BRADLEY University

Speech Intelligibility Enhancement using Microphone Array via Intra-Vehicular Beamforming

Devin McDonald, Joe Mesnard
Advisors: Drs. In Soo Ahn and Yufeng Lu

April 28th , 2018

Table of Contents

Introduction	Page 4
System Block Diagram, Specification, and Subsystems	Page 5
Simulink Models	Page 7
Calibration	Page 9
Engineering Efforts	Page 12
Parts List	Page 12
Deliverables	Page 13
Preliminary Study	Page 13
Simulation Study	Page 17
Experimental Study	Page 22
Conclusion	Page 27
References	Page 29

List of Figures

Figure 1 - System Block Diagram	Page 5
Figure 2 - Real Time Simulink Model	Page 7
Figure 3 - Simulink Simulation Input Block	Page 8
Figure 4 - Simulation Calibration Model	Page 9
Figure 5 - AGC Model	Page 10
Figure 6 - AGC Coefficient Plot	Page 10
Figure 7 - Zero Crossing Detection for a 150 Hz Calibration Signal	Page 11
Figure 8 - Gannt Chart of Engineering Efforts	Page 12
Figure 9 - Preliminary Testing Setup	Page 13
Figure 10 - Linear Translation Test	Page 14
Figure 11 - Spectral Sweep Test	Page 14
Figure 12 - Beamforming a Simple Sine Wave in MATLAB	Page 15
Figure 13 - Three Microphone Instantaneous Power	Page 16
Figure 14 - Three Microphones Summed and Normalized	Page 16
Figure 15a - Simulated 400 Hz (before beamforming)	Page 18
Figure 15b - Simulated 400 Hz (after beamforming)	Page 18
Figure 15c - Simulated 400 Hz power increase of 20.57 dB	Page 18
Figure 16a - Simulated 1 kHz (before beamforming)	Page 19
Figure 16b - Simulated 1 kHz (after beamforming)	Page 19
Figure 16c - Simulated 1 kHz power increase of 14.83 dB	Page 19
Figure 17a - Simulated 3 kHz (before beamforming)	Page 20
Figure 17b - Simulated 3 kHz (after beamforming)	Page 20
Figure 17c - Simulated 3 kHz power increase of 5.002 dB	Page 20
Figure 18a - Simulated 6 kHz (before beamforming)	Page 21
Figure 18b - Simulated 6 kHz (after beamforming)	Page 21
Figure 18c - Simulated 6 kHz power increase of 7.473 dB	Page 21
Figure 19 - Testing Setup with Computer, Interface, Array, and Speaker	Page 22
Figure 20a - Experimental 400 Hz (before beamforming)	Page 23
Figure 20b - Experimental 400 Hz (after beamforming)	Page 23
Figure 20c - Experimental 400 Hz power increase of 11.11 dB	Page 23
Figure 21a - Experimental 1 kHz (before beamforming)	Page 24
Figure 21b - Experimental 1 kHz (after beamforming)	Page 24
Figure 21c - Experimental 1 kHz power increase of 10.58 dB	Page 24
Figure 22a - Experimental 3 kHz (before beamforming)	Page 25
Figure 22b - Experimental 3 kHz (after beamforming)	Page 25
Figure 22c - Experimental 3 kHz power increase of 3.654 dB	Page 25
Figure 23a - Experimental 6 kHz (before beamforming)	Page 26
Figure 23b - Experimental 6 kHz (after beamforming)	Page 26
Figure 23c - Experimental 6 kHz power increase of 7.093 dB	Page 26

Introduction

Problem Background

According to the National Safety Council [1], there are approximately 1.6 million crashes each year due to distracted driving involving mobile phones. Drivers often hold their phone while making or taking a call, which causes their eyes to leave the road. In an attempt to discourage the handheld use of mobile phones while driving, hands-free Bluetooth calling connectivity has become the auto-industry standard. This hasn't entirely solved the problem, however.

