Project
Design of Binuaral Hearing Aid

Krishan (210020021)

Samyak Sanjay Parakh (210020043)

Under the Supervision of:

Professor Naveen Kadayinti

Professor Nitya Tiwari

Evaluator:

Professor Rajshekhar Bhat

Professor Naveen MB

November 2023

1. Introduction

What is Hearing Aid?
A hearing aid is an electronic, battery-operated device that amplifies and changes sound to allow for improved communication.

What is hearing loss and deafness?
A person who is not able to hear as well as someone with normal hearing – hearing thresholds of 20 dB or better in both ears – is said to have hearing loss.

1.1. Existing Hearing Aids

Types of hearing aids

Figure 1: Types of Hearing Aid

- Behind-the-ear
- In-the-ear
- In-the-canal

Limitation of Exiting Hearing Aid

- Poorly fitting earmolds may cause feedback, a whistle sound
- Damaged by buildup of earwax and ear drainage
- Patient specific
- Expensive
- Need to be replace after a period of time

1.2. Significance of the Project

- According to WHO over 5% of the world's population – or 430 million people – require rehabilitation to address their disabling hearing loss (432 million adults and 34 million children).
- Nearly 80% of these patients live in countries with low and middle income. The number of people having some form of hearing loss is projected to increase to 2.5 billion by the year 2050 [50]. This is about 1 of every 4 people.

1.3. Major Objectives

- Fixing the limitation of existing hearing aids like prevention from unhygienic condition
- Providing additional benefits like comfortable to wear over prolonged periods of use by the hearing-impaired users specially old-aged persons and children
- Value for money product and less maintainance cost

1.4. Specific objectives

a) A binaural hearing-aid consisting of digital signal processing (DSP) chip, array of microphones, speakers with required insulation, power amplifiers, audio codec with ADC and DAC, Bluetooth, etc. will be designed to fit over the ear with a head-band. Over-the-ear design maximizes user comfort and headband provides sufficient area for placement of electronic components.

b) Signal processing algorithms for use in hearing aids to enhance the speech perception of hearing-impaired listeners will be developed. Signal processing techniques including

- Beamforming for reducing acoustic feedback by creating nulls in the direction of the output speakers
- The device will be designed to compensate hearing loss of up-to 60-80 dB in the first prototype.
- Direction of arrival estimation, speech enhancement, and beamforming for suppressing background noise,
- Frequency dependent amplification for compensating frequency dependent loss,
- Dynamic range compression for compensating loudness recruitment.

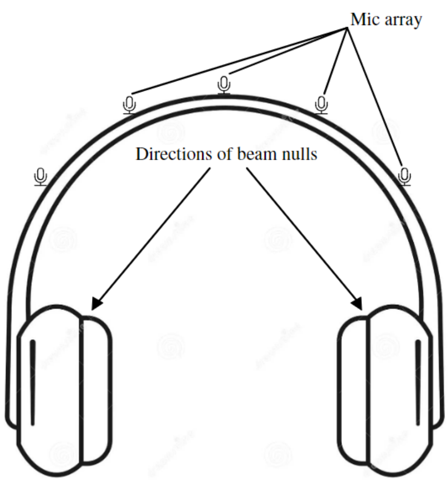


Figure 2: Illustration of an Over the ear type headphone. These are the most comfortable for long use and provide enough room for housing a large battery and multiple microphones and associated electronics.

2. Directional Pattern of Microphone

AIM: To find the directional Pattern of Microphone used in our project.

Design Microphone Pre Amplifier circuit

↓

Place the designed circuit at a certain height using mono-pod stand

↓

Put any type of speaker at the same level of height, maintaining a distance 1m between the designed circuit and speaker

↓

Play a sine tone of single frequency (2Khz) into the speaker

↓

Microphone catches the played sound from headphone's speakers and gives output

↓

Record the output or sound using mobile phone's recorder

↓

Revolve the speaker around the designed circuit maintaining the same level of height and same distance between the two

↓

Following the same previous steps, record the data or sound at the interval of 30°on a circle of 1m radius

↓

Process the recorded data in matlab to get the directional pattern of microphone

Directional pattern in Matlab

STEP-1: Calculate the power of received data using the below formulae

$$\text{Power} = \frac{1}{N}\sum_{i=1}^{N} |x(i)|^2$$

STEP-2: Similarly calculate the power of input signal.

STEP-3: Calculate the gain in dB

$$\text{Gain(dB)} = 10 \cdot \log_{10}\left(\frac{\text{Power}_{\text{out}}}{\text{Power}_{\text{in}}}\right)$$

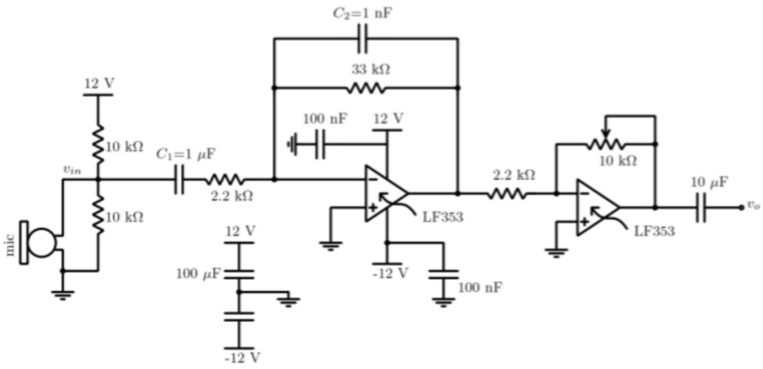


Figure 3: Circuit for microphone interface

Hardware Setup:

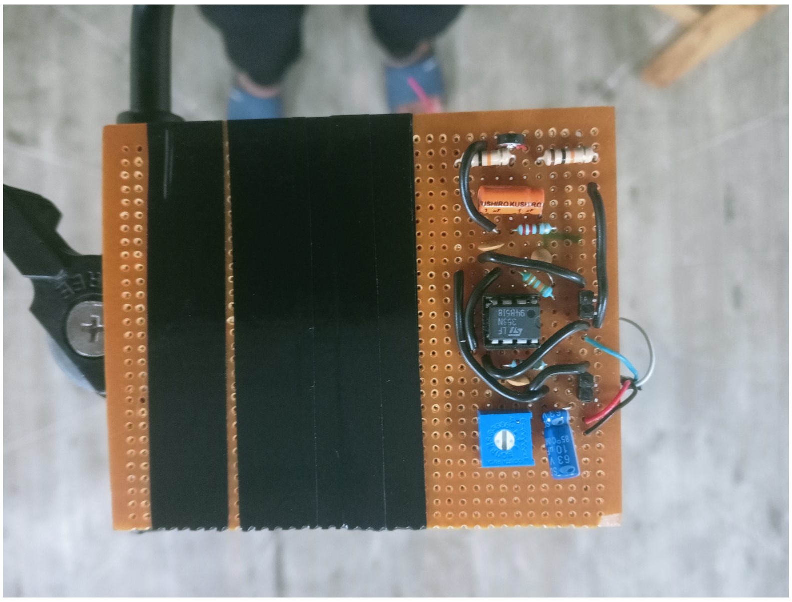


Figure 4: Microphone with pre-amplifier

Figure 5: Microphone at certain Height

Result

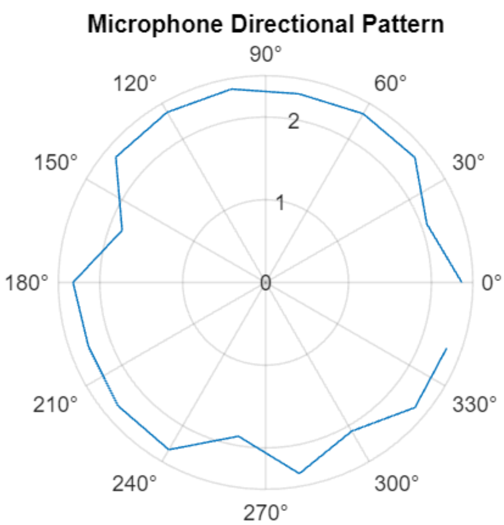


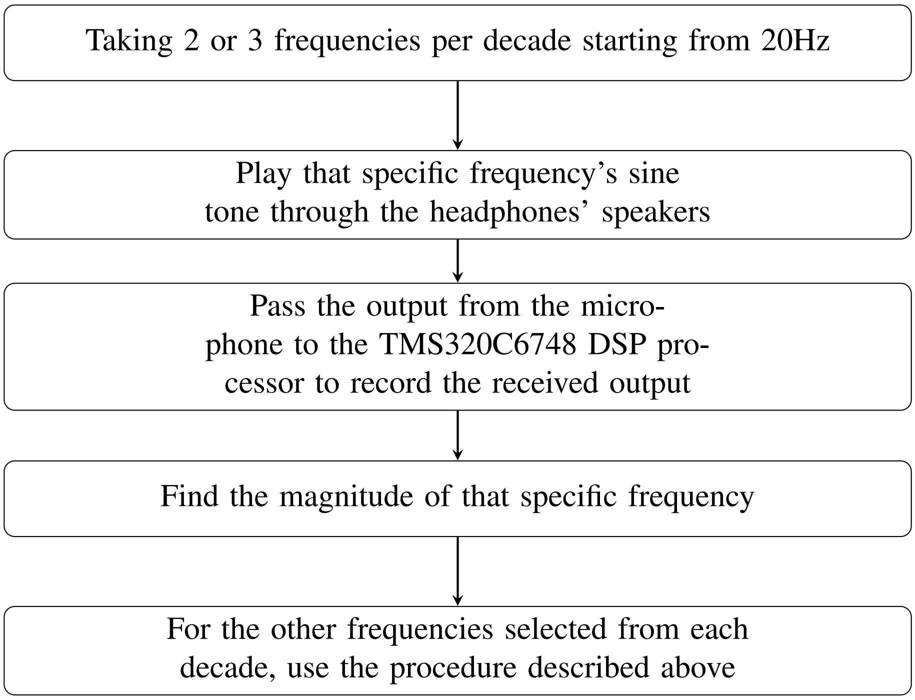
Figure 6: Microphone's Directional Pattern

Conclusion: Microphones are omni-directional

3. Channel Estimation

AIM: To find the channel response of the system.
Objective to estimate the channel response of system: To do acoustic feedback cancellation, we need to identify the behaviour of the system.

3.1. Method 1



3.1.1 Method to find the magnitude response using Matlab

Direct FFT method
STEP-1: Take FFT of recorded sine tone of a specific frequency.
STEP-2: Take magnitude and get the maximum value of magnitude as it gives magnitude response for that specific frequency.
STEP-3: Follow the same above steps to get magnitude response for all other taken frequencies.

Hardware Setup

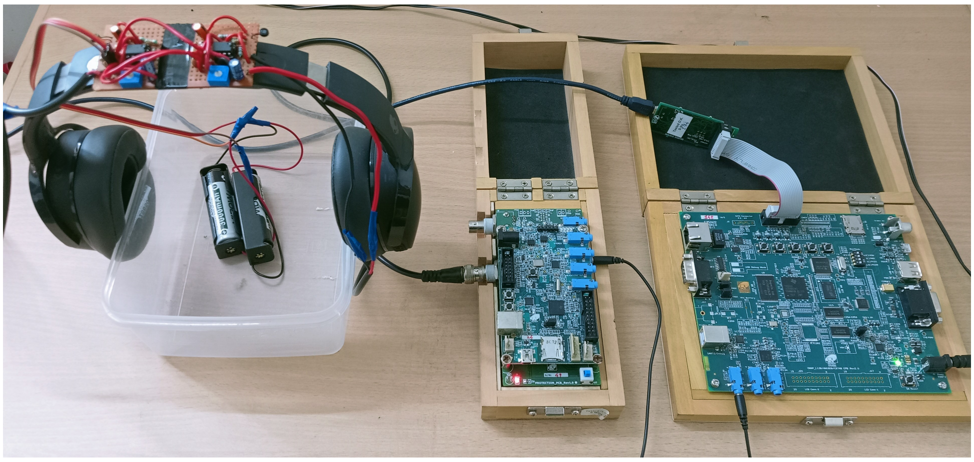


Figure 7: Mics + headphone with DSP Processor

Result

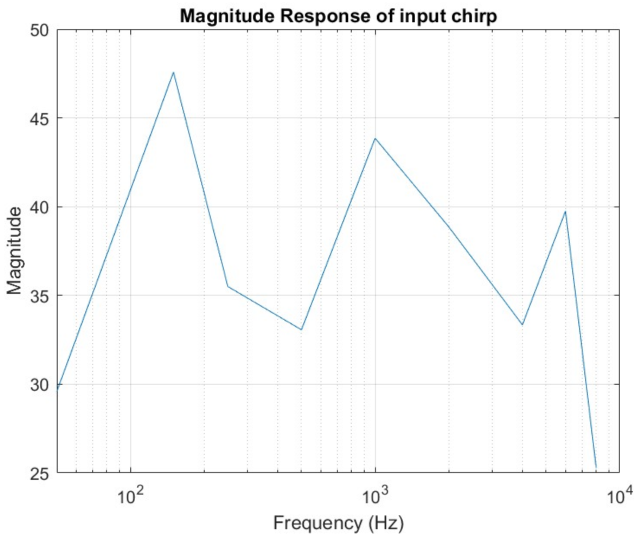


Figure 8: Magnitude Response of system

Problems

- Not able to get phase response, only magnitude response given by this method
- Time consuming
- Recorded data is of very small length(62.5ms) as DSP processor can't have too much memory to record the long data
- Not able to get rid off from noise

3.2. Method 2

Generating the chirp signal with starting frequency 50Hz and ending frequency 16kHz from matlab.

↓

Play this generated chirp signal into the headphones' speakers

↓

Record the output from the microphone through mobile phone's recorder using Aux-cable

↓

Number of Recordings=30

Hardware Setup

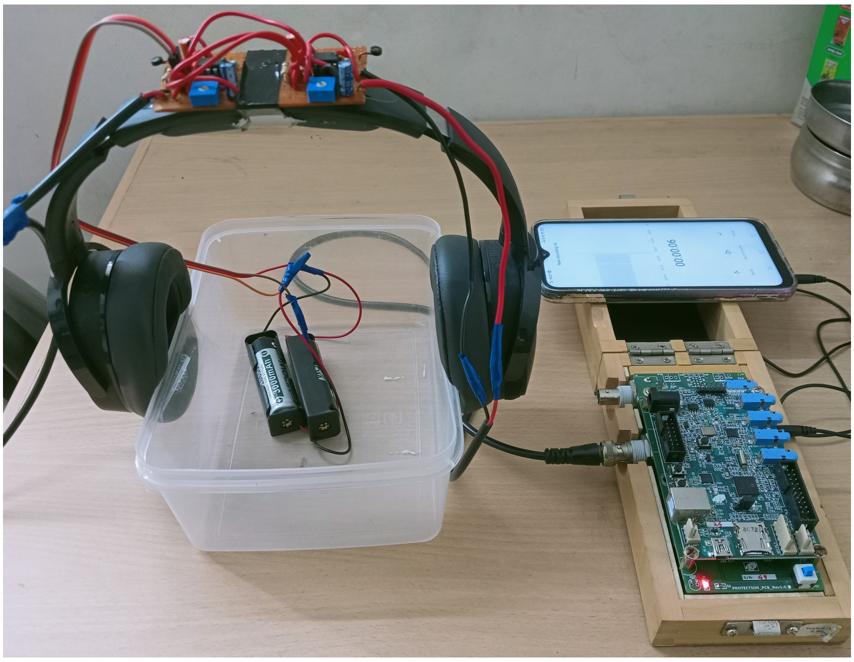


Figure 9: Headphone + Mics Recording through Mobile phone

$$
\begin{gathered}
y_1(t) = x(t) * h(t) + n_1(t) \\
y_2(t) = x(t) * h(t) + n_2(t) \\
y_3(t) = x(t) * h(t) + n_3(t) \\
. \\
. \\
. \\
y_k(t) = x(t) * h(t) + n_k(t)
\end{gathered}
$$

Here, x(t) is the input chirp signal
$y_i(t)$ are the output, i=1,2,3,...
h(t) is the channel
$n_i(t)$ is the noise, i=1,2,3,...

