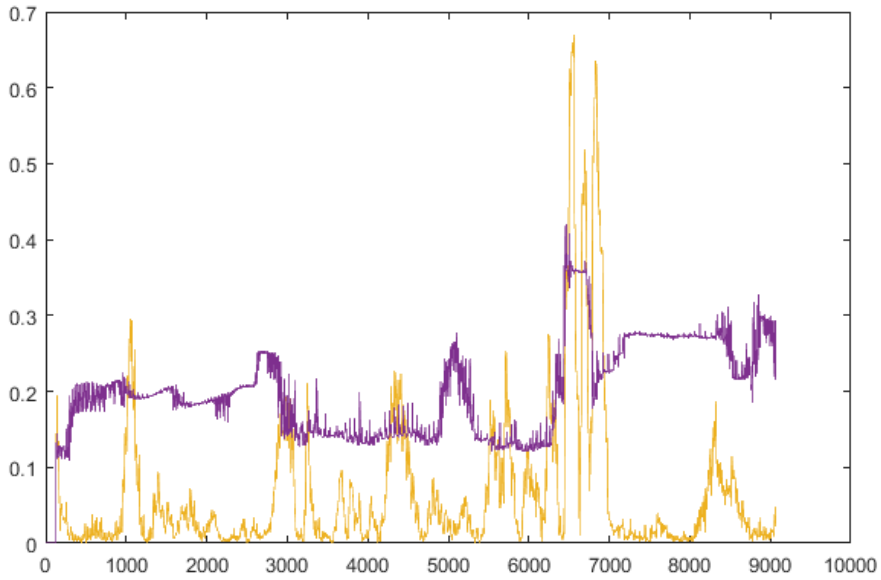

Рисунок 28 – Решающая функция (оранжевый) и порог (фиолетовый) для слова «музыка»

Отсегментируем сигнал по 1 субполосному методу сегментации:

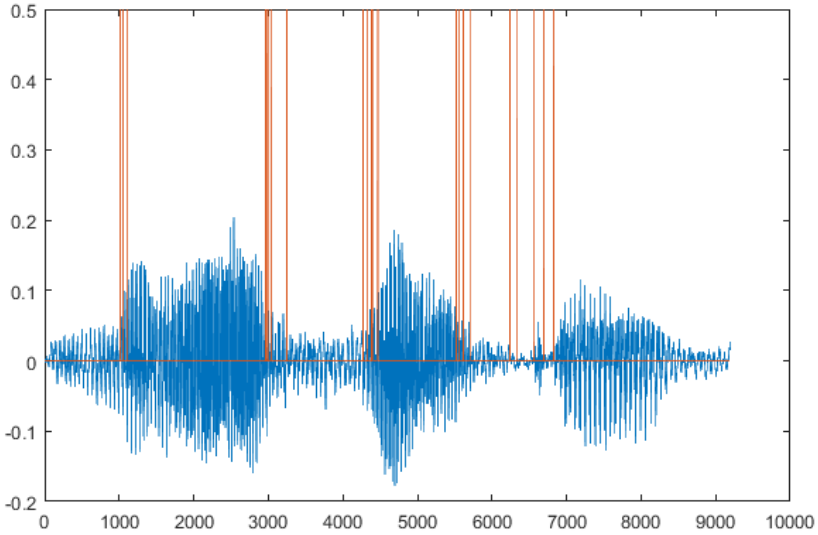

Рисунок 29 – Определение границ сегментов (красный) субполосным методом сегментации 1 для слова «музыка»

Отсегментируем сигнал по 2 субполосному методу сегментации:

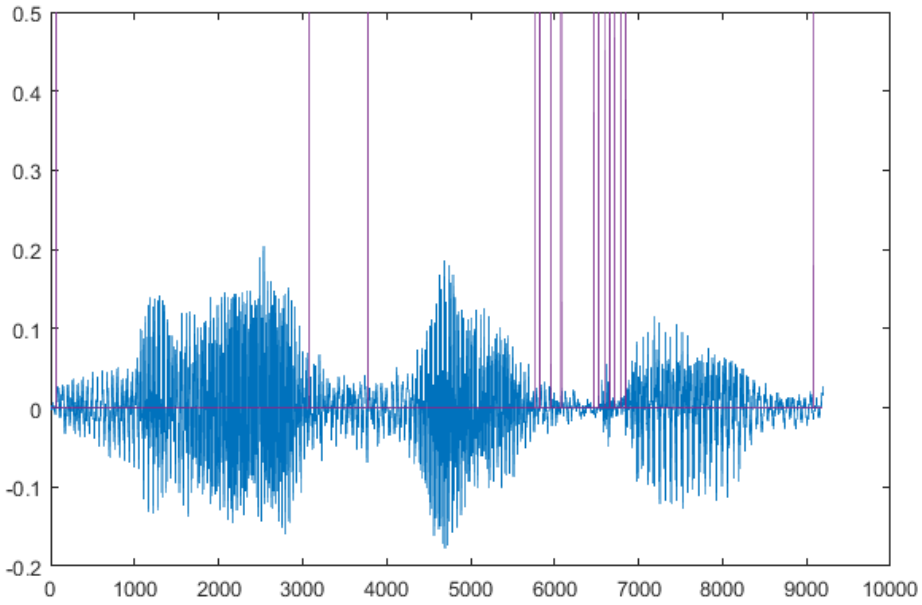

Рисунок 30 – Определение границ сегментов (фиолетовый) субполосным методом сегментации 2 для слова «музыка»

Отсегментируем сигнал по методу локальных минимумов распределения энергии:

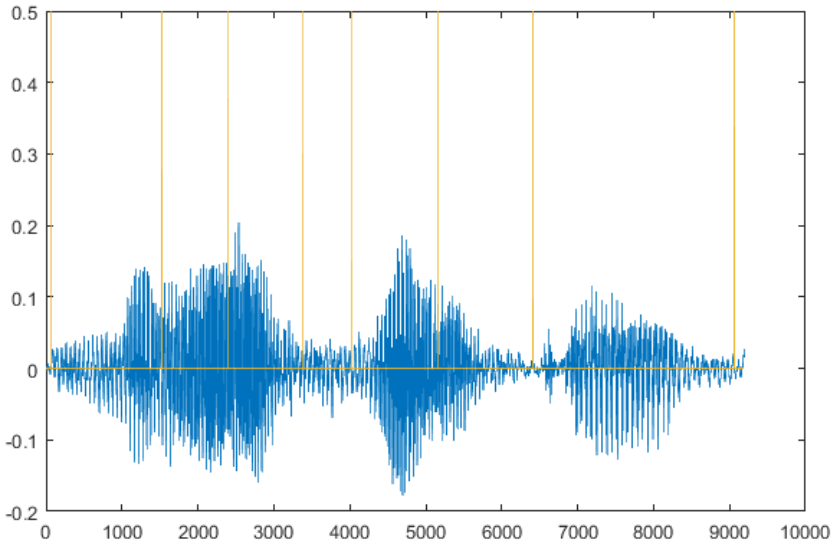

Рисунок 31 – Определение границ сегментов (оранжевый) по методу локальных минимумов для слова «музыка»

Совмещая границы полученные после сегментации тремя методами получаем:

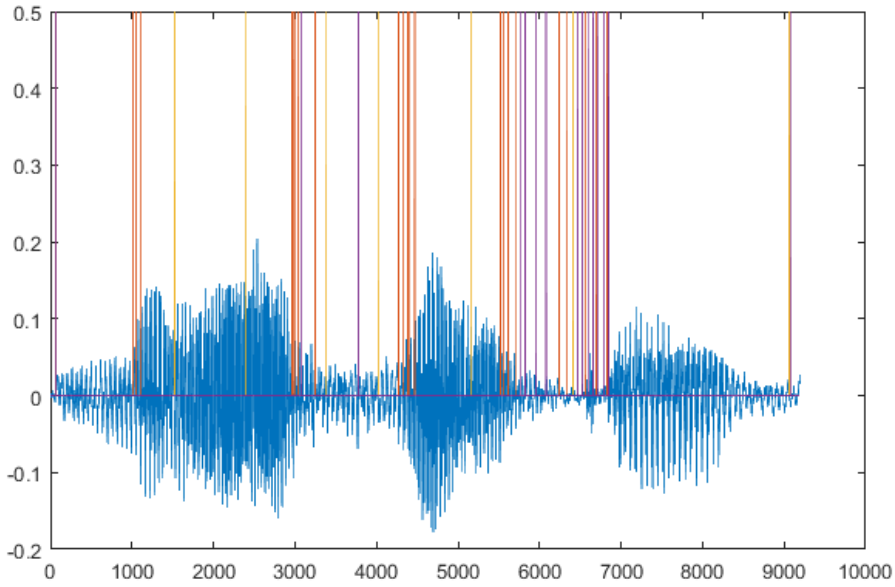

Рисунок 32 – Совмещенные границы сегментов, полученные тремя методами сегментации

