**Лабораторная работа №15**

**«Фонетика»**

**Уровень 1**

**Ответы на вопросы по теме «Задачи автоматического анализа речи»:**

1. Какие задачи речевой обработки являются наиболее актуальными?

Наиболее актуальными задачами речевой обработки являются: 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

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

Преобразование звучащей речи в текст применяется в:

* Голосовых помощниках (Siri, Alexa, Google Assistant)
* Системах голосового ввода текста
* Транскрипция аудиозаписей (интервью, лекций, совещаний)
* Автоматизация колл-центров

Определение, кто говорит (по биометрическим характеристикам голоса) используется в биометрической защите и авторизации

Определение эмоционального состояния говорящего применяется в:

* В call-центрах для анализа удовлетворённости клиента
* В психологии и медицине (мониторинг депрессии, стресса)
* В интерактивных системах (роботы, персонализированные ИИ-ассистенты)

Синтез речи применим в:

* В навигаторах (например, Яндекс.Навигатор)
* В книгах с озвучиванием (аудиокниги)
* В образовательных приложениях для детей и людей с ограниченными возможностями

Анализ речи используется для диагностики и мониторинга заболеваний:

* Неврологические расстройства (например, болезнь Паркинсона)
* Психические заболевания (шизофрения, депрессия)
* Логопедия и восстановление после инсультов

1. Как происходит процесс верификации говорящего?

Под верификацией понимают определение системой говорящего как «своего» или «чужого». В качестве примера верификации можно рассмотреть следующую ситуацию: 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

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

Процесс верификации подтверждает личность говорящего, а процесс идентификации распознает, исходя из поступившего речевого сигнала, кто из пользователей вступает в контакт.

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

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 (Google Assistant, Яндекс Алиса, Siri, Alexa).

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

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.

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

Немаловажную роль в сокращении количества ошибок при распознавании играет внедрение машинного обучения, которое позволяет системам STT постоянно совершенствоваться.

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

Можно выделить три этапа преобразования речи в текст:

* Анализ сигнала. Полученный на вход речевой сигнал очищается от лишних шумов и делится на минимальные фрагменты.
* Распознавание сигнала. Фрагменты проходят через эталонную модель, определяющую, какие звуки были произнесены; после чего система пытается объединить их в значимые единицы.
* Преобразование в текст. При помощи баз данных и модели языка происходит подбор нераспознанных единиц, исходя из контекста, и преобразование полученных результатов в графический вид.

**Практическое задание:**

Оцените качество работы систем распознавания речи на русском и английском языках. Проверьте, ухудшится ли результат, если на вход системы STT сообщить: несвязный речевой сигнал (случайные наборы слов), несогласованные друг с другом слова (например, по роду) или речевой сигнал с параллельной речью нескольких людей (или посторонним шумом). Можете использовать системы распознавания речи, внедренные в онлайн переводчики (Яндекс Переводчик, Google Translation) или в голосовые помощники (Google Assistant, Яндекс Алиса, Siri, Alexa).

**Выполнение практического задания:**

Для оценки качества работы систем распознавания речи я выбрала голосовые помощники Siri и Яндекс Алиса:

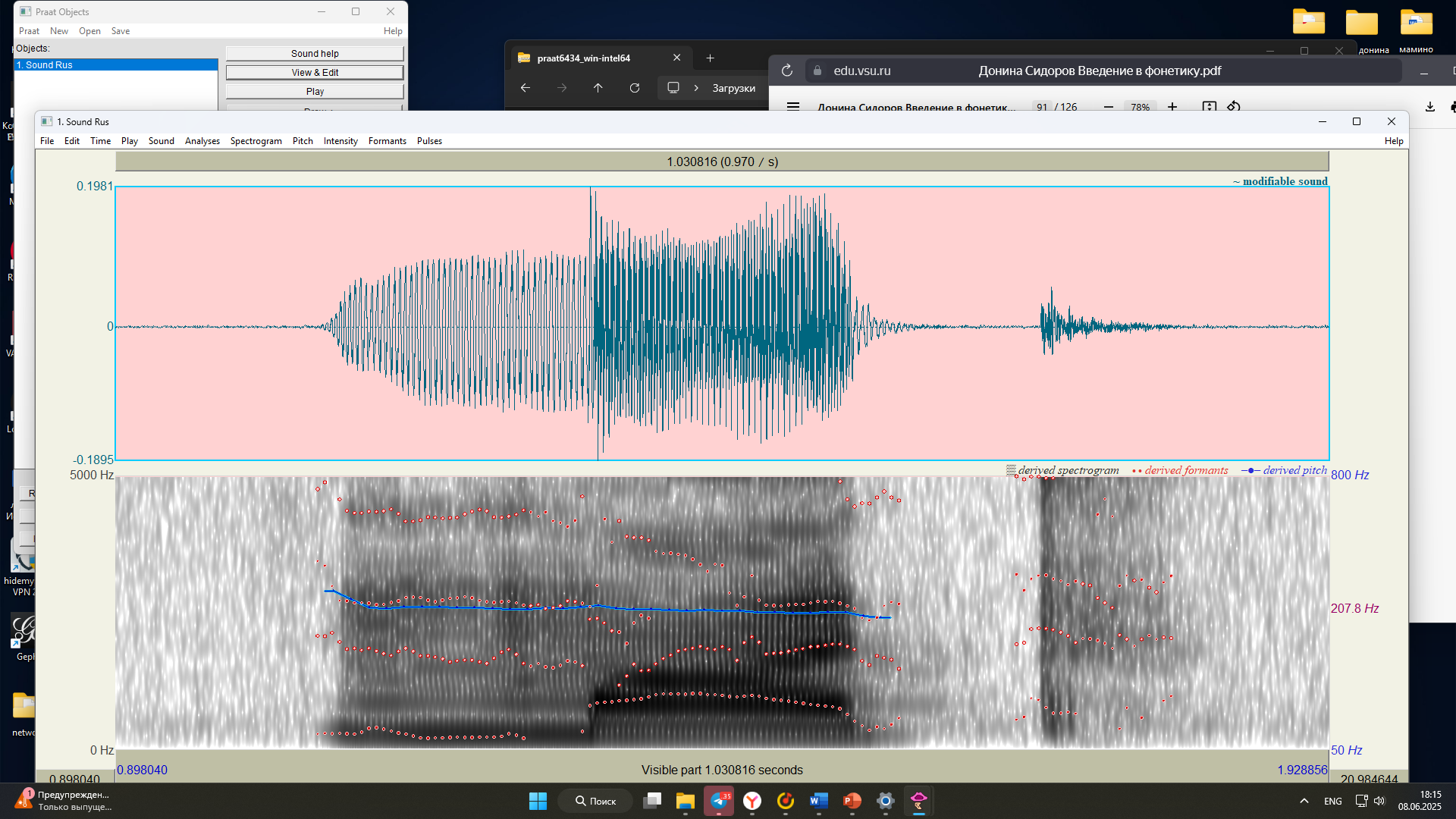
|  |  |  |
| --- | --- | --- |
| условие | google assistant | яндекс алиса |
| говорю связно и четко на русском языке | точно распознает речь | точно распознает речь |
| говорю связно и четко на английском языке | неплохо распознает английскую речь, но иногда делает ошибки | не очень хорошо распознает речь, делает ошибки в словах |
| говорю случайные наборы слов | неплохо распознает речь | неплохо распознает речь, но делает ошибки |
| говорю несогласованные друг с другом слова | система может переформулировать | передает как есть, но иногда может переформулировать |
| говорю с посторонним шумом | смешивает речь | смешивает речь |

