

Министерство цифрового развития

Федеральное государственное бюджетное образовательное учреждение высшего
образования

«Сибирский государственный университет телекоммуникаций и
информатики»
(СибГУТИ)

Кафедра прикладной математики и кибернетики

Отчёт

по лабораторной работе № 7
«Распознавание голосовых сообщений с использованием
готовой модели Whisper от OpenAI»

Выполнил:

студент группы ИП-312

Дорогин Н.С.

Работу проверил: старший преподаватель
кафедры ПМиК
Дементьева К.И.

Новосибирск 2025 г.

Введение (задание)

Цель: освоить на практике применение предобученных моделей для транскрибации аудио, оценки качества распознавания и визуализации результатов.

Задание:

Выбираю для работы свой набор голосовых сообщений (практически все из Telegram).

Задание 1. Подготовка данных:

1. Устанавливаем необходимые библиотеки.

Гуглите как установить Whisper и что может помочь в анализе файлов (librosa).

```
PS C:\Users\Danik\Desktop\STUDY_SIBSUTIS_DOROGIN\MMO\Лаба7> pip install librosa
Collecting librosa
  Downloading librosa-0.11.0-py3-none-any.whl.metadata (8.7 kB)
Collecting audioread>=2.1.9 (from librosa)
```

```
PS C:\Users\Danik\Desktop\STUDY_SIBSUTIS_DOROGIN\MMO\Лаба7> pip install openai-whisper
Collecting openai-whisper
  Downloading openai_whisper-20250625.tar.gz (803 kB)
                                         803.2/803.2 kB 5.0 MB/s  0:00:00
    Installing build dependencies ... done
    Getting requirements to build wheel ... done
    Preparing metadata (pyproject.toml) ... done
```

Так же для работы Whisper надо установить FFmpeg. Делаем по гайду с интернета.



Установка FFmpeg на Windows

Эта инструкция поможет установить **FFmpeg** — утилиту, необходимую для обработки видео и аудио, в том числе для работы таких ИИ-инструментов, как OpenAI Whisper.

[Показать оглавление](#)