Удалим лишние границы сегментов согласно процедуре (стр.58)

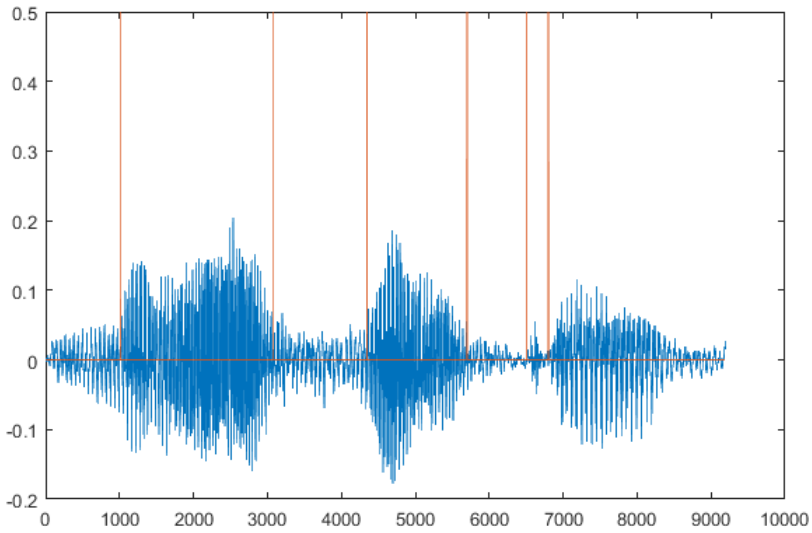

Рисунок 33 – Сегментация слова «музыка» после процедуры по удалению лишних границ (стр.58)

Рассчитаем корреляцию полученных сегментов с эталонным набором звуков: За произнесенный звук принимаем в соответствие эталонный звук, корреляция сегмента с которым наибольшая

Таблица 3– Значения корреляций между сегментами и эталонными звуками для слова «музыка»

	М	У	З	Ы	-	К	А
1	2	3	4	5	6	7	8
В	0,039	0,019	0,007	0,066	0,071	0,019	0,009
ВО	0,017	0,084	0,014	0,029	0,033	0,026	0,0058
З	0,0015	0,023	0,0447	0,131	0,003	0,056	0,01
ЗВ	0,045	0,035	0,058	0,043	0,02	0,04	0,019
И	0,008	0,033	0,037	0,053	0,019	0,092	0,038
ИТЬ	0,08	-0,025	-0,009	0,07	0,02	0,037	0,01
Н	0,017	0,047	0,07	0,023	0,038	0,09	0,01
НИ	0,008	0,032	0,026	0,12	0,005	0,0016	0,012
О	0,002	0,025	0,05	0,002	0,015	0,017	0,063
ОЗ	0,02	0,0015	0,01	0,11	0,012	0,043	0,02
ОН	0,01	0,089	0,026	0,038	0,02	0,04	0,023
П	0,055	0,015	0,018	0,046	0,076	0,035	0,007
ПО	0,062	0,012	0,003	0,073	0,037	0,096	0,107
ТЬ	0,034	0,024	0,006	0,006	0,096	0,037	0,006
МУ	0,006	0,026	0,028	0,0007	0,006	0,053	0,012
А	0,006	0,02	0,04	0,02	0,018	0,0754	0,16
К	0,098	0,012	0,014	0,075	0,038	0,11	0,005
ИЯ	0,0032	0,0069	0, 017	0,025	0,0068	0,052	0,036
ЦИ	0,022	0,05	0,04	0,0005	0,006	0,02	0,016

Продолжение таблицы 3

1	2	3	4	5	6	7	8
АЦ	0,024	0,0165	0,003	0,025	0,008	0,012	0,036
ГА	0,0055	0,002	0,015	0,0005	0,012	0,0575	0,012
ВИ	0,0053	0,025	0,06	0,019	0,003	0,007	0,014
АВ	0,013	0,02	0,0323	0,006	0,018	0,017	0,015
НА	0,0042	0,018	0,018	0,21	0,012	0,026	0,022
Я	0,002	0,012	0,13	0,12	0,013	0,06	0,014
Ц	0,047	0,012	0,025	0,014	0,17	0,105	0,037
Г	0,119	0,033	0,01	0,106	0,018	0,09	0,032
УЗ	0,0015	0,239	0,0134	0,0284	0,0527	0,03	0,044
Ы	0,025	0,0228	0,055	0,008	0,0336	0,033	0,0234
У	0,00246	0,0044	0,036	0,037	0,02	0,074	0,0215
М	0,17	0,003	0,0424	0,056	0,09	0,07	0,153
КА	0,045	0,024	0,004	0,064	0,007	0,03	0,004
ЫК	0,022	0,03	0,009	0,008	0,04	0,0096	-0,0108
ЗЫ	0,0074	0,033	0,012	0,023	0,0034	0,0007	0,06

Согласно таблице 3 сопоставляя сегменты эталонным звукам с максимальным значением корреляции получаем слово «МУЗЗВГКА», где 5 из 6 звуков распознано верно (83,3%), однако число букв после распознавания увеличилось до 8.

Проведем подобный эксперимент для слова «навигация»:

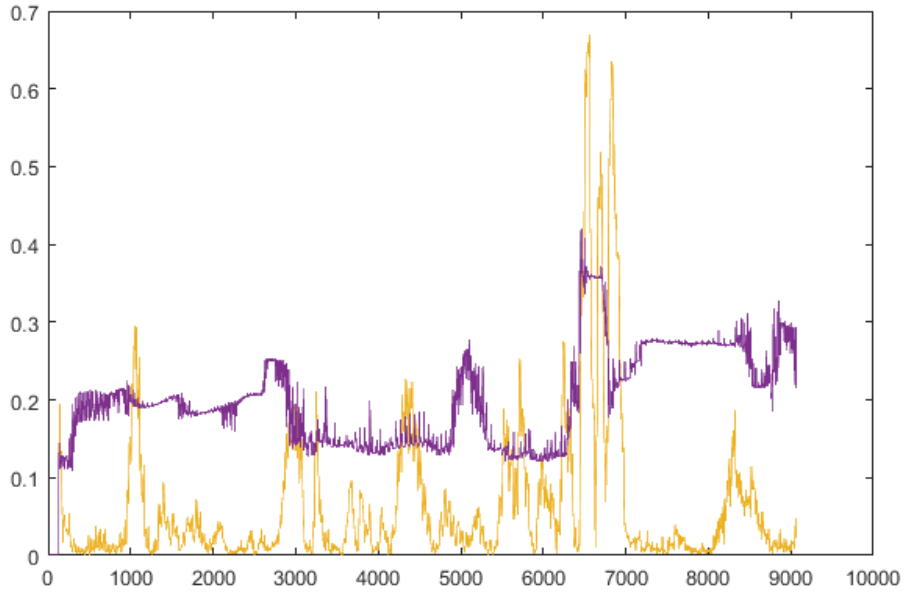

Рисунок 34 – Решающая функция (оранжевый) и порог (фиолетовый) для слова «навигация»

Отсегментируем сигнал по 1 субполосному методу сегментации:

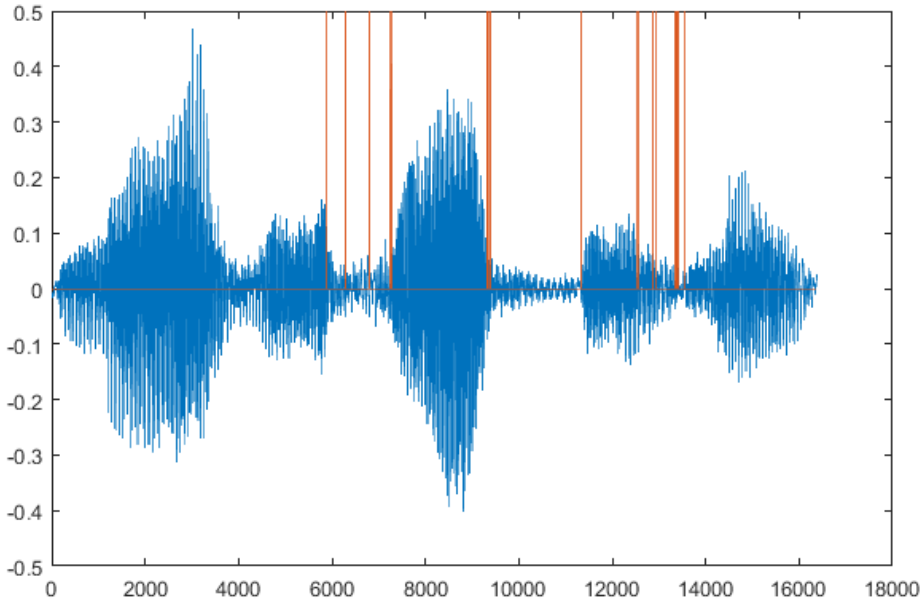

Рисунок 35 – Определение границ сегментов (красный) субполосным методом сегментации 1 для слова «навигация»

