REDUCING SPEECH NOISE

Ali DAĞDELEN, Mert TAMCAN, Moayad RAJJOUB
04102401, 171024032, 171024078
adagdelen@gtu.edu.tr , mert.tamcan2017@gtu.edu.tr , moayad@gtu.edu.tr

ABSTRACT

Our project aims to obtain a noise-free audio signal at the output as much as possible by reducing the noise in the noise-added audio signal.

For this purpose, we created different sound signals containing frying noise by using three different twenty-second speeches and three different frying sounds.

By analyzing the frying sounds and the conversations we created with added frying sounds in the frequency space, we tried to suppress the frying sound noises we added to the conversations by determining the filter type and filter parameters.

The document includes the definition of the problem addressed in the project, the solution approach to this problem, and evaluations of how the project can help solve similar problems that we may encounter in daily life.

KEYWORDS

Amplitude Spectrum Analysis, Noise Reduction, Low Pass Filter

1. Defining the Problem (Introduction)

Whether it is a conversation recorded to be listened to later or instant communication such as telephone communication, effectively suppressing (filtering) the noise in the conversation is of great importance for the efficiency of communication in order for the information to be conveyed to be accurate and understandable.

Noise refers to sound signal components that have no informational value in communication and are therefore actually unwanted.

The conditions in which the conversation takes place and is recorded may naturally cause some unwanted sounds to be mixed into the speech sound. When talking on a mobile phone while walking in windy weather, the other party has difficulty understanding what is being said due to the wind noise mixing with the speaking voice. This is a good example to express the importance of reducing noise. In order to create a noise-free speech recording outside of environments specially designed for sound recording, such as studios, a solution to filter out the noise is required.

There are different solutions for noise filtering or reducing speech noise.

For example, some special microphones can filter out unwanted sounds in speech and suppress the noise before the physical sound wave turns into an electrical sound signal. In devices such as microphones, filtering is generally carried out by electronic components for this purpose.

Other noise added to the signal in the communication channel and all distortions occurring in the channel can be digitally filtered at the output, and as a result, the sound signal that comes with noise from one end and is inevitably distorted throughout the channel can be obtained at the other end, free from all noise, distortion and unwanted sounds.

In digital communication, noise suppression / filtering processes are often carried out with computer software, so that the voice signal, that is, speech, can be transmitted with low cost and minimum power consumption, without depending on electronic elements and complex circuit designs.

In this project, these two great advantages of digital communication were taken into consideration, the sound signals were read by a computer program, frequency analysis was made, and appropriate filter parameters were determined, and the noise added to the speech was tried to be reduced through the software.

As we briefly mentioned above, signal processing and filtering operations carried out through software provide a great cost advantage since they do not require electronic components and extra power consumption. In addition, another great advantage of signal processing performed through software is ease of development.

At almost zero cost, you can experiment using different flow diagrams and algorithms until you achieve results that are very close to ideal.

However, in physically implemented signal processing or filter structures, designing a different circuit each time is both a costly and tiring process. In addition, such physically implemented circuits have unique challenges that must be taken into account in the design of electronic circuits, such as excessive power consumption and keeping the temperature under control so that the circuit elements remain in the efficient operating zone.

In this respect, processing the signals through software using a computer program provides great convenience in practice.

2. Solution Methodology of the Problem

Each sound has inherently different characteristics. While even the voices of two people speaking the same language are different from each other and can be easily distinguished, it is not expected that the sound produced by a person speaking and a frying pan of potatoes have the same characteristics.

We prefer to analyze different sounds in frequency space to see their distinctive details. In this respect, frequency is a very important feature for audio data.

High-frequency sound waves have small periods, low-frequency sound waves have large periods.

