

Sound Processing with pyaudio (microphone\_streaming\_with\_spectrum.py)

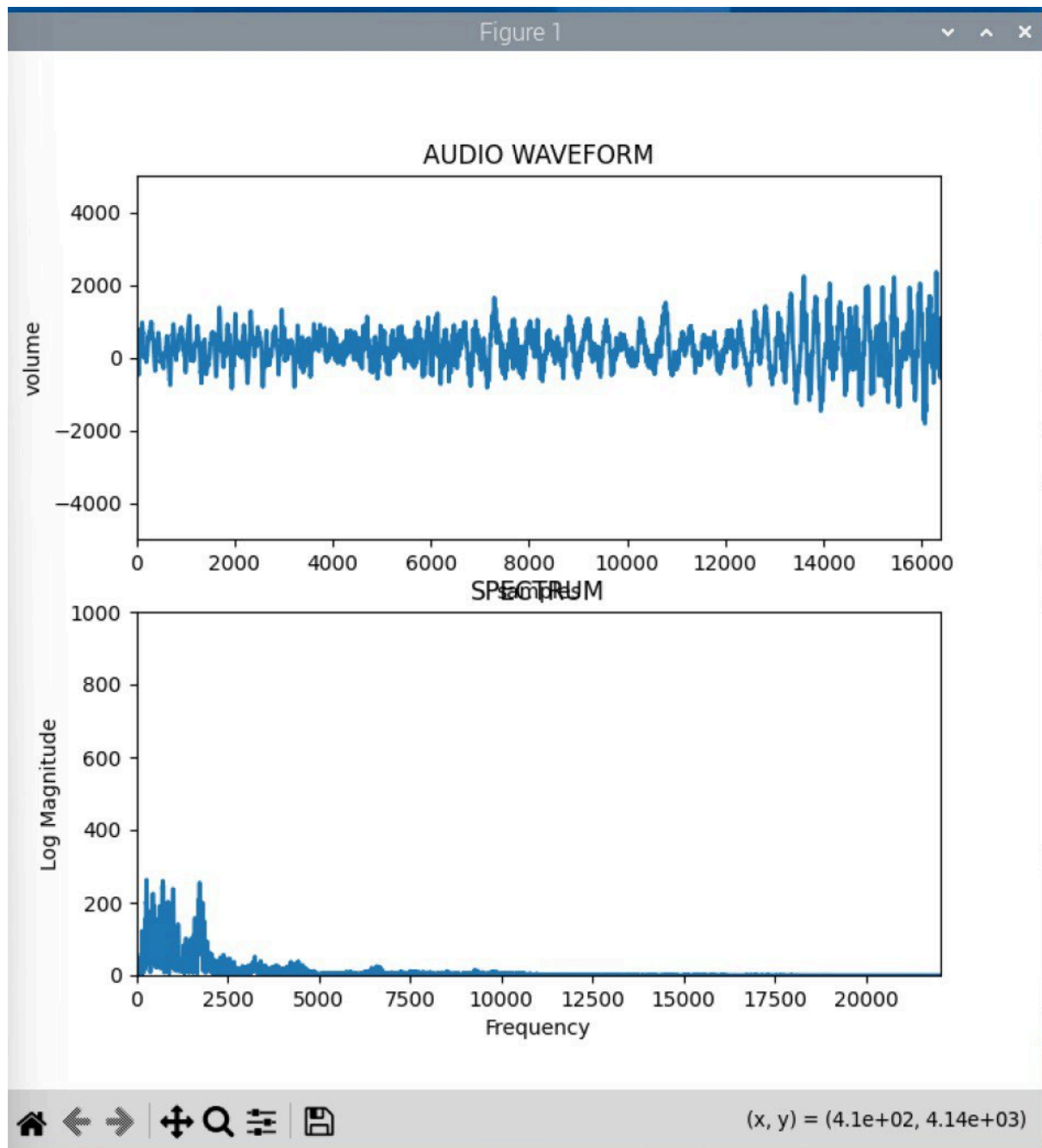


Fig 1. Sample Captured Speech using pyaudio

Speech will be captured through the microphone of the webcam, and visualised onto the audio waveform spectrum as seen in Figure 1.