3.2.1 Method 2.a: Averaging in time domain

CONSIDERING THE NOISE

Averaging all the outputs, we get

$$\frac{\sum_{i=0}^{30} y_i(t)}{30} = x(t) * h(t) + \frac{\sum_{i=0}^{30} n_i(t)}{30}$$

Let's define $y(t) = \frac{\sum_{i=0}^{30} y_i(t)}{30}$ and $n(t) = \frac{\sum_{i=0}^{30} n_i(t)}{30}$.

$$y(t) = x(t) * h(t) + n(t)$$

By taking FFT,

$$Y(j\omega) = X(j\omega)H(j\omega) + N(j\omega)$$
$$\implies H(j\omega) = \frac{Y(j\omega) - N(j\omega)}{X(j\omega)}$$

IGNORING THE NOISE

$$Y(j\omega) = X(j\omega)H(j\omega)$$
$$\implies H(j\omega) = \frac{Y(j\omega)}{X(j\omega)}$$

Result

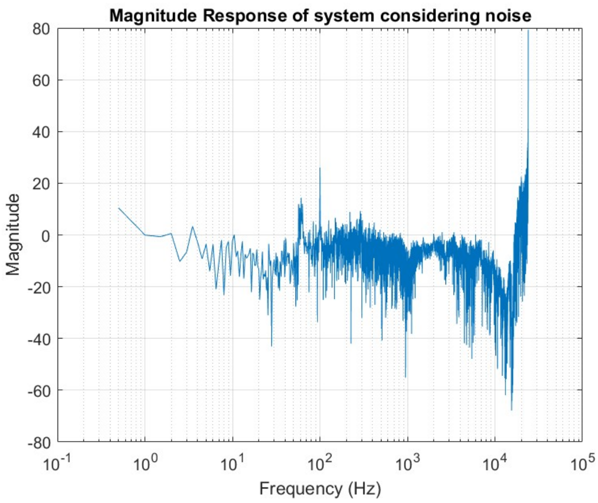


Figure 10: Magnitude Response of the system considering noise

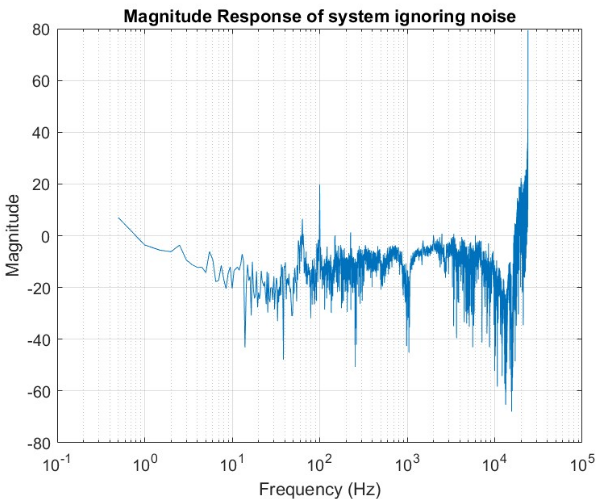


Figure 11: Magnitude Response of the system ignoring the noise

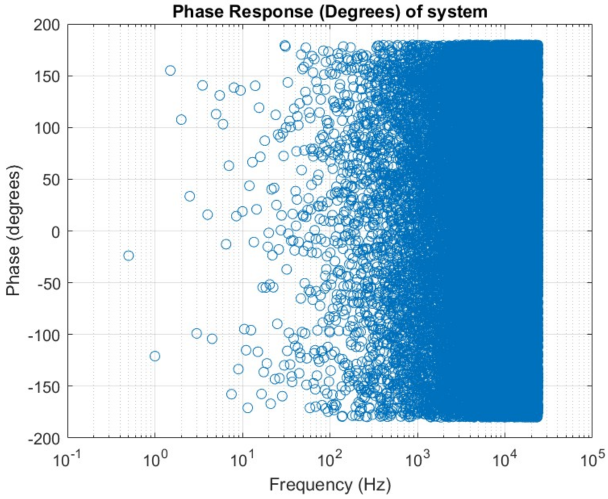


Figure 12: Phase Response (Degrees) of the system

3.2.2 Method 2.b : Averaging in Frequency domain

Ignoring the noise part. Taking FFT of all outputs $y_1(t), y_2(t), ...$ and taking the magnitude of all, we get

$$
\begin{aligned}
|Y_i(j\omega)| &= |X(j\omega)H(j\omega)| \\
&= |X(j\omega)| \cdot |H(j\omega)|
\end{aligned}
\tag{1}
$$

where $i = 1, 2, ...$ Now averaging all the outputs, we get

$$
\frac{\sum_{i=0}^{30} |Y_i(j\omega)|}{30} = |X(j\omega)| \cdot |H(j\omega)| \tag{2}
$$

Let's say $|Y(j\omega)| = \frac{\sum_{i=0}^{30} |Y_i(j\omega)|}{30}$.

$$
|Y(j\omega)| = |X(j\omega)| \cdot |H(j\omega)| \tag{3}
$$

$$
|H(j\omega)| = \frac{|Y(j\omega)|}{|X(j\omega)|} \tag{4}
$$

Result

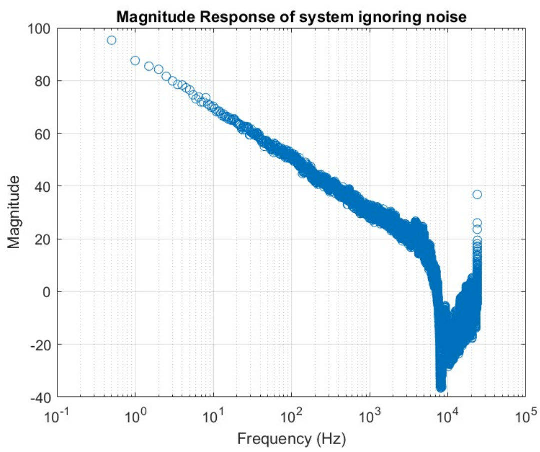


Figure 13: Magnitude Response of system ignoring noise

3.3. Matlab Code Validation

Validating the code by taking an artificial channel

- Taken an artificial channel defined with numerator and denominator coefficients b and a as follow

$$b = [1111] * 1/4$$
$$a = [1]$$

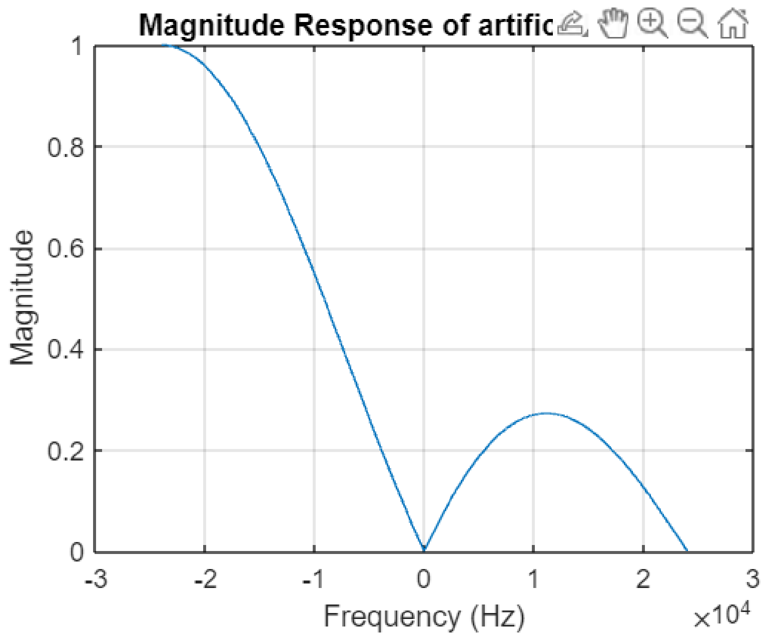


Figure 14: Magnitude Response of artificial channel

- Passing the chirp signal through taken artificial channel without adding the noise