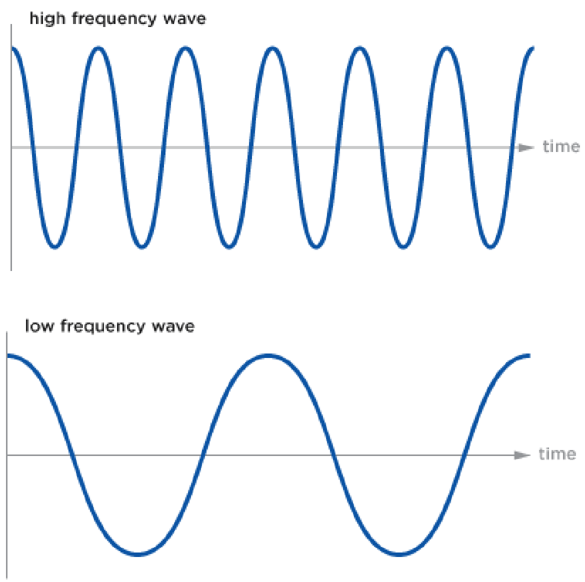


Figure 1 - High and Low Frequency Sounds [1]

For example, the frequency of an adult male's speaking voice is between 85 and 180 Hz. [2] Using this information, we can detect and filter out unwanted sounds (noise) of different frequencies that interfere with an adult male's speech.

With the basic approach I explained above, the aim of this project is to filter out the frying sounds as much as possible, that is, to reduce the noise, from the sounds created by adding the sounds of different foods fried in oil to the speaking voices of different adult males.

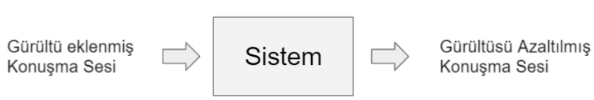


Figure 2 – General Block Diagram of the System

The function of the system shown as a closed box in Figure 2 is to filter out the noise as much as possible, thus ensuring a clearer understanding of the speech. Therefore, the system should basically consist of a filter.

As an application in this project, the frying noises of different foods were added to the speaking voices of different adult men. Our aim is to suppress the frying noise in the background of conversations as much as possible.

Figure 3 – Solution Steps of the Problem

Filter properties will be determined by examining the amplitude spectra of frying noise sounds and speech sounds to which these noises are added.

3. Analysis of Data Related to the Problem

In our project, we used three different speech data, each twenty seconds long. The first speech section was taken from a speech given by Steve Jobs to a crowd. The second speech is taken from a BBC interview with Mark Zuckerberg. The third speech was taken from Elon Musk's speech to a crowd.

We added different frying sounds as noise to each of these speech data. French fries were added as noise for the first speech, fish fries were added as noise for the second speech, and chicken fries were added as noise for the third speech.

We used Camtasia Studio software to add frying sounds as noise to the conversations.

A) Examination of Frying Sounds Added as Noise in Frequency Space

Frying sound files were read with Python and discrete time graphs of these signals were drawn using the obtained data.

Discrete Time Plot of French Fries Sound

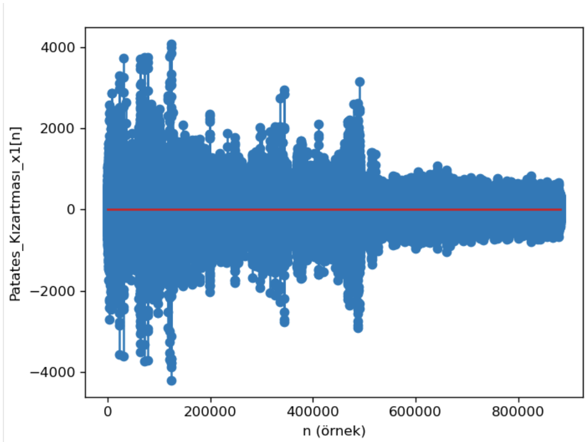
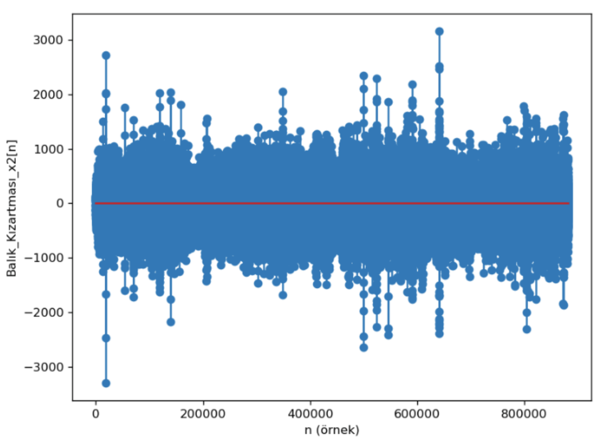