## Basic Sound Analytics with pyaudio (filtering\_audio.py)

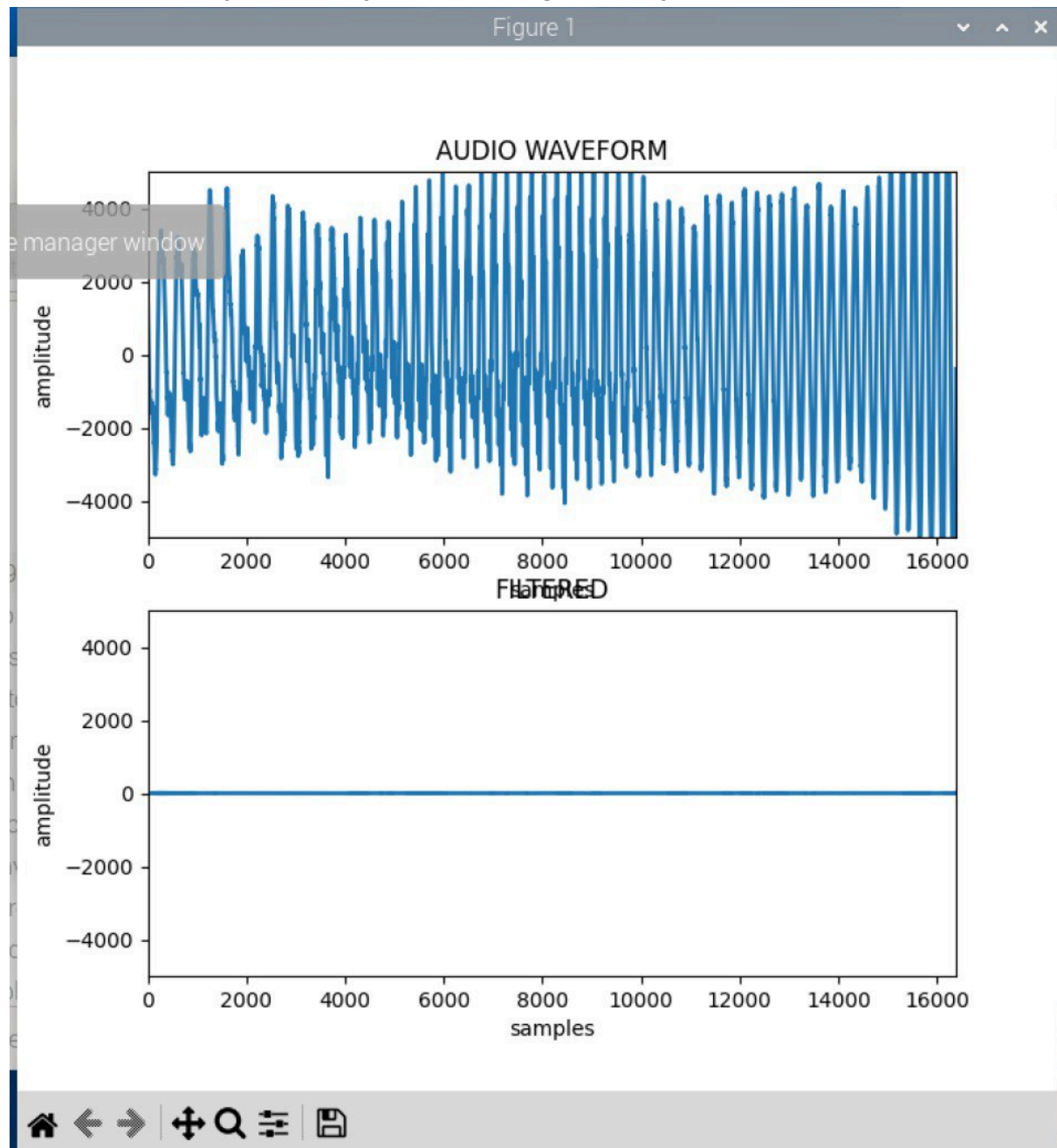


Fig 2. Filtered audio spectrum visualisation with pyaudio

The filtering removes noise and unwanted frequencies, which can be seen in Figure 2 whereby the filtered audio spectrum of our voice is at a amplitude of 0 due to the code specifying the frequency range of 19400Hz to 19600Hz.