Отсегментируем сигнал по 2 субполосному методу сегментации:

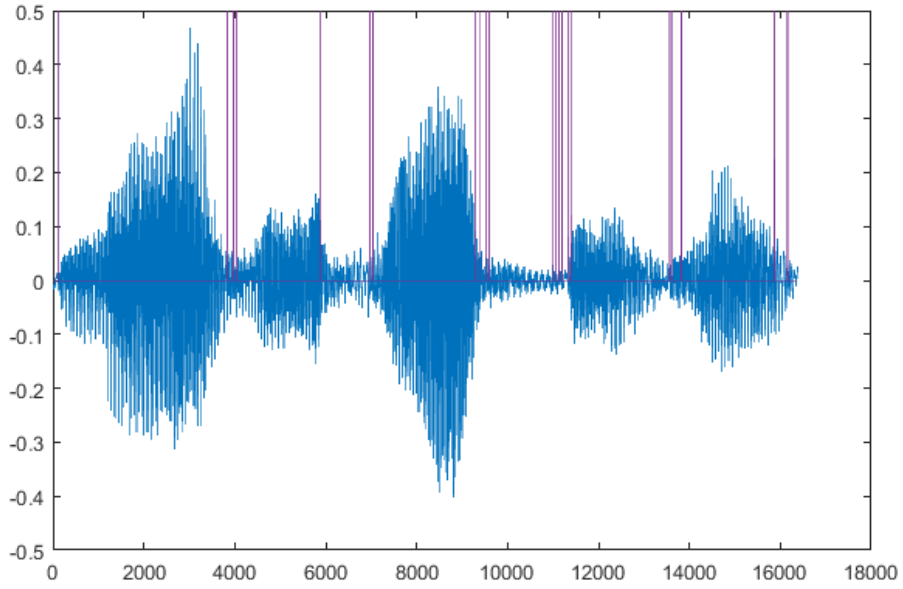

Рисунок 36 – Определение границ сегментов (фиолетовый) субполосным методом сегментации 2 для слова «навигация»

Отсегментируем сигнал по методу локальных минимумов распределения энергии:

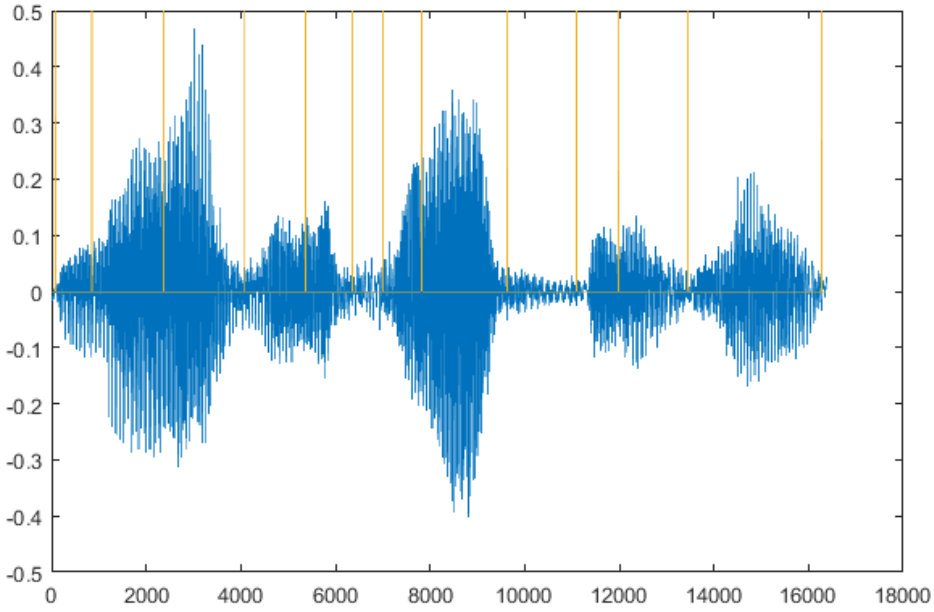

Рисунок 37 – Определение границ сегментов (оранжевый) по методу локальных минимумов для слова «музыка»

Совмещая границы полученные после сегментации тремя методами получаем:

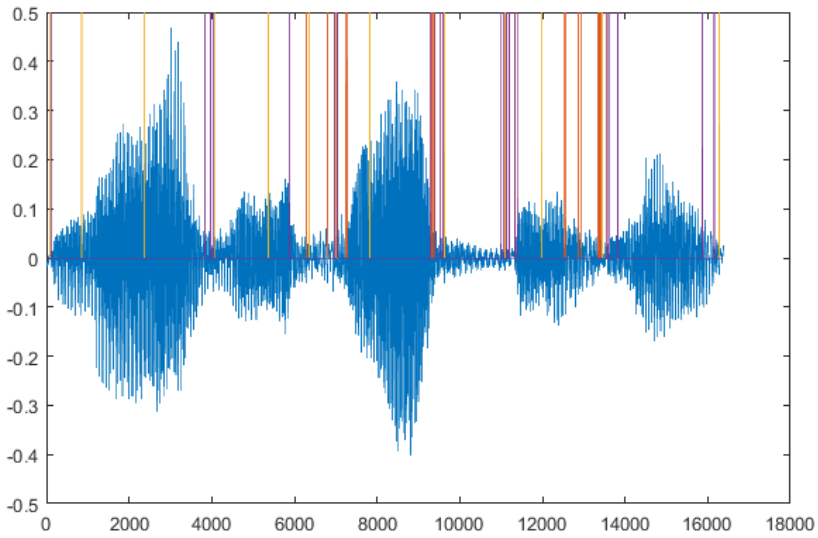

Рисунок 38 – Совмещенные границы сегментов, полученные тремя методами сегментации

Удалим лишние границы сегментов согласно процедуре (стр.58):

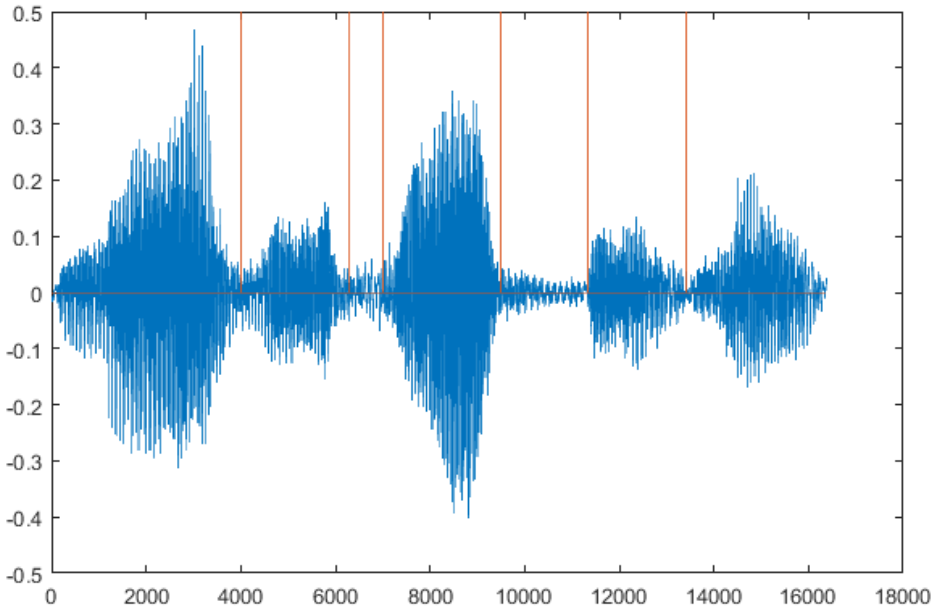

Рисунок 39 – Сегментация слова «музыка» после процедуры по удалению лишних границ (стр.58)

Таблица 4– Значения корреляций между сегментами и эталонными звуками для слова «музыка»