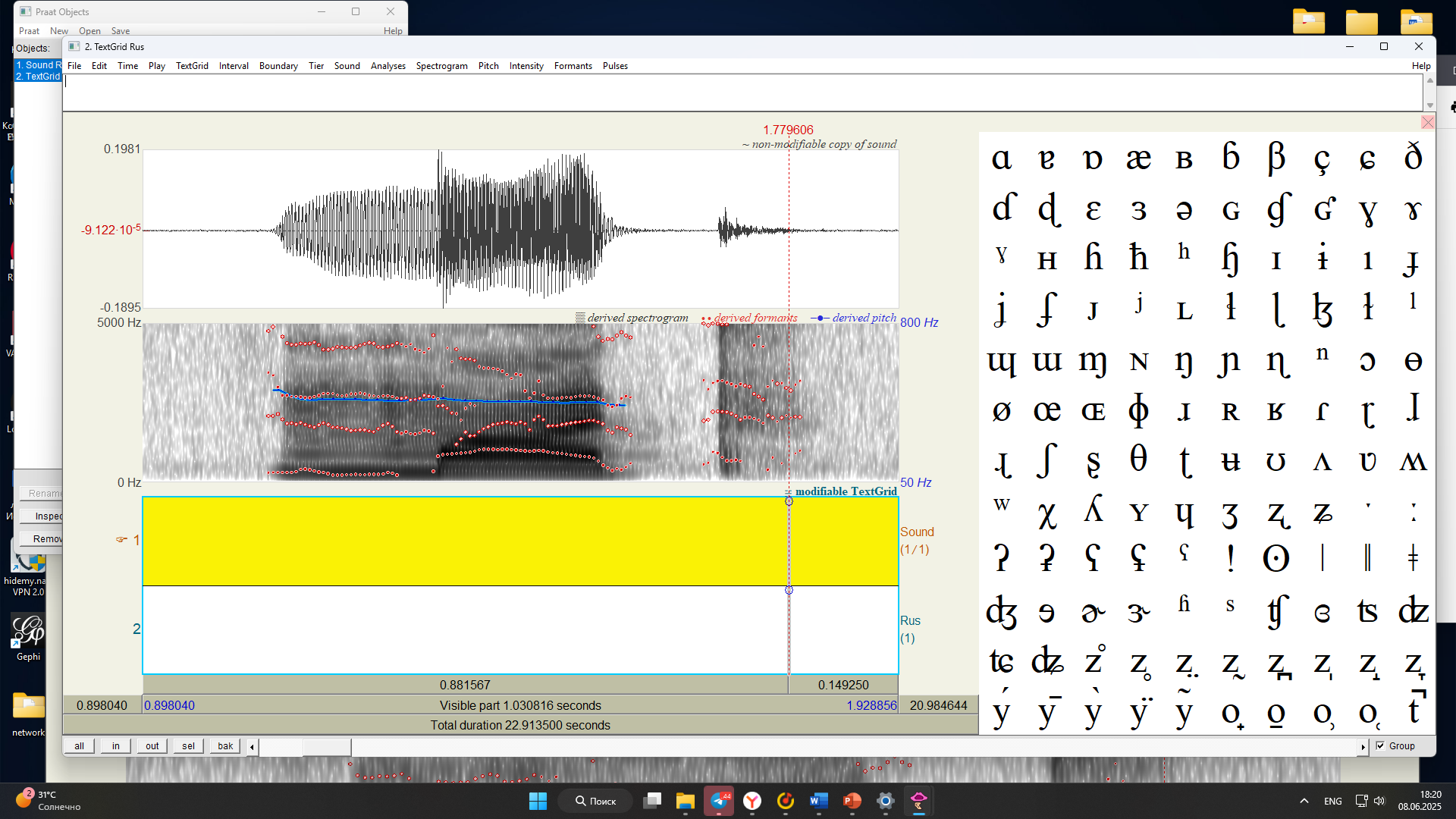
**Лабораторная работа**

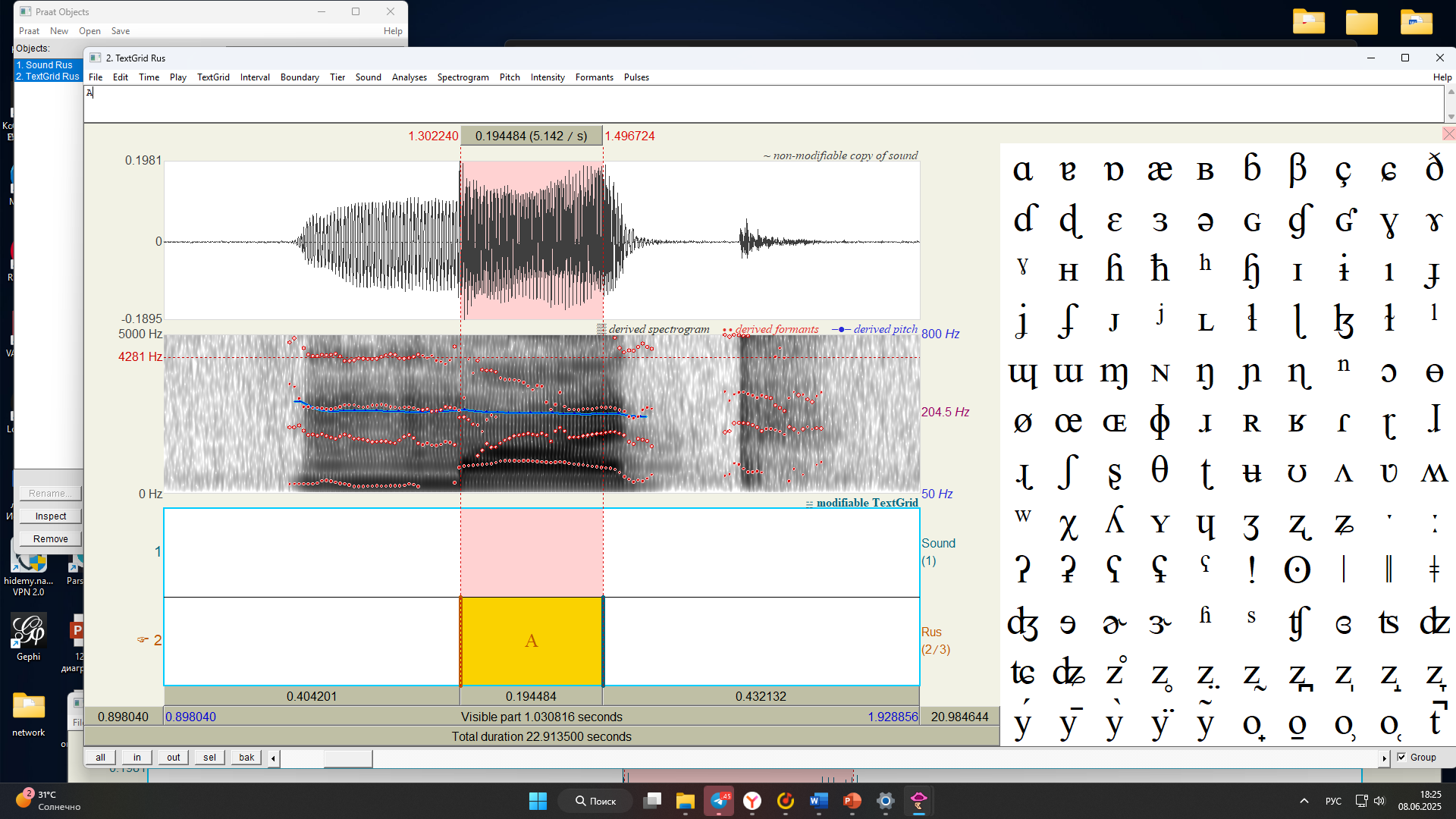
1. По аналогии с приведенным в параграфе примером, запишите речевой фрагмент с гласными звуками в ударной позиции между твердыми и мягкими согласными, проведите аннотирование звуков и создайте собственную диаграмму гласных. Сделайте анализ полученных результатов: изменение одного и того же звука в разных позициях (между твердыми/мягкими согласными); изменение конкретного звука в процессе произнесения. Сопоставьте свою диаграмму с диаграммой гласных МФА.

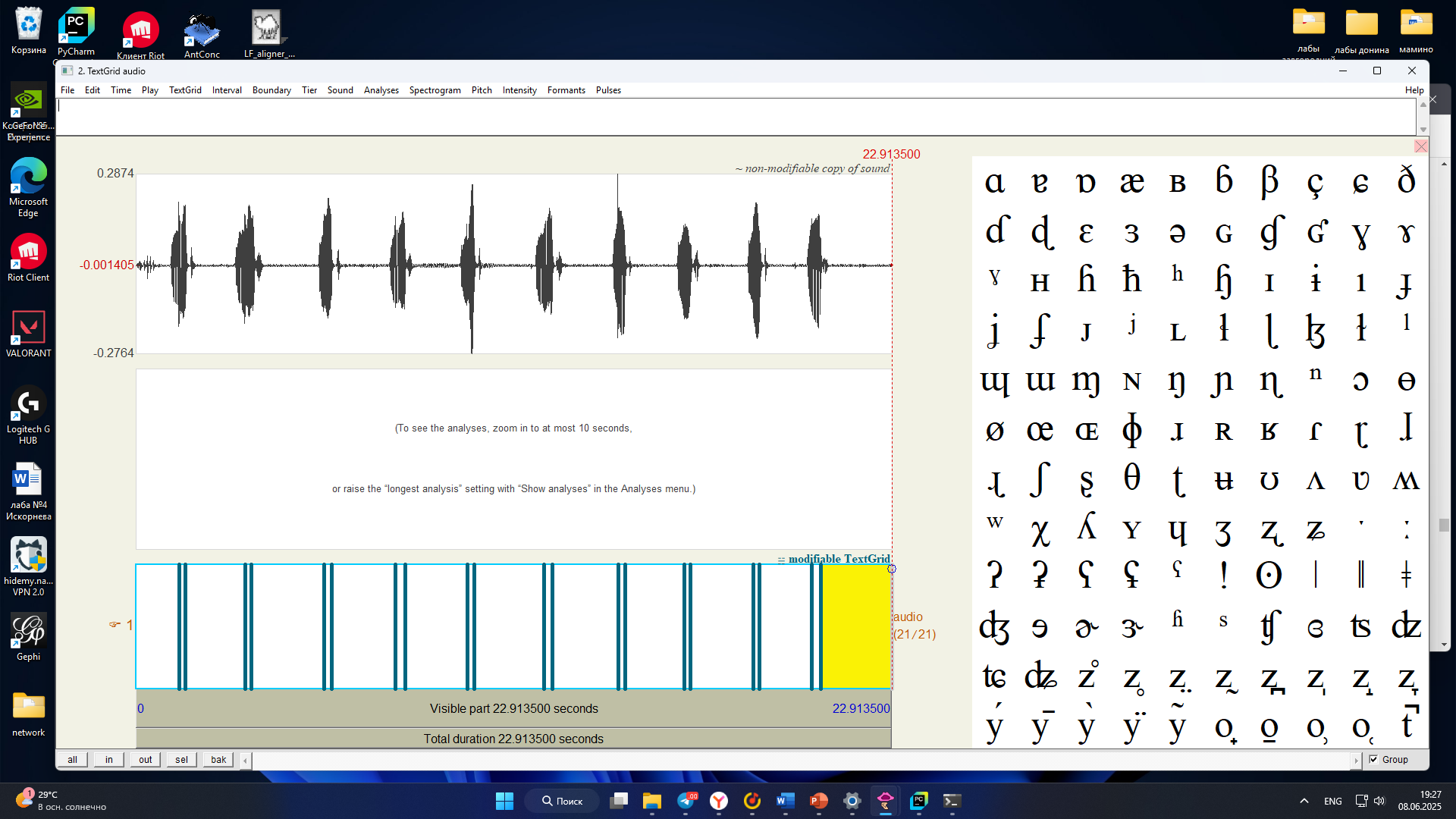
Использовались словоформы: «мат»-«мять», «мот»-«мёть», «мэт»-«меть», «мыт»-«мить», «мут»-«мють»