## Librosa Spectrogram (audio\_features.py)

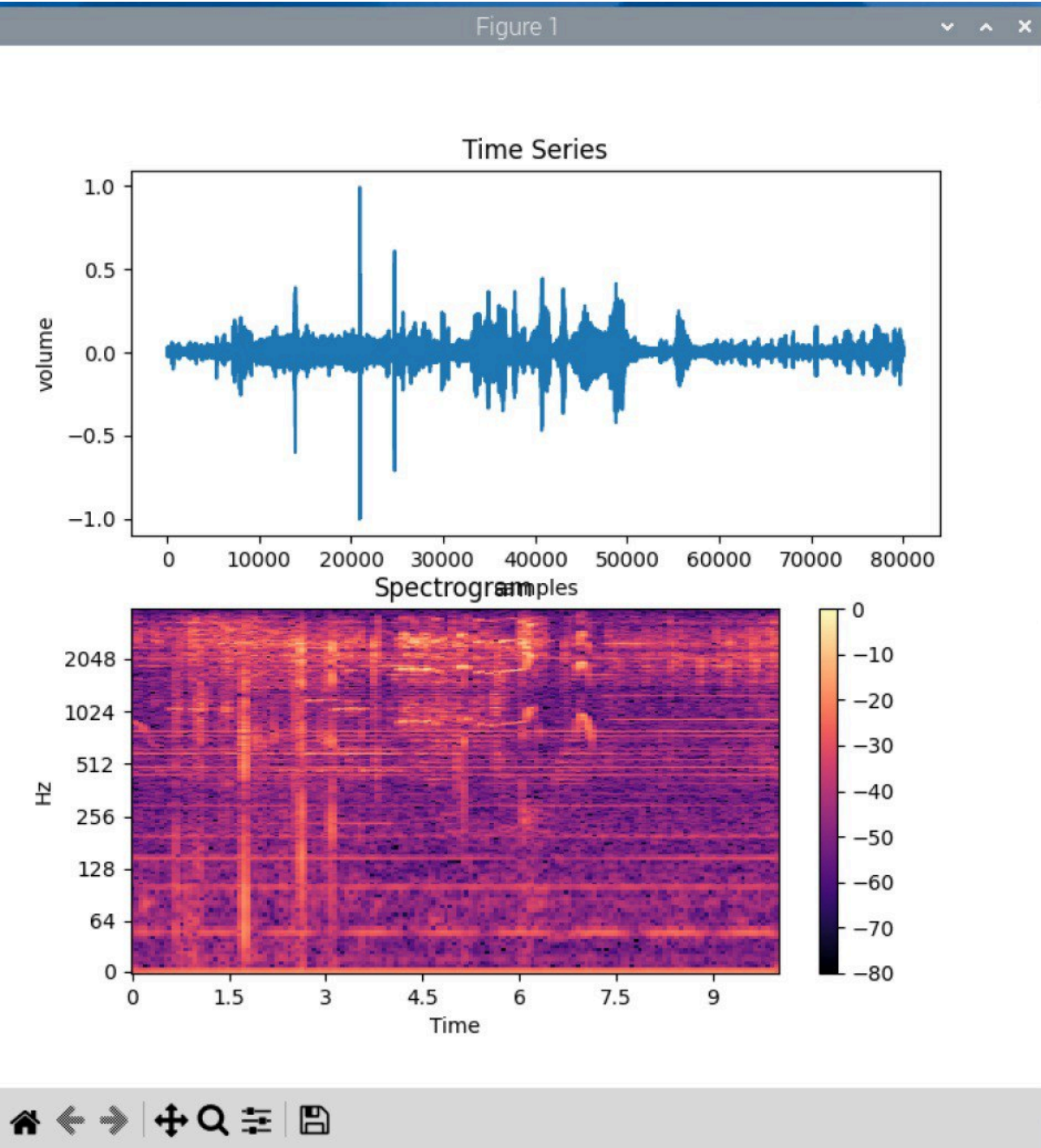