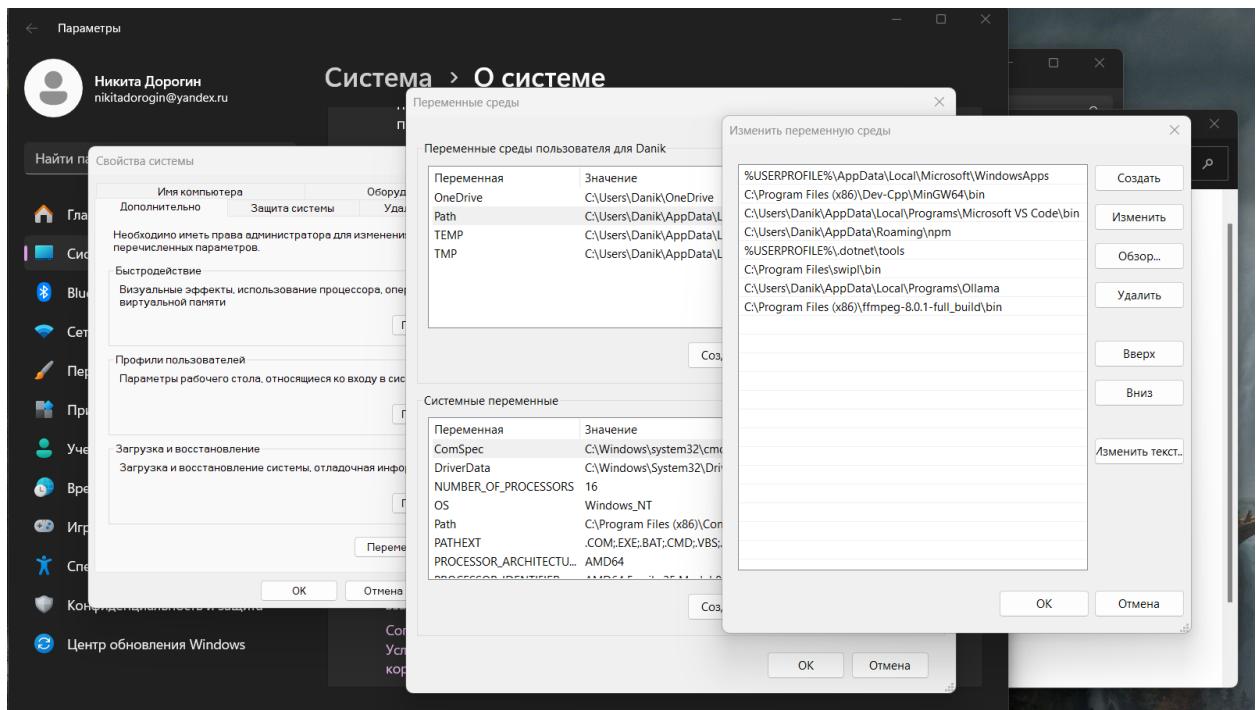
Шаг 1. Скачивание FFmpeg

Перейдите на официальный сайт сборок FFmpeg:

👉 <https://www.gyan.dev/ffmpeg/builds/>

В разделе `release builds` скачайте:

- `ffmpeg-release-full.7z` — полная сборка с максимальной поддержкой кодеков



2. Загружаем наши аудио.

Я заранее преобразовал их в wav через онлайн-конвертер:

Файл добавлен! Запустите задачу или добавьте еще файлы

Добавить еще файлы Настройки НАЧАТЬ →

START → Θ 1 / ⏱ 30 s + ДОБАВИТЬ ПРИМЕР ФАЙЛА

MyMes1.ogg 141.64 KB X

MyMes2.ogg 125.55 KB X

MyMes3.ogg 87.73 KB X

Дополнительные настройки Конвертер WAV

Sounds

- Confess1.wav
- Confess2.wav
- KatyaMes.wav
- MomMes.wav
- MyMes1.wav
- MyMes2.wav
- MyMes3.wav
- MyMes4.wav
- PeasantMes.wav
- RomMes.wav

~WRL0628.tmp
main.py
Лаб7. Дорогин Никита ИП-312.docx
Лабораторная 7. Транскрибация аудиозаписей с использов...

Задание 2. Транскрибация аудио с помощью модели Whisper:

Создаём дефолтную модель и прогоняем через неё наши аудио по одному и тому же сценарию:

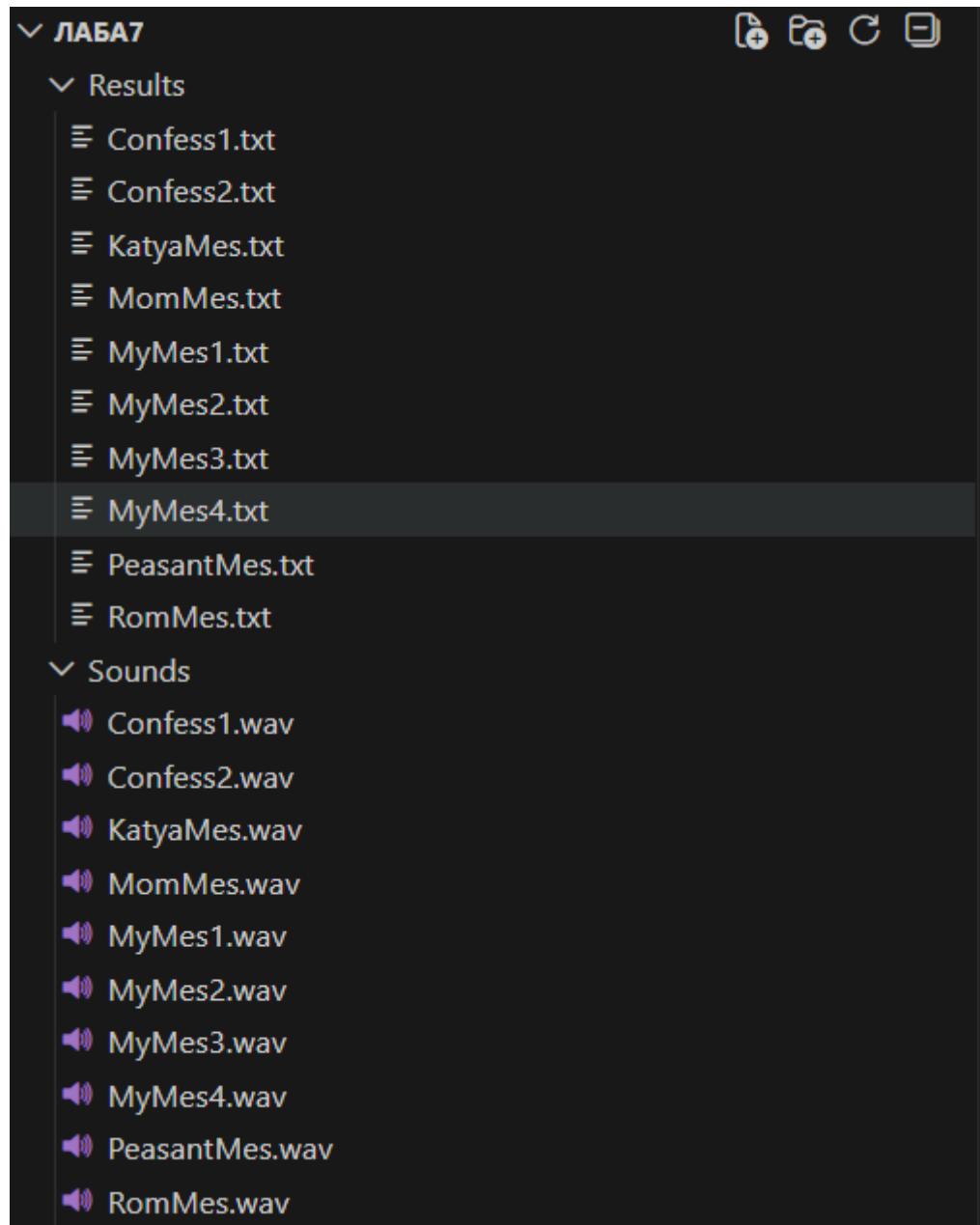
```
model = whisper.load_model("base")

MyMes1 = model.transcribe(".\Sounds\MyMes1.wav")
MyMes1_file = open(".\Results\MyMes1.txt", "w")
MyMes1_file.write(MyMes1["text"])
MyMes1_file.close()
```

При работе есть предупреждение, но оно не критичное:

```
in.py"
C:\Users\Danik\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.11_qbz5n2kfra8p0\LocalCache\local-packages\Python311\site
-packages\whisper\transcribe.py:132: UserWarning: FP16 is not supported on CPU; using FP32 instead
    warnings.warn("FP16 is not supported on CPU; using FP32 instead")
```

Результат:



A code editor window is shown with the tab 'Confess1.txt' selected. The content of the file is: 'Results > Confess1.txt'. Below the tabs, there is a status bar with the number '1' and a series of question marks and diamonds.

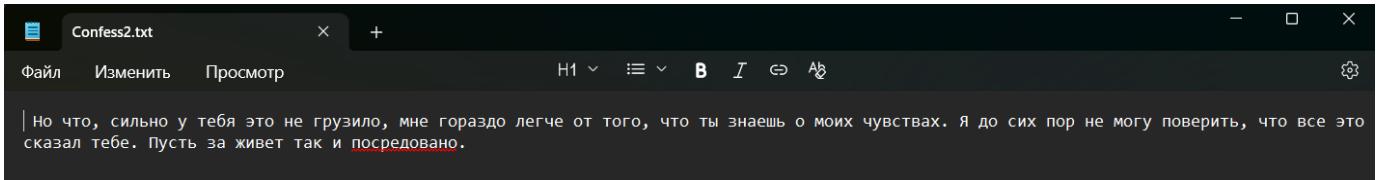
Упс, кажется, кто-то не знает русский язык. Пойдём гуглить как это исправить.

А, нет, не пойдём, я понял.

A screenshot of the Visual Studio Code interface. The title bar says 'Confess1.txt'. The editor shows the same text as the previous image: 'Results > Confess1.txt'. The status bar at the bottom indicates the file is 1 page long. A message in the status bar reads: 'Обидно, что за эти полтора года я не смог заставить себя поговорить с тобой о чем-то кроме учёбы Обидно, что за полтора года нашего знакомства я тебе практически ничего не знаю и мы считаем друг другу чужие люди То есть ну ну вообще я ведь даже не могу назвать себя твоим другом и мне очень за это обидно'.

