

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН

Казахский национальный исследовательский технический университет имени  
К. И. Сатпаева

Институт автоматике и информационных технологий

Кафедра «Электроники, телекоммуникации и космических технологий»

Специальность 6В06201 – Телекоммуникация

Абдинаби Мадина Абдуалиевна

**ДИПЛОМНАЯ РАБОТА**

«Исследование биометрических особенностей речевых сигналов»


Алматы 2024

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ  
КАЗАХСТАН

Некоммерческое акционерное общество «Казахский национальный  
исследовательский технический университет имени К.И.Сатпаева»

Институт Автоматики и информационных технологий

Кафедра Электроники, телекоммуникации и космических технологий

  
**ДОПУЩЕН К ЗАЩИТЕ**  
Заведующий кафедрой ЭТиКТ  
кандидат технических наук,  
ассоциированный профессор  
*[Signature]* Таштай Ерлан  
«30» *май* 2024 г.

**ДИПЛОМНАЯ РАБОТА**

На тему: «Исследование биометрических особенностей речевых сигналов»

6B06201 – Телекоммуникация

Выполнила: *[Signature]*

Рецензент

к.т.н. доцент АУЭС им. Г.Даукеева

*[Signature]* Вайкенов А.С.

«30» *мая* 2024 г.

*[Signature]* Блдинаби Мадина

Научный руководитель

к.т.н. старший преподаватель

*[Signature]* Джунусов Н.А.

«05» *мая* 2024 г.

Алматы 2024

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН

Казахский национальный исследовательский технический университет  
имени К.И. Сатпаева

Институт автоматизации и информационных технологий

Кафедра «Электроники, телекоммуникации и космических технологий»

ОП «6В06201 Телекоммуникация»



**ЗАДАНИЕ**

**на выполнение дипломной работы**

Обучающемуся Абдинаби Мадина Абдуалиевне

Тема: Исследование биометрических особенностей речевых сигналов

Утверждена приказом ректора университета № 5487/0 от «4 дек.»

Срок сдачи законченной работы «30» мая 2024 г.

Исходные данные к дипломной работе: Основная частота вибрации голосовых связок составляет около 110 Гц, 200 Гц и 300 Гц; LTV модель для представления речи; выбирается БПФ 2048 точек; основная частота -10,79 Гц; блок FORMAT имеет 24 байта; кепстральный метод речевого сигнала для измерения периода шага; казахские фонемы замечаются в интервале 400-800 Гц.

**Краткое содержание дипломной работы:**

- а) Аналитический обзор основных показателей речевого сигнала
- б) Анализ биометрических параметров и показателей речевого сигнала
- в) Методы и алгоритмы обработки, используемые в анализах речевого сигнала
- г) Расчет параметров речевого сигнала

**Перечень графического материала (с точным указанием обязательных чертежей):**

1. Модель представления речевого сигнала
2. LTV модель для представления речи
3. Кепстральный метод речевого сигнала
4. Алгоритм определения комплексного кепстра
5. Алгоритм процесса принятия кепстра
6. Моделирование речевого сигнала в среде MATLAB
7. Рисунок – Временной и спектральной характеристики казахских фонем


## ГРАФИК

подготовки дипломной работы (проекта)

| Наименования разделов, перечень разрабатываемых вопросов               | Сроки представления научному руководителю и консультантам | Примечание |
|--|---|------------|
| Аналитический обзор основных показателей речевого сигнала              | 1.02.2024   | Выполнил   |
| Анализ биометрических параметров и показателей речевого сигнала        | 1.03.2024   | Выполнил   |
| Методы и алгоритмы обработки, используемые в анализах речевого сигнала | 25.03.2024  | Выполнил   |
| Расчет параметров речевого сигнала                                     | 15.04.2024  | Выполнил   |

### Подписи

консультантов и нормоконтролера на законченную дипломную работу с указанием относящихся к ним разделов работы

| Наименования разделов | Консультанты, И.О.Ф. (уч. степень, звание)                  | Дата подписания | Подпись   |
|-----------------------|---|-----------------|---|
| Нормоконтролер        | Старший преподаватель каф. ЭТиКТ, Доктор Ph.D. Досбаев Ж.М. | 30.05.2024      |  |

Научный руководитель  Джунусов Н.А.

Задание принял к исполнению студент  Абдинаби М.А.

Дата «20» декабря 2023 г.

## АННОТАЦИЯ

В дипломной работе была детально рассмотрена тема «Исследование биометрических особенностей речевых сигналов», направленная на изучение уникальных характеристики голоса человека и их применение для идентификации личности, аутентификации, обнаружения эмоционального состояния и оценки состояния здоровья.

В ходе исследований проведен анализ спектральных, временных, амплитудных и других параметров речевых сигналов с использованием методов обработки сигналов и машинного обучения. Полученные данные позволяют создавать биометрические системы, способные точно идентифицировать говорящего по его уникальным акустическим особенностям.

## АҢДАТПА

Диплом адам дауысының бірегей сипаттамаларын зерттеуге және оларды жеке сәйкестендіруге, аутентификацияға, эмоционалдық жағдайды анықтауға және денсаулықты бағалауға бағытталған «Сөйлеу сигналдарының биометриялық ерекшеліктерін зерттеу» тақырыбы егжей-тегжейлі қарастырылды.

Зерттеу барысында сигналдарды өңдеу және машиналық оқыту әдістерін қолдану арқылы сөйлеу сигналдарының спектрлік, уақытша, амплитудалық және басқа параметрлерін талдау жүргізілді. Алынған деректер динамикті оның бірегей акустикалық сипаттамалары негізінде дәл анықтай алатын биометриялық жүйелерді жасауға мүмкіндік береді.

## ANNOTATION

The thesis examined in detail the topic «Study of biometric features of speech signals», aimed at studying the unique characteristics of a person's voice and their use for personal identification, authentication, detection of emotional state and health assessment.

During the research, the analysis of spectral, temporal, amplitude and other parameters of speech signals was carried out using signal processing and machine learning methods. The data obtained makes it possible to create biometric systems that can accurately identify a speaker based on his unique acoustic characteristics.

## СОДЕРЖАНИЕ

|   |    |
|---|----|
| Введение  |    |
| 1 Аналитический обзор основных показателей речевого сигнала   | 8  |
| 1.1 Анализ биометрических параметров и показателей речевого сигнала   | 8  |
| 1.2 Особенности формирования речевых сигналов   | 12 |
| 1.2.1 Артикуляторная фонетика   | 15 |
| 1.2.2 Акустическая фонетика   | 16 |
| 1.2.3 Слуховая фонетика   | 19 |
| 1.2.4 Классификация речевых сигналов  | 20 |
| 1.3 Корреляционный анализ сигналов  | 22 |
| 2 Методы и алгоритмы обработки, используемые в анализах речевого сигнала  | 27 |
| 2.1 LTV модель для представления речи   | 27 |
| 2.2 Кепстральный метод речевого сигнала   | 30 |
| 2.2.1 Алгоритм определения комплексного кепстра   | 30 |
| 2.2.2 Алгоритм процесса принятия кепстра  | 31 |
| 2.3 Модель представления речевого сигнала в среде MATLAB  | 33 |
| 2.3.1 Применение MATLAB для обработки речевого сигнала  | 36 |
| 2.3.2 Моделирование временных, спектральных характеристик и линейных кодовых значений речевого сигнала в среде MATLAB | 37 |
| 3 Расчет опорных параметров речевого сигнала  | 57 |
| 3.1 Расчет временной и спектральной характеристики казахских фонем  | 57 |
| 3.2 Расчет оценки энергетического спектра речевого сигнала  | 58 |
| 3.3 Расчет равномерного распределения громкости речевого сигнала  | 59 |
| Заключение  | 60 |
| Перечень принятых сокращений, терминов  | 61 |
| Список использованной литературы  | 62 |
| Приложение А  | 65 |

## ВВЕДЕНИЕ

Актуальность темы исследования. Под речевой деятельностью понимается процесс говорения, протекающий во времени и облеченный в звуковую или письменную форму. С помощью речи человек познает мир в полной мере и способен передавать накопленные знания другим людям. Человек издает звуки благодаря строению голосового аппарата.

Использование голосового аппарата для аутентификации личности является удобным и надежным методом, основанным на уникальных голосовых характеристиках каждого человека. Распознавание личности по голосу, задача, поставленная практически полвека назад, продолжает привлекать внимание исследователей. Эти задачи являются одной из форм биометрической аутентификации.

Вместе с бурным развитием информационных технологий возникает потребность в специализированных монографиях по обработке речевых сигналов.

Именно поэтому исследование биометрических особенностей речевых сигналов является актуальной темой. Время от времени появляются новые методы и технологии, которые помогают улучшить точность и надежность распознавания личности по голосу. Это важно для обеспечения безопасности информации и для развития более удобных и эффективных систем аутентификации.

Исследования в данной области способствуют развитию медицинских технологий. Результаты могут помочь в диагностике различных заболеваний, связанных с голосовыми характеристиками. Кроме того, популярность голосовых интерфейсов и виртуальных ассистентов растет в геометрической прогрессии. Так, чтобы создать более естественные и удобные пользовательские интерфейсы, необходимо понимать биометрические особенности речи.

Степень разработанности темы.

Вопрос исследования биометрических особенностей речевых сигналов является предметом интереса для таких специалистов, как инженеры-разработчики программного обеспечения, специалисты по цифровой обработке сигналов, лингвисты и фонетисты, психологи и исследователи в области биометрии и телекоммуникации.

В работе рассмотрены пособия, публикации, научные статьи авторов, исследующих данную тему: В.П. Морозов, В.Н. Сорокин, А.С. Леонов, Н.Б. Покровский, Л.Р. Рабинер, А.А. Афанасьев, А.И. Иванов, Н.К. Смоленцев, А.Н. Гучай и других.

Целью дипломной работы является исследование биометрических особенностей речевых сигналов, в том числе анализ казахских фонем.

Задачи исследования были поставлены в соответствии с выдвинутой актуальностью и целью исследования:



- 1) провести аналитический обзор основных показателей речевого сигнала;
- 2) проанализировать биометрические параметры и показатели речевого сигнала;
- 3) рассмотреть особенности формирования речевых сигналов;
- 4) выполнить корреляционный анализ сигналов;
- 5) исследовать кепстральный метод речевого сигнала;
- 6) представить речевой сигнал в среде MATLAB;
- 7) произвести расчет временной и спектральной характеристики казахских фонем.

Предметом исследования является анализ уникальных биометрических параметров, присутствующих в речевых сигналах.

Объектом исследования являются разнообразные параметры и характеристики голоса, используемые для уникальной идентификации личности.

В процессе работы применялись общенаучные и специальные способы познания, среди которых: сравнительно-правовой метод, экономический анализ, логический анализ, а также экспертное оценивание. Использование данных исследований позволяет комплексно и всесторонне изучить объект и предмет дипломной работы.

Теоретическая значимость исследования состоит в теоретическом понимании биометрических особенностей речевых сигналов. В процессе ознакомления с работой читатель сможет расширить знания о биологических и акустических особенностях голоса. Результаты исследования могут быть использованы для анализа более точных методов аутентификации и идентификации личности на основе речевых данных.

Практическая значимость исследования заключается в том, что результаты проведенной работы могут использоваться в качестве рекомендательного пособия для студентов высших учебных заведений, а также являться практическим руководством для педагогического состава университетов.

Структура и объем исследования. Дипломная работа включает в себя введение, три главы, заключение, список использованной литературы и приложение.



# 1 Аналитический обзор основных показателей речевого сигнала

## 1.1 Анализ биометрических параметров и показателей речевого сигнала

Обработка речевых сигналов появилась в 70-х годах. Некоторыми из важных аспектов цифровой обработки речи являются:

- высококачественное кодирование (перцептивное кодирование) речи и аудио;
- распознавание речи;
- улучшение и модификация речи и аудио;
- синтез текста в речь;
- синтез речи в текст [1, с.15].

В начале проведения исследований необходимо измерить первичные характеристики и границы зон параметров речевого сигнала, передающих (создающих) ощущения звуков речи соответствующего языка. Как известно, ощущение звуков речи можно создать, генерируя их как, естественными системами (речь человека или, например, попугая), так и искусственными. При естественной или искусственной генерации речи в речевом сигнале изменяются физические параметры, которые, воздействуя на мембрану уха, возбуждают группы рецепторов. Изменения этих параметров во времени создают звуковые образы (траектории параметров во времени), воспринимающиеся как соответствующие звуки языка.

Исследователи рассматривают речевые и языковые показатели как признаки присутствия говорящего, состоящие в основном из слушания и речи. Симптомы речевых и языковых нарушений, характеризующиеся грамматическими, словарными и синтаксическими ограничениями, возникают в основном из-за неправильного положения языка. Модель структурной мультистратификации языковых симптомов – это взаимосвязь между положением и функционированием языка [2, с.64].

Распознавание речевых эмоций (РРЭ), подотрасль эмоциональных вычислений, существует уже более двух десятилетий. Она привела к появлению значительного количества исследовательских работ. Она предполагает распознавание эмоциональных аспектов речи независимо от ее смыслового содержания.

Типичную систему РРЭ можно рассматривать как набор методик, которые выделяют, извлекают и классифицируют речевые сигналы для обнаружения заложенных в них эмоций. Примеров использования РРЭ в реальных приложениях бесчисленное множество, некоторые из которых продемонстрировали, что включение эмоциональных атрибутов во взаимодействие человека и машины может значительно улучшить опыт взаимодействия пользователей. Например, система РРЭ может оценивать работу сотрудников колл-центра, выявляя эмоции клиентов (гнев или радость). Эта информация может помочь компаниям улучшить качество обслуживания

или провести целенаправленное обучение. Данный фактор приведет к повышению удовлетворенности клиентов и эффективности работы колл-центра [3, с. 655].

Распознавание речевых эмоций стало важной основой для многих интеллектуальных сервисных систем (здравоохранение, умные дома и интеллектуальные развлечения). Центры экстренной помощи могут использовать анализ речевых эмоций для выявления опасных или угрожающих жизни обстоятельств. Некоторые пользователи умных автомобилей также могут использовать интерактивную систему голосового реагирования для предотвращения аварий из-за усталости водителей. В клинических условиях РРЭ может способствовать укреплению психического здоровья, например, для выявления признаков потенциальных суицидальных мыслей. Для образовательных онлайн-сервисов является ценным инструментом, поскольку позволяет преподавателям оценивать степень овладения учащимися новыми навыками, проведя анализ эмоциональной наполненности их ответов. Это может быть использовано для точной настройки учебного плана и оптимизации процесса обучения [4, с. 3210].

Точность работы систем по распознаванию речи повышается, если учитывать особенности анатомии человека, связанные с полом говорящего лица. Такие системы используются при сортировке телефонных звонков, поиске образцов речи в базах данных, существенно сокращая время получения результата.

Применение систем в медицинском исследовании позволяет отслеживать динамику болезни при реабилитации больных, упрощает задачу ранней диагностики опухолей гортани по речевому сигналу. Наиболее распространенный на сегодняшний день зеркальный осмотр дает долю ошибок от 30% до 50%. Использование томографии на ранних стадиях не позволяет выявить изменений тканей и является дорогостоящей процедурой. Между тем, на ранних стадиях заболевания отмечается изменение голоса пациента, связанное с изменениями параметров речевого аппарата [5, с. 37].

Отличительные особенности речи, спектрограммы и другие атрибуты человеческой речи лежат в основе использования методов машинного обучения для распознавания речевых эмоций. Традиционно ML включает в себя процесс изучения паттернов и вычисления параметров функций на основе необработанных данных (таких как речь, изображения, ЭКГ и видео). Эти функции используются для обучения модели, которая учится выдавать желаемую выходную метку в задаче прогнозирования и/или классификации. Тестирование множества отдельных функций, интеграция различной деятельности в общий вектор функций или использование альтернативных стратегий выбора функций, могут дать некоторое представление о том, какие функции обеспечивают наиболее эффективную кластеризацию данных по классам. Кроме того, современные методы (графики/глубокие нейронные сети) предоставляют более элегантные способы обхода проблем при выборе оптимального признака.

Обработка речевых сигналов является активным направлением исследований в области цифровой обработки сигналов. Биометрический метод – это система, созданная для распознавания определенных биометрических признаков. Примерами биометрических признаков являются лицо, отпечатки пальцев, геометрия кисти, отпечаток ладони, радужная оболочка, голос, подпись, походка и динамика нажатия клавиш.

Что такое анализ биометрических параметров речевого сигнала? Программное обеспечение, предназначенное для речевой аналитики, использует распознавание речи, обработку естественного языка и машинное обучение для преобразования произнесенных слов из разговоров клиентов в текст. Затем программное обеспечение может анализировать этот текст, чтобы получить представление о настроениях, предпочтениях и потребностях клиентов [6, с. 74].

Инструменты речевого анализа позволяют анализировать голосовые записи в режиме реального времени, обеспечивая мгновенную обратную связь для сотрудников контакт-центра и способствуя постоянному совершенствованию. Это позволяет агентам лучше понимать потребности клиентов и соответствующим образом корректировать общение, тем самым повышать производительность и удовлетворенность клиентов.

Голос напрямую сочетает в себе биологические и поведенческие характеристики. Звук, который человек издает при разговоре, зависит от физических особенностей его тела (рот, нос, губы, голосовые связки и т.д.) и может зависеть от возраста, эмоционального состояния, родного языка и состояния здоровья. Качество записывающего устройства и окружающий шум также влияют на скорость распознавания [7, с. 36].



Рисунок 1.1 – Классификация методов биометрических показателей

Пол, возраст, интеллект, региональное и этническое происхождение, образование и род занятий – все это является биометрическим фактором, которые могут быть использованы для распознавания человеческой речи. Он содержит информацию о многочисленных речевых нарушениях, которые могут быть диагностированы путем выявления неправильных показателей речевого сигнала.

К показателям речевого сигнала относятся высота тона или громкости речи, которые не соответствуют возрасту, полу, культурному происхождению или географическому региону человека. Другие заболевания, связанные с речью, включают когнитивно-коммуникативные расстройства, трудности в общении, проблемы с глотанием и заболевания аутистического спектра. В диагностике и мониторинге различных заболеваний решения, основанные на речевых технологиях, сыграли решающую роль. В этой области модели обучения в настоящее время являются передовым подходом.

Благодаря своей низкой стоимости и неинвазивности автоматическая диагностика заболеваний голосовых связок вызвала интерес у широкой публики исследователей и психологов. Данные технологии могут использоваться для ранней диагностики нарушений речи. При исследовании модели «machine learning» и DNNS (Deep neural networks) при анализе показателей нормальной и ненормальной речи из распространенных речевых расстройств, показатели модели DNNS превзошли другие модели машинного обучения, такие как SVMs (Машина опорных векторов) и Hmms (Скрытая Марковская модель). Ученые, использовавшие сверхточную сеть искусственного интеллекта для исследования патологии голоса, показали, что сверхточная нейронная сеть может извлекать характеристики из отпечатков голосовых связок, полезные при обнаружении речевых аномалий.

Согласно исследованиям патологии речевого сигнала с использованием глубоких нейронных сетей, которые сочетают в себе нейронную сеть и сеть долговременной и кратковременной памяти, выяснили, что они дают многообещающие результаты.

Несмотря на то, что в последние годы модели обучения значительно усовершенствовались, некоторые из них не так хорошо известны. Детские НММ-ASR-модули (например, ASR-системы с глубокими нейронными сетями), требуют большого объема данных для обучения и чрезвычайно зависят от используемых данных. Получение данных о клинической речи гораздо сложнее, чем данные о речи в среднем. Необходимы дополнительные исследования для разработки систем анализа с минимальным количеством данных. Ограниченность базы данных, содержащих большие речевые показатели, являются еще одним фактором, препятствующим разработке системы и точности и анализа биометрических параметров и показателей речевого сигнала [9, с. 204].

Анализ биометрических параметров и показателей речи были проведены с помощью распознавания и классификации патологии гортани, чтобы выявить аномалию голосового тракта. Патология вокальных данных прямо

пропорциональна качеству речи. Эта пропорциональность сравнивается с индивидуальными акустическими характеристиками качества голоса людей.

Распознавание эмоций, отчаяния, тревожного расстройства и стресса - все это может быть полезно благодаря широкому спектру просодических и спектральных характеристик, предоставляемых человеческой речью. Чтобы диагностировать клиническую депрессию у подростков, были проанализированы такие акустические параметры, как спектральные, просодические, цепстральные, голосовые и энергетические. По сравнению со всеми другими характеристиками авторы обнаружили, что характеристики, основанные на энергии, дают более оптимистичные результаты. По мнению авторов исследования, характеристики, связанные с качеством речи (ударение и интонация), могут быть использованы для различения подростков, склонных к суициду, и подростков, не склонных к суициду.

Любой язык можно описать в терминах набора отличительных звуков, называемых фонемами. В русском языке 41 фонема, а в казахском языке Сочетания согласных фонем в основном двучленны, их около 312. Они включают гласные, дифтонги, согласные и другие звуки.

Фонетика – это отрасль лингвистики, которая специализируется на изучении звуков человеческой речи. Речь идет о физических свойствах звуков речи, а именно о телефонах, их физиологическом производстве, искусственных свойствах и слуховом восприятии. Изучение фонетики является предметом лингвистики, которая фокусируется на речи.

Фонетика подразделяется на три вида:

- артикуляторная фонетика. Занимается изучением произношения речи артикуляторном и голосовым трактом докладчиком;

- акустическая фонетика. Занимается изучением передачи речи от говорящего к слушающему. Акустическая фонетика - это изучение звуковых волн, создаваемых человеческими вокальными органами для общения;

- слуховая фонетика. Занимается изучением фонетики восприятия и восприятия речи слушателем.

Все части фонетики взаимосвязаны, потому что процесс человеческого общения – это система слуховых механизмов (длина волны, высота тона и другие физические свойства звука).

## **1.2 Особенности формирования речевых сигналов**

Речевой сигнал, исходящий изо рта, носа и щек говорящего, является одномерной функцией (давления воздуха). Микрофоны преобразуют колебания давления воздуха в электрические сигналы, напряжения или токи, в виде которых мы обычно имеем дело с речевыми сигналами при обработке речи. Цифроаналоговые преобразователи преобразуют аналоговые напряжения в двоичные (или n-арные) цифровые сигналы. Речевые сигналы с ограниченной полосой пропускания (например, с помощью телефонной системы) с полосой

пропускания менее 4000 Гц могут быть представлены, согласно теореме о дискретизации, 8000 выборками в секунду. Каждая выборка может быть квантована до 256 уровней (8 бит) с незначительным ухудшением слышимости, если уровни должным образом соответствуют диапазону напряжений сигнала. (Один или два бита на выборку могут быть сэкономлены за счет разумного, неравномерного выбора уровней за счет лишь незначительных слышимых искажений).

Таким образом, общая скорость передачи информации, необходимая для высококачественного представления речевого сигнала в полосе частот до 4 кГц, составляет 8 бит/выборка, умноженная на 8000 выборок в секунду, или 64 Кбит/с. (Для сравнения, скорость передачи данных на стереокомпакт-диске (CD) превышает 1,4 Мбит/с.) Целью сжатия речи является максимальное снижение этой скорости передачи данных для более эффективного хранения и передачи.

Исследования в формировании речевых технологий начались в начале 90-х годов, когда методы распознавания образов и машинного обучения были еще в начальном состоянии. Было проведено несколько исследований с использованием таких фундаментальных методов, как измерение расстояния, векторное квантование и статистический анализ. Однако в последние годы методы машинного обучения стали популярными для обнаружения проблемных голосов с использованием рассчитанных акустических свойств входного сигнала.

Речевой сигнал затем представляется последовательностью чисел  $x_n$ , которые представляют относительное давление воздуха в определенный момент времени  $n \in \mathbb{N}$  (рисунок 1.2). Это представление известно как импульсно-кодовая модуляция, часто сокращаемая как РСМ. Точность этого представления определяется двумя факторами: 1) частотой дискретизации (шагом во времени между  $n$  и  $n+$ ) и 2) точностью и распределением амплитуд  $x_n$ .

