Faculty of Engineering & Technology

Electrical & Computer Engineering Department

CIRCUITS AND ELECTRONICS LABORATORY

ENCS5344

Rivan jaradat ID:1200081

Aya Dahbour ID:1201738

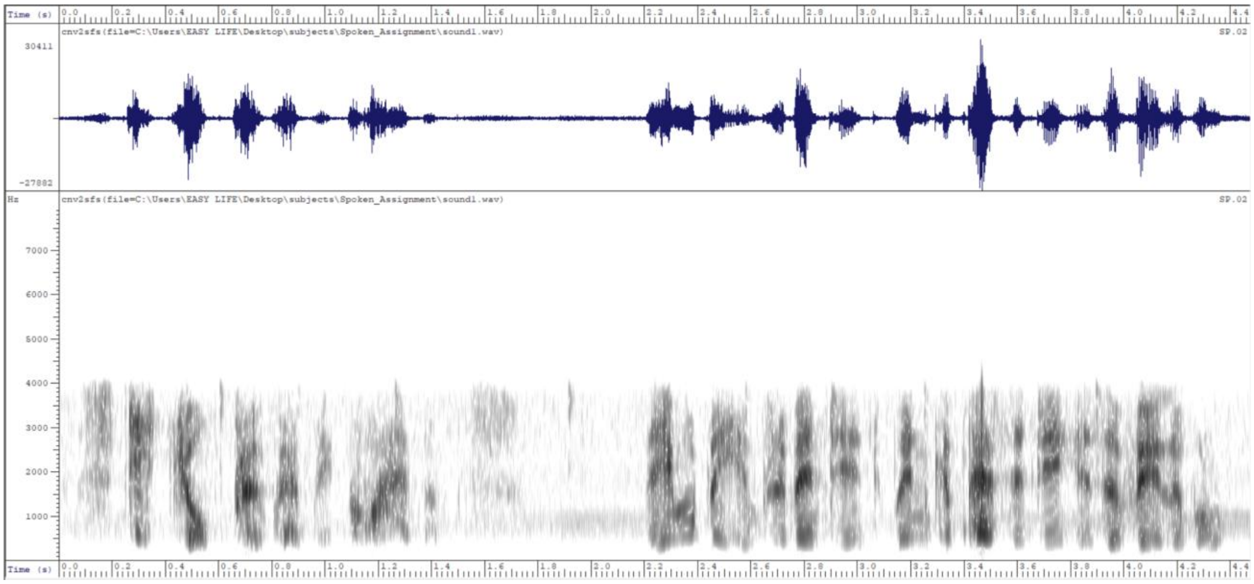
Section :1

Date :5/4/2024

Part 1:

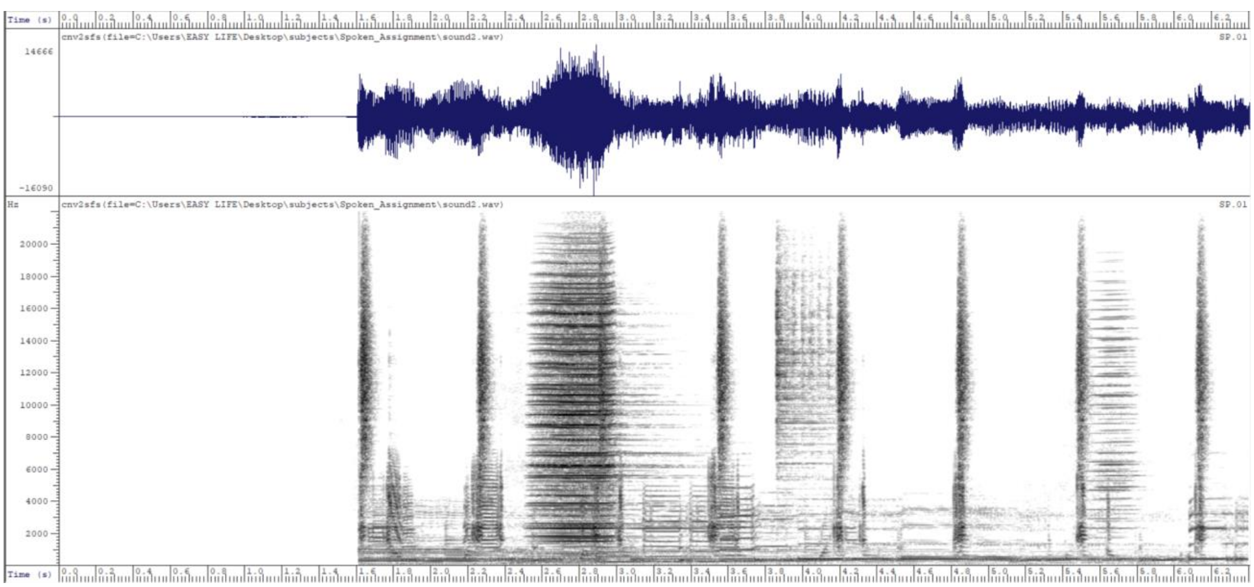
a.

- Sound1



As shown in the figure, the frequency range for the sound1 displayed in the spectrogram would be roughly from 0 Hz to 4000 Hz.

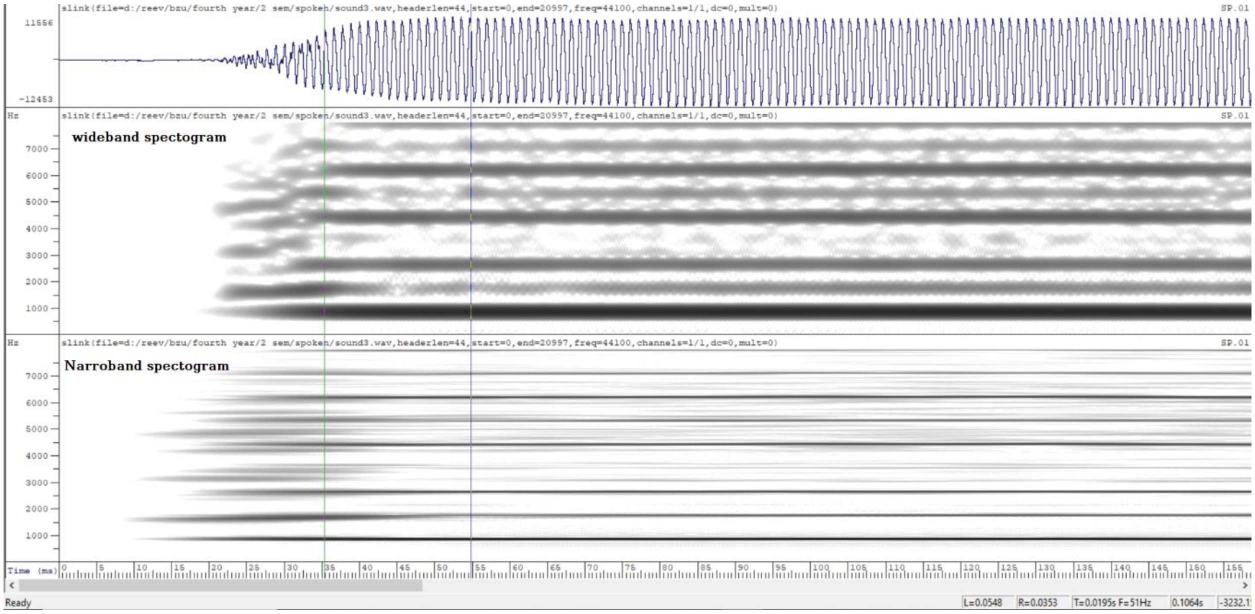
- Sound2



As for sound2, the estimated frequency range for the sound in this spectrogram is approximately from 0 Hz to 2500 Hz.

b.

