МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН

Некоммерческое акционерное общество

«АЛМАТИНСКИЙ УНИВЕРСИТЕТ ЭНЕРГЕТИКИ И СВЯЗИ»

Кафедра "Электроника"

«ДОПУЩЕН К ЗАЩИТЕ»

Зав.кафедрой _____

(ученая степень, звание, Ф.И.О.)

_____ «____» _____201___г.

(подпись)

ДИПЛОМНЫЙ ПРОЕКТ

На тему: Разработка цифровой системы голосового управления на базе ARM-микроконтроллера

Специальность 5В071600 Приборостроение

Выполнил(а) Аязбай Абу-Алим Ерикулы Группа ПС-11-3

(Ф.И.О.)

Научный руководитель к.т.н., доц. Монсизов К.М.

(ученая степень, звание, Ф.И.О.)

Консультанты:

по экономической части: ст. пр. Бабич А.А.

(ученая степень, звание, Ф.И.О.)

_____ «12» мая 2015 г.

(подпись)

по безопасности жизнедеятельности: ст. пр. Бечембетова А.С.

(ученая степень, звание, Ф.И.О.)

_____ «3» июня 2015 г.

(подпись)

(подпись)

Нормоконтролер: к.т.н., проф. Копесбаева А.А.

(ученая степень, звание, Ф.И.О.)

_____ «09» июня 2015 г.

(подпись)

Рецензент: к.т.н., доц. Утебаев Р.М.

(ученая степень, звание, Ф.И.О.)

_____ «____» _____201___г.

(подпись)

Алматы 2015

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН

Некоммерческое акционерное общество

«АЛМАТИНСКИЙ УНИВЕРСИТЕТ ЭНЕРГЕТИКИ И СВЯЗИ»

Факультет Аэрокосмических и информационных технологий

Кафедра Электроника

Специальность 5В071600 Приборостроение

ЗАДАНИЕ

на выполнение дипломного проекта

Студенту Аязбай
(Ф.И.О.)

Тема проекта Разработка цифровой системы голосового управления на базе ARM-микропроцессора

Утверждена приказом по университету № 124 от «29» 09 2014 г.

Срок сдачи законченного проекта « 15 » 05 2015 г.

Исходные данные к проекту (требуемые параметры результатов исследования (проектирования) и исходные данные объекта):

одноплатный ARM-компьютер Pi1 model B, веб-камера Logitech C270, плата ESP8266

Перечень вопросов, подлежащих разработке в дипломном проекте, или краткое содержание дипломного проекта: технологическая часть (описание существующих технологий), конструкторская часть (описание используемых устройств и их взаимодействия), программное обеспечение (алгоритм и листинг программы), общие вопросы охраны труда (расчет освещения и вентиляции), технико-экономическая часть (расчет экономи-

ческого эффекта и срока окупаемости).

Перечень графического материала (с точным указанием обязательных чертежей): в данной работе содержится 26 рисунка и 15 таблиц.

Основная рекомендуемая литература:

1) Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. – Киев: Наукова думка, 1987. – 264 с.
2) Магда Ю.С. Разведчик Pi. Руководство по настройке и применению. – М.: ДМК Пресс, 2014. – 188 с.

Консультации по проекту с указанием относящихся к ним разделов проекта

Раздел	Консультант	Сроки	Подпись
Технолог. часть	Мансуров К.М.	25.03.15	
Констр. часть	Мансуров К.М.	10.04.15	
Произ обеспеч.	Мансуров К.М.	20.04.15	
БЖД	Бекмурзаева А.С.	30.04.15	
Экон. часть	Бабич А.А.	05.05.15	

График
Подготовки дипломного проекта

Наименование разделов, перечень разрабатываемых вопросов	Сроки представления научному руководителю	Примечание
1 Технологическая часть	25.03.15	
2 Конструкторская часть	10.04.15	
3 Программное обеспечение	20.04.15	
4 Безопасность жизнедеят.	30.04.15	
5 Экономическая часть	05.05.15	

Дата выдачи задания «01» 02 2015 г.

Заведующий кафедрой _____ (Конев А.А.)
(подпись) (Ф.И.О.)

Научный руководитель проекта _____ (Молодров К.М.)
(подпись) (Ф.И.О.)

Задание принял к исполнению студент _____ (Аязбай А.В.)
(подпись) (Ф.И.О.)

Аннотация

В данной дипломной работе рассмотрены вопросы о технологиях, используемых в системах голосового управления. Описаны одноплатный ARM-миникомпьютер и Wi-Fi плата на микросхеме ESP8266, и разработана система голосового управления на данных устройствах.

Рассмотрены вопросы безопасности жизнедеятельности, составлен бизнес-план и рассчитан срок окупаемости разработанного проекта.

Summary

In this work concepts and technologies of voice recognition systems were defined, also this work includes descriptions of ARM-minicomputer and ESP 8266 Wi-Fi board. The author has made intellectual voice control system using these devices and described the process step by step.

Activity safety issues are considered, the business plan is made and the price of development of the project pays off.

Аңдатпа

Осы жұмыста дауыстпен басқару жүйелеріінде пайдаланылатын технологиялар туралы баяндалған. ARM-миникомпьютер мен ESP8266 Wi-Fi платаның сипаттамасы берілген, сондай-ақ осы құрылғылар негізінде дауыспен басқару жүйесі дайындалды.

Сонымен қатар, осы жұмыста өмір қауіпсіздігі мәселелері қаралды, бизнес-жоспар мен жобаның өзін-өзі ақтау мерзімі есептелді.

Содержание

	Введение	7
1	Технологическая часть	8
1.1	Состояние проблемы автоматического распознавания речи	8
1.2	Основные понятия распознавания речи	9
1.3	Обзор устройств чтения аудио сигналов	12
1.4	Подбор устройства захвата речи	20
1.5	Поиск оптимальной аудио-системы распознавания речи на базе открытого исходного кода	24
2	Конструкторская часть	29
2.1	одноплатный ARM-миникомпьютер pcDuino1	29
2.2	Wi-Fi-плата на базе микросхемы ESP8266	30
2.3	Модули реле	32
2.4	Беспроводной маршрутизатор TP-LINK TL-WR702N	33
2.5	Конструкция системы голосового управления	34
3	Программное обеспечение	36
3.1	Операционная система Ubuntu	36
3.2	Создание загрузочной карты памяти с Lubuntu	36
3.3	Программный пакет CMU sphinx и его установка	37
3.4	Настройка и запуск pocketsphinx	38
3.5	Обработка результатов распознавания	41
4	Безопасность жизнедеятельности	45
4.1	Анализ условий труда, выявление опасных и вредных производственных факторов	45
4.2	Расчет искусственного освещения в цеху	47
4.3	Расчет системы вентиляции	50
5	Технико-экономическое обоснование проекта	55
5.1	Введение	55
5.1.1	Обоснование необходимости разработки	55
5.1.2	Цель и задачи	55
5.1.3	Анализ рынка сбыта	55
5.2	Технико-экономическое обоснование проекта	55
5.2.1	Расчет инвестиционных затрат	56
5.2.2	Расчет себестоимости товарной продукции	57
5.2.3	Расчет доходов от реализации систем управления голосом	60
5.2.4	Расчет экономической эффективности	61
5.3	Вывод	63
	Заключение	64
	Список литературы	65

Введение

Проблема создания средств устного диалога человека с машинами является одной из наиболее актуальных проблем кибернетики, информатики и вычислительной техники[1], поэтому целью данной дипломной работы является разработка опытного образца системы голосового управления. Для достижения цели было поставлено несколько задач:

- выбор программного пакета распознавания речи;
- выбор устройства аудиозахвата;
- изучение принципа работы ARM-миникомпьютеров;
- установка и настройка выбранной программы распознавания речи.

В первой главе рассмотрены различные аудиосистемы распознавания речи с открытым исходным кодом. Среди них был выбран наиболее подходящий по описанным в работе критериям. Также в главе описаны различные типы устройств захвата звука и их особенности. Во второй главе описаны характеристики, приведена структурная схема ARM-миникомпьютера pcDuino1, который является основой системы голосового управления. В третьей главе описан алгоритм установки, настройки и запуска программного пакета sphinx, который является движком распознавания речевых команд.

1 Технологическая часть

1.1 Состояние проблемы автоматического распознавания речи

Попытки научить компьютеры общаться с людьми при помощи естественного голосового интерфейса предпринимались с первых лет истории компьютерной техники. В процессе многолетних исследований было выяснено, что к решению проблемы необходимо привлекать не только программистов, но и специалистов по языкознанию (лингвистике), радиоинженеров, математиков, биологов и даже психологов.

В самом деле, для создания системы распознавания речи нужно решить множество задач.

Прежде всего, нужно преобразовать колебания воздуха в электрические сигналы при помощи микрофона, отфильтровав при этом помехи и шумы.

Далее каким-то образом сигнал необходимо представить в цифровой форме, доступной для обработки при помощи компьютера (оцифровать). Здесь есть разные возможности: можно вводить в компьютер информацию об амплитуде звукового сигнала, а можно анализировать спектральный состав сигнала, выделяя из сигнала набор основных частот. Эту информацию можно комбинировать [2].

Специалисты по языкознанию необходимы для изучения принципов построения речи, выделения тех элементов речи, которые программа должна распознавать во входном потоке, — фонем, морфем, слогов, слов и т.п. Языкознание изучает такие аспекты речи, учет которых просто необходим при создании систем распознавания и синтеза речи.

Надо сказать, что извлечение из речевого потока лингвистических конструкций представляет собой непростую задачу. Только ребенок, который учится говорить или читать, произносит слова по слогам, отделяя слоги и слова паузами. В реальной жизни речь представляет собой сплошной поток звуков. В процессе формирования сплошной речи звуки, соответствующие одним и тем же буквам, могут изменяться при соединении с другими звуками.

Звуки сплошной речи содержат постоянно изменяющийся спектр гармонических частот, а также шум. Громкость и темп речи также постоянно изменяются. Более того, одна и та же фраза, сказанная разными людьми, или даже одним человеком, находящимся в разных психических состояниях, может иметь разную спектрально-временную окраску. Это сильно затрудняет создание универсальных систем распознавания, «понимающих» речь разных людей.

Для того чтобы выделить из оцифрованного звука лингвистические конструкции, применяются различные математические методы в сочетании со специальным компьютерным оборудованием, таким, например, как аппаратные или программные нейронные сети. На протяжении всей истории систем распознавания речи эти методы постоянно изменялись. При этом одни

методы отмирали из-за их неэффективности, а другие разрабатывались и совершенствовались.

Исследования биологов помогают изучить механизмы образования и распознавания речи, которыми пользуется человек (а может быть, и животные). Пытаясь решить проблему речевого интерфейса, многие исследователи пытаются моделировать работу человеческих органов, таких как речевой тракт, ухо и нейронные системы головного мозга.

И, наконец, при решении задачи создания речевого интерфейса компьютерных систем нужны психологи, так как без учета психологических особенностей человека невозможно создать по-настоящему удобный речевой интерфейс.

Проблемы общения человека и компьютера выходят за рамки чисто технических наук. Это происходит, потому что современные компьютеры пытаются общаться с людьми с применением тех же способов, с помощью которых общаются между собой люди. И точно также как бывает трудно общаться человеку с больными, страдающими расстройствами психики, так же трудно будет ему общаться с компьютерной программой, демонстрирующей неадекватное психическое поведение.

1.2 Основные понятия распознавания речи

Речь – комплексное явление. Люди плохо понимают как она возникает и воспринимается. Считается, что речь состоит из слов, а каждое слово из фонов. В действительности все иначе. Речь – это динамический процесс без явновыраженных частей. Воспользовавшись аудиоредактором, можно увидеть форму сигнала при записи речи – рисунок 1.1.

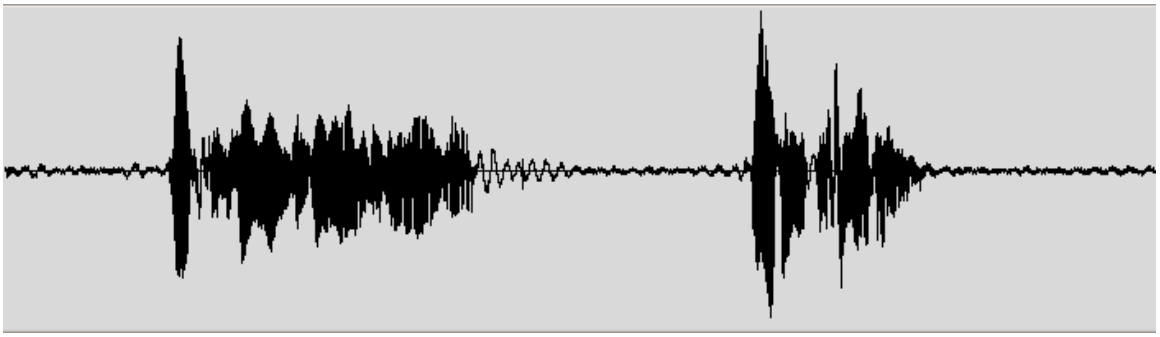


Рисунок 1.1 – Форма сигнала при записи речи

Все современные описания речи являются в некоторой степени вероятностными, то есть нет явных границ между разными словами. Распознавание речи не могут быть абсолютно верны. Такой подход является необычным для разрабочтчиков программного обеспечения, которые обычно работают с детерминированными системами.

Речь – это непрерывный звуковой поток, где устойчивые состояния сменяются динамическими и наоборот. В этой последовательности состояний, некоторые могут определять более или менее похожие единицы речи – фоны. Считается, что слова состоят из фон, но это не совсем так. Акустические параметры звуковой волны одной фоны могут различаться в зависимости от различных факторов – от контекста фоны, от стиля речи говорящего и т.д., поэтому фоны могут звучать отлично от "канонического" звучания. Т.к. переходы между фонами более информативны, используется такое понятие, как дифоны. Дифоны – часть фон между двумя последовательными фонами. Также используется понятие части фоны. Часто можно определить три или более частей фоны.

Наиболее часто выделяют три части фоны. Первая часть фоны зависит от предыдущей фоны, вторая часть является относительно стабильной, а третью часть фоны определяет последующая фона, поэтому в распознавании речи чаще всего определяют три части фоны.

Иногда фоны рассматриваются в контексте. Некоторые фоны в контексте можно рассматривать как трифоны или даже как пентафоны. Например, фона "u" в слове "bad" следует после фоны "b" и предшествует фоне "d", а в слове "ban" следует после фоны "b" и предшествует фоне "n". В двух этих словах фона "u" звучит по-разному [3].

При вычислениях, целесообразней определять части трифонов, чем целиком. Для этого нужно создать детектор начала трифон. Большое количество разнообразных звуковых детекторов основываются на некоторых отдельных небольших детекторах. Обычно используется около 4000 небольших детекторов для создания детекторов для трифон. Такие детекторы называются "сеноны". Сеноны могут определять контекст намного обширнее и комплексно, чем простое "предыдущий-следующий". Принцип работы сенонов можно представить как дерево решений.

Несколько последовательных фон рассматривают как слоги. Характерной чертой слогов является то, что их свойства не зависят от скорости речи, в отличие от фон, но они зависят от интонации речи. Слоги в свою очередь составляют слова, которые ограничивают количество возможных комбинаций фон. Если существует 40 фон и слово в среднем состоит из 7 фон, то должно быть 40^7 различных слов, однако, даже самый образованный человек редко использует более 20 тысяч слов, что немного облегчает задачу распознавания речи.

Простой способ распознавания речи происходит следующим образом: записывается звуковой сигнал, который делится на фрагменты, которые разделены периодами тишины, далее происходит распознавание каждого фрагмента. Для этого перебираются все возможные комбинации слов, и выбирается та, которая наиболее совпадает с записанным фрагментом. Есть несколько важных моментов в данном процессе.

Во-первых – это принцип признака. Так как количество параметров велико, требуется их оптимизация. Речь делится на кадры длинной около 10

мс, и из каждого кадра выделяется 39 чисел, которые описывают речь. Эти числа называются векторами признака. Принцип вычисления таких векторов – тема для отдельной работы, но если описать его двумя словами, то это – производная спектра.

Во-вторых – это принцип модели. Модель описывает некоторый математический объект, который является совокупностью параметров некоторого слова, то есть содержит наиболее вероятный вектор признака слова. Главные проблемы принципа модели – как хорошо модель соответствует действительности, можно ли улучшить модель и как изменяется адекватность модели в разных условиях. Модель речи является скрытой марковской моделью. Эта модель описывает процессы как последовательность состояний, которые сменяют друг друга с некоторой вероятностью. Скрытая марковская модель применима не только в распознавании речи, но и в других аналогичных процессах.

В-третьих – это процесс сравнивания. Поскольку процесс перебирания всех возможных векторов признака может занять слишком много времени, для оптимизации процесса используется множество приемов.

В соответствии со структурой речи, для распознавания речи используются три модели.

Акустическая модель содержит акустические свойства (наиболее вероятные векторы признака) и параметры для каждого слога. Используют контекстно-независимые и контекстно-зависимые модели.

Фонетический словарь указывает соответствие слов последовательности фон. Такой метод не очень эффективен, так как фонетическая модель обычно содержит два или три варианта произношения, но этого достаточно для поддержания точности распознавания на удовлетворительном уровне.