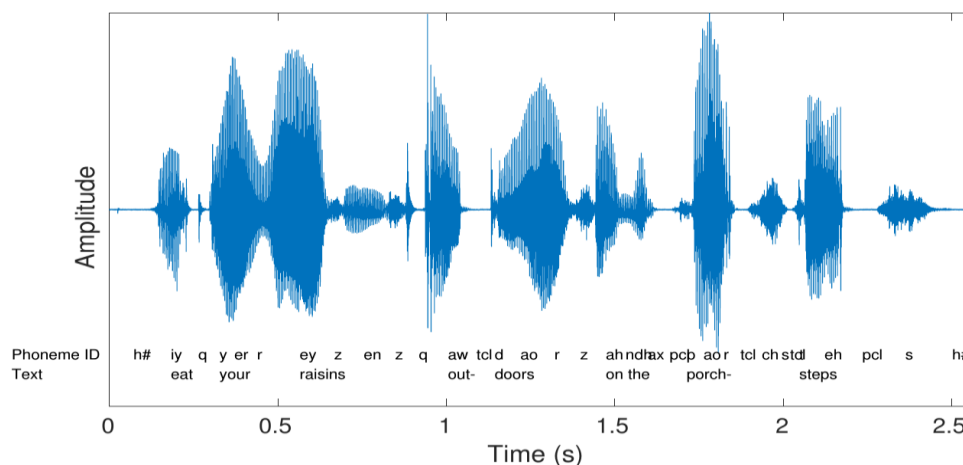


Рисунок 1.2 – Речевой сигнал

Наиболее важной информацией в речевых сигналах являются форманты, которые находятся в диапазоне от 300 Гц до 3500 Гц, так что нижний предел частоты дискретизации составляет около 7 или 8 кГц. Фактически, первые цифровые речевые кодеки (например, AMR-NB), используют частоту дискретизации 8 кГц, известную как узкополосная. Однако некоторые согласные, особенно фрикативные, такие как /s/, содержат значительную энергию, превышающую 4 кГц, поэтому узкополосной частоты недостаточно для обеспечения высокого качества речи. Однако большая часть энергии остается ниже 8 кГц, так что широкополосной частоты дискретизации 16 кГц достаточно для большинства целей. Сверхширокополосный и полнополосный диапазоны также соответствуют частотам дискретизации 32 кГц и 44,1 кГц (или 48 кГц) соответственно. Последняя частота дискретизации также используется в компакт-дисках (CD). Такие более высокие показатели полезны при рассмотрении также неречевых сигналов, таких как музыка и общий звук [11, с.77].

Характеристики модели зависят от того, как произносится гласный или согласный звук. В первую очередь, акцент делается на механизме образования гласных звуков. Когда голосовые связки подвергаются напряжению за счет окружающей мускулатуры, давление воздуха из легких заставляет голосовые связки вибрировать. Например, чтобы наглядно представить себе этот эффект, нужно поднести резинку губам. Если держать ее открытой, когда выдувается воздух через нее, он проходит более или менее свободно, такое действие соответствует «процессу дыхания».

Если держать резинку туго и плотно прижатой друг к другу, то при дуновении через отверстие боковые стороны резинки начинают вибрировать. Этот эффект лучше всего достигается при использовании широкой резинки. Этот процесс похож на поток воздуха на противоположной стороне резинки или на голосовых связках. Сила легких - это простой источник, который можно смоделировать как постоянную подачу воздуха под давлением. Голосовые связки реагируют на такой сигнал вибрацией, что означает, что на выходе этой системы происходит некоторая периодическая функция.

Исследования о биомеханике шепота раскрывают удивительную динамику, которая происходит в органах речи человека. При шепоте голосовые связки приближаются к узкому треугольному отверстию между хрящами. Проходя через это отверстие, воздух создает турбулентность, что является источником возбуждения для шепота. Когда вокальный тракт находится в другом положении, воздушный поток снова становится турбулентным, создавая характерный широкополосный шум. Интересно, что частотный спектр этого шума отражает местоположение сужения. Звуки, порождаемые этим возбуждением, называются шипящими или щелевыми.

В процессе шепота возможны два типа звуков: остановки и аффрикаты. Остановки происходят, когда поток воздуха полностью прекращается закрытием голосовых связок. Аффрикаты же характеризуются постепенным и турбулентным потоком воздуха.



Речевой сигнал, в основном, представляет собой свернутый сигнал, возникающий при взаимодействии возбуждения с импульсной характеристикой голосового тракта. Генерация речи происходит в звуковом ящике, который состоит из голосовых связок. Вибрация голосовых связок создает основное возбуждение. Оно проходит через голосовой тракт и превращается в речевой сигнал.

Хотя исследователи еще не смогли полностью синтезировать естественный речевой сигнал, они активно изучают его параметры и используют методы синтеза для создания сигналов схожего качества. Тем не менее, существующие синтезированные сигналы не идентичны естественным. Это свидетельствует о сложности математического описания речевого сигнала, который не является детерминированным и не поддается прогнозированию на основе предыдущих образцов.

### 1.2.1 Артикуляторная фонетика

Артикуляция – конфигурация голосового тракта (гортани и глоточной, ротовой и носовой полостей), возникающая в результате расположения подвижных органов голосового тракта (например, языка) относительно других частей голосового тракта, которые могут быть жесткими (например, твердого неба). Эта конфигурация модифицирует воздушный поток для воспроизведения звуков речи. Основными артикуляторами являются язык, верхняя губа, нижняя губа, верхние зубы, верхний десневой отросток (альвеолярный отросток), твердое небо, перистальтика (мягкое небо), язычок (свободно свисающий конец мягкого неба), стенка глотки и голосовая щель (пространство между голосовыми связками).

Артикуляционная фонетика рассматривает анатомио-физиологическую базу артикуляции (речевой аппарат) и механизмы речепроизводства.

Артикуляционная фонетика – это раздел фонетики, изучающий физиологические компоненты звукообразования. Когда человек издает звук, он использует свой язык, зубы, твердое небо, мягкое небо, голосовую щель, голосовые связки, диафрагму. Артикуляционная фонетика изучает органы голосового тракта и то, как они используются для воспроизведения определенных звуков. Традиционные описания согласных (т.е. произношение, место произношения, манера произношения) требуют тех же знаний об артикуляторах, что и в этой области фонетики [12, с.102].

Артикуляционная фонетика имеет дело с физическим воспроизведением звуков речи. В нее входят фазы речи, компоненты (механизмы подачи воздуха, инициация, фонация, высота звука и форма голосового тракта), способы артикуляции (стопа, носовой, трель, постукивание / хлопанье, фрикативный, приближающий), места произношения артикуляция (ротовая и глоточно-гортанная), видоизмененная и двойная артикуляция, а также гласные (округлые/ неокругленные, близкие/средние/открытые, передние /центральные/задние, «кардинальные», монофтонги, дифтонги). Все

компоненты акцента могут быть описаны в артикуляционных терминах (сегментация, динамика голоса, качество голоса).

Артикуляционные действия составляют физическую основу распознавания звуков речи на слух. Фонетические символы, которые используются для представления звуков речи, имеют собственное артикуляционное отображение.

Активный артикулятор – это артикулятор, который перемещается по направлению к другому артикулятору при воспроизведении звука речи. Этот артикулятор перемещается по направлению к другому артикулятору, образуя в голосовом тракте какое-либо замыкание (например, открытое приближение, закрытие).

Пассивный артикулятор – это артикулятор, который остается неподвижным при произнесении звуков речи. Часто активный артикулятор движется именно к этому месту (например, к твердому небу).

Манера артикуляции относится к способу произнесения звука, а не к месту его произнесения. Звуки различаются по способу их произношения. Когда артикуляторы приближаются друг к другу, поток воздуха меняется в зависимости от конкретного типа звука. Например, поток воздуха можно полностью перекрыть или сделать турбулентным.

Когда речь заходит о гласных, мы используем другие определения для их описания. Мы обращаем внимание на вертикальное положение языка, горизонтальное положение языка и положение губ.

Гласные произносятся при свободном прохождении воздушного потока по средней линии голосового тракта. Они обычно озвучены и производятся без трения [13, с.178].

### 1.2.2 Акустическая фонетика

Изучение акустической фонетики значительно расширилось в конце XIX века. Акустическая фонетика – это раздел фонетики, который изучает акустические аспекты звуков речи. Речь идет об амплитуде звукового сигнала, его продолжительности, основной частоте и других свойствах частотного спектра. Изучение акустической фонетики имеет большое значение для учителей иностранных языков при обучении произношению, а логопедов - при оценке и планировании коррекционных мероприятий.

Фонетисты могут использовать такое оборудование, как осциллограммы и спектрограммы, для анализа частоты и продолжительности издаваемых звуковых волн.

Существует четыре акустических свойства звуков речи: частота, время, амплитуда и форманта.

Частота относится к отдельным пульсациям, производимым колебаниями голосовых связок в течение единицы времени. Частота колебаний зависит от длины, толщины и натяжения голосовых связок и, следовательно, различна для речи ребенка, взрослого мужчины и женщины. Речевой звук содержит два типа

частот: основную частоту, которая относится к функции голосовых связок и отражает частоту вибрации голосовых связок во время произнесения (высоту тона), и формирующую частоту, которая относится к конфигурации голосового тракта.

Время как свойство звуков речи отражает длительность произносимого звука [14, с. 2100].

Амплитуда определяется томностью полос: чем больше интенсивность звуковой энергии в данный момент времени и на определенной частоте, тем темнее будет отметка в соответствующей точке экрана.

Форманта или Волна – это концентрация акустической энергии вокруг определенной частоты в речевой волне. Существует несколько формант, каждая из которых имеет разную частоту, примерно по одной в каждом диапазоне частот 1000 Гц. Иными словами, форманты встречаются с интервалом примерно в 1000 Гц. Каждая форманта соответствует резонансу в голосовом тракте.

Звук создается вибрирующим предметом, например камертоном, расположенным ниже.

Когда рычаг камертона перемещается наружу, он сталкивает соседние молекулы воздуха (повышая давление воздуха в этой области). Скопившиеся молекулы воздуха будут стремиться переместиться в менее плотные области (с более низким давлением), что приводит к их перепополнению.

Когда рычаг камертона перемещается внутрь, соседние молекулы воздуха становятся менее плотными (т.е. создается частичный вакуум). Молекулы воздуха из окружающих областей будут стремиться переместиться в новую область с низким давлением.

В результате возникает чередующаяся картина областей высокого и низкого давления, удаляющихся от вибрирующего объекта. Это звуковая волна.

На рисунке 1.3 представлена сложная волна, возникающая в результате сложения трех простых волн. На нем показаны волны сигналов для слов «ага», «ада» и «ата». Как можно заметить, между двумя гласными есть область молчания, представляющая звуки г, д и т. Несмотря на то, что энергия сигнала самая низкая или незначительная, его продолжительность важна для его восприятия. В этом случае мы также можем наблюдать промежутки молчания между несколькими звуками. Они естественны и необходимы для восприятия содержащейся в нем информации. Предположим, что если удалить все области молчания и объединим только озвученные и невокализованные сегменты, то сигнал будет выглядеть так, как показано на том же рисунке 3. Воспринимать информацию из этого сигнала будет очень сложно.

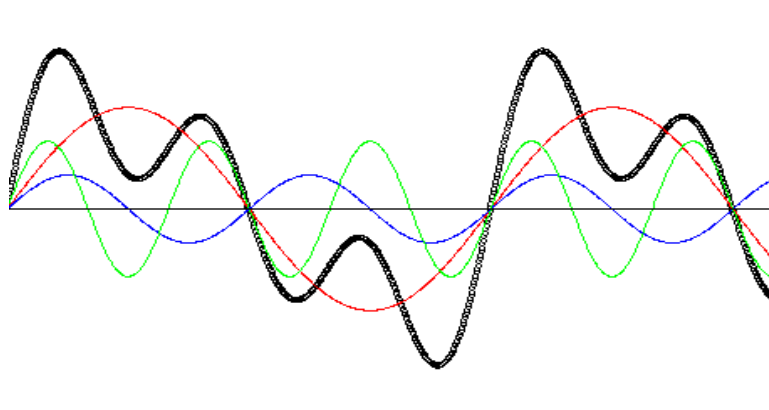


Рисунок 1.3 – Сложная волна

Если перед нами стоит задача изучения свойств простых волн, мы можем изобразить их по отдельности (рисунок 1.4, 1.5, 1.6).

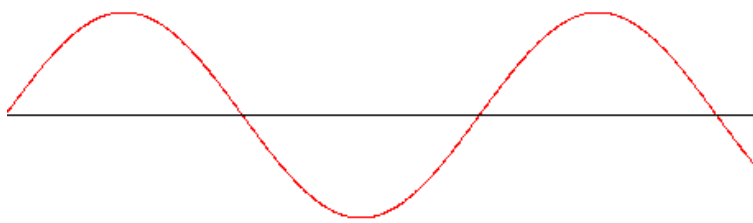


Рисунок 1.4 – Волна А

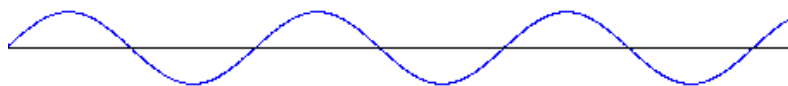


Рисунок 1.5 – Волна Б

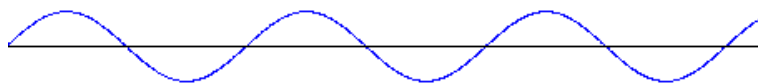


Рисунок 1.6 – Волна В

Рассмотрит тот случай, если бы сложная волна состояла из сотни простых волн.

К счастью, исследователей не совсем волнует, что делает каждая простая волна в каждый момент времени. Единственное, что их действительно интересует в каждой простой волне, – это ее частота и амплитуда.

Концепции сигнала и системы связаны классической структурной волной. Система реагирует на входной сигнал/возбуждение и выдает выходной

сигнал/отклик. Для данной конструкции системы выходной отклик зависит от типа входного возбуждения. Соответственно, мы можем получить разные выходные сигналы. Для изучения этого эксперимента можно использовать одну и ту же структурную волну. В случае речи она может быть изменена, как показано на рисунках выше. Система воспроизведения речи реагирует на входное возбуждение, генерируя речевой сигнал.

Процесс создания речи включает последовательное генерирование озвученной и невокализованной речи, разделенных так называемой областью молчания. Во время области молчания голосовой тракт не возбуждается и, следовательно, речь не выводится. Однако тишина является неотъемлемой частью речевого сигнала [20, с. 85]. Без наличия зоны молчания между озвученной и невокализованной речью речь будет неразборчивой. Кроме того, продолжительность молчания наряду с другой озвученной или невокализованной речью также является показателем определенной категории звуков.

Акустически вокальный тракт представляет собой трубку с неравномерным поперечным сечением длиной около 17 см, которая обычно открывается на одном конце и закрывается на другом. Трубка вокального тракта имеет много собственных частот. Если вокальный тракт имел равномерное поперечное сечение, эти частоты были бы заданы формулой (1.1).

$$F_n = \frac{C(2n - 1)}{4l} \quad (1.1)$$

где  $C = 350$  м/с;  
 $l = 17$  см.

Несмотря на то, что с точки зрения амплитуды/энергии область молчания не имеет значения, ее продолжительность очень важна для разборчивой речи.

### 1.2.3 Слуховая фонетика

Слуховая фонетика это то, как мы воспринимаем и слышим звуки, а также то, как ухо, мозг и слуховой нерв воспринимают звуки. Этот раздел посвящен физиологическим процессам, связанным с восприятием речи.

Акустическая фонетика имеет дело с физическими свойствами звуков речи, в то время как слуховая фонетика имеет дело с тем, как эти звуки воспринимаются и обрабатываются слуховой системой человека.

Слуховая фонетика рассматривает слушателя в более широком аспекте. Воспринимаемое значение измеряется в единицах высоты тона и рассчитывается в шкалах Mel или bark.

Основной особенностью звуков речи является способ возбуждения голосового тракта. Голосовое и невокализованное возбуждение являются основными компонентами возбуждения. При нормальном произнесении речи

под гортанью за счет усилий мышц и легких повышается давление воздуха, что приводит к раздвиганию голосовых связок. Голосовая щель начинает открываться, и скопившийся воздух выходит из легких через голосовую щель (пространство между голосовыми складками) и через возможное сужение голосового тракта. Увеличенный поток воздуха снижает местное давление, и голосовые связки пытаются сомкнуться из-за их эластичного натяжения. Этот цикл повторяется, и голосовые связки располагаются таким образом, чтобы воздух проходил через голосовую щель, вызывая вибрацию голосовых складок. Звук, производимый в процессе, называется звонкими звуками, а период времени между последовательными раскрытиями голосовых связок (частота вибрации) называется основной частотой или высотой тона.

Как объяснялось выше, когда источник издает звук, распространяется звуковая волна, которая заставляет молекулы на своем пути собираться вместе и снова расходиться, то есть вибрировать. Это колебание молекул воздуха достигает нашего уха и одновременно вызывает вибрацию барабанной перепонки в среднем ухе, и эта вибрация передается от барабанной перепонки к трем маленьким косточкам: молоточку, наковальне и стремени. Теперь эта вибрация передается во внутреннее ухо и в улитку - маленький, похожий на спираль орган, наполненный жидкостью. Внутри улитки находятся две мембраны: вестибулярная и базилярная. Именно последняя играет центральную роль в процессе прослушивания, поскольку именно здесь расположены слуховые рецепторные клетки [21, с.88].

Клетки базилярной мембраны, стимулируемые частотой поступающего звука, преобразуют эти колебания в нервные сигналы, которые передаются по слуховым нервам к центральному рецептору и регулятору - головному мозгу. Однако эта стимуляция зависит от частоты, то есть, чем выше частота, тем выше вибрация мембраны. Затем мозг помогает нам идентифицировать поступающий звук как реальный звук с их собственными специфическими значениями высоты звука.

Кратким примером манипулирования звуковыми характеристиками является реализация физических свойств звука в качестве входных данных для фонологии конкретного языка. Например, для изучения тональности языка исследуются вариации омофонным слов, которые могут создавать разницу в значении и приводить к контрастным тональным парам. Слово, которое произносится с пятью различными значениями высоты звука, может иметь пять различных значений в тональном языке, таком как китайский, в то время как в нетональном языке, таком как английский, оно может не иметь никакого значения и означать одно и то же, когда произносится и воспринимается с 10 различными значениями высоты звука. Чудеса создания, восприятия и использования речи людьми гораздо глубже, и все они неразделимы.

#### 1.2.4 Классификация речевых сигналов

Метод загрузки возвращает два значения, первое из которых - фактическая звуковая волна, а второе - «частота дискретизации». Как вы знаете, «звук» - это аналоговый сигнал, и для того, чтобы преобразовать его в цифровую форму и представить в виде чего-то вроде числового массива, нужно произвести сэмплирование исходного сигнала. Короче говоря, сэмплирование - это просто «выбор» конечного числа точек из исходного сигнала и отбрасывание остальных. Можно сохранить эти выбранные точки в массиве и выполнять с ними различные дискретные операции. Теорема Найквиста-Шеннона о дискретизации показала, что при достаточно высокой частоте дискретизации можно уловить всю информацию, содержащуюся в сигнале, и даже полностью восстановить ее.

Длина кадра обычно выбирается такой, чтобы длительность кадра по времени  $T=N/v$  (сек.) составляла 10-20 мс. Пусть на текущем кадре длины  $N$  наблюдается последовательность отсчетов  $s_1, \dots, s_k, \dots, s_N$ . Рассмотрим основные параметры речевого сигнала, используемые ниже.

1. Кратковременная энергия речевого сигнала определяется по формуле (1.2).

2.

$$E = \frac{1}{N} \sum_{k=1}^N s_k^2 \quad (1.2)$$

3. Число нулей интенсивности  $Z$  рассчитывается по формуле (1.3).

$$Z = \frac{1}{2} \sum_{K=2}^N |\text{sign}(s_K) - \text{sign}(s_{K-1})| \quad (1.3)$$

где

$$\text{sign}(S) = \begin{cases} 1, & s \geq 0; \\ -1, & s < 0 < n - 1. \end{cases}$$

4. Коэффициенты разложения в ряд Фурье  $c_0, c_1, \dots, c_{N/2}$  представлены в формуле (1.4) и (1.5).

Кадр определяет периодическую функцию с периодом 1, заданную на сетке из точек вида :

$$f_l = f(x_l) = s_{k+1}, l = Nt + k \quad (1.4)$$

Такую функцию можно разложить в ряд Фурье, т.е. представить в виде:



$$f_l = \sum_{q=0}^{N-1} A_q \exp(2\pi i q x_1) \quad (1.5)$$

Скалярное произведение для функций на сетке определяется по формуле (1.6).

$$(f, q) = \frac{1}{N} \sum_{l=0}^{N-1} f_l \bar{q}_l \quad (1.6)$$

Функции  $g_q(x_l) = \exp\{2\pi i q x_l\}$  при  $0 \leq q < N$  образуют ортонормированную систему относительно так введенного скалярного произведения. Коэффициенты Фурье можно найти по формуле (1.7):

$$A_q = (f, g_q) = \frac{1}{N} \sum_{l=0}^{N-1} f_l \exp(2\pi i q x_l) \quad (1.7)$$

Непосредственное осуществление этих преобразований требует  $O(N^2)$  арифметических операций. Для сокращения этого числа применяется алгоритм быстрого преобразования Фурье. Алгоритм основан на том, что при  $N=2^m$  в слагаемых правой части выражений (\*) можно выделить группы, входящие в выражения различных коэффициентов  $A_q$ . Вычисляя каждую группу только один раз можно сократить число операций до  $O(N \times \log_2 N)$ . Если  $N \neq 2^m$ , то в нашем случае можно добавить нулевые отсчеты. Разложение в ряд Фурье дает представление речевого сигнала в виде суммы гармонических колебаний с частотами  $v(q)$ . Запишем соотношение между частотой  $v(q)$  и индексом  $q$  (формула 1.8):

$$v(q) = \frac{v_d q}{N} = 2^{-m} q v_d \quad (1.8)$$

где  $v_d$  - частота дискретизации.

Значения спектра от  $q = N/2 + 1$  до  $N-1$  не содержат новой информации, т.к. значения  $f_l$  действительны. Более точно в формуле (1.9).

$$A_q = A_{N-q} \text{ при } q=0, 1, \dots, N/2 \quad (1.9)$$

Удвоенное значение  $A_q$  - это комплексная амплитуда. Вещественные амплитуды получаются из них по формул t (1.10).

$$c_0 = A_0, c_q = 2|A_q| \text{ при } q=1, \dots, N/2 \quad (1.10)$$

### 5. Распределение энергии сигнала по частотным группам $p_1, \dots, p_{20}$

Но сначала нужно подготовить наборы данных. Как мы видели ранее, существуют формулы произнесенных цифр тремя говорящими. Разбиение данных на классическое распределение по 80-20 тренировочным тестам может оказаться не очень хорошей идеей, поскольку наши тренировочные данные будут содержать те же динамики, что и в тестовых данных, и, вероятно, дадут очень хорошие результаты, в то время как другие динамики будут показывать очень плохие результаты [22, с.71].

## 1.3 Корреляционный анализ сигналов

Корреляция двух функций, сигналов или осциллограмм определяется как мера сходства между этими сигналами. Существует два типа корреляций:

- взаимная корреляция;
- автокорреляция.

Взаимная корреляция между двумя различными сигналами, функциями или формами сигналов определяется как мера сходства или когерентности между одним сигналом и версией другого сигнала с задержкой по времени. Взаимная корреляция между двумя различными сигналами указывает на степень взаимосвязи между одним сигналом и версией другого сигнала с задержкой по времени [23, с. 77].

Взаимная корреляция энергетических (или непериодических) сигналов и энергетических (или периодических) сигналов определяется отдельно.