Fig 3. Spectrogram visualisation

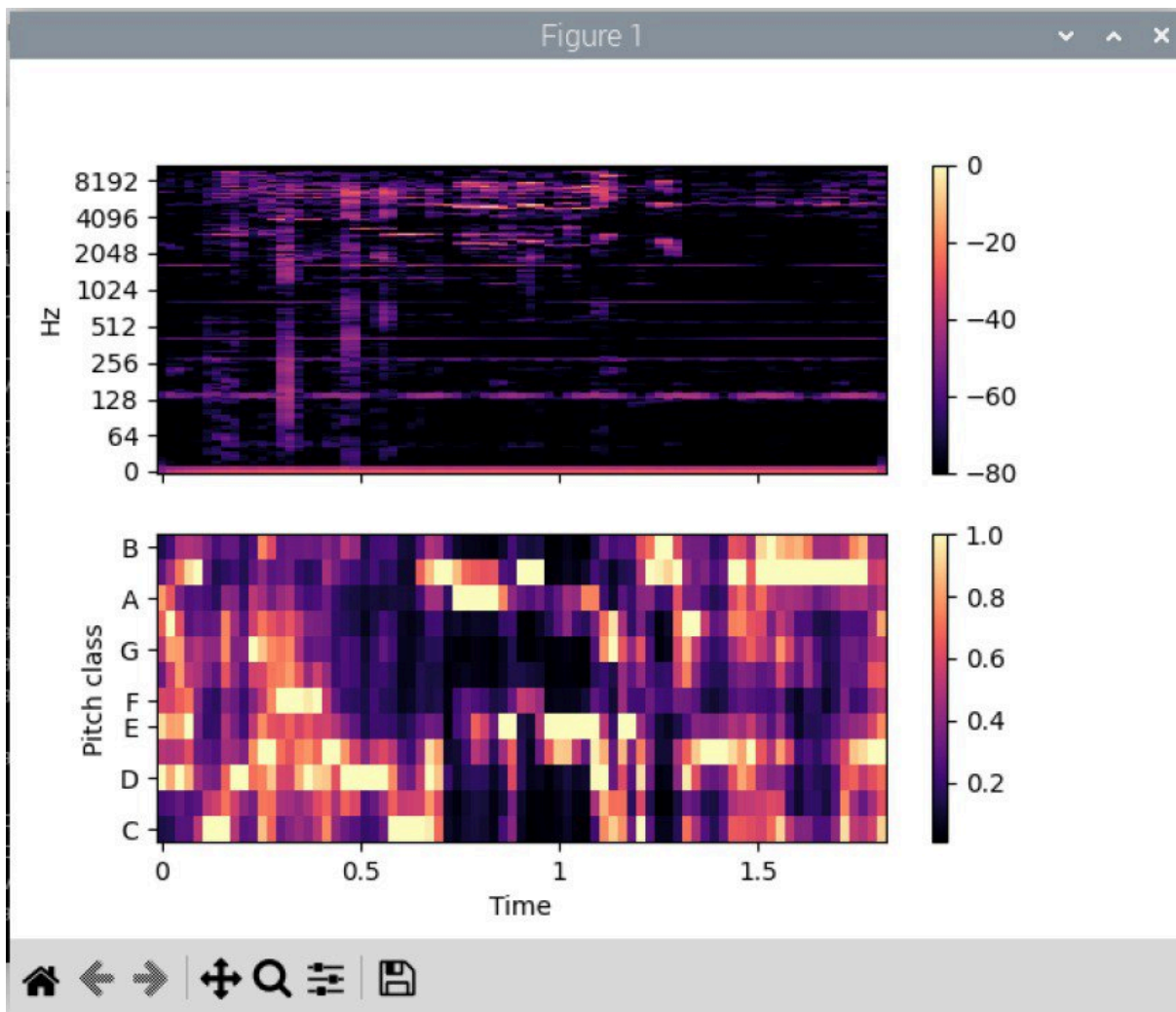


Fig 4. Chromogram visualisation