Языковая модель используется для оптимизации процесса поиска слов. Языковая модель содержит возможные комбинации слов, и определяет какое слово может следовать за предыдущим.

При разработке систем распознавания речи, разработчик стремится сделать ее как можно более точной и оптимизированной (быстрой). Для тестирования систем распознавания речи обычно используются специальные тестовые записи. Для оценки систем распознавания речи используются параметры, приведенные ниже.

Частота ошибочных слов определяется по формуле 1.1:

$$WER = \frac{I + D + S}{N}, \qquad (1.1)$$

где WER (word error rate) – частота ошибочных слов;
I – количество лишних слов;
D – количество отсутствующих слов;
S – количество перемещенных слов;
N – количество слов в оригинальном тексте.

Точность определяется по формуле 1.2:

$$Accuracy = \frac{N - D - S}{N}. \qquad (1.2)$$

Скорость определяется как отношение продолжительности распознавания к продолжительности чтения текста. Например, текст читался 2 часа, а его распознавание заняло 6 часов, следовательно скорость распознавания составит 0,33 от реального времени [3].

1.3 Обзор устройств чтения аудио сигналов

Первые записи звуков происходили методом непосредственного нарезания записи: через рогоподобные микрофоны вибрация воздуха передавалась на граммофонную иглу, которая и производила нарезание этих колебаний на поверхности вращающегося воскового цилиндра. Сегодняшние микрофоны преобразуют колебания воздуха в электрический сигнал, и хотя эта теория лежит в основе работы всех микрофонов, их различия заключаются в процессах производящих эти преобразования. Вопрос о классификации микрофонов не так прост, как может показаться. Они различаются:

- по принципу преобразования звуковых колебаний в электрические (механико-электрические характеристики);
- по принципу воздействия звуковых колебаний на диафрагму (механико-акустические характеристики);
- по зависимости уровня на выходе от пространственной ориентации (характеристики направленности);
- по принципу подключения к аудио-тракту (коммутационные характеристики).

К тому же, микрофон, сочетая в себе вышеуказанные характеристики в самых разных комбинациях, может еще и иметь разный дизайн и предназначение - петличный, подвесной, ручной, накамерный, прикрепляемый к музыкальному инструменту, настольный и т.д.

С точки зрения механико-электрического принципа существует не так много разновидностей – в настоящее время используются только конденсаторные и динамические микрофоны. Все прочие плохо подходят к применения в профессиональных реалиях.

Устройство динамического микрофона напоминает устройство динамического громкоговорителя (по этой причине последние часто применяются и в роли микрофона - в переговорных устройствах, рациях, то есть там, где компактность предпочтительнее качества звука). Диафрагма динамического микрофона связана с катушкой, находящаяся в зазоре вокруг магнита (рисунок 1.2). Продольные колебания прилегающего воздуха воздействует на диафрагму с катушкой относительно магнита с постоянным

магнитным полем, что служит причиной к возникновению на концах катушки переменного электрического напряжения и тока, амплитуда и частота которых пропорциональны уровню давления и частоте звука, действующего на диафрагму.

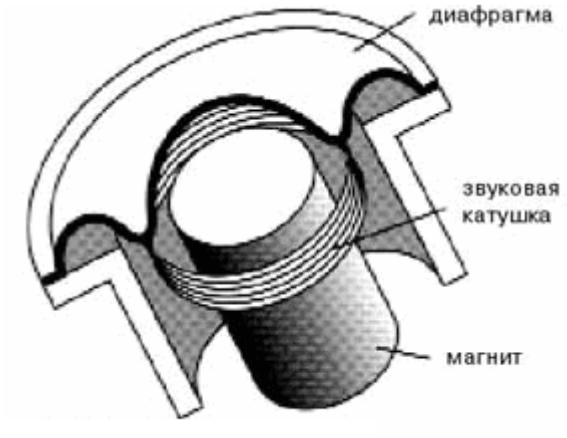


Рисунок 1.2 – Устройство динамического микрофона

В конденсаторном микрофоне звуковые колебания приводят в движение мембрану, которая является одной из обкладок конденсатора. (Рисунок 1.3) Этот конденсатор подключен последовательно с источником постоянного тока. При воздействии звука на мембрану, она начинает двигаться поступательно, вызывая изменение емкости, которое превращает постоянный ток источника в переменный. В силу ряда особенностей использования конденсатора в роли звуко-электрического преобразователя, конденсаторный микрофон всегда снабжается встроенным предусилителем, согласующим выход микрофона с входом нагрузки. Так как предложение включить конденсатор на вход усилителя низкой частоты вызовет у инженера-электронщика неадекватную реакцию.

Рисунок 1.3 – Устройство конденсаторного микрофона

Подавляющее большинство микрофонных предусилителей основаны на транзисторных микросхемах. Однако, существует ряд студийных моделей с ламповыми усилителями. Их неточно называют "ламповыми микрофонами". Некоторые звукорежиссеры отдают предпочтение ламповым усилителям, хотя коэффициент нелинейных искажений таких устройств выше чем у аналогичных транзисторных.

Конденсаторные микрофоны делятся на микрофоны с большой и малой диафрагмой. Конденсаторные микрофоны с большой мембраной в силу размеров, дизайна и цены применяются в основном в студиях звукозаписи, а микрофоны с малой мембраной имеют, соответственно меньший размер, и потому более универсальны.

Существует также основанный на конденсаторном микрофоне электретный микрофон, у которого пластины конденсатора, изготовленные из специального материала, постоянно заряжены и не нуждаются в постоянном источнике питания. Этот источник в электретных микрофонах все же имеется, но только для питания микрофонного усилителя, который так же необходим в электретных микрофонах, как и в обычных конденсаторных микрофонах.

Большинство современных конденсаторных микрофонов требуют для работы постоянное напряжение 48 В, которое подается от специального источника питания, либо с микшерного пульта, имеющего функцию "фантомного питания". некоторые видеокамеры также имеют возможность подачи фантомного питания для использования в видеосъемке внешних конденсаторных микрофонов.

По принципу устройства микрофоны делятся на приемники давления и приемники градиента давления. В микрофонах-приемниках давления звуковые колебания действуют только на фронтальную сторону диафрагмы, а в приемнике градиента давления (разности) - и на фронтальную, и на тыловую стороны. Отличия в конструкции акустических приемников сказывается, в основном, на их пространственных характеристиках.

По пространственным характеристикам микрофоны, сперва, делятся на ненаправленные и направленные. Направленность определяется как изменение чувствительности микрофона при перемещении источника звука неизменной интенсивности относительно оси, перпендикулярной плоскости диафрагмы. Естественно, что наиболее чувствителен микрофон именно по этой оси. Однако поведение микрофона по мере отклонения источника от этой оси различно.

В случае если чувствительность меняется очень слабо, микрофон является ненаправленным, и его характеристика направленности графически изображается в виде круга (рисунок 1.4).

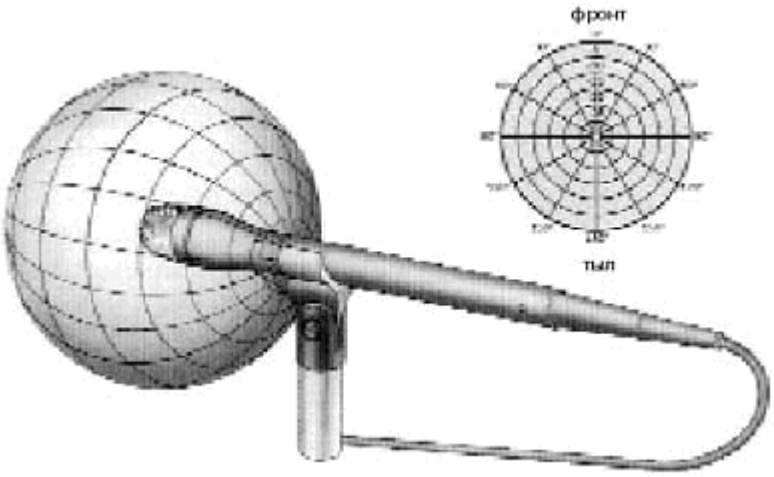


Рисунок 1.4 – Графическое изображение направленности ненаправленного микрофона

Если чувствительность в пределах фронтальной полусферы меняется мало, а чувствительность со стороны тыльной полусферы резко падает, микрофон является односторонненаправленным. Поскольку график характеристики направленности напоминает сердце ("крендель"), то такой микрофон называется кардиоидным (рисунок 1.5).

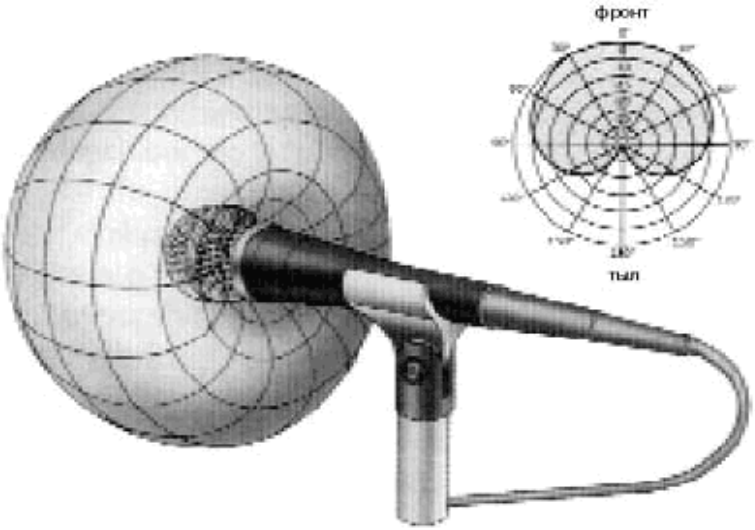


Рисунок 1.5 – Графическое изображение направленности кардиоидного микрофона

Если у кардиоидного микрофона чувствительность при отклонении от оси сильно ослабляется, образуя вытянутую кардиоиду ("грушу"), это суперкардиоидный микрофон (рисунок 1.6).

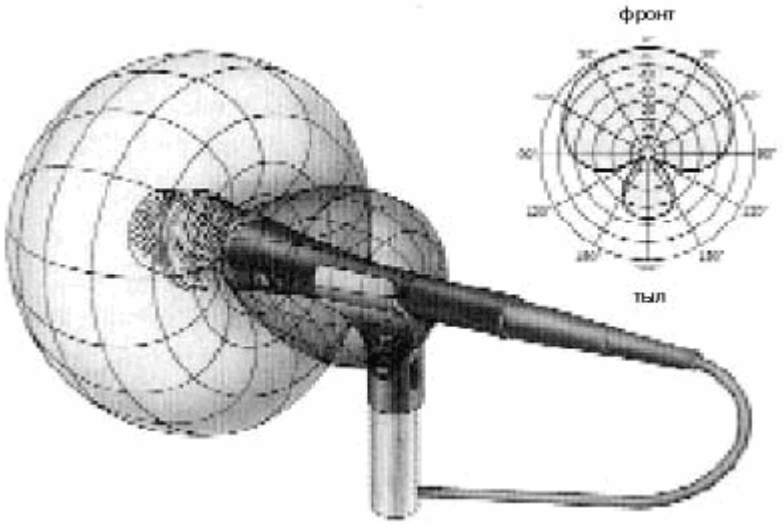


Рисунок 1.6 – Графическое изображение направленности суперкардиоидного микрофона

В случае резкого падения чувствительности микрофона при отклонении от оси, этот микрофон является гиперкардиоидным, или остронаправленным. Существуют также двусторонненаправленные микрофоны, график характеристики которых представляет собой "восьмерку".

При этом следует учитывать, что на характеристики направленности сильно влияет соотношение частоты звуковой волны или ее длины и размеров микрофона. Для звуковых волн низкой частоты (большой длины) направленность микрофонов проявляется меньше, а для высоких частот (коротких волн) – больше. Следует отметить, что человеческое ухо обладает такими же особенностями.

По физическому принципу подключения микрофоны делятся на традиционные проводные и беспроводные (радиомикрофоны). Беспроводные микрофоны представляют собой устройства, включающие в одном корпусе микрофонную головку, передатчик (трансмиттер) и приемник (ресивер). Петличные беспроводные микрофоны состоят из двух частей: самого микрофона, закрепленного на одежде и передатчика, находящегося на поясе или в кармане и соединенного с помощью аудиокабеля.

Беспроводные микрофоны создаются на базе стандартных микрофонных головок (капсюлей), поэтому их акустические характеристики практически идентичны базовым проводным аналогам.

Параметры микрофонов охватывают ряд характеристик, отраженных, как правило, в их технической документации.

Одной из важных характеристик является номинальный диапазон частот. Номинальный диапазон частот – это диапазон, в котором звуковой сигнал вызывает электрический сигнал на выходе. Чем шире диапазон частотт, тем выше класс микрофона.

Также стоит отметить такой параметр микрофона как неравномерность частотной характеристики, то есть неравномерность чувствительности микрофона от частоты сигнала. Чем линейнее частотная характеристика микрофона, то есть чем меньше разность между максимальной и минимальной чувствительностью микрофон, в номинальном диапазоне частот, тем лучше микрофон.

Чувствительность микрофона – это отношение выходного напряжения к звуковому давлению, оказываемому на диафрагму микрофона, и выражается в милливольтах на паскаль (мВ/Па). Так как звуковое давление на диафрагму микрофона может быть самым разным в зависимости от частоты, измерение чувствительности стандартизировано: оно производится в условиях действия прямой звуковой волны на частоте 1 кГц.

Как правило, чувствительность конденсаторных микрофонов намного выше чувствительности динамических.

Чувствительность микрофона может быть измерена как спереди микрофона, так и сзади. Отношение этих величин называется перепадом чувствительности "фронт/тыл". Очевидно, что данный параметр сильно зависит от типа направленности микрофона.

Также следует согласовывать выходное сопротивление микрофона и входное сопротивление усилителя. Эти пармаетры, как правило, тоже указываются для частоты 1 кГц. Эти сопротивления должны быть одного порядка, но входное сопротивление усилителя должно быть несколько выше выходного сопротивления микрофона. Это нужно для получения высокого значения отношения сигнал/шум.

Следует отличать чувствительность от уровня чувствительности, зависящего от номинального сопротивления нагрузки.

Стандартный уровень чувствительности выражается в децибелах (дБ) и отражает уровень мощности, развиваемой микрофоном в номинальную нагрузку при давлении в один паскаль. При этом, чем меньше сопротивление нагрузки (и, стало быть, выходное сопротивление микрофона), тем выше уровень чувствительности микрофона.

Еще один параметр – предельное звуковое давление. Оно измеряется в среднем диапазоне частот и показывает, при каком уровне коэффициент нелинейных искажений превысит 0,5%. Для микрофонов высокого класса предельное звуковое давление может достигать 140 дБ..

Микрофоны также имеют и собственные шумы. Уровень собственных шумов микрофона определяется как уровень звукового давления, эквивалентный уровню сигнала на выходе микрофона при отсутствии какого либо воздействия (в условиях тишины). Чем ниже уровень собственных шумов микрофона, тем выше отношение сигнал/шум и, соответственно качество выходного сигнала. Для микрофонов высокого класса уровень собственных шумов составляет 20 дБ и менее.

Динамический диапазон микрофона - это разность между предельным звуковым давлением и уровнем собственных шумов микрофона, т.е. уровень

собственных шумов является пороговым значением, т.к. полезный сигнал будет теряться в шуме, если его уровень ниже уровня собственных шумов. При превышении уровня звука предельного звукового давления микрофона, будут появляться нелинейные искажения.

Отношение/сигнал шум обычно не указывается в характеристиках микрофона, так как данный параметр не нормируется. Нельзя напрямую сравнивать данный параметр для разных микрофонов, т.к. каждый производитель производит такие замеры в разных условиях, но, в основном, это значение вычисляется как разность собственных шумов микрофона и уровня звука в 94 дБ.

В зависимости от производителя, технический паспорт микрофона может содержать множество параметров. Крупные производители даже могут прилагать к паспорту разнообразные графики испытаний, таких, как АЧХ, ФЧХ и др.

В зависимости от условий записи, следует использовать микрофон подходящей конструкции. Микрофоны могут быть на ножках, ручные, накамерные (для видеосъемки), крепящиеся на одежду и т.д. Подходящая конструкция микрофона позволит гибко располагать его в зависимости от его диаграммы направленности.

Наиболее распространенным типом является ручной микрофон, так как его конструкция является наиболее универсальной. Ручной микрофон можно расположить на стойке или держать в руке. Ручной микрофон состоит из головки, где располагается сам конденсаторный капсюль, рукоятки цилиндрической формы и разъема для подключения кабеля. Как правило, в ручных микрофонах используется разъем XLR, а к предусилителю или устройству записи подключается посредством 6,3 мм разъема. В случае, если это радиомикрофон, то в рукоятке располагается предусилитель, передатчик и автономный источник питания (аккумулятор или щелочная батарея). Ручные микрофоны обладают кардиоидной или суперкардиоидной диаграммой направленности.

Как было сказано ранее, ручные микрофоны имеют схожую форму, а вот в других типах формы микрофонов могут быть самыми разными. Высококачественные студийные микрофоны имеют форму плоского диска. Такие устройства крепятся на стойке с помощью пружин, резинок или других эластичных приспособлениях, что позволяет изолировать микрофон он колебаний здания, пола или стойки. Микрофоны, предназначенные для разговоров могут располагаться на гибких шлангах и крепиться к головным телефонам или одежде.