Автокорреляционная функция определяется как мера сходства или когерентности между сигналом и его версией с задержкой во времени. Таким образом, автокорреляция - это корреляция сигнала с самим собой.

Как и взаимная корреляция, автокорреляция также определяется отдельно для энергетических (или аперидических) сигналов и силовых (периодических) сигналов [24, с.57].

Главными задачами корреляционного анализа являются выявление зависимости или сходства между какими-либо сигналами. Корреляционный анализ имеет широкое применение в различных областях: например, поиск изображения по образцу, поиск полезных сигналов «под шумами», обработка сигналов радаров и т.п.

Корреляция в простом виде – это поэлементное произведение двух сигналов. Если получим большое число – сигналы похожи, если маленькое – нет. Рассмотрим формулу (1.11).

$$r_{12} = \sum_{n=0}^{N-1} x_1[n] * x_2[n] \quad (1.11)$$

Однако, результат выражения по формуле (1.11) будет зависеть от количества отсчётов сигналов. Чтобы это учесть, результат нормируют на количество отсчётов по формуле (1.12).

$$r_{12} = \frac{1}{N} \sum_{n=0}^{N-1} x_1[n] * x_2[n] \quad (1.12)$$

Рассмотрим пример. Возьмём два меандра, представленных на рисунке 1.7 и сдвинутых по фазе и посчитаем их корреляцию по формуле (1.12).

```

1. clear;
2.
3. fs = 100;
4. ts = 0:1/fs:5-1/fs;
5.
6. x1 =
   (square(2*pi*ts)+1)/2;
7. x2 =
   (square(2*pi*ts+pi)+1)/2;
8.
9. subplot(2,1,1);
10. plot(x1), grid on;
11. xlabel('Временные
   отсчёты, n'),
   ylabel('x1[n]');
12.
13. subplot(2,1,2);
14. plot(x2), grid on;
15. xlabel('Временные
   отсчёты, n'),
   ylabel('x2[n]');
16.
17. r12 = sum(x1.*x2)

```

Рисунок 1.7 – Меандра, сдвинутая по фазе

В результате выполнения скрипта получим графики сигналов  $x_1$  и  $x_2$  и значение корреляции  $r_{12} = 0$  (рисунок 1.8).

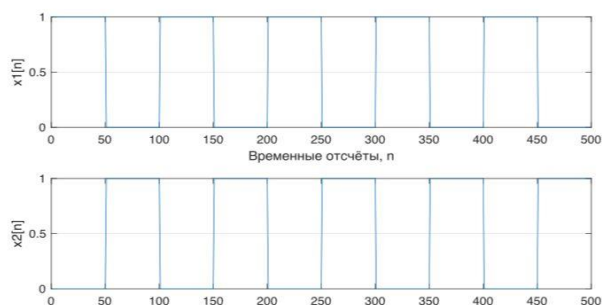


Рисунок 1.8 – Два меандра, сдвинутые на 180 градусов

Получается, что сигналы похожи, а корреляция равна нулю. Значит, выражение (1.12) необходимо модифицировать и добавить временной сдвиг одного сигнала относительно другого по формуле (1.13).

$$r_{12}[j] = \frac{1}{N} \sum_{n=0}^{N-1} x_1[n] * x_2[n+j] \quad (1.13)$$

Результат выполнения модифицированного кода показан на рисунке 1.9.

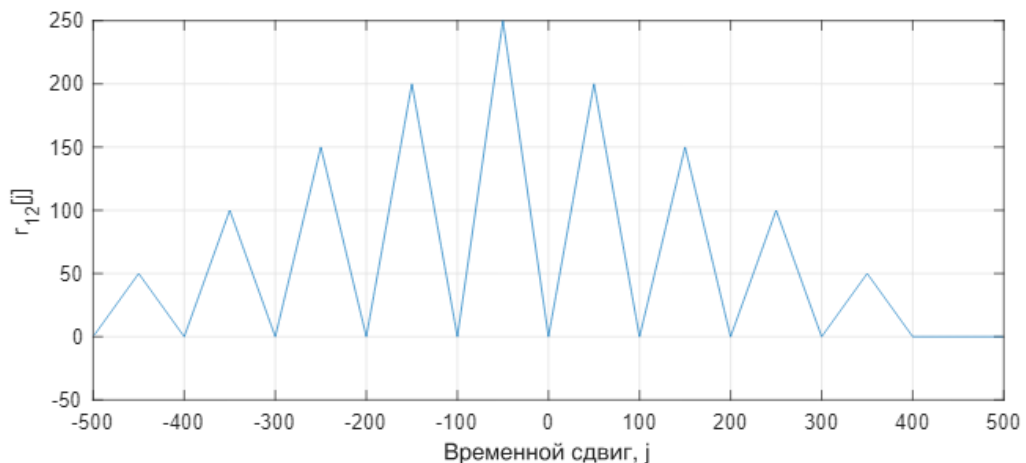


Рисунок 1.9 – Корреляционная функция двух меандров

По графику можно понять, что при нулевом сдвиге корреляция нескольких сигналов пропорциональна нулю  $r_{12}[0] = 0$ , при дальнейшем движении влево или вправо показатели  $r_{12}$  увеличиваются, затем снова уменьшаются. Такое явление происходит часто, однако амплитуда пиков  $r_{12}$  постепенно уменьшается к краям графика. Наверняка многие ожидали увидеть в качестве  $r_{12}(j)$  периодический треугольный сигнал.

Однако, из-за того, что сигналы  $x_1[n]$  и  $x_2[n]$  имеют конечную длину, получается, что при сдвиге одного сигнала относительно другого, наступает момент, когда они не перекрываются и вместо парных произведений отсчётов получаем произведения на «пустые отсчёты», или на ноль. Чем дальше два сигнала сдвигаются друг относительно друга, тем больше получается произведений с нулевым результатом, тем меньше получается результирующая  $r_{12}$ . Эта неприятность называется краевой эффект [28, с.54].

Согласно теории воспроизведения речи с использованием источника-фильтра, голосовая речь представляется как реакция фильтра голосового тракта на речевой источник. Голосовой источник состоит из квазипериодических импульсов, которые создаются колебаниями голосовых связок. Он характеризуется двумя критическими моментами: глобальным закрытием (ГЗ) и открытием (ГО). ГЗ и ГО необходимы для точной оценки многих обращений в различных областях речевой деятельности, таких как оценка качества голоса,

анализ и кодирование речи, идентификация говорящего и оценки голосового источника [29, с.103].

Исходным параметром гортани, широко используемым для оценки ГЗ и ГО, является коэффициент открытости (КО). Он определяется как соотношение между продолжительностью фазы открытой гортани и периодом речи.

Фаза раскрытия – это часть голосового цикла, в течение которой голосовая щель открыта. Таким образом, это продолжительность между одним ГЗ и последующим ГО. Речевой период - это интервал, ограничивающий два последовательных ГО.

Голосовой диапазон представляет значительный интерес, поскольку, он связан с качеством голоса, таким как «хриплый» и «сдавленный» тембр. Голос с придыханием возникает, когда голосовые складки не полностью смыкаются во время голосового цикла и, следовательно, голосовой диапазон велик. Сдавленный голос вырабатывается при суженной голосовой щели и соответствует небольшому голосовому голосу.

Голосовые и невокализованные сегменты могут быть дискриминированы с использованием нормализованного отношения до аномальной энергии, определяемого формулой (1.14).

$$P_t = \frac{\sum_{i=1}^N |s(i) - s(i-1)|}{\sum_{i=1}^N s^2(i)} \quad (1.14)$$

где,  $P_t$  - предварительно выделенное энергетическое отношение;

$s(i)$  - образцы  $i$ -го сигнала;

$s(i-1)$  - образцы  $(i-1)$ -го сигнала.

Дисперсия разницы между соседними отсчетами для голосовой речи будет очень малой и что для невокализованной речи будет выше. Голосовые и невокализованные сегменты могут быть дискриминированы с использованием отношения длинной полосы к полному диапазону. В озвученном речевом сигнале энергия сосредоточена в нижней полосе. Мы можем измерить отношение энергии в первой полосе 1 кГц к полной полосе. Это соотношение является высоким для голосовой речи и довольно низким для невокализованной речи.

## 2 Методы и алгоритмы обработки, используемые в анализах речевого сигнала

### 2.1 LTV модель для представления речи

Прежде чем ознакомимся с моделью представления речи LTV, рассмотрим применение быстрого преобразования Фурье (БПФ) в измерении шага. БПФ представляет собой эффективный алгоритм для вычисления дискретного преобразования Фурье (DFT), который находит широкое применение в обработке сигналов и анализе данных. Например, если мы возьмем DFT из 128 точек, мы получим 128 комплексно-сопряженных коэффициентов. График амплитуды БПФ, построенный на основе номера частоты, отражает распределение амплитуд сигнала в различных частотных компонентах. Это позволяет анализировать частотные характеристики сигнала, в том числе и при измерении шага, где изменения частоты могут быть связаны с шаговым ритмом или другими движениями человека. Коэффициент дискретного преобразования Фурье имеет вид, приведенный в формуле (2.1):

$$X(k) = \sum_{n=0}^{N-1} x(n)W^{nk} \quad (2.1)$$

где  $W$  задается формулой (2.2):

$$W = e^{-j2\pi/N} \quad (2.2)$$

Вычисление коэффициентов дискретного преобразования Фурье (DFT) является сложной задачей, требующей выполнения множества умножений и комплексных операций. Для упрощения этого процесса были разработаны алгоритмы быстрого преобразования Фурье (БПФ). Существует два основных алгоритма БПФ: прореживание по частоте и прореживание по времени.

Например, если вычислить 512-точечное БПФ для последовательности из 512 выборок, получится график амплитуды, который будет симметричен относительно центрального значения (в данном случае 256) из-за комплексно-сопряженного свойства коэффициентов DFT. Этот график отображает амплитуду в зависимости от номера частоты.

Для 512-точечного БПФ, если частота дискретизации сигнала составляет 22,100 Гц, то весь диапазон частот 22100 Гц будет разделен на 512 частотных компонентов, каждая из которых будет соответствовать частоте 43,16 Гц. Таким образом, мы можем легко преобразовать ось номера коэффициента DFT в фактические значения частоты в герцах [30, с.36].

Область спектра анализирует амплитуду спектральных компонентов сигнала, позволяя оценить их вклад в общий спектральный состав сигнала.

Для начала берется вокализированный сегмент речи, содержащий 2048 выборок. Затем применяется 2048-точечное преобразование Фурье (FFT), и

первые 1024 значения FFT записываются. Поскольку спектр симметричен, используются только первые половины значений FFT.

Далее отслеживается первый пик на выходе FFT, что позволяет найти основную частоту. Разрешение этого измерения определяется количеством точек в БПФ. Например, если использовать БПФ из 2048 точек, разрешение частоты будет составлять 10,79 Гц. Однако точность измерения зависит от разрешения в домене DFT.

Основная частота вычисляется как номер точки БПФ, где происходит первый пик, умноженный на частотное разрешение. Например, если первый пик имеет значение 12560 и находится на 16-й позиции, то основная частота составит  $10,79 * 16 = 172,26$  Гц.

Результаты анализа отображаются на графиках сигнала и выходного спектра БПФ, что позволяет наглядно представить основную частоту и ее характеристики. В приложении приведена программа на языке MATLAB для спектральной характеристики с использованием преобразования Фурье.

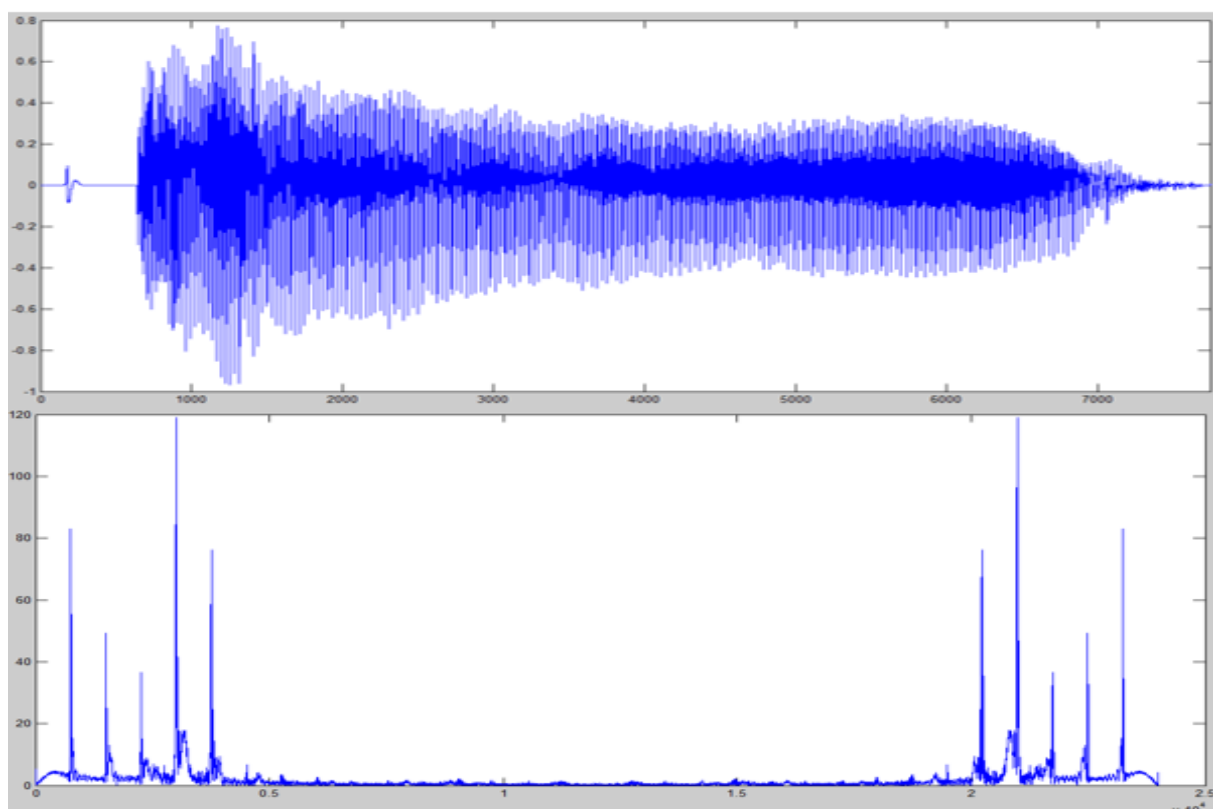


Рисунок 2.1 – Участок вокализованного сигнала / а / и его БПФ

Метод и алгоритм обработки LTV (Long-Term Spectral Variability) для речевых сигналов используются для анализа изменчивости спектральных характеристик речевого сигнала на протяжении времени. Этот метод может быть полезен, например, для изучения произносительных особенностей или динамики речи в различных условиях.



Этот метод обладает широким спектром применений. Широко используется в лингвистике, клинической фонетике и разработке систем распознавания и синтеза речи. Речевой сигнал формируется голосовыми связками и голосовыми каналами. Исследователи обычно не создают речевой сигнал самостоятельно, а анализируют его параметры, используя методы синтеза для генерации аналогичных сигналов. Однако синтезированный сигнал не всегда точно соответствует естественному, поскольку нет математической модели, способной полностью описать речевой сигнал, который является случайным и изменчивым по времени.

Проблема синтеза речи становится сложной из-за изменяющегося характера речевого сигнала. Человек может мгновенно изменять резонанс вокальной полости и модифицировать тональность голоса, создавая различные гласные звуки. Линейно-временная (LTV) модель способна адаптироваться к изменениям во времени, чтобы более точно описывать разнообразие речевых сигналов.

Линейная модель (LTV) изменяется таким образом, чтобы учесть вариации во времени, как показано на рисунке 2.2. Необходимо отслеживать изменение периода основного тона и изменение импульсной характеристики голосового тракта. Обычно для речевого сигнала эти параметры считаются постоянными в течение 20 мс. Используется короткое окно времени для просмотра сигнала и над окном, предполагается, что параметры речевого сегмента остаются постоянными.

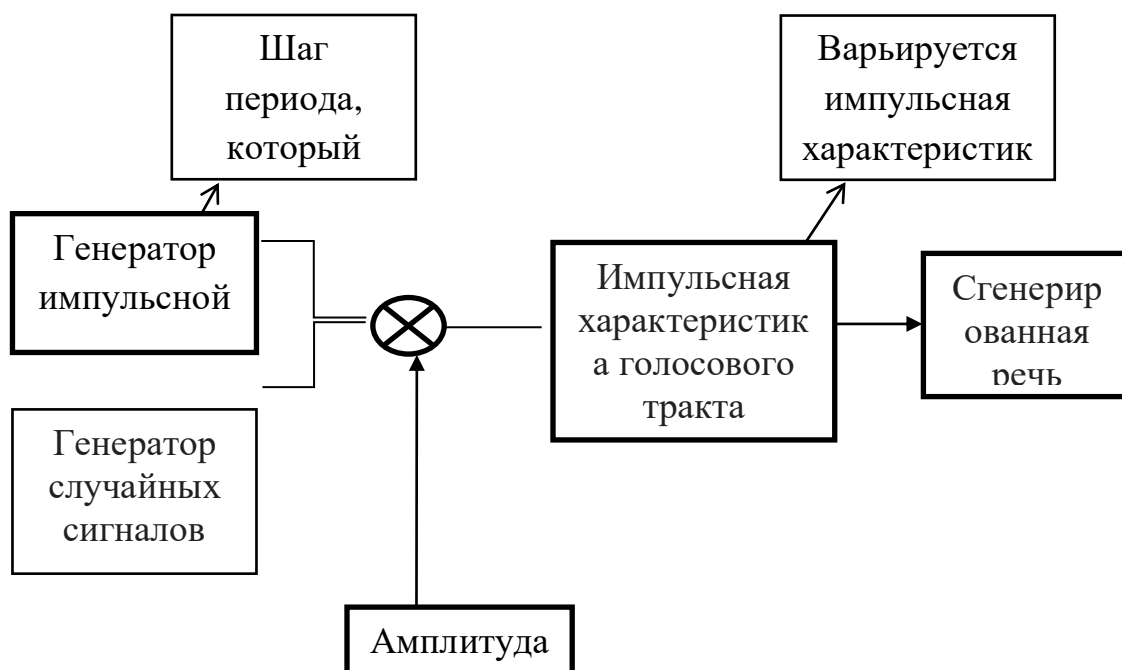


Рисунок 2.2 – LTV модель для представления речи

## 2.2 Кепстральный метод речевого сигнала

Термин кепстр (cepstrum) был введен советским математиком А.Н. Колмогоровым. Кепстр является одним из видов гомоморфных сигналов, функция обратного преобразования Фурье от логарифма спектра мощности сигнала. Кепстр записывается в виде формулы (2.3):

$$C_s(q) = \frac{1}{2\pi} \left( \frac{\ln}{S} (\dot{w}) 1 \right) e^{-i\omega q} d\omega \quad (2.3)$$

где  $S(w)$  — спектр входного сигнала.

На рисунке 2.2 представлен алгоритм определения кепстра.

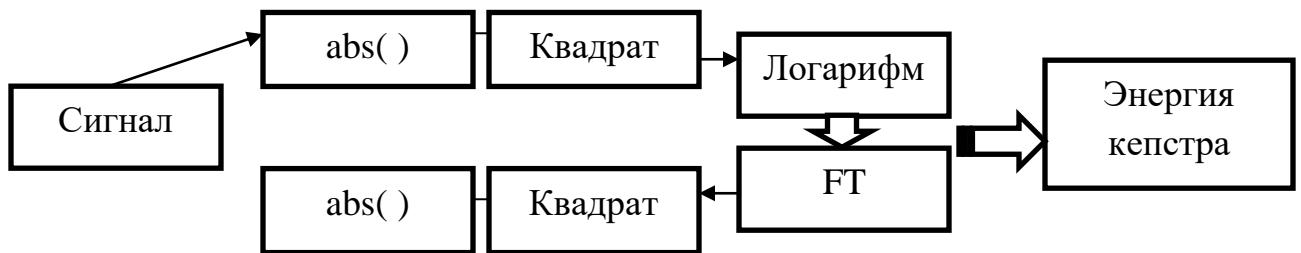


Рисунок 2.2 – Алгоритм определения кепстра

### 2.2.1 Алгоритм определения комплексного кепстра

Изучение спектральных характеристик сигналов – это важная составляющая обработки сигналов, распознавания образов и аудиообработки. В этом контексте комплексный кепстр выступает важным инструментом анализа, который был определен Оппенгеймом в контексте развития теории гомоморфных систем. Комплексный кепстр представляет собой преобразование Фурье логарифма сигнала, полученного с развернутой фазой преобразования Фурье исходного сигнала. Однако его важное свойство заключается в том, что он способен анализировать долгосрочные изменения спектральных характеристик сигнала. Это особенно полезно в области аудио и речевой обработки, где важно изучить динамику изменений звуковых характеристик во времени. Комплексный спектр определим алгоритмически как на рисунке 2.3.

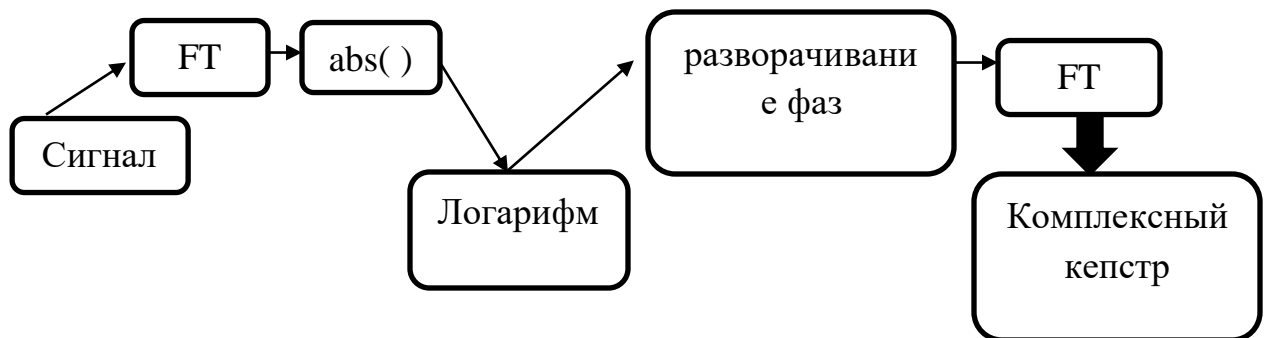


Рисунок 2.3 – Алгоритм определения комплексного кепстра

Реальный кепструм будет использовать функцию журнала, определенную для реальных значений.

Настоящий кепстр связан с энергией кепструма следующим образом по формуле (2.4):

$$4(\text{Реальный спектр})^2 = \text{Спектр мощности} \quad (2.4)$$

Реальный кепстр также определяют также определяют через амплитудный спектр  $A_x(n)$  по формуле (2.5):

$$CC_x[m] = DFT^{-1}[\ln A_x(n)] \quad (2.5)$$

в кадрах сигнала реальный кепстр определяется соотношением по формуле (2.6):

$$CC_x(m, k) = IDFT \ln|x|[n, k] \quad (2.6)$$

Комплексный спектр определяется по формуле (2.7):

$$X_n = X_n \exp(j\Phi_X n)$$
$$CC_x(m) = DFT^{-1}(\ln X)(n) + JDFT^{-1}(\Phi_X)(n) \quad (2.7)$$

Один из методов анализа сигналов, особенно в области обработки речевых сигналов, связан с использованием авторегрессионных моделей.

В этом подходе сигнал моделируется как авторегрессионный процесс, где текущее значение сигнала зависит от предыдущих значений сигнала и случайной ошибки.

Для анализа таких сигналов часто используется метод линейного предсказания (ЛП), при котором на основе предыдущих значений сигнала строится модель, предсказывающая следующее значение.

Коэффициенты этой модели могут быть использованы для вычисления спектра сигнала.

Таким образом, если сигнал моделируется авторегрессионным процессом, его спектр может быть вычислен по коэффициентам линейного предсказания. Однако есть еще один способ вычисления кепстра – непосредственно по спектру линейного предсказания.

