Лабораторная работа 16 (уровень 1)

*Задания уровня 1*

*Ответы на вопросы:*

1. Какие задачи речевой обработки являются наиболее актуальными?

* Машинный перевод: автоматическое переведение текста с одного языка на другой.
* Анализ тональности: определение тональности текста (положительная, отрицательная, нейтральная).
* Классификация текста: определение категории или класса, к которому принадлежит текст.
* Генерация текста: создание текста на основе заданных параметров или контекста.
* Вопросно-ответные системы: ответ на вопросы, заданные пользователем на естественном языке.
* Суммаризация текста: создание краткого обзора или резюме большого объема текста.
* Автоматическое распознавание речи: преобразование речи в текст.
* Сентимент-анализ: определение эмоциональной окраски текста.

2. В каких областях решение этих задач находит свое применение?

* Медицина: анализ медицинских записей, автоматизация диагностики, медицинский текст-майнинг для исследований.
* Финансы: анализ новостей и финансовых данных, прогнозирование рынка, обработка финансовых отчетов.
* Компьютерные игры: создание персонализированных игровых сценариев, управление персонажами на основе естественного языка.
* Клиентский сервис и чат-боты: автоматизированный ответ на вопросы клиентов, обработка жалоб и обратной связи.
* Образование: персонализированные образовательные платформы, проверка заданий и тестов.
* Исследования и наука: обработка больших объемов текстов для извлечения знаний, анализ научных статей.
* Мультимедиа и развлечения: автоматическое создание субтитров для видео, анализ музыкальных текстов.

3. Как происходит процесс верификации говорящего?

А) Запись голоса: пользователь записывает свой голос, произнося определенные фразы или слова.

Б) Извлечение биометрических данных: при записи голоса извлекаются уникальные биометрические данные, такие как особенности тональности, скорости речи, амплитуды и т. д.

В) Создание голосового отпечатка: на основе извлеченных данных создается уникальный голосовой отпечаток пользователя.

Г) Сравнение с сохраненными данными: полученный голосовой отпечаток сравнивается с заранее сохраненными данными для проверки совпадения.

Д) Принятие решения: после сравнения система принимает решение о подтверждении или отклонении личности говорящего.

4. В чем заключается отличие верификации от идентификации говорящего по голосу?

Верификация говорящего:

- Цель: проверка подлинности утверждения человека о своей личности.

- Процесс: пользователь утверждает, что он является определенным человеком, и его голос сравнивается с заранее сохраненными данными для подтверждения или опровержения утверждения.

Идентификация говорящего:

- Цель: определение личности говорящего без предварительного утверждения.

- Процесс: голос пользователя сравнивается с базой данных голосов различных людей для определения, кто именно говорит. Система ищет соответствие в базе данных и определяет личность говорящего.

Таким образом, верификация говорящего используется для проверки утверждений о личности, а идентификация говорящего применяется для определения личности без предварительного утверждения. Оба процесса могут быть полезны в различных областях, требующих аутентификации по голосу.

5. Чем обусловлена актуальность решения такой задачи как преобразование речи в текст?

* Доступность информации: преобразование речи в текст делает аудио- и видеоконтент более доступным для людей с нарушениями слуха или для тех, кто предпочитает текстовый формат.
* Удобство и эффективность: текстовый формат удобен для быстрого поиска, анализа и цитирования информации. Преобразование речи в текст позволяет легко сохранять и передавать содержание.
* Автоматизация процессов: технологии распознавания речи позволяют автоматизировать процессы, такие как транскрибирование интервью, заседаний, лекций и других аудиозаписей.
* Улучшение пользовательского опыта: виртуальные помощники, системы распознавания голоса и другие приложения, преобразующие речь в текст, значительно улучшают пользовательский опыт и удобство взаимодействия с устройствами.
* Развитие искусственного интеллекта: преобразование речи в текст является важным компонентом развития технологий искусственного интеллекта, таких как обработка естественного языка и распознавание речи.
* Медицинская сфера: в медицинской области преобразование речи в текст может быть использовано для документирования медицинских записей, консультаций и других процессов.
* Повышение производительности: преобразование речи в текст позволяет увеличить производительность и эффективность работы, особенно в случаях, когда необходимо быстро перевести устную информацию в письменный формат.

6. В чем заключается отличие первых систем Speech-to-Text от тех, которые разрабатываются в настоящее время?

Первые системы распознавания речи (Speech-to-Text) имели ограниченные возможности и качество работы по сравнению с современными разработками. Вот несколько основных отличий между первыми системами и современными технологиями распознавания речи:

* Точность: первые системы распознавания речи были менее точными из-за ограничений в алгоритмах и доступных данных. Современные системы используют глубокое обучение и нейронные сети, что позволяет им достигать значительно более высокой точности.
* Скорость: современные системы работают намного быстрее благодаря улучшенным алгоритмам и вычислительной мощности современных компьютеров.
* Распознавание различных языков и диалектов: современные системы обладают возможностью распознавать большее количество языков и диалектов, что делает их более универсальными.
* Адаптивность к голосам разного типа: современные системы способны лучше адаптироваться к разным голосам, включая акценты, интонации и скорость речи.
* Контекстуальное понимание: некоторые современные системы распознавания речи способны учитывать контекст и предыдущие фразы для более точного распознавания.
* Интеграция с другими технологиями: современные системы распознавания речи часто интегрируются с другими технологиями, такими как искусственный интеллект, автоматизация процессов и анализ данных.