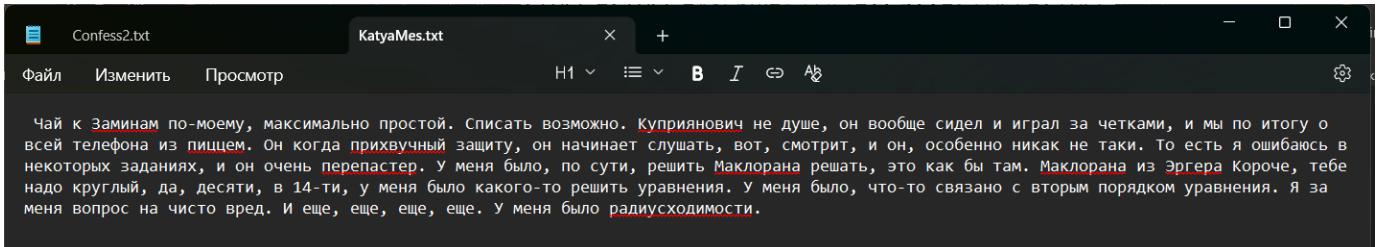
У меня VS Code не поддерживает русский.

Всё, тогда всё работает.

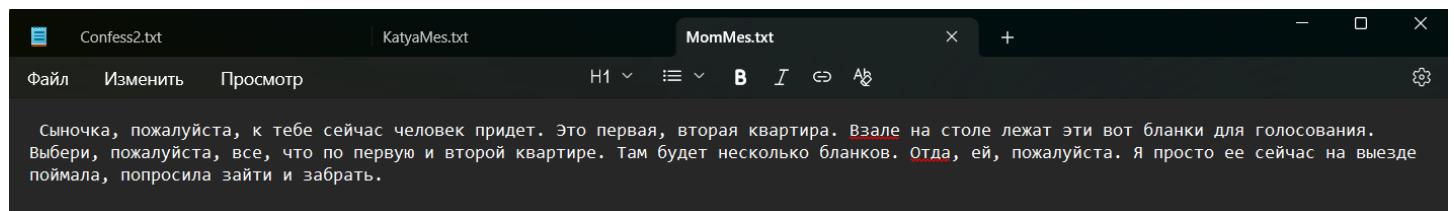


В двух сообщениях выше голос был громким, но его немного приглушал ветер – на улице была метель.

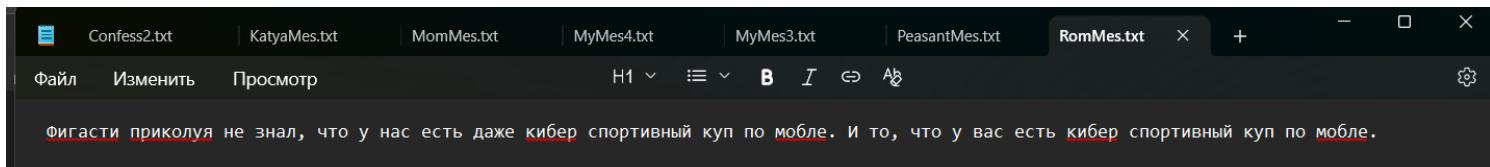
Это сообщение было очень тихим:



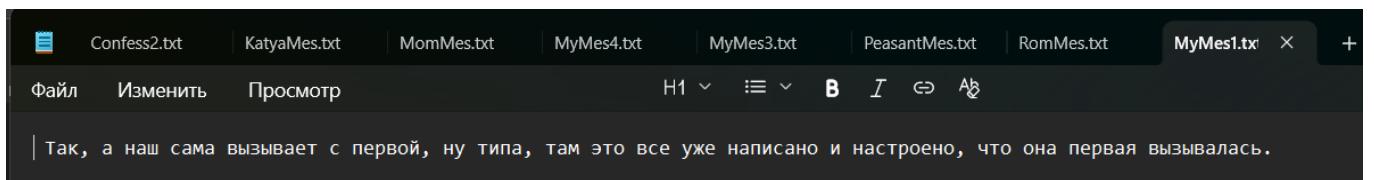
Это сообщение было записано с нормальной громкостью в помещении.
Плюс оно единственное из WhatsApp:

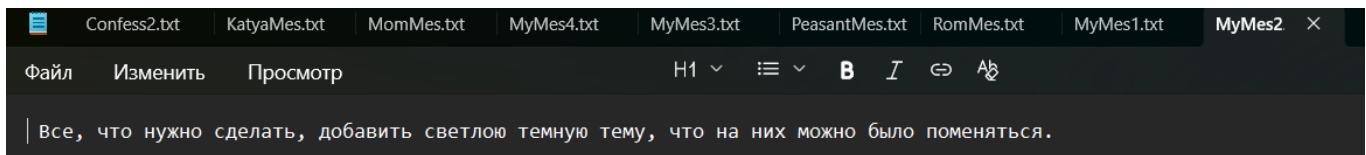


Это тоже без посторонних шумов:

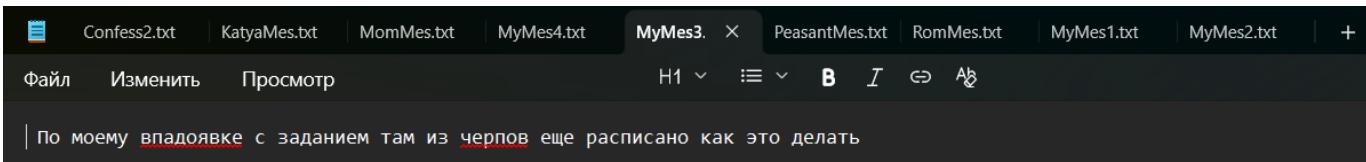


Следующие 4 сообщения были записаны на улице, но в спокойную погоду. На фоне мог быть разве что звук от проезжающих машин.

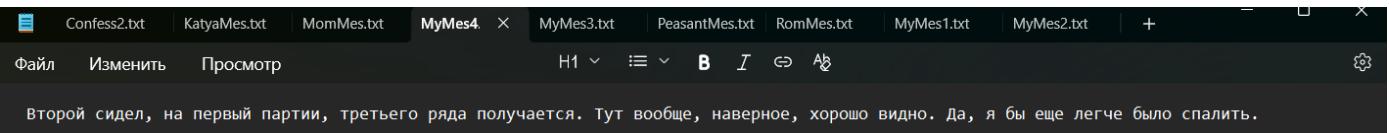




Confess2.txt | KatyaMes.txt | MomMes.txt | MyMes4.txt | MyMes3.txt | PeasantMes.txt | RomMes.txt | MyMes1.txt | **MyMes2** X
Файл Изменить Просмотр H1 ≡ B I ↴ ↵
| Все, что нужно сделать, добавить светлою темную тему, что на них можно было поменяться.

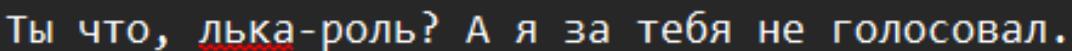


Confess2.txt | KatyaMes.txt | MomMes.txt | MyMes4.txt | **MyMes3** X | PeasantMes.txt | RomMes.txt | MyMes1.txt | MyMes2.txt | +
Файл Изменить Просмотр H1 ≡ B I ↴ ↵
| По моему впадоявке с заданием там из черлов еще расписано как это делать



Confess2.txt | KatyaMes.txt | MomMes.txt | **MyMes4** X | MyMes3.txt | PeasantMes.txt | RomMes.txt | MyMes1.txt | MyMes2.txt | +
Файл Изменить Просмотр H1 ≡ B I ↴ ↵
| Второй сидел, на первый партии, третьего ряда получается. Тут вообще, наверное, хорошо видно. Да, я бы еще легче было спалить..

И последнее было записано довольно тихо. Я пытался звук с ноута записать на телефон и вышло паршивенько, но модель транскрибировала довольно неплохо.



| Ты что, лька-роль? А я за тебя не голосовал.