	НА	ВИ	Г	А	Ц	И	Я
1	2	3	4	5	6	7	8
В	0,0029	0,1614	0,00244	0,0123	0,0571	0,049	0,0012
ВО	0,003	0,0108	0,0077	0,0427	0,00017	0,003	0,0195
З	0,0014	0,0006	0,00002	0,037	0,027	0,03	0,0162
ЗВ	0,025	0,0084	0,044	0,04	0,0076	0,067	0,083
И	0,026	0,031	0,012	0,015	0,005	0,035	0,01
ИТЬ	0,0028	0,014	0,0026	0,028	0,02	0,037	0,1355
Н	0,015	0,094	0,0021	0,017	0,0212	0,082	0,02
НИ	0,0534	0,01	0,0046	0,04	0,018	0,124	0,0127
О	0,057	0,009	0,01	0,037	0,044	0,018	0,021
ОЗ	0,019	0,08	0,009	0,056	0,027	0,06	0,0063
ОН	0,006	0,009	0,002	0,08	0,0045	0,014	0,074
П	0,023	0,017	0,22	0,015	0,0146	-0,048	0,02
ПО	0,02	0,076	0,017	0,05366	0,068	0,051	0,038
ТЬ	0,011	0,044	0,0058	0,0347	0,016	0,0065	0,0024
МУ	0,0004	0,06	0,003	0,05	0,013	0,0026	0,041
А	0,026	0,027	0,018	0,038	0,018	0,002	0,0533
К	0,007	0,076	0,07	0,002	0,097	0,017	0,0056
ИЯ	0,012	0,02	0,018	0,06	0,03	0,02	0,0046
ЦИ	0,017	0,09	0,003	0,05	0,018	0,023	0,106
АЦ	0,015	0,0093	0,0074	0,009	0,01	0,0226	0,0084
ГА	0,023	-0,025	0,006	0,065	0,0056	0,027	0,03
ВИ	0,009	0,02	0,0053	0,024	0,012	0,024	0,0362
АВ	0,0003	0,015	0,0035	0,013	0,015	0,012	0,0063
НА	0,059	0,119	0,0004	0,031	0,0115	0,069	0,034
Я	0,012	0,016	0,007	0,044	0,022	0,029	0,09

Продолжение таблицы 4

1	2	3	4	5	6	7	8
ЦИ	0,017	0,09	0,003	0,05	0,018	0,023	0,106
АЦ	0,015	0,0093	0,0074	0,009	0,01	0,0226	0,0084
ГА	0,023	-0,025	0,006	0,065	0,0056	0,027	0,03
ВИ	0,009	0,02	0,0053	0,024	0,012	0,024	0,0362
АВ	0,0003	0,015	0,0035	0,013	0,015	0,012	0,0063
НА	0,059	0,119	0,0004	0,031	0,0115	0,069	0,034
Я	0,012	0,016	0,007	0,044	0,022	0,029	0,09
Ц	0,0024	0,036	0,036	0,0006	0,032	0,0027	0,01
Г	0,02	0,09	0,0005	0,0006	0,082	0,028	0,049
УЗ	0,0018	0,012	0,0128	0,05	0,005	0,009	0,04
Ы	0,039	0,052	0,02	0,02	0,02	0,088	0,09
У	0,0028	0,017	0,004	0,045	0,013	0,17	0,0015
М	0,026	0,022	0,03	0,05	0,027	0,0313	0,0195
КА	0,023	0,105	0,007	0,055	0,05	0,0185	0,05
ЫК	0,036	0,01	0,04	0,03	0,015	0,033	0,043
ЗЫ	0,051	0,075	0,009	0,025	0,002	0,04	0,004

Согласно таблице 4 сопоставляя сегменты эталонным звукам с максимальным значением корреляции получаем слово «НАВИКГАКУЦИ», соответственно 8 из 9 звуков\букв распознано верно (88,8%), однако число букв после распознавания увеличилось до 11.

Дальнейшим этапом является распознавание следующих групп слов, указанных на рисунке 14 (стр.50). В результате распознавания, для каждой из подгруппы, был выявлен свой примерный процент правильного распознавания звука :

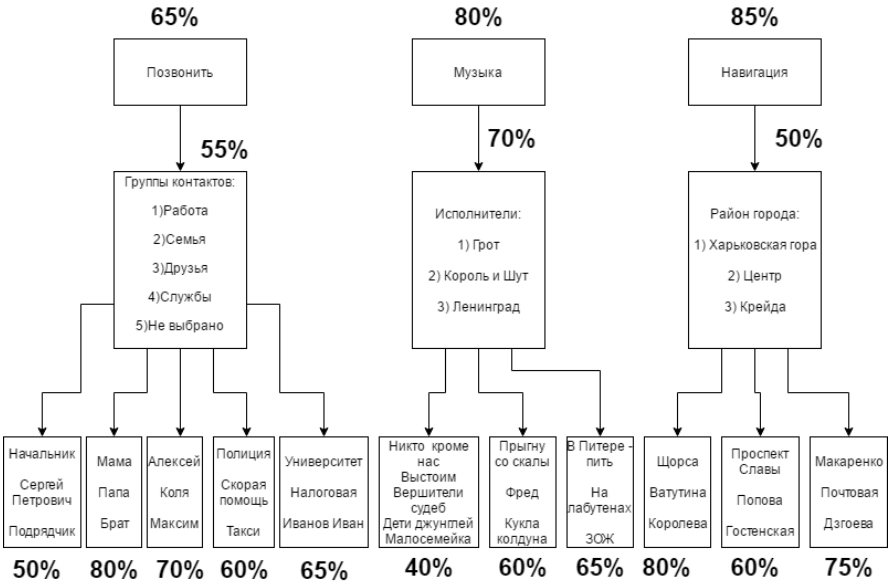

Рисунок 40 – Группы словаря с указанием процента правильного распознавания букв

Также было проведено распознавание слов (позвонить\музыка\навигация) сказанных другим диктором.