Это позволяет получить информацию о спектральных характеристиках сигнала и их изменениях во времени.

## 2.2.2 Алгоритм процесса принятия кепстра

Процесс принятия кепстра определяется алгоритмически. Он представлен на рисунке 2.4.

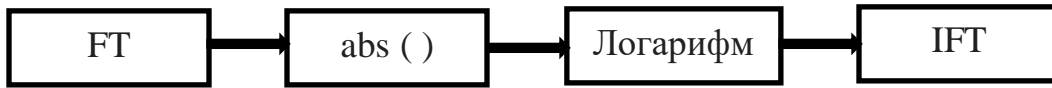


Рисунок 2.4 – Алгоритм процесса принятия кепстра

Уравнение, включающее обратное преобразование Фурье (IFT) вместо прямого преобразования Фурье (FT) применяется в анализе и обработке цифровых сигналов. Несмотря на то, что это изменение может показаться незначительным, оно имеет фундаментальное значение для точности и качества анализа.

При работе с цифровыми сигналами преобразование Фурье и его обратное преобразование являются критически важными элементами. Для дискретных сигналов даже небольшие изменения в фазовой информации могут существенно повлиять на результат обратного преобразования Фурье и, следовательно, на восстановленный сигнал.

Величина может измениться, поскольку уравнение для DFT не будет постоянной, а уравнение для IDFT имеет постоянный множитель  $1 / N$ , как указано ниже в формулах (2.8) и (2.9):

$$x(k) = \sum_0^1 x(n)w^{nk} \quad (2.8)$$

$$x(n) = \frac{1}{N} \sum_0^{-1} x(k)w^{-nk} \quad (2.9)$$

Можно сделать константы равными в обоих уравнениях для DFT и IDFT следующим образом: DFT  $X(k)$  и IDFT  $x(n)$  можно записать в виде формулы (2.10) и (2.11):

$$x(k) = \frac{1}{\sqrt{N}} \sum_0^{-1} x(n) \quad (2.10)$$

$$x(n) = \frac{1}{\sqrt{n}} \sum_0^{-1} x(K)W^{-nk} \quad (2.11)$$

Представленный алгоритм основан на преобразовании Фурье (FT) и обратном преобразовании Фурье (IDFT), применяемом к сигналам. После применения FT к сигналу, последует его абсолютное значение и логарифм, что позволяет оставаться в области спектра, где ось  $x$  обозначает частоту. При применении обратного преобразования IDFT сигнал возвращается во временную область. Однако, присутствует логарифмический блок, который означает, что выходной домен не является ни частотным, ни временным. Этот домен известен как частотный домен, где происходят изменения по частоте.

График кепстрального сигнала состоит из амплитуды и графика погрешности. Когда сигнал анализируется в кепстральном домене, этот процесс называется кепстральным анализом. Короткое время кепстрального анализа было предложено Шредером и Ноллом для определения высоты речи человека.

В кепстре периодического импульсного сигнала (звуков речи, звуков работающих механизмов) формируется пик с частотой, соответствующей частоте следования импульсов.

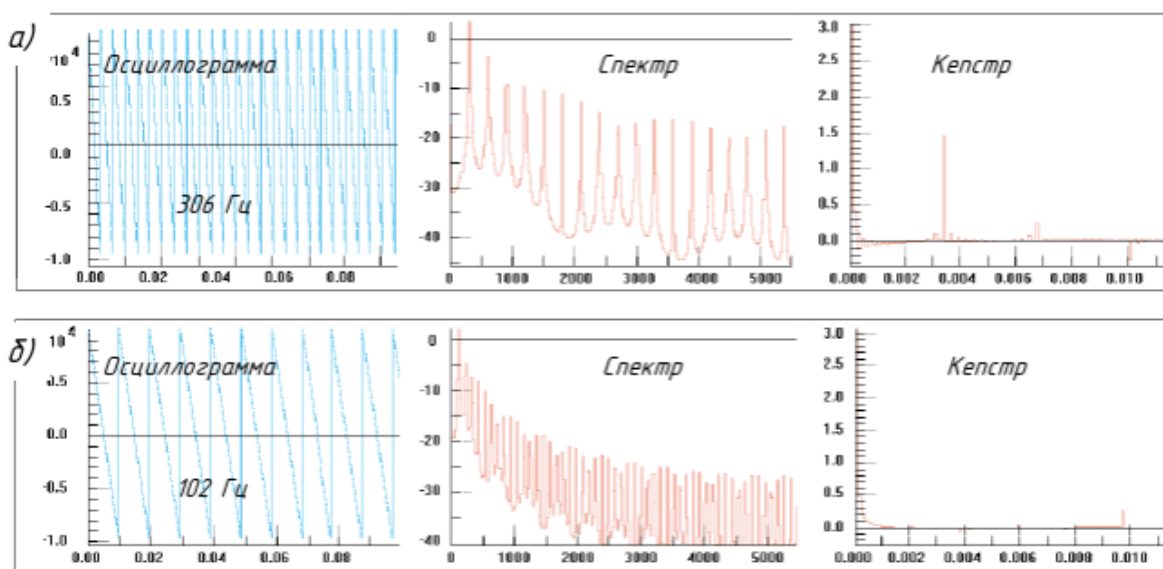


Рисунок 2.5 – Осциллограммы периодических импульсных сигналов, логарифм среднего спектра сигнала, средний спектр сигнала, а) – частота периода 306 Гц, б) частота периода 102 Гц

Исследования в области кепстрального анализа продолжают развиваться, расширяя его применение не только на определение высоты речи, но и на анализ других акустических характеристик, таких как интонация, эмоциональное состояние и идентификация говорящего.

При необходимости определения частоты шага как первого пика в спектре, возникает проблема перекрытия первой форманты и основной частоты. Это затрудняет отделение этих двух компонентов. Однако в кепстральной области возможно эффективное выделение основного тона и первой форманты.

## 2.3 Модель представления речевого сигнала в среде MATLAB

Человеческий голос – это сложный комплексный феномен, состоящий из нескольких параметров, таких как частота, сила, длительность и тембр. Однако при анализе отдельных качеств голоса необходимо понимать, что они образуют неделимый комплекс. Например, высота звука определяется количеством колебаний голосовых складок в секунду, и эти колебания могут происходить как целиком, так и частично. В результате человеческий голос может производить широкий диапазон тональности, от 75 до 1100 Герц, включая различные голосовые типы (бас, тенор, сопрано). Сравнивая его с музыкальными инструментами, видим, что голос перекрывает диапазон их звучания и обладает разнообразием динамического диапазона [37, с.3].

Речь, как основное средство коммуникации, обладает высокой информативностью, но эта информация распределена неравномерно среди различных звуков. Например, глухие согласные обычно несут большую часть смысловой нагрузки, в то время как гласные звуки вносят меньший вклад в понимание речи. Это объясняет, почему разборчивость речи снижается при наличии шумов, особенно из-за маскировки глухих звуков.

Для передачи речевого сообщения требуется значительно больше пропускной способности по сравнению с телеграфным сообщением из-за высокой частоты дискретизации и динамического диапазона речевого сигнала. Образование звуков речи происходит благодаря командам, поступающим от речевого центра мозга к мускулам артикуляционных органов. Этот процесс осуществляется с невысокой скоростью, не превышающей в среднем 100 бит в секунду, и весь остальной объем информации в речевом сигнале считается сопутствующим.

В MATLAB существуют инструменты для работы со звуковыми файлами формата WAV, включая их воспроизведение, запись и анализ. Функция `wavread` используется для чтения WAV-файлов, а функция `wavwrite` - для их записи. При чтении файлов в MATLAB помимо отсчетов сигнала извлекается и служебная информация, такая как частота дискретизации и число бит на отсчет. Графическое представление речевого сигнала позволяет визуально анализировать его особенности и структуру.

В современной цифровой эпохе аудиоинформация играет важную роль, и возможности ее обработки в программной среде, такой как MATLAB, становятся все более значимыми. Одним из ключевых аспектов работы с аудиофайлами является их воспроизведение и анализ.

Функции `sound` и `wavplay` предоставляют удобные способы воспроизведения аудиосигналов в MATLAB. Первая функция, `sound`, позволяет передать аудиоданные непосредственно на звуковую карту компьютера для воспроизведения с учетом заданных параметров, таких как частота дискретизации и число бит на отсчет. Вторая функция, `wavplay`, дает

больше контроля над режимом воспроизведения, позволяя выбирать между синхронным и асинхронным режимом.

Для записи звука используется функция «wavrecord», которая позволяет записать аудиосигнал с микрофона прямо в переменную MATLAB. Это открывает возможности для дальнейшего анализа и обработки записанного звука.

Спектрограмма является мощным инструментом для анализа спектральных характеристик аудиосигнала во времени. В MATLAB для ее создания используется функция «spectrogram», которая вычисляет мгновенные спектры сигнала в различные моменты времени, образуя матрицу спектрограммы.

Анализ и обработка аудиосигналов - это важная область в современной цифровой обработке сигналов. В системе MATLAB доступны различные инструменты для работы с звуковыми файлами, исследования и анализа голосов.

Функции «wavread» и «wavwrite» позволяют читать и записывать файлы формата WAV, что обеспечивает удобство работы с аудиоданными. С помощью функции «wavread» звуковая информация из файла считывается в матрицу, а параметры, в числе которых частота дискретизации и число двоичных разрядов, сохраняются в соответствующие переменные [38, с.16].

Для анализа и обработки аудиосигналов можно использовать различные операции, например, добавление белого шума к аудиозаписи. Это можно сделать, используя функцию «randn» для генерации случайных чисел, распределенных по закону Гаусса, и добавления их к оригинальной аудиозаписи. Для визуализации результатов анализа доступны различные графические инструменты. Например, функция «spectrogram» позволяет построить спектрограмму аудиосигнала, отображая его спектр в зависимости от времени. Это удобно для анализа изменения спектра сигнала во времени.

Проведенные исследования позволяют делать выводы о характеристиках голосов и различиях между ними. Например, анализ амплитудных и фазовых спектров голосов позволяет определить различия в частотных и временных характеристиках голосов мужчин и женщин. На рисунке 2.6 представлена предварительная обработка речевых сигналов в программе MATLAB.



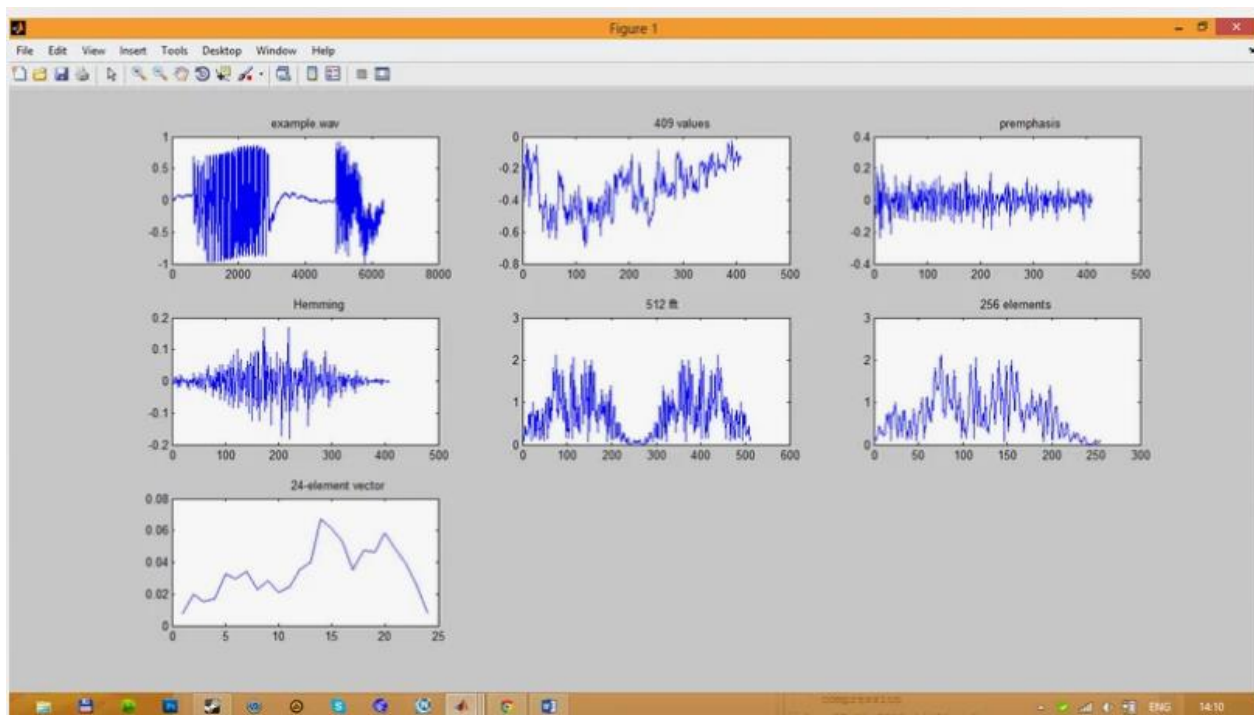


Рисунок 2.6 – Предварительная обработка речевых сигналов с помощью MATLAB

### 2.3.1 Применение MATLAB для обработки речевого сигнала

Блок RIFF состоит из 12 байтов. Первые четыре байта выражены четырьмя символами ASCII, а именно RIFF. Следующие четыре байта указывают общую длину пакета в двоичном формате. Последние четыре байта на самом деле представляют собой четыре символа ASCII, а именно WAVE. В таблице 1 приведены данные для этого блока.

Таблица 1– Блок RIFF (всего 12 Байтов)

| Номер Байта  | Спецификация блока        |
|--|---------------------------|
| 0-3  | RIFF (ASCII)символы       |
| 4-7  | Общая длина пакета данных |
| 8-11   | Символы «WAVE» ASCII      |
| Примечание – составлено на основе источника [40, с.63] |                           |

Блок FORMAT имеет 24 байта, как показано в таблице 2. Первые четыре байта – «fmt». Следующие четыре байта указывают длину блока FORMAT. Он обычно указывается как «0x10». Байты с 12 по 15 указывают частоту дискретизации в Гц и байтах 16-9, байты 13-23 представляют байты в секунду и бит в секунду, соответственно.

Таблица 2 – Блок FORMAT (всего 24 байтов)

| Номер байта  | Спецификация блока   |
|--|--|
| 0-3  | 'fmt' (Символы ASCII)  |
| 4-7  | Длина Блока FORMAT (всегда указывается как 0X10)   |
| 8-9  | Указывается как 0X01   |
| 10-11  | Номера каналов ( 0X01 = Моно, 0X02 = Стерео)   |
| 12-15  | Частота Дискретизации (двоичная, в Гц)   |
| 16-19  | Байт в секунду   |
| 20-21  | Байты представленные в образце : 1 = 8 бит моно, 2 = 8 бит стерео или 16 бит моно, 4 = 16 бит стерео |
| Примечание – составлено на основе источника [41, с.25] |  |

Блок DATA имеет переменное количество байтов в зависимости от длины данных (таблица 3). Спецификация этого фрагмента представлена в таблице 2.3. Байты 4-7 представляют длину данных. Таким образом, длина заголовка для любого WAV-файла составляет 44 байта (12 байтов RIFF-фрагмента + 24 байта блока FORMAT + 8 байт блока DATA).

Таблица 3 – Блок DATA

| Номера байта   | Спецификация блока       |
|--|--------------------------|
| 0-3  | DATA (символы ASCII)     |
| 4-7  | Продолжительность данных |
| 8  | Значения данных          |
| Примечание – составлено на основе источника [42, с.17] |                          |

### 2.3.2 Моделирование временных, спектральных характеристик и линейных кодовых значений речевого сигнала в среде MATLAB

Формантные частоты, присутствующие в речевом сигнале, являются областями спектральной протуберанцы, для которых улучшена акустическая энергия, и обеспечивают сигналы для восприятия как гласных, так и согласных. Форманты являются результатом эффекта взаимодействия резонансов вокального тракта с источником звука и, таким образом, обеспечивают «акустическое окно» в форме голосового тракта, хотя и косвенного. Частично это объясняется тем, что достаточные гармоники, создаваемые голосом, адекватно определяют функцию передачи голосового тракта и четко выражают ее как огибающую речевого спектра.

В казахском языке имеются 37 фонемы, из них 12 гласных и 25 согласных. Классификация гласных фонем казахского языка приведена на рисунке 2.7.

**Дауысты дыбыстар** (Гласные звуки в казахском языке)

| Дауысты дыбыстар түрлері |                  |                         |                  |                          |                     |
|--------------------------|------------------|-------------------------|------------------|--------------------------|---------------------|
| Виды гласных звуков      |                  |                         |                  |                          |                     |
| Тілдің қатынасына қарай  |                  | Жақтың қатынасына қарай |                  | Еріннің қатынасына қарай |                     |
| По подъему языка         |                  | По положению челюсти    |                  | По участию губ           |                     |
| жуан                     | жіңішке          | ашық                    | қысаң            | еріндік                  | езулік              |
| твердые                  | мягкие           | открытые                | сжатые           | губные                   | неогубленные        |
| а, о, ы, ұ, э, у         | ә, ө, і, ү, е, и | а, ә, е, о, ө, э        | ы, и, і, у, ұ, ү | о, ө, у, ұ, ү            | а, ә, е, э, ы, і, и |

**Дауыссыз дыбыстар** (Согласные звуки в казахском языке)

| Дауыссыз дыбыстар түрлері |                     |                                 |
|---------------------------|---------------------|---------------------------------|
| Виды согласных звуков     |                     |                                 |
| Ұяң                       | Үнді                | Қатаң                           |
| Звонкие                   | Сонорные            | Глухие                          |
| б, в, г, ғ, д, ж, з, һ    | й, л, м, н, ң, р, у | к, қ, п, с, т, ф, х, ч, ц, ш, щ |

Рисунок 2.7 – Классификация гласных фонем казахского языка

На рисунке 2.7 показаны временные составляющие казахской фонемы «а», а выборки замечаются в интервале 400-800 Гц. Взята спектральная характеристика казахской фонемы «а» с помощью команды «spectrum.welch» в среде MATLAB для анализа акцентологических особенностей казахского и английских языков.

1) Казахская фонема «а».

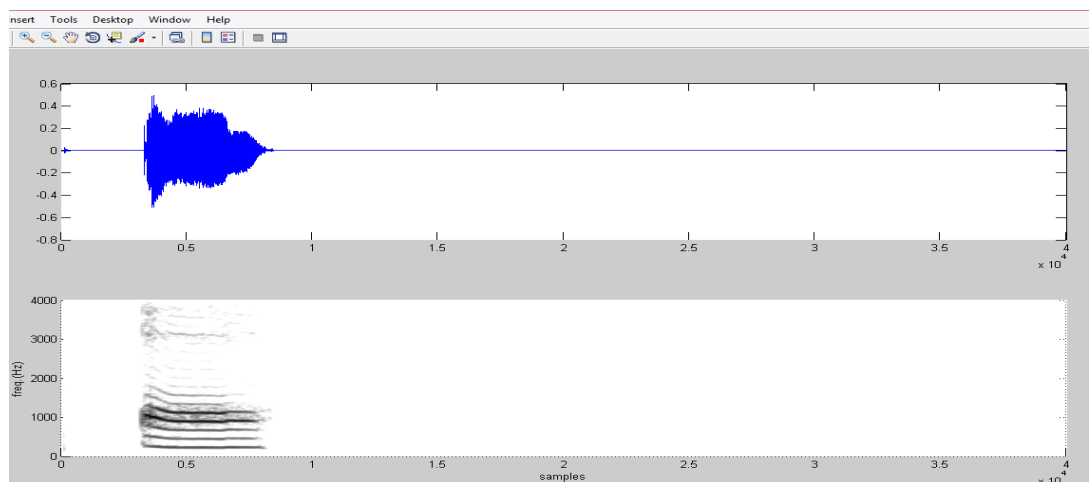


Рисунок 2.8 – Временная и спектральная характеристика казахской фонемы «а»

На рисунке 2.8 взята спектральная характеристика казахской фонемы «а» с помощью команды «spectrum.welch». Приведена программа в среде MATLAB для анализа акцентологических особенностей казахского языка.

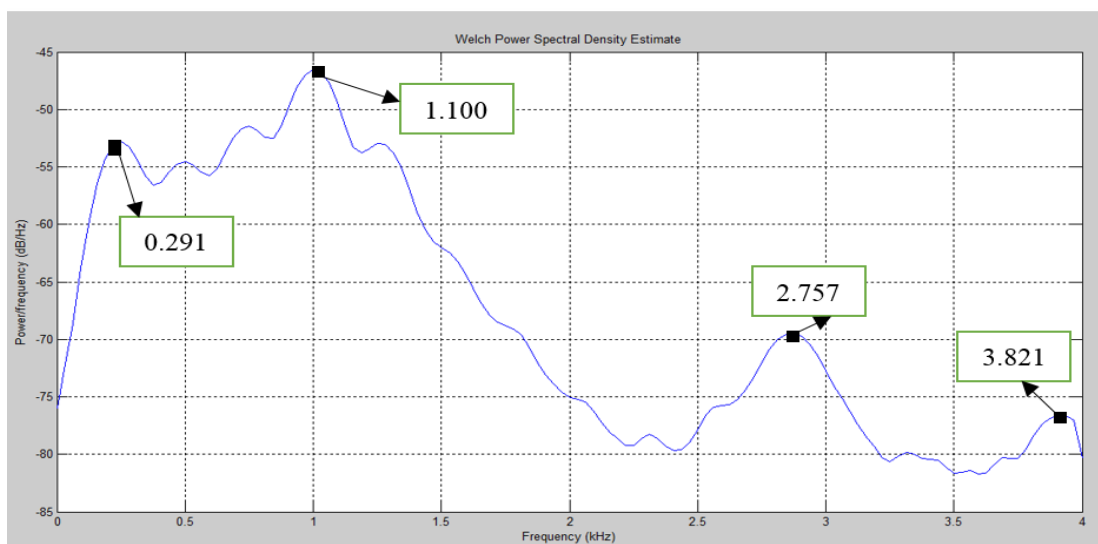


Рисунок 2.9 – Спектральная характеристика казахской фонемы «а»

где  $F_1 = 291$  Гц – первая форманта;  
 $F_2 = 1100$  Гц – вторая форманта;  
 $F_3 = 2757$  Гц – третья форманта;  
 $F_4 = 3821$  Гц – четвертая форманта.

В речевой науке и фонетике формант используется для обозначения акустического резонанса человеческого голосового тракта. Таким образом, в фонетике формант может означать либо резонанс, либо спектральный максимум, создаваемый резонансом. На рисунке 2.10 представлены линейные кодовые значения предсказания фонемы «а».

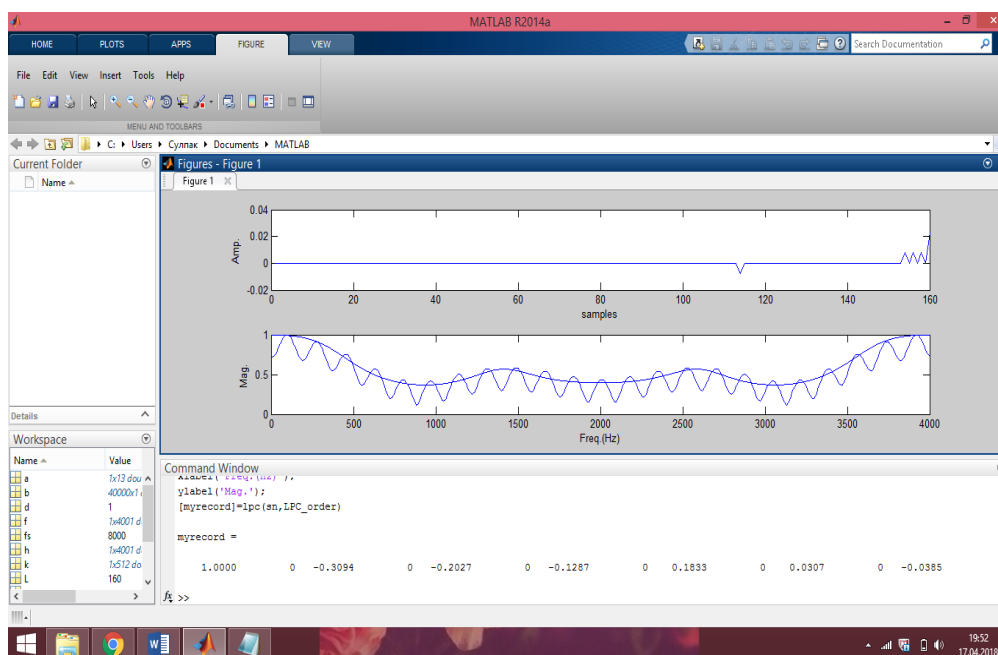


Рисунок 2.10 – Линейные кодовые значения предсказания фонемы «а»

2) Рассмотрим казахскую фонему «э» на рисунке 2.11.