The level of near-end speech intelligibility being sent is reduced due to multiple sources of noise. Some noises occur outside the car cabin such as engine noise, wind noise, conductive vibration, and road noise such as tires against pavement. Others occur inside the cabin including talking passengers, air conditioning, and music. Regardless of their source location, all of these noise sources and others combine to reduce the intelligibility of phone conversations. This causes frustration and often affects the driver's concentration. They simply pick up the cellphone and use it as normal.

In audio signal processing applications, beamforming can be applied to selectively emphasize audio signals based on the direction-of-arrival (DOA) in the relationship to an array of microphones. Acoustic beamforming is a process by which multiple signals from a microphone array are filtered and combined in order to increase the amplitude of a target source's signal at a static DOA without increasing the amplitude of signals with differing DOAs.

Problem Statement

This project aims to enhance speech intelligibility using microphone array via intra-vehicular beamforming, where the beamforming technique is used to combat near-end noise, and a uniform linear array (ULA) of microphones is used for data acquisition. The processed signal is then sent to a far-end user over a hands-free Bluetooth system with increased near-end speech intelligibility.

The proposed solution for this project is beamforming. Specifically, we will use a technique called Delay and Sum beamforming. This type of beamforming takes advantage of the fact that a uniform linear array of microphones will detect a signal at different times, due to the space between them. Any signal that is centered among the array will have the strongest correlation. If the microphones are summed and then normalized by the number of microphones in the array, any signal coming from directly in front of the array will stay at its original volume. Any signal coming from an angle will be attenuated.

The algorithm, in its natural state, has a beam steered at 0 degrees. Steering the beam in other directions is accomplished by tricking the system. If the microphone data is delayed by a certain number of samples, the strongest correlation will no longer be directly in front of the array, but off to a specific angle.

Delay and Sum Beamforming is a simple, yet very effective way to create a distinction between data that is coming from a desired angle and data that is coming from other sources. This project will use this information to enhance the speech content of a certain person in the vehicle, and attenuate all others.

The algorithm, in its natural state, has a beam steered at 0 degrees. Steering the beam in other directions is accomplished by tricking the system. If the microphone data is delayed by a certain length of time, the strongest correlation will no longer be directly in front of the array, but off to a specific angle.

Delay and Sum Beamforming is a simple, yet very effective way to create a distinction between data that is coming from a desired angle and data that is coming from other sources. This project will use this information to enhance the speech content of a certain person in the vehicle, without increasing the presence of noise.

Scope

The aim of the project is to complete a system that increases near-end intelligibility before being sent via Bluetooth to a cell phone and then to the far-end receiver. This system is to be integrated with existing technology already implemented in vehicles. It is assumed that these vehicles already contain Bluetooth technology. Thus, the scope of this project does not include the Bluetooth processing.