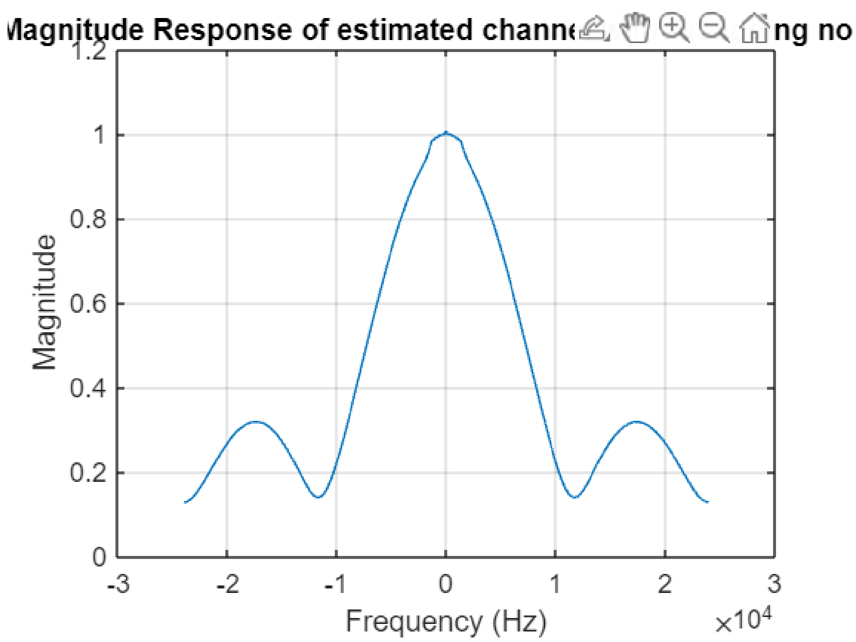


Figure 15: Magnitude Response of estimated channel without adding noise

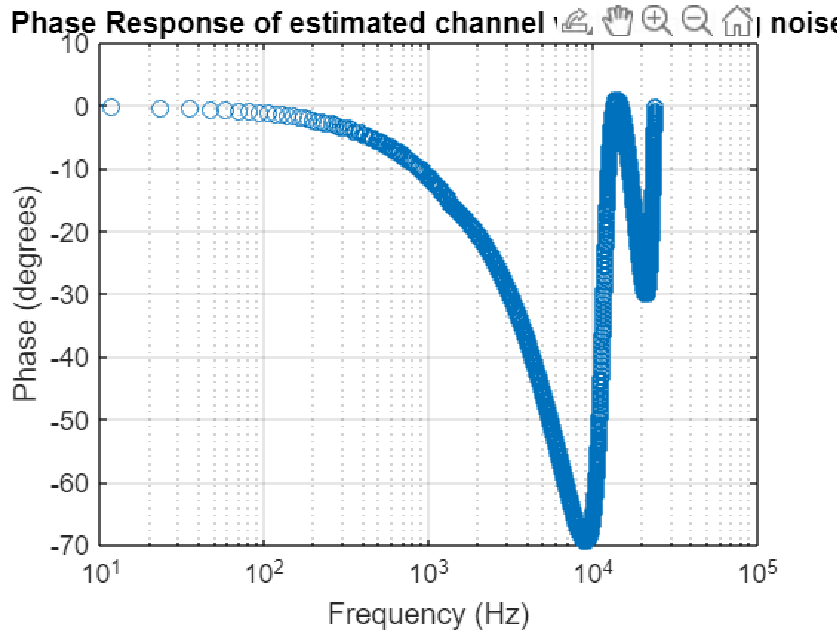


Figure 16: Phase Response of estimated channel without adding noise

- Add the AWGN noise of SNR less than 50dB, let say 45dB

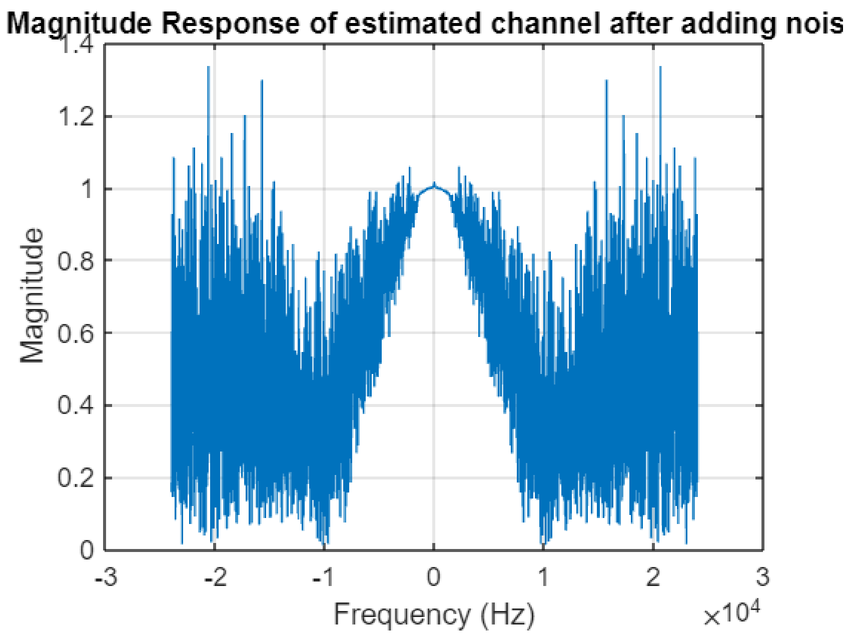


Figure 17: Magnitude Response of estimated channel after adding noise

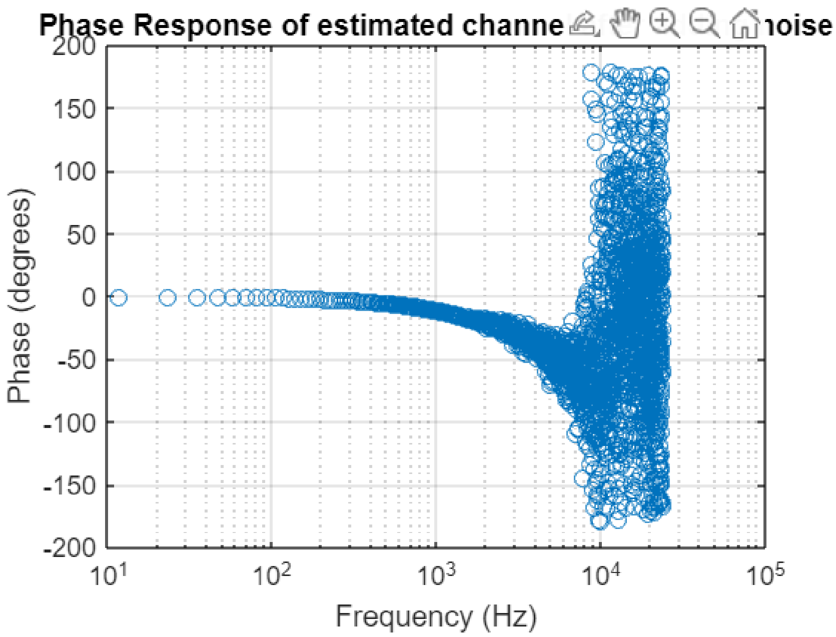


Figure 18: Phase Response of estimated channel after adding noise

Observation: Addition of noise of less than 50dB lead to distort the channel response

Spectrogram of recorded data

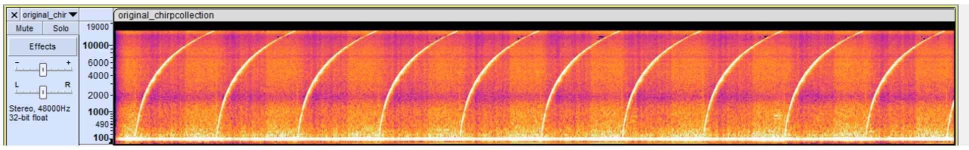


Figure 19: Spectrogram of recorded data

Conclusion from the Observation and results of Method 2:

- Some background Noise is present in recordings
- *Assumption:* Automatic Gain Control (AGC) in a mobile phone's recorder may amplify background noise during quiet parts of the recording. When the input signal is low, the AGC might increase the gain, including any ambient noise present.