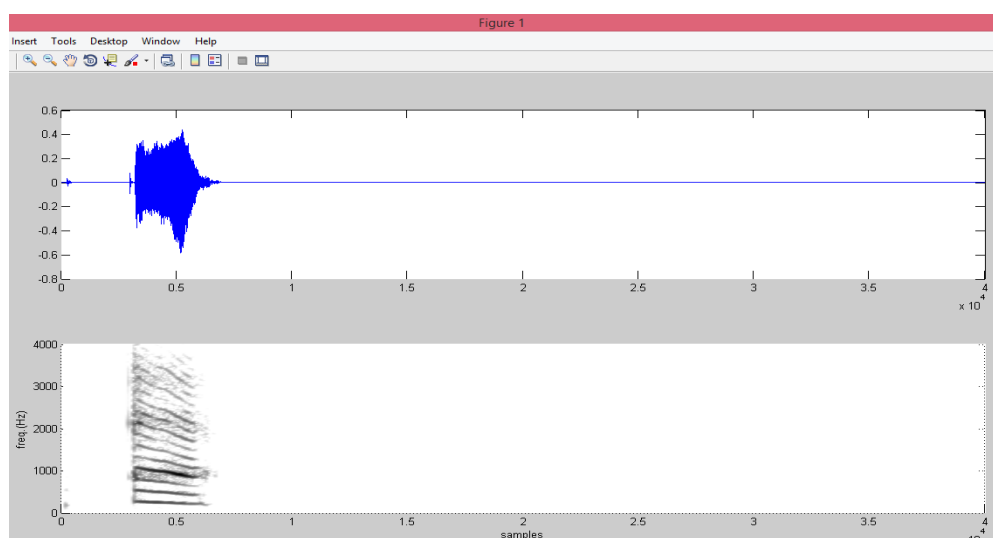


Рисунок 2.11 – Временная и спектральная характеристика казахской фонемы «э»

Тайм-слоты казахской фонемы «э» замечаются до 600 Гц (рисунок 2.12).

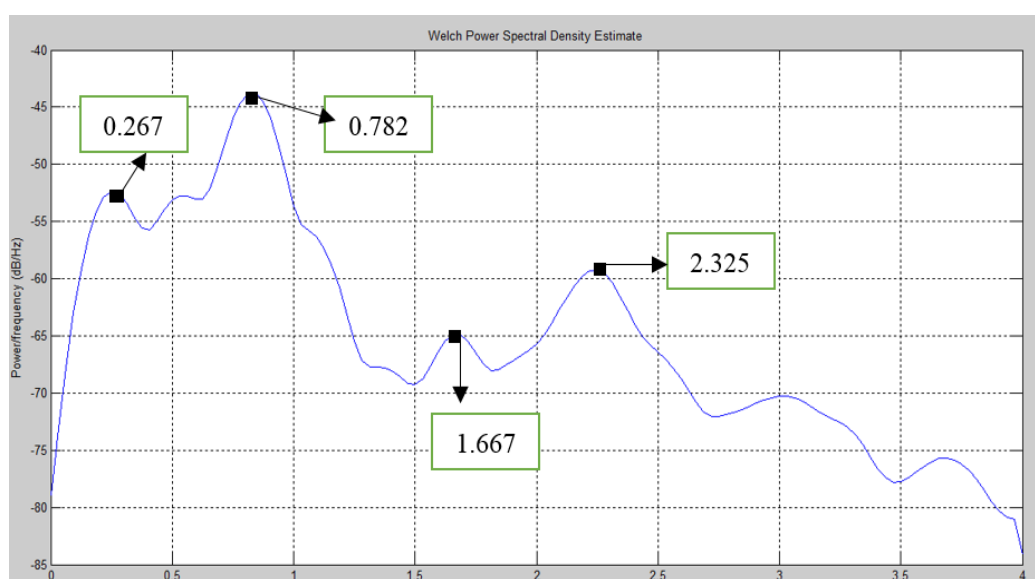


Рисунок 2.12 – Спектральная характеристика казахской фонемы «э»

На рисунке 2.13 показаны формантные частоты казахской фонемы «э»: F1 = 227 Гц; F2 = 782 Гц; F3 = 1667 Гц и F4 = 2325 Гц.



На рисунке 2.15 представлена спектральная характеристика казахской фонемы «о».

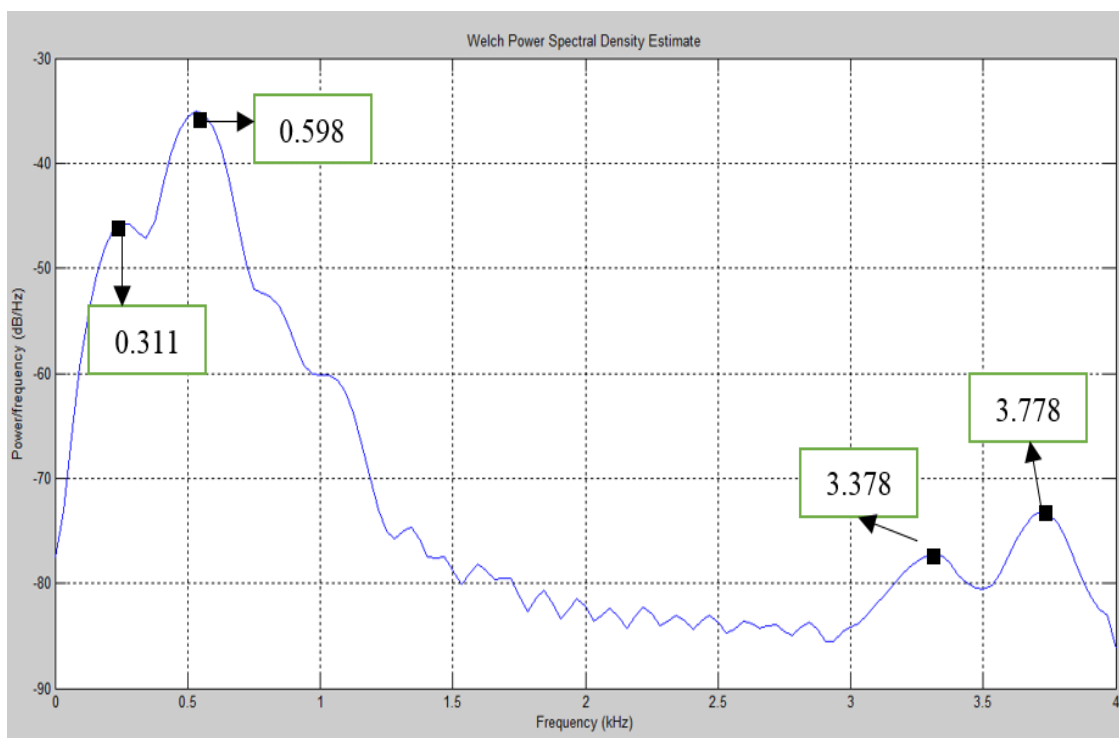


Рисунок 2.15 – Спектральная характеристика казахской фонемы «о»

Здесь форманты резонансных частот: F1=311Гц, F2=588 Гц, F3=3378 Гц и F4=3778 Гц (рисунок 2.16).

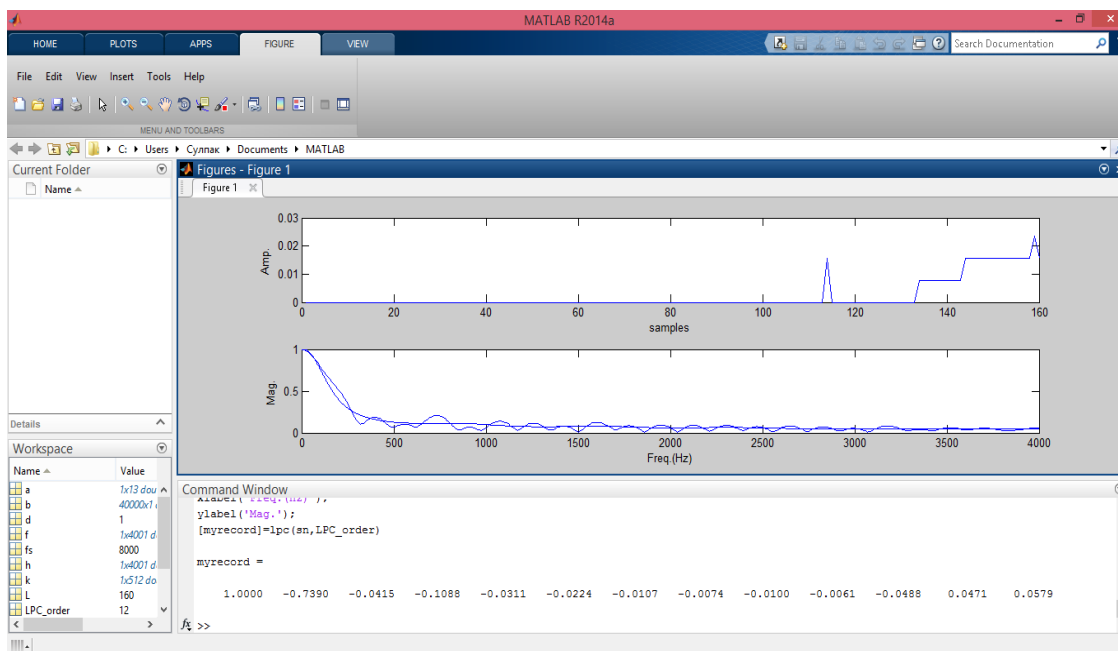


Рисунок 2.16 – Линейные кодовые значения предсказания фонемы «о»

4) Казахская фонема «ө» На рисунке 2.17 показана временная характеристика фонемы «ө».

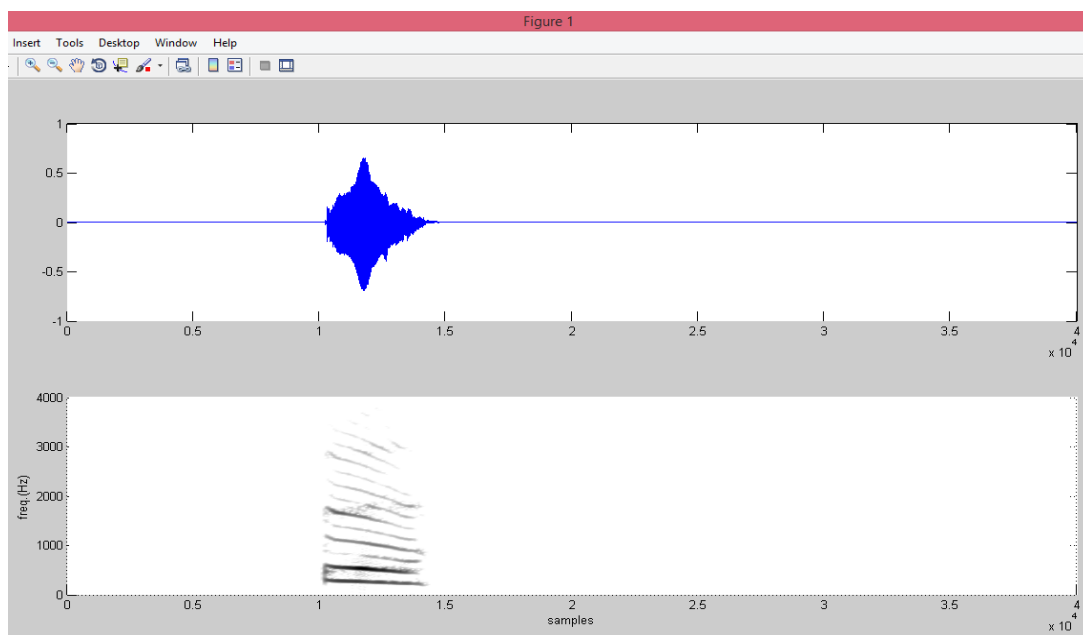


Рисунок 2.17 – Временная и спектральная характеристика казахской фонемы «ө»

Здесь форманты резонансных частот:  $F_1=350$  Гц,  $F_2=501$  Гц,  $F_3=1935$  Гц и  $F_4=2705$  Гц (рисунок 2.18).

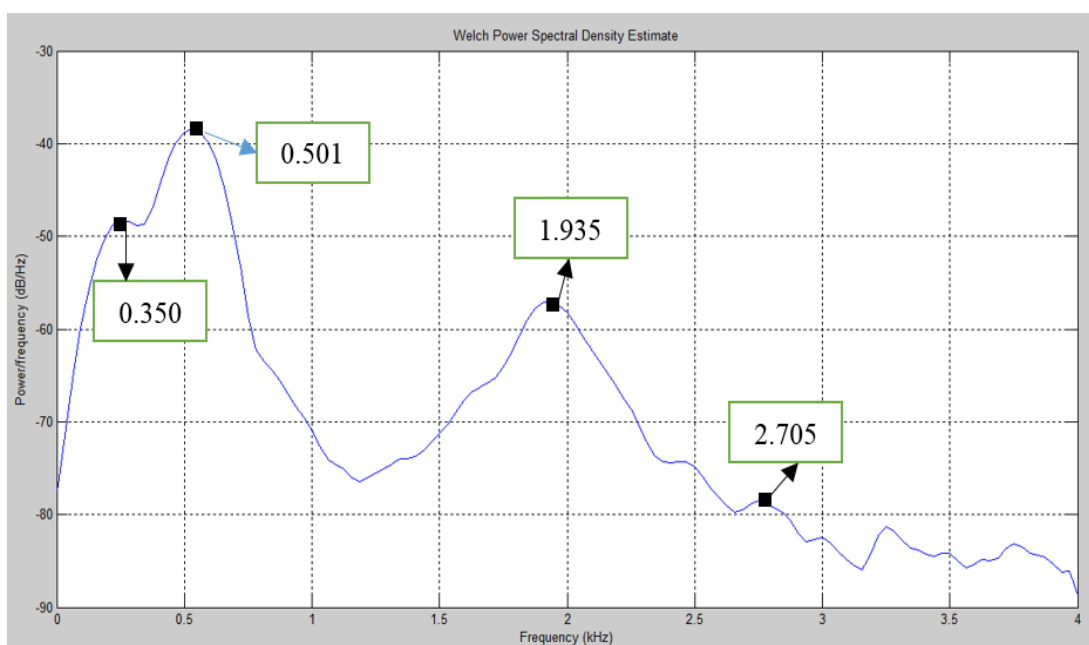


Рисунок 2.18 – Спектральная характеристика фонемы «ө»



На рисунке 2.19 представлены линейные кодовые значения предсказания фонемы «ө».

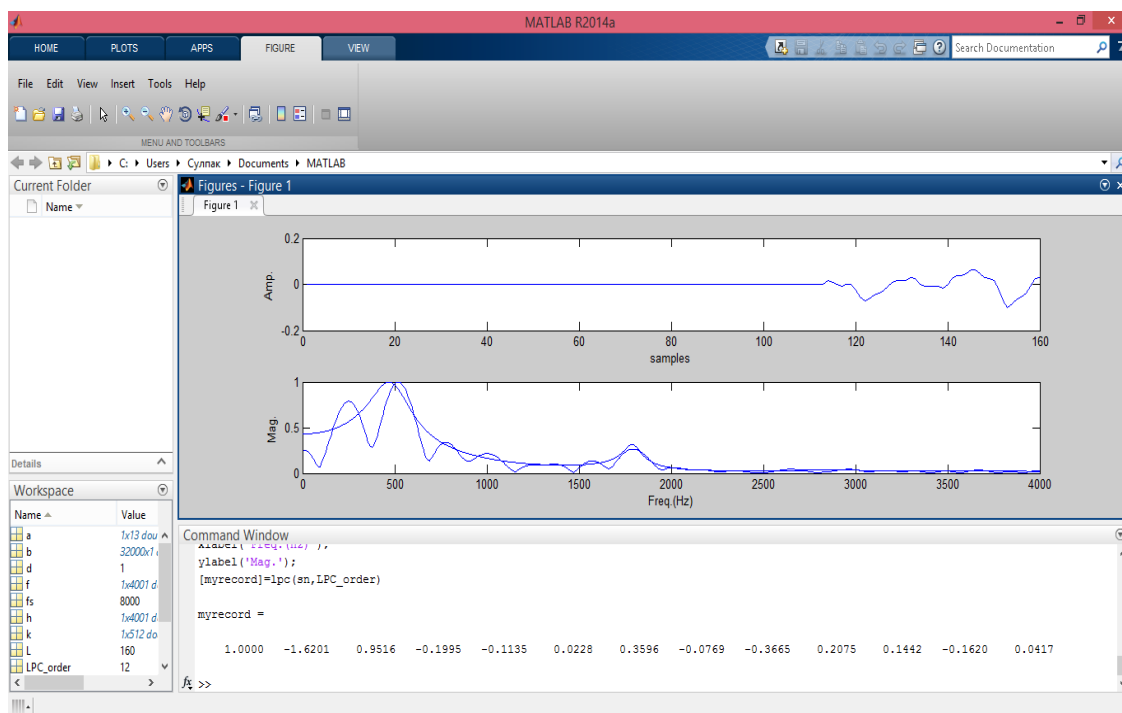


Рисунок 2.19 – Линейные кодовые значения предсказания фонемы «ө»

Казахская фонемы «ы». На рисунке 2.20 показаны временные составляющие казахской фонемы «ы».

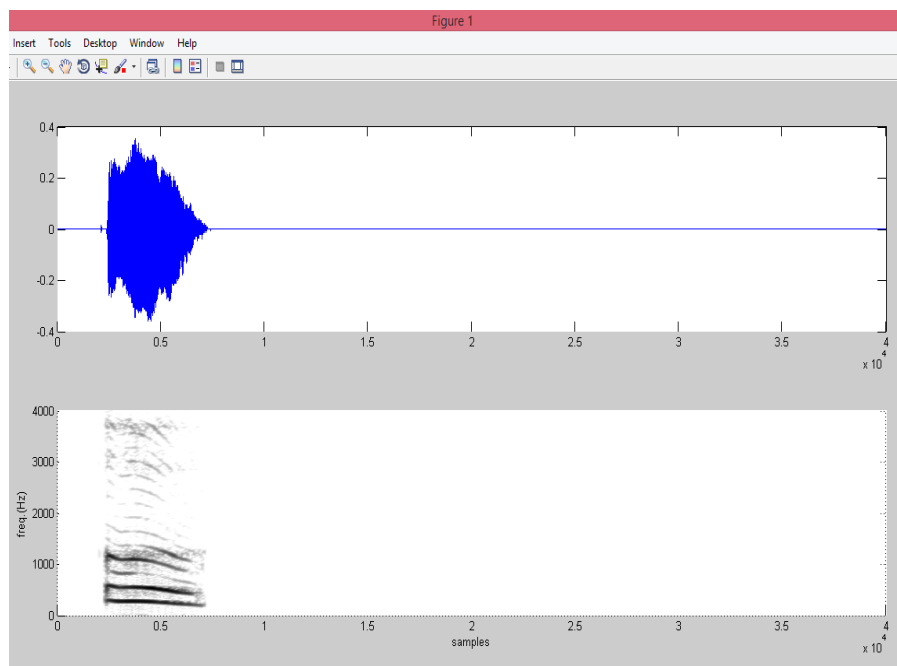


Рисунок 2.20 – Временная и спектральная характеристика казахской фонемы «ы»

На рисунке 2.21 представлена спектральная характеристика казахской фонемы «Ы».

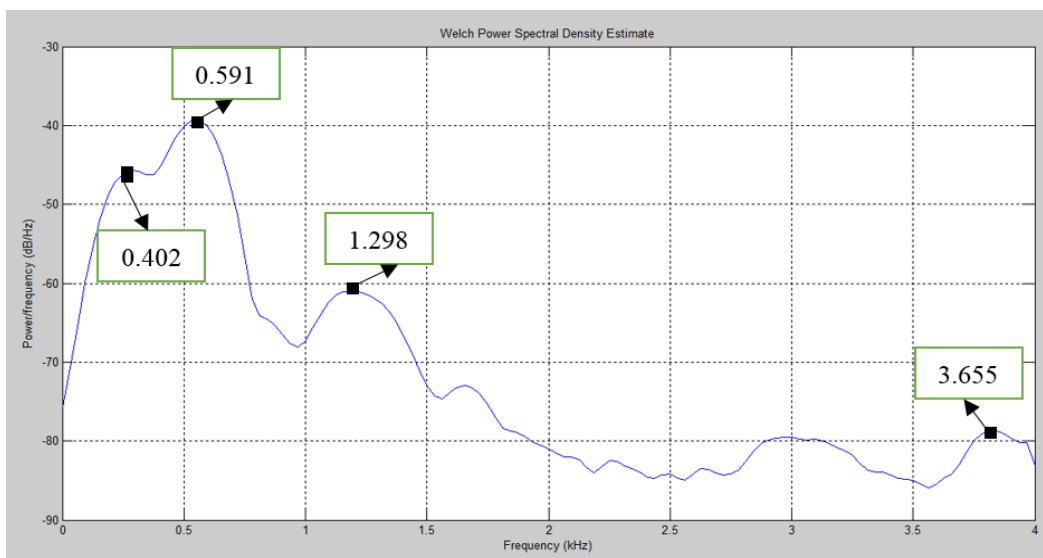


Рисунок 2.21 – Спектральная характеристика казахской фонемы «Ы»

Спектральная характеристика фонемы «Ы» показывает следующих формантов: F1 = 402 Гц, F3 = 591 Гц, F4 = 1298 Гц и F4 = 3655 Гц (рисунок 2.22).

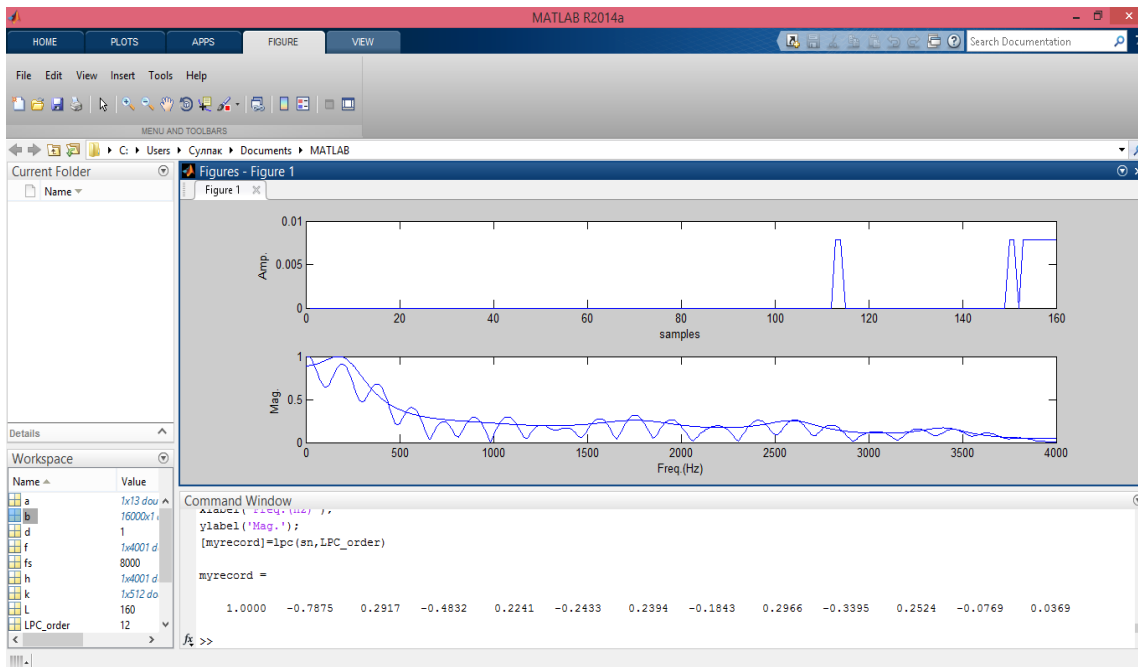


Рисунок 2.22 – Линейные кодовые значения предсказания фонемы «Ы»

б) Казахская фонема «і» представлена на рисунке 2.23.