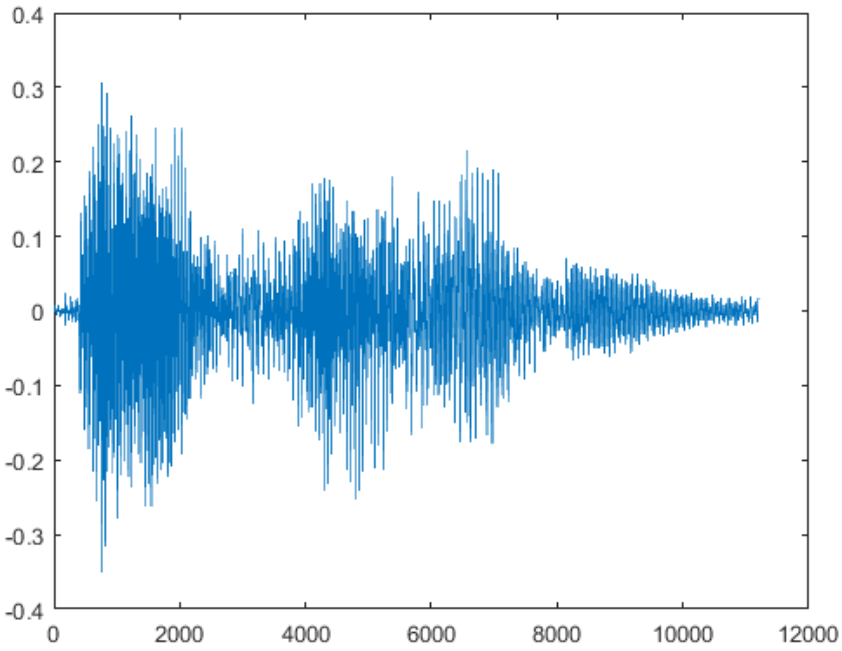

Рисунок 41– График слова «позвонить» произнесенного другим диктором

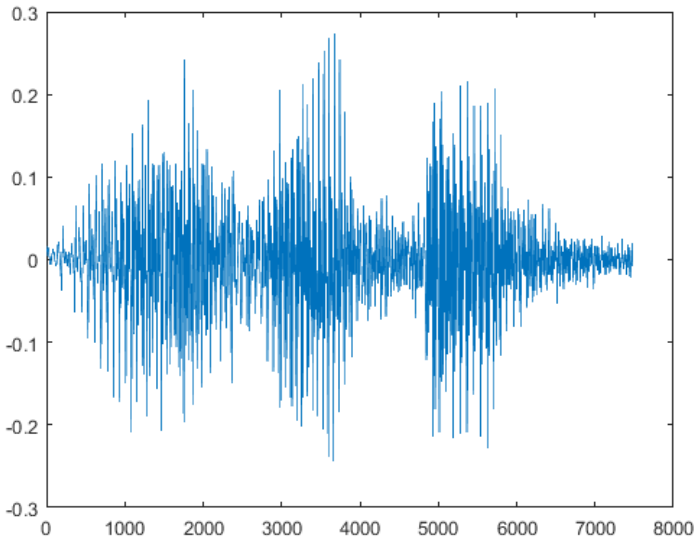

Рисунок 42– График слова «музыка» произнесенного другим диктором

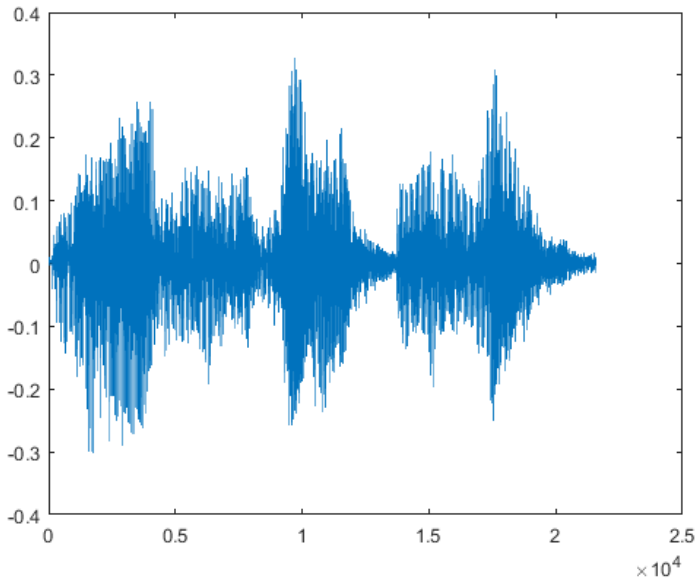

Рисунок 43– График слова «навигация» произнесенного другим диктором

После проведения аналогичных вычислений с данными сигналами был получен вывод, что качество распознавания ухудшилось на 15-20%, но тем не менее, этого достаточно чтобы сопоставить распознанное слово со словом из словаря.

Так в слове «позвонить» было распознано - 4 звука, в слове «музыка» - 3 звука, в слове навигация – 5 звуков.

ЗАКЛЮЧЕНИЕ

В результате магистерской диссертации была реализована модель голосового управления автомобилем, включающая в себя управление: звонками мобильного телефона, выбором треков в магнитоле, навигацию. Были решены поставленные в магистерской диссертации задачи.

Был проведен анализ литературных источников, который показал, что в настоящее время существуют различные методы и алгоритмы распознавания речи, однако, все они не лишены недостатков.

Так же в работе была осуществлена реализация модели голосового управления в системе Matlab.

На первом этапе реализации произнесённое слово (позвонить\музыка\навигация) сегментировалось тремя методами сегментации. Затем полученные границы сегментов объединялись, и используя некоторые принципы, происходило сокращение лишних границ. Далее для каждого из полученных сегментов вычислялась корреляция с каждым из эталонных звуков или сочетанием звуков и принималось решение о том какой именно звук был произнесен. В итоге, слово из словаря наиболее похожее на слово, полученное эмпирически, принималось как произнесенное диктором.

На втором этапе происходила такая же процедура для следующих групп слов:

1) Работа, Семья, Друзья, Службы, Не выбрано

2) Грот, Король и Шут, Ленинград

3) Харьковская гора, Центр, Крейда

На третьем этапе происходила аналогичная процедура распознавания для группы слов указанной на рисунке 14.

В результате проделанных экспериментов можно сделать следующий вывод:

Данный метод подходит для распознавания речи диктора в задачах голосового управления диктором и обеспечивает хороший уровень распознавания команд. Однако, данный метод зависим от размера словаря. Расширение базы слов, приводит к уменьшению вероятности правильного распознавания звука и снижению быстродействия системы. Но при этом, точности данного метода достаточно для распознавания слов при изменении диктора.

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. Винцюк, Т.К., Анализ, распознавание и интерпретация речевых сигналов [Текст]/ Т.К. Винцюк - Киев: Наук.думка, 1987. - 264с.

3. Голд, Б. Цифровая обработка сигналов [Текст] / Б. Голд, Ч. Рейдер. – М.: Сов. Радио, 1973. – 368 с.

4. Сорокин В.Н., Цыплихин А.И. Сегментация и распознавание гласных. // Информационные процессы, т. 4 , № 2, с. 202-220, 2004 г.

5. Жиляков, Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным [моногр.]/Е.Г. Жиляков. – Белгород: Изд-во БелГУ, 2007. – 160 с.

6. Жиляков, Е.Г. Методы обработки речевых данных в информационно-телекоммуникационных системах на основе частотных представлений [Текст]/ Е.Г. Жиляков, С.П. Белов, Е.И. Прохоренко. – Белгород: Изд-во БелГУ, 2007. - 136 с.

7. Жиляков, Е.Г. Сегментация речевых сигналов на основе анализа распределения энергии по частотным интервалам [Текст] / Е.Г. Жиляков, Е.И. Прохоренко, А.В. Болдышев, А.А. Фирсова, М.В. Фатова // Научные ведомости Белгородского государственного университета. Серия: История. Политология. Экономика. Информатика, Том 18 – 2011. - №7-1 (102). – С. 187-196

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

9. Куприянов, М.С. Цифровая обработка сигналов: процессоры, алгоритмы, средства проектирования [Текст] / М.С. Куприянов. – СПб.: Политехника, 1999. – 592 с.

10. Лайонс, Р. Цифровая обработка сигналов / Лайонс Р; - 2-е изд. ; Пер. с англ. − М.: ООО "Бином-Пресс", 2006 − 656 с.: ил.

11. Марпл-мл, С.Л. Цифровой спектральный анализ и его приложения / Марпл-мл. С.Л.; Пер. с англ. −М.: Мир, 1990.

12. Михайлов В.Г., Златоусов Л.В. Измерение параметров речи [текст]/ В.Г. Михайлов, Л.В. Златоусова; Под.ред. М.А. Сапожникова. – Москва: Радио и связь, 1987. – 168с.: ил.

13. Рабинер Л. Р., Шафер Р.В. Цифровая обработка речевых сигналов = Digitalprocessingofspeechsignals[текст]/ Л.Р. Рабинер, Р.В. Шафер.;Пер. с англ.М.В. Назарова, Ю.Н. Прохорова; Под ред. М.В.Назарова, Ю.Н. Прохорова. – Москва: Радио и связь, 1981. – 496с.:ил.

15. Сергиенко А.Б. Цифровая обработка сигналов: учебное пособие. — 3-е изд.— М.: БХВ-Петербург, 2011. — 768 с.

16. Солонина, А.И. Основы цифровой обработки сигналов [Текст]/ А.И. Солонина, Д.А. Улахович, С.М. Арбузов, Е.Б.Соловьева. – СПб.: БХВ-Петербург, 2005. – 768с.: ил.

17. Ю.Лабунец В. Г. Алгебраическая теория сигналов и систем. Красноярск: Изд-во Краснояр. ун-та, 1984.

18. Цемелъ Г. И. Опознавание речевых сигналов. М., Наука, 1971.

19. Шелепов В.Ю. Новые алгоритмы сегментации речевого сигнала и распознавания некоторых классов фонем / В.Ю. Шелепов, А.В. Ниценко // Искусственный интеллект. – 2007. – № 1. – С. 213-224.

20. Шелепов В.Ю. Новые алгоритмы распознавания фонем и их классов, поиск слова по его смешанной транскрипции при распознавании слов большого словаря / В.Ю. Шелепов, А.В. Ниценко,

А.В. Жук // Искусственный интеллект. – 2007. – № 2. – С. 139-147.

21. Шелепов В.Ю. О распознавании фонем с помощью анализа речевого сигнала в частотной и временной областях. Приложение к распознаванию синтаксически связных фраз / В.Ю. Шелепов, А.В. Ниценко, А.В. Жук, Д.С. Азаренко // Речевые технологии. – 2008.– № 2. – С. 43-52

22. Жиляков Е.Г., Фирсова А.А.. Сегментация речевых сигналов на основе субполосного анализа // Вестник НТУ "ХПИ", № 39 (1012). - 2013г. - С.73-81 .

23. Фирсова А.А. Разработка и исследование субполосных методов и алгоритмов сегментации речевых сигналов / Фирсова А.А.//Автореферат диссертации на соискание ученой степени кандидата технических наук. 17 мая 2013 г. - Белгород. - С.15-19.

24. Музычук Д.С. Сегментация, шумоподавление и фонетический анализ в задаче распознавания речи [Текст] / Музычук Д.С., Медведев М.С. // Молодой ученый. - 2013. - №6. - С. 86-96.

25. Дремин И.М., Иванов О.В., Нечитайло В.А. Вейвлеты и их использование. //Успехи физических наук, т. 171, №5 с. 465-500, 2001 г.

26. Ермоленко Т.Н. Алгоритмы сегментации с применением быстрого вейвлет- преобразования / Т.Н. Ермоленко, В.И. Шевчук // Статьи, принятые к публикации на сайте международной конференции Диалог’2003.

27. Шелухин О.И. Цифровая обработка и передача речи [Текст]. / Лукьянцев Н.Ф..– М.: Радио и связь, 2000.– 454 с.