Chart 1 – French Fries (Discrete Time)

The French fries sound has 2 channels and there are 88200 samples in each channel. The sampling frequency of this audio file is 44100 Hertz.

Discrete Time Plot of Fish Fry Sound



Graph 2 – Fish Fry Sound (Discrete Time)

The fish fry sound has 2 channels and there are 88200 samples in each channel. The sampling frequency of this audio file is 44100 Hertz.

Discrete Time Plot of Fried Chicken Sound

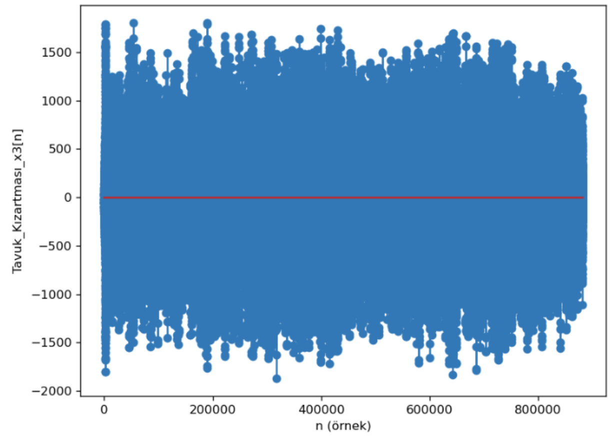
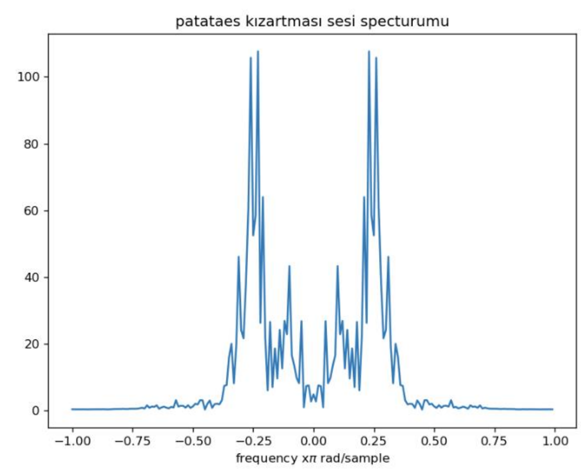


Chart 3 – Frying Chicken Sound (Discrete Time)

The chicken fry sound, like our other two sounds, has 2 channels and there are 88200 samples in each channel. The sampling frequency of this audio file is 44100 Hertz like the others.

Moving the signals to the frequency space with the Fourier transform provides great convenience in the analysis of the signals. It is not possible to determine the properties of this signal by commenting on the discrete time graphs of the signals above and to decide on the properties required for a suitable filter design. Therefore, the graphs we will draw in the frequency space must be at hand when making such evaluations.

