Directed Microphone Array Optimal Linear Filter for Noise Suppression

Nicolas Alarcon
Electrical Engineering
Columbia University
New York, USA
na2946@columbia.edu

Atri Ray
Electrical Engineering
Columbia University
New York, USA
arr2232@columbia.edu

***Abstract*—The purpose of this project is to create and demonstrate beamforming in a simulated noise environment. The results will be directly used in our senior design project, the LENSCRIBE real-time subtitle augmented reality (AR) glasses. We will simulate environmental noise and use a pure 261.6 Hz tone as the target signal. We will then implement several beamforming procedures to pick out the target signal from environmental noise. We then classify the resolution of the beamformed signal by measuring the variance of the output.**

I. Introduction

The goal of LENSCRIBE glasses is to serve as a communication aid. The glasses will feature a real-time subtitle system that transcribes spoken words into text and displays them on the lens. The system will utilize a beamforming directed microphone array and an AI-integrated speech-to-text machine (STM) to capture and process the speech. The transcribed text will be displayed onto AR glasses. This project will be broken up into 3 main subsystems: the microphone array, the STM, and the display. All components will be integrated into a pair of 3D-printed glasses, with the necessary interfaces and power systems. [6]

Originally developed for sonar and radar applications, beamforming has become a critical component in modern wireless communication, including multi-input multi-output (MIMO) systems and 5G networks. It improves signal-to-noise ratio (SNR) and supports efficient spectrum and energy usage, which are vital in environments with high user density or interference. [7]

Beamforming is a signal processing technique that enables directional transmission or reception of signals by utilizing arrays of antennas or sensors. By adjusting the phase and amplitude of signals across the array, beamforming focuses energy in a desired direction, enhancing signal strength while suppressing interference and noise from other directions.

Figure 1 shows a graphic displaying an example of *delay and sum* beamforming.

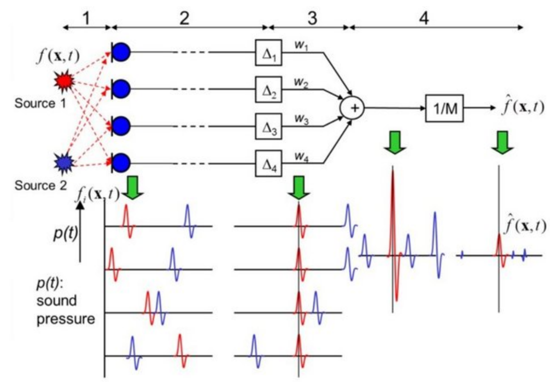

Fig. 1. A simple graphic displaying delay and sum beamforming [1]

II. Technical Approach

A. Matlab

MATLAB was selected for this project due to its powerful computing capabilities and its extensive support for signal processing applications. Its programming structure allows for efficient manipulation of multi-dimensional data, which is perfect for the implementation of beamforming algorithms that rely heavily on matrix operations and vectorized computations. The Signal Processing Toolbox in MATLAB provides several pre-built functions for Fourier transforms, spectral analysis, and filter design, significantly reducing the development time for both time-domain and frequency-domain beamforming techniques.

Moreover, MATLAB's visualization capabilities and easy-to-use audio output allowed for the efficient analysis and comparison of beamforming performance metrics.

B. Simulated Hardware

The IMP23ABSU microphone was selected as the simulated sensor for this project due to our intention of using it in the LENSCRIBE system. [8]

We simulated this system by recreating the frequency response from the datasheet. We show this graph in Figure 2. From here, we interpolated data, and fit the frequency response to a 5th degree polynomial, applying them to the microphone input signal.