Figure 1

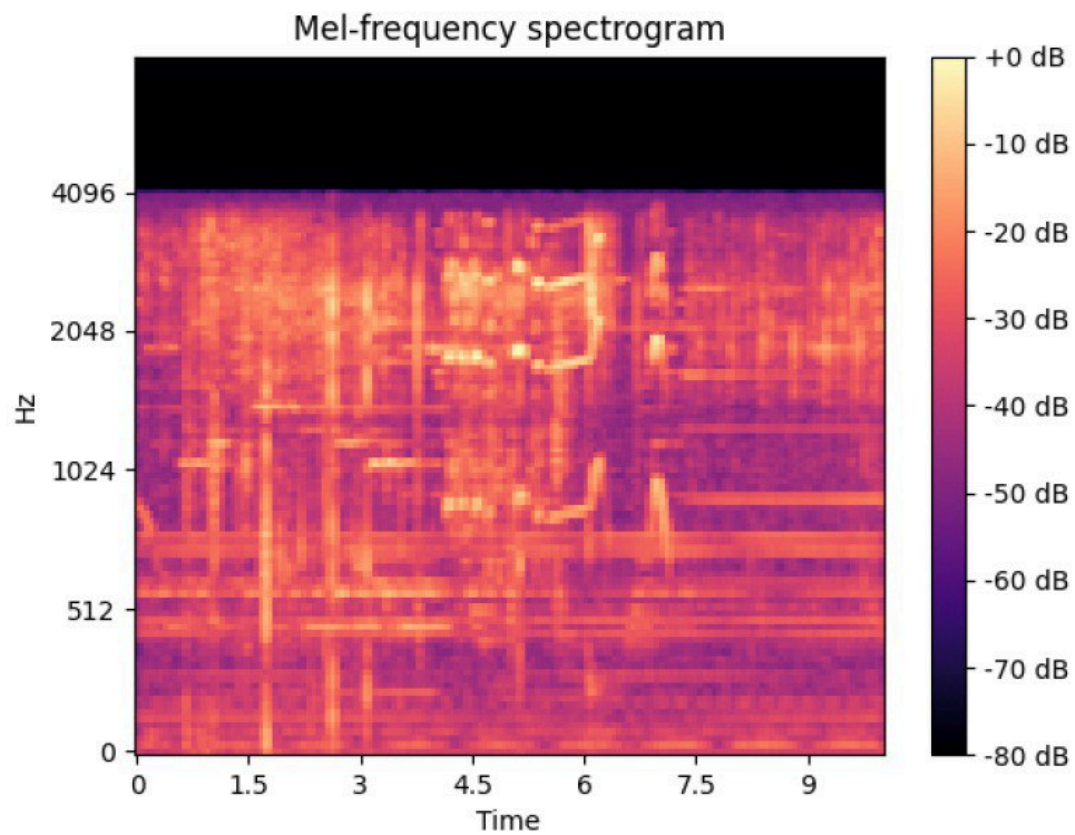


Fig 5. Mel-frequency spectrogram