28. Жиляков, Е.Г. Вариационные методы частотного анализа звуковых сигналов [Текст]. / Белов С.П., Прохоренко Е.И. // Труды учебных заведений связи. СПб, 2006.-№ 174.-С.163-170.

29. Жиляков Е.Г. Бабаринов С.Л. Чадюк П.В. Исследование сервиса компании Google Inc. по распознаванию русской речи / Жиляков Е.Г. Бабаринов С.Л. Чадюк П.В // Научные ведомости Белгородского государственного университета. Серия: История. Политология. Экономика. Информатика № 15-1 (158)/ - том 27. - 2013

30. Жиляков Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным на основе частотных представлений. - Белгород: Изд-во БелГУ, 2007. - с. 160

31. Цыплихин А.И. Системный анализ, управление и обработка информации,- 2006.

32. Фирсова А.А. О различиях распределения энергии звуков русской речи и шума / А.В. Болдышев, А.А. Фирсова // Материалы 12-ой Международной конференции и выставки "Цифровая обработка сигналов и еѐ применение. –"DSPA'2010". – Москва. – 2010. – С. 204–207.

33. Сорокин В.К. Синтез речи. М. : Наука, 1992. С. 392

34. Цыплихин А.И. Анализ и автоматическая сегментация речевого сигнала: дис. канд. тех. наук / ИППИ РАН. – М., 2006. – 149 с.

35. А. С. Колоколов. Обработка сигнала в частотной области при распознавании речи.- 2006. – с. 13–18

36. Конев А. А. Параметрическое описание сегментов речевого сигнала / В. И. Голубев, А. А. Конев // Научная сессия ТУСУР – 2005: Материалы Всероссийской научно-технической конференции студентов, аспирантов и молодых специалистов – Томск: Издательство ТУСУРа, 2005. – С. 113- 116.

37. Кочаров Д.А. Автоматическая интерпретация звуков речи // Диссертационная работа.- СПбГУ 2008

38. Утробин В.А., Гай В.Е. Алгоритм выделения вокализованных участков речевого сигнала // Вестник Нижегородского университета им. Н.И. Лобачевского, 2012, № 6 (1). С. 175–179

39. Мясникова Е.Н. Объективное распознавание звуков речи Л.: «Энергия», 1967. - 150 с.

40. Черник, Н. Н. Сегментация спонтанной речи в языках различных типов / Н. Н. Черник // Вестник Белорусского государственного экономического университета. - 2009. - N 4. - С. 101-107.

41. Ли У. А. Методы автоматического распознавания речи. М., Мир, 1983.

42. Огнев И.В., Огнев А.И., Парамонов П.А., Классификация речевых образов на основе анализа распределений их локальных экстремумов //131труды XXI международной научно-технической конференции"Информационные средства и технологии". - М.: МЭИ, 2013 - с. 53-57.

43. Винцюк Т.К., Анализ, распознавание и интерпретация речевых сигналов, — Киев: Наукова думка, 1987. – 264 стр.

44. Бондаренко Л. В., Вербицкая Л. А., Гордина М. В., Основы общей фонетики. – М.: Академия, 2004. – 160 с.

45. Шарий Т.В., О проблеме параметризации речевого сигнала в современных системах распознавания речи // Вісник Донецького Національного Универсітету, 2008, вып. 2, стр. 536-541

46. Маркел Дж. Д., Грэй А. Х., Линейное предсказание речи. – М.: Связь,1980. – 308 с.

47. Загоруйко Н. Г., Методы распознавания и их применение. – М.: Советское радио, 1972. – 208 с.

48. Агашин О.С., Корелин О.Н., Методы цифровой обработки речевого сигнала в задаче распознавания изолированных слов с применением сигнальных процессоров // Труды Нижегородского государственного технического университета им. Р.Е. Алексеева № 4, 2012, с. 32-44

49. . Ронжин А.Л., Ли И. В. Автоматическое распознавание русской речи //Вестник Российской академии наук, 2007, том 77, № 2, с. 133-138.

50. Огнев И. В., Парамонов П.А. Исследование способов представления числа для реализации арифметических операций в ассоциативной среде с командным управлением // Информационные средства и технологии: труды Международной научно-технической конференции (19 – 21 октября 2010 г.): в 3 т. – М.: МЭИ, 2010. – 1 т. – с. 54-60.