В отличии от ручных, "неручные" микрофоны могут иметь как направленную, так и ненаправленную диаграммы направленности. Микрофоны, имеющие ненаправленную характеристику, называют микрофонами "общего пространства зала".

Несколько позже, по сравнению с описанными выше типами, были изобретены микрофоны настольного типа, с диаграммой направленности,

имеющей форму полусферы. Такие микрофоны широко применяются на телевидении и радио. Также такой тип микрофонов используется в театрах, но их устанавливают уже на полу. Микрофоны настольного типа представляют собой плоский или слегка выпуклый, прямоугольный или круглый предмет. Благодаря такой диаграмме направленности, настольные микрофоны используются и на конференциях, т.к. может захватывать речь каждого человека за столом.

Микрофоны, предназначенные для крепления на одежду, называются петличными. Петличные микрофоны имеют маленькую чувствительную головку. Так как микрофон крпится на одежду, его сложно направить на источник звука, поэтому петличные микрофоны имеют ненаправленную диаграмму направленности. Во время совершения движений человеком, микрофон может регистрировать низкочастотные колебания, поэтому в петличных микрофонах используются фильтры.

Накамерные микрофоны, используемые в видеосъемке и репортажах, имеют узкий сектор направленности, так как в видеосъемке требуется записывать звук только с области кадра. Некоторые видеокамеры имеют предусилители с переменным коэффициентом усиления, которые прямо пропорционален фокусному расстоянию объектива. Т.е. при увеличении фокусного расстояния объектива, коэффициент усиления микрофонного усилителя увеличивается, что позволяет увеличивать громкость звука, при увеличении отдаленных объектов.

Отдельно следует отметить микрофоны, с переключаемой направленностью. Такие микрофоны подходят в тех случаях, когда нельзя точно сказать, в каких условиях будет эксплуатироваться микрофон. Переключаемая направленность достигается за счет того, что в головке микрофона расположены несколько капсюлей, соединяя которые в различных комбинациях, достигаются направленная, узконаправленная, ненаправленная и другие характеристики направленности. Такие микрофоны имеют переключатель, который располагается на корпусе микрофона.

Для расширения номинального диапазона частот, в некоторых микрофонах используют два или более капсюлей, которые используются в разных диапазонах частот, по аналогии с наушниками, имеющими несколько арматурных звуковых излучателей, или с многополосными акустическими системами. Особенностью такого приема заключается в том, что необходимо согласовывать частоту среза и фазы на граничных частотах. В случае, если сигналы с двух капсюлей на граничной частоте будут в разных фазах, может образоваться спад в АЧХ микрофона.

В микрофонах реализуют не только переключаемую направленность, но и переключаемую частотную характеристику. Такие переключатели обычно используются в ручных микрофонах для выступлений. Это позволяет быстро изменить частотную характеристику микрофона: уменьшить низкие или высокие частоты или выровнять характеристику. В других ситуациях это не

совсем оправдано, т.к. после записи уровни тех или иных гармоник можно изменить программно.

Диафрагма микрофона – очень хрупкая деталь, поэтому производители конструируют головку микрофона таким образом, чтобы предотвратить повреждение диафрагмы и самого капсюля. В ручных микрофонах капсюль защищена от механических повреждений плетёной металлической сеткой. Под сеткой располагается пропускающая звук ткань, для защиты от мелких частиц и слюны. Микрофоны, предназначенные для использования вне помещения, имеют экран из синтетического ворса, который защищает микрофон от задуваний и, соответственно, от нежелательных шумов.

Сигнал на выходе микрофона, как правило, не превышает несколько десятков милливольт, поэтому сильно подвержен внешним электромагнитным наводкам. В случае, если соединительные кабели длинные, то шум от наводок будет даже превышать уровень выходного сигнала микрофона, в таких случаях используют балансное соединение. Для этого микрофон должен обладать балансным выходом, а предусилитель балансным входом. Балансный кабель должен быть экранированным, которое соединяется с корпусом предусилителя.

Все эти сведения, казалось бы, очевидны и общеизвестны, однако плохая или неправильная экранировка, не говоря уже о ее отсутствии - главный источник проблем, как в студии, так и в звукоусилении.

Практически все современные профессиональные микрофоны имеют разъем XLR. Разъемы, использовавшиеся ранее, сейчас не используются. Разъем XLR имеет три контакта: первый – общий провод, второй – прямая полярность (плюс), третий – обратная полярность (минус).

При использовании нескольких микрофонов, нужно чтобы их фазовые характеристики были одинаковыми. Даже применяя микрофоны одного производителя, нельзя быть уверенным в том, что их фазовые характеристики совпадают. Для проверки синфазности микрофонов, достаточно суммировать их сигналы. Если громкость будет возрастать, то можно сказать, что микрофоны имеют схожие фазовые характеристики.

1.4 Подбор устройства захвата речи

При изучении руководства к использованию микрофона, одним из основных параметров является понятие как «сопротивление» или «импеданс». Это параметр, который показывает значение выходного сопротивления микрофона для переменного тока. Сопротивления измеряется в Омах (Ом). В зависимости от величины сопротивления различают высокоомные (10-50 кОм) и низкоомные (50-600 Ом) микрофоны. Выходное сопротивление микрофона должно соответствовать входному сопротивлению усилителя, иначе может снизиться отношение полезного сигнала к шуму. На это следует обращать внимание. Стоит отметить, что эти сопротивления не должны быть совершенно одинаковыми, выходное сопротивление микрофона должно быть

в допустимых границах величины, указываемой в параметрах микрофонного усилителя. Если усилитель обладает низкоомным входом, то выходное сопротивление микрофона должно быть так же низкоомным. Если подключить высокоомный микрофон к низкоомному входу усилителя, то это приведет к снижению уровня сигнала. С другой стороны, при соединении низкоомного микрофона с высокоомных входом усилителя, звук может приобрести окраску дисторшена, хотя, в общем смысле, это и не является большой проблемой.

Преимущество использования низкоомного микрофона заключается в том, что у таких микрофонов более низкий уровень шума, даже при использовании длинных кабелей.

При выборе микрофона следует учитывать не только всю совокупность его технических характеристик, но и условия записи, по этой причине конкретные рекомендации дать довольно трудно. Однако общие правила выбора микрофона все же существуют.

Ненаправленный микрофон можно применять при записи голоса и звуков в сильно заглушенном помещении. Его же следует использовать для передачи общей акустической обстановки при многомикрофонной записи.

Односторонне направленный микрофон с характеристикой типа «кардиоида» желательно применять при записи в помещении с большим количеством звуковых отражений. Применяют его и в том случае, когда в помещение, где проводят запись, проникают посторонние шумы. Микрофон следует устанавливать тыльной стороной к источнику звуковых помех. Такой микрофон рекомендуется использовать при широком фронте размещения исполнителей. Этот микрофон применяют при маловероятной в любительских условиях многомикрофонной записи для четкого разделения групп звуков, а также при размещении источника звука близко к микрофону, чтобы снизить низкочастотные искажения, присущие в этом случае ненаправленному и двусторонне направленному микрофонам.

Двусторонне направленный микрофон с диаграммой типа «восьмерка» следует применять при записи в заглушенном помещении, когда необходимо увеличить относительный уровень переотраженных сигналов, а также при записи отдельных звуков и голосов для выделения низких частот в условиях близкого размещения, источников звуков у микрофона. Используют такой микрофон и в том случае, когда необходимо отстроиться от направленных источников шума. Для этого микрофон ориентируют зоной нулевой чувствительности к источнику шума. Двусторонне направленный микрофон, сориентированный в горизонтальном направлении, оказывается полезным для ослабления звуковых волн, отраженных от пола, потолка и боковых стен помещения. Это позволяет применить акустическую обработку только двух стен: за источником звуковых волн и напротив него.

Для окончательного выбора нужно определиться в каком звуковом поле будет использоваться микрофон.

Если звук достигает микрофон, не отражаясь предварительно от других поверхностей (стен, потолка, пола), то применяется понятие «прямое звуковое поле» (рисунок 1.7).

Для достижения хороших результатов записи в условиях прямого звукового поля, необходимо направить зоны низкой чувствительности микрофона в сторону предполагаемых источников отраженного звука. Это позволит уменьшить искажения, т.к. отраженный звук может иметь отличную от неотраженного звука фазу.

В случае, когда звук, перед тем как достигнуть микрофон, отражается о какие-либо поверхности, применяется понятие «отражённое (реверберационное) звуковое поле» описывает ситуацию, (рисунок 1.8).

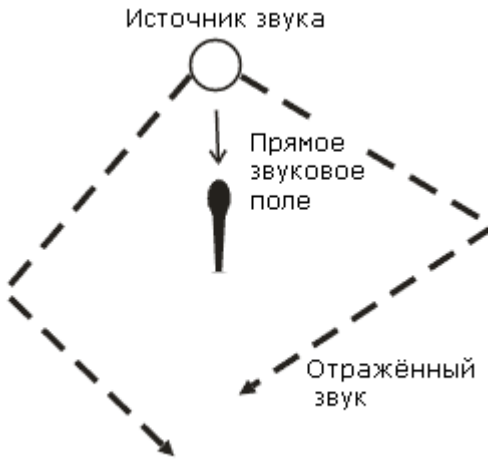


Рисунок 1.7 – Прямое звуковое поле и отраженный звук

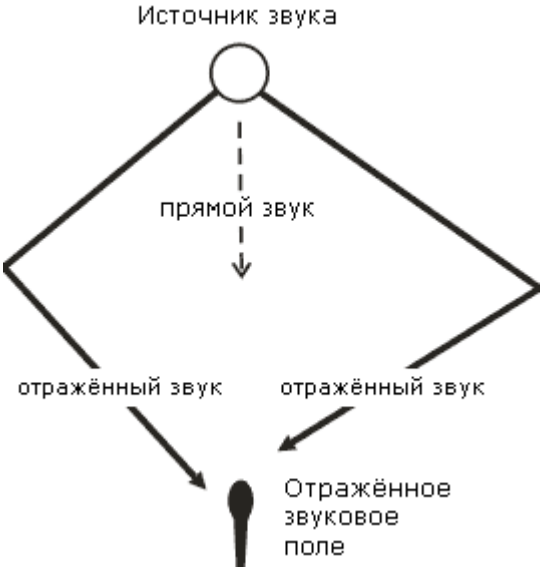


Рисунок 1.8 – Отраженное звуковое поле

В случае отраженного звукового поля, для полного подавления нежелательного шума недостаточно манипуляций с зоной низкой чувствительности микрофона, но это так же снижает уровень искажений, хоть и в меньшей степени. Поэтому в отражённом звуковом поле так же следует использовать однонаправленные микрофоны, которые обеспечивают более надёжную защиту от эффекта обратной связи (за счет отраженного звука иной фазы) и посторонних сигналов по сравнению с всенаправленными. Чем меньше сектор направленности микрофона, тем лучше он справляется с задачей. В этой связи, наиболее хорошо проявляют себя микрофоны с суперкардиоидной направленностью, затем идут кардиоидные микрофоны. Эффективность микрофона в отраженном звуковом поле выражается численной мерой, которой является коэффициент направленности, который находится в обратной зависимости от сектора направленности микрофога. Значения коэффициента для микрофонов с различными диаграммами направленности представлены в таблице 1.1:

Т а б л и ц а 1.1 – Диаграммы направленности.

Тип микрофона	Коэффициент направленности	Отсечение шумов, дБ
Всенаправленный	1	0 дБ
Кардиоидный	1.7	4.8 дБ
Суперкардиоидный	1.9	5.7 дБ

Данные таблицы свидетельствуют о том, что коэффициент направленности суперкардиоидного микрофона в 1.9 раза превышает соответствующий показатель для всенаправленного микрофона. В реальных условиях это выражается в следующем: в отражённом звуковом поле суперкардиоидный микрофон снимает на 5.7 дБ меньше реверберационных шумов.

Для реализации голосового управления, требуется микрофон высокого качества. Так как нужно производить захват речи только оператора, целесообразно использовать узконаправленный микрофон, что позволит отсечь посторонние звуки и голоса. Также для отсечения посторонних вуков, лежащих вне диапазона голоса человека, требуется фильтрация звукового сигнала.

Исходя из изложенного выше стоит остановить выбор на микрофоне, встроенном в веб-камеру Logitech C270 (рисунок 1.9). Это узконаправленный микрофон высокого качества, с небольшим внутренним сопротивлением, представляющий собой конденсаторный микрофон. Устройство имеет встроенные уилитель, АЦП и частотный фильтр, позволяющий снизить уровень шумов. Кроме того, частотная характеристика модели оптимизирована для записи человеческой речи. Характеристики микрофона веб-камеры Logitech C270 представлены в таблице 1.2.

Рисунок. 1.9 – Веб-камера Logitech C270

Т а б л и ц а 1.2 – Характеристики микрофона камеры Logitech C270

Параметр	Значение
Диаграмма направленности	узконаправленный
Частотная характеристика	200-7000 Гц
Тип микрофона	конденсаторный
Чувствительность	1.4 мВ / Па
Сопротивление	75 Ом
Макс. звуковое давление	100, 1 КГц на 3 % КНИ
Соотношение сигнал/шум	68 дБ, при 1 КГц
Масса	80 г

1.5 Поиск оптимальной аудио-системы распознавания речи на базе открытого исходного кода

Из наиболее распространенных программных продуктов распознавания речи можно выделить следующие: CMU Sphinx, Julius, Simon software, iATROS, RWTH ASR (как разновидность Q Public License (QPL) лицензии), SHoUt, VoxForge (как разновидность — Open source acoustic models and speech corpus, то есть речевой модели как корпуса). Рассмотрим их более подробно.

CMU Sphinx – для краткости далее будет называться Sphinx, была написана разработчиками программ распознавания речи университета Carnegie Mellon. Программный пакет Sphinx включает в себя серию распознавателей речи (Sphinx 2-4), программу для адаптации звуковых моделей (Sphinx train).

В 2000 году разрработчики Sphinx университета Carnegie Mellon утвердила компоненты системы распознавания речи с открытым исходным кодом, включая Sphinx 2 и Sphinx 3 (позже, в 2001 году). Речевой декодер включал в себя простые приложения и акустические модели. Имеющиеся ресурсы включали в себя языковую модель, добавочное программное обеспечение для тренировки акустической модели, словарь произношения и лингвистическую модель комплиляции, который является общественным достоянием (cmudict).

Sphinx – это распознаватель непрерывной речи, который работает с любым диктором, использует n-граммную статистическую языковую модель и

Скрытую Марковскую модель. Она была разработана Kei-Fu Li. Sphinx способен распознавать продолжительную речь, имеет дикторонезависимый огромный словарь распознавания, то есть те возможности, которые в 1986 году вызвали большие разногласия в среде распознавания речи. Sphinx в историческом развитии примечателен тем, что в своем развитии затмил все предыдущие версии по своей производительности.

Sphinx2 – ориентированная на производительность и самая быстрая программа распознавания речи, разработанная Чудон Хуаном в университете Carnegie Mellon, и выпущенная с открытым исходным кодом на основе BSD-лицензии на SourceForge Кевином Ленсо в Linux World в 2000 году. Sphinx2 ориентирован на распознавание речи в режиме реального времени и идеально подходит под создание различных мобильных приложений. Sphinx2 включает в себя такую функциональность как конечный указатель, частичная генерация гипотез, подключение динамичной языковой модели и так далее. Она использует диалоговую систему и языковую систему обучения. Она может использоваться на компьютерах по модели PBX, таких как Asterisk. Код Sphinx2 был инкорпорирован в многочисленные коммерческие продукты.

К сожалению, Sphinx2 недолго развивался (не более чем плановое техническое обслуживание). Текущий декодер для распознавания в режиме реального времени интегрирован в PocketSphinx.

Sphinx3 представлял собой наполовину непрерывную акустическую модель распознавания речи (использовали Гауссовые смеси всех видов с индивидуальной моделью, который учитывал вес вектора над этими Гауссами). Sphinx3 принял распространенную продолжительную модель построенную на скрытых Марковских моделях и первоначально использовался для точного распознавания речи, которое осуществлялось в "пост-фактум" режиме (не реальноо времени). Последние разработки (алгоритмы и программное обеспечение) способствовали тому, что Sphinx3 мог распознавать в режиме близком к режиму реального времени, хотя приложение еще не подходило для качественного использования в роли приложения. Sphinx3 после активного развития и воссоединения с SphinxTrain обеспечил доступ к многочисленным современным методикам и моделям, таким как LDA/MLLT, MLLR и VTLN, которые улучшили точность распознавания речи.

Sphinx4 – полностью переписанный речевой движок Sphinx, главная цель которого обеспечить гибкий каркас для для осуществления разработок в распознавании речи. Sphinx4 полностью написан на языке программирования Java. Существенный вклад в развитие Sphinx4 внесла Sun Microsystems, и оказала помощь в программной экспертизе проекта. Отдельные разработчики проекта – это специалисты из MERL, MIT, CMU. Текущие цели развития включают в себя:

- развитие новых акустических моделей для тренировки;
- реализация системы речевой адаптации (MLLR);
- улучшения менеджмента конфигурации;

выпуск ConfDesigner – графической оболочки для PocketSphinx – эта версия Sphinx может быть встроена в любые другие системы на базе ARM-микропроцессоров, такие как pcDuino, Raspberry Pi, Odroid. PocketSphinx активно развивается и встраивается в различные системы с арифметикой фиксированной запятой и в эффективные модели на базе смешанной модели вычислений.