7. За счет чего системы распознавания речи стали допускать меньше ошибок?

* Глубокое обучение и нейронные сети: современные системы распознавания речи используют методы глубокого обучения, такие как нейронные сети, которые способны автоматически извлекать признаки из аудиоданных и обучаться на больших объемах информации. Это позволяет им более точно распознавать речь.
* Большие объемы данных: с появлением больших объемов аудиоданных, доступных для обучения, системы могут учиться на более разнообразных примерах речи, что повышает их точность.
* Улучшенные алгоритмы и модели: исследования в области машинного обучения привели к созданию более эффективных алгоритмов и моделей, способных лучше учитывать различные аспекты речи, такие как интонация, акценты и контекст.
* Комбинирование различных подходов: современные системы распознавания речи часто комбинируют несколько подходов, таких как использование языковых моделей, контекстуальное понимание и адаптивность к разным голосам, что помогает снизить количество ошибок.
* Обратная связь и коррекция ошибок: некоторые системы используют механизмы обратной связи и коррекции ошибок, что позволяет им постепенно улучшать свою точность в процессе использования.

8. Перечислите этапы работы систем распознавания говорящего и преобразования речи в текст.

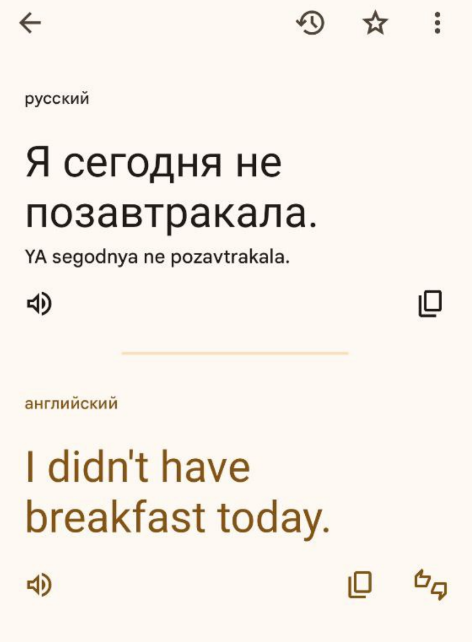
* Захват аудиосигнала: на этом этапе система захватывает аудиосигнал, который содержит речь говорящего. Это может быть сделано с помощью микрофона или другого аудиоустройства.
* Предварительная обработка: аудиосигнал может быть предварительно обработан для улучшения качества звука, фильтрации шумов или уменьшения эха.
* Функция преобразования речи в текст (STT): на этом этапе алгоритмы распознавания речи преобразуют аудиосигнал в текстовую форму. Это включает в себя анализ звуковых волн, выделение особенностей и соотнесение их с известными шаблонами для определения слов
* Обработка и анализ текста: полученный текст может быть дополнительно обработан для исправления опечаток, уточнения контекста или интерпретации специфических терминов.
* Постобработка и улучшение: некоторые системы могут использовать методы машинного обучения для постобработки текста, уточнения результатов распознавания и повышения точности системы в целом.