LENSCRIBE will have an array of four microphones arranged in a single line along one of the legs, so a linear microphone array was configured to capture acoustic signals for the beamforming simulation. The array consists of $N = 4$ microphones arranged along the y-axis with uniform spacing

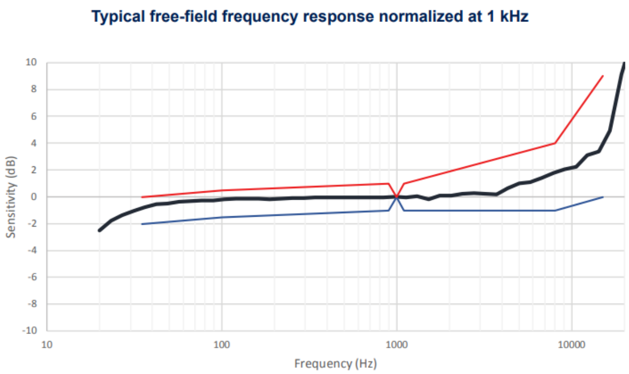

Fig. 2. The frequency response of the IMP23ABSU microphone.

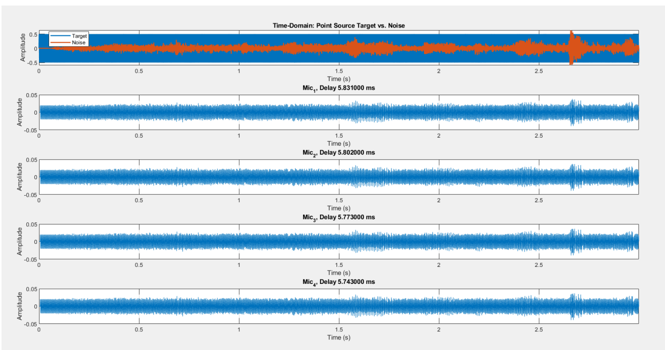

Fig. 3. The Delays Each Microphone Experiences.

$d = 0.01m$. This spacing was chosen to be below the spatial Nyquist criterion for the frequencies of interest. For signals below $8000kHz$ and below for an approximate speed of sound of $343m/s$, a microphone spacing of less than $0.02m$ minimizes spatial aliasing in received signals. The position of each microphone is given by (x, y) coordinates.

The target audio source was positioned at $(0, 2)m$ to represent a point source located directly in front of the microphone array at a distance of 2 meters.

Figure 3 shows the delay each microphone experiences in receiving the signal.

The target audio used was a straight tone of 261.63 Hz, which is commonly known as "Middle C" or "C4".

C. Noise

The noise environment for the beamforming simulation was designed to model realistic acoustic interference, incorporating both spatial distribution and temporal effects. The audio file that was used for environmental noise was collected by walking through Uris Library with a microphone and was then truncated in length to match the target audio signal.

The noise sources were spatially distributed along the parabolic curve

$$y = 2x^2 + 2x \quad \{x \in \mathbb{Z} \mid -2 \leq x \leq 2\}$$

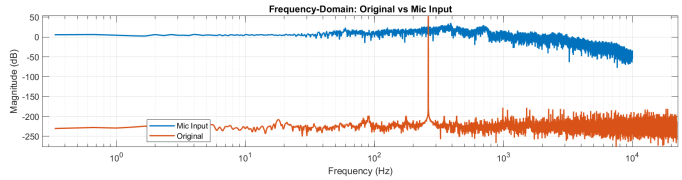

Fig. 4. The Target Signal and the Pre-Beamformed Signal with Noise.

with microphones positioned at the origin. This arrangement was chosen to mimic a typical real-world scenario where noise sources are distributed non-uniformly around a target source.

Both the original signal and the signal with noise are shown in figure 4.

D. Simulated Audio to Digital Conversion

The audio-to-digital conversion pipeline was simulated to approximate the practical behavior of a microphone and its associated analog front-end prior to sampling by an analog-to-digital converter (ADC). The simulation incorporated a measured or specified microphone frequency response, followed by an anti-aliasing filter (AAF), ensuring that the signal entering the digital stage accurately reflects realistic hardware characteristics.

Starting from the signal and noise sources, modeled as point sources, the audio signal was modeled as an air pressure wave, with attenuation inversely proportional to distance and a sample delay equal to $Fs * d/c$ where $c = 343m/s$ and Fs is the sampling frequency of the input signal. This results in the microphone input signal that is attenuated by H_{mic}.