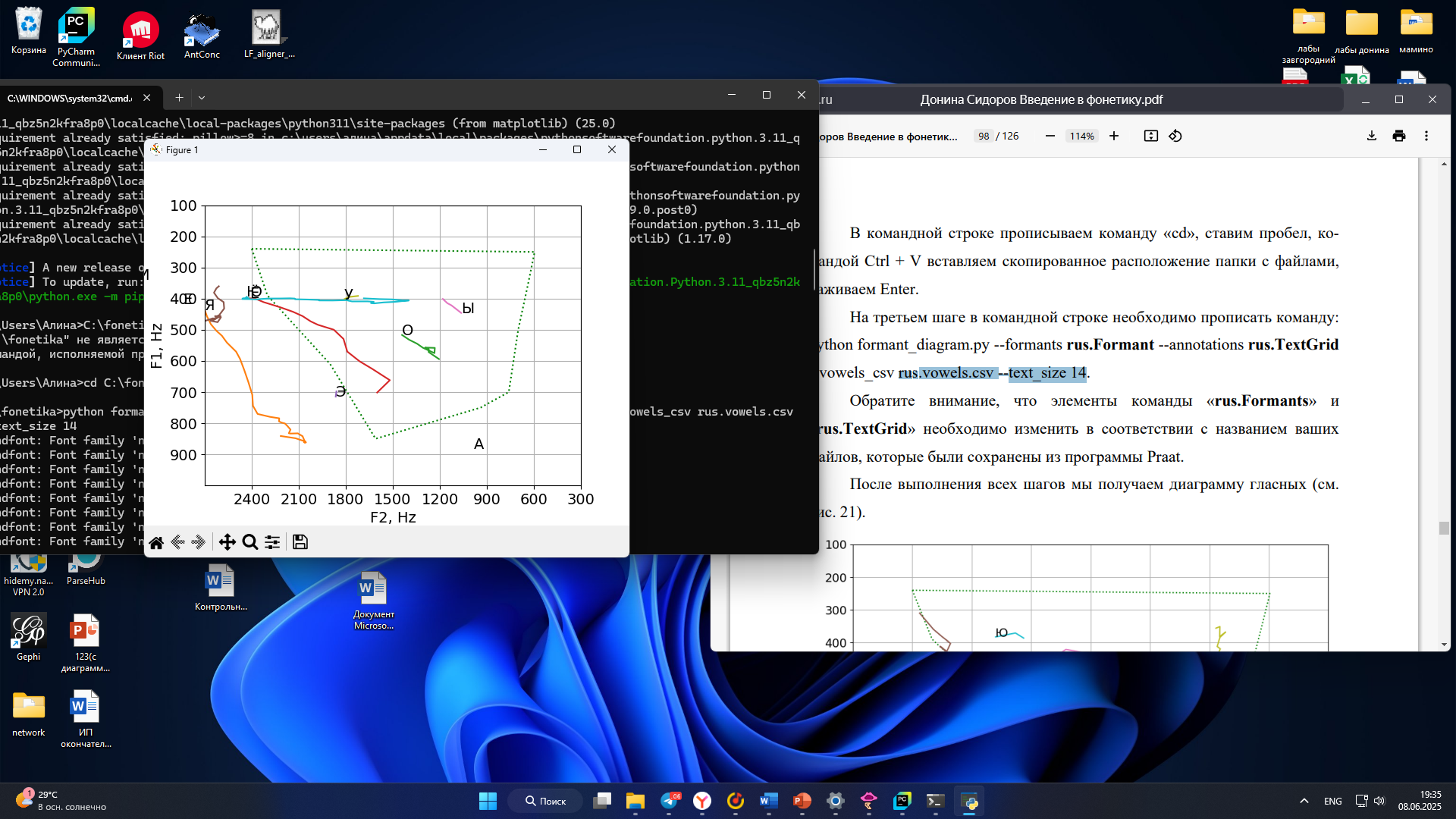


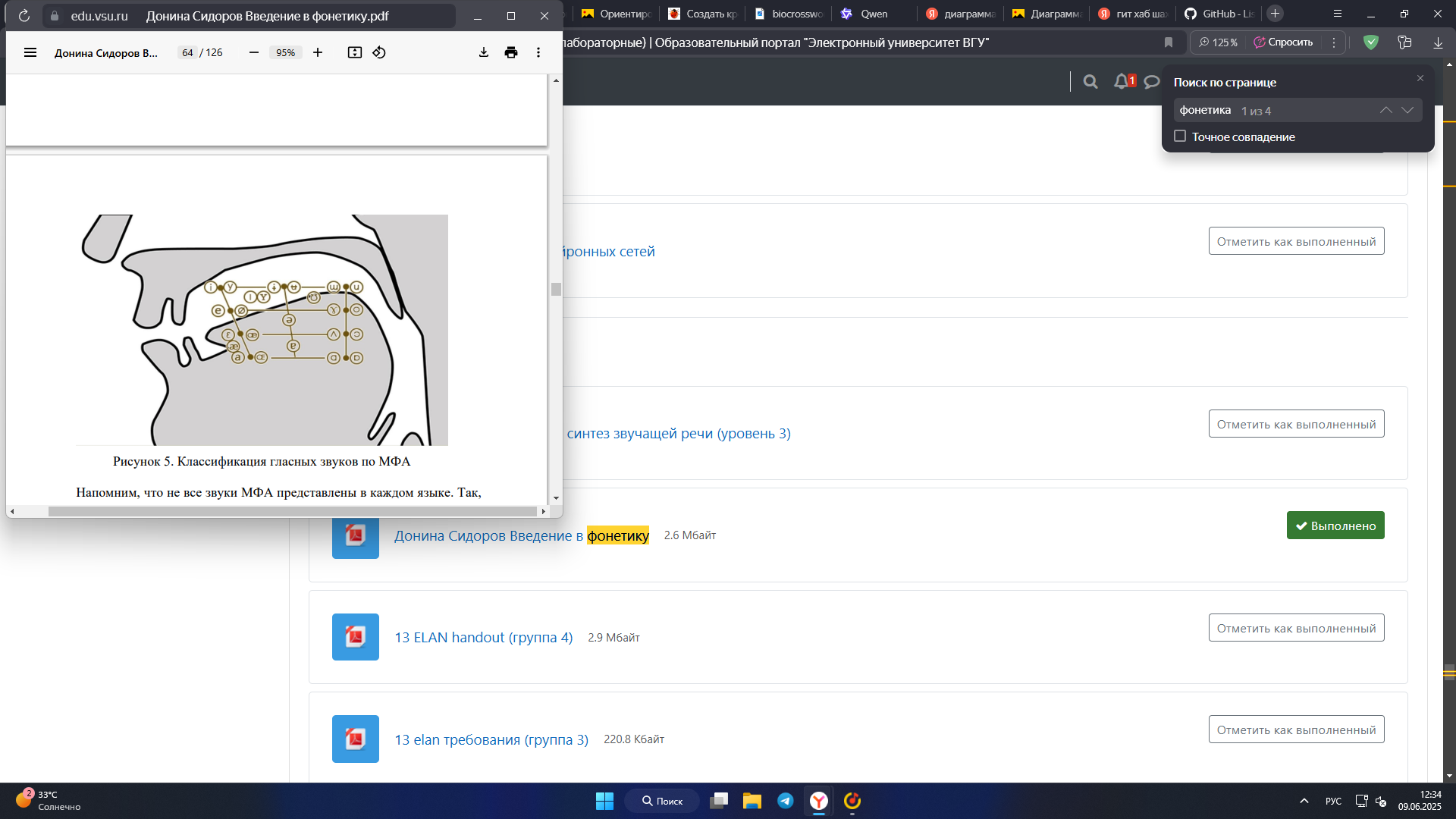






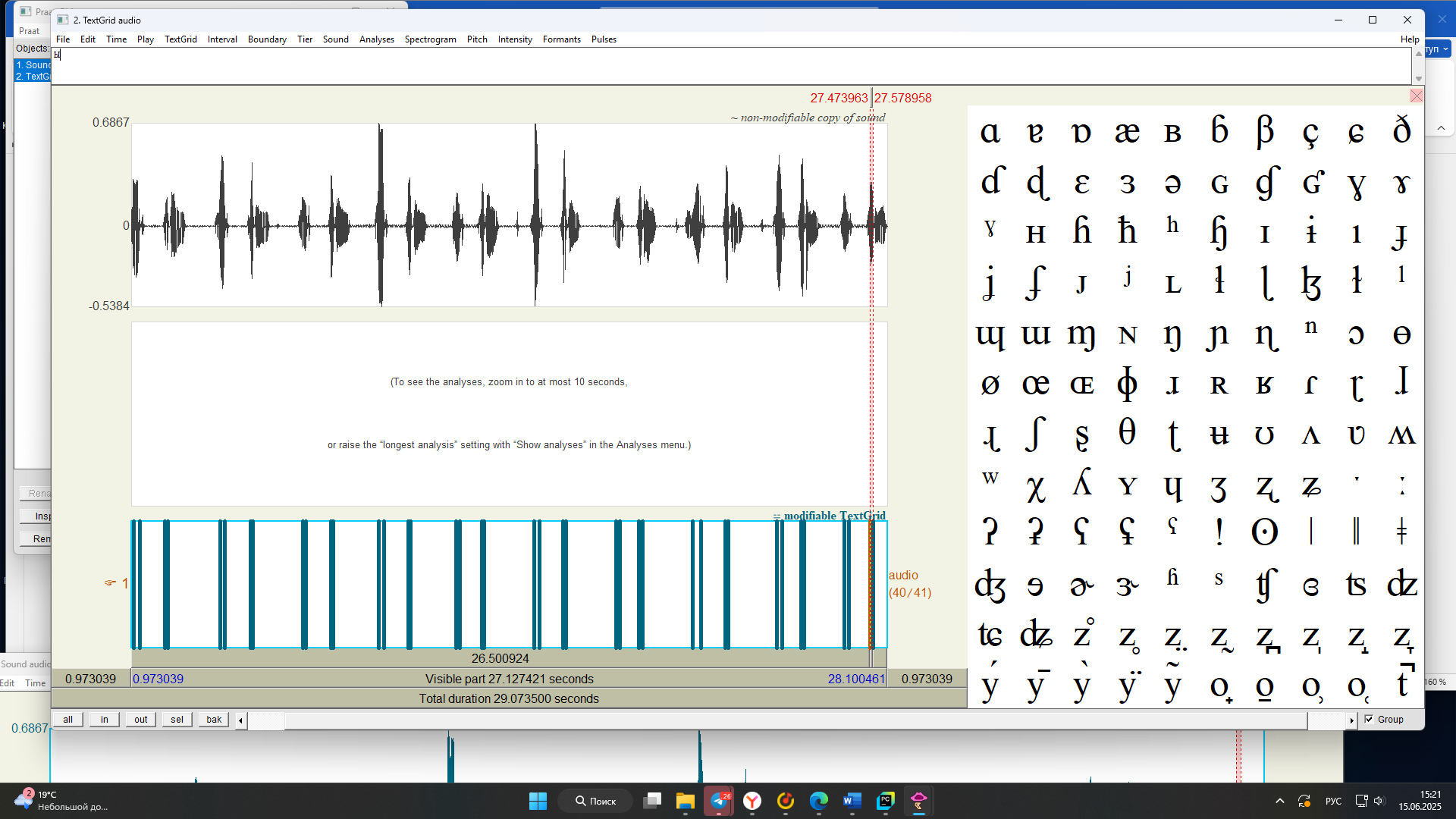


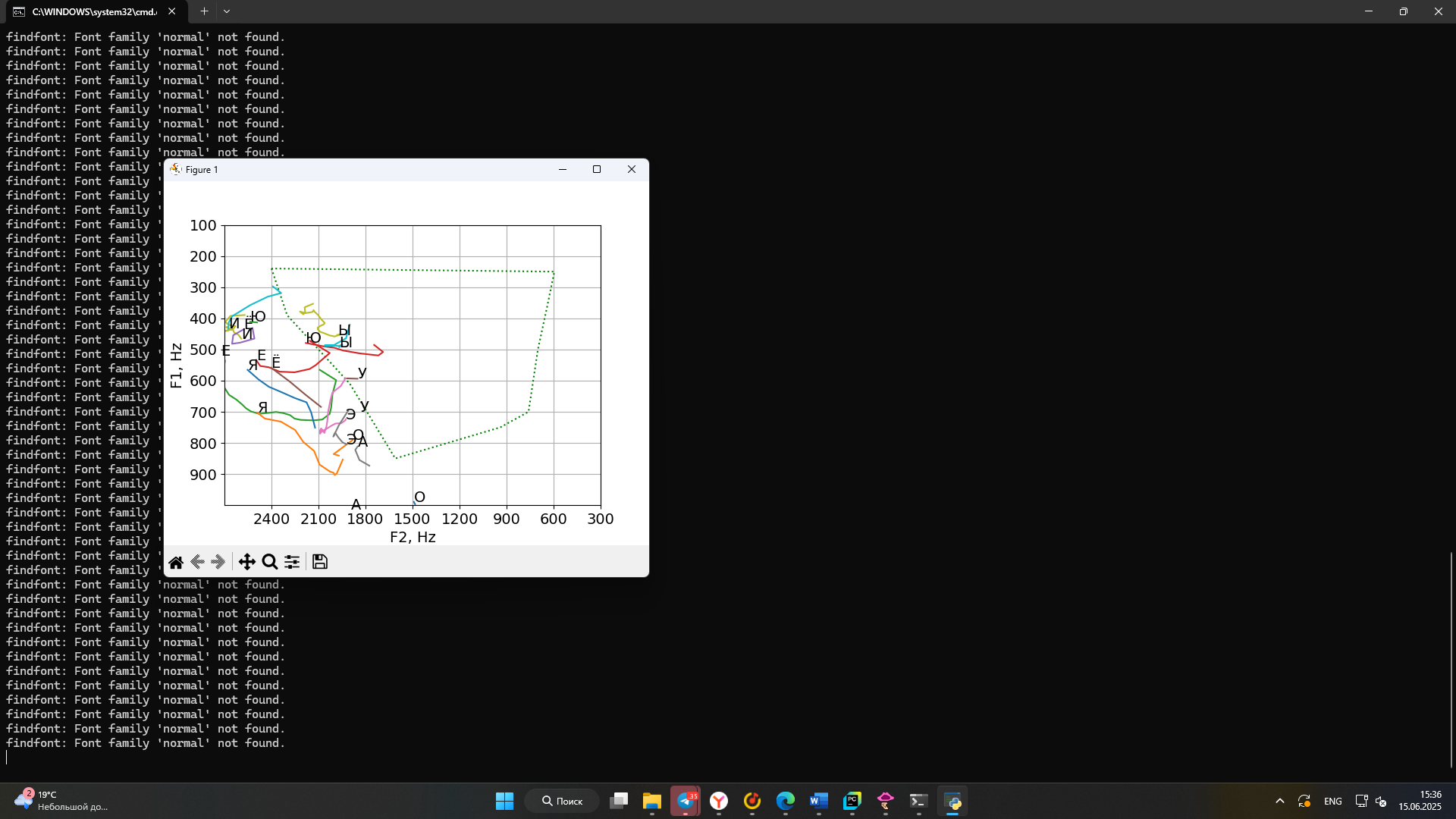