ПРИЛОЖЕНИЕ А

Код matlab 1:

```
set(0,'DefaultFigureColor','w')
clc
 clear
tic

[X1,Fd]=audioread('C:\Users\Вова\Desktop\Магистратура\друзья 8000.wav');

Nn=0;%начальное смещение обрабатываемого отрезка
R=32;

N=2*(2*R+1);
% N=258;
 Mx=length(X1);
%
 NNx=floor((Mx-Nn-2*N));%общая длительность речевого материала для обработки
Nxo=NNx;%длительность обрабатываемого отрезка без учета концов
   Nx=min(NNx,Nxo+2*N);
N2=2*N;
%  Nx=2*N;

  Om0=pi/(2*R+1);%полуширина частотного интервала
  %считыванием обрабатываемого отрезка
for k=1:Nx+2*N
   x(k)=X1(k+Nn);
```

```
end
 %Формирование субполосных матриц

for r=1:R+1
   Om(r)=2*(r-1)*Om0;
end

  skN(1)=0;%сыма квадратов элементов
for i=1:N
   for k=1:N
      if k==i A1(i,k,1)=Om0/pi;
      else A1(i,k,1)=sin (Om0*(i-k))/(pi*(i-k));

      end;
skN(1)=skN(1)+A1(i,k,1)^2;
   end
end
 skN2(1)=0;
for i=1:N2
   for k=1:N2
      if k==i A2(i,k,1)=Om0/pi;
         As(i,k)=Om0/pi;
      else A2(i,k,1)=sin (Om0*(i-k))/(pi*(i-k));As(i,k)=sin (Om0*(i-
k))/(pi*(i-k));
      end;
skN2(1)=skN2(1)+A2(i,k,1)^2;
   end
end
for r=2:R+1
skN(r)=0;
```

```
    for i=1:N
        for k=1:N
            A1(i,k,r)=2*A1(i,k,1)*cos(Om(r)*(i-k));
skN(r)=skN(r)+A1(i,k,r)^2;
        end

    end
end
    for r=2:R+1
skN2(r)=0;
    for i=1:N2
        for k=1:N2
            A2(i,k,r)=2*A2(i,k,1)*cos(Om(r)*(i-k));
skN2(r)=skN2(r)+A2(i,k,r)^2;
        end
    end
    end
   Nmax=8;

   ch=zeros(1,Nx);
   Nsh=N/2;
   MMc=floor(Nx/Nsh);
   betf=zeros(MMc,Nmax);
   beti=zeros(MMc,Nmax);
   uf=zeros(MMc,Nmax);
    ui=zeros(MMc,Nmax);
  % for tk=1:MMc
 Dolsq=zeros(1,Nx+2*N);
DolsqI=zeros(1,Nx+2*N);
Psir122sqI=zeros(1,Nx+2*N);
```

```
    for t=1:Nx-1%цикл сегментации
%t=(tk-1)*Nsh+1;
     PEDRF(t)=1;
 %средние значения сравниваемых отрезков
smt1(t)=0;
smt2(t)=0;

       for i=1:N
          ti=t-1+i;
           x1(i)=x(ti);%отрезок первый
           smt1(t)=smt1(t)+x1(i);
           x2(i)=x(N+ti);%отрезок последующий
           smt2(t)=smt2(t)+x2(i);

       end
       smtOb(t)=smt1(t)+smt2(t);
       smtOb(t)=smtOb(t)/(2*N);%общее среднее
       smt1(t)=smt1(t)/N;%первое среднее
       smt2(t)=smt2(t)/N;%второе среднее
       s1(t)=0;
       s2(t)=0;

       for i=1:N
          %центрирование общим средним
          xOb(i)=x1(i)-smtOb(t);
          xOb(i+N)=x2(i)-smtOb(t);
          x1(i)=x1(i)-smtOb(t);
          x2(i)=x2(i)-smtOb(t);
          %квадраты норм центрированных отрезков
          s1(t)=s1(t)+x1(i)^2;
```

```
        s2(t)=s2(t)+x2(i)^2;

      end
%  норма объединенного отрезка
     sOb(t)=s1(t)+s2(t);



%средние иквадраты отклонений
   disp1(t)=s1(t)/N;
   disp2(t)=s2(t)/N;
   dispOb(t)=sOb(t)/N2;

%средние части норм в нулевых интервалах
     Psr1(t)=Om0*s1(t)/pi;
     Psr2(t)=Om0*s2(t)/pi;
     PsrOb(t)=Om0*sOb(t)/pi;
     %суммарные матрицы
          AS1=zeros(N,N);
        AS2=zeros(N,N);
        ASOb=zeros(N2,N2);
         ASOb2=zeros(N2,N2);
         ASOb2pr=zeros(N2,N2);
         %определяемые характеристики
        PL12(t)=0;
        chOb(t)=0;
         chOb2(t)=0;
        PsirOb(t)=0;
         PsirOb2(t)=0;
        Psir12(t)=0;
```

```
 Psir122(t)=0;
 Psir122sq(t)=0;
 Psir122M(t)=0;
  Psir1222(t)=0;
   Psir1222sq(t)=0;
  PsirNorm(t)=0;
PS1(t)=0;
PS2(t)=0;
PSOb(t)=0;
 PSObsq(t)=0;
 PSOb2(t)=0;
 PSOb2sq(t)=0;
PP1(t)=0;
PP2(t)=0;
PS12(t)=0;
PS21(t)=0;
PSob1(t)=0;
PSob2(t)=0;
PSOb2sq(t)=0;
PSOb2sqI(t)=0;
Sper2(t)=0;
 Sper1(t)=0;
 Sob(t)=0;
 Sob2(t)=0;
 Sob2sqI(t)=0;
 Sob2sq(t)=0;
 PObs(t)=0;
 PCor(t)=0;
  shirOb2(t)=0;
  shirOb2sq(t)=0;
```

```
        shirOb2sqI(t)=0;
        shir12(t)=0;
        shir21(t)=0;
    PSOb2sqq(t)=0;

    %удлиннение сравниваемых отрезков для интерполяции спектров

   for i=1:N
   xs1(i)=x1(i);
   xs1(i+N)=0;
xs2(i)=x2(i);
   xs2(i+N)=0;
end
NS=16*(2*R+1);%количество вычислчяемых спектральных компонент

dsom=pi/NS;%шаг по частоте
%цикл вычисления спектров сравниваемых отрезков
Omi=zeros(1,NS);
for i=1:NS
   Omi(i)=(i-1)*dsom;
  sc1=0;
ss1=0;
sc2=0;
ss2=0;
scOb=0;
ssOb=0;
for k=1:N
    sc1= sc1+x1(k)*cos(Omi(i)*(k-1));
  ss1=ss1+x1(k)*sin(Omi(i)*(k-1));
 sc2= sc2+x2(k)*cos(Omi(i)*(k-1));
```

```
    ss2=ss2+x2(k)*sin(Omi(i)*(k-1));

  end
  dci=cos(N*Omi(i));
  dsi=sin(N*Omi(i));
  S1(t,i)=(sc1^2+ss1^2)/NS;
  S2(t,i)=(sc2^2+ss2^2)/NS;
  SObS(t,i)=((sc1+dci*sc2-
dsi*ss2)^2+(ss1+dci*ss2+dsi*sc2)^2)/NS;%спектр объединенного отрезка
  end
  %вычисления частей хэнергий в частотных интервалах через
суммирование
  %для всех возможных частотных интервалов и отдельных отрезков
  NSr=floor(NS/(2*R+1));%количество суммируемых компонент в
нулевом интервале
  Nk=2*16;
  Dsred(t)=0;
  Dsredsq(t)=0;
  Nsp=NS-4*NSr;
  ind=zeros(1,Nsp);
  for i=1:Nsp
    ST1(t,i)=0;
     ST2(t,i)=0;
     STOb(t,i)=0;
    for k=1:4*NSr
      kk=(i+k-1);
      ST1(t,i)=ST1(t,i)+S1(t,kk);
      ST2(t,i)=ST2(t,i)+S2(t,kk);
      STOb(t,i)=STOb(t,i)+SObS(t,kk);
    end
```

```
        ST1sq(t,i)=sqrt(ST1(t,i));
        ST2sq(t,i)=sqrt(ST2(t,i));
        STObsq(t,i)=sqrt(STOb(t,i));
        ST1n(t,i)=ST1(t,i)/s1(t);
         ST2n(t,i)=ST2(t,i)/s2(t);
         Dsred(t)=Dsred(t)+STOb(t,i);
         Dsredsq(t)=Dsredsq(t)+STObsq(t,i);
      end
      NSR=NS-Nk;
      DsredS(t)=Dsred(t)/NSR;
       DsredSsq(t)= Dsredsq(t)/NSR;
      %компонент спектров

       P1(t,1)=0;
       P2(t,1)=0;
       PObI(t,1)=0;
      NSr=floor(NS/(2*R+1));%количество суммируемых компонент в
нулевом интервале
      for i=1:NSr
        P1(t,1)=P1(t,1)+S1(t,i);
         P2(t,1)=P2(t,1)+S2(t,i);
          PObI(t,1)=PObI(t,1)+SObS(t,i);

      end
      POb(t,1)=P1(t,1)+P2(t,1);%оценка части энергии общего на основе частей
сравн. отрезков
      for r=1:R
        nr=(2*r-1)*NSr;
        P1(t,r+1)=0;
        P2(t,r+1)=0;
```

```
PObI(t,r+1)=0;
PObs(t,r+1)=0;
for i=1:2*NSr
 P1(t,r+1)=P1(t,r+1)+S1(t, nr+i);
 P2(t,r+1)=P2(t,r+1)+S2(t, nr+i);
 PObI(t,r+1)=PObI(t,r+1)+SObS(t, nr+i);
end
  POb(t,r+1)=P1(t,r+1)+P2(t,r+1); %оценка по сумме
end
SRsq(t)=0;
for r=1:R+1
 PObsq(t,r)=sqrt(PObI(t,r));
SRsq(t)= SRsq(t)+PObsq(t,r);
end
PsrObsq(t)=SRsq(t)/(2*R+1);
Nfr(t)=0;
Nfrsq(t)=0;
for i=1:Nsp
 if
STOb(t,i)>DsredS(t)Nfr(t)=Nfr(t)+1;ind(i)=1;PCor(t)=PCor(t)+sqrt(ST1(t,i)*ST2(t,i));PP1(t)=PP1(t)+ST1(t,i);PP2(t)=PP2(t)+ST2(t,i);
  Psir1222(t)= Psir1222(t)+(sqrt(ST2(t,i))-sqrt(ST1(t,i)))^2;PSOb(t)=PSOb(t)+STOb(t,i);end
%end
 if STObsq(t,i)>DsredSsq(t)Nfrsq(t)=Nfrsq(t)+1;ind(i)=1;
  Psir1222sq(t)= Psir1222sq(t)+(sqrt(ST2(t,i))-sqrt(ST1(t,i)))^2;PSObsq(t)=PSObsq(t)+STOb(t,i);end
end
  if PObI(t,1)> 1*PsrOb(t)PCor(t)=PCor(t)+sqrt(P1(t,1)*P2(t,1));
PP1(t)=PP1(t)+P1(t,1); PP2(t)=PP2(t)+P2(t,1);
```

```
Psir122(t)= (sqrt(P1(t,1))-sqrt(P2(t,1)))^2; PsirNorm(t)= (sqrt(P1(t,1)/s1(t))-sqrt(P2(t,1)/s2(t)))^2;chOb2(t)=1; Sob2(t)=Sob2(t)+PsrOb(t);
PSOb2(t)=PObI(t,1); shirOb2(t)=Om0; Psir122M(t)=PObI(t,1)*(sqrt(P1(t,1))-sqrt(P2(t,1)))^2;end
for r=2:R+1
ASOb2pr=A2(:,:,r);
if PObI(t,r)> 2*PsrOb(t) ASOb2=ASOb2+ASOb2pr; PCor(t)=PCor(t)+sqrt(P1(t,r)*P2(t,r)); PP1(t)=PP1(t)+P1(t,r); PP2(t)=PP2(t)+P2(t,r);
Psir122(t)=Psir122(t)+ (sqrt(P1(t,r))-sqrt(P2(t,r)))^2;PsirNorm(t)= PsirNorm(t)+(sqrt(P1(t,r)/s1(t))-sqrt(P2(t,r)/s2(t)))^2;
chOb2(t)=chOb2(t)+1;
Sob2(t)=Sob2(t)+2*PsrOb(t);Sob2M(t)=Sob2(t)+2*PsrOb(t)*PObI(t,r);
PSOb2(t)= PSOb2(t)+PObI(t,r); shirOb2(t)=shirOb2(t)+2*Om0;Psir122M(t)=Psir122M(t)+PObI(t,r)*(sqrt(P1(t,r))-sqrt(P2(t,r)))^2;
end
end
if PObsq(t,1)> 1*PsrObsq(t)
Psir122sq(t)= (sqrt(P1(t,1))-sqrt(P2(t,1)))^2; Sob2sq(t)=Sob2sq(t)+PsrOb(t); shirOb2sq(t)= shirOb2sq(t)+Om0;
PSOb2sqq(t)= PSOb2sqq(t)+PObsq(t,1); PSOb2sq(t)=PObI(t,1);end
for r=2:R+1

if PObsq(t,r)> 2*PsrObsq(t)
Psir122sq(t)=Psir122sq(t)+ (sqrt(P1(t,r))-sqrt(P2(t,r)))^2;
Sob2sq(t)=Sob2sq(t)+2*PsrOb(t);
PSOb2sq(t)= PSOb2sq(t)+PObI(t,r); shirOb2sq(t)= shirOb2sq(t)+2*Om0; PSOb2sqq(t)= PSOb2sqq(t)+PObsq(t,r);
end
```

```
 end
PsrOb2Isq(t)=PSOb2sq(t)/shirOb2sq(t)*Om0;
 PsrObIsq(t)=PSOb2sqq(t)/shirOb2sq(t)*Om0;
if PObsq(t,1)> 1*PsrObIsq(t)
        Psir122sqI(t)=                    (sqrt(P1(t,1))-sqrt(P2(t,1)))^2;
Sob2sqI(t)=Sob2sqI(t)+PsrOb(t); shirOb2sqI(t)= shirOb2sqI(t)+Om0;
        PSOb2sqI(t)=PObI(t,1);end
 for r=2:R+1

    if PObsq(t,r)> 2*PsrObIsq(t)
       Psir122sqI(t)=Psir122sqI(t)+ (sqrt(P1(t,r))-sqrt(P2(t,r)))^2;
       Sob2sqI(t)=Sob2sqI(t)+2*PsrOb(t);
    PSOb2sqI(t)=     PSOb2sqI(t)+PObI(t,r);     shirOb2sqI(t)=
shirOb2sqI(t)+2*Om0;
    end
 end
 DolPsq(t+N)=PSObsq(t)/Dsred(t);
 DolP(t+N)=PSOb(t)/Dsred(t);
 DolsqI(t+N)=shirOb2sqI(t)/pi;
 Dolsq(t+N)=shirOb2sq(t)/pi;
%   PsrObIsqI(t)=PSOb2sqq(t)/shirOb2sqI(t)*Om0;
 CofCor(t)=PCor(t)/sqrt(PP1(t)*PP2(t));
PshirOb2(t+N)=shirOb2(t)/pi;%общая информац доля частотной полосы
PDolOb2(t+N)=PSOb2(t)/sOb(t);%lдоля энергии
cof=1+1/R;
Dplsq(t+N)=(Nfrsq(t)*DsredS(t));
Dpl(t)=(Nfr(t)*DsredS(t));
Porog12(t+N)=Sob2(t)/PSOb2(t);
PRF12(t+N)=Psir12(t)/sOb(t);
% Porog122s(t+N)=(Sob2(t)/PSOb2(t))^2;
```

```
% PRF122s(t+N)=(Psir122(t)/sOb(t))^2;
%Porog122(t+N)=(Sob2(t)/PSOb2(t));
%Porog122(t+N)=Nfr(t)*DsredS(t)/PSOb(t);
Porog1222(t+N)=Dsred(t)/PSOb(t);%.Nfr(t)*DsredS(t)/PSOb(t);
Porog1222sq(t+N)=(Nfrsq(t)*DsredS(t))/Dsred(t);
Porog1222sqS(t+N)=Nfrsq(t)/NSR;
%PRF122(t+N)=(Psir122(t)/sOb(t));
%PRF122(t+N)=(Psir122(t)/Dsred(t));
PRF1222(t+N)=Psir1222(t)/(Nfr(t)*DsredS(t));
PRF1222sq(t+N)=Psir1222sq(t)/Nfr(t);
PRF1222sqS(t+N)=Psir1222sq(t)/Dsred(t);
PRF122M(t+N)=Psir122M(t)/sOb(t);
Porog122M(t+N)=Sob2(t);
Porog122Ms(t+N)=Sob2M(t)/sOb(t);
PRF122M2(t+N)=Psir122M(t)/(PSOb2(t));
Porog122M2(t+N)=PSOb2(t)/shirOb2(t);
PRFNorm(t+N)=PsirNorm(t);
PRF1221(t+N)=(Psir122(t)/Sob2(t));
PRF122sq1(t+N)=(Psir122sq(t)/PSOb2sq(t));
Porog1221(t+N)=1/(sOb(t)/PSOb2(t));
Porog122sq1(t+N)=Sob2sq(t)/(sOb(t));
PRF122sqI(t+N)=(Psir122sqI(t)/Sob2sqI(t));
Porog122sqI(t+N)=PSOb2sqI(t)/PSOb2sq(t);
    end
    RFR1222sqS=zeros(1,NNx+N);
    for t=1:NNx-1
       Porog1222RFR(t+N)=1;
       s(t)=PRF1222sqS(t)/Porog1222sqS(t);
      if s(t)>=1 RFR1222sqS(t+N)=s(t);
      end
```

```
    end
    %определение границ звуков
is=0;ii=0;
is2=0;ii2=0;
for t=N+1:Nx-1
 if PRF1222sqS(t)>Porog1222sqS(t);
 if is==0 is=1;id=0;ii=ii+1;in(ii)=t; dif1(in)= PRF1222sqS(t)-
PRF1222sqS(t-1);kch(ii)=1;
 else id=id+1;dif2=  PRF1222sqS(in(ii)+id)-PRF1222sqS(in(ii));
    if dif2>dif1(in(ii)) dif1(in)=dif2;dt(in(ii))=id;kch(ii)=kch(ii)+1;
 end;  end; else is=0;
 end
end
Ur=3;%floor(N/30);
w2=zeros(1,Nx);
 vv22=zeros(1,Nx);
for i=1:ii
   itt=in(i)+dt(in(i));
 w2(itt)=PRF1222sqS(itt)-Porog1222sqS(itt);
 if kch(i)>=Ur vv2(in(i)+dt(in(i)))=1;end
end
NM=Nx;
toc
```