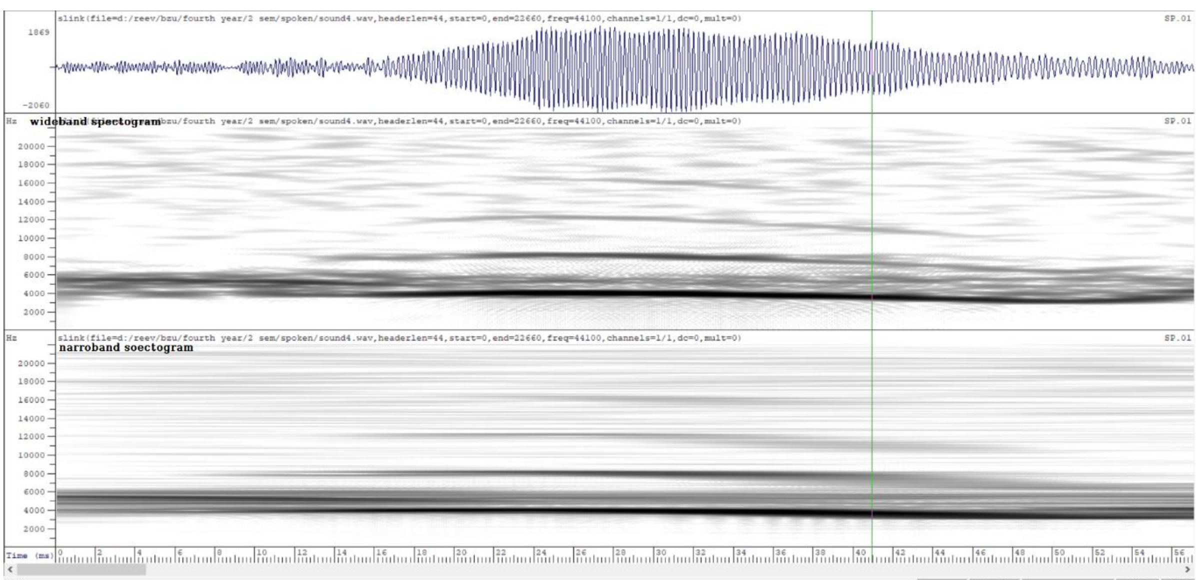
- Sound3



<u>Wide-band Spectrogram:</u> Because of its broad, blended frequency bands, it has poorer frequency resolution and is best suited for studying quick temporal changes rather than specific frequency content.

<u>Narrow-band Spectrogram:</u> High frequency resolution with clear, identifiable harmonic lines is ideal for extensive inspection of frequency components and harmonic analysis.

- Sound4

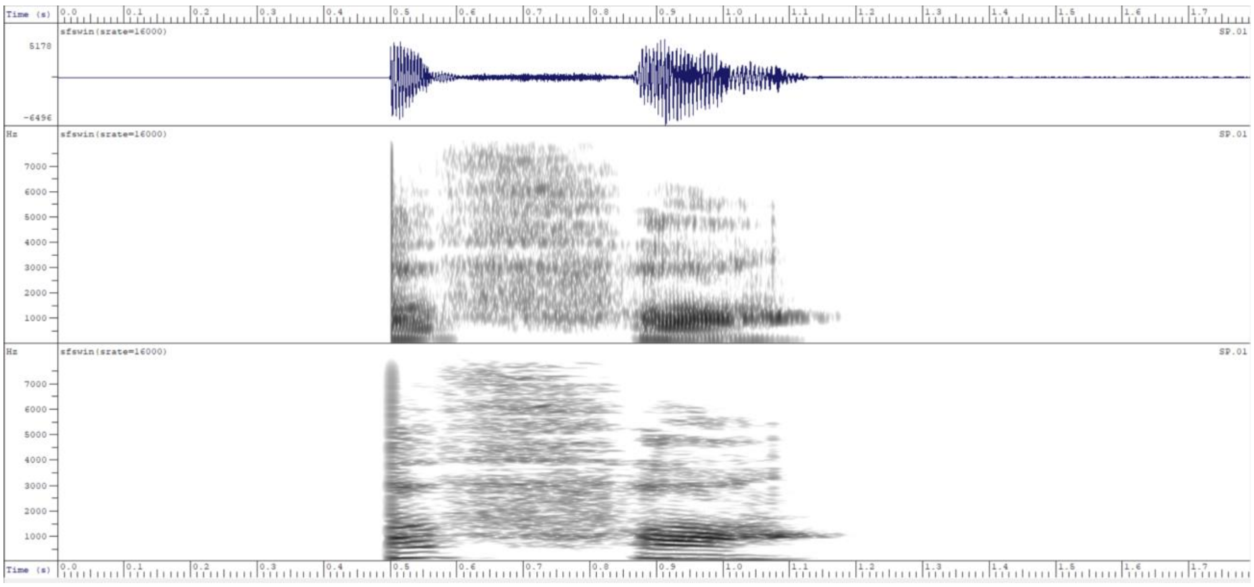


Wide-band Spectrogram: Similarly demonstrates reduced frequency resolution with a focus on temporal dynamics, which is good for capturing fast changes in sound.

Narrow-band Spectrogram: High frequency resolution with well-defined harmonic structures, perfect for detailed frequency analysis and complex sound tests.

- In the two pictures, narrow-band spectrograms provide comprehensive frequency insights, whereas wide-band counterparts excel at following fast temporal occurrences.

c.



1. **Waveform**

- Voiced Parts: Display a consistent, periodic rhythm that suggests a constant shaking of the vocal cords.
- Unvoiced Parts: Show an erratic, lower-amplitude pattern that reflects the absence of vocal cord vibration.