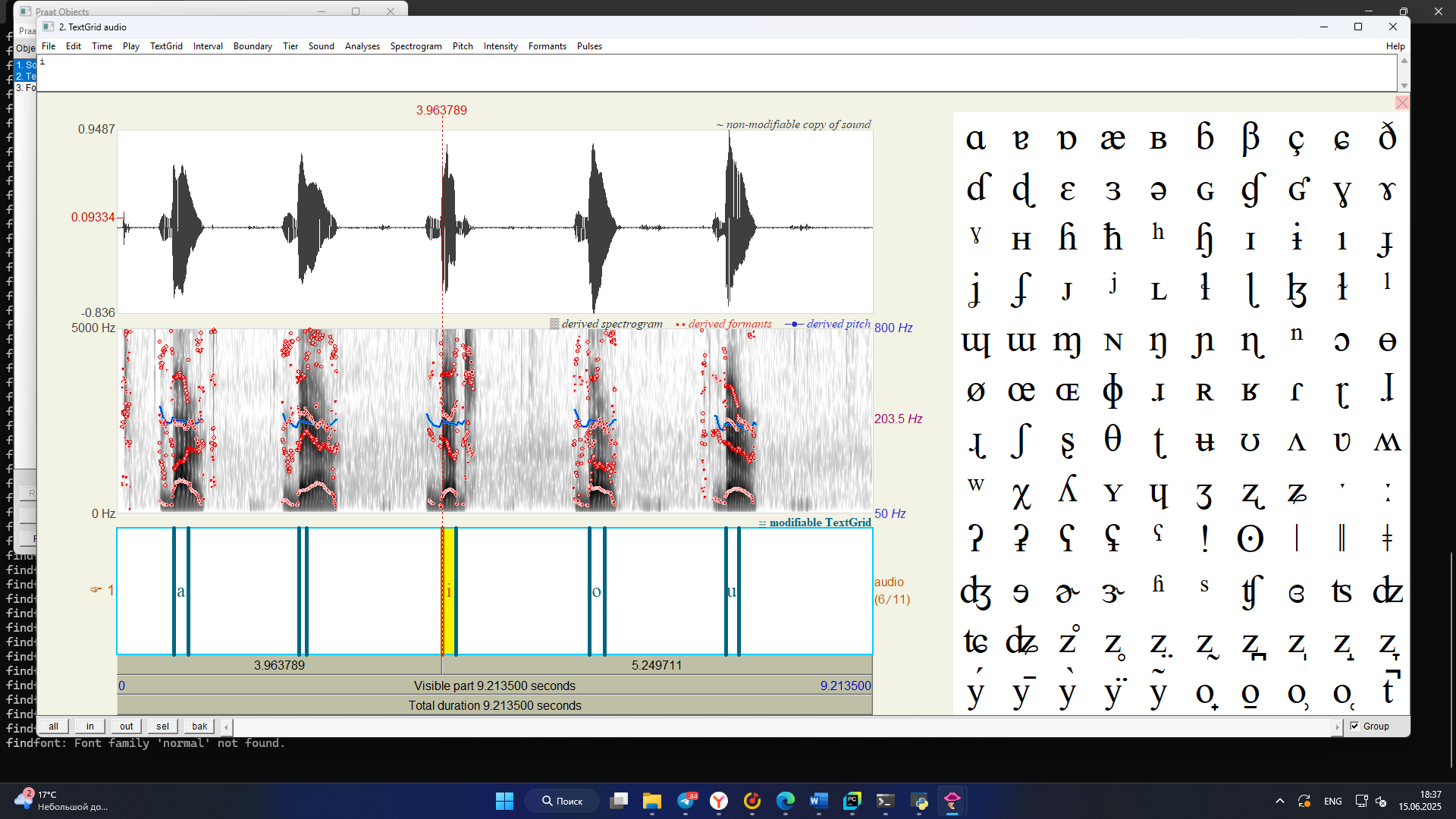
1. Проведите исследование качественной редукции русских гласных звуков. Запишите ряды словоформ, где один и тот же гласный