To prevent spectral overlap and signal distortion during the sampling stage, a fourth-order Bessel low-pass filter was implemented as the AAF. Bessel filters are known for their maximally flat group delay, which preserves the temporal integrity of the audio signal. By selecting a cutoff frequency near the upper limit of the audio bandwidth, 8 kHz, the filter effectively attenuates higher-frequency components that would otherwise alias into the baseband after sampling. This filter will be implemented in hardware using two second-order sections per microphone.

Both the microphone frequency response $H_{mic}(f)$ and the AAF transfer function $H_{AAF}(f)$ were applied to the microphone array data in the frequency domain. The Bessel filter was designed in continuous time and converted to discrete time, such that the correct frequencies were attenuated in the discrete time signal.

By multiplying the spectra of the microphone input signals with $H_{mic}(f)$ and $H_{AAF}(f)$, the simulation accounted for the amplitude and phase modifications introduced prior to sampling.

The end product of this stage of the code is the microphone transfer function. We need this transfer function to be accurate to the continuous time response of the input audio signal as it reaches the ADC to be representative of the real signal that LENSCRIBE would have to process.

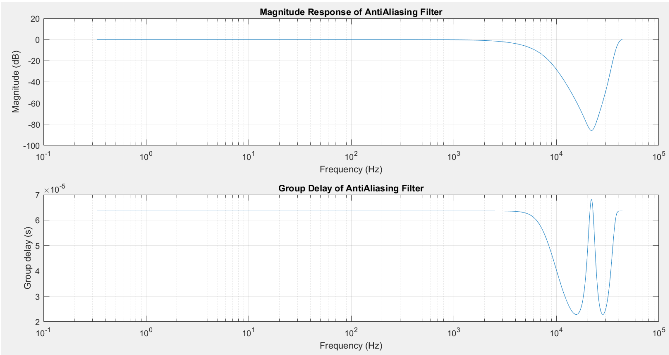

Fig. 5. The Magnitude Response and Group Delay of the AAF.

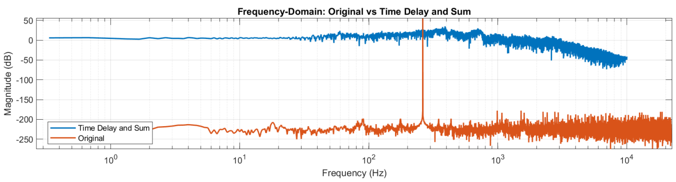

Fig. 6. The Result of the Time Delay Figure and Sum Beamforming.

Figure 5 shows the magnitude response (above) and the group delay (below) of the anti-aliasing filter.

E. Time Domain Delay and Sum Beamforming

Beamforming using the Time Domain Delay and Sum strategy is the simplest form of beamforming. It involves shifting each of the signals received from the microphone by a predetermined amount for static steering and then averaging the signals. If these delays are correct, we will produce a single high output signal.

First, the propagation delays, which we measured earlier in seconds, are converted to sample indices using the sampling frequency fs. For these non-integer delays, fractional sample shifts were applied using linear interpolation. This approach approximates the signal values at fractional sample points by weighting the contributions of adjacent samples. This process involves:

- Index Mapping, which is shifting the time axis by the fractional delay and clamping the resulting indices within valid sample bounds.
- Linear Interpolation, which is computing a weighted average of the two nearest samples, where the weights are proportional to the fractional distance from the shifted index to the surrounding integer indices.

We then average the signals to create a single high signal that is the reconstructed target signal.

The signal reconstructed with the time domain delay and sum strategy is show in figure 6.

F. Frequency Domain Delay and Sum Beamforming

The Frequency Delay and Sum Beamforming method is the frequency domain implementation of the time domain delay and sum. An ideal delay by k samples, $D_k = \delta\left[n-k\right]$ can be

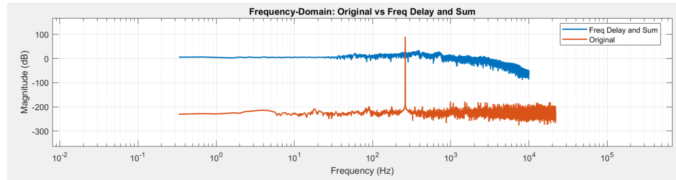

Fig. 7. Result of the Frequency Delay Beamforming

represented in the frequency domain as $e^{-j2\pi k}$. Computing the relative distance d between the uniform linear array elements and the source results in the relative number of delays needed for each microphone $\Delta k = \Delta f_s * d/c$, which can then be applied to the signal from each microphone and averaged.