Observing this, let's move to different recording method.

3.4. Using DAQ(Data Acquisition System) for recording the data

Advantage: Reduces the chance of getting more noise in recorded signal because it has no AGC to amplify the noise.

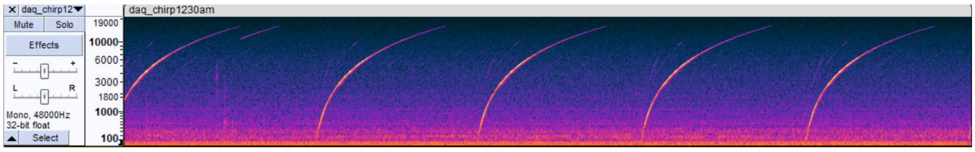


Figure 20: Spectrogram of recorded data

3.4.1 Method 1.0

Take 2 or 3 frequencies per decade starting from 20Hz, generate a single audio file playing sine tone of different taken frequencies one by one

↓

Play this audio file from the matlab through headphones' speakers

↓

Microphone catches the played sound from headphones' speakers and gives output

↓

Record the output from the DAQ

↓

Extract data of each frequency sine tone and find magnitude of that frequency

Direct FFT Method to find the magnitude response in Matlab method

STEP-1: Extract the different frequency sine tone from the recorded data using cross-correlation function of matlab.
STEP-2: Find the index of sample at which cross-correlation gives maximum value. This point indicate the starting of the input sine tone.
STEP-3: Extract the data of same length as of input sine tone.
STEP-4: Take FFT of extracted data.
STEP-5: Take magnitude and get the maximum value of magnitude as it gives magnitude response for that particular frequency.
STEP-6: Follow the same above steps to get magnitude response for all taken frequencies.

Hardware Setup

Figure 21: Mics + Headphone with DAQ

Figure 22: Mics + Headphone

Figure 23: DAQ

Results

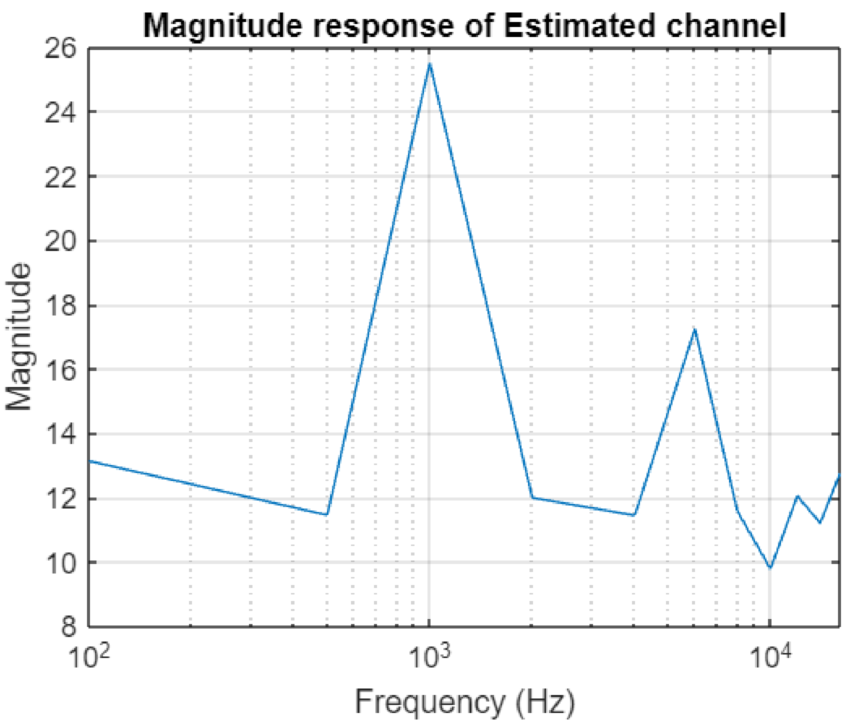


Figure 24: Magnitude response of Estimated channel

3.4.2 Method 2.0

Generate multiple copies of chirp signal with starting frequency 50Hz and ending frequency 16kHz from matlab in a single audio file

↓

Play this generated audio file through headphones' speakers

↓

Record the output from the microphone through DAQ

↓

Extract data of each frequency sine tone using the same cross correlation method defined above