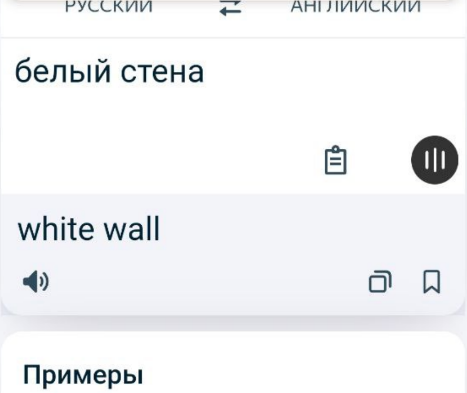
1.1

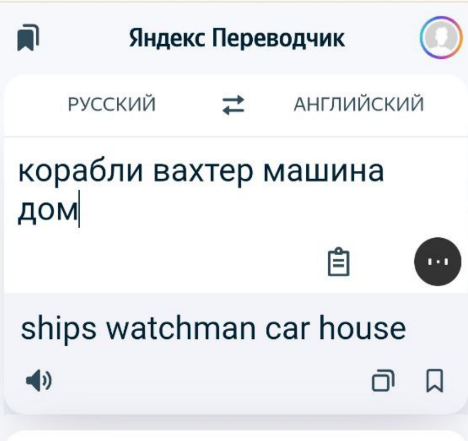
Используем переводчик от гугл и яндекс

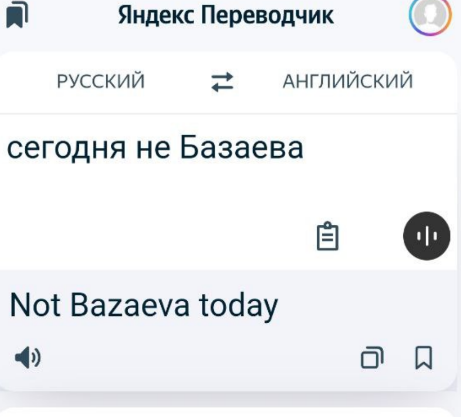






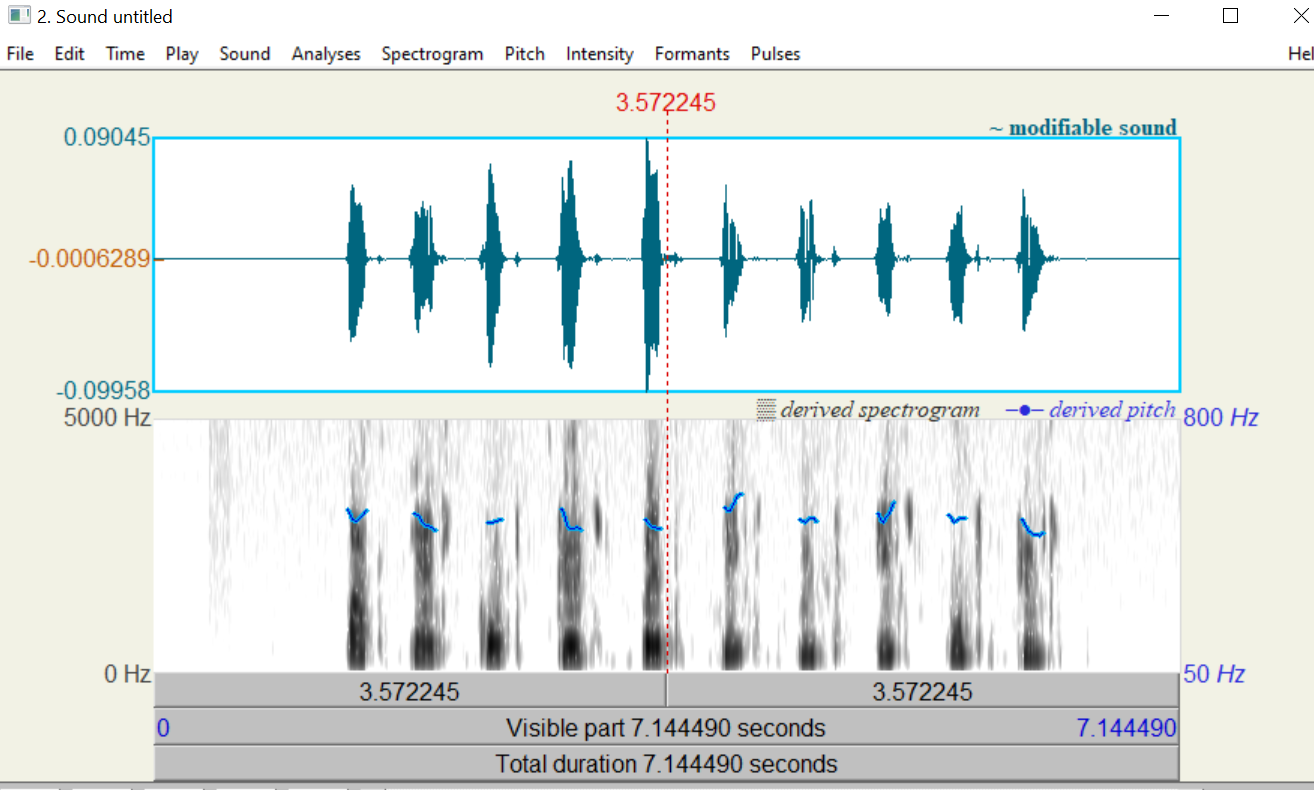




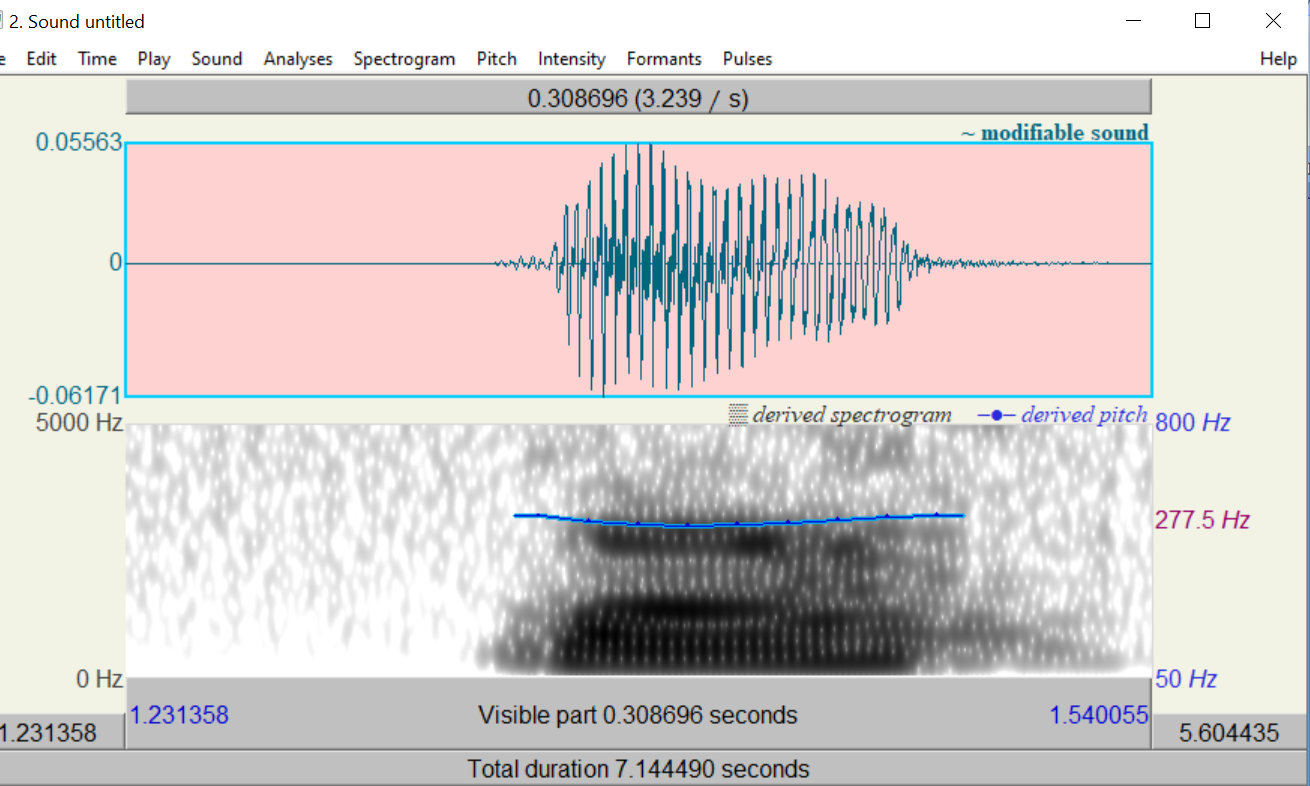


1.2 (1)

Загружаем звук