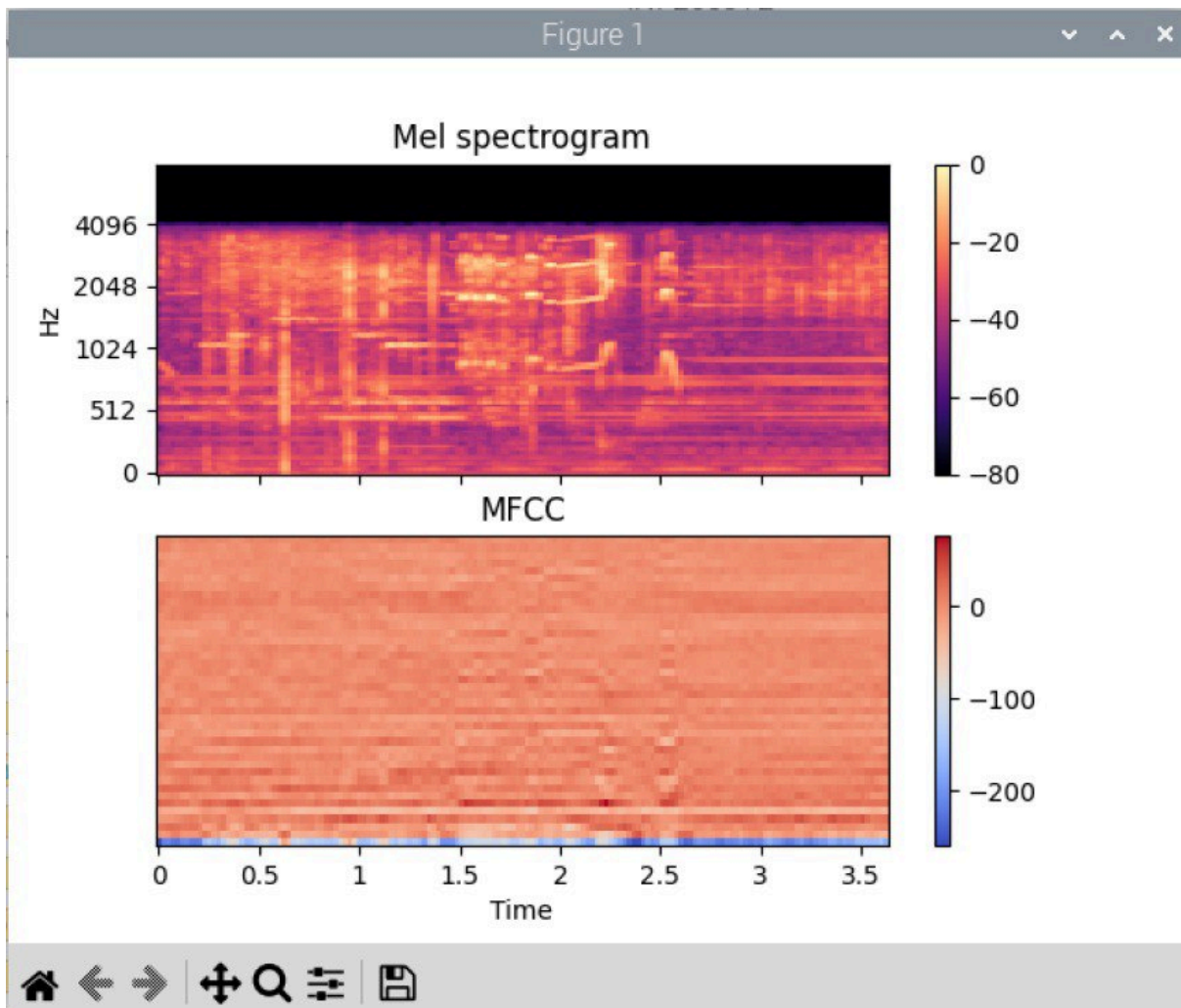
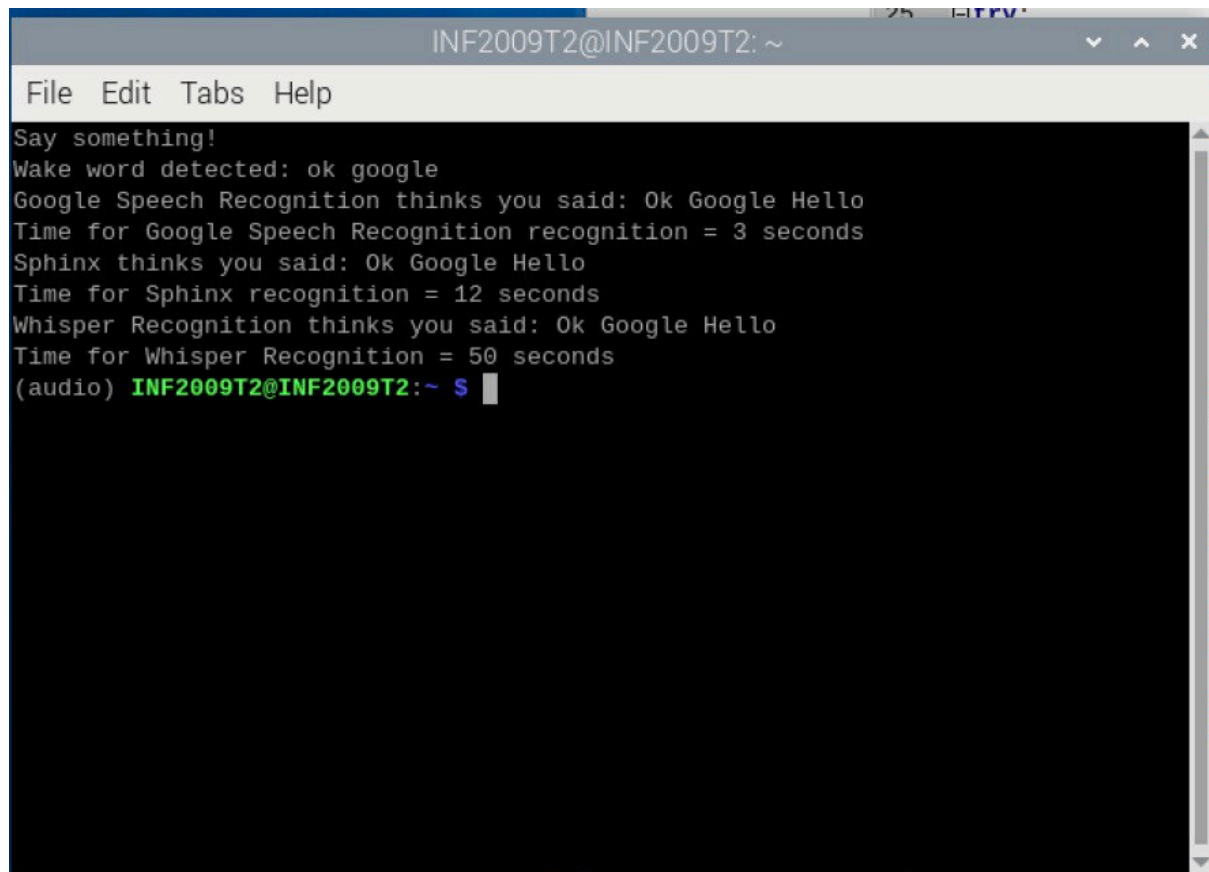