↓

Find magnitude of that frequency

↓

Number of Recordings=12

Results

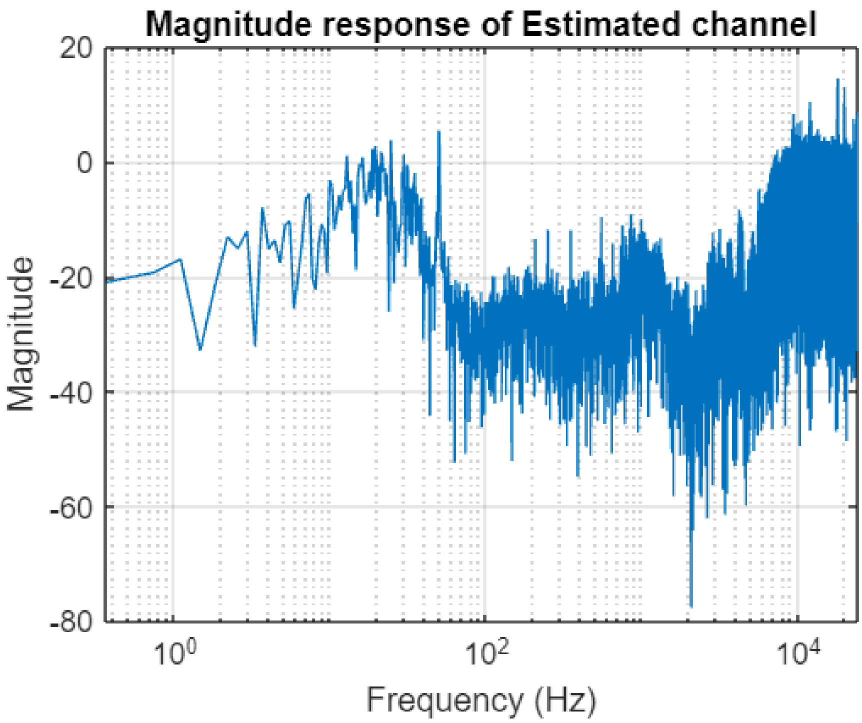


Figure 25: Magnitude response of Estimated channel

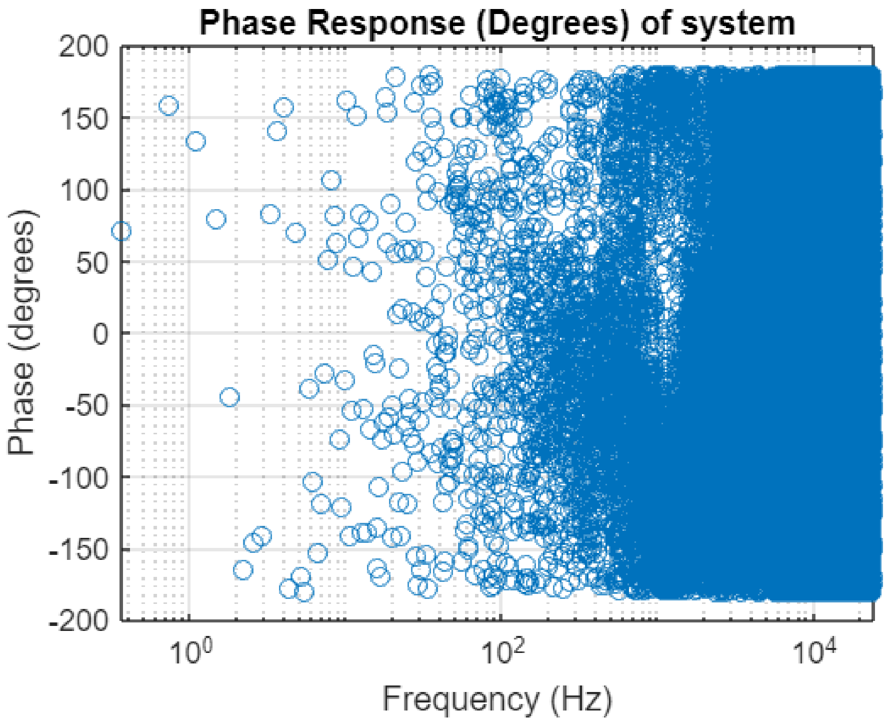


Figure 26: Phase response of Estimated channel

Conclusion

Due to addition of some kind of background noise, the channel response of the system get distorted.

4. BeamForming and Direction of Arrival Calculation

We aim to suppress noise of received audio signal,calculate Direction of arrival using data from 2 microphones

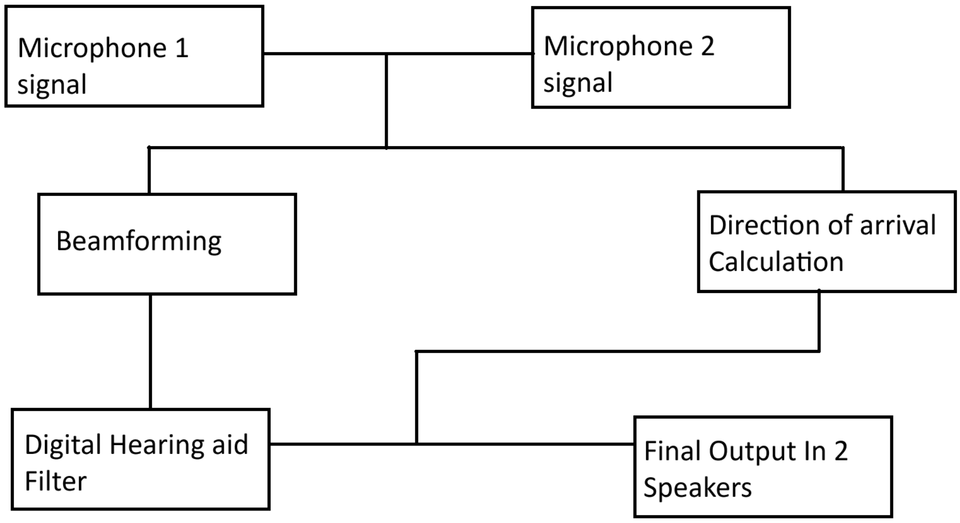


Figure 27: Plan of Action

The first step in this formal process involves the acquisition of two distinct noisy audio signals from two individual microphones, effectively creating a binaural recording. Subsequently, the recorded signals are subjected to analysis to estimate the direction of arrival (DOA) of the sound source and to calculate the time delay between the signals received by the two microphones. This DOA and time delay information becomes instrumental for the subsequent stages of signal processing.

Moving forward, the recorded signals undergo Multi-Channel Minimum Variance Distortionless Response (MVDR) beamforming, a technique designed to enhance the target signal while concurrently suppressing noise. The DOA and time delay information is crucially used to optimize and refine the beamforming process, ensuring the most effective results.

Following the MVDR beamforming, the processed signal is further subjected to a digital hearing aid filter. This filter is meticulously designed to enhance speech intelligibility and mitigate background noise, thereby significantly improving the overall quality of the listening experience.

Additionally, based on the calculated time delay, an appropriate delay is introduced to synchronize the two enhanced signals. This synchronization guarantees precise alignment for the listener's perception of the sound.

Finally, the enhanced and synchronized signals are routed to two separate speakers, completing the formal process. This approach ensures that the listener experiences the improved and clarified sound in a spatially accurate manner, allowing for a superior auditory experience.

4.1. Direction of Arrival Calculation

4.1.1 Using Cross Correlation Method

With two microphones spaced 10 centimetres apart, cross-correlation can estimate the direction of arrival (DOA) of a sound source. The time delay between signals received by the microphones is calculated by the lag with the highest cross-correlation. This time delay can then be used to determine the DOA based on the array's geometry.

For Delay-5		**For Delay-10**	
SNR	Angle of Arrival	SNR	Angle of Arrival
-5	0	-5	0
0	0	0	0
5	0	5	-45.099
10	-20.742	10	-45.099
inf	-20.742	inf	-45.099
Expected Angle	-21		-45

Table 1: Angle of arrival using cross correlation

Observation At lower signal-to-noise ratios, it becomes challenging to reliably discern the time delay using cross-correlation.

4.1.2 Using Least Mean Square Algorithm

Consider two signals $x[n]$ and $d[n]$, and then consider the filter $w[n]$ such that:

$$d[n] = x[n] * w[n]$$

where $*$ represents the convolution operator. Applying the LMS algorithm to $x[n]$ and $d[n]$ will theoretically give $w[n]$ as an output. As shown in the picture below, the LMS algorithm has two inputs, $x[n]$ and $d[n]$, and three outputs, $y[n]$, $e[n]$, and $w[n]$.

The algorithm has two main parameters, $M = 50$ and $\mu = 0.0117$. M is the order of the filter, and μ, called the step-size, controls the convergence of the algorithm. The coefficients of the filter $h[n]$ are called weights and will be represented as a vector $w[n]$. LMS consists of two main processes, the filtering and the adaptive process. The maximum delay possible with the setup is $N_{\text{max}} = 15$. This extra delay is added to the signal called DESP, which is the third parameter of the system.

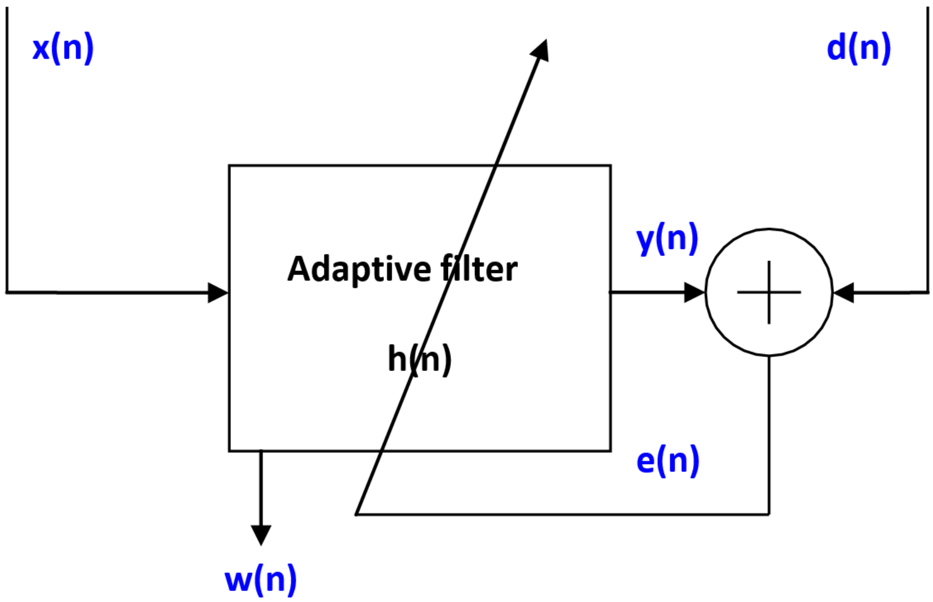


Figure 28: LMS Algorithm

Filtering Process

First, $w[n]$ must be initiated with an arbitrary value $w[0]$. At each instant k, the input $x[n]$ we are working with is:

$$x[n] = [x[k], x[k-1], x[k-2], \ldots, x[k-M+1]]$$

which are the last M samples of the signal $x[n]$. Then $y[n]$ is obtained by filtering $x[n]$ with $h[n]$:

$$y[n] = w[n]H * x[n]$$

where $w[n]H$ is the conjugate transpose of $w[n]$. After that, the estimated error function $e[n]$ is calculated as:

$$e[n] = d[n] - y[n]$$

Adaptive Process

The new weights are then calculated, based on the previous $w[n]$ and the error function $e[n]$:

$$w[n+1] = w[n] + \mu \cdot x[n] * e^T[n]$$

The successive corrections of the weight vector eventually lead to the minimum value of the mean square error.

As shown in the equation, the parameter μ is of crucial importance. It controls the convergence of the algorithm, which is essential in a real-time system. Considering $R[n]$ the correlation matrix of $x[n]$, which is obtained as:

$$R[n] = x[n] * x[n]H$$

and let $\lambda_{\max}$ be the largest eigenvalue of the matrix R. It is proved that the LMS algorithm converges and stays stable if the step-size satisfies the following condition:

$$0 < \mu < \frac{1}{\lambda_{\max}}$$

The delay calculation is composed of three steps: First obtaining the Fast Fourier Transform of $x[n]$, then its phase, and finally, the delay N. The FFT is an efficient algorithm used to calculate the Discrete Fourier Transform (DFT), so it will help to switch from the time domain to the frequency domain. Considering an ideal situation, the filter would be a pure Dirac delta delayed N' samples.

$$h[n] = \delta[n - N']$$

$$h[n] = \{0, 0, \dots, 1, \dots, 0\}$$

This transform has two main components: modulus and phase. Since the desired information is the position of the delta (N'), only the phase will be useful (the modulus only has information about amplitude).

$$H_j = 1 \cdot e^{-2\pi jN'/n} \quad \Omega = -2\pi jN'/n$$

With the derivative, the variable j disappears, and the slope S is obtained.

$$S = -2\pi N'/n$$

$$N' = -S \cdot n/(2\pi)$$

At this point, the parameter DESP must be subtracted to get the right delay.

$$N = N' - \text{DESP}$$

After calculating the value of N, the delay in time τ is

$$\tau = N/f_s$$

‘

For Delay-5		For Delay-10	
SNR	Angle of Arrival	SNR	Angle of Arrival
-5	1.25	-5	2.253
0	1.843	0	3.331
5	3.012	5	4.375
10	3.214	10	4.825
20	3.456	20	6.634
inf	5.312	inf	9.606

Table 2: Table for Delay-5 and Delay-10 using LMS algorithm

Conclusion Results obtained from our study indicate that correlation-based methods are effective primarily at higher Signal-to-Noise Ratios (SNR), exhibiting limitations in accurately estimating time delays at lower SNR values. Conversely, the LMS (Least Mean Squares) algorithm, owing to its adaptive nature, demonstrates better performance at low SNR values.Whereas Cross correlation method performs better at high SNR value.

4.2. Minimum Variance Distortionless Response (MVDR) Beamforming

$$Y(f,t) = d(f)S(f,t) + N(f,t),\ Y, S, N \epsilon C^{M\times F\times T}$$

where Y, S and N are the STFTs of the noisy speech, target clean speech, and the noise,respectively. d is the steering vector.

$$\hat{S} = w^H Y,\ w \epsilon C^{M\times F}$$
$$w_{MVDR}(f) = arg_w min w^H(f)\phi_{YY}(f)w(f)$$
$$\text{s.t. } w(f)^H d(f) = 1$$

The MVDR beamformer is obtained by minimising the power of the interfering signal while preserving the distortionless source signal.
Here $\phi_{YY}(f)$ is the covariance matrix of the noisy STFT Y at the frequency bin f.

4.2.1 Reference channel selection MVDR beamforming

$$w_{MVDR}(f) = \frac{\phi_{NN}^{-1}(f)\phi_{SS}(f)}{Trace(\phi_{NN}^{-1}(f)\phi_{SS}(f))}u$$

where ϕ_{NN} and ϕ_{SS} are the covariance matrices of the noise and speech respectively. u is a one-hot vector representing the selected reference channel (Usually the first microphone is set as the reference one, i.e., u[0]=1).
The covariance matrix SS of the target speech could be obtained as

$$\phi_{SS}(f) = \frac{\sum_{t=1}^{T}\hat{S}(t,f)\hat{S}^H(t,f)}{\sum_{t=1}^{T}M_s(t,f)M_s^H(t,f)}$$
$$\hat{S}(t,f) = M_s(t,f)Y(t,f)$$

Where M_s is the Time-Frequency mask of the target speech. The covariance matrix of Φ_{NN} can be estimated in a similar way by replacing the speech mask M_s with the noise mask M_n. Φ_{SS} could be normalized by the number of frames T, in this case, we normalize it by the sum of the mask power since there is not always speech present in the frames.

The task of estimating the MVDR beamforming weight thus is transformed to estimating the Time-Frequency masks of the noisy STFT. We can apply statistical methods to estimate it, or we can train a neural network to predict the mask. Even more, we can combine the mask prediction network with the beamforming method to make the whole beamforming process differentiable (For this, we have used the work of Prof. Nitya Tiwari, which estimates speech and noise).

To evaluate speech Quality, the PESQ indicator is used.

PESQ stands for Perceptual Evaluation of Speech Quality. Spearline uses this in its global in-country number testing. This objective and recognized industry-standard audio quality measure takes into consideration characteristics such as: Audio sharpness, Call volume, Background noise, Variable latency or lag in audio, Clipping, Audio interference. It gives a rating for given speech from -0.5 to 4.5.

SNR	Original	Noisy	Single Enhancement	Reference Channel MVDR Beamforming
-5	4.5	0.9149	1.1710	1.0771
0	4.5	1.1781	1.5479	1.4063
5	4.5	1.4680	2.0263	1.7260
10	4.5	1.7913	2.3691	2.0001
20	4.5	2.3895	2.6686	2.1323

Table 3: PESQ Results for Different SNR Levels using Single Channel MVDR

The quality of the reference-based MVDR-enhanced signal is observed to be inferior to that of single-channel enhancement but demonstrates a slight improvement over the quality of the noisy signal.This occurrence can most likely be attributed to the inherent noise present in our reference signal, thereby leading to inaccuracies in the MVDR weight estimation. To address this issue, we employed a steering vector estimation MVDR model to compute the weights.

4.2.2 Steering vector MVDR Beamforming

The basic idea is to directly estimate the steering vector using the covariance matrix of a microphone image of a target speech signal. Specifically, we utilise the principal eigenvector of an estimate of the covariance matrix as an estimate of the steering vector. The covariance matrix can be estimated by using time-frequency masks as described below.

Let $M_n(f, t)$ denote the time-frequency mask that represents the probability of the time-frequency point (f, t) containing only noise.