будет стоять в ударной, предударной и «прочей» позиции. Создайте диаграмму гласных и проанализируйте полученные результаты.

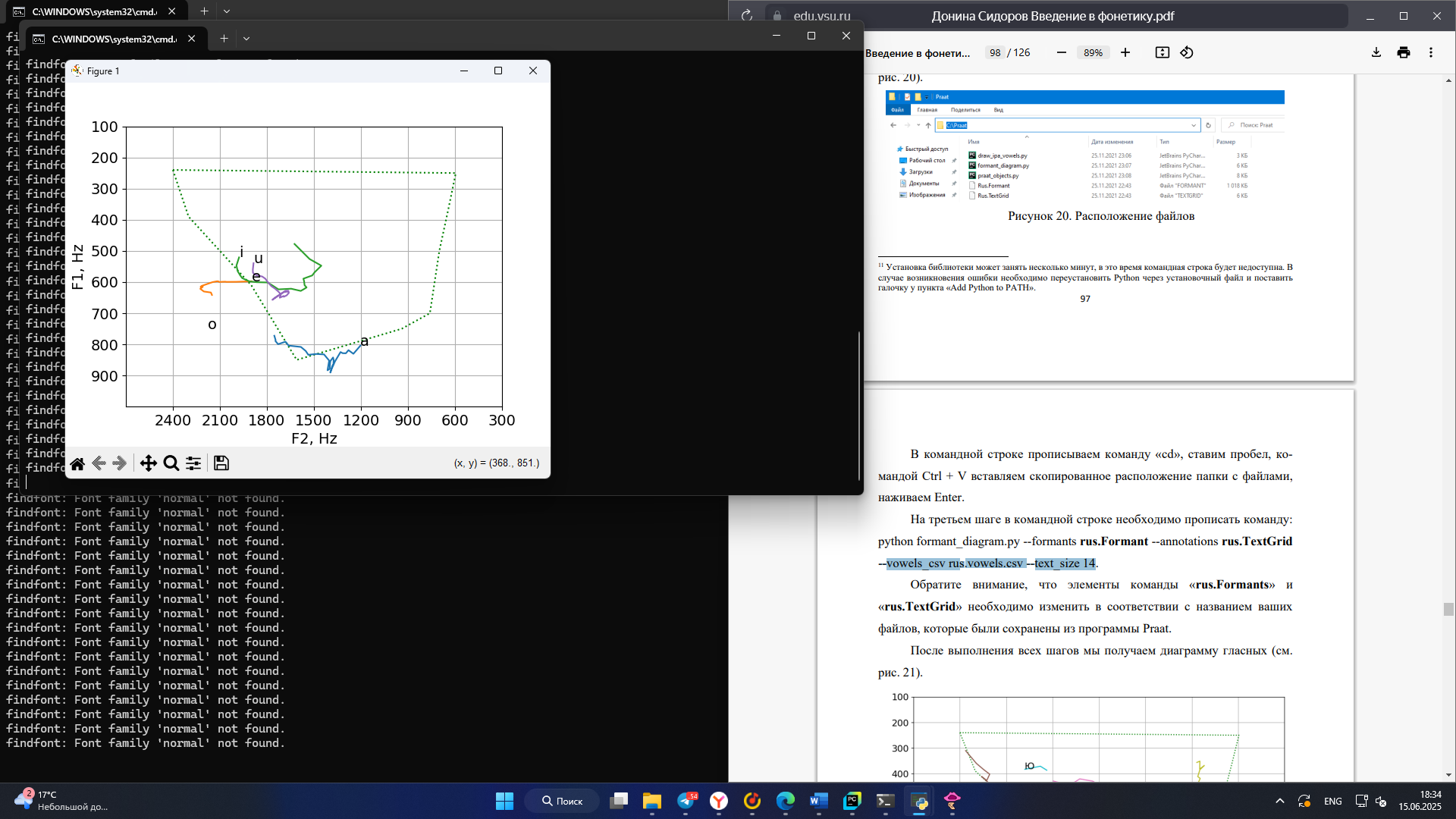
Использовались такие словоформы: «нОт»-«нотЫ», «нЕт»-«нетЫ», «нУт»-«нутЫ», «нАт»-«натЫ», «нИт»-«нитЫ», «нЯт»-«нятЫ», «нЮт»-«нютЫ», «нЁт»-«нётЫ», «нЭт»-«нэтЫ», «нЫт»-«нытЫ»