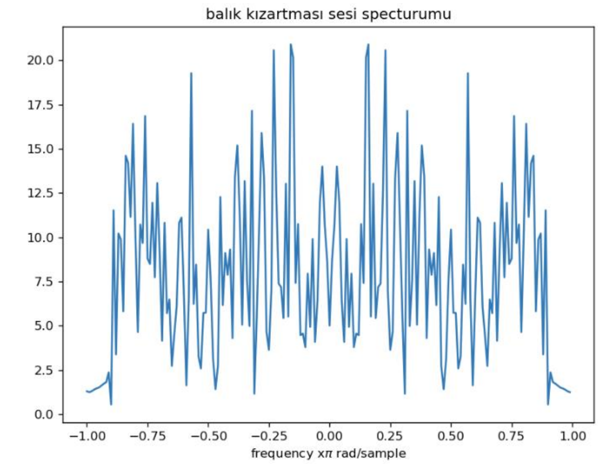
Amplitude Spectrum of French Fries Sound



Graph 4 – French Fries Amplitude Spectrum

When we examine the amplitude spectrum above, it can be seen that the sound of French fries has amplitude components in the frequency range of $-\pi/2$ to $+\pi/2$, and especially its amplitude reaches the highest value around $+\pi/4$ / $-\pi/4$ frequencies.

Amplitude Spectrum of Fish Fry Sound



Graph 5 – Amplitude Spectrum of Fried Fish

When we examine the amplitude spectrum of the fish fry sound, we see that the components of the fish fry sound are distributed in the -π to +π frequency range, instead of being concentrated in a certain frequency range. With this feature, it clearly differs from the other two frying sounds.

Amplitude Spectrum of the Sound of Fried Chicken

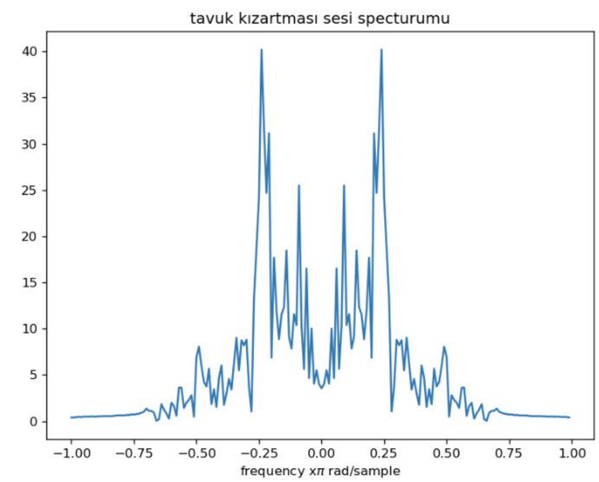


Chart 6 – Amplitude Spectrum of Fried Chicken

The amplitude spectrum of the sound of fried chicken is similar to the sound of French fries. When we examine the graph, it can be seen that the amplitude components are concentrated in the -3π/4 to +3π/4 frequency range, and especially around the -3π/5 / +3π/5 frequencies, the amplitude reaches its highest value.

B) Examination of Conversations with Added Frying Noise in Frequency Space

The graph below was obtained by plotting the sound created by adding French fries' noise to Steve Jobs' speech in discrete time.

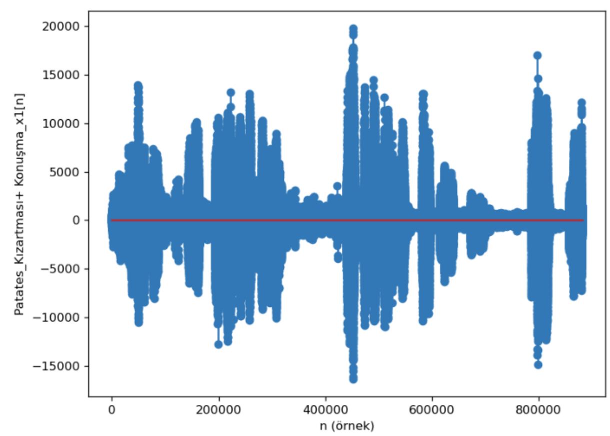


Chart 7 – Speech with French Fries Noise Added (Discrete Time)

The graph below was obtained by adding fish fry noise to the speech of Mark Zuckerberg's interview with the BBC and plotting it in discrete time.