Then, we can estimate the covariance matrices of noise as

$$\phi_{nn} = \frac{1}{\sum_t M_n(f, t)} \sum_t M_n(f, t) Y(f, t) Y^H(f, t)$$

respectively. Then, the desired covariance matrix for the target speech signal is obtained by

$$R_f^x = \frac{1}{T} \sum_t Y(f, t) Y^H(f, t) - \phi_{nn} \tag{5}$$

An estimate of the steering vector can be obtained by first performing eigenvector decomposition on R(x)f and then extracting the eigenvector associated with the maximum eigenvalue.Let V be an eigenvector associated with the maximum eigenvalue.

$$W_{MUDR} = \frac{\phi^{-}1_{nn}(f)V(f)}{V(f)^{H}\phi_{nn}(f) * V(f)} \tag{6}$$

All the data presented herein has been computed as the average over a set of five audio samples.

SNR	Original	Noisy	First Mic Single Enhancement	Reference Channel MVDR Beamforming	Steering Beam MVDR
-5	4.5	0.9149	1.1710	1.0771	1.9241
0	4.5	1.1781	1.5479	1.4063	2.4812
5	4.5	1.4680	2.0263	1.7260	2.7345
10	4.5	1.7913	2.3691	2.0001	2.8529
20	4.5	2.3895	2.6686	2.1323	3.1399

Table 4: PESQ Results for Different SNR Levels using Steering MVDR

This data encompasses two distinct types of noise samples, specifically, white noise and babble noise. Furthermore, the calculations are based on five distinct SNR values, specifically, -5 dB, 0 dB, 5 dB, 10 dB, and 20 dB. The utilization of these averaged samples and variations in noise levels contributes to a comprehensive analysis of the audio processing performance under various conditions.

4.2.3 Conclusion

In summary, our endeavor has yielded a successful outcome, as we have effectively suppressed noise and notably improved the audio quality. The robust approach of multi-channel MVDR beamforming, coupled with digital hearing aid filtering, has demonstrated its efficacy in enhancing speech intelligibility and reducing background noise. The refined audio output resulting from this process holds great promise for further applications, particularly in the deployment of hearing aid filters. This accomplishment underscores the potential for our methods to significantly enhance the auditory experience in various real-world scenarios.

5. EDL Project work

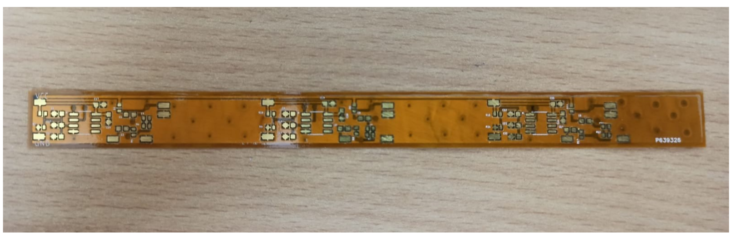


Figure 29: Microphone array with amplifier circuit on Flexible-PCB

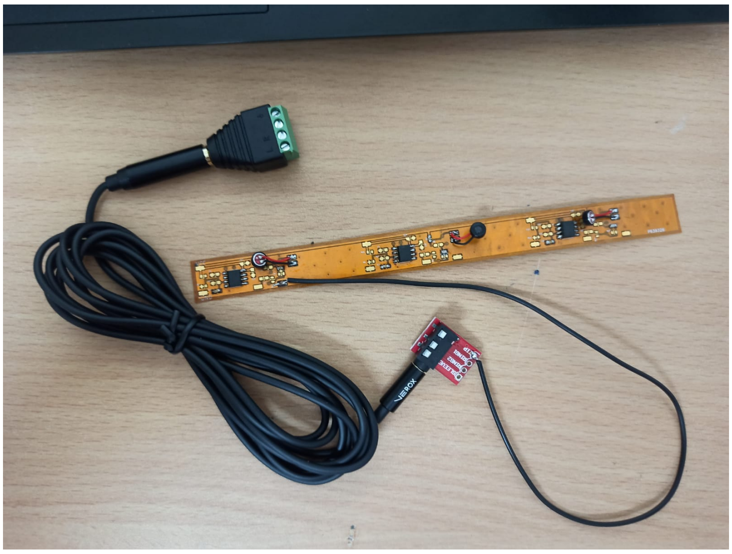


Figure 30: Microphone array with amplifier circuit on Flexible-PCB

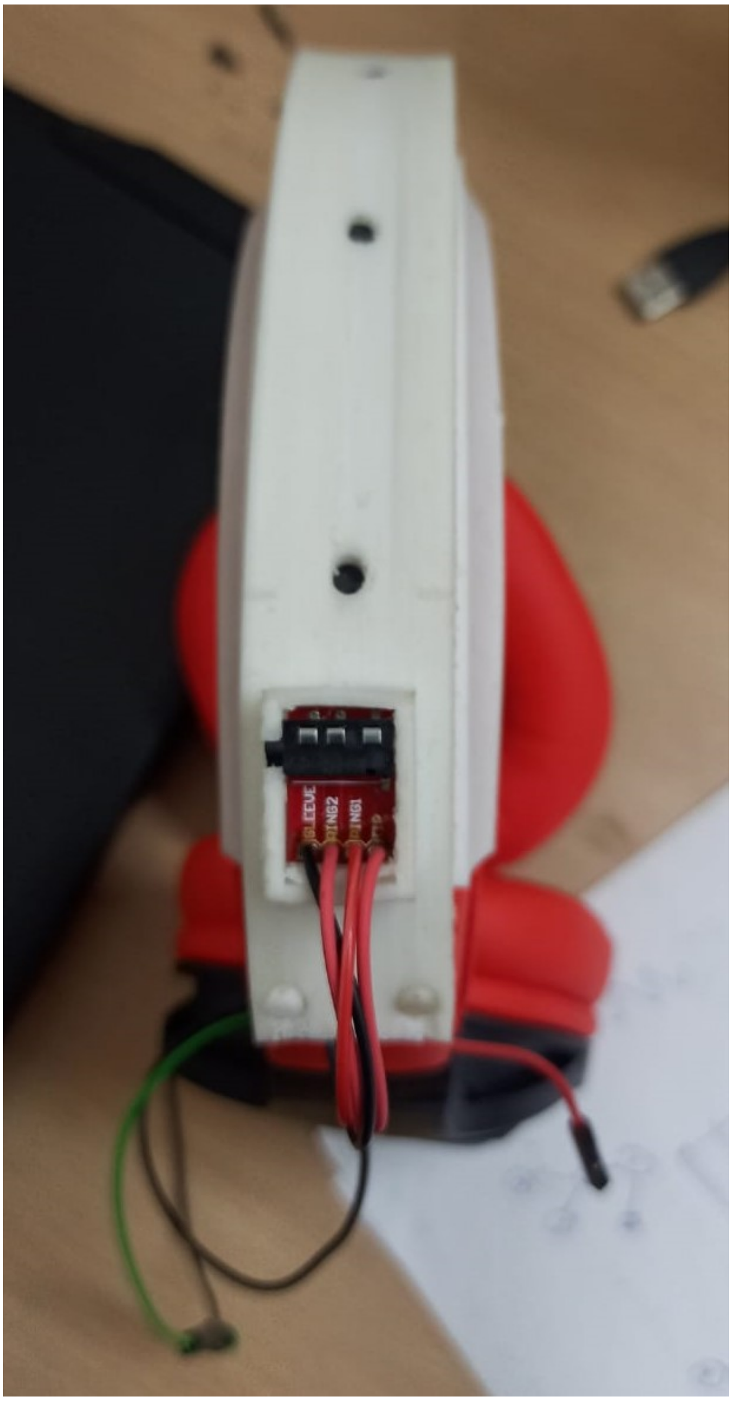


Figure 31: Microphone array with amplifier on Flexible-PCB mounted on top-of headphone

Figure 32: Microphone array with amplifier on Flexible-PCB mounted on top-of headphone

6. Future Strategies

1. Begin by employing the Least Mean Squares (LMS) algorithm to assess the system's channel response through an adaptive filter.

2. Gain insights into hearing aid filters and integrate them into our system.

3. Conduct calculations for Direction of Arrival and implement Beamforming through Digital Signal Processing Kit (TMS320C6748).