G. Bartlett Beamforming

Bartlett Beamforming is a beamscan algorithm that consists of using a steering vector $v(\theta)$ to collect the array response at each angle. By calculating the power at every angle, the angles where power is maximized represent the direction of signal sources. This can be used for Angle of Arrival Estimation and Source Separation. However, with a known angle of arrival, the original signal can be reconstructed $y = v^H(/theta)X$ without steering, where X is the input signal to the beamformer and v^H is the complex transpose of the steering vector, and the steering vector can be computed

$$\mathbf{v}(\theta) = \begin{bmatrix} 1 \\ e^{-j2\pi f_s \tau_1(\theta)} \\ e^{-j2\pi f_s \tau_2(\theta)} \\ \vdots \\ e^{-j2\pi f_s \tau_{M-1}(\theta)} \end{bmatrix},$$

where $\tau_m(\theta) = \frac{d \cdot m \cdot \sin(\theta)}{c}$ represents the time delay for the m-th element.

The power output $v^H * R_{xx} * v$ in dB for each steering vector ranging from -90 to 90 degrees results in figure 7, where v^H is the conjugate transpose of the steering vector, R_{xx} is the autocorrelation matrix of the signal, and the peaks correspond to respective source signals. This was verified experimentally in figure 8 by comparing the microphone array when steered towards, and perpendicular to, the target signal. The attenuation when facing away from the signal demonstrates that the beamformer was directing the signal.

III. Experimentation: Resolution Measurements

Various methods were explored for measuring the effectiveness of the various beamformers. signal to noise ratio (SNR) and variance were two of the metrics we explored, however, conclusive results were not reached with the SNR and our method for quantifying it needs to be refined. Variance was then our metric, as lower variance for a normalized audio signal power spectrum should indicate less noise interference. However, our results were inconclusive there as well, as all

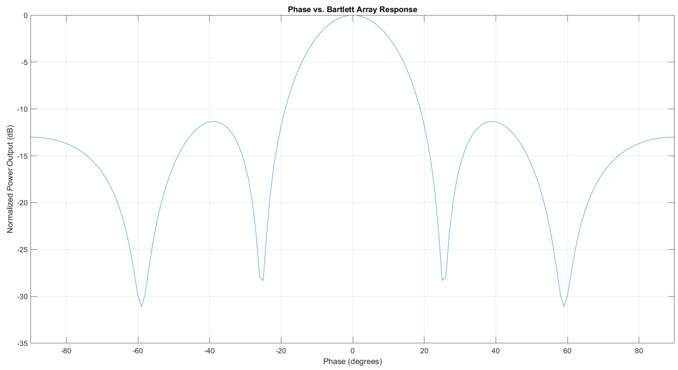

Fig. 8. Bartlett Beamformer Spatial Spectrum Outputs

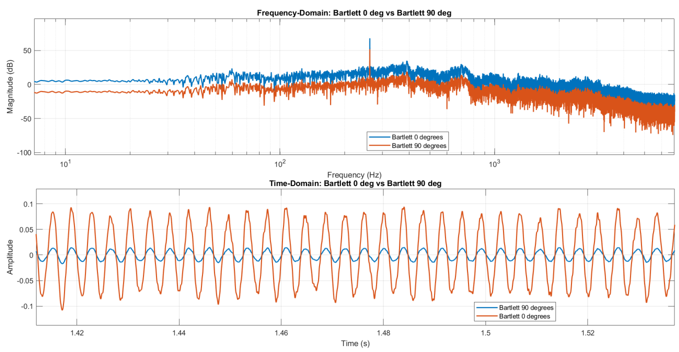

Fig. 9. Caption

three beamforming methods yielded the same variance of $3.333335 * 10^{-5}$ in figure 9.