Fig 6. Mel Frequency Cepstral Coefficients (MFCC) and Mel Spectrogram comparison

### Advanced Sound Analytics (microphone\_recognition.py)

The performance of the OpenAI Whisper speech recognition API is compared to the Google and Sphinx speech recognition API as seen in Figure 7.

A terminal window titled 'INF2009T2@INF2009T2: ~' with a menu bar containing 'File', 'Edit', 'Tabs', and 'Help'. The terminal output shows the following text:

```
Say something!  
Wake word detected: ok google  
Google Speech Recognition thinks you said: Ok Google Hello  
Time for Google Speech Recognition recognition = 3 seconds  
Sphinx thinks you said: Ok Google Hello  
Time for Sphinx recognition = 12 seconds  
Whisper Recognition thinks you said: Ok Google Hello  
Time for Whisper Recognition = 50 seconds  
(audio) INF2009T2@INF2009T2:~ S
```

The text is displayed in a monospaced font on a black background. The prompt '(audio)' is in green, and the username 'INF2009T2' is in green. The prompt character 'S' is in blue. A vertical scrollbar is visible on the right side of the terminal window.

```
INF2009T2@INF2009T2: ~  
File Edit Tabs Help  
Say something!  
Wake word detected: ok google  
Google Speech Recognition thinks you said: Ok Google Hello  
Time for Google Speech Recognition recognition = 3 seconds  
Sphinx thinks you said: Ok Google Hello  
Time for Sphinx recognition = 12 seconds  
Whisper Recognition thinks you said: Ok Google Hello  
Time for Whisper Recognition = 50 seconds  
(audio) INF2009T2@INF2009T2:~ S
```

Fig 7. Performance comparison between the speech recognition APIs