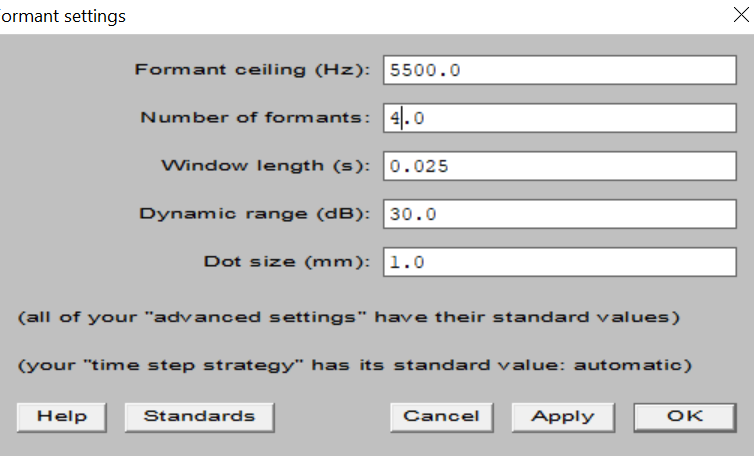
Увеличим выделенный фрагмент:

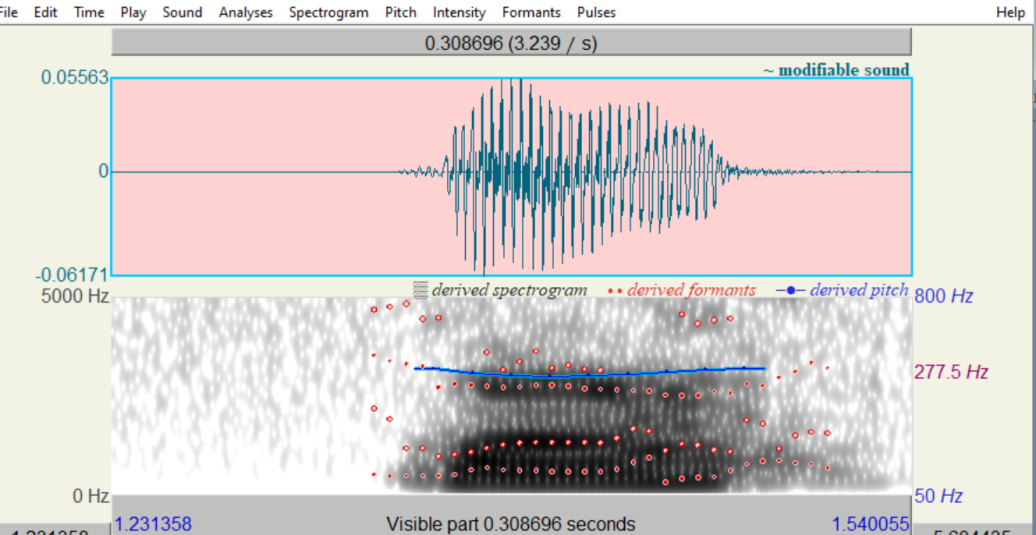


В нижней части окна появится спектрограмма. При нажатии на область спектрограммы по горизонтали отразится время, а по вертикали – частота

Синими точками в Praat отображается Частота основного тона.