1. Запишите речевой фрагмент с гласными звуками иностранного языка. Внимательно выбирайте окружение для каждого конкретного гласного, желательно, одни и те же согласные. Обратите внимание на особенности гласных в языковой системе: редуцированные/нередуцированные, долгие/короткие, губные/негубные наличие дифтонгов и т.д. Сделайте диаграмму гласных, сопоставьте ее с диаграммой русских гласных, проанализируйте полученные результаты.





**Ответы на вопросы по теме «Основы работы систем преобразования текста в речь»**

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

Основная задача систем преобразования текста в речь (TTS – Text-ToSpeech) – создание голосового сообщения без участия человека напрямую

1. По каким критериям оценивают сгенерированную речь?

Основные критерии для оценки качества синтеза речи: разборчивость, т.е. возможность человеком декодировать речевой сигнал, и естественность – автоматически созданное высказывание не должно значительно отличаться от речи человека

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

В основе компилятивного синтеза речи заложены заранее записанные сегментные единицы (от звуков речи до целых высказываний), т.е. на вход системы подаются подготовленные человеком образцы речевых сигналов, которые система объединяет и озвучивает в требуемой последовательности.

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

Формантный (синтез речи происходит путём моделирования формант — частотных полос, характеризующих речевые звуки (особенно гласные). Форманты определяют тембр и тип звука, и они напрямую связаны с акустическими характеристиками сигнала.) и статистический (речь строится на основе статистических моделей , обученных на реальных речевых данных. Эти модели предсказывают акустические параметры (спектр, длительность, просодию и т. д.) на основе текста.)

1. В чем заключается особенность артикуляционного способа синтеза речи?

Артикуляционный синтез основывается на создании модели человеческого речевого аппарата и ее реализации с помощью некоторого механизма – «говорящей» машины, которая порождает речевой сигнал путем артикуляции

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

* Компилятивный: Если записанные человеком фрагменты речи достаточно велики

(например, целые высказывания), подобный синтезатор не сможет воспроизвести множество разнообразных высказываний; в противном случае, при записи минимальных сегментных единиц (звуков речи), возможности системы гораздо выше, но возникают проблемы с разборчивостью и естественностью речевого сигнала на выходе. Для разрешения проблемы использования минимальных сегментных единиц, исследователи пришли к использованию не отдельных фонов, а дифонов – речевых участков, которые берут свое начало в середине первого звука, а заканчиваются в середине второго звука. Использование дифонов

позволяет удачно синтезировать переходные участки между звуками, но при

этом возникают помехи (шумы) при дифонных соединениях вследствие перепадов частот разных сегментов, что вызывает необходимость в дополнительной обработке. Распространено и использование интонационных модификаций для определенных участков синтезируемой речи, например – конец повествовательно высказывания произносится с понижающейся тональностью, а вопросительные – с восходящей.

* Формантный: Трудности с передачей естественной интонации и сложных согласных.
* Статистический: Менее естественная речь.
* Артикуляционный: Моделирование физических процессов требует мощных вычислений: решение уравнений аэродинамики, упругости тканей, взаимодействия воздуха и мягких тканей. Нет полной и точной базы знаний о том, как именно двигаются артикуляторы при произнесении каждого звука.

**Практическое задание**

Проанализируйте как минимум четыре синтезатора речи. Оцените разборчивость и естественность сгенерированного речевого сигнала. Проверьте, учитывают синтезаторы речи суперсегментные характеристики речи – интонацию (паузы при перечислении, интонирование в восклицательных и вопросительных предложениях) и ударение (способен ли отличить одинаковые словоформы с разными ударными слогами (пример: Все зависит от того, к чему у тебя лежит душа′; После принятия ду′ша не стоит сразу выходить на улицу)). Для анализа можете использовать синтезаторы речи, внедренные в онлайн переводчики (Яндекс Переводчик, Google Translation), речь голосовых помощников (Google Assistant, Яндекс Алиса, Siri, Alexa), а также следующие онлайн синтезаторы речи:

• Acapela: https://www.acapela-group.com/

• Oddcast: https://ttsdemo.com/

• Text-to-speech: https://text-to-speech.imtranslator.net/speech.asp

• PiliApp: https://ru.piliapp.com/text-to-speech/

• iSpeech: https://www.ispeech.org/text.to.speech

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| условие | Acapela | Oddcast | PiliApp | Яндекс Переводчик |
| Восклицательное предложение | Делает акцент на восклицание, но не сильный | Нет акцента на восклицание | Нет акцента на восклицание | Нет акцента на восклицание |
| Вопросительное предложение | Произносит с вопросительной интонацией | Произносит с вопросительной интонацией | Произносит с вопросительной интонацией | Не соблюдает вопросительную интонацию |
| Пунктуация | Соблюдает паузы | Соблюдает паузы | Соблюдает паузы | Соблюдает паузы |
| ударение | Не отличает одинаковые словоформы с разными ударными слогами | Различает одинаковые словоформы с разными ударными слогами | Не отличает одинаковые словоформы с разными ударными слогами | Различает одинаковые словоформы с разными ударными слогами |
| естественность | 7/10 | 3/10 | 2/10 | 2/10 |
| разборчивость | 10/10 | 10/10 | 10/10 | 9/10 |