System Block Diagram, Specifications, and Subsystems

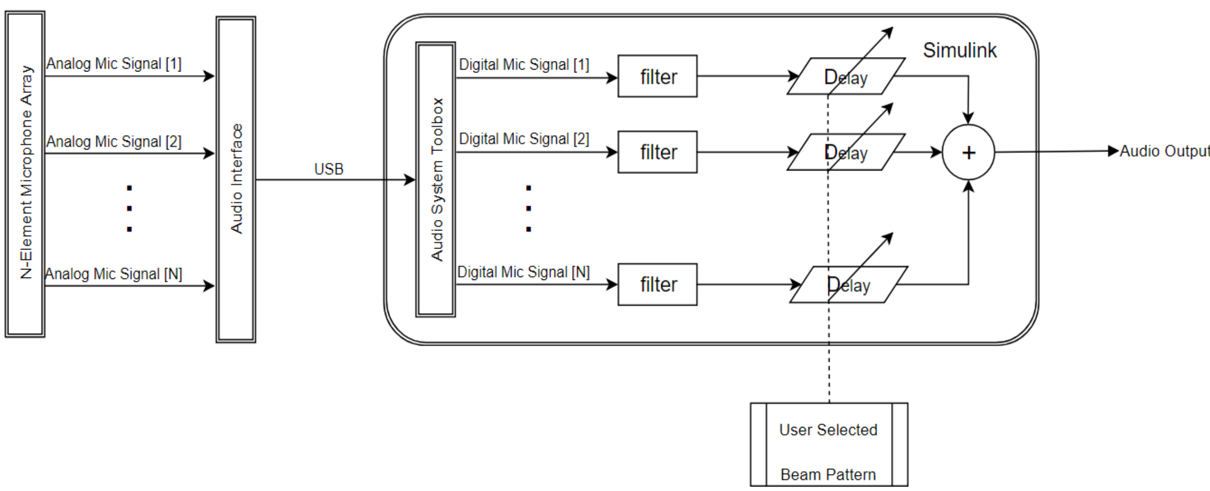


Figure 1 – System Block Diagram

System Description

N-Element Microphone Array

ULA of microphones will output signal via XLR.

Filters

A-Weighting Filters implemented in MATLAB/Simulink are designed to focus on the prominent frequencies of human speech (~500Hz to ~4kHz).

Delay

Delays will work as a part of the “Delay” and Sum beamforming algorithm

User input

The end user will be able to switch beam patterns to control where the beam is steered and who in the vehicle can be heard.

Audio Interface

The Focusrite Scarlett 18i20 will send digitized audio data from the microphones to the computer via USB.

Audio System Toolbox

The audio system toolbox in Simulink will be used to communicate with the audio interface and get stream data into Simulink.

Nonfunctional Requirements

- The system will increase the intelligibility of near-end speech sent to the far-end user.
- The system requires little user manipulation.
- The system can be integrated within a vehicle.

Functional Requirements

- The system is tested and demonstrated in intra-vehicular or similar environment.
- The system includes a ULA microphone array.
- Each microphone is routed to a system (such as MATLAB) for data acquisition.
- Beamforming is implemented in real-time.