Julius – это высокопроизводительная программа распознавания непрерывной речи с большим словарем (large vocabulary continuous speech recognition), декодер программного обеспечения для исследования в области связанной речи и разработки. Julius идеально подходит для декодирования в режиме реального времени почти на всех современных компьютерах, со словарем 60 тысяч слов, используя задачи тиаграмма слова и контекстно независимую Скрытую Марковскую модель. Главной особенностью проекта является его полная встраиваемость, также стоит отметить то, что безопасная модуляция может быть независима от модельных структур и различных типов Скрытых Марковских моделей, которая поддерживает общее состояние трифонов и связанной смеси-моделей со множеством микстур, фонем и утверждений. Стандартные форматы активны за счет свободного моделирования инструментария. В основном, Julius рассчитан на работу на платформах Linux и других UNIX-подобных станциях, но система работает также на Windows. Julius является open-source системой и распространяется с BSD типом лицензии.

Julius – разрабатывается как часть свободного программного обеспечения для исследования в области распознавания японского языка, начиная с 1997 года и эта работа продолжалась в рамках Continuous Speech Recognition Consortium (консорциум распознавания непрерывной речи) с 2000 по 2003 год в Японии.

Начиная с версии 3.4 грамматическая база системы распознавания речи анализатора называется Julian и интегрирована в Julius. Julian – модифицированная форма Julius, которая использует собственную спроектированную форму грамматики Конечного автомата (Finite-state machine) как языковую модель. Она может использоваться для строительства систем голосовой навигации с небольшим словарем или других разговорных систем для распознавания разного рода диалогов.

Для запуска распознавателя речи Julius требуется произвести подбор языковой модели и акустической модели для требуемого языка. Julius адаптирует акустическую модель кодировочного формата HTK ASCII, базу данных произношения формата HTK, и 3-х слойных диаграм построения языковой модели стандарта ARPA (2 прямых и 3 реверсивных модели тренировки обернутые в речевой корпус с вывернутым заказчиком слов).

Хотя Julius распространяется только для модели японского языка, проект VoxForge работает над созданием акустической модели для английского языка, используя движок системы распознавания речи Julius.

RWTH ASR (кратко RASR) – это инструментарий распознавания речи с

открытым исходным кодом. Инструментарий включает в себя технологию умения распознавать речь для создания автоматических систем распознавания речи. Данная технология развивается Технологическим центром Естественного языка и Образцовой распознавательной группой в Рейнско-Вестфальском техническом университете Ахена.

RWTH ASR включает в себя инструментарий для разработки акустических моделей и декодеры а также компоненты для адаптации речи спикера, адаптивные системы обучения речи спикера, неконтролируемые системы обучения, дифференциальные системы обучения и решетчетые словообразные формы обработки. Данное программное обеспечение работает на ОС Linux и Mac OS X. Домашняя страница проекта предлагает уже готовые для использования модели для исследования с задачами, обучающими системами и всесторонней документацией.

Инструментарий публикуется под лицензией open source, которая называется «RWTH ASR лицензия», которая является производной от лицензии QPL (Q Public License). Эта лицензия представляет возможность свободного использования, включая повторное распространение и модификацию для некоммерческого использования.

Simon – система распознавания речи, основанная на речевых движках Julius и HTK. Система Simon спроектирована таким образом, что она довольна удобна для работы с различными языками и разного рода диалектами. При этом реакция распознавания речи полностью настраиваемая и она не подходит для исключительного распознавания единичных голосовых запросов и не может быть сконфигурирована под нужды пользователей.

Чтобы легко использовать систему необходимо выполнить определенные «сценарии». Пакеты Simon сконфигурированы для специальных задач. Среди возможных сценариев Simon, например «Firefox» (запуск и управление браузером «Firefox») или «пакет управления окном» (закрытие, движение, изменение размеров окон) и так далее. Сценарии легко могут быть создаваться пользователями и раздаваться в сообществе через Get Hot New Stuff систему. На сегодняшнее время написано более 39 сценариев и опубликовано 3 языка на репозитарии opendesktop.org.

Simon также поддерживает простые генерические и основные модели подобные GPL моделям из Voxforge, которые применяют пользователи для распознавания на английском, немецком и португальских языках. При этом нет необходимости в адаптации или тренировке модели для того, чтобы она начала корректно работать. С демонстрацией Simon 0.3.0 можно ознакомиться на официальном сайте разраотчиков. При этом речь пользователя включает в себя техническую терминологию – это главная особенность реализации в simon – и тем самым она демонстрирует как можно выгодно использовать simon для пользователей и как simon может быть интегрирован пользователями для усиления своих собственных разработок.

iATROS – это переделанная программа распознавания речи предыдущего поколения ATROS, которая подходит для распознавания как

речи, так и для распознавания текста, написанного от руки. iATROS основан на модулярной структуре и может использоваться как для построения дифференцированных моделей, чья цель осуществить Ветибри поиск на основе скрытой Марковской модели. iATROS может работать как с использованием сети интернет, так и в режиме оффлайн (основываясь на ALSA модулях).

iATROS состоит из 2-х модулей предварительной обработки (для речевого сигнала и изображений написанных от руки) и модуля ядра распознавания. Предварительная обработка данных и черты извлечения модулей обеспечиваются векторами распознавания модулей, которые используют Скрытые Марковские модели и языковые модели, которые исполняются поиском предположений из лучших систем распознавания речи. Все эти модули выполнены на языке программирования “C”.

SHoUt – это инструментарий, написанный Марином Хубрехтсом в Нидерландском институте образа и звука, который предназначен для распознавания непрерывной речи с большим словарем данных. SHoUt состоит из приложения для тренировки статистических моделей и для (не)речевого детекционирования, диаризации и декодирования речи.

VoxForge – это свободный речевой движок и языковая акустическая модель, представленная в общедоступном хранилище данных. VoxForge был собран как репозиторий транскрипций речи, имеющий свободный GPL корпус для использования с речевыми движками, которые имеют открытый исходный код. Речевые аудио-файлы могут быть собраны в акустические модели для использования с системами распознавания речи с открытым исходным кодом типа Julius, Sphinx и HTK (но HTK имеет ограничение по дистрибьюции).

HTK – это инструментарий для распознавания речи, использующий скрытую марковскую модель. Его главным образом предназначают для распознавания речи, однако он также используется для других иных задач, в которых используется скрытая Марковская модель – синтез речи, распознавание визуальных образов (мимики, черт лица и т.д.).

Таким образом, рассмотрев самые распространенные системы распознавания речи с открытым исходным кодом необходимо заметить, что наиболее сильно представлено решение на базе CMU Sphinx, особенно PocketSphinx. Оно более всего подходит для задач распознавания на базе ARM-миникомпьютера.

2 Конструкторская часть

2.1 одноплатный ARM-миникомпьютер pcDuino1

В данный момент широкое распространение получили системы на основе ARM-миникомпьютеров, таких как Raspberry Pi, BeagleBone, Odrioid, pcDuino и другие. Они различаются производительностью процессора, объемами ОЗУ и ПЗУ, количеством портов ввода-вывода, поддерживаемыми интерфейсами, ценой и другими показателями. Система голосового управления была реализована на миникомпьютере pcDduino V1 производства компании SparkFun.

Рассмотрим эту платформу подробнее и более внимательно, т.к. именно в ней и заключается решение поставленной задачи. Разработчики обозначили свой продукт как PC + Arduino, т.е. Arduino-совместимый миникомпьютер. Оправданный ход, т.к. SparkFun также производит множество плат расширения для Arduino, и большинство из них можно использовать с pcDuino. Ядром платы является система на кристалле Allwinner A10, содержащее одно ядро ARM Cortex-A8 на частоте 1 ГГц и графический процессор Mali 400. В таблице 2.1 приведены основные технические характеристики pcDuino V1 [4].

Т а б л и ц а 2.1 –Характеристики pcDuino V1

Параметр	Значение
Процессор	1 ГГц ARM Cortex A8
Графический процессор	Mali 400
ОЗУ	1 Гб
ПЗУ	2 Гб + слот для карт расширения памяти
Видеовыход	HDMI
ОС	Lubuntu, Android
Связь	Ethernet 100 mbps, UART, I2C, SPI
Порты ввода-вывода	14 цифровых, 2 ШИМ, 2 6-битных и 4 12-битных аналоговых входа, 4 SPI, 2 I2C
Порты USB	2

Архитектура миникомпьютера pcDuino V1 и его интерфейс приведены на рисунке 2.1.

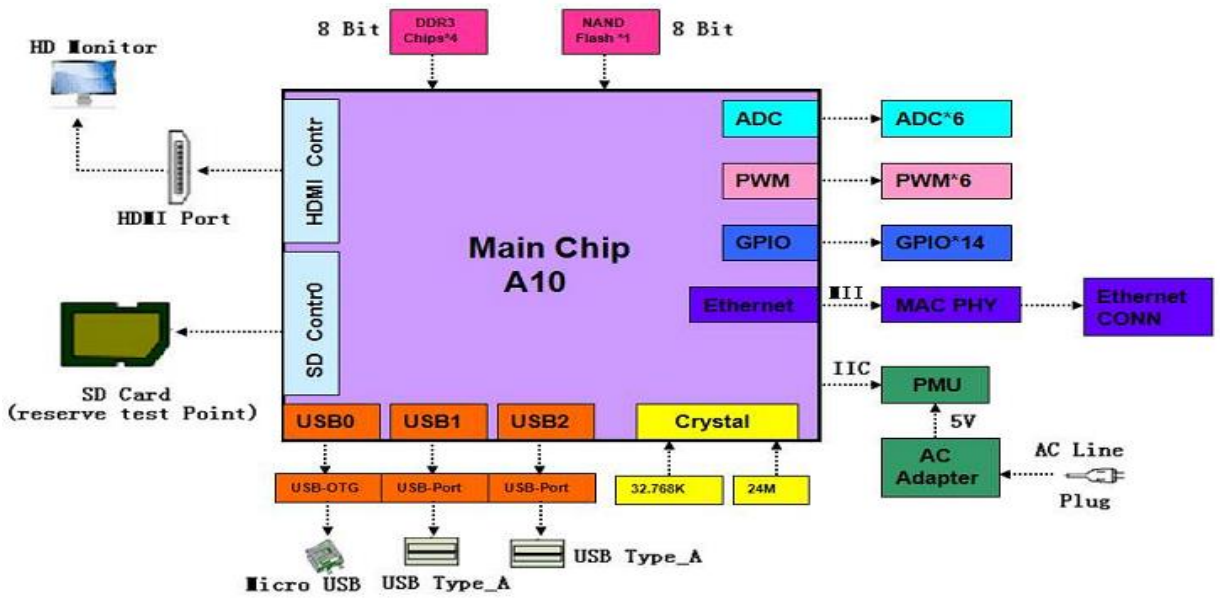


Рисунок 2.1 – Архитектура миникомпьютера pcDuino V1

2.2 Wi-Fi-плата на базе микросхемы ESP8266

Плата ESP8266 создана для использования в умных розетках, mesh-сетях, IP-камерах, беспроводных сенсорах, носимой электронике и так далее. Одним словом, ESP8266 появился на свет, чтобы стать мозгом грядущего «Интернета вещей».

Предусмотрено два варианта использования чипа: 1) в виде моста UART-WIFI, когда модуль на базе ESP8266 подключается к существующему решению на базе любого другого микроконтроллера и управляется AT-командами, обеспечивая связь решения с инфраструктурой Wi-Fi; 2) реализуя новое решение, использующее сам чип ESP8266 в роли управляющего микроконтроллера.

Чип ESP8266 является одним из самых высокоинтегрированных решений для работы с WiFi. Внутри чипа уместилась куча всего того, что в конкурирующих решениях часто является частью внешней обвязки. Структурная схема микросхемы ESP8266 приведена на рисунке 2.2.

В итоге типовая обвязка чипа состоит всего из нескольких элементов. Меньше элементов - меньше цена компонентов, меньше стоимость пайки, меньше площадь размещения, меньше стоимость печатной платы. Что прекрасно подтверждается актуальными ценами модулей на базе героя нашего сегодняшнего обзора. Управляет всем этим интегрированным хозяйством расширенная версия 32-битного процессора Tensilica L106 Diamond series.

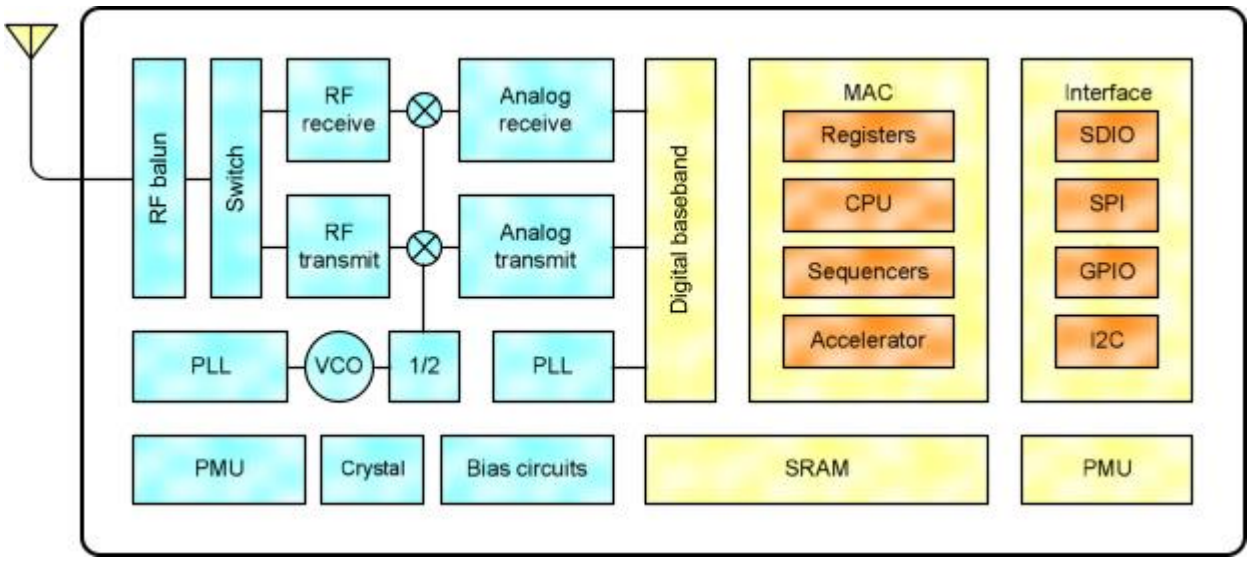


Рисунок 2.2 - Структурная схема микросхемы ESP8266

Tensilica L106 Diamond series - 32-битный микроконтроллер с RISC-архитектурой. Микроконтроллеh работает на частоте 160 МГц, и использует 24- и 16-битные инструкции.

Плата (рисунок 2.3) также обладает низким энергопотреблением. В режиме передачи она потребляет 215 мА, в режиме глубокого сна - 60 мкА, а в режиме поддерживания связи с точкой доступа - 1 мА.

Рисунок 2.3 – Wi-Fi плата на базе ESP8266

2.3 Модули реле

Для реализации интерфейса между объектами управления и микроконтроллером, были использованы два релейных модуля, схема и внешний вид которых приведены на рисунках 2.4 и 2.5.

