**Задание 1.**

Вопросы по теме «Задачи автоматического анализа речи»:

1. К наиболее важным задачам речевой обработки и анализа можно отнести распознавание говорящего по голосу и преобразование речи в текст.
2. Автоматическое создание транскрипции, синхронное преобразование живой речи в текстовый формат, управление голосовыми помощниками, выявление особенностей речи говорящего, выделение наиболее значимых слов в речи, идентификация и верификация личности, создание и разработка технологий для людей с ограниченными возможностями, создание диалоговых систем и чат-ботов.
3. Говорящий производит речевой сигнал – сообщает данные о своей личности (инициалы или специальный код), а система автоматического распознавания индивидуальных особенностей голоса и речи должна подтвердить личность говорящего, исходя из эталонной модели, т.е. имеющейся записи голоса говорящего.
4. Идентификация включает в себя определение личности говорящего из конкретного списка пользователей10. Система, исходя из поступившего речевого сигнала, должна определить, кто из пользователей вступает с ней в контакт.
5. Данная технология активно применяется: при автоматическом создании субтитров; при разработке приложений и технических средств для людей с ограниченными возможностями, у которых возникают сложности при наборе текста вручную; для распознавания команд голосовыми помощниками.
6. Первые образцы систем STT работали исключительно на основе эталонных моделей, при этом речь членилась системой на отдельные крупные блоки, после чего происходил поиск наиболее близкого акустического сигнала из уже записанных в базе данных, что приводило к большому количеству ошибок. В настоящее время, благодаря использованию баз данных коллокаций (устойчивых сочетаний) и N-gramm (последовательностей из N символов), системы STT способны предугадывать целые фразы, исходя из правильно распознанного фрагмента речи.
7. Внедрение машинного обучения, которое позволяет системам STT постоянно совершенствоваться.
8. Анализ сигнала, Распознавание сигнала, Преобразование в текст.

Практическое задание:

Яндекс переводчик – на русском языке: хорошо распознает речь связанных предложений и отдельных слов (если делать паузы), если предложение не связано логически или слова несогласованны, то в результате получим отличное предложение от начального (будет пытаться связать слова в предложении). На английском языке: если делать определенные паузы в несвязанных предложениях, то ничего не будет исправлено, всё будет корректно записано.

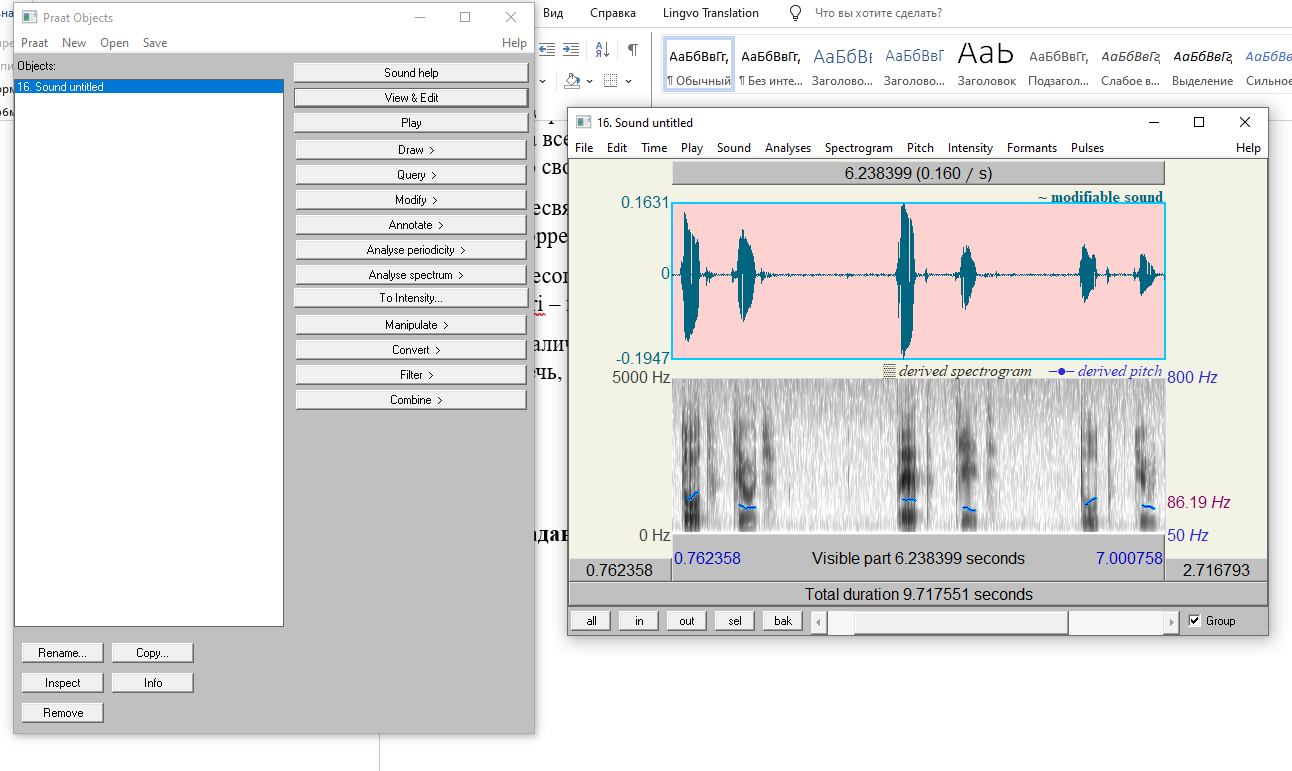
Siri – направлена на смысловое понимание контекста, поэтому делает больше исправлений, если в речи присутствуют неточности. Однако благодаря этому, на все свои вопросы получаю нужный ответ, поэтому siri хорошо справляется со своей задачей.