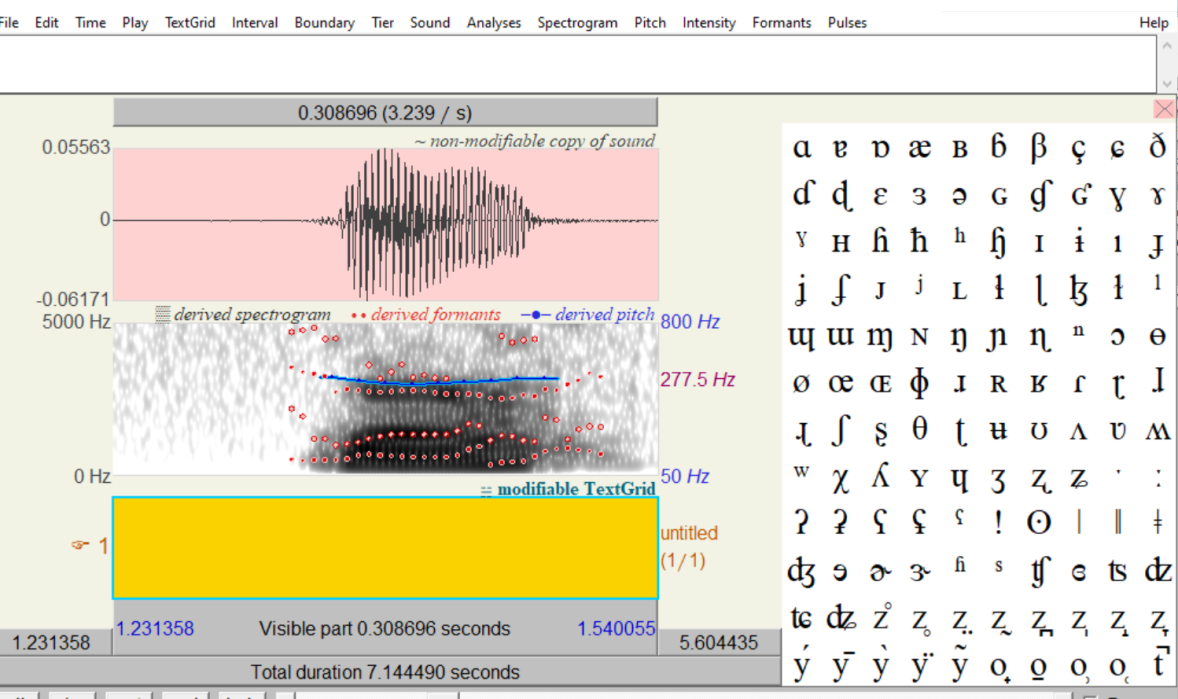
Выделим формант:



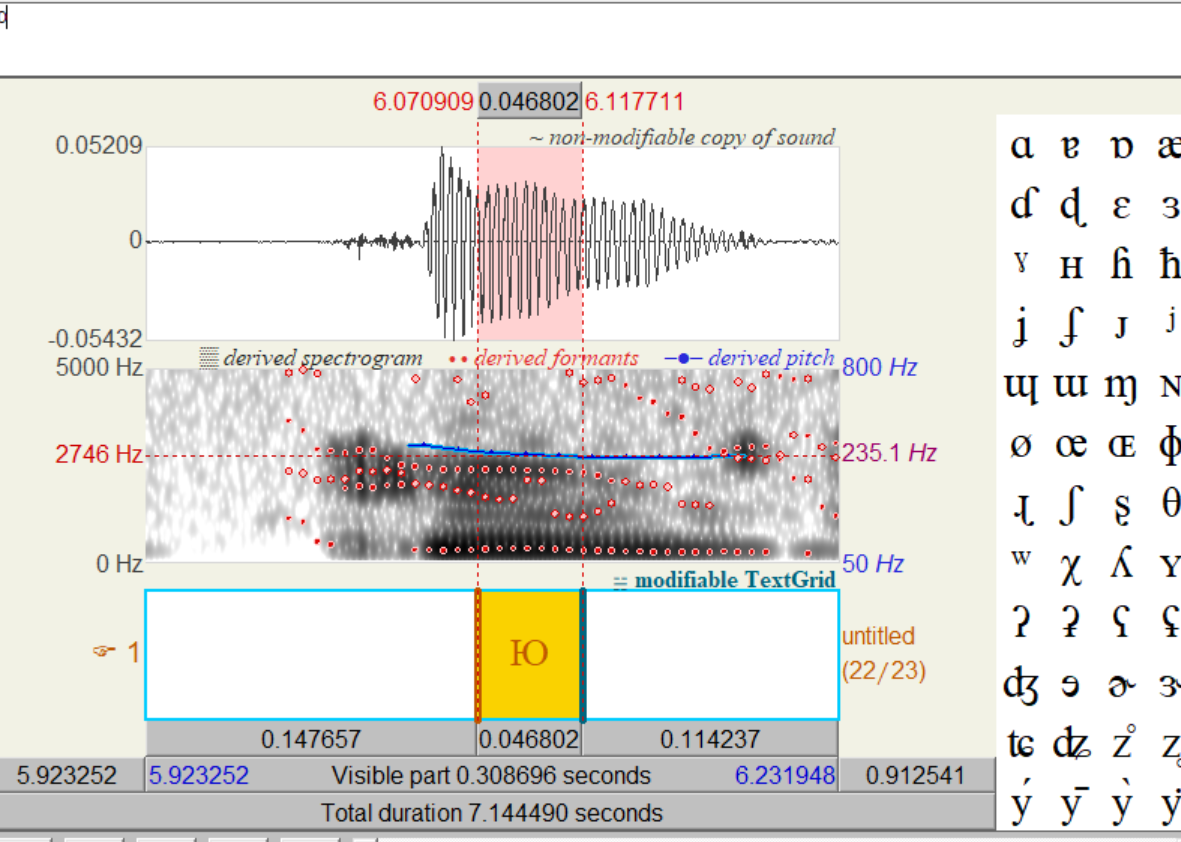


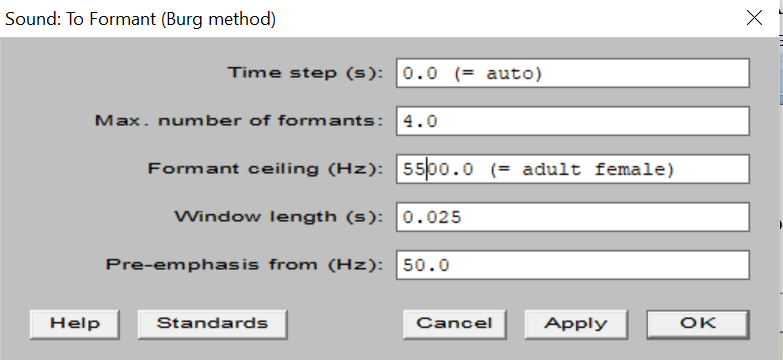
Для построения диаграммы гласных нам необходимо разметить гласные в речевом фрагменте.

Появился ещё один файл



Далее мы будем выделять каждый участок речевого фрагмента и обозначать на нем гласные