Simulink Models

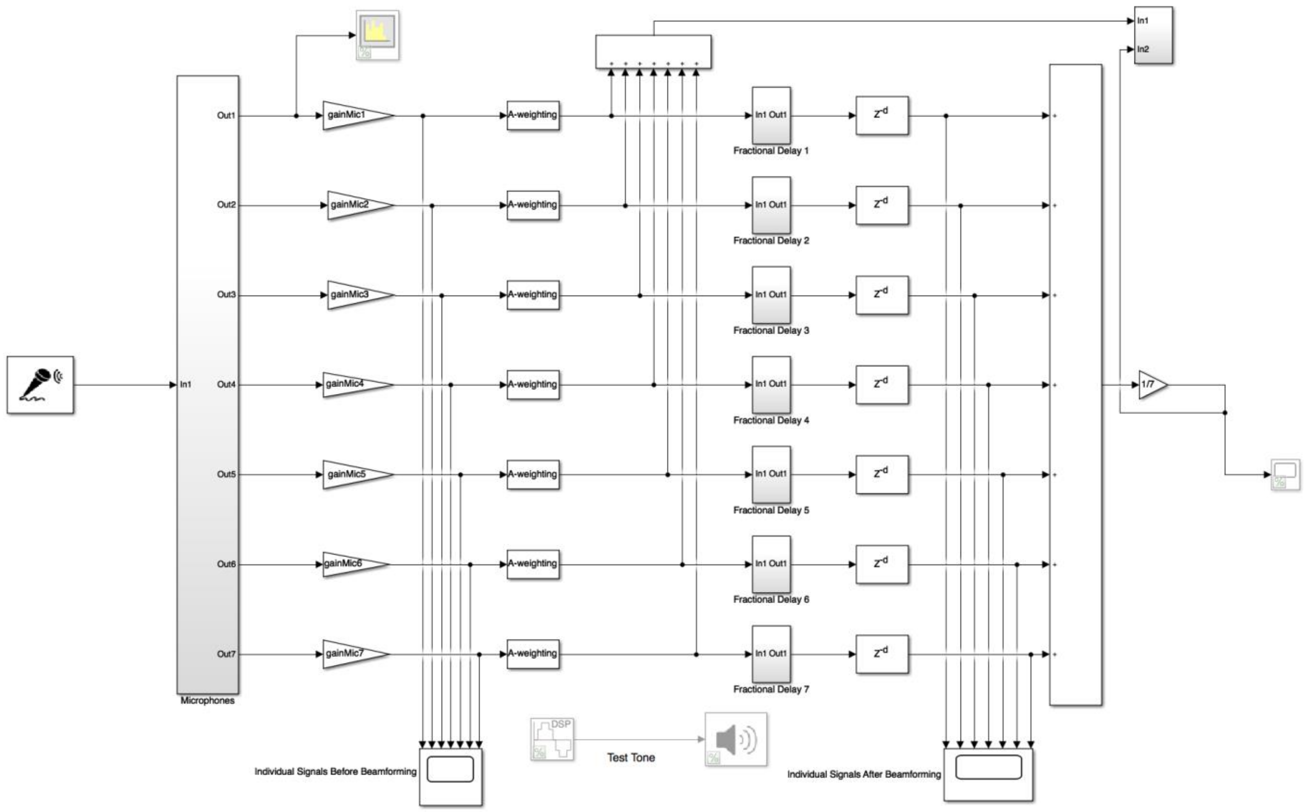


Figure 2 - Real Time Simulink Model

Figure 2 shows the Simulink model that implements delay and sum beamforming. For real-time operation, the input is as shown. For simulation, the microphone signals are replaced by sine waves and delay blocks that simulate the physical delays that the signals experience when travelling through air. This can be seen in Figure 3.

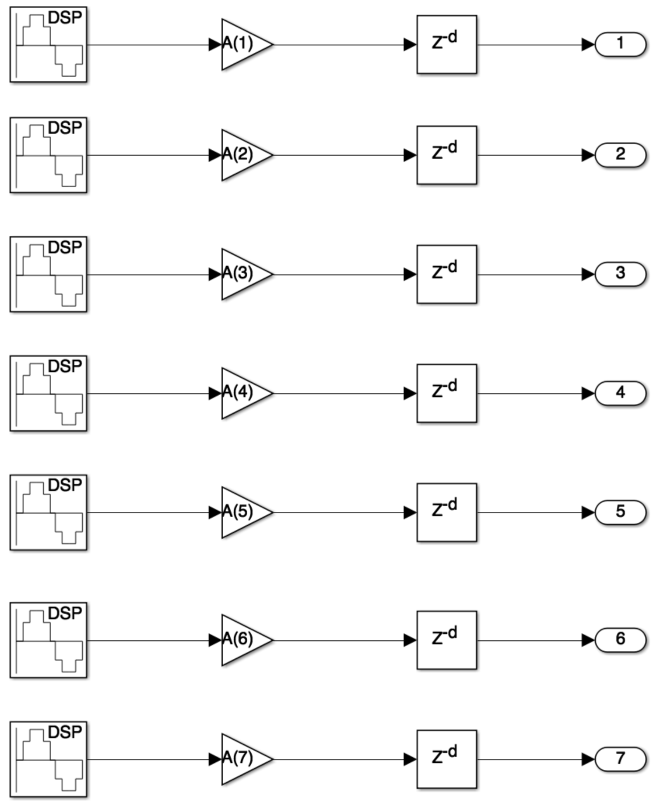


Figure 3 - Simulink Simulation Input Block

Figure 3 shows the input block for simulated results. Gains and delays are applied to simulate the different input responses, and sample delays are used to simulate the time it takes for the sound waves to travel through air.