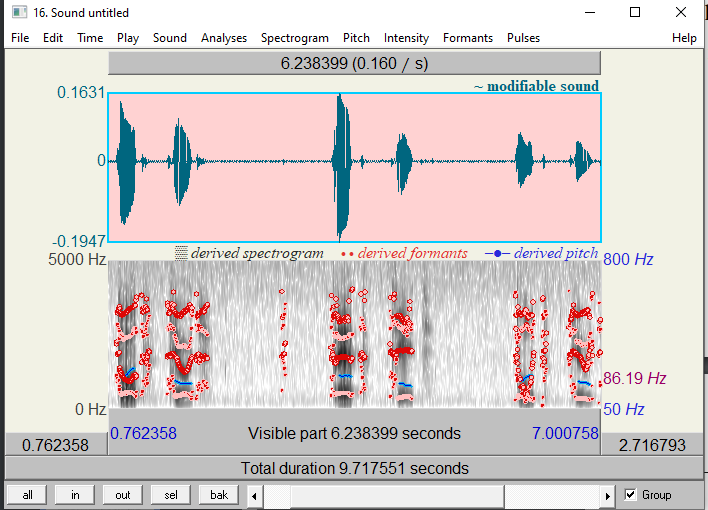
Несвязный речевой сигнал: яндекс – если делать паузы, то всё будет корректно обработано, siri – в любом случае постарается исправить текст.

Несогласованные друг с другом слова: яндекс – корректно принимает слова, siri – понимает контекст, однако сразу исправляет ошибки.

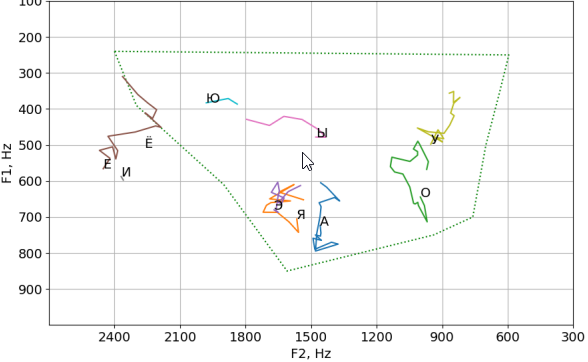
Наличие шума (параллельной речи): яндекс – воспринимает параллельную речь, получаем набор слов, siri – принимает только наши слова.

**Задание 2.**



Форманты:  


Построил диаграму



Тот – тёть, тут – тють, … (что использовал)

Ряды словоформ:

А: ждать, часы, кружка.

О: солнце, стопа, тело.

У: лук, суфлёр, папирус.

Ю: люк, сюда, мною.