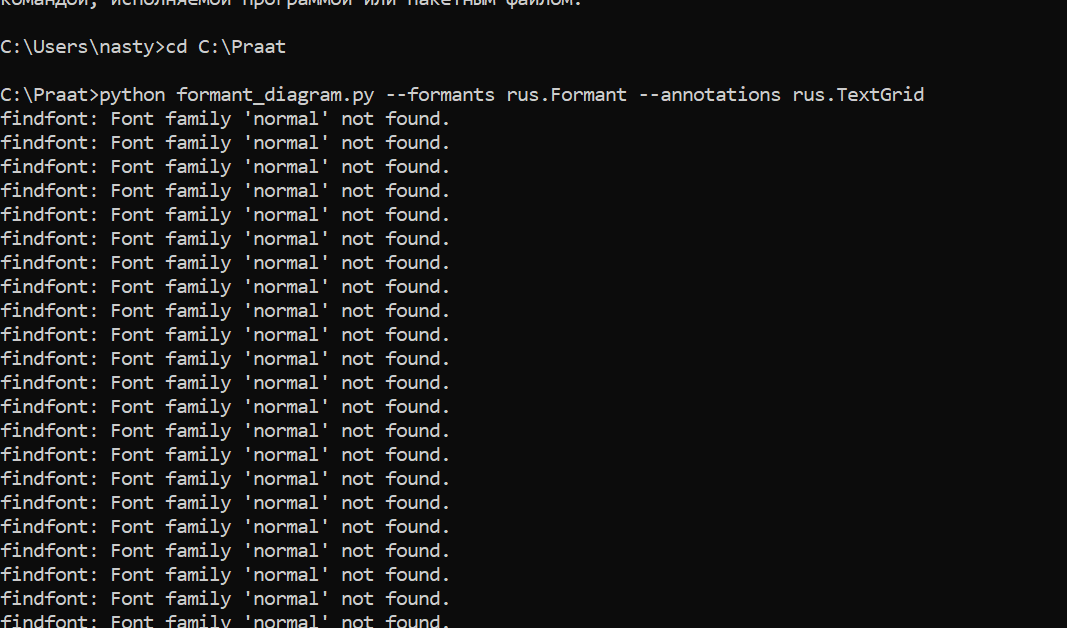


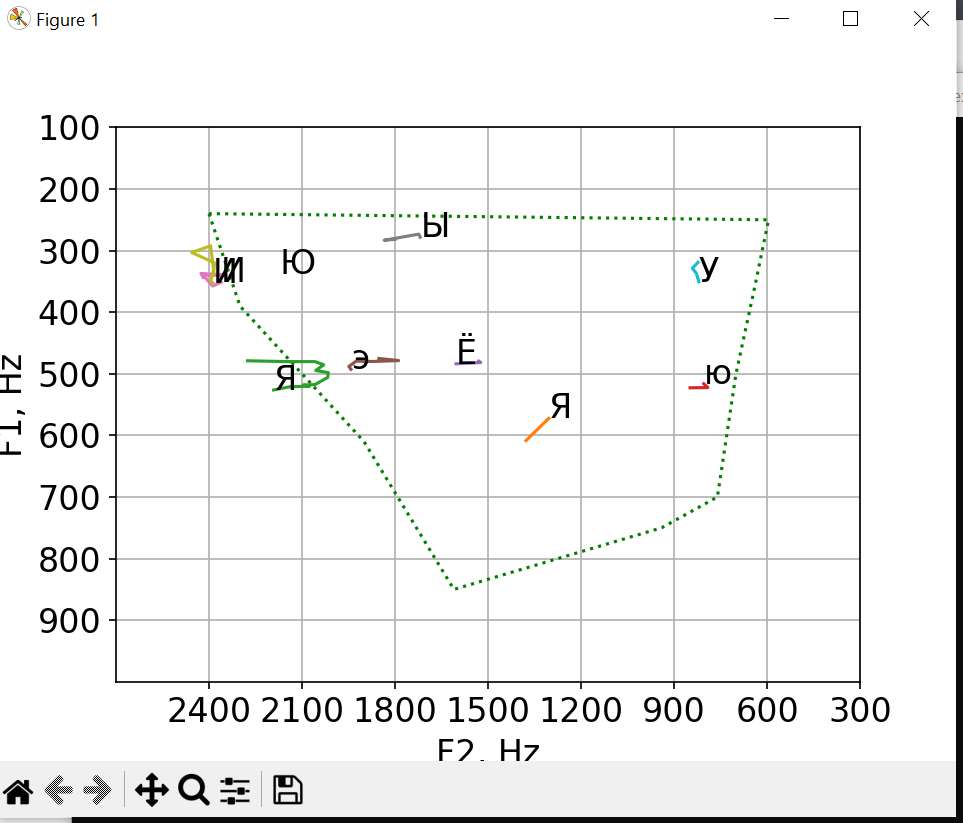


Появился ещё один файл

Сохраняем два последних файла в текстовом формате

Далее скачиваем программы с Гугл Диска и сохраняем в папки с файлами из Praat





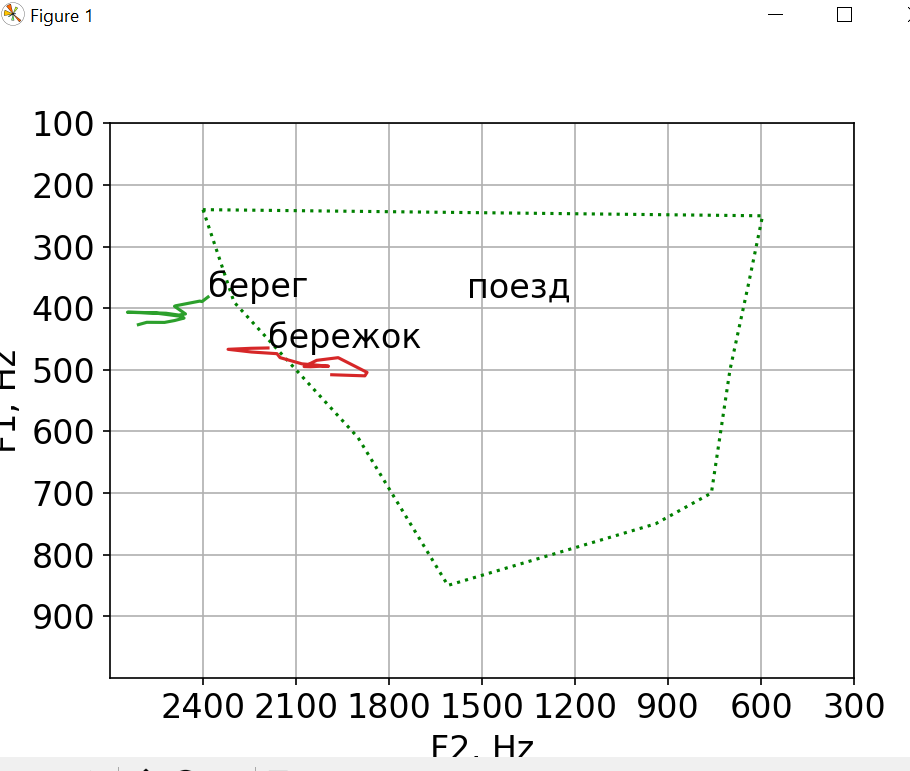
Расположение каждого гласного зависит от усредненного значения формант, зеленой пунктирной линией обозначено приблизительное пространство гласных, разноцветными линиями обозначено изменение гласного со временем, по вертикали отражаются значения для F1, по горизонтали – для F2.

1.2 (2)

Возьмём слова: поезд, берег, бережок (буква е в разных ударных позициях)

Делаем все то же самое, как и в прошлом примере

Получаем:



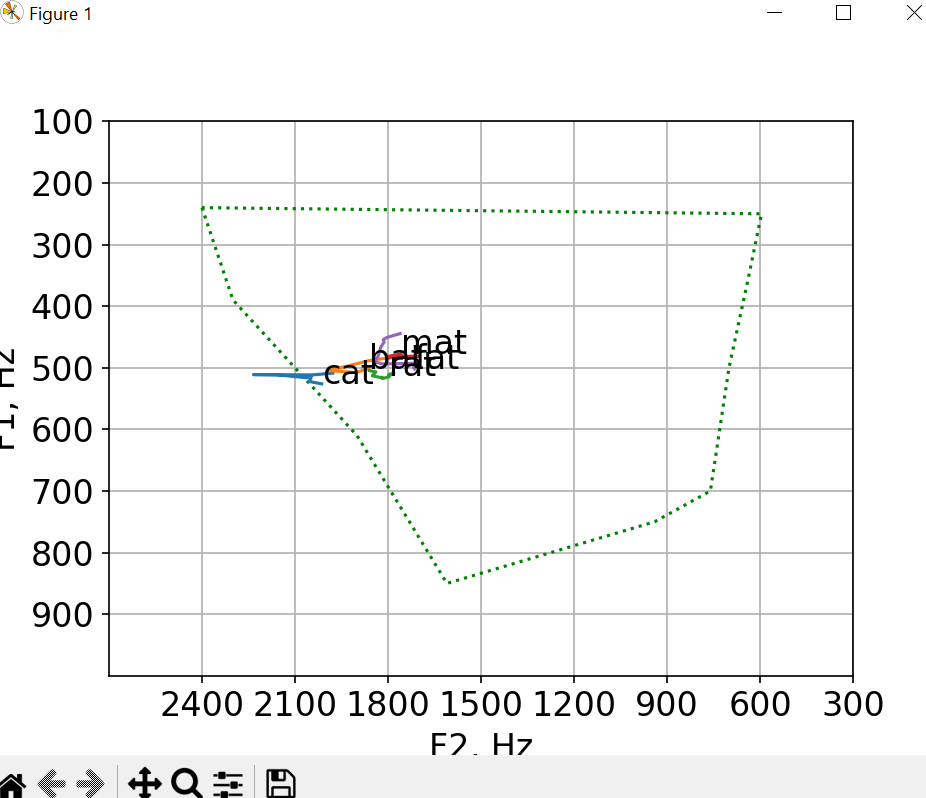
Глядя на график, можно сказать, что звук разный.