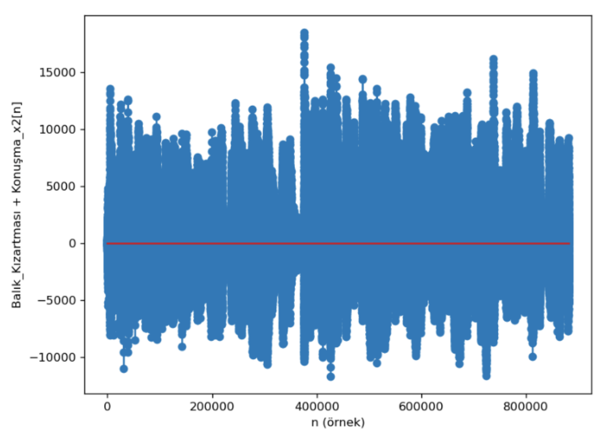
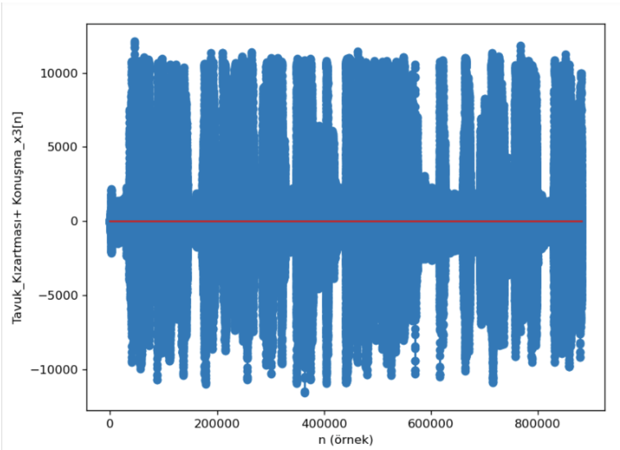


Chart 8 – Speech with Added Fish Fry Noise (Discrete Time)

The graph below was obtained by adding chicken fry noise to Elon Musk's speech and plotting the resulting sound in discrete time.

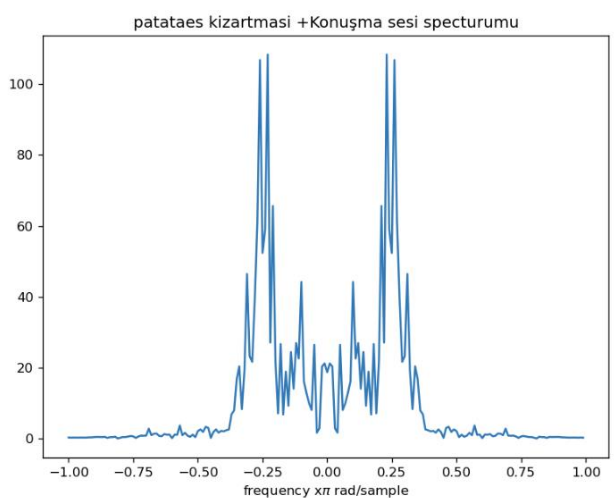


Graph 9 – Speech with Added Noise of Frying Chicken (Discrete Time)

Let's examine the amplitude spectra of these voice signals by taking the Fourier transforms of the speech with added frying noise.

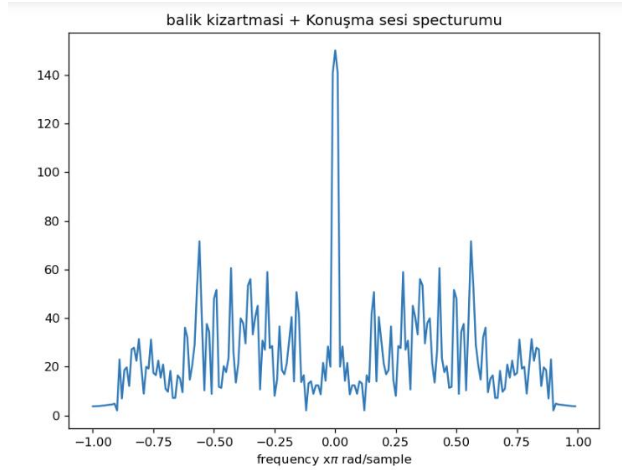
Examining the Amplitude Spectra of Speeches and Determining the Filter Features Required to Reduce Noise

The graph below is the amplitude spectrum of the sound created by adding French fry noise to Steve Jobs' speech.