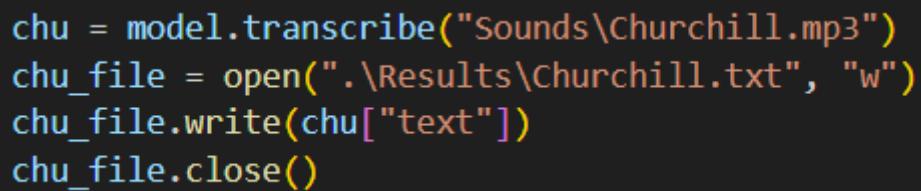
Также я побалуюсь за кадром и протестирую распознание нецензурной брани. Извините.

В первом файле он не распознал речь, потому что я слишком сильно кривил голосом и там не так просто распознать речь.

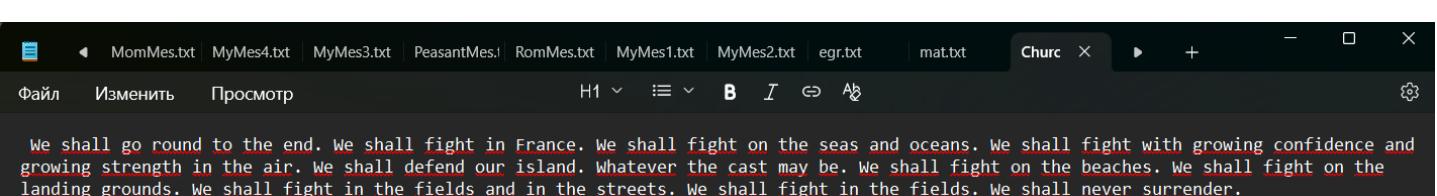
Второй он в принципе распознал правильно, ну, насколько правильно он ранее распознавал фразы.

Ну и напоследок пара экспериментов.

Попробую закинуть Черчилля в формате mp3.



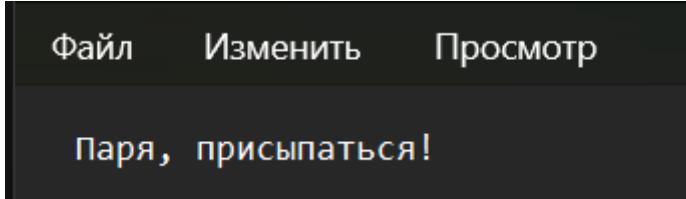
```
chu = model.transcribe("Sounds\Churchill.mp3")
chu_file = open(".\Results\Churchill.txt", "w")
chu_file.write(chu["text"])
chu_file.close()
```



MomMes.txt | MyMes4.txt | MyMes3.txt | PeasantMes.txt | RomMes.txt | MyMes1.txt | MyMes2.txt | egr.txt | mat.txt | **Churc** X | +
Файл Изменить Просмотр H1 ≡ B I ↴ ↵
| We shall go round to the end. We shall fight in France. We shall fight on the seas and oceans. We shall fight with growing confidence and growing strength in the air. We shall defend our island. Whatever the cost may be. We shall fight on the beaches. We shall fight on the landing grounds. We shall fight in the fields and in the streets. We shall fight in the fields. We shall never surrender.

Ну и последний эксперимент – закинуть сообщение в том же формате, что и в мессенджере.

```
wake = model.transcribe("Sounds\Wakeup.ogg")
wake_file = open(".\Results\Wakeup.txt", "w")
wake_file.write(wake["text"])
wake_file.close()
```



В принципе сработало.