2. **Narrow-band Spectrogram**

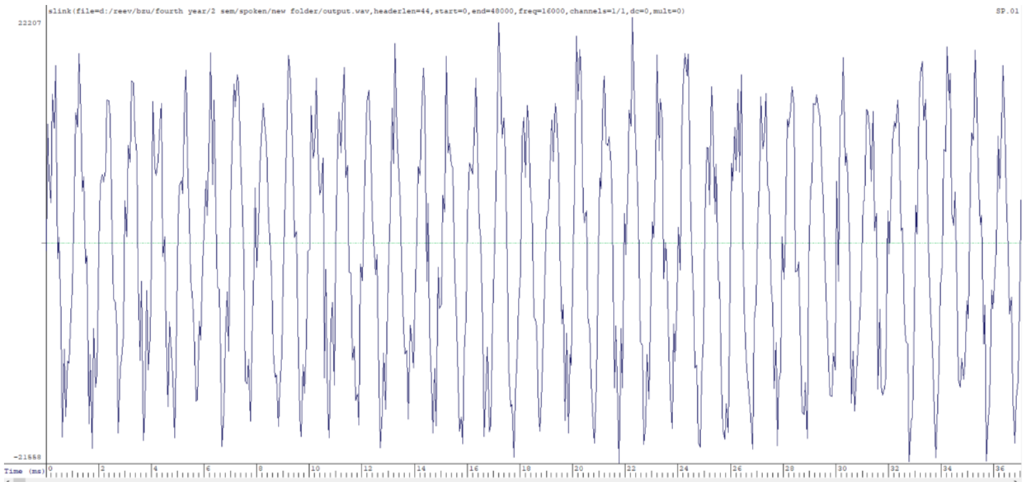
- Voiced Parts: Exhibit a distinct tonal structure with clearly defined harmonic lines.
- Unvoiced Parts: Show a broader, more dispersed appearance that is characteristic of noise-like sounds, with energy dispersed across a wider frequency range.

Part2:

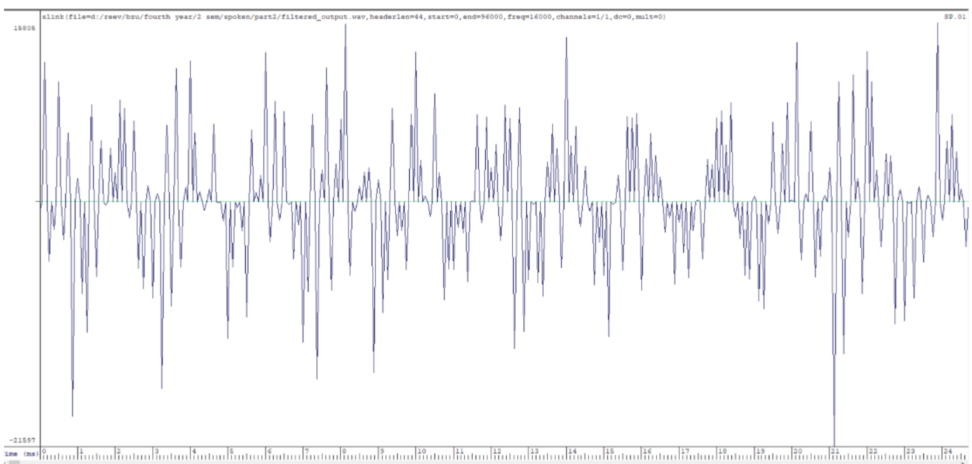
FIR filter:

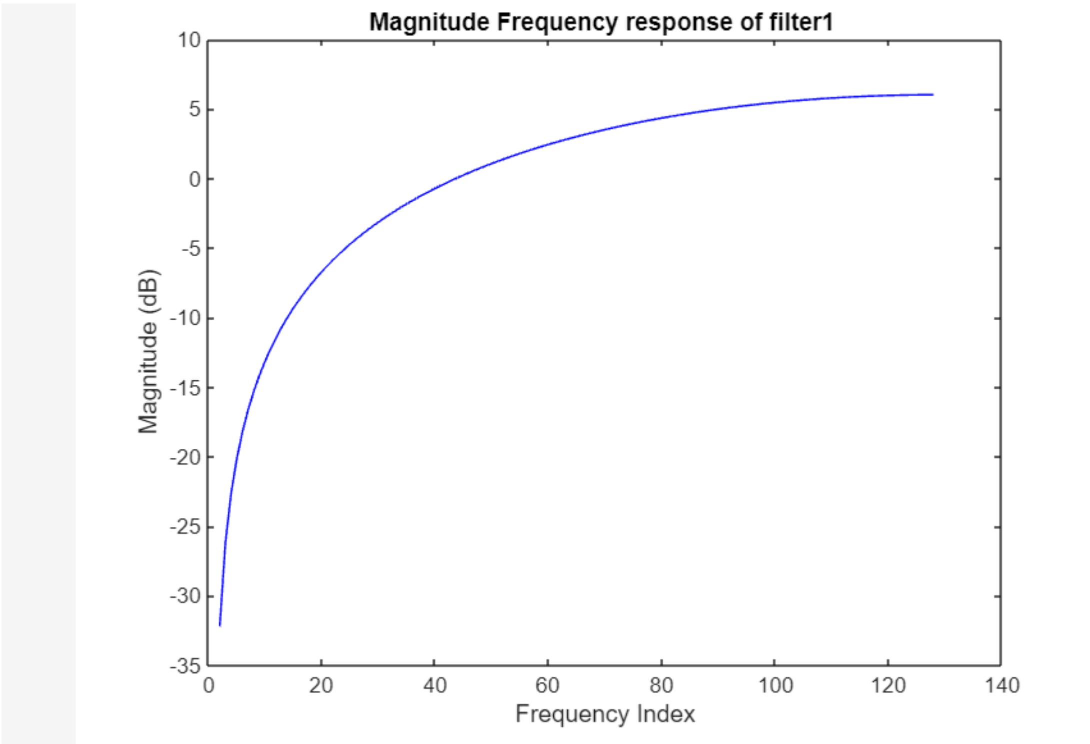
Input:

In this figure its represent the input of the FIR filter



Output:



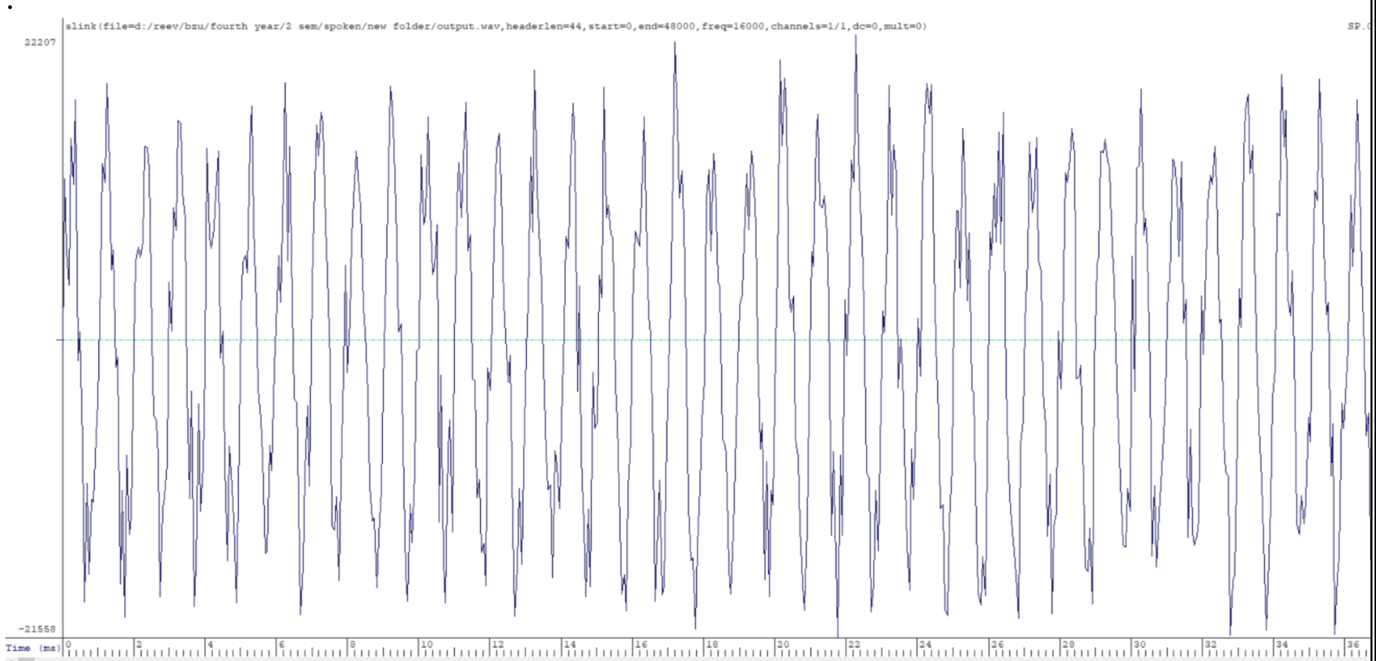


Second filter:

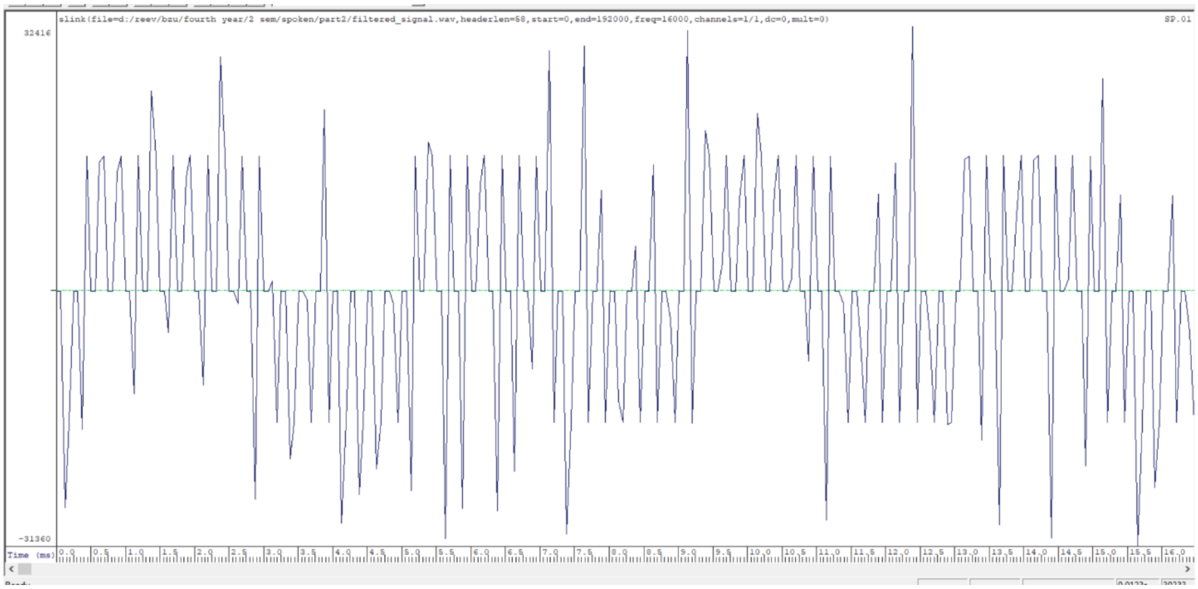
y(n) = b0*x(n)+b1*x(n-1) +b2*x(n-2) +b3*x(n-3) +b4*x(n-4) +b5*x(n-5) +b6*x(n-6) +b7*x(n-7)

input

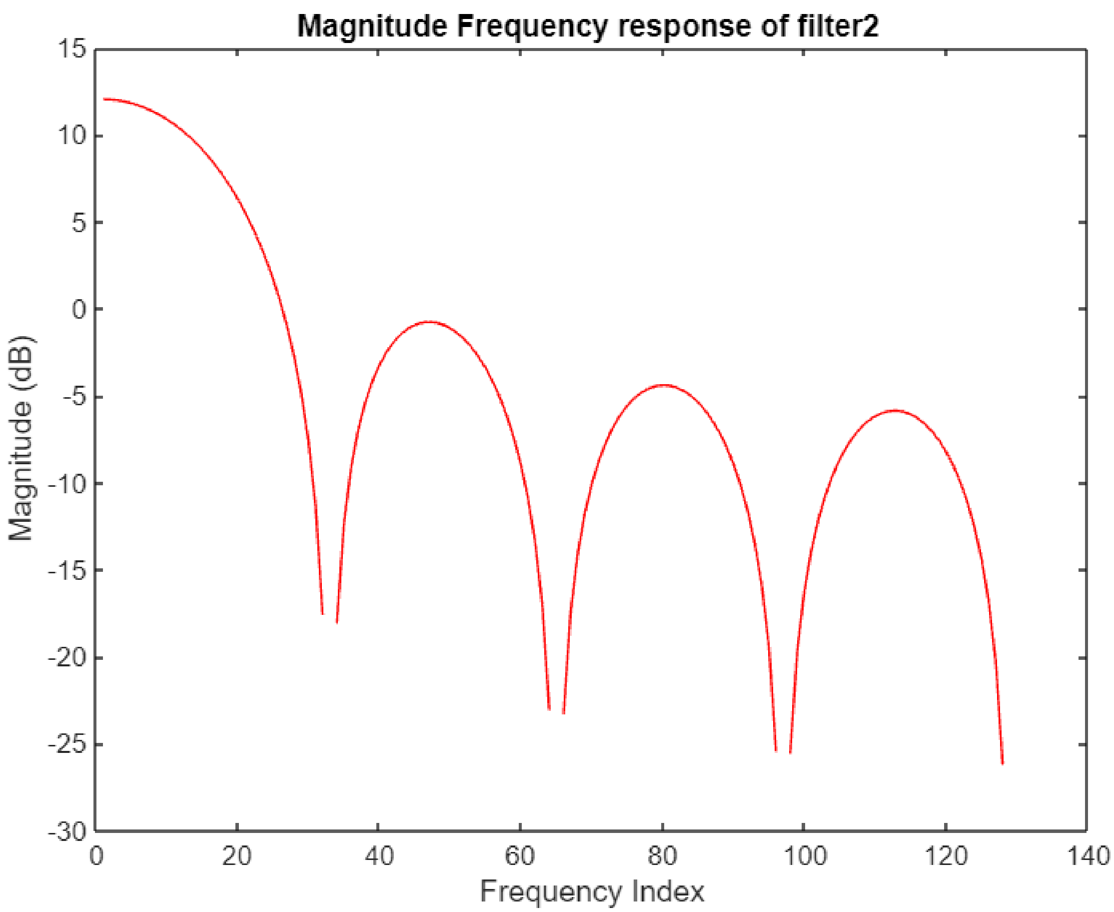
:



Output:



slink(file=d:/reev/bzu/fourth year/2 sem/spoken/part2/filtered_signal.wav,headerlen=58,start=0,end=192000,freq=16000,channels=1/1,dc=0,mult=0)
SP.01
32416
-31360
Time (ms)



Magnitude Frequency response of filter2
Magnitude (dB)
Frequency Index
15
10
5
0
-5
-10
-15
-20
-25
-30
0
20
40
60
80
100
120
140

Here, in this part, we combined Signals 1 and Signals 2 and obtained a sound that combines them. After that, we inserted it into two filters with the aim of filtering out the white noise from the 1 kHz sinusoidal signal, and then we noticed the difference between the inputs and outputs of each filter, as well as the difference between the two filters.