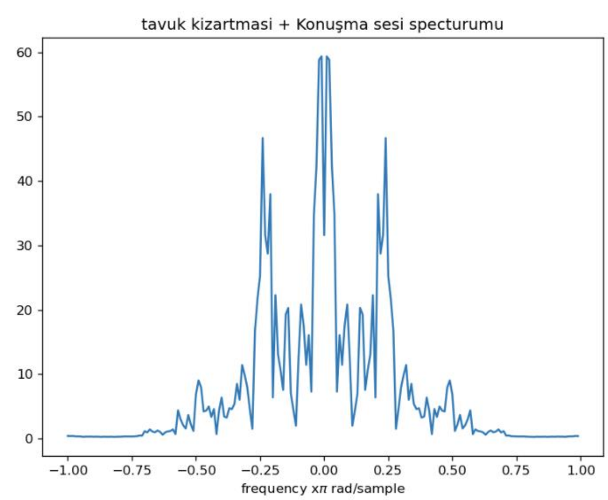
Graph 10 – Speech with French Fries Noise Added (Amplitude Spectrum)

The graph below is the amplitude spectrum of the sound obtained by adding fish fry noise to the speech taken from Mark Zuckerberg's BBC interview.



Graph 11 – Speech with Added Fish Frying Noise (Amplitude Spectrum)

The graph below is the amplitude spectrum of the sound created by adding chicken fry noise to Elon Musk's speech.

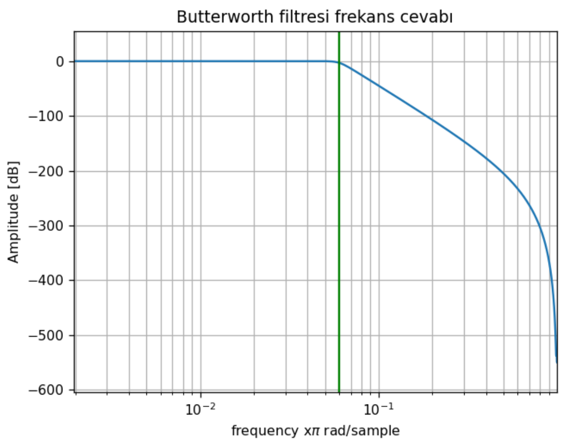


Graph 12 – Speech with Added Noise of Fried Chicken (Amplitude Spectrum)

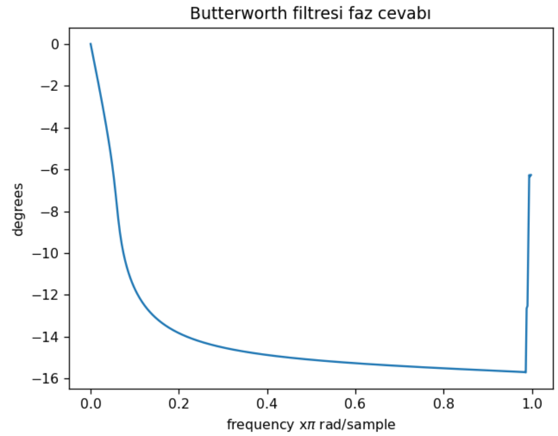
Since the large amplitude frequencies of the human voice are at low frequencies, it is obvious that a low-pass filter with a low cut-off frequency must be designed. Based on the common features of the 3 frying graphs, a value close to the 0.1 pi frequency can first be selected as the cut-off frequency. However, since the 3 frying sounds are different from each other, the amplitudes of the 3 frying sounds at frequencies lower than 0.1pi are not the same. In this case, the performance of the filter in 3 different situations will not be the same. At frequencies lower than the 0.1pi cutoff frequency, sounds with larger amplitude values will be louder. After various frequency trials to see the performance of the filter in 3 different situations, 0.06pi was set as the cut-off frequency. [3]