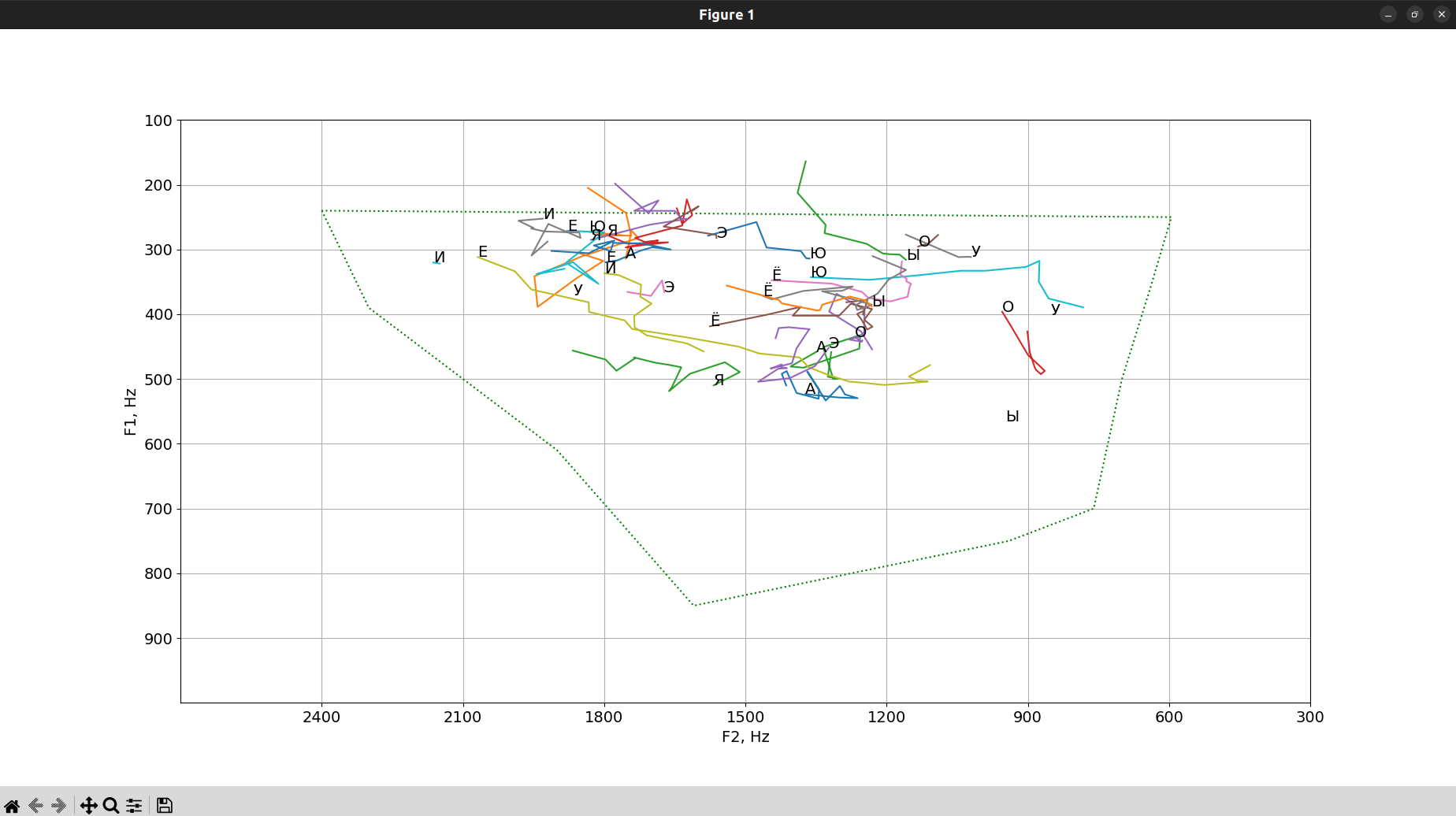
Я: яхта, язык, туфля.

Ё: ёжик, зачёт, дёготь.

Е: вена, река, овощем.

Ы: дымка, дыхание, горы.

Э: мэр, этап, эшафот.

И: вид, пиар, армия.

**Задание 3.**

Вопросы по теме «Основы работы систем преобразования текста в речь»:

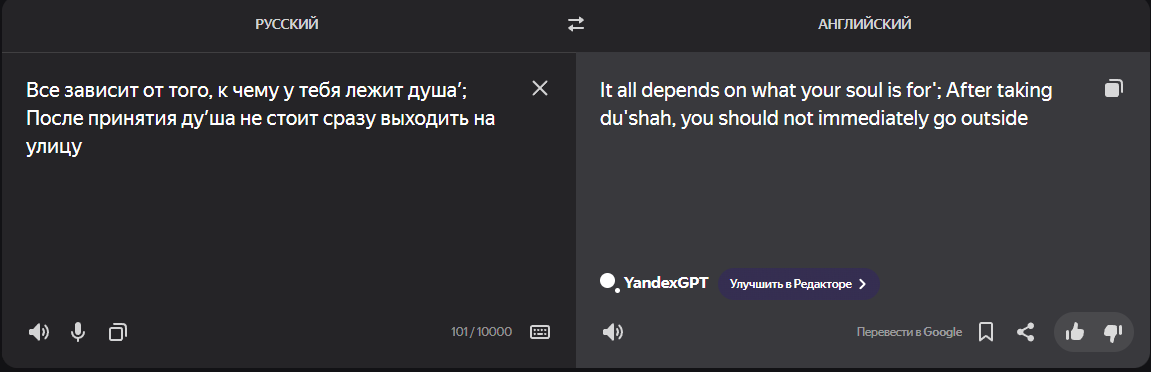
1. Системы Text-To-Speech (TTS) представляют собой технологию, которая преобразует текстовую информацию в аудиофайл с синтезированной речью. Основная задача TTS - создание естественно звучащей речи на основе введенного текста.
2. Сгенерированную речь оценивают по нескольким критериям, таким как естественность (насколько звучит речь как у человека), понятность (насколько легко понять произнесенный текст), выразительность (способность передавать эмоции и интонации) и т.д.
3. Способ синтеза речи, использующий записанные фрагменты человеческой речи, называется конкатенативным синтезом. Он работает путем объединения заранее записанных фрагментов речи для создания целого предложения или текста.
4. Два способа синтеза речи, основанные на акустических свойствах речи, это синтез на основе формантов и синтез на основе скрытых марковских моделей (HMM). Основное отличие между ними заключается в подходе к моделированию и генерации звуков.
5. Артикуляционный синтез основывается на создании модели человеческого речевого аппарата и ее реализации с помощью некоторого механизма – «говорящей» машины, которая порождает речевой сигнал путем артикуляции.
6. Основные сложности для синтеза речи каждым способом включают в себя естественность звучания, высокую степень детализации, управление просодией (интонация, темп, паузы) и др. Пути их решения включают в себя использование глубоких нейронных сетей, улучшение баз данных для синтеза, разработку более точных моделей артикуляции и другие технические усовершенствования.