Код matlab 2;

```
clc
clear
[X,Fs]=audioread('C:\Users\Вова\Desktop\Магистратура\службы
8000.wav'); % загрузка файла
```

```
N=64;
R=16;

h=0.4;

a=zeros(N,N,R);
v=[0:(pi/R):pi]
for r=1:R
    for i=1:N
      for k=1:N
        if i==k
          a(i,k,r)=(v(r+1)-v(r))/pi;
        else
          a(i,k,r)=(sin(v(r+1)*(i-k))-sin(v(r)*(i-k)))/(i-k)/pi;
        end;
      end;
    end;
end;
A=a; clear a

k=0;
for i=1:floor(length(X)/N-1)
   x1=X(i*N-(N-1):i*N);
   x2=X(i*N:i*N+N-1);
%     x1=X(i*N-(N-1):i*N);
%     x2=X(i*N+1:i*N+N);

for r=1:R
       Pr1(r)=x1'*A(:,:,r)*x1;
       Pr2(r)=x2'*A(:,:,r)*x2;
```

```
end
E1=sum(x1.^2);
E2=sum(x2.^2);
for r=1:R
   d1(r)=Pr1(r)/E1;
   d2(r)=Pr2(r)/E2;
end

pr2=0;
for r=1:R
   pr1=d1(r)*d2(r);
   pr1=pr1^(1/2);
   pr2=pr2+pr1;
end
pr3=1-pr2;
pr3=pr3^(1/2);
ANS=pr3; clear pr1 pr2 pr3

aaa=[E1 E2];
s1=max(aaa);
s2=min(aaa);
clear aaa

W=s1/s2*ANS;
en(i)=ANS;
if  W>h
   k=k+1;
   gr(k)=i*N;
end
```

```
end


X=X';
plot(X)
hold on
% for i=1:length(gr)
%     aa=gr(i)
% %     d(i)=aa;
% line([aa aa],[min(X) max(X)],'LineStyle','-','Color',[0 0 0])
% end
TR=X';
TR1=gr';
   gran=zeros (1,length (TR));
   for i=1:length (TR1)
      gran(1,gr(1,i))=1
   end
   ensr=mean(en);
% b=zeros (1,floor(length(X)/N));
% for   i=1:length (X)-1;
%
%
%
%
%
y = sgolayfilt(en,3,27);
y = sgolayfilt(y,3,27);
```

Код matlab 3:

```
clear
[x,Fs]=audioread('C:\Users\Вова\Desktop\Магистратура\службы 8000.wav')
x=x';
h=hilbert(x);
ogib=abs(h);
y = sgolayfilt(ogib,3,91);
xd = diff(y);
xds = sign(xd);
ix = (xds(1:end-1)~=xds(2:end)); % all extrema
ix = ix & (xds(1:end-1)<0); % only minima
mask(2:length(ix)+1) = ix;
ind = find(mask);
% ind1=ind*N-ind;
a=zeros(1,length (x));
for i=1:length (ind)
   a(1,ind(1,i))=1;
end
```