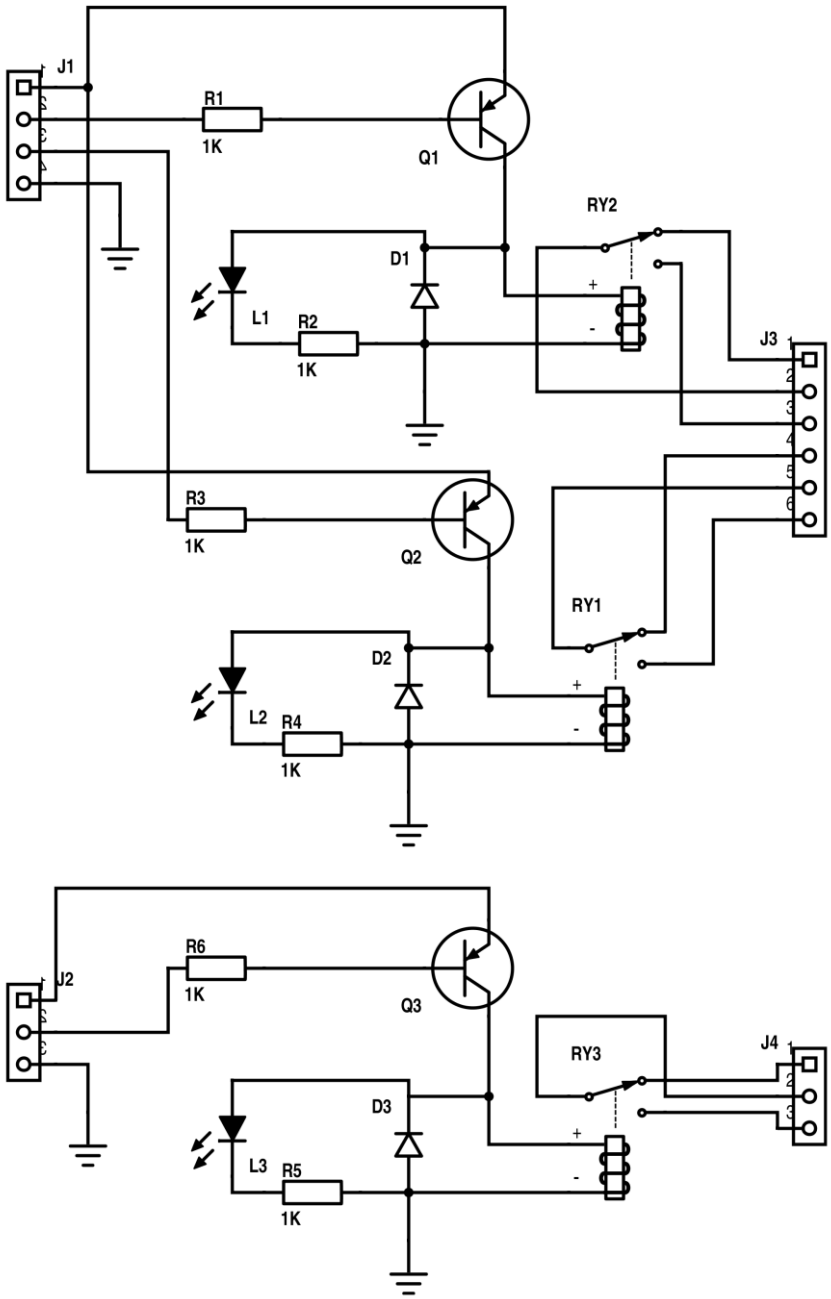


Рисунок 2.4 – Электрическая схема релейных модулей

Реле рассчитаны на управляющее напряжение в 5 В, и может коммутировать переменные напряжение и ток до 250 В и 10А соответственно. В режиме удержания якоря катушка реле потребляет 80 мА.

Рисунок 2.5 – 2-канальный релейный модуль

2.4 Беспроводной маршрутизатор TP-LINK TL-WR702N

Беспроводной маршрутизатор TL-WR702N (рисунок 2.6) является многофункциональным беспроводным сетевым устройством, поддерживает рабочие режимы точка доступа, маршрутизатор, мост, клиент и ретранслятор, позволяя пользователям более эффективно работать с различными беспроводными приложениями.

Рисунок 2.6 – Маршрутизатор TP-LINK TL-WR702N

Маршрутизатор использует режимы шифрования WPA-PSK/WPA2-PSK, которые позволяют эффективно защитить беспроводную сеть. Функция предварительного шифрования, которая устанавливает начальное имя беспроводной сети и пароль для пользователей в целях защиты беспроводной сети, повышает его защищенность. Модель TL-WR702N оборудована портом Micro USB и может получать питание от внешнего адаптера питания или

через подключение к порту USB компьютера. Полные характеристики устройства приведены в таблице 2.2.

Т а б л и ц а 2.2 - Характеристики маршрутизатора TP-LINK TL-WR702N

Параметр	Значение
Стандарты беспроводной передачи данных	IEEE 802.11n, IEEE 802.11g, IEEE 802.11b
Размеры (ШхДхВ)	57 x 57 x 18 мм
Тип антенны	Встроенная
Диапазон частот	2,4-2,4835 ГГц
Защита беспроводной сети	64/128/152-битный WEP, WPA/WPA2, WPA-PSK/WPA2-PSK
Сетевая безопасность (firewall)	Защита от DoS-атак, межсетевой экран SPI, фильтрация по IP-адресу/MAC-адресу/доменному имени, связывание IP- и MAC-адресов
Параметры окружающей среды	Рабочая температура: 0°C~40°C (32°F~104°F) Температура хранения: -40°C~70°C(-40°F~158°F)

2.5 Конструкция системы голосового управления

Основой системы является ARM-миникомпьютер pcDuino1. К нему посредством интерфейса USB подключена веб-камера со встроенным микрофоном Logitech C270, а посредством Ethernet миникомпьютер подключен к маршрутизатору TP-LINK TL-WR702N для связи с платой ESP8266, к выводам которой подключены модули реле. Структурная схема системы голосового управления приведена на рисунке 2.7.

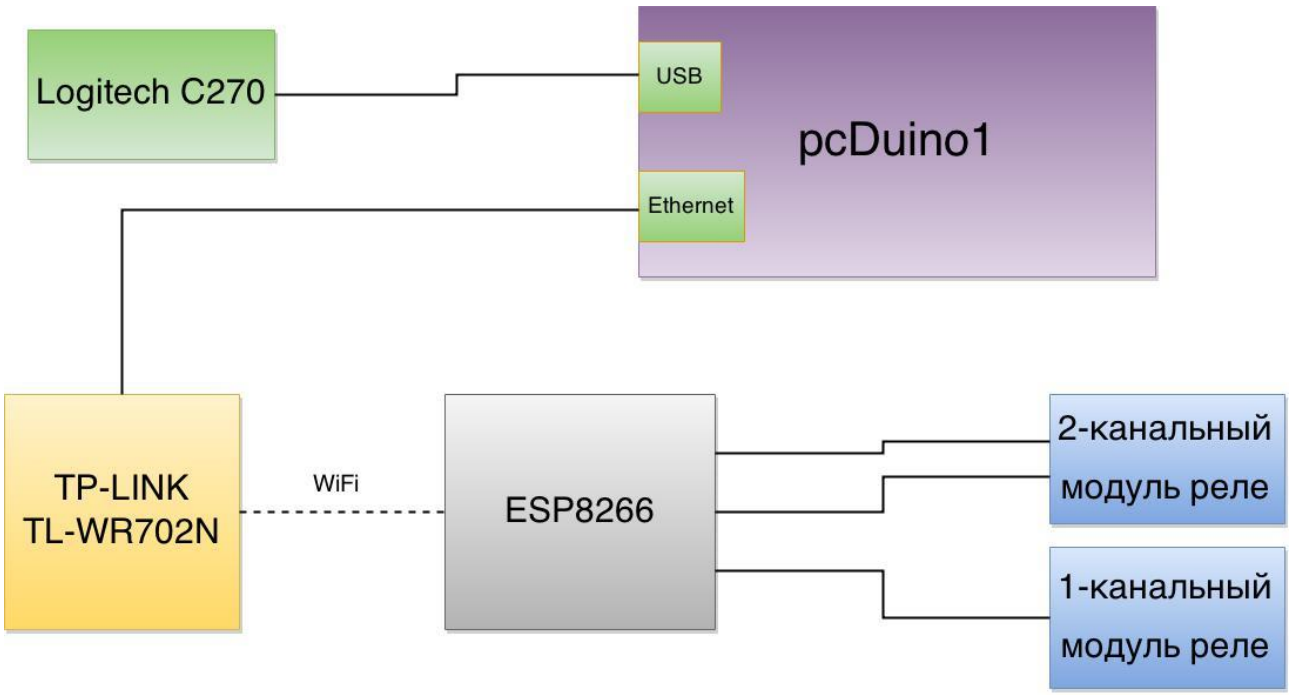


Рисунок 2.7 – Структурная схема системы голосового управления

3 Программное обеспечение

3.1 Операционная система Ubuntu

На миникомпьютер pcDuino 1 была установлена операционная система Lubuntu. Операционная система Lubuntu идентична системе Ubuntu, основное отличие Lubuntu — малые требования к ресурсам компьютера за счет использования среды рабочего стола LXDE, поэтому причине вся документация к Ubuntu применима и к Lubuntu. Ubuntu — несомненно, популярнейший дистрибутив операционной системы Linux. Ubuntu — это серьезный соперник в мире настольных систем и Webсерверов. Эта система была вызвана к жизни Марком Шаттлуортом (Mark Shuttleworth). Он основал компанию Canonical, которая и руководит разработкой Ubuntu Linux.

Эта система базируется на идее свободного, как экономически, так и юридически, программного обеспечения. Это означает, что можно устанавливать его бесплатно, а также изменять и распространять без каких-либо вознаграждений или авторских отчислений, что достигается выпуском Ubuntu на условиях специальной лицензии, защищающей эти права.

Таким образом, вы можете устанавливать Ubuntu на любой компьютер, на котором эта система сможет работать, и обновлять ее до самой последней версии так часто, как вам захочется, не беспокоясь о лицензиях, активации продукта или специальных ключах, которые вы должны ввести. Вы также можете ее распространять — раздавать своим друзьям или родственникам и помогать им в установке системы на их компьютеры.

Кроме того, Ubuntu очень легко устанавливается, автоматически распознавая используемые вами устройства и устанавливая для них самые подходящие драйверы. От других операционных систем, установка которых занимает от одного часа до половины дня, Ubuntu отличается по-настоящему выгодно, требуя на установку в среднем менее часа.

Уже более десяти лет ОС Linux обеспечивает вычислительные мощности миллионов Web-серверов, что не удивительно, учитывая объем работ, вложенных в разработку графических интерфейсов. В наши дни настольные установки Linux составляют 1% от общего числа установок (и эта цифра быстро растет) и львиную долю в них занимает Ubuntu.

Операционную систему Lubuntu можно установить во встроенную NAND-память pcDuino, или можно создать загрузочную карту памяти. Т.к. объем встроенной NAND-памяти всего 2 Гб, выбор был сделан в пользу карты microSD.

3.2 Создание загрузочной карты памяти с Lubuntu

Для подготовки загрузочной карты памяти, необходимо иметь загрузочный образ операционной системы Lubuntu и карту памяти microSD.

Для создания загрузочной карты была установлена программа Win32 Disk Image на персональный компьютер под управление ОС Windows. С помощью программы Win32 Disk Image был записан образ операционной системы Lubuntu на карту памяти. Процесс создания загрузочной карты памяти показан на рисунке 3.1.

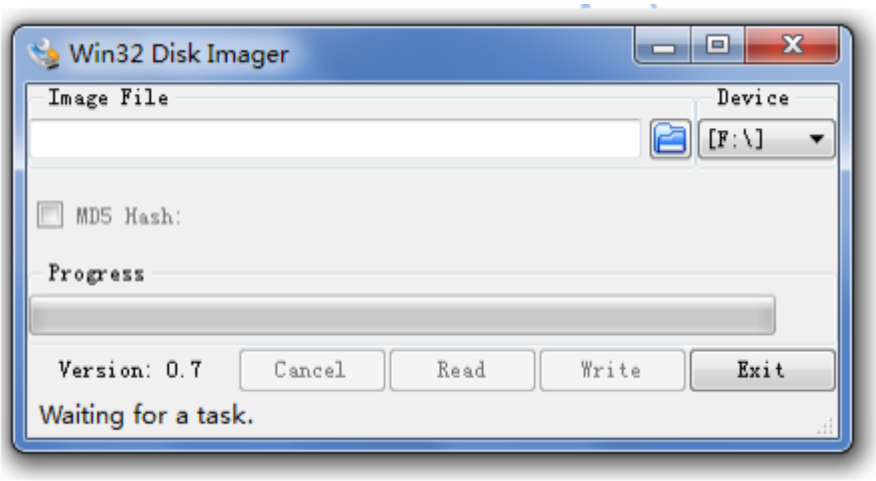


Рисунок 3.1 – Окно программы Win32 Disk Image

3.3 Программный пакет CMU sphinx и его установка

Программный пакет CMU sphinx – ведущий инструмент в области распознавания речи. Программный пакет CMU sphinx включает в себя несколько программ для различных задач и применений:

Pocketsphinx — небольшая библиотека написанная на языке Си;

Sphinxbase —дополнительная библиотека, требуемая для работы Pocketsphinx;

Sphinx4 — гибкая программа распознавания речи, написанная на Java;

Sphinxtrain — программа для адаптации звуковых моделей.

На миникомпьютер pcDuino V1 были установлены pocketsphinx-0.8, sphinxbase-0.8, акустическая модель русского языка, а также созданы языковая модель и словарь. Для работы программы требуются акустическая модель языка, языковая модель и словарь, где акустическая модель описывает звуки слов, а языковая – априорную вероятность каждого фрагмента речи, например, что словосочетание "секретный код" является более вероятной, чем "секретный кот". Словарь – это набор слов, которые могут быть распознаны. Словарь состоит из следующих слов: "включить", "выключить", "реле", "номер", "один", "два" и "три", т.к. требуется управление тремя модулями реле.

Т.к. пакет программ CMU sphinx отсутсвует в стандартных репозиториях, pocketsphinx-0.8 и sphinxbase-0.8 были установлены вручную. Для этого с официального сайта были загружены архивы пакетов pocketsphinx-0.8 и sphinxbase-0.8, распакованы с помощью команды "tar" через

терминал операционной системы Lubuntu. Далее для каждого пакета были выполнены команды:

% ./configure
% make
% make install

3.4 Настройка и запуск pocketsphinx

Для запуска программы выполняется команда pocketsphinx_continuous с набором необходимых параметров:

- hmm – путь к директории, содержащей файлы акустической модели;
- jsgf – путь к языковой модели;
- dict – путь к файлу словаря.

Эти параметры записываются в файл аргументов, для того, чтобы не прописывать их при каждом запуске pocketsphinx. Тогда команда, набираемая в терминале Lubunut будет иметь следующий вид: "pocketsphinx_continuous -argfile /home/ubuntu/pocketsphinx-0.8/argfile", где -argfile – путь к файлу, содержащего все аргументы. Содержимое файла аргументов приведена на рисунке 3.2.

Рисунок 3.2 – Содержимое файла аргументов argfile

Вместо статической языковой модели, был использован JavaScript Grammar File (JSGF), содержимое которого показано на рисунке 3.3.

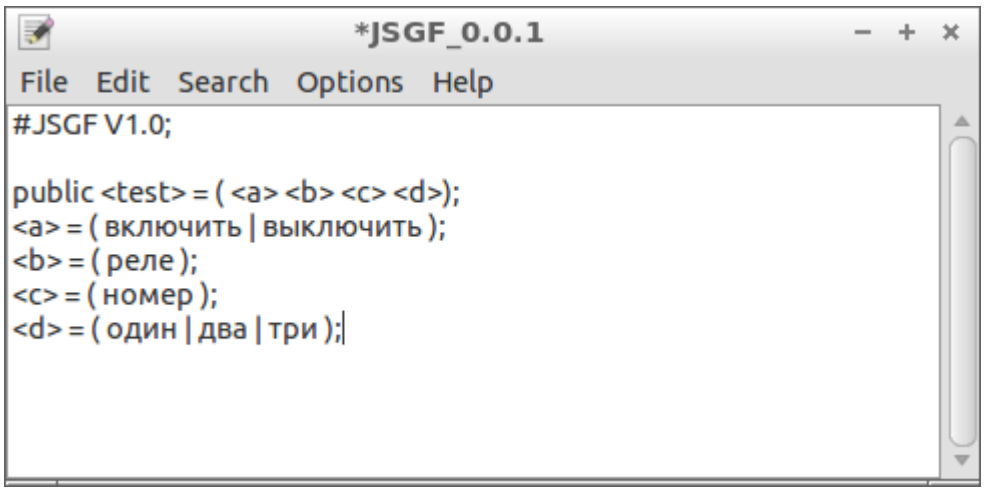


Рисунок 3.3 – Содержимое файла языковой модели jsgf

В этом файле указано то, что команда состоит из четырех слов. Первое слово может быть "включить" или "выключить", второе – "реле", третье – "номер", а четвертое – "один", "два" или "три".

Акустическая модель русского языка была загружена с репозитория voxforge. Voxforge – репозиторий акустических моделей, где можно загрузить акустические модели русского, английского, болгарского, французского, итальянского и многих других языков для sphinx и julius.

Набрав в терминале команду

% pocketsphinx_continuous -argfile /home/ubuntu/pocketsphinx-0.8/argfile, получаем окно, показанное на рисунке 3.4. После инициализации программы, появляется строчка "READY...", которая показывает, что программа начала работу, и готова распознавать команды. При появлении какого либо сигнала на микрофоне, начинается захват и обработка этого сигнала, и в течение двух секунд выдается результат. Результат распознавания команды "включить реле номер один" приведен на рисунке 3.5.