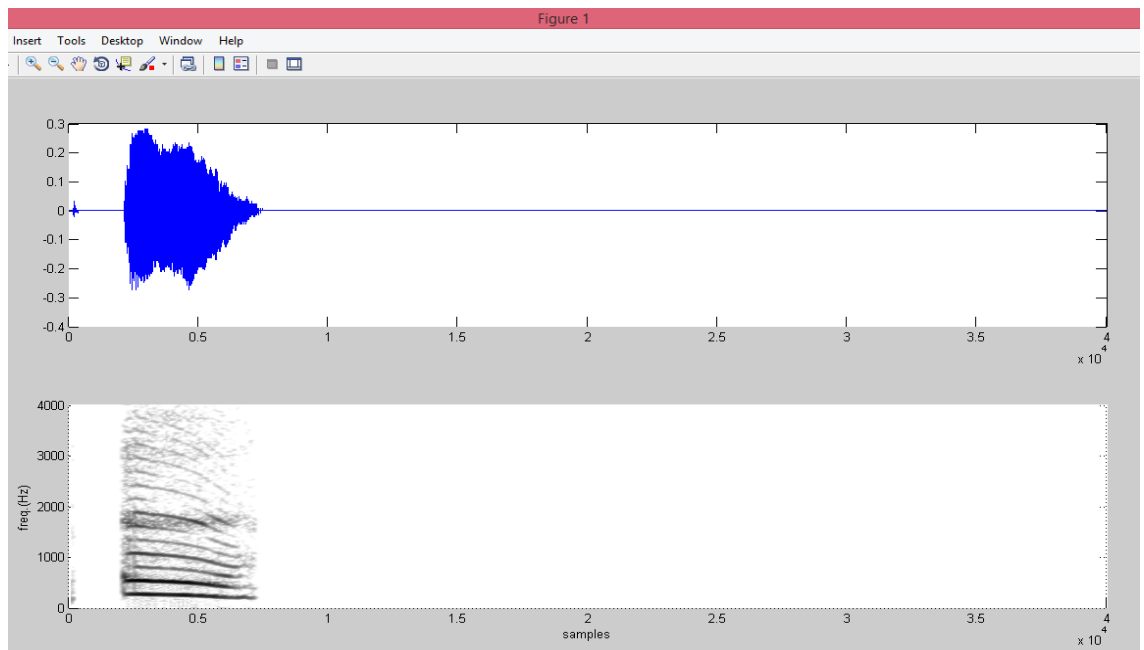


Рисунок 2.23 – Временная и спектральная характеристика казахской фонемы «i»

На рисунке 2.23 показана временная характеристика казахской фонемы «i». На рисунке 2.24 показана спектральная характеристика казахской фонемы «i».

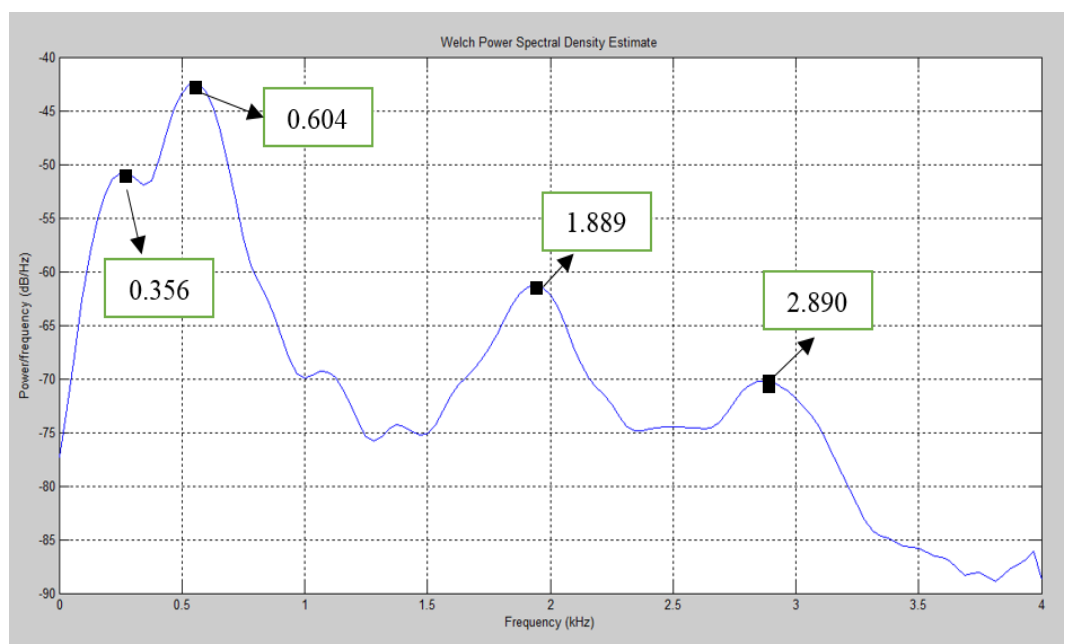


Рисунок 2.24 – Спектральная характеристика казахской фонемы «i»

Формантные частоты казахской фонемы «i»: F1 = 356 Гц, F2 = 604 Гц, F3 = 1888 Гц и F4 = 2890 Гц.



На рисунке 2.27 представлена спектральная характеристика казахской фонемы «ү».

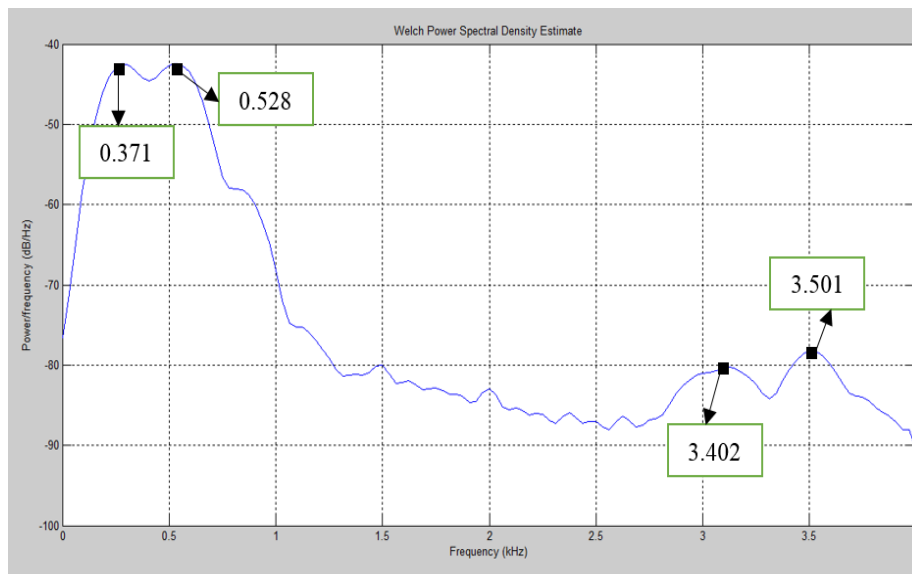


Рисунок 2.27 – Спектральная характеристика казахской фонемы «ү»

Форманты фонемы «ү»: F1 = 371 Гц, F2 = 528 Гц, F3 = 3402 Гц и F4 = 3501 Гц (рисунок 2.28).

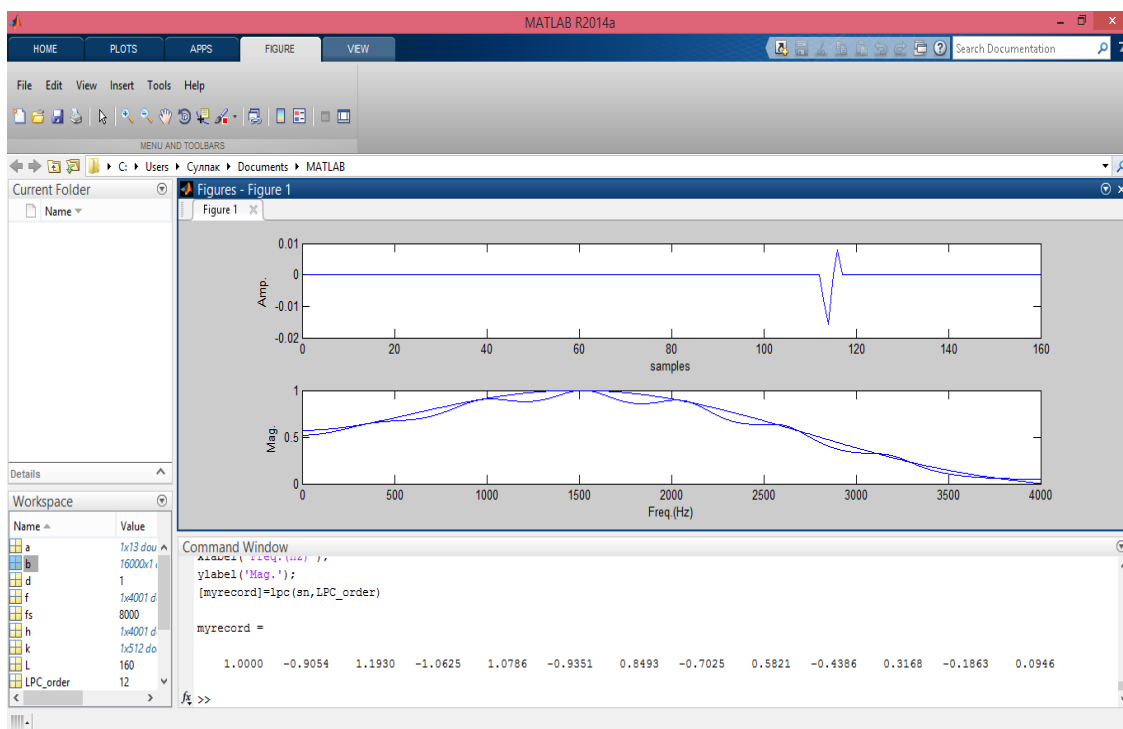


Рисунок 2.28 – Линейные кодовые значения предсказания фонемы «ү»

8) Казахская фонема «ү» представлена на рисунке 2.29.

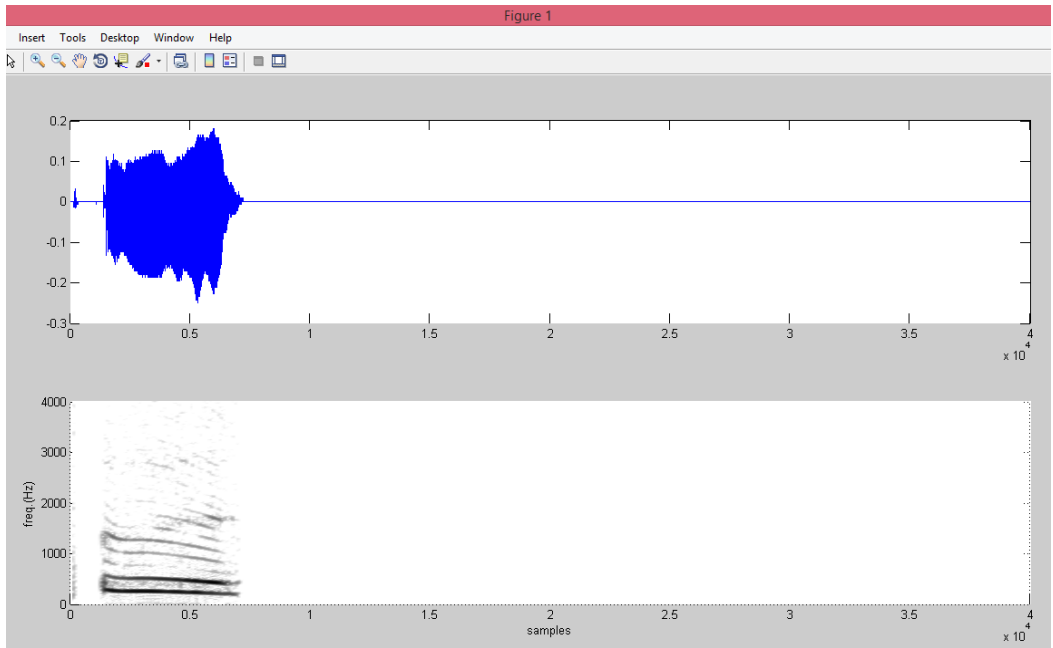


Рисунок 2.29 – Временная и спектральная характеристика казахской фонемы «ү»

На рисунке 2.30 представлен спектральный анализ фонемы.

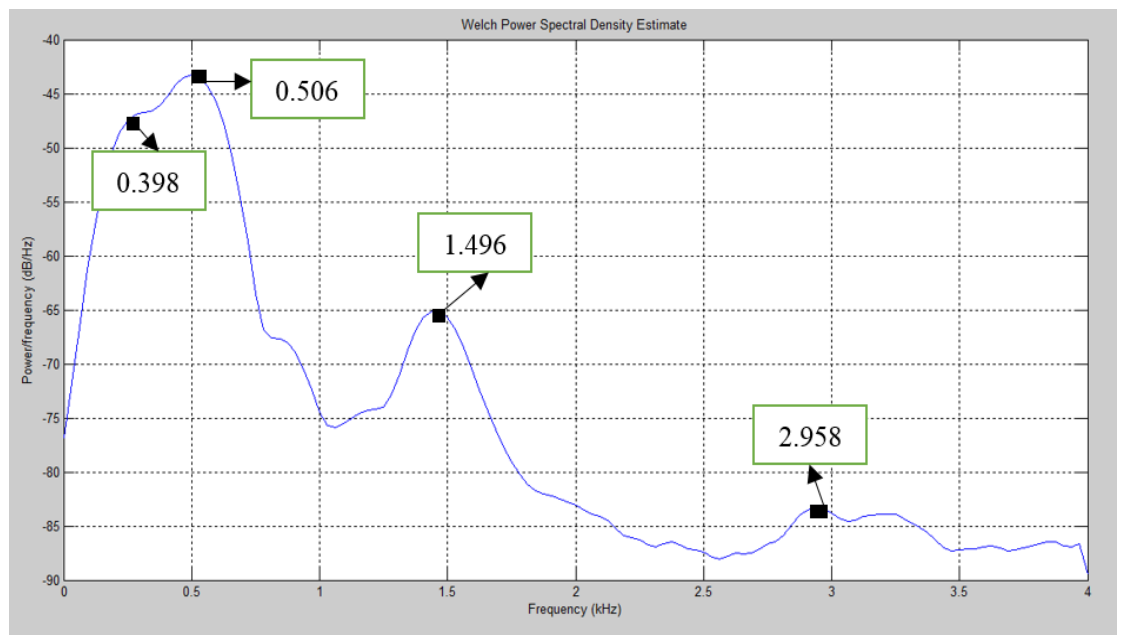


Рисунок 2.30 – Спектральная характеристика казахской фонемы «ү»

Форманты фонемы «ү»: F1 = 398 Гц, F2 = 506 Гц, F3 = 1496 Гц и F4= 2958 Гц (рисунок 2.31).



Рисунок 2.31 – Линейные кодовые значения предсказания фонемы «у»

8) Казахская фонема «э» представлена на рисунке 2.32.

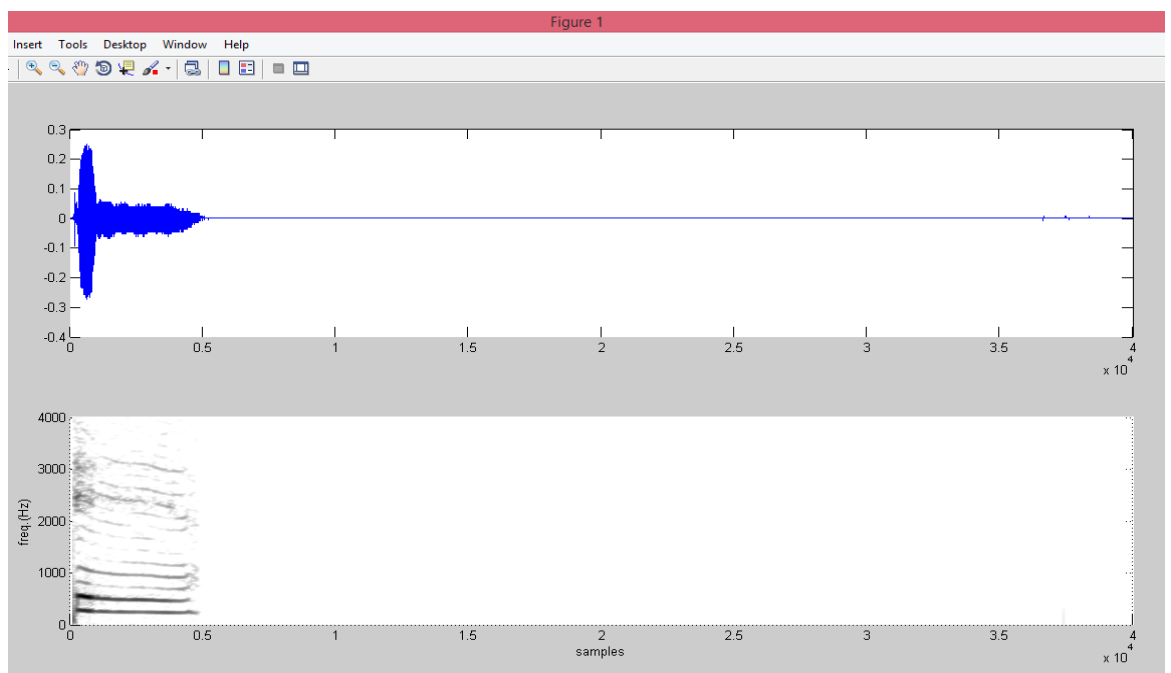


Рисунок 2.32 – Временная и спектральная характеристика казахской фонемы

На рисунке 2.32 показаны тайм-слоты фонемы «э» до 500 Гц.

На рисунке 2.33 приведена спектральная характеристика казахской фонемы «э».

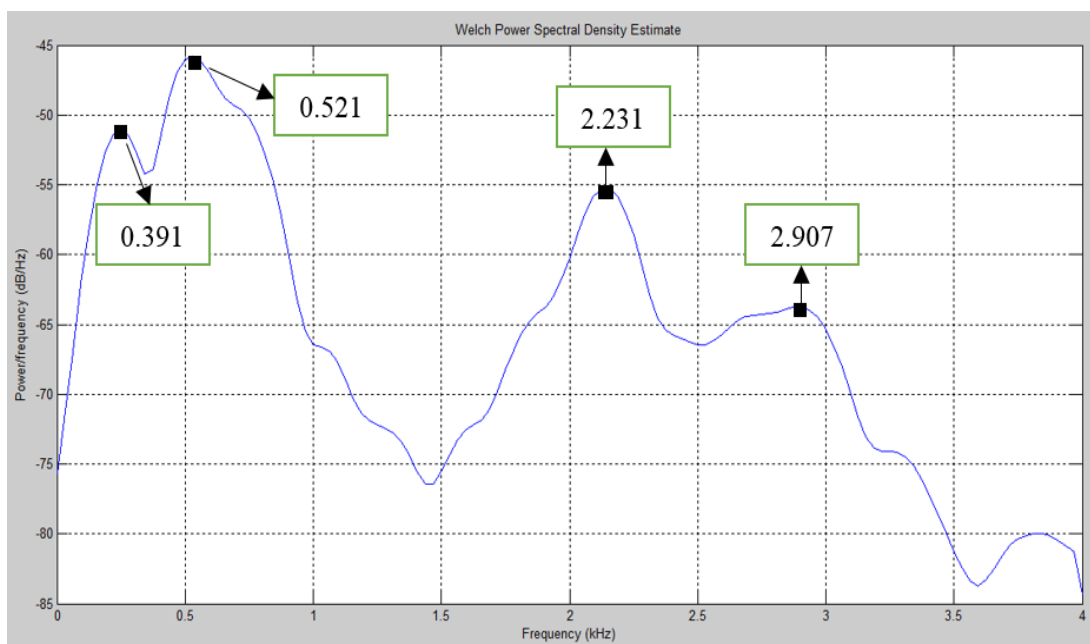


Рисунок 2.33 – Спектральная характеристика казахской фонемы «э»

Спектральная характеристика фонемы «э» показывает формантов  $F1 = 391$  Гц,  $F2 = 521$  Гц,  $F3 = 2231$  Гц и  $F4 = 2907$  Гц.



Рисунок 2.34 – Линейные кодовые значения предсказания фонемы «э»



10) Казахская фонемы «е» представлена на рисунке 2.35

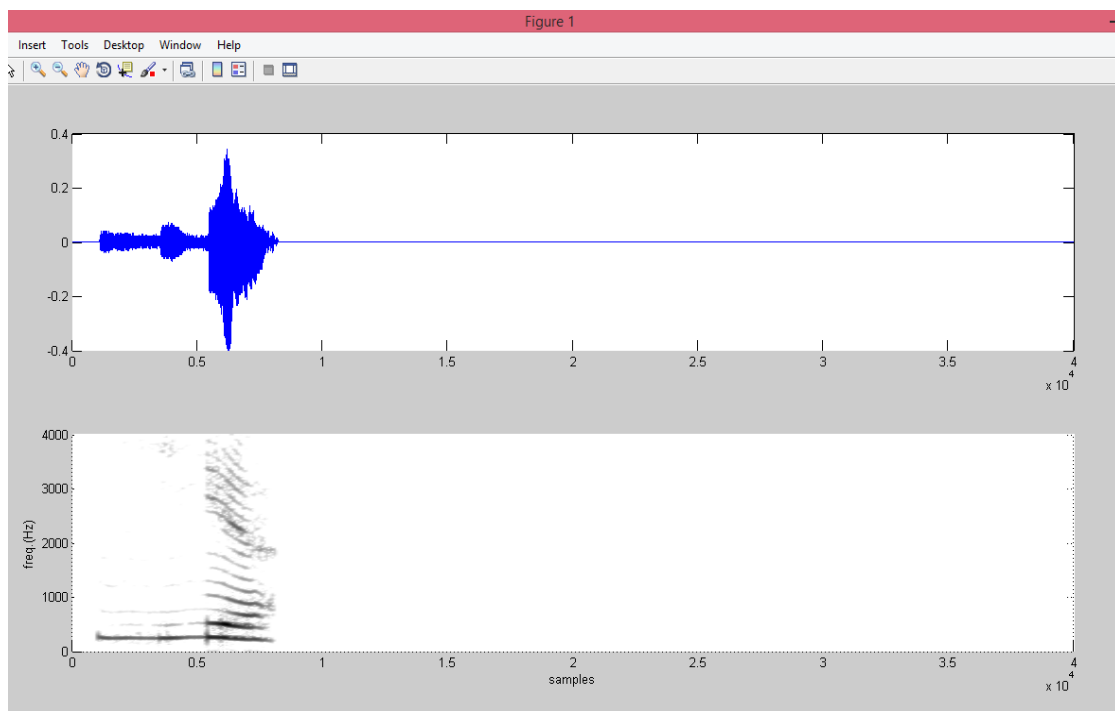


Рисунок 2.36 – Временная и спектральная характеристика казахской фонемы «е»

На рисунке 2.37 показаны тайм-слоты казахской фонемы «е».