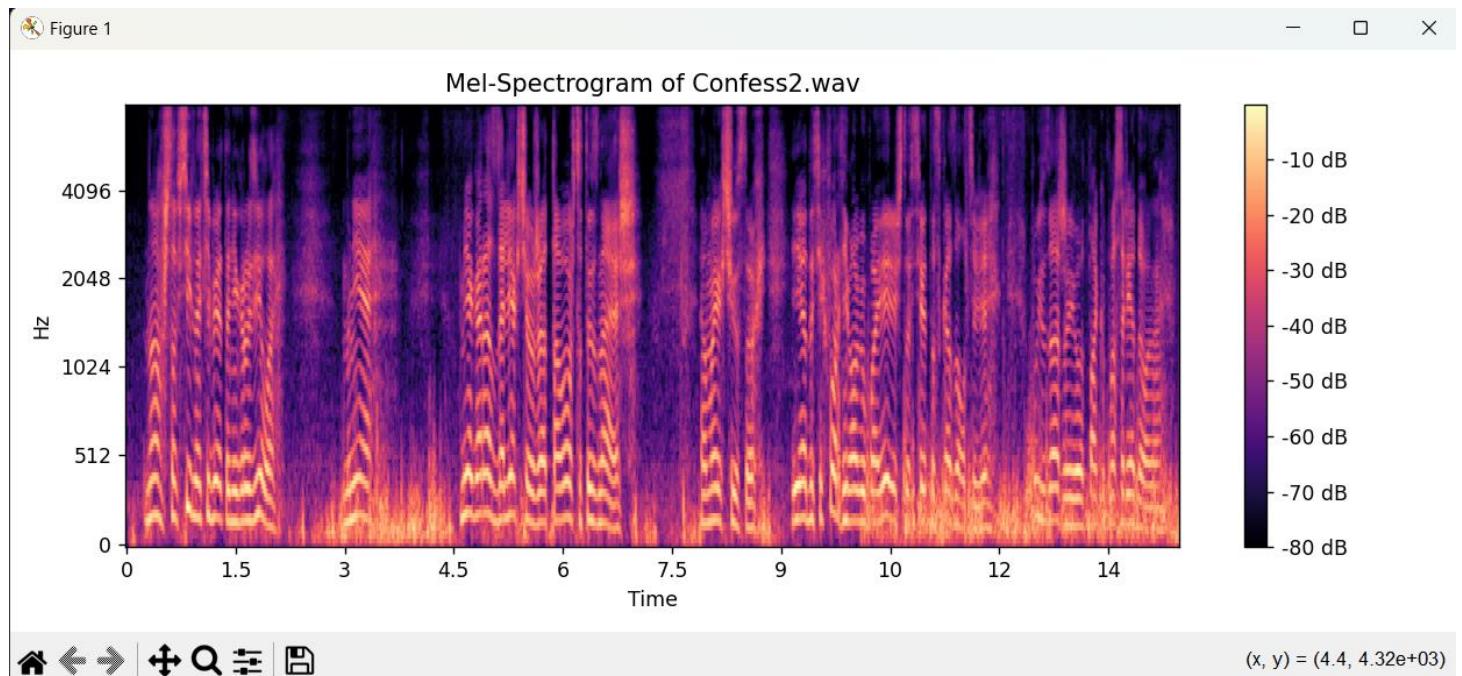
Задание 3. Визуализация и анализ результатов:

```
76 ✓ import librosa
77     import librosa.display
78     import matplotlib.pyplot as plt
79     import numpy as np
80
81     audio_path = r".\Sounds\Confess2.wav"
82     y, sr = librosa.load(audio_path, sr=16000)
83
84     print("Sample rate:", sr)
85     print("Length (sec):", len(y) / sr)
86
87 ✓ S = librosa.feature.melspectrogram(
88     y=y,
89     sr=sr,
90     n_fft=1024,
91     hop_length=256,
92     n_mels=128,
93     fmin=20,
94     fmax=8000
95 )
96
97     S_db = librosa.power_to_db(S, ref=np.max)
```

```

plt.figure(figsize=(10, 4))
librosa.display.specshow(
    S_db,
    sr=sr,
    hop_length=256,
    x_axis='time',
    y_axis='mel'
)
plt.colorbar(format='%+2.0f dB')
plt.title("Mel-Spectrogram of Confess2.wav")
plt.tight_layout()
plt.show()

```



Confess2.txt

Файл Изменить Просмотр

| Но что, сильно у тебя это не грузило, мне гораздо легче от того, что ты знаешь о моих чувствах. Я до сих пор не могу поверить, что все это сказал тебе. Пусть за живет так и посредовано.

Сложно сказать, почему транскрипция на последнем предложении дала сбой.

“Пусть даже и вот...” (не “Пусть за живет...”) в аудиофайле звучит достаточно чётко и лишь на слове опосредовано имеет место дуновение

ветра на фоне. Возможно, на моменте “вот так опосредованно” имело место небольшое ускорение речи, сделавшее её менее распознаваемой.

Отчёт:

GoogleCollab:

<https://colab.research.google.com/drive/1InYZSFY0xjGShodyJXLZvcIH-hd156SS#scrollTo=aF28ht0uim28>