We notice in the images of the filtered output signals that both filters succeeded in reducing the noise level in the original signal. A sharper sinusoidal signal was obtained by the filter with an impulse response of {1, -1}, which was able to significantly suppress noise. However, the sinusoidal signal was somewhat distorted by this filter. Conversely, the second filter, whose coefficients were set to 0.5, provided a good compromise between signal distortion and noise reduction. Compared to the first filter, it was more successful in preserving the integrity of the sinusoidal signal, although not completely removing the noise. As a result, the second filter provided a more balanced approach, maintaining a somewhat stronger sine wave, while the first filter excelled at noise reduction.

Part 3:

A)

‘sample1.wav

vowels	F1	F2	F3
/ i: /	394Hz	2259Hz	2779Hz
/ { /	587Hz	2019Hz	2708Hz
/ u: /	406Hz	1532Hz	2529Hz
/ A: /	639Hz	1483Hz	2132Hz

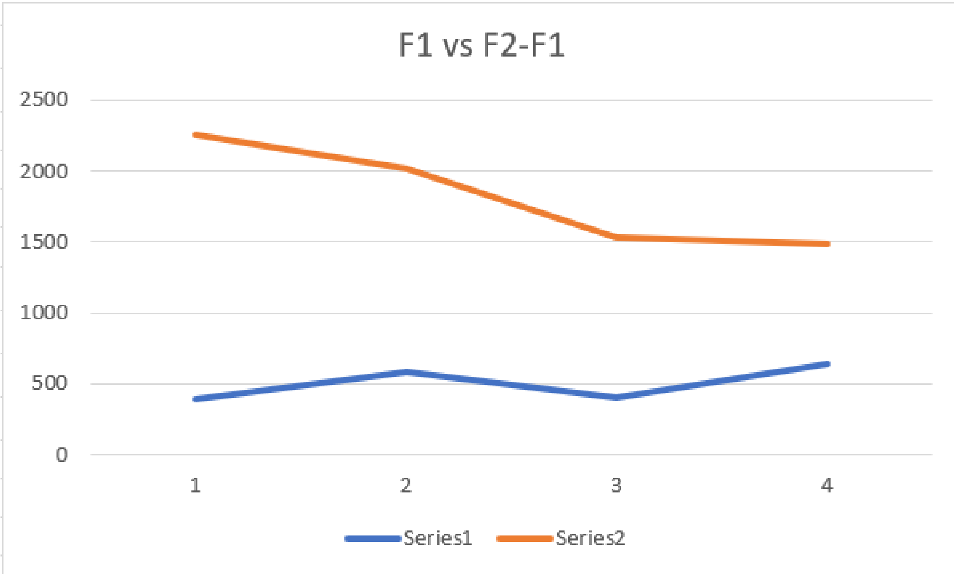
rivan.wav

vowels	F1	F2	F3
/ i: /	360Hz	2496Hz	3018Hz
/ { /	889Hz	1855Hz	2905Hz
/ u: /	385Hz	767Hz	2172Hz
/ A: /	744Hz	1741Hz	4191Hz

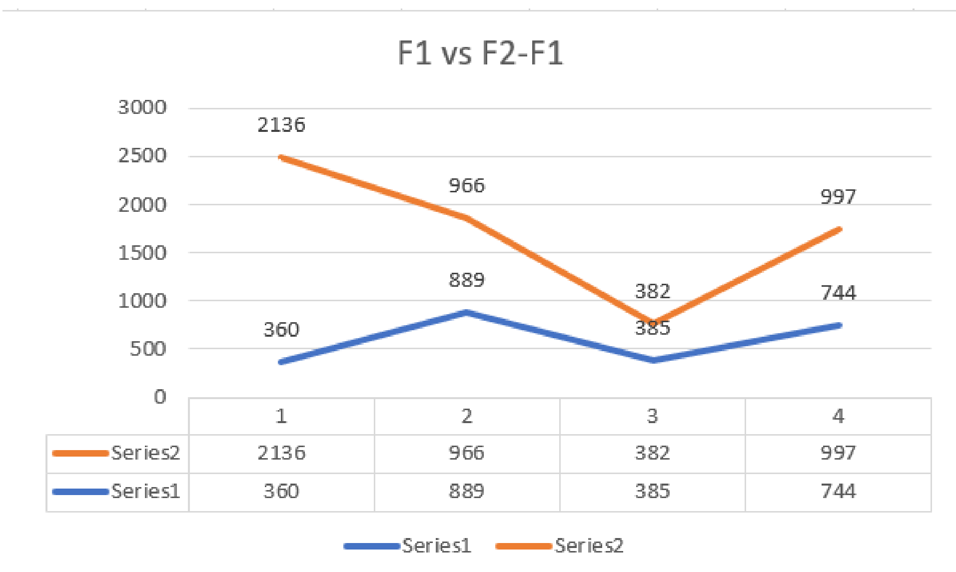
aya.wav

vowels	F1	F2	F3
/ i: /	370Hz	2640Hz	4488Hz
/ { /	954Hz	2814Hz	4226Hz
/ u: /	304Hz	956Hz	4119Hz
/ A: /	820Hz	1570Hz	2675Hz