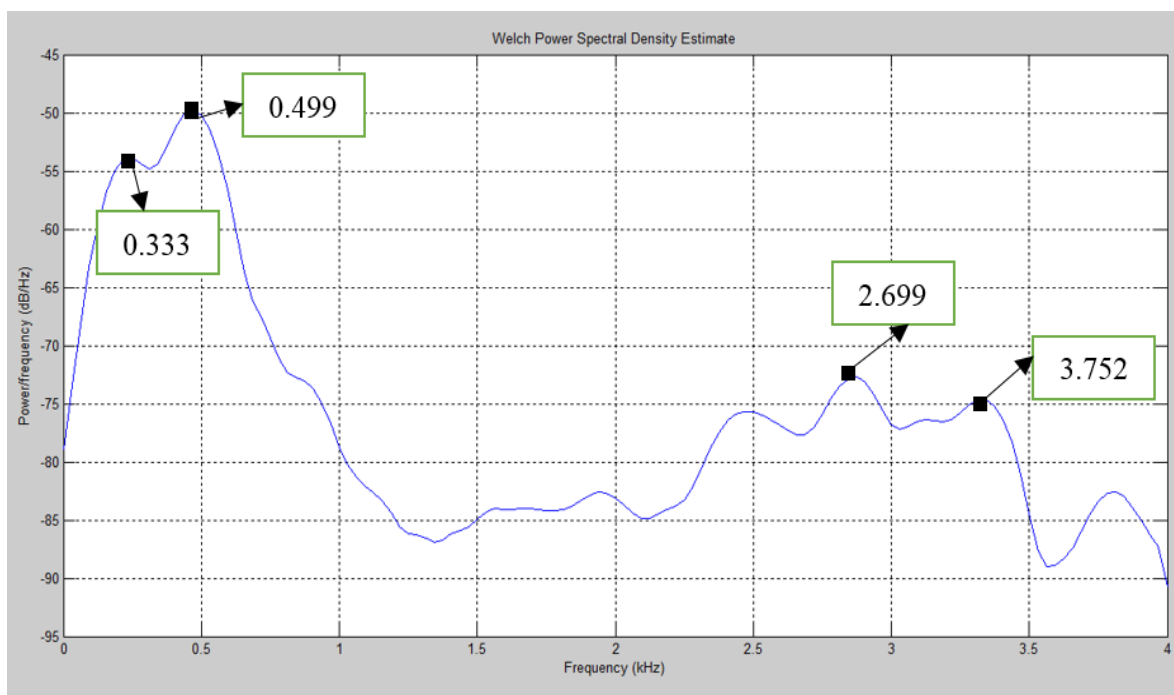


Рисунок 2.38 – Спектральная характеристика казахской фонемы «е»

Спектральная характеристика фонемы «е» показывает следующих формантов  $F1 = 333$  Гц,  $F2 = 499$  Гц,  $F3 = 2694$  Гц и  $F4 = 3752$  Гц (рисунок 2.39).

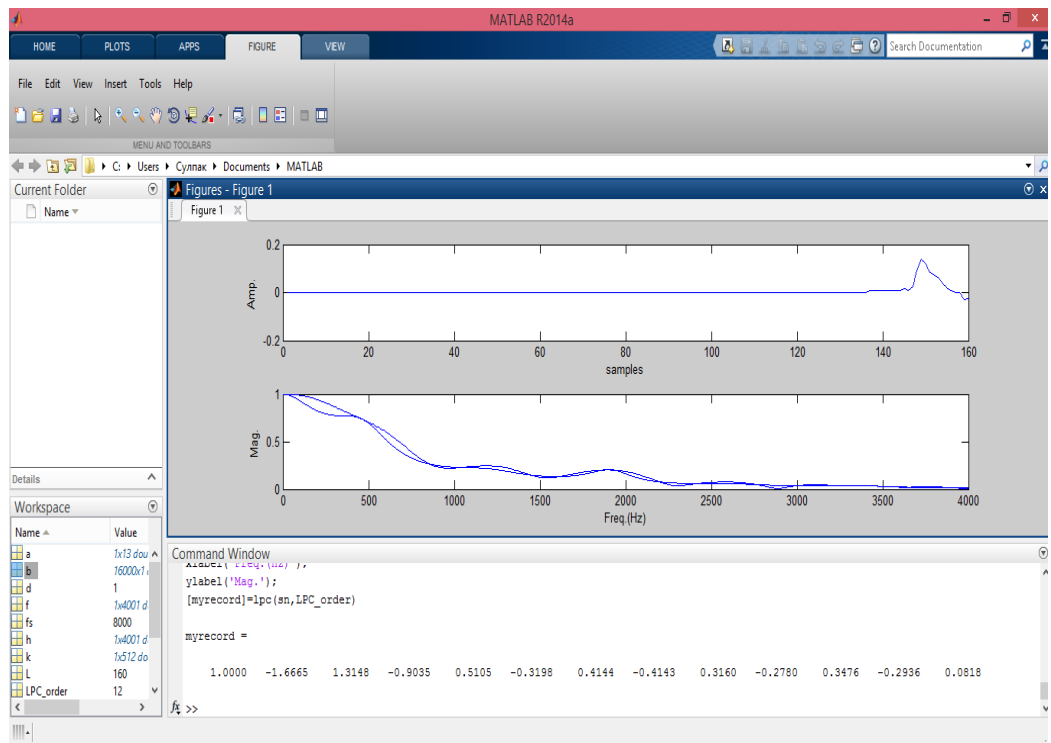


Рисунок 2.39 – Линейные кодовые значения предсказания фонемы «е»  
 11) Казахская фонема «у» представлена на рисунке 2.40.

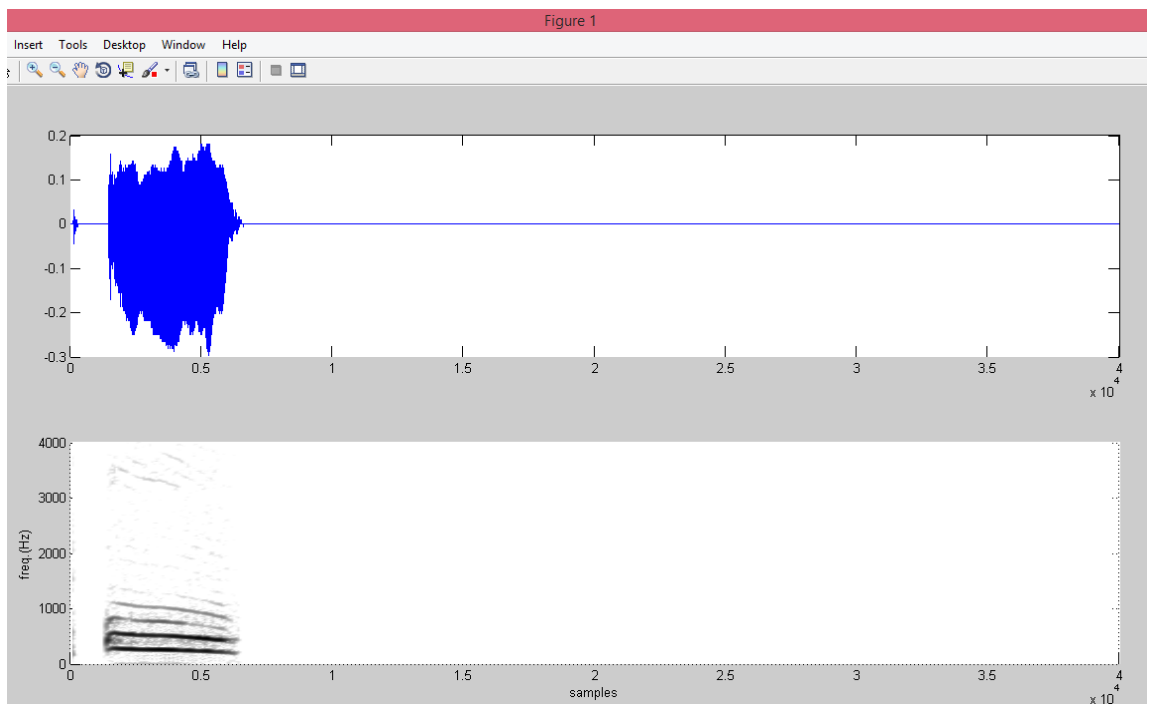


Рисунок 2.40 – Временная и спектральная характеристика казахской фонемы «у»

На рисунке 2.41 показана формантная (спектральная) характеристика казахской фонемы «у».

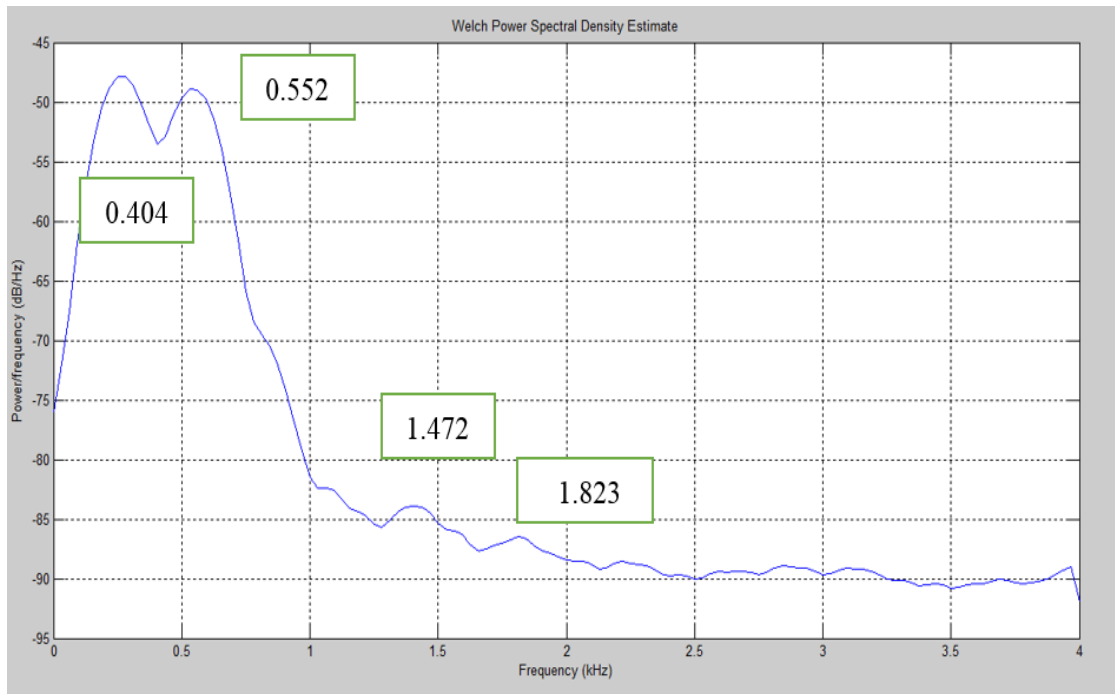


Рисунок 2.41 – Спектральная характеристика казахской фонемы «у»

Спектральная характеристика фонемы «у» показывает следующих формантов  $F1 = 404$  Гц,  $F2 = 552$  Гц,  $F3 = 1472$  Гц и  $F4 = 1823$  Гц.

12) Казахская фонема «и» представлена на рисунке 2.42.

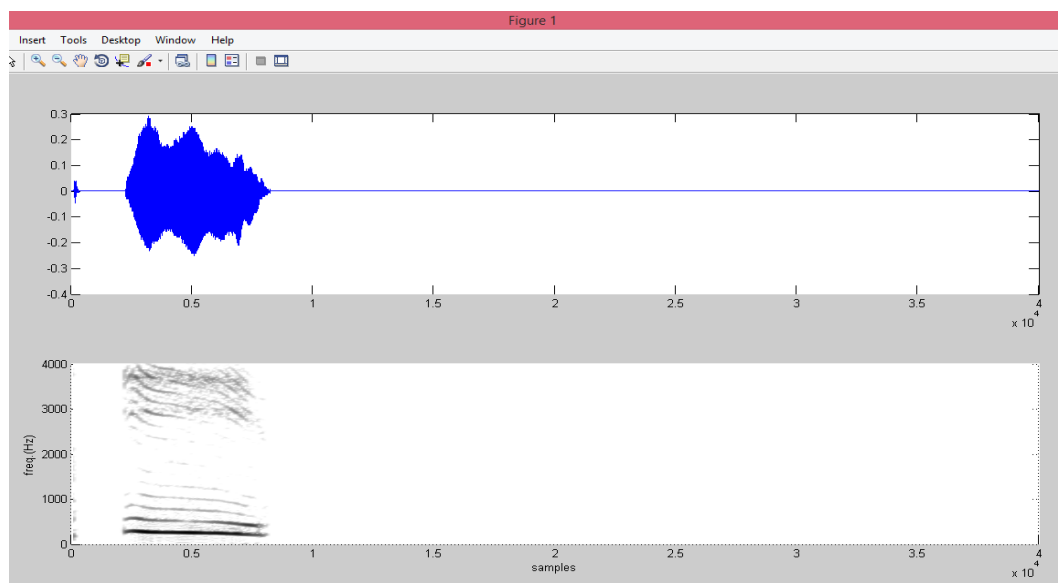


Рисунок 2.42 – Временная и спектральная характеристика казахской фонемы «и»

На рисунке 2.432 показаны временные слоты казахской фонемы «и». А спектральная характеристика показана на рисунке 2.43.

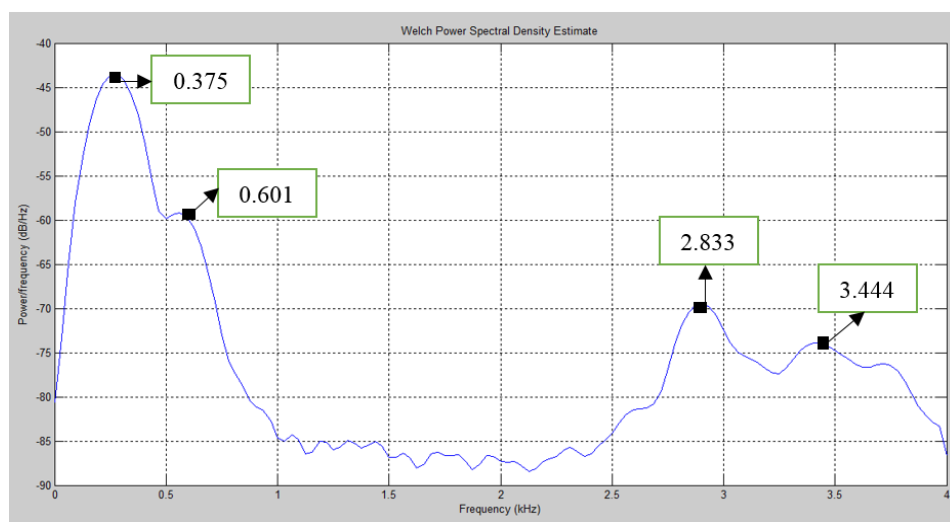


Рисунок 2.43 – Спектральная характеристика казахской фонемы «и»

Спектральная характеристика фонемы «И» показывает следующих формантных частот  $F1 = 375$  Гц,  $F2 = 601$  Гц,  $F3 = 2833$  Гц и  $F4 = 3444$  Гц.

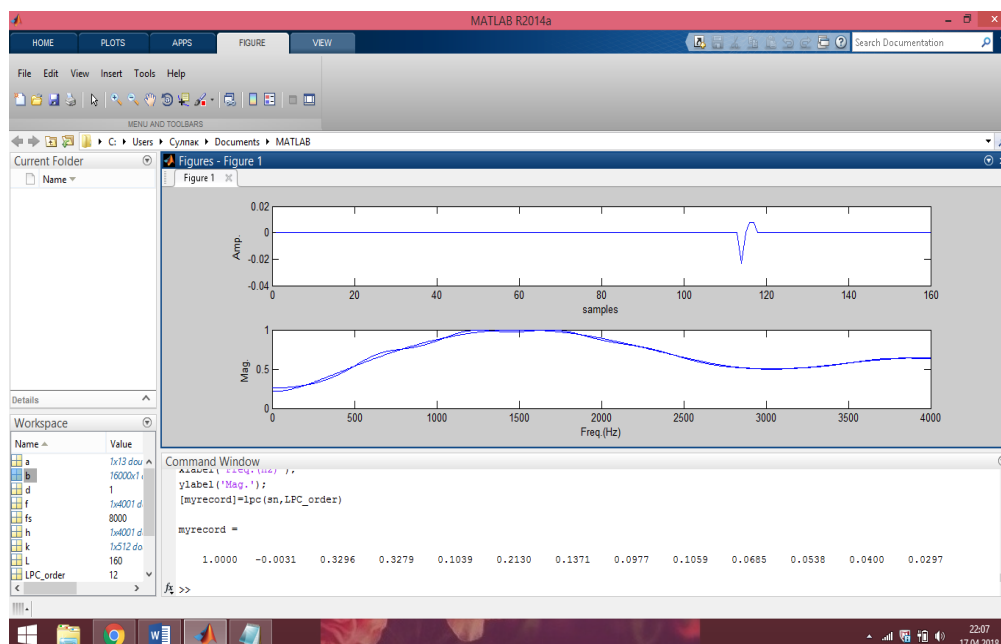


Рисунок 2.44 – Линейные кодовые значения предсказания фонемы «и»

Таким образом, в фонетике формант может означать либо резонанс, либо спектральный максимум, создаваемый резонансом. Форманты часто измеряются как амплитудные пики в частотном спектре звука, используя

спектрограмму или анализатор спектра, а в случае голоса это дает оценку резонансов вокального тракта.

В таблице 4 приведены формантные частоты всех гласных казахских фонем для анализа их произношения.

Таблица 4 – Формантные частоты гласных казахских фонем

| Фонемы | F1, Гц | F2, Гц | F3, Гц | F4, Гц |
|--------|--------|--------|--------|--------|
| «а»    | 291,1  | 1100   | 2757   | 3821   |
| «ә»    | 267    | 782    | 1667   | 2325   |
| «о»    | 311    | 588    | 3378   | 3779   |
| «ө»    | 350    | 501    | 1935   | 2705   |
| «ы»    | 402    | 591    | 1298   | 3753   |
| «і»    | 356    | 604    | 1888   | 2890   |
| «ұ»    | 371    | 528    | 3402   | 3501   |
| «ү»    | 398    | 506    | 1496   | 2958   |
| «э»    | 391    | 521    | 2231   | 2907   |
| «е»    | 333    | 499    | 2694   | 3752   |
| «у»    | 404    | 552    | 1472   | 1825   |
| «и»    | 375    | 601    | 2823   | 3444   |

Известно, что первые два форманта важны для определения качества гласных. Таким образом, первый формант F1 имеет более высокую частоту для открытого гласного [а] и более низкую частоту гласных [ә], [о], [ө], [ұ]; и второй формант F2 имеет более высокую частоту твердого и сжатого гласного [у] и более низкую частоту для гласных [о] и [ө]). При применении к сигналу с изменяющимися во времени резонансами и вокального тракта подход спектральной фильтрации обеспечивает четкое спектрографическое представление формантного изменения во время хода сигнала и облегчает отслеживание частот формантов для дальнейшего анализа

Формантные значения берутся из пиков спектральной характеристики. В нашей работе показано, как оценивать частоты формальных гласных с использованием линейного предсказательного кодирования (LPC) для английского языка. Обычно, формантные частоты получаются путем нахождения корней полинома прогнозирования.

### 3 Расчет опорных параметров речевого сигнала

#### 3.1 Расчет временной и спектральной характеристики казахских фонем

Мел-частотный кепстр – это мощный инструмент анализа звука, который используется в распознавании речи, обработке аудиосигналов и других областях. Он представляет собой кратковременный спектр мощности речевого сигнала, преобразованный на мел-шкалу частот.

Суть метода MFCC заключается в том, чтобы анализировать уникальные особенности речевого сигнала - обертоны и дополнительные частоты, которые содержатся в звуке. Используя нелинейное преобразование шкалы частот в шкалу мел, учитываются свойства слуха человека. Речь идет о критических полосах, в которых сигналы одинаковой частоты воспринимаются одинаково. Этот подход позволяет снизить избыточность информации в спектре сигнала и выделить основные характеристики речи, делая ее более удобной для анализа и обработки компьютерными алгоритмами [49, с.11].

MFCC учитывают нелинейные особенности восприятия звука человеческим ухом. Преобразование в мел-частотную шкалу позволяет снизить размерность признакового пространства, выделить основные акустические характеристики звукового сигнала и сгладить шумы.

Шкала мел вычисляется по формуле (26).

$$M(f) = 2595 \cdot \ln(1 + f/700) \quad (3.1)$$

где  $f$  - частота в Гц;

$M$  – частота в мелах.

Рассчитаем мелы для всех 12 фонем казахского алфавита.

В качестве частоты Гц для каждой фонемы примем общую частоту

$F_{\text{общ}} = F_1 + F_2 + F_3 + F_4$  (таблица 5).

Таблица 5 – Итоговая частота фонем, Гц

| Фонемы | F, Гц  |
|--------|--------|
| «а»    | 7969,1 |
| «э»    | 5041   |
| «о»    | 8056   |
| «ө»    | 5491   |
| «ы»    | 6044   |
| «і»    | 5738   |
| «ү»    | 7802   |
| «ү»    | 5358   |
| «э»    | 6050   |
| «е»    | 7278   |
| «у»    | 4253   |
| «и»    | 7243   |

1. Определим шкалу мел для казахской гласной буквы «а»:

$$M(f) = 2595 \cdot \ln(1 + 7969,1 / 700) = 6512 \text{ мел.}$$

2. Определим шкалу мел для казахской гласной буквы «ә»:

$$M(f) = 2595 \cdot \ln(1 + 5041 / 700) = 5460 \text{ мел.}$$

3. Определим шкалу мел для казахской гласной буквы «о»:

$$M(f) = 2595 \cdot \ln(1 + 8056 / 700) = 6554 \text{ мел.}$$

4. Определим шкалу мел для казахской гласной буквы «ө»:

$$M(f) = 2595 \cdot \ln(1 + 5491 / 700) = 5643 \text{ мел.}$$

5. Определим шкалу мел для казахской гласной буквы «ы»:

$$M(f) = 2595 \cdot \ln(1 + 6044 / 700) = 5869 \text{ мел.}$$

6. Определим шкалу мел для казахской гласной буквы «і»:

$$M(f) = 2595 \cdot \ln(1 + 5738 / 700) = 5758 \text{ мел.}$$

7. Определим шкалу мел для казахской гласной буквы «ү»:

$$M(f) = 2595 \cdot \ln(1 + 7802 / 700) = 6448 \text{ мел.}$$

8. Определим шкалу мел для казахской гласной буквы «ұ»:

$$M(f) = 2595 \cdot \ln(1 + 5358 / 700) = 5599 \text{ мел.}$$

9. Определим шкалу мел для казахской гласной буквы «э»:

$$M(f) = 2595 \cdot \ln(1 + 6050 / 700) = 5869 \text{ мел.}$$

10. Определим шкалу мел для казахской гласной буквы «е»:

$$M(f) = 2595 \cdot \ln(1 + 7278 / 700) = 6315 \text{ мел.}$$

11. Определим шкалу мел для казахской гласной буквы «у»:

$$M(f) = 2595 \cdot \ln(1 + 4253 / 700) = 5075 \text{ мел.}$$

12. Определим шкалу мел для казахской гласной буквы «и»:

$$M(f) = 2595 \cdot \ln(1 + 7243/700) = 6292 \text{ мел.}$$

### 3.2 Расчет оценки энергетического спектра речевого сигнала

Для расчета энергетического спектра речи воспользуемся оконной функцией Хэмминга, приведенная в формуле (3.2)

$$w(n) = 0,54 * 0,46 \cos\left(\frac{2\pi n}{N}\right) \quad (3.2)$$

где  $n$  – шаг отсчета;

$N$  – количество выборок сигнала.