TABLE I
VARIANCE MEASUREMENTS FOR EACH BEAMFORMER

Metric	Variance
Target + Noise	1.511719×10^{-5}
Time Delay and Sum	3.333335×10^{-5}
Freq Delay and Sum	3.333335×10^{-5}
Bartlett	3.333335×10^{-5}

IV. DISCUSSION

A. *Effectiveness of Beamforming Approaches*

We determined that there is not a significant difference in the effectiveness in our implementations of beamforming. Given that the Delay and Sum strategy takes significantly fewer resources for a static direction than the alternatives, it would appear that this is the optimal method to use for the beamforming in LENSCRIBE. Due to these tests being so close, however, it is impossible to declare these results as anything other than inconclusive.

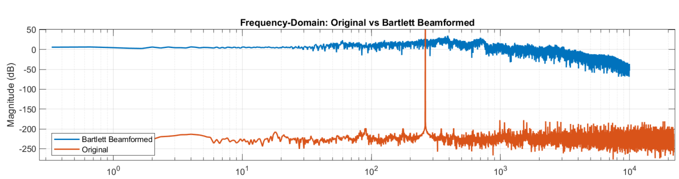

Fig. 10. Bartlett Beamforming Result

B. *Brief Notes on Effectiveness of Measurements*

We would like to clarify that, despite MATLAB having extensive toolkits that we could have used to perform the beamforming, such as the phase-array package, we implemented all of the beamforming methods ourselves. This may have led to less significant data or inefficient data collection.

It is likely that our methods of measuring SNR and Variance were not as accurate as we had hoped, leading to worse data than we had intended. Our metrics were designed to capture improvements in signal clarity and consistency, but the lack of standardized reference conditions or carefully controlled validation procedures might have led to inaccuracies. In practice, this may have resulted in the underestimation of beamforming performance.

We only tested the beamforming on the note ”middle C.” One test is not enough to draw any meaningful conclusions, but we were able to demonstrate somewhat successful noise suppression. Real speech involves a wide spectrum of frequencies so future tests should incorporate multiple signal types and a range of acoustic environments to validate the scalability of the demonstrated methods.

C. *Future Goals*

As the end goal for this beamforming microphone array is use in the LENSCRIBE system, this simulation project will not end with the Digital Signal Processing class. We intend to sharpen the measurements of the variance and the SNR. In addition, we intend to improve the noise simulation, as we think our solution had a significant impact on the effectiveness of our system. We intend to demonstrate even more effective noise suppression through beamforming in the future.

REFERENCES

[1] Polytec, “Beamforming,” [Online]. Available: https://www.polytec.com/us/acoustics/technology/beamforming. [Accessed: Dec. 20, 2024].

[2] I. Berrios, “Introduction to Beamforming Part 1,” Medium, Dec. 12, 2018. [Online]. Available: https://medium.com/@itberrios6/introduction-to-beamforming-part-1-92767408f483. [Accessed: Dec. 20, 2024].

[3] S. Ravi et al., “Distributed and Efficient Beamforming for Next-Generation Wireless Networks,” arXiv, Nov. 4, 2022. [Online]. Available: https://arxiv.org/abs/2211.02165. [Accessed: Dec. 20, 2024].

[4] L. Wang et al., “Enhanced Adaptive Beamforming Techniques for 6G Communication,” arXiv, Mar. 31, 2024. [Online]. Available: https://arxiv.org/abs/2403.17376. [Accessed: Dec. 20, 2024].

[5] MathWorks, “MATLAB,” [Online]. Available: https://www.mathworks.com/products/matlab.html. [Accessed: Dec. 20, 2024].

[6] N. Alarcon et al., ”Senior Capstone Project Proposal LenScribe: Beamforming Subtitle Glasses,” [Online] [Accessed: Dec. 20, 2024]

[7] Wikipedia, “Beamforming,” [Online]. Available: https://en.wikipedia.org/wiki/Beamforming. [Accessed: Dec. 20, 2024].

[8] STMicroelectronics, “IMP23ABSU: Analog MEMS Microphone,” [Online]. Available: https://www.st.com/en/mems-and-sensors/imp23absu.html. [Accessed: Dec. 20, 2024].