As the filter degree increases, the sharpness of the filter increases. However, as the higher filter degree increases the side lobes in the impulse response, the number of noise components in the time domain also increases. Additionally, increasing the filter degree also increases the phase delay. After trying various filter degrees, the 10th degree was chosen. [4] [5]

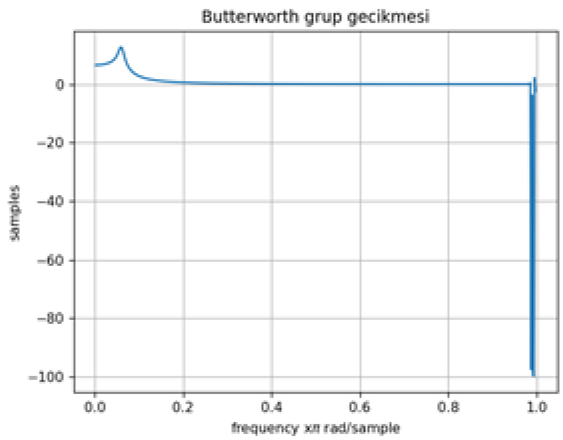
Filter Charts



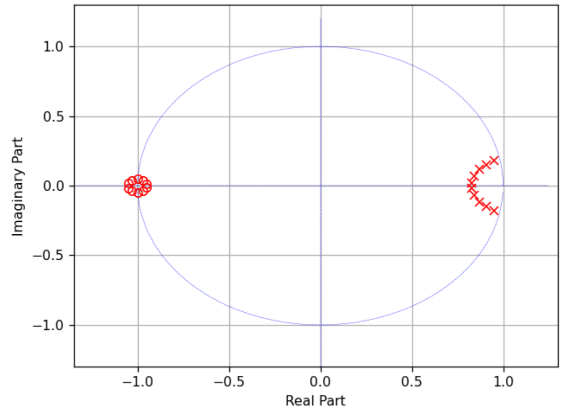
Graph-13 Filter Frequency Response



Graph-14 Filter Phase Response



Graphic-15 Filter Group delay



Graph -16 Pole zero diagram

Graph -13 gives us the frequency response of our filter in decibels. It is clear from the graph that the filter we designed is a low-pass filter.
The green line shows us the part where the 0.06*pi rad/sample value corresponds.

It can be seen from Graph-14 that the phase graph is not linear. Since FIR filters are linear phase, this shows us that we have designed an IIR filter. Increasing the filter degree negatively affects the phase delay.

Graph-15 shows us that there is no fixed group delay. This feature belongs to IIR filters. Since our filter is not linear in phase, the group delay is not constant. This means that the output signals will be delayed by different sample amounts at different frequency values. The fact that the zeros are at -1 in the graph -16 indicates that the filter will filter out high frequency values. This tells us that the filter is a low pass filter. Polar angles are related to the frequency of the filter. The filter degree affects the angle value separating the poles. The formula 180/N shows that there are 18 degrees between the poles. [4]

CONCLUSION AND EVALUATION

The low pass filter we used reduced the frying noise to a certain extent but could not completely suppress the noise.

One of the reasons why the filter cannot completely suppress noise may be that some components of the noise are located in the pass band of the filter.

On the other hand, since some components of the speech voice were outside the pass band of the filter, some distortion occurred in the speech voice.

REFERENCES

[1] https://afsharphysics.files.wordpress.com/
[2] https://en.wikipedia.org/wiki/Voice_frequency
[3] https://docs.scipy.org/doc/scipy/reference/l
[4] https://www.allaboutcircuits.com
[5] https://www.earlevel.com/main/2003/02/27/pole-zero-placement/