Практическое задание.

Текст.

Так бережно разливается по воздуху звук торопливого ручейка, весело бегущего извилистой ленточкой среди травы. Словно играя на арфе, падают росинки с деревьев, перепрыгивают еле слышно на цветы, а потом беззвучно исчезают в траве. На поляне жужжат деловые шмели, саранча поет свою незамысловатую песню. То здесь, то там слышно щебетанье птиц.  А тон всему задает тихий шелест ветра. Словно дирижер он руководит журчанием ручья, следит за тональностью росинок и периодически утихомиривает насекомых. Лида лежит на мягкой траве, прикрывая голову от солнца большим листом лопуха, и наслаждается музыкой природы. Она ничего не может слышать уже несколько лет. Ей помогает в этом память.

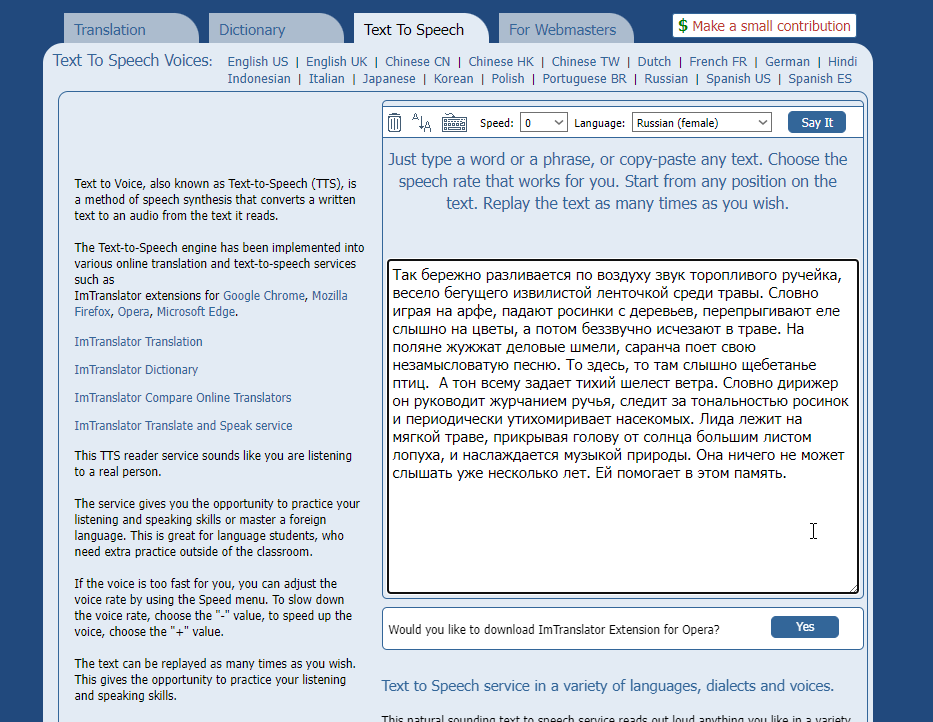
Яндекс переводчик. Есть нотки интонации, но сухо. Наличие пауз при запятых и точках. Не различает словоформы с разными ударениями.



Oddcast. Большие проблемы с ударением и интонацией. Наличие пауз перед знаками препинания.



Text-to-speech. Отсутствует интонация, не справляется со сложными ударениями. Паузы присутствуют.



PiliApp. Есть только паузы, самая ужасная озвучка. 