Для управления устройствами, необходимо эти команды выделить. Для этого при запуске добавляем параметр "> /home/ubuntu/sphinx_pg/out", после чего результат распознавания будет записываться в текстовый файл для последующей обработки.

```
ubuntu@ubuntu: ~
File  Edit  Tabs  Help
INFO: jsgf.c(581): Defined rule: <test.g00002>
INFO: jsgf.c(581): Defined rule: <test.a>
INFO: jsgf.c(581): Defined rule: <test.g00004>
INFO: jsgf.c(581): Defined rule: <test.b>
INFO: jsgf.c(581): Defined rule: <test.g00006>
INFO: jsgf.c(581): Defined rule: <test.c>
INFO: jsgf.c(581): Defined rule: <test.g00008>
INFO: jsgf.c(581): Defined rule: <test.d>
INFO: fsg_model.c(215): Computing transitive closure for null transitions
INFO: fsg_model.c(270): 36 null transitions added
INFO: fsg_model.c(421): Adding silence transitions for <sil> to FSG
INFO: fsg_model.c(441): Added 27 silence word transitions
INFO: fsg_search.c(366): Added 0 alternate word transitions
INFO: fsg_lextree.c(108): Allocated 2916 bytes (2 KiB) for left and right contex
t phones
INFO: fsg_lextree.c(253): 67 HMM nodes in lextree (40 leaves)
INFO: fsg_lextree.c(255): Allocated 7236 bytes (7 KiB) for all lextree nodes
INFO: fsg_lextree.c(258): Allocated 4320 bytes (4 KiB) for lextree leafnodes
INFO: continuous.c(371): pocketsphinx_continuous COMPILED ON: Apr 21 2014, AT: 1
0:26:31

Warning: Could not find Capture element
READY....
```

Рисунок 3.4 – Окно терминала, после запуска pocketsphinx

```
ubuntu@ubuntu: ~
File  Edit  Tabs  Help
000000002: включить реле номер три
READY....
Listening...
Recording is stopped, start recording with ad_start_rec
Stopped listening, please wait...
INFO: fsg_search.c(1032): 186 frames, 2055 HMMs (11/fr), 6008 senones (32/fr), 1
350 history entries (7/fr)

INFO: fsg_search.c(1417): Start node <sil>.0:2:32
INFO: fsg_search.c(1417): Start node <sil>.0:2:32
INFO: fsg_search.c(1417): Start node <sil>.0:2:32
INFO: fsg_search.c(1417): Start node <sil>.0:2:32
INFO: fsg_search.c(1456): End node <sil>.183:185:185 (-518)
INFO: fsg_search.c(1456): End node <sil>.183:185:185 (-518)
INFO: fsg_search.c(1456): End node <sil>.183:185:185 (-518)
INFO: fsg_search.c(1456): End node <sil>.183:185:185 (-518)
INFO: fsg_search.c(1456): End node <sil>.183:185:185 (-518)
INFO: fsg_search.c(1456): End node один.146:167:185 (-509)
INFO: fsg_search.c(1680): lattice start node <s>.0 end node </s>.186
INFO: ps_lattice.c(1365): Normalizer P(O) = alpha(</s>:186:186) = -172388
INFO: ps_lattice.c(1403): Joint P(O,S) = -172388 P(S|O) = 0
000000003: включить реле номер один
READY....
```

Рисунок 3.5 – Результат распознавания команды

3.5 Обработка результатов распознавания

После действий, проделанных в главе 3.4, распознанные речевые команды записываются в текстовый файл "out". Содержимое этого файла показана на рисунке 3.6.

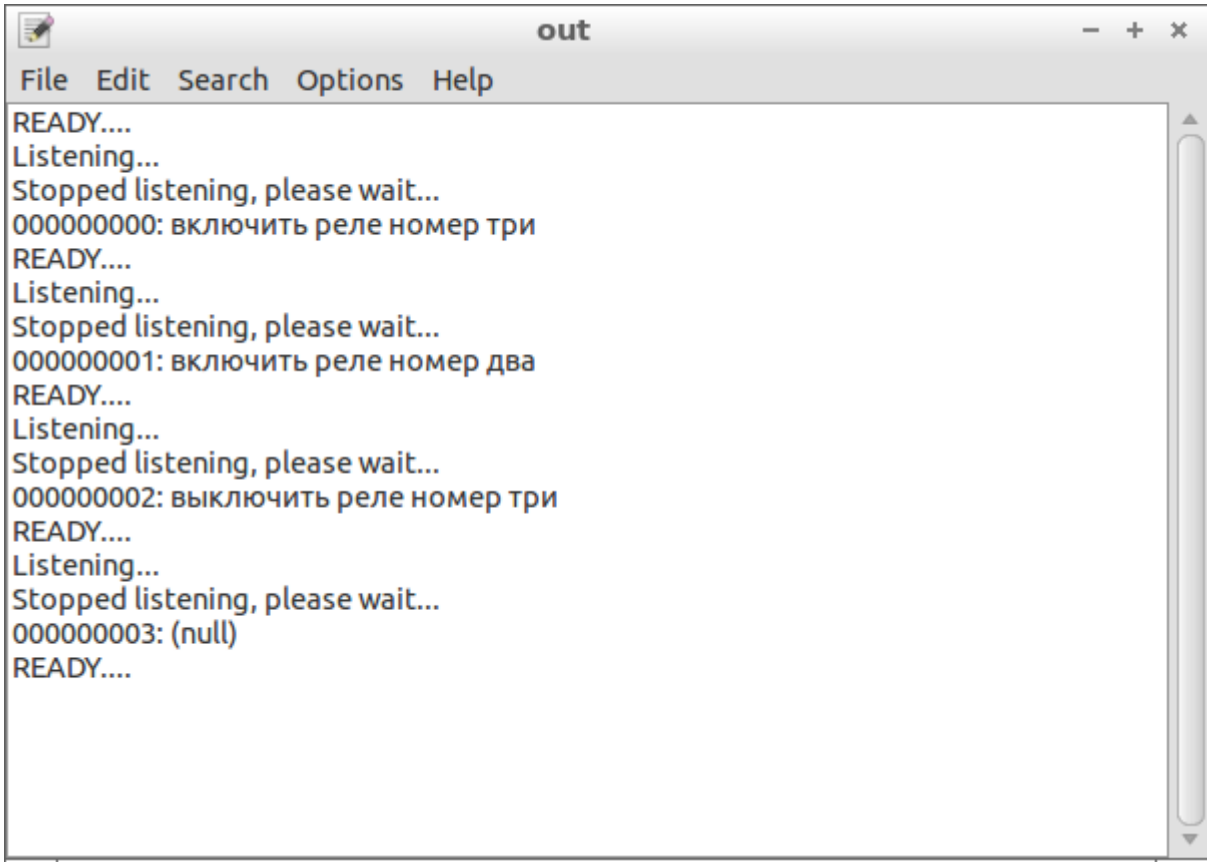


Рисунок 3.6 – Содержимое файла out

Из рисунка можно понять, что каждая речевая команда имеет девятизначный номер, после которого следует сама распознанная команда. Для выделения и обработки команд была написана программа на языке С. Исходный текст программы указан в листинге 3.1.

Листинг 3.1 – Программа обработки результатов распознавания

```
#include <stdio.h>
#include <stdlib.h>
#include <iostream>
#include <fstream>

using namespace std;

char buffer[100];
char str1[]="включить реле номер один";
```

```
char str2[]="включить реле номер два";
char str3[]="включить реле номер три";
char str4[]="выключить реле номер один";
char str5[]="выключить реле номер два";
char str6[]="выключить реле номер три";
int i,j,prev_n,b,lol,maximum,command;
int match[6];
double n;

void main(void)  {
  std::ifstream sphinx_out("/home/ubuntu/sphinx_pg/out");

  while(1)
   {
  while(! sphinx_out.eof())
  {
  sphinx_out.getline(buffer,sizeof(buffer));
  if(buffer[0]=='0')
 {
     b=0;
     for(i=0;i<6;i++) match[i]=0;
     n++;
     for(i=0;i<=100;i++) {
          if(buffer[i]==' ' && lol==0) lol =i+1;
          printf("%c",buffer[i]);
          }
     for(i=0;i<=100;i++)
          {
          if(str1[i]==buffer[lol+i]) match[0]++;
          if(str2[i]==buffer[lol+i]) match[1]++;
          if(str3[i]==buffer[lol+i]) match[2]++;
          if(str4[i]==buffer[lol+i]) match[3]++;
          if(str5[i]==buffer[lol+i]) match[4]++;
          if(str6[i]==buffer[lol+i]) match[5]++;
          }

maximum = match[0];
command=0;

  for (i = 1; i <=5; i++)
  {
    if (match[i] > maximum)
    {
       maximum  = match[i];
       command = i;
    }
  }

if(match[command]<40 || match[command]==50) command=6;

switch(command)
{
```

```
     case 0:
          system("curl                --request                GET
'http://192.168.0.112/gpio?st=1&pin=5'");
          break;
     case 1:
          system("curl                --request                GET
'http://192.168.0.112/gpio?st=1&pin=4'");
          break;
     case 2:
          system("curl                --request                GET
'http://192.168.0.112/gpio?st=1&pin=2'");
          break;
     case 3:
          system("curl                --request                GET
'http://192.168.0.112/gpio?st=0&pin=5'");
          break;
     case 4:
          system("curl                --request                GET
'http://192.168.0.112/gpio?st=0&pin=4'");
          break;
     case 5:
          system("curl                --request                GET
'http://192.168.0.112/gpio?st=0&pin=2'");
          break;

}
     printf(" %d",command);
     printf("\n");

  }
  for(i=0; i<100; i++) buffer[i]='\0';
  }
  sphinx_out.clear();
  sphinx_out.seekg(0,std::ios::beg);
  for(i=1;i<=(n*4);i++)
sphinx_out.getline(buffer,sizeof(buffer));
  }
 }
```

Программа выделяет речевые команды из текстового файла, и сравнивает их с одной из шести возможных, при их совпадении, посылается GET-запрос на ESP8266, который в свою очередь управляет модулями реле. Блок-схема программы приведена на рисунке 3.7.

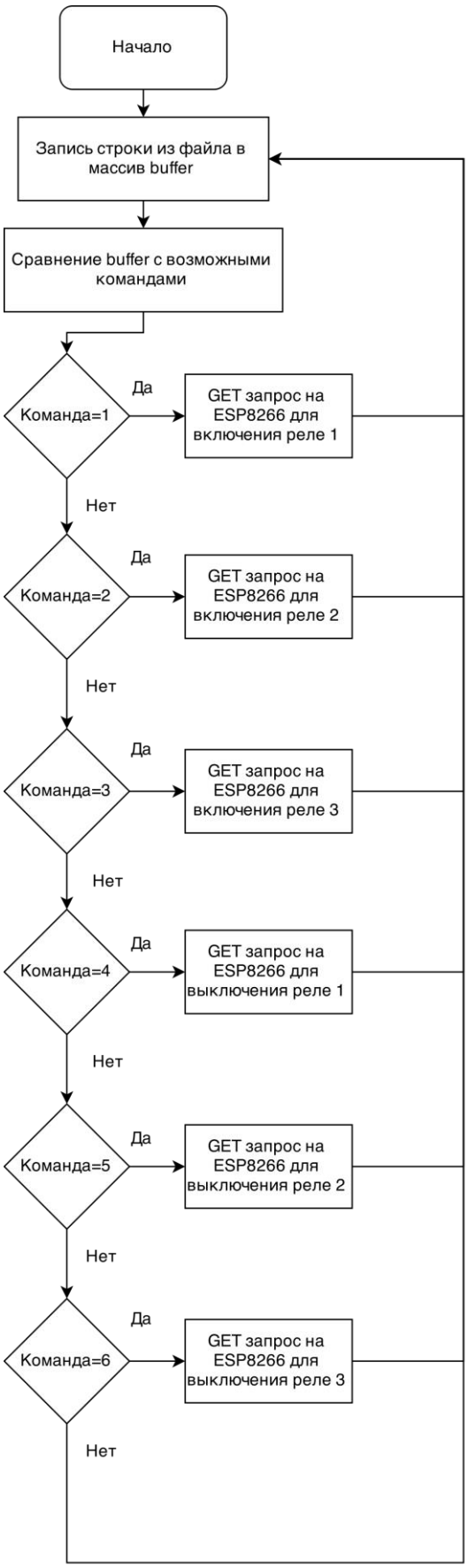
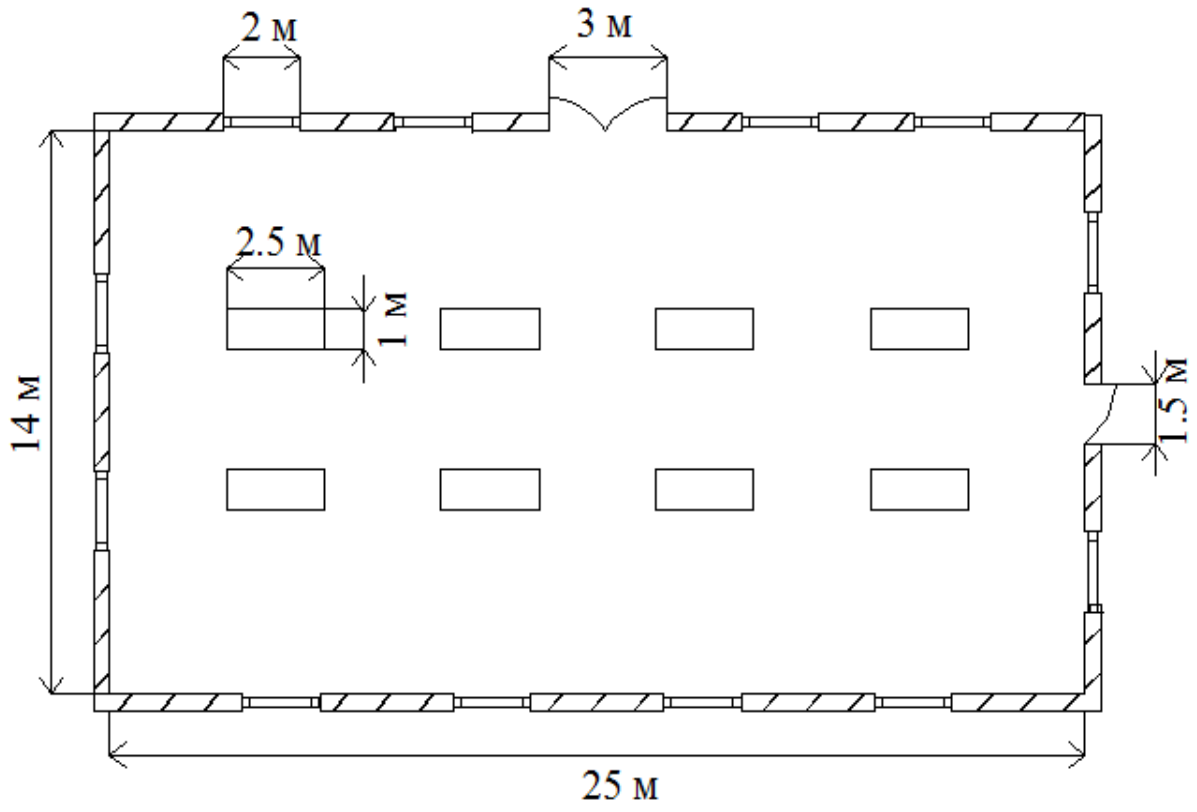


Рисунок 3.7 – Блок-схема программы обработки результата распознавания речи

4 Безопасность жизнедеятельности

4.1 Анализ условий труда, выявление опасных и вредных производственных факторов

Сделаем анализ помещений и условий труда в нем. Сначала приведем схему цеха (рисунок 4.1).



	окно
	дверь
	рабочее место
	стена

Рисунок 4.1 – План монтажно-сборочного цеха

Для более полного и подробного анализа условий труда необходимо учитывать следующие факторы:

- количество людей в производственном цеху – 12 человек, из них 8 мужчин и 4 женщин;
- площадь цеха – 25х14х 4 м;
- размер окна – 2х1,5 м;
- размер дверей – 3 и 1,5 м;
- размер рабочего места 2,5х1 м.

Источники опасности:

- опасность ранения – части рабочего места с острыми краями, режущие кромки и режущие инструменты;
- электрическая опасность – поражение электрическим током может привести к тяжелейшим травмам;
- при пайке и лакировке, выделяются опасные вредные вещества. Над рабочими местами устанавливаем фанкойлы тип, которых мы выберем потом после расчета вентиляции. Наиболее опасными для здоровья являются тяжелые пары свинца, канифоли, применяемых при сборке оборудования.

Во избежание и предотвращение производственных травм, каждый работник проконсультирован по технике безопасности. Цех оснащен электрическим автоматом защиты, который срабатывает при коротком замыкании и отключает все электричество в цехе.

При производстве оборудования применяются следующие технологические операции:

- комплектование;
- установка элементов на плату (сборка);
- пайка и лакирование;
- тестирование;
- маркирование.

Все эти операции выполняются в сборочно-монтажном цехе, со следующими санитарными нормами:

- площадь на одного рабочего не менее 4,5 м2;
- объем помещения на одного рабочего не менее 15 м3;
- в цехе применяется искусственное освещение;
- категория тяжести работ в цехе II (ГОСТ 12.1.005-88), так как работы производятся сидя, стоя и сопровождаются физическими напряжениями (энергозатраты до 120 ккал/час).

В помещении в течение года метеоусловия находятся в следующих пределах:

- температура воздуха от 20 до 26 °С;
- относительная влажность от 40 до 60 %;
- подвижность воздуха не более 0,2 м/с;
- расстояние между рабочими местами не менее 3 м.

Расположение рабочих мест приведено на рисунке 4.1. При сборке используется паяльные станции Quick-702ESD.

Т а б л и ц а 4.1 – Технические характеристики паяльной станции Quick-702ESD

Характеристика	Значение
Длина, мм	335
Ширина, мм	253
Высота, мм	160
Коэффициент перехода в помещение тепла от двигателя - η	0,8
Напряжение питания, В	220
Потребляемая мощность, Вт	520
Температурный режим пайки, °С	200…480
Температурный режим распайки, °С	150…500
Температурный диапазон оловоотсоса, °С	320…480
Стабильность температуры, °С	±1
Напряжение паяльника, В	24
Напряжение оловоотсоса, В	36
Вес, кг	13,3

Электропитание осуществляется от трехфазной сети переменного тока с напряжением питания 220В/380В и частотой f=50Гц.