1.2 (3)

Возьмём слова: cat, bat, rat, mat, fat

Делаем все то же самое, как и в прошлом примере

Получаем:



Глядя на график можно сказать, что произношение созвучно.

1.3

*Ответы на вопросы:*

1. Что из себя представляют системы Text-To-Speech и что является их основной задачей?

Системы Text-To-Speech (TTS) представляют собой технологию, которая преобразует написанный текст в звуковую речь. Основная задача TTS - создание естественно звучащей речи, которая звучит так, будто её произносит человек. Это позволяет компьютерам и устройствам воспроизводить текст в виде аудио, что может быть полезно для людей с ограничениями в зрении, а также для создания аудиокниг, голосовых помощников и других приложений, где требуется озвучивание текста.

2. По каким критериям оценивают сгенерированную речь?

* Естественность звучания: оценка того, насколько звучит сгенерированная речь естественно, как человеческая речь.
* Четкость и понятность: насколько хорошо каждое слово и фраза воспринимается и понимается слушателем.
* Интонация и эмоциональная окраска: способность передавать различные интонации, эмоции и акценты в речи.
* Скорость речи: возможность контролировать скорость произношения текста для оптимального восприятия.
* Артикуляция и произношение: как точно и четко произносятся отдельные звуки, слова и фразы.
* Производство пауз и переходов: как естественно вставляются паузы между словами и фразами, а также как происходят переходы между ними.
* Акцент и диалект: умение воспроизводить различные акценты и диалекты, если это необходимо.

3. Какой способ синтеза речи подразумевает использование записанных фрагментов человеческой речи? Опишите принцип его работы.

Способ синтеза речи, который подразумевает использование записанных фрагментов человеческой речи, называется конкатенативным синтезом речи. Принцип работы заключается в том, что заранее записанные звуки, фразы или слова собираются и объединяются в нужном порядке, чтобы создать натуральный звучащий поток речи.

4. Какие два способа синтеза речи основываются на акустических свойствах речи? В чем их отличие?

Конкатенативный синтез (Waveform concatenation)

Формантный синтез (Formant synthesis)

Конкатенативный синтез: соединяет вместе предварительно записанные фрагменты речи, называемые дифонами или трифонами.

Формантный синтез: генерирует речь путем манипулирования формантами - резонансными частотами вокального тракта.

Основное отличие между этими двумя методами заключается в том, что конкатенативный синтез использует предварительно записанные фрагменты речи, а формантный синтез генерирует речь на основе акустических моделей.

5. В чем заключается особенность артикуляционного способа синтеза речи?

Особенностью артикуляционного способа синтеза речи является моделирование физических процессов артикуляции и формирования речи человеком. Этот метод стремится имитировать анатомические и физиологические особенности производства звуков человеческой речи.

6. Перечислите основные сложности для синтеза речи каждым способом. Какие существуют пути их решения?

Основные сложности для синтеза речи каждым способом могут быть связаны с недостаточной натуральностью звучания, достижением высокой качества и свойственной для человеческой речи экспрессивности.

Для их решения могут применяться различные техники и алгоритмы улучшения качества синтеза, такие как использование более сложных моделей, улучшение алгоритмов синтеза и обработка сигналов.

Практическое задание:

Будем использовать: переводчик от гугл, переводчик от яндекса, PiliApp и iSpeech.

Возьмем предложения из примера:

Все зависит от того, к чему у тебя лежит душа′.

После принятия ду′ша не стоит сразу выходить на улицу.

*Гугл переводчик:*

Речь разборчива, но поставить правильно ударение синтезатор речи не смог (в предложением со слвом ду′ша).

*Янедкс переводчик:*

Речь разборчива, ударение поставлено верно в обоих случаях.

*PiliApp:*

Речь разборчива, но поставить правильно ударение синтезатор речи не смог (в предложением со слвом ду′ша). Речь очень роботизирована.

*iSpeech:*

Речь разборчива, но текст произносится очень быстро, поставить правильно ударение синтезатор речи не смог (в предложением со слвом ду′ша). Речь очень роботизирована.