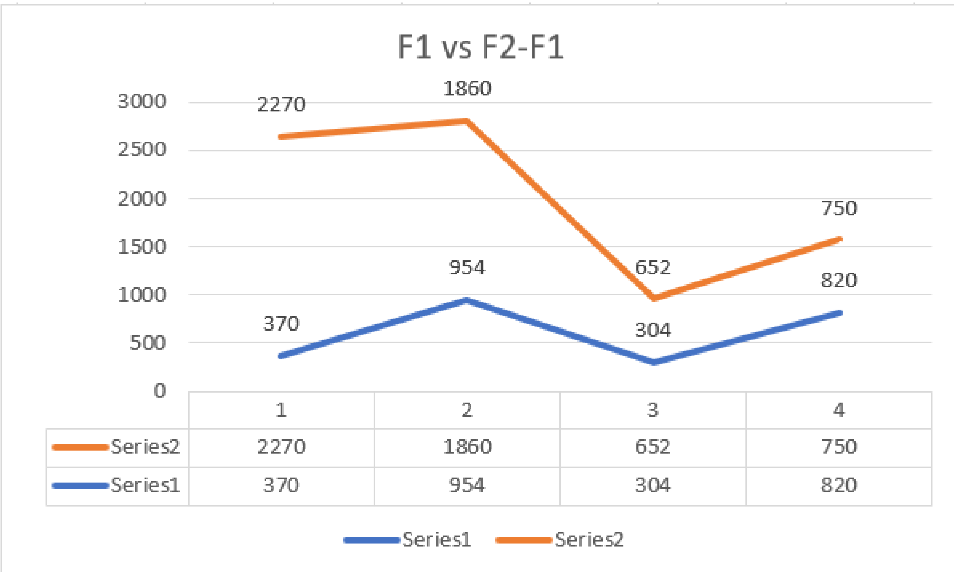
F1 vs F2-F1 for sample1.wav



F1 vs F2-F1 for rivan.wav



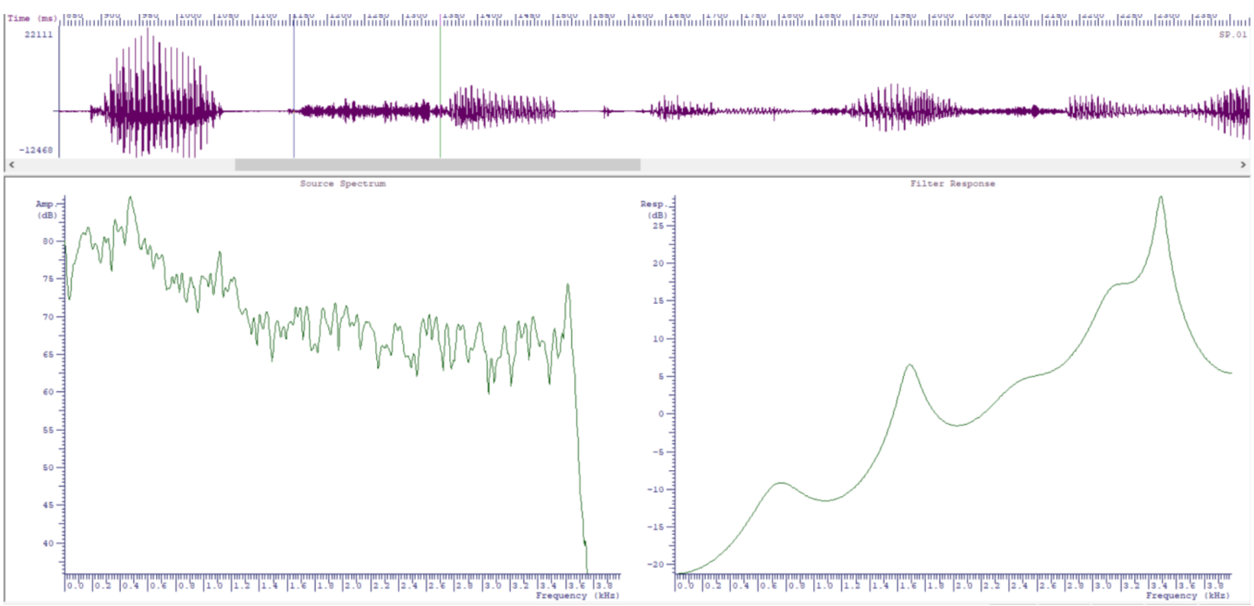
F1 vs F2-F1 for aya.wav



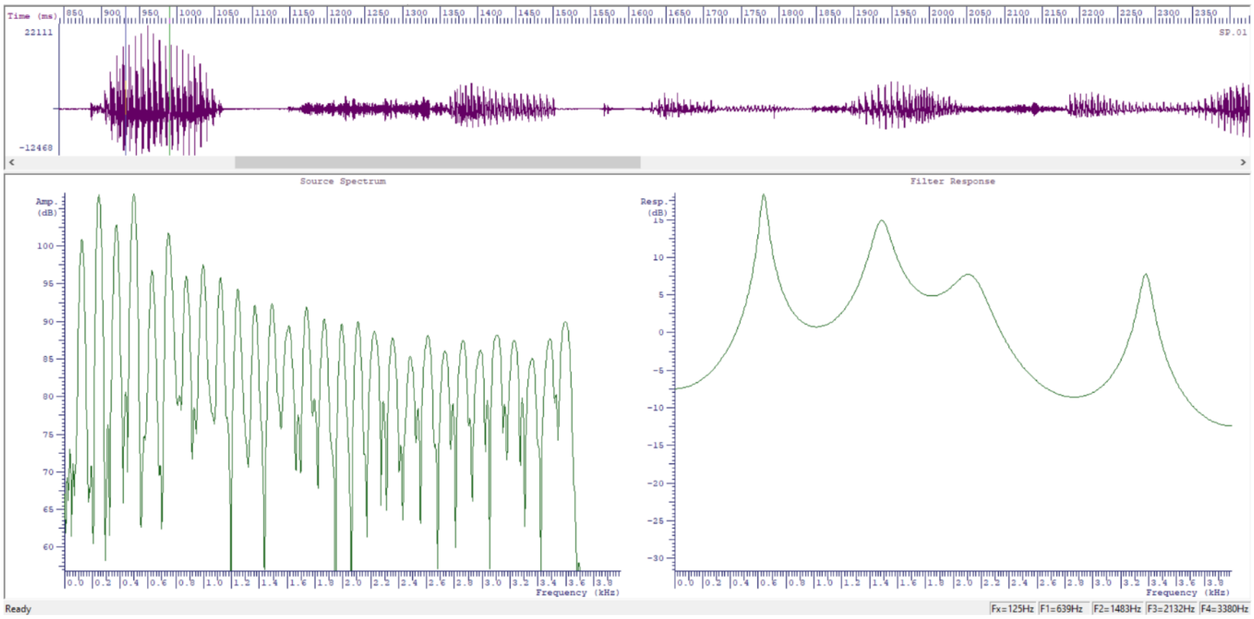
B)

Using the supplied audio file "sample1.wav," source-filter analysis was carried out to examine the creation of the phonemes /s/ and /A:/. For both phonemes, the source spectrum (S) and filter spectrum (F) were shown.