Рабочие работаю в цеху с 9.00 до 18.00 в одну смену. Так как работают с деталями малого размером, требуется средняя точность 0,5–1 мм, что соответствует III б разряду зрительной работы. А также для непрерывной работы во время пасмурной погоды и зимний период года (так как дни короткие) в цеху должно быть искусственное освещение.

При сборке оборудование, рабочие паяют, вследствие выделяются тяжелые пары олова и канифоли, а также для поддержания микроклимата. В цеху должна располагаться вентиляционная система.

4.2 Расчет искусственного освещения в цехе

Задачей расчета является определение потребной мощности электрической осветительной установки для создания в производственном помещении заданной освещенности.

Для расчета общего равномерного освещения при горизонтальной рабочей поверхности основным является метод светового потока (коэффициента использования), учитывающий световой поток, отраженный от потолка и стен. Световой поток лампы $\Phi_л$ (лм) при лампах накаливания рассчитывается по формуле (4.1):

$$\Phi_л = \frac{E_H \cdot S \cdot z \cdot k}{N \cdot n \cdot \eta}, \qquad (4.1)$$

где E_H – нормированная минимальная освещенность, 300 лк (для разряда зрительной работы III б для искусственного освещения , при системе общего освещения);

S – площадь освещаемого помещения, 350 $м^2$ (25x14);

z – коэффициент минимальной освещенности, равный отношению $E_{ср}/E_{мин}$, значение которого для ламп накаливания – 1,15;

k – коэффициент запаса 1,5 (для сборочных и сборочно-монтажных цехов при газоразрядных лампах);

N – число светильников в помещении;

n – число ламп в светильнике

η – коэффициент использования светового потока ламп, зависящий от КПД и кривой распределения света светильника, коэффициентот отражения потолка $\rho_{пот}$, пола $\rho_{пол}$ и стен $\rho_с$, высоты подвески светильников и показателя помещения i.

Показатель помещения i определяется по формуле (4.2):

$$i = \frac{A \cdot B}{h \cdot (A + B)}, \qquad (4.2)$$

где A – длина помещения, 25 м;

B – ширина помещения, 14 м;

h – высота светильника над рабочей поверхностью, м.

Рассчитываем высоту подвеса над рабочей поверхностью по формуле (4.3).

$$h = H - h_c - h_p, \qquad (4.3)$$

где h_c – расстояние от светильника до перекрытия, 1 м;

h_p – расстояние рабочей поверхности над полом, 1 м;

H – высота помещения, 4 м.

$h = 4 - 1 - 1 = 2$ м.

i = 25·14/2(25+14)=4,487.

По ГОСТ 2239-79 подбираем ближайшую стандартную лампу и определяем электрическую мощность всей осветительной системы. В практике допускается отклонение светового потока выбранной лампы от расчетного до минус 10% и плюс 20%, в противном случае выбирают другую схему расположения светильников.

Отражения от потолка $\rho_{ПОТ}$, пола $\rho_{ПОЛ}$ и стен $\rho_{С}$, примем минимальное значение 30, 10 и 10 соответственно. Выберем светильник ЛСП01 2х150, для которого значение η при i = 4,487 (округляем до 4,5) равняется 54.

Т а б л и ц а 4.2 – Технические характеристики светильник ЛСП01 2х150

Тип светильника	Количество ламп	Мощность, Вт	Длина, мм	Ширина, мм	Высота, мм	Вес, кг
ЛСП01	2	300	1536	674	184	15,5

В каждом светильнике ЛСП01 2х150 находится по 2 лампы типа ЛХБ150

Т а б л и ц а 4.3 – Технические характеристики лампы типа ЛХБ150

Тип лампы	Мощность, Вт	Напряжение на лампе, В	Ток лампы, А	Световой поток, лм	Длина лампы, мм	Диаметр лампы, мм
ЛХБ 150	150	90	1,90	8000	1500	40

Рассчитаем из формулы (4.1) число необходимых для освещения светильников (формула 4.4):

$$N = \frac{E_{Н} \cdot S \cdot z \cdot k}{\Phi_{л} \cdot n \cdot \eta}, \qquad (4.4)$$

$$N = \frac{300 \cdot 330 \cdot 1{,}15 \cdot 1{,}5}{800 \cdot 2 \cdot 0{,}54} = 20{,}94 \approx 21 \ шт.$$

Возьмем 21 шт и проверим допустимость наших действий по формуле (4.1). В этом случае фактическая освещенность составит:

$$N = \frac{800 \cdot 2 \cdot 0{,}54 \cdot 21}{350 \cdot 1{,}15 \cdot 1{,}5} = 300{,}52 \ лк.$$

Отклонение светового потока ламп от допустимого значения не превышает 10%:

$$\frac{300{,}52 - 300}{300{,}52} = 0{,}17\%.$$

Сделаем расчет расстояния между светильниками и схему их расположения.

$$L_{AB} = \lambda_{AB} \cdot h, \qquad (4.5)$$

где $\lambda_A = 1{,}6;\ \lambda_B = 1{,}8.$

$L_A = \lambda_A \cdot h = 1{,}6 \cdot 2 = 3{,}2\ м;$

$L_B = \lambda_B \cdot h = 1{,}8 \cdot 2 = 3{,}6\ м.$

Расстояние от крайних светильников до стены:

$$l_A = \frac{L - 6 \cdot L_A}{2} = \frac{25 - 6 \cdot 3{,}2}{2} = 2{,}9\ м;$$

$$l_B = \frac{L - 2 \cdot L_B}{2} = \frac{14 - 2 \cdot 3{,}6}{2} = 3{,}5\ м.$$

Расположение светильников приведено на рисунке 4.2.

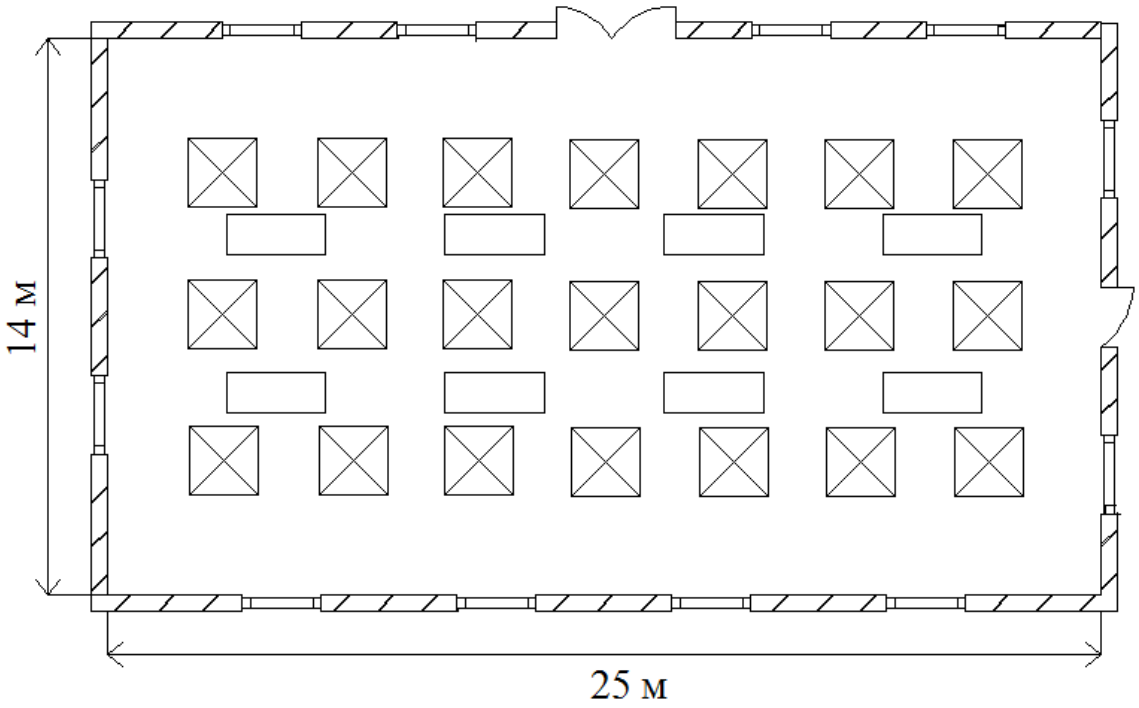


Рисунок 4.2 – Схема расположения осветительных ламп

4.3 Расчет системы вентиляции цеха

Теплопоступления от оборудования (формула (4.6)):

$$Q_{обр} = n \cdot P_{обр} \cdot \eta, \qquad (4.6)$$

где n – количество оборудования, шт;

$P_{обр}$ – мощность, потребляемая оборудованием, кВт/ч;

η=0,75 – 0,95 коэффициент перехода в помещение тепла.

$$Q_{обр} = n \cdot P_{обр} \cdot \eta = 12 \cdot 520 \cdot 0{,}8 = 4992 \ \ Вт.$$

Теплопоступления от осветительных приборов (формула (4.7)):

$$Q_{осв} = \eta \cdot N_{осв} \cdot F_{пол}, \qquad (4.7)$$

где η – от 0,5 до 0,6 для люминесцентных ламп,
$N_{осв}$ – мощность лампы. Для хорошо освященных помещений $N_{осв}$ от 50 до 100 Вт/м^2.
$F_{пол}$ – площадь пола.

$$F_{пол} = A \cdot B = 25 \cdot 14 = 350 \ \ м^2;$$

$$Q_{осв} = 0{,}6 \cdot 100 \cdot 350 = 21000 \ \ Вт.$$

Теплопоступление от людей:

$$Q_л^я = \eta \cdot Q_ч^я, \qquad (4.8)$$

где $Q_ч^я$ – явное число от человека, принято считать что женщина выделяет 85% тепла от нормы тепловыделений взрослого мужчины, летом при 26 ^{0}С один мужчина выделяет явного тепла $Q_ч^я$=61 Вт, зимой при 20 ^{0}С один мужчина выделяет явного тепла $Q_ч^я$=82 Вт;
n – число людей.

Теплопоступление летом от мужчин и женщин:

$$Q_м^я = 8 \cdot 61 = 488 \ \ Вт;$$

$$Q_ж^я = 4 \cdot 61 \cdot 0{,}85 = 207{,}4 \ \ Вт;$$

$$Q_л^я = Q_м^я + Q_ж^я = 488 + 207{,}4 = 695{,}4 \ \ Вт.$$

Теплопоступление зимой от мужчин и женщин:

$$Q_м^я = 8 \cdot 82 = 656 \ \ Вт;$$

$$Q_ж^я = 4 \cdot 82 \cdot 0{,}85 = 278{,}8 \ \ Вт;$$

$$Q_л^я = Q_м^я + Q_ж^я = 656 + 278{,}8 = 934{,}8 \ \ Вт.$$

Теплопоступления и теплопотери в результате разности температур

Теплопоступления и теплопотери в результате разности температур определяется по формуле (4.9):

$$Q_{отд} = V_{пом} \cdot X_0 \cdot (t_{вн} - t_н), \qquad (4.9)$$

где $V_{пом}$ – объем помещения;

X_0– удельная тепловая характеристика, $X_0 = 0{,}42\ \frac{Вт}{м^3 \cdot °C}$;

$t_в$ – температура внутри помещения:

летом $t_в$= 28°C ;

зимой $t_в$ = 20°C;

$t_н$ – температура снаружи помещения:

летом $t_н$ = 28°C;

зимой $t_н$ = –25 °C.

$$V_{пом} = A \cdot B \cdot H = 25 \cdot 14 \cdot 4 = 1400\ м^3.$$

Для теплого периода года:

$$Q_{отд} = 1400 \cdot 0{,}42 \cdot 0 = 0\ Вт.$$

Для холодного периода года:

$$Q_{отд} = 1400 \cdot 0{,}42 \cdot (20 - (-25)) = 26460\ Вт.$$

Общее тепло выделяемое внутри помещения определим по формуле (4.10):

$$Q_{изб} = Q_{обр} + Q_{осв} + Q_л^я + Q_{отд}. \qquad (4.10)$$

Общее тепло выделяемое летом:

$$Q_{изб} = 4992 + 21000 + 695{,}4 + 0 = 26687{,}4\ Вт.$$

Общее тепло выделяемое зимой:

$$Q_{изб} = 4992 + 21000 + 934{,}8 - 26460 = 466{,}8\ Вт.$$

Количество тепла, необходимого для удаления (или поступления) воздуха в помещение

$$L = \frac{Q_{изб} \cdot 860}{C \cdot \Delta t \cdot \gamma}, \qquad (4.11)$$

где C –теплоемкость воздуха, C=0,24 ккал/кг °C;

Δt – выбираем в зависимости от Q_n. ГОСТ 12.1.014-84 «Воздух рабочей зоны». $\Delta t = 6°C$ при $Q_n \leq 20$ ккал/м^3 и

$\Delta t = 8°C$ при $Q_n > 20$ ккал/м^3;
γ – удельная масса приточного воздуха, $\gamma = 1{,}206$ кг/м^3.

$$Q_n = \frac{Q_{изб} \cdot 860}{V_{пом}} . \qquad (4.12)$$

$$Q_n = \frac{26687{,}4 \cdot 860}{1400} = 16{,}393 \ ккал/м^3,$$

тогда $\Delta t = 6 \ °C$,

$$L = \frac{26687{,}4 \cdot 860}{0{,}24 \cdot 10^3 \cdot 6 \cdot 1{,}206} = 13250 \ м^3/ч.$$

4.3.6 Кратность воздухообмена

$$n = \frac{L}{V_{пом}}, \qquad (4.13)$$

$$n = \frac{13250}{1400} = 9.5 \ 1/ч.$$

То есть необходимый расход воздухообмена 13681,638 м3/ч. Выберем нужное оборудование: фанкойлы 42NFS65C с шумоизоляцией 8 шт. технические характеристики указаны в таблице 4.4, чиллер WRAT101 технические характеристики указаны в таблице 4.5, и насосная станция GP2-12.

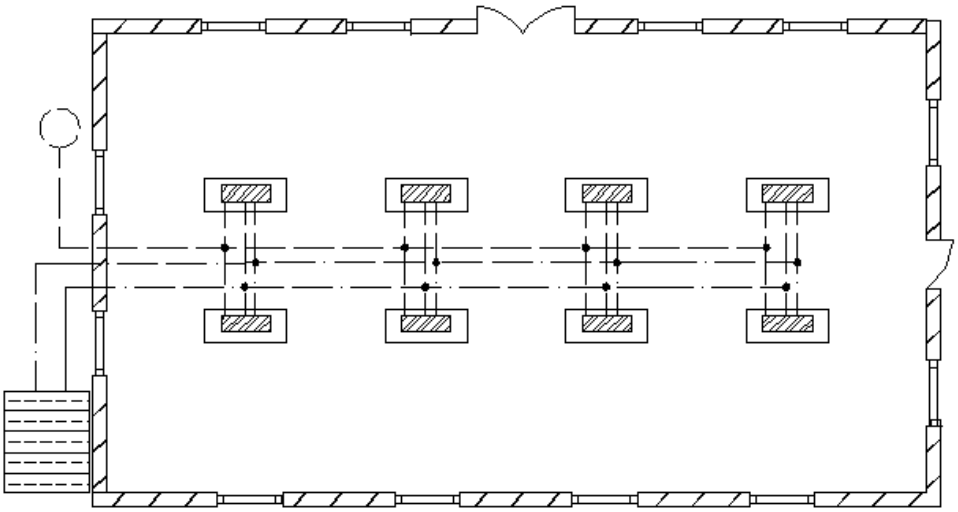
Т а б л и ц а 4.4 – Технические характеристики фанкойла 42NFS65C

Наименование	Производитель	Мощность по холоду, кВт	Мощность по теплу, кВт	Расход воздуха, м^3/с
42NFS65C	CARRIER	3,5	4,1	1720

Т а б л и ц а 4.5 – Технические характеристики чиллера WRAT101

Наименование	Мощность по холоду, кВт	Мощность по теплу, кВт
WRAT101	28,3	31,1

Приведем схему вентиляции (рисунке 4.3).



	окно
	дверь
	рабочее место
	стена
	фанкойл
	чиллер и насосная станция
	воздуховоды
	дренажный трубопровод
	канализация

Рисунок 4.3 – Схема вентиляции

5 Технико-экономическое обоснование проекта

5.1 Введение

5.1.1 Обоснование необходимости разработки

В рамках данного дипломного проекта был разработан опытный образец системы голосового управления, которая может быть использована в различных сферах: автоматизация жилых и производственных помещений, автоматизация лифтов, бытовых приборов, т.е. везде, где присутствует человеко-машинный интерфейс. В Казахстане мало внимания уделено теме голосового управления, поэтому причине целесообразно двигаться в данном направлении. На отечественном рынке средств автоматизации нет аналогичных продуктов, и спрос на такие устройства будет достаточно высок.

5.1.2 Цель и задачи

Цель данной главы – расчет экономической эффективности коммерческого производства систем голосового управления. Для достижения цели, необходимо выполнение следующих задач:

- расчет инвестиционных затрат;
- расчет себестоимости товарной продукции;
- расчет доходов от реализации.