Определим оценку энергетического спектра для казахской гласной буквы «ө»:

$$w(n) = 0,54 * 0,46 \cos\left(\frac{2 \cdot 3,14 \cdot 0,35}{1024}\right) = 0.24$$

Определим оценку энергетического спектра для казахской гласной буквы «ә»:

$$w(n) = 0,54 * 0,46 \cos\left(\frac{2 \cdot 3,14 \cdot 0,28}{1024}\right) = 0.25$$

Определим оценку энергетического спектра для казахской гласной буквы «і»:

$$w(n) = 0,54 * 0,46 \cos\left(\frac{2 \cdot 3,14 \cdot 0,256}{1024}\right) = 0.25$$

Определим оценку энергетического спектра для казахской гласной буквы «э»:

$$w(n) = 0,54 * 0,46 \cos\left(\frac{2 \cdot 3,14 \cdot 0,38}{1024}\right) = 0.248$$

### 3.3 Расчет равномерного распределения громкости речевого сигнала

Речевой сигнал производит неравное восприятие громкости на разных частотах. Для этого выполняется условие с использованием кривой равной громкости, заданной формулой (3.3).

$$E(\omega) = \frac{\omega^2 + 56,8 \cdot 10^6 \omega^4}{(\omega^2 + 6,3 \cdot 10^6)(\omega^2 + 0,38 \cdot 10^9)(\omega^6 + 9,58 \cdot 10^{26})} \quad (3.3)$$



где  $\omega$  —угловая линейная частота равная  $2\pi f$ , Гц .

Определим равномерное распределение громкости для казахской гласной буквы «ү»:

$$E(\omega) = \frac{((2 \cdot 3,14 \cdot 10000)^2 + 56,8 \cdot 10^6) 2 \cdot 3,14^4}{((2 \cdot 3,14 \cdot 10000)^2 + 6,3 \cdot 10^6)((2 \cdot 3,14 \cdot 10000)^2 + 0,38 \cdot 10^9)((2 \cdot 3,14 \cdot 10000)^6 + 9,58 \cdot 10^{26})} = 0,07$$

Для учета неравного восприятия громкости на разных частотах мы можем использовать кривую равной громкости, которая обычно выражается с использованием формулы Звукового давления (SPL) в децибелах, связанного с частотой. Одной из таких кривых равной громкости является кривая А-частотной характеристики, измеренная в фон [51, с.23].

Формула А-частотной характеристики представлена как (3.4):

$$L_A(f) = 20 \log_{10}(f) + 94 \quad (3.4)$$

где:  $L_A(f)$  - уровень звука в децибелах (dB) на частоте (f) Гц,

(f) - частота звука в Гц.

Эта формула представляет кривую равной громкости для человеческого слуха. Определим характеристику распределения речевого сигнала для трех казахских фонем. Возьмем для примера: э, ө, і.

1. Э:  $L_A(f) = 20 \log_{10}(f) + 94 = 142,52$  дБ, где  $f_1 = 267$  Гц.

2. Ө:  $L_A(f) = 20 \log_{10}(f) + 94 = 144,88$  дБ, где  $f_1 = 350$  Гц.

3. І:  $L_A(f) = 20 \log_{10}(f) + 94 = 145,02$  дБ, где  $f_1 = 356$  Гц..

## ЗАКЛЮЧЕНИЕ

В рамках данной дипломной работы были рассмотрены основные понятия биометрических параметров речевого сигнала. Биометрические особенности речевых сигналов представляют собой уникальные характеристики голоса человека. Эти особенности могут быть использованы для идентификации или верификации личности. Речевые сигналы содержат информацию о различных аспектах голоса (тональность, интонация, скорость произношения, амплитуда и частотные характеристики). Данные параметры могут быть анализированы и использованы в биометрических системах.

Одним из основных преимуществ использования речевых биометрических особенностей является их уникальность и стабильность. Голос каждого человека уникален. Даже с течением времени, с учетом возрастных особенностей и состояния здоровья, биометрические особенности голоса остаются относительно постоянными.

Биометрические особенности речевых сигналов также обладают высокой удобством использования. Они могут быть собраны незаметно для пользователя в реальном времени, например, при разговоре по телефону или при использовании голосовых устройств аутентификации.

Кроме того, биометрические особенности речевых сигналов могут быть успешно комбинированы с другими биометрическими методами.

В дипломной работе представлен анализ аудиозаписей гласных фонем казахских языков с использованием программ, разработанных в среде MATLAB. Основное внимание уделено получению линейных кодовых значений и определению формантных спектральных характеристик.

В ходе исследования проведен сравнительный анализ формантных характеристик гласных звуков казахского языка. Выявлено, что для казахского языка характерны отличия в частотах первого и второго формантов в зависимости от типа гласного звука.

Также был проведен расчет опорных параметров речевого сигнала. Для этого использовались частотные характеристики на основе мел-кепстральных коэффициентов и оценка энергетического спектра речевых фонем с учетом равномерного распределения громкости.

## Перечень принятых сокращений, терминов

РРЭ - Распознавание речевых эмоций  
ЭКГ - Электрокардиография  
LTV- модель представления речи  
AMDF- функция разности средних значений  
БПФ (FFT )- быстрое преобразование Фурье  
DFT- дискретное преобразование Фурье  
FFT- преобразование Фурье  
IFT-обратное преобразование Фурье  
IDFT-обратного дискретное преобразование Фурье  
MFCC- Мел-частотный кепстр  
LSP- линейная спектральная пара  
ML - машинное обучение «machine learning»  
DNNS (Deep neural networks)  
SVMs - Машина опорных векторов  
HMM-ASR-модули – искусственный интеллект и машинное обучение  
ASR-системы Hnms - Скрытая Марковская модель  
CD - стереокомпакт-диск  
PCM - импульсно-кодовая модуляция  
MATLAB – язык программирования и пакет прикладных программ

## СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

- 1 Морозов В.П. Невербальная коммуникация. Экспериментально-психологические исследования. М: Институт психологии РАН 2011, 250 с.
- 2 Гуртуева И.А., Бжихатлов К.Сh. Аналитический обзор и классификация методов выделения признаков акустических сигналов в речевых системах. Известия Кабардино-Балкарского научного центра РАН. 2022. № 1, 105 с.
- 3 Сорокин В. Н., Ижнин А. Н., Цыплихин А. К, Чепелев Д. Н. Артикуляторноориентированная система распознавания речи // Труды Международного семинара «Диалог-2022», 2022, стр. 657-662.
- 4 Sorokin V.N., Leonov A.S., Makarov I.S., Tsyplikhin A.I. Speech inversion and resynthesis // Proceedings of InterSpeech-2015, Lisbon, 2005, P. 3209-3212.
- 5 Цыплихин А. К Импульсы голосового источника: гетерархический алгоритм синхронизации TWIN // Труды XVI сессии РАО, 2015, Т. 3, С. 36-40.
- 6 Леонов А.С., Макаров КС., Сорокин В.Н., Цыплихин А.И. Артикуляторный ресинтез гласных // Информационные процессы, 2018, Т. 3, № 2, С. 73-82.
- 7 Репин В. Г., Цыплихин А. И. Определение точной верхней грани ошибок метода наименьших квадратов // Радиотехника и электроника, 2017, Т. 48, № 1, С. 9199.
- 8 Леонов А.С., Макаров КС., Сорокин В.Н., Цыплихин А.К Артикуляторный ресинтез фрикативных // Информационные процессы, 2014, Т. 4, № 2, С. 141-159.
- 9 Сорокин В. Н., Цыплихин А. К. Сегментация и распознавание гласных // Информационные процессы, 2014, Т. 4, № 2, С. 202-220.
- 10 Идентификация дикторов на основе сравнения статистик основного тона голоса / С.Л. Коваль, П.В. Лабутин, Е.В. Малая и др. // Информатизация и информационная безопасность правоохранительных органов. Тр. XV Межд.науч. конф. М.: Академия управления МВД России, 2006. С. 324–327.
- 11 Аграновский А. В. Теоретические аспекты алгоритмов обработки и классификации сигналов / А. В. Аграновский, Д. Л. Леднов. М. : Радио и связь, 2004. 164 с.
- 12 Леонов А.С., Макаров КС., Сорокин В.Н., Цыплихин А.К Кодовая книга для речевых обратных задач // Информационные процессы, 2015, Т. 5, № 2, С. 101-119.
- 13 Цыплихин А. И., Сорокин В. Н. Сегментация речи на кардинальные элементы // Информационные процессы, 2016, Т. 6, № 3, С. 177-207.
- 14 Воронцов К. В. Комбинаторные обоснования обучаемых алгоритмов // ЖВ-МиМФ, 2014, Т. 44, № 11, С. 2099-2112.
- 15 Макаров К. С. Построение и исследование артикуляторных кодовых книг для решения речевых обратных задач. Диссер. на соиск. степ, к.т.н., ИППИ РАН, 2015. 500 с.
- 16 Покровский Н. Б. Расчет и измерение разборчивости речи. М.: Связь издат, 2017, 250 с.

- 17 Столбов М.Б., Основы анализа и обработки речевых сигналов – СПб.: НИУ ИТМО, 2021. – 101 с.
- 18 Иванов А. И. Биометрическая идентификация личности по динамике подсознательных движений / А. И. Иванов. Пенза : Изд-во Пенз. гос. ун-та, 2000. 188 с.
- 19 Сорокин В. Н., Выогин В. В., Тананыкин А. А. // Информ. процессы. 2012. Т. 12, № 1. С. 1-30.
- 20 Ралшшвили Г. С. Автоматическое опознавание говорящего по голосу. М.: Радио и связь, 2017.-224 с.
- 21 Смирнов В. И. Курс высшей математики. Том 2. М: Государственное издательствотехнико-теоретической литературы, 2007, 500 с.
- 22 Ковальчук А.Н. Формула для перехода из частотной области в шкалу барка и наоборот. Информатика. 2011. № 4(32). С. 71.
- 23 Рабинер Л.Р., Цифровая обработка речевых сигналов: [пер. с англ.] / Л.Р. Рабинер, Р.В.Шафер. – М.: Радио и связь, 2007. – 251 с.
- 24 Афанасьев А.А. Модель системы предобработки речевого сигнала на основе накопления базы биспектров сигналов речи / А.А. Афанасьев, Д.Д. Сергеев, О.Н. Титов // Информационные системы и технологии. - 2018.- №2. - С.56-67.
- 25 A.R.Jayan., Speech and Audio Signal processing,Pte.Ltd.,India, 2016, с.412.
- 26 Руководство по биометрии / Р. М. Болл, Дж. Х. Коннел, Ш. Пан- канти и др. М. : Техносфера, 2007. 368 с.
- 27 Галушкин А. Нейрокомпьютеры в биометрических системах. Кн. 26. / А. И. Галушкин. М. : Радиотехника, 2007. 192 с.
- 28 Павлов, А. Н. Методы анализа сложных сигналов / А. Н. Павлов. Саратов : Науч. кн., 2008. 120 с.
- 29 Лобанов, Б. М. Компьютерный синтез и клонирование речи / Б. М. Лобанов, Л. И. Цирульник. Минск : Белорус, наука, 2008. 316 с.
- 30 Афанасьев А.А. Система обработки речевого сигнала на основе модернизации метода прямого предсказания / Афанасьев, А.А. // Проектирование и технология электронных средств. - 2017.- №3.- С.60
- 31 Кобзарь,А. И. Прикладная математическая статистика. Для инженеров и научных работников / А. И. Кобзарь. М. : Физматлит, 2006. 816 с
- 32 Иванов А. Высоконадежная биометрическая аутентификация пользователя: последний дюйм первой мили / А. Иванов, А. Малыгин // Первая миля. 2007. № 2. С. 20-24.
- 33 Ахмад, Х. М. Введение в цифровую обработку речевых сигналов : учеб. пособие / Х. М. Ахмад, В. Ф. Жирков ; Владим. гос. ун-т. – Владимир:Изд-во Владим. гос. ун-та, 2007 – 192 с.
- 34 Болл, Р.М. Руководство по биометрии / Р.М. Болл. - М.: Техносфера, 2018. - 273 с.
- 35 Гучай, А. Н. Текстозависимая верификация диктора / А. Н. Гучай. Саарбрюккен : БАГ LAMBERT Academic Publishing, 2012. 144 с.

- 36 Киселев, А. Н. Оценка параметров диеторонезависимых признаков фонем с применением адаптивного частотного временного анализа : дис. ... канд. техн. наук / А. Н. Киселев. Тула : Тульский гос. ун-т, 2005. 189 с.
- 37 Исследование алгоритмов обработки сигналов в системе MATLAB: метод, указания к лабораторным работам / Владим. гос. ун-т; сост. Е.К. Левин. - Владимир: Изд-во Владим. гос. ун-та, 2011. - 78 с.
- 38 Распознавание речи. Выделение человеческого голоса среди бытовых шумов: Курсовая работа \ А.В. Буторина; Моск. Гор. Псих.-Педагог. ун-т. 2007. - 43 с.
- 39 Ануфриев И. Самоучитель MatLab 5.3/6.x; БХВ-Петербург - Москва, 2004. - 736 с.
- 40 Гринев А. Ю., Ильин Е. В. Основы электродинамики с MATLAB; Логос - Москва, 2013. - 176 с
- 41 Сергиенко, А. Б. Цифровая обработка сигнала / А. Б. Сергиенко. СПб. : Питер, 2003. 604 с..
- 42 Алексеев Е. Р., Чеснокова О. В. MATLAB 7; ИТ Пресс - Москва, 2006. - 464 с.
- 43 Есентемирова А.А. О фонетической природе казахского словесного ударения // Вопросы казахской фонетики и фонологии / Отв. ред. Ж.А. Аралбаев. – Алма-Ата: Наука, 1979. С. 183-189.
- 44 Кухарев, Г. А. Биометрические системы. Методы и средства идентификации личности человека / Г.А. Кухарев. - М.: Политехника, 2020. - 240 с.
- 45 Лакин, Г.Ф. Биометрия / Г.Ф. Лакин. - М.: Высшая школа; Издание 3-е, перераб. и доп., 2017. - 293 с.
- 46 Плохинский, Н. А. Биометрия: моногр. / Н.А. Плохинский. - М.: Издательство МГУ, 2020. - 368 с.
- 47 Синай, Я.Г. Динамические системы - 2 / Я.Г. Синай, А.М. Вершик, Р.Л. Добрушин. - М.: [не указано], 2020. - 819 с.
- 48 Генетика и биометрия. - Москва: СПб. [и др.] : Питер, 2016. - 790 с.
- 49 Дворянкин С. В., Кханг. В. В. Голосовые признаки индивидуальной информативности в речевых системах управления доступом // Тезисы докладов научно-технической конференции профессорско-преподавательского состава. - М.: МТУСИ, 2001. С. 11-12.
- 50 Дворянкин С. В. Взаимосвязь цифры и графики, звука и изображения // Открытые системы 3/2000. С. 34-40
- 51 Тропченко А Ю., Тропченко А.А. Цифровая обработка сигналов. Методы предварительной обработки. Учебное пособие по дисциплине «Теоретическая информатика». – СПб.: СПбГУ ИТМО, 2009 – 100 с.

## Приложение А

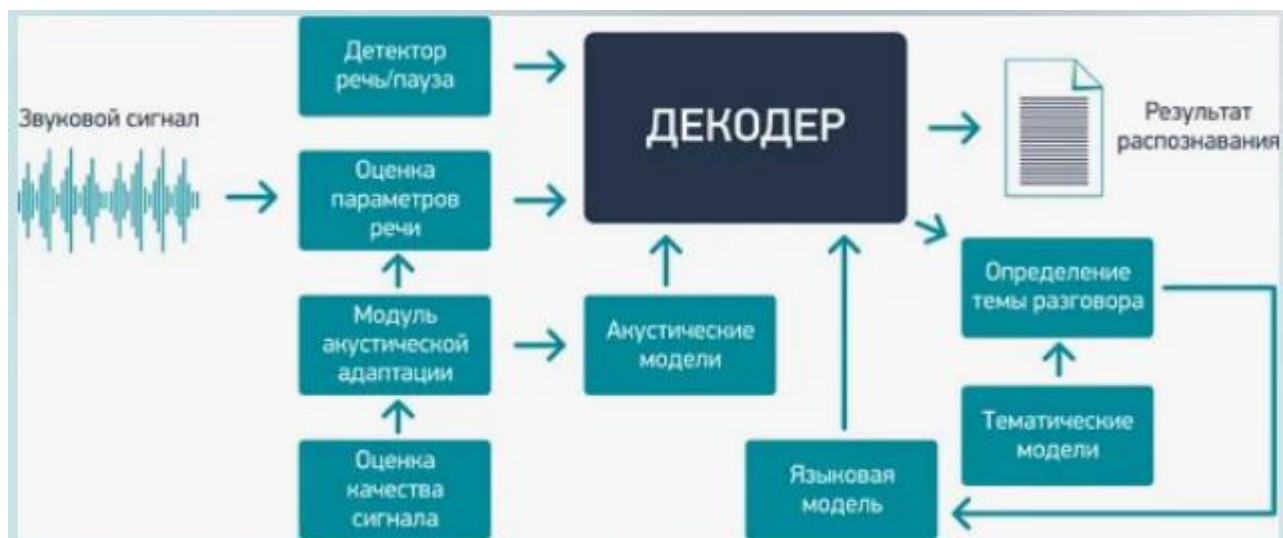


Рисунок А.1 – Алгоритм результата распознавания

## ОТЗЫВ НАУЧНОГО РУКОВОДИТЕЛЯ

Дипломная работа

Абдинаби Мадина Абдуалиевна

6B06201-Телекоммуникация

Тема: «Исследование биометрических особенностей речевых сигналов»

В данной работе исследованы биометрические особенности речевых сигналов. А также сделан анализ биометрических параметров и показателей речевого сигнала. Представлена LTV-модель для представления речи.

В разделе методах и алгоритмах обработки, используемых в анализах речевого сигнала приведены и все методы, а также алгоритмы обработки.

В разделе расчета параметров речевого сигнала все расчеты произведены верно. Моделирование речевого сигнала сделано успешно.

В целом, общие требования к созданию, изложению, оформлению текстовых и графических материалов работы выполнены в соответствии с заданными стандартами.

А дипломная работа отвечает требованиям, предъявляемым к дипломной работе, и заслуживает оценки «отлично» (90%), а дипломант Абдинаби Мадина Абдуалиевна достоин академической степени «бакалавр» техники и технологии по специальности 6B06201-Телекоммуникация.

**Научный руководитель:**

Ст. преподаватель каф. ЭТиКТ

Джунусов Н.А.

10 июня 2024 ж.





МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РЕСПУБЛИКИ КАЗАХСТАН

**РЕЦЕНЗИЯ**

На дипломную работу

Абдинаби Мадина Абдуалиевна

Специальность 6В06201 – Телекоммуникация

На тему: «Исследование биометрических особенностей речевых сигналов»

Выполнено:

- а) теоретическая часть на 36 листах
- б) практическая часть на 27 листах

**ЗАМЕЧАНИЯ К РАБОТЕ**

В работе студента Абдинаби Мадины Абдуалиевна была рассмотрена тема «Исследование биометрических особенностей речевых сигналов». Нет ссылок на литературу по тексту.

**Оценка работы**

Дипломная работа выполнена в соответствии с ГОСТ на оценку /90 /«отлично», а дипломант, Абдинаби Мадина Абдуалиевна достойна степени бакалавра специальности 6В06201 – Телекоммуникация.

**Рецензент**

**к.т.н., профессор**

**АУЭС имени Г. Даукеева**

**Байкенов А.С.**

«23»



|   |            |
|---|------------|
| Қолтаңбаны растаймын<br>Подпись заверяю |            |
| Қызметі                                 | аты-жөні   |
| « 23 »                                  | 05 2024 ж. |

**Университеттің жүйе администраторы мен Академиялық мәселелер департаменті  
директорының ұқсастық есебіне талдау хаттамасы**

Жүйе администраторы мен Академиялық мәселелер департаментінің директоры көрсетілген еңбекке қатысты дайындалған Плагиаттың алдын алу және анықтау жүйесінің толық ұқсастық есебімен танысқанын мәлімдейді:

**Автор: Абдинаби Мадина Абдуалиевна**

**Тақырыбы: Исследование биометрических особенностей речевых сигналов**

**Жетекшісі: Нуридин Джунусов**

**1-ұқсастық коэффициенті (30): 1.4**

**2-ұқсастық коэффициенті (5): 0**

**Дәйексөз (35): 1.1**

**Әріптерді ауыстыру: 9**

**Аралықтар: 0**

**Шағын кеңістіктер: 0**

**Ақ белгілер: 0**

**Ұқсастық есебін талдай отырып, Жүйе администраторы мен Академиялық мәселелер департаментінің директоры келесі шешімдерді мәлімдейді :**

Ғылыми еңбекте табылған ұқсастықтар плагиат болып есептелмейді. Осыған байланысты жұмыс өз бетінше жазылған болып санала отырып, қорғауға жіберіледі.

Осы жұмыстағы ұқсастықтар плагиат болып есептелмейді, бірақ олардың шамадан тыс көптігі еңбектің құндылығына және автордың ғылыми жұмысты өзі жазғанына қатысты күмән тудырады. Осыған байланысты ұқсастықтарды шектеу мақсатында жұмыс қайта өңдеуге жіберілсін.

Еңбекте анықталған ұқсастықтар жосықсыз және плагиаттың белгілері болып саналады немесе мәтіндері қасақана бұрмаланып плагиат белгілері жасырылған. Осыған байланысты жұмыс қорғауға жіберілмейді.

**Негіздеме:**

31.05.2024  
Күні

Кафедра меңгерушісі



## Протокол

### о проверке на наличие неавторизованных заимствований (плагиата)

**Автор:** Абдинаби Мадина Абдуалиевна

**Соавтор (если имеется):**

**Тип работы:** Дипломная работа

**Название работы:** Исследование биометрических особенностей речевых сигналов

**Научный руководитель:** Нуридин Джунусов

**Коэффициент Подобия 1:** 1.4

**Коэффициент Подобия 2:** 0

**Микропробелы:** 0

**Знаки из других алфавитов:** 9

**Интервалы:** 0

**Белые Знаки:** 0

**После проверки Отчета Подобия было сделано следующее заключение:**

- Заимствования, выявленные в работе, является законным и не является плагиатом. Уровень подобия не превышает допустимого предела. Таким образом работа независима и принимается.
- Заимствование не является плагиатом, но превышено пороговое значение уровня подобия. Таким образом работа возвращается на доработку.
- Выявлены заимствования и плагиат или преднамеренные текстовые искажения (манипуляции), как предполагаемые попытки укрытия плагиата, которые делают работу противоречащей требованиям приложения 5 приказа 595 МОН РК, закону об авторских и смежных правах РК, а также кодексу этики и процедурам. Таким образом работа не принимается.
- Обоснование:

31.05.2024  
Дата

Заведующий кафедрой



## Протокол

### о проверке на наличие неавторизованных заимствований (плагиата)

**Автор:** Абдинаби Мадина Абдуалиевна

**Соавтор (если имеется):**

**Тип работы:** Дипломная работа

**Название работы:** Исследование биометрических особенностей речевых сигналов

**Научный руководитель:** Нуридин Джунусов

**Коэффициент Подобия 1:** 1.4

**Коэффициент Подобия 2:** 0

**Микропробелы:** 0

**Знаки из здругих алфавитов:** 9

**Интервалы:** 0

**Белые Знаки:** 0

**После проверки Отчета Подобия было сделано следующее заключение:**

- Заимствования, выявленные в работе, является законным и не является плагиатом. Уровень подобия не превышает допустимого предела. Таким образом работа независима и принимается.
- Заимствование не является плагиатом, но превышено пороговое значение уровня подобия. Таким образом работа возвращается на доработку.
- Выявлены заимствования и плагиат или преднамеренные текстовые искажения (манипуляции), как предполагаемые попытки укрытия плагиата, которые делают работу противоречащей требованиям приложения 5 приказа 595 МОН РК, закону об авторских и смежных правах РК, а также кодексу этики и процедурам. Таким образом работа не принимается.
- Обоснование:

31.05.2024  
Дата

Мархуси С  
проверяющий эксперт