/s/



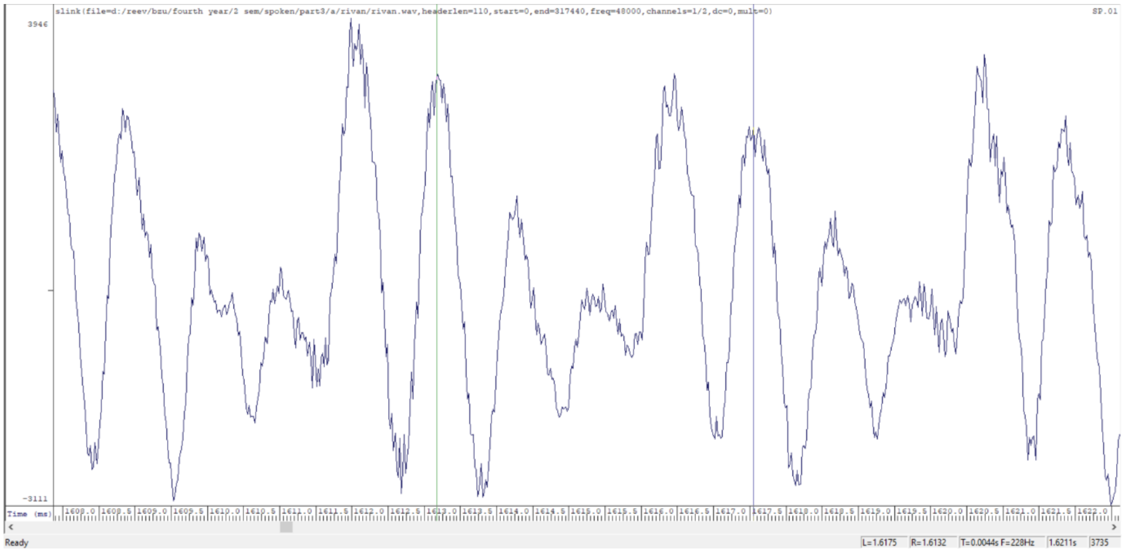
/A:/



Through the images we can notice the differences in how these sounds are produced in their individual spectra. The source spectrum of the /s/ sound, classified as a fricative, reflects energy dissipated over a wide frequency range, indicating turbulent airflow resulting from constriction at the site of articulation. On the other hand, the source spectrum focused energy at the low frequencies of the sound /A:/, where it is classified as a vowel. This means that the vocal cords vibrate periodically. The resonance properties of the vocal tract are reflected in the filter spectrum of both sounds. When /s/ is used, the filter spectrum displays peaks and troughs that correspond to the resonant frequencies of the vocal tract, but when /A:/ is used, the response is smoother and reflects the relatively open configuration of the vocal tract

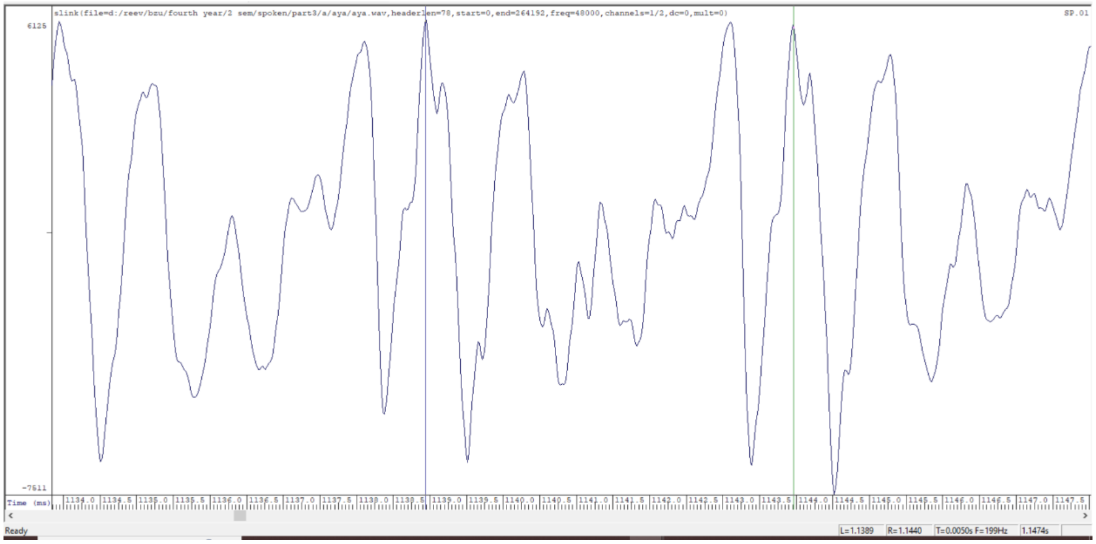
C)

Rivan (a)



F=228Hz

Aya (a)



F=199Hz

The fundamental frequency is derived from the duration of one complete cycle of the waveform, which represents one vibration cycle of the vocal folds. Since the period is the time it takes for one complete cycle, inversely related to frequency

Part4:

Phoneme	Start Time	End Time
/k/	0	35
/ə/	35	72
/p/	72	136
/æ/	136	215
/s/	215	295
/ɪ/	295	329
/t/	329	410
/i/	410	510

Part 5:

In this part, while collecting the sounds to obtain the desired sound, there were some difficulties, such as making sure that parts of the waveform moved smoothly to prevent errors in the final audio file. High-quality synthesized speech is produced by properly adjusting amplitude levels and aligning waveforms, among other critical factors.

Part 6:

In creating a new word from the phonemes extracted from "capacity" to form "cats" (K AE T S) which is contained in file “cats.wav”, the resulting speech quality may have a few comments and potential areas for improvement as the following:

1. **Quality:**
- <u>Transitions:</u> The spliced phonemes may have sharp transitions, which disrupt the fluidity of speech.
- <u>Intonation:</u> The rhythm and pitch may not be consistent with a natural pronunciation of "cats", making it sound unnatural.

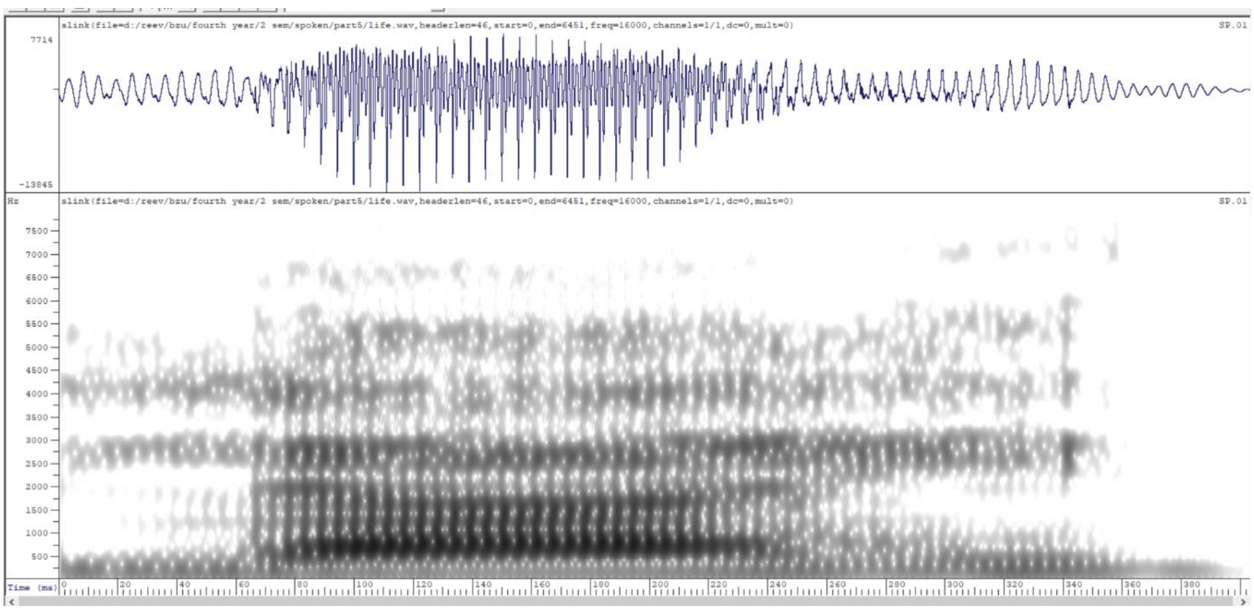
2. **Improvements:**
- <u>Smoothing:</u> Use cross-fading techniques at phoneme boundaries to make the sounds mix more naturally.
- <u>Adjusting Pitch and Duration:</u> Use audio editing software to adjust the pitch and duration of phonemes to better resemble the natural qualities of the word "cats".
- <u>Noise Reduction:</u> Apply noise reduction algorithms to remove any background noise or artifacts.