5.1.3 Анализ рынка сбыта

Наша продукция может заинтересовать как физических, так и юридических лиц. Среди физических лиц особый спрос на такую продукцию могут проявить люди с ограниченными физическими возможностями, т.к. управление освещением, отоплением и другими устройствами и процессами в доме вызывает сложность. Также продукция заинтересует и обычных людей, интересующихся новинками техники. Среди юридических лиц система голосового управления может заинтересовать строительные компании, для установки их в жилые дома и коттеджи, компании, занимающиеся автоматизацией предприятий.

5.2 Финансовый план

5.2.1 Расчет инвестиционных затрат

Рассчитаем объем капитальных вложений, необходимых для организации производства. Перечень необходимого оборудования, количество и общая стоимость представлены таблице 5.1.

Т а б л и ц а 5.1 – Стоимость оборудования необходимого и программного обеспечения для реализации проекта по производству систем голосового управления

Наименование	Количество, шт.	Сумма, тенге
Фрезерный станок High-Z S-1400T	2	1000202
Пакет программ WIN PCNC Professional (CD)+ карта I/O	1	67246
Стол с Т-образными пазами для S-1400T	2	87604
Монтажный стол	2	34188
Вакуумный стол 400x1000mm	2	42240
Вакуумный насос для вакуумного стола 0,70 KW	2	57350
Персональный компьютер	2	184000
Итого		1472830

Общие капиталовложения по созданию сети включают в себя капиталовложения на оборудование и программное обеспечение, капиталовложения на монтаж и капиталовложения на транспортные расходы, и определяются по формуле (5.1).

$$\sum K = K_O + K_M + K_T, \qquad (5.1)$$

где $\sum K$ – общие капитальные вложения;

K_O – капитальные вложения на оборудования;

K_M – капитальные вложения на монтажные работы (20% от стоимости оборудования);

K_{TP} – капитальные вложения на транспортные расходы (5% от стоимости оборудования).

K_M = 1472830 · 0,2 = 294566 тенге.

K_{TP} = 1472830 · 0,05= 73642 тенге.

$\sum K = 1472830 + 294566 + 73642 = 1841038$ тенге.

5.2.2 Расчет себестоимости товарной продукции

В состав себестоимости товарной продукции входят следующие статьи затрат:

- затраты на материалы и комплектующие;
- заработная плата работников;
- социальный налог;
- электроэнергия для производственных нужд;
- амортизационные отчисления;
- накладные расходы.

Себестоимость товарной продукции определяются по формуле(5.2):

$$C = C_{мат.} + ФОТ + C_C + A + Э + H, \qquad (5.2)$$

где $C_{мат.}$ – затраты на материалы и комплектующие;
ФОТ – фонд оплаты труда;
C_C – отчисления на социальные нужды;
А – амортизационные отчисления;
Э – электроэнергия на производственные нужды;
Н – накладные расходы.

Перечень и стоимость материалов и комплектующих, необходимых для производства продукции, приведены в таблице 5.2.

Т а б л и ц а 5.2 - Затраты на материалы и комплектующие

Наименование	Количество, шт.	Сумма, тенге
Стеклотекстолит фольгированный 200х300мм	20	23800
Паяльная паста	10	15600
Пластик АБС	20	18950
Микросхемы Allwinner A10	50	100000
Итого		158350

Первоначальная численность сотрудников для организации производства прибора представлена в таблице 5.3. Приведем среднемесячные оклады персонала.

Т а б л и ц а 5.3 – Заработные платы сотрудников компании

Исполнители	Количество, человек	Заработная плата за месяц, тенге
Руководитель	1	150000
Программист	2	120000
Инженер	2	80000
Итого	5	450000

Фонд оплаты труда определяется по формуле (5.3):

$$ФОТ = З_{осн} + З_{доп}, \quad (5.3)$$

где $З_{осн}$ – основная заработная плата;

$З_{доп}$ – дополнительная заработная плата.

Основная заработная плата за год составит:

$З_{осн} = 450000 \cdot 12 = 5400000$ тенге.

Дополнительная заработная плата составляет 10% от основной заработной платы:

$З_{доп} = З_{осн} \cdot 0{,}1 = 5400000 \cdot 0{,}1 = 540000$ тенге.

Тогда ФОТ составит:

$ФОТ = 5400000 + 540000 = 5940000$ тенге.

Отчисления на социальные налоги напрямую зависят от фонда оплаты труда. На сегодня действующие нормы отчислений характеризуются следующими данными: социальный налог составляет 13% от ФОТ согласно налоговому кодексу РК, за вычетом отчислений в пенсионный фонд в размере 10% от фонда оплаты труда согласно закону «О пенсионном обеспечении в Республике Казахстан».

Отчисления на социальные нужды определяются по формуле(5.4):

$$С_С = 0{,}13 \cdot (ФОТ - (0{,}1 \cdot ФОТ)). \quad (5.4)$$

$С_С = 0{,}13 \cdot (5940000 - (0{,}1 \cdot 5940000)) = 694980$ тенге.

Сумма амортизационных отчислений начисляется по единым нормам, которые устанавливаются в процентах от стоимости основных фондов. Норма амортизации, согласно Кодексу РК «О налогах и других обязательных платежах в бюджет» за 15 января 2015 года, на машины и оборудование составляет 25%, на компьютеры, программное обеспечение и оборудование для обработки информации 40%. Амортизационные отчисления для оборудования и ПО рассчитываются по формуле (5.5) [6]:

$$A = \frac{Н_{амор т} \cdot С_{пер}}{100*12} \cdot div\left[\frac{N}{30}\right] \cdot K, \qquad (5.5)$$

где N – количество дней эксплуатации;

$Н_{АМОРТ}$ – норма амортизации;

$С_{ПЕР}$ – первоначальная стоимость оборудования, $С_{ПЕР} = С_{ТЕХ}$;

К – количество единиц оборудования.

$$A_{Комп.иПО} = \frac{40 \cdot 251246}{100 \cdot 12} \cdot \left(\frac{365}{30}\right) \cdot 1 = 101894 \text{ тенге.}$$

$$A_{ОБОР} = \frac{25 \cdot 1221584}{100 \cdot 12} \cdot \left(\frac{365}{30}\right) \cdot 1 = 309637 \text{ тенге.}$$

$А = А_{Комп. и ПО} + А_{ОБОР}$= 101894 + 309637 = 411531 тенге.

Затраты на электроэнергию рассчитываются по формуле (5.6):

$$Э = W \cdot T \cdot S, \qquad (5.6)$$

где *W* - потребляемая мощность;

T - количество часов работы оборудования (8 часов, 245 рабочих дней);

S - стоимость киловатт-часа электроэнергии.

W = 1,2 · 2 + 1 · 2 + 0,8 · 2 = 6 кВт.

Т = 8·245 = 1960 ч.

S = 15,4 тенге/кВт · час.

Таким образом

$Э = 6 \cdot 1960 \cdot 15{,}4 = 181104$ тенге.

Накладные расходы составляют 10% от общей суммы и определяются по формуле (5.7):

$$H = 0{,}10 \times C. \qquad (5.7)$$

Определим накладные расходы:

$C = 158350 + 5940000 + 694980 + 411531 + 181104 = 7385965$ тенге.

$H = 0{,}1 \cdot 7385965 = 738567$ тенге.

Рассчитаем себестоимость товарной продукции по формуле (5.1):

$C = 158350 + 5940000 + 694980 + 411531 + 181104 + 738567 = 8124562$ тенге.

Рассмотрим структуру себестоимости товарной продукции, которая представлена в таблице 5.4.

Т а б л и ц а 5.4 – Структура себестоимости товарной продукции, тенге

Наименование показателей	План	Уд. вес, %
Материалы и комплектующие	158350	1,9
ФОТ	5940000	73,1
Отчисления на социальные нужды	694980	8,6
Амортизационные отчисления	411531	5,1
Затраты на электроэнергию	181104	2,2
Накладные расходы	738567	9,1
Полная себестоимость	8124562	100
Выпущено продукции, единиц	50	
Себестоимость единицы продукции	162491	

5.2.3 Расчет доходов от реализации систем управления голосом

Доход от реализации продукта (5.8):

$$Д = N_P \cdot Ц, \qquad (5.8)$$

где N_P – число реализованных систем голосового управления (50 единиц);

$Т_П$ – стоимость продукта без НДС, т.е. 83,3% от рыночной стоимости.

Рыночная стоимость продукции рассчитывается по формуле (5.9):

$$Ц = С + П + НДС, \qquad (5.9)$$

где С – себестоимость единицы продукции;

П – прибыль (15% от себестоимости);

НДС – налог на добавленную стоимость (12% от С+П).

Рассчитаем рыночную стоимость продукции:

$Т_П = 162491 + 162491 \cdot 0,15 = 186864$ тг;

НДС = $186864 \cdot 0,12 = 22424$ тг;

Ц = 186864 + 22424 = 209288 тг.

Таким образом, доход с продажи 50 единиц продукции составит:

Д = 50 · 186864 = 9343200 тенге.

По плану количество произведенной и реализованной продукции составит 50 единиц.

5.2.4 Расчет экономической эффективности

В результате расчётов, проведённых в разделе 5.2, была получена коммерческая стоимость реализации проекта, равная 1472830 тенге.

Годовой экономический эффект будет рассчитываться по формуле (5.10):

$$Ea = (Д – С)/К = П/К, \quad (5.10)$$

где С – себестоимость продукции;
Д – доходы от основной деятельности;
П – чистый доход;
К – разовые капитальные вложения.

Нормативный коэффициент экономической эффективности капитальных вложений принимается равным = 0,2. Оборудование основных фондов внедряется новым, поэтому капитальные затраты будут равны коммерческой стоимости производства систем управления голосом.

Д–С = 9343200 – 8124562 = 1218638 тенге.

Годовой экономический эффект по формуле (5.10) составит:

E_a = (9343200 – 8124562)/ 1472830 = 0,82.

Срок окупаемости рассчитываем по формуле (5.11):

$$T = 1/ E_a. \quad (5.11)$$

Т = 1/0,82 = 1,2 года.

На основе проведенных выше вычислений можно сказать о том, что внедрение разработанной системы позволит получить значительный экономический эффект. Годовой экономический эффект составил 1218638 тенге, при сроке окупаемости системы 1,2 года, что говорит о целесообразности производства данной системы.

Коэффициент дисконтирования рассчитывается по формуле (5.12)

$$a_t = \frac{1}{(1+E)^t}, \quad (5.12)$$

где a_t – коэффициент дисконтирования;
E – норма дисконта;
t – номер шага расчета (года).

Рассчитаем коэффициент дисконтирования на три года:

$$a_1 = \frac{1}{(1+0,2)^1} = 0,83;$$

$$a_2 = \frac{1}{(1+0{,}2)^2} = 0{,}69;$$

$$a_3 = \frac{1}{(1+0{,}2)^3} = 0{,}58.$$

PV – современная стоимость денежного потока на протяжении экономической жизни проекта (5.13)

$$PV = \sum_{t=1}^{n} P_t/(1+E)^t, \qquad (5.13)$$

где E – норма дисконта;

n– число периодов реализации проекта;

P_t – чистый поток платежей в периоде t.

Чистая текущая стоимость проекта (NPV) рассчитывается по формуле (5.14).

$$NPV = \sum_{t=1}^{n} \frac{P_t}{(1+E)^t} - I_0, \qquad (5.14)$$

где I_0 – сумма первоначальных затрат, сумма инвестиций на начало проекта.

PI – индекс рентабельности инвестиций; рассчитывается по формуле (5.15).

$$PI = \sum_{t=1}^{n} \frac{P_t}{(1+E)^t} / I_0 \,. \qquad (5.15)$$

Если PI > 1, то проект следует принять.

DPP – Дисконтированный срок окупаемости проекта (5.16)

$$\mathrm{DPP} = K/P_t, \qquad (5.16)$$

где К– капиталовложение;

P_t – чистый поток платежей в периоде t.

С помощью формул (5.12) - (5.15) рассчитаем показатели предприятия на три года, и запишем их в таблицу 5.5.

Т а б л и ц а 5.5 – Показатели предприятия

Показатели	1 г.	2 г.	3 г.
Инвестиционные затраты, млн. тг	1,473		
Чистый денежный поток, млн. тг	1,219	1,219	1,219
Норма дисконта, %	20	20	20
Коэффициент дисконтирования	0,833	0,694	0,579
Чистая текущая стоимость (PV), млн. тг	1,016	0,846	0,705
Чистый дисконтированный доход (NPV), млн. тг	1,094		
Индекс доходности (PI)	1,743		
Чистая текущая стоимость с нарастающим итогом, млн. тг	-0,457	0,389	1,094

5.3 Вывод

Т а б л и ц а 5.6 – Сводная таблица экономической эффективности проекта

№	Показатели	
1	Капитальные вложения	1478300 тенге
2	Доходы	9343200 тенге
3	Себестоимость продукции	8124562 тенге
4	Прибыль	1218638 тенге
5	Коэффициент экон. эффективности	0,82
6	Срок окупаемости	1,2 года
7	Срок окупаемости с учетом дисконтирования	1,5 года

В данной главе были проделаны расчеты для определения экономической эффективности проекта. Согласно расчетам, для реализации проекта потребуется 1,478 млн. тг. разовых капиталовложений. Себестоимость единицы продукции составит 162491. Доход от реализации 50 единиц продукции, выпущенных в течение года, составит 9,343 млн. тг.. Согласно плану, проект окупится через 1 год и 3 месяца. Сводные данные приведены в таблице 5.6. Исходя из результатов расчета, можно считать проект прибыльным и эффективным.

Заключение

В данном дипломном проекте были описаны основные понятия речи, распознавания речи, устройств аудиозахвата.

Были рассмотрены различные программы и инструменты для разработки систем распознавания речи. Были описаны одноплатный ARM-миникомпьютер, его операционная система, работа и программирование на миникомпьютерах.

В качестве "движка" системы голосового управления был выбран программный пакет CMU Sphinx, разработанный в университете Carnegie Mellon. Данный программный пакет имеет широкие возможности настройки и оптимизирован под работу на устройствах, основанных на ARM-архитектуре и работающих на nix-системах.

CMU Sphinx была установлена и настроена на одноплатном ARM-миникомпьютере pcDuino 1. Управление исполнительным механизмами реализовано через реле, которые включается Wi-Fi платой ESP8266 посредством http-GET запросов.

Список литературы

1. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. – Киев: Наук. думка, 1987. – 264 с.

2. Фролов А., Фролов Г. Синтез и распознавание речи. Современные решения [Электронный ресурс]. – Электрон. журн. – 2003. – Режим доступа: http://www.frolov-lib.ru

3. CMU Sphinx. Open Basic concepts of speech. [Электронный ресурс]. – Режим доступа: http://cmusphinx.sourceforge.net/wiki/tutorialconcepts

4. Аязбай А.Е., Конуркульжин Д.А., Орынбай А.А. Система голосового управления на одноплатном ARM-миникомпьютере // Высшая школа Казахстана. – 2004. – № 4. – С. 186-189.

5. Агуров П.В. Последовательные интерфейсы ПК. Практика программирования. – СПб.: БХВ-Петербург, 2004. – 496 с.

6. Базылов К.Б., Алибаева С.А., Бабич А.А. Выпускная работа бакалавров. Экономический раздел. – Алматы: АИЭС, 2008. – 20 с.

7. Блэк Ю. Сети ЭВМ: протоколы, стандарты, интерфейсы. – М: Мир, 1990. – 465 с.

8.Агуров П.В. Интерфейсы USB. Практика использования и программирования. – СПб.: БХВ-Петербург, 2004. – 576 с.

9. Воротников С.А. Информационные устройства робототехнических систем. – М.: Изд – во МГТУ им. Н.Э. Баумана, 2005. – 384 с.

10. Евстифеев А.В. Микроконтроллеры семейства Classic фирмы ATMEL. – М.: Додэка – XXI, 2006. – 288 с.

11. Комягин В. Ubuntu Linux 10.04. Русская версия. - М.: Триумф, 2011. – 208 с.

12. Князевский Б.А. Охрана труда. – М.: Высшая школа, 2002. – 365 с.

13. Кодекс Республики Казахстан «О налогах и других обязательных платежах в бюджет». – Алматы: Центральный Дом Бухгалтера, 2008.-82 с.

14. Микросхемы АЦП и ЦАП. Справочник. – М.: Изд. дом «Додэка – XXI», 2005. – 432 с.

15. Магда Ю.С. Raspberry Pi. Руководство по настройке и применению. - М.: ДМК Пресс, 2014. – 188 с.

16. Новиков Ю.В., Скоробогатов П.К. Основы микропроцессорной техники. – М.: БИНОМ. Лаборатория знаний, 2006. – 359 с.

17. Ревич Ю.В. Практическое программирование микроконтроллеров Atmel AVR на языке ассемблера. – СПб.: БХВ-Петербург, 2011. – 352 с.

18. СанПин 1.01.004-01 «Гигиенические требования к организации и условиям работы с видеодисплейными терминалами и персональными вычислительными машинами».

19. Техническая документация. Однокристальные 8-разрядные FLASH микроконтроллеры. – М.: ООО «Микро-Чип», 2002. – 163 с.

19. Угрюмов Е.П. Цифровая схемотехника. – СПб.: БХВ-Петербург, 2010. – 578 с.