Part7:

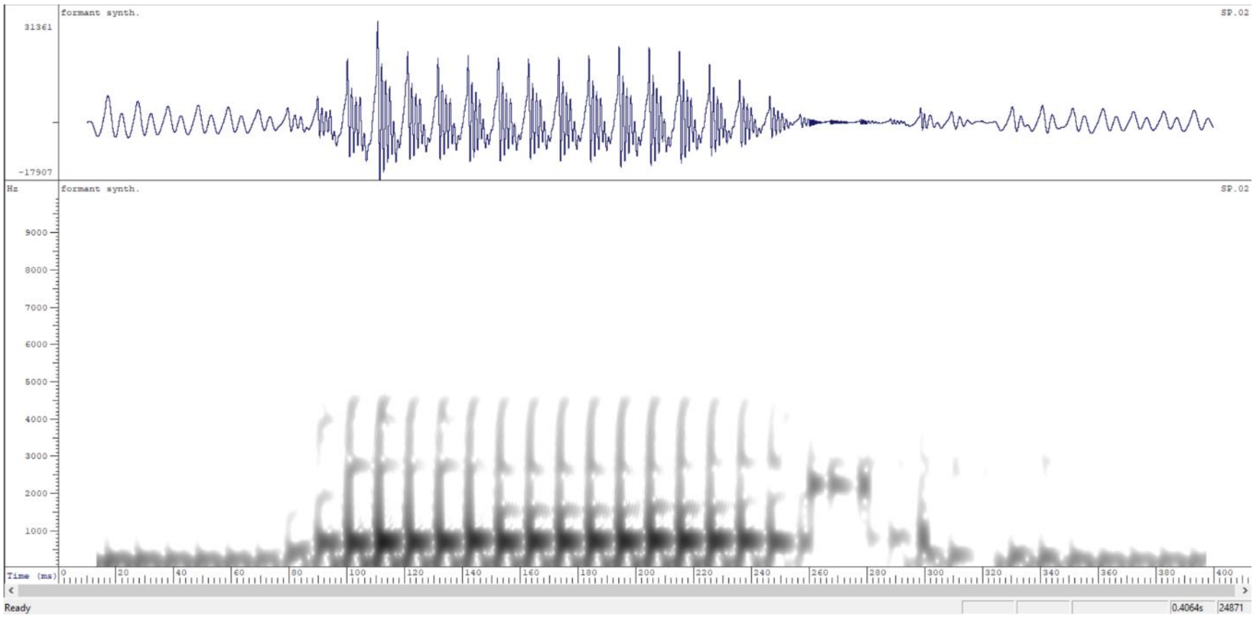
After applying this part to achieve the “Generation of vowel sounds using the source-filter model”, We obtained the expected results for each of the required vowels. It was also described and attached sequentially to each of the audio files. The “low” files, which depicts the sound after it has been passed through a band bass filter, has its own frequencies, with 200 subtracted from either side. As for the “high” files, reflects an addition of 200. And the las, the “output” represents the results of the combined signal.

Part 8:

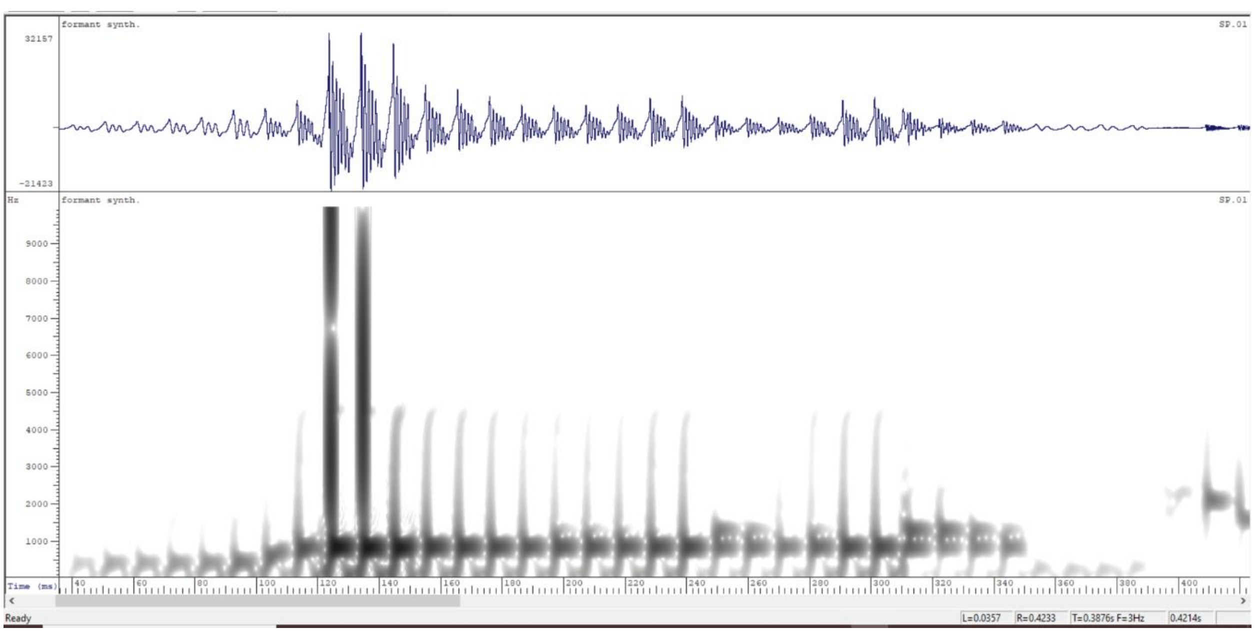
Original:



synthesized original:



synthesised modified:



In general:

Original speech: A model of the human voice, with inherent amplitude and variation in energy distribution across frequencies.

Original synthesized speech: Typical characteristics of synthesized speech are consistency and lack of detail compared to native speech.

Modified and compound speech: Because there is constant formant frequency and high uniformity, it sounds more mechanical than natural.

From the images we see that the modified speech appears unnatural due to fixed formant values, resulting in a prolonged /a/ sound instead of a natural assimilation. The original and composite text should sound closer to typical speech, with the composite version being slightly less dynamic.

B)

Fundamental frequency for pitch changes and formant frequencies for timbre changes are the two basic parameters that are adjusted while changing a speaker's voice in a formant-based synthesizer. This allows the sound synthesizer to replicate the distinctive vocal features of different speakers.