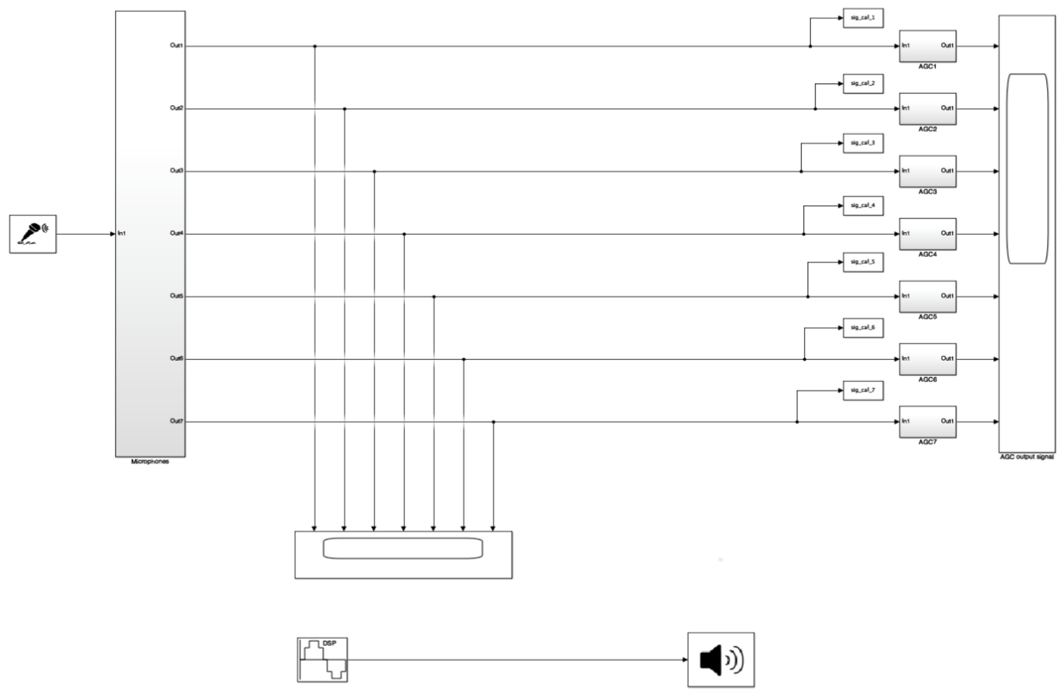


Figure 4 - Simulink Calibration Model

Figure 4 shows the calibration Simulink model that calibrates the delay times and the gain coefficient values.

Calibration

Calibration of the system is performed using the model shown in figure 4. The calibration consists of two processes: AGC calibration, and delay calculation.

Automatic Gain Controller (AGC) Calibration

An AGC is used to match the gains of each microphone. Because the microphones are not matched, and the gains for each microphone are set by hand at the interface, there exists a large variation in the response of each microphone. In order to create a more even response form the microphones, an AGC is used. A 1 kHz sine wave is played for 10 seconds. This allows time for the AGC to converge. The final values that the AGCs converge to are used as coefficients for the microphone signals. The AGC model is shown in Figure 5, and a sample plot of the coefficients is also shown in Figure 6.

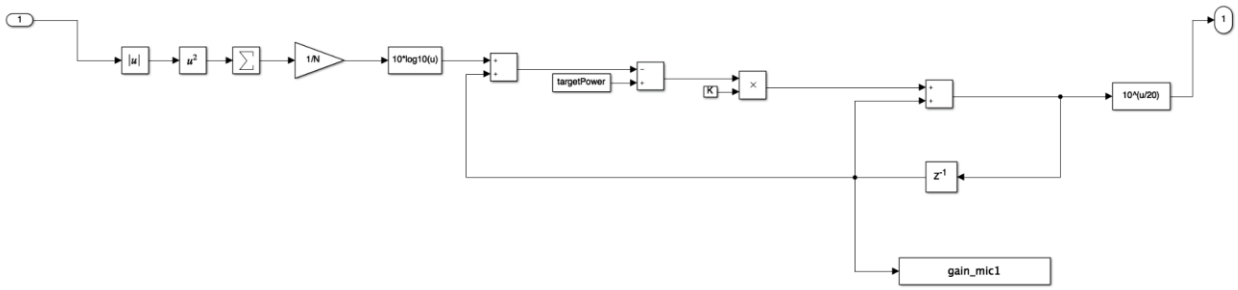


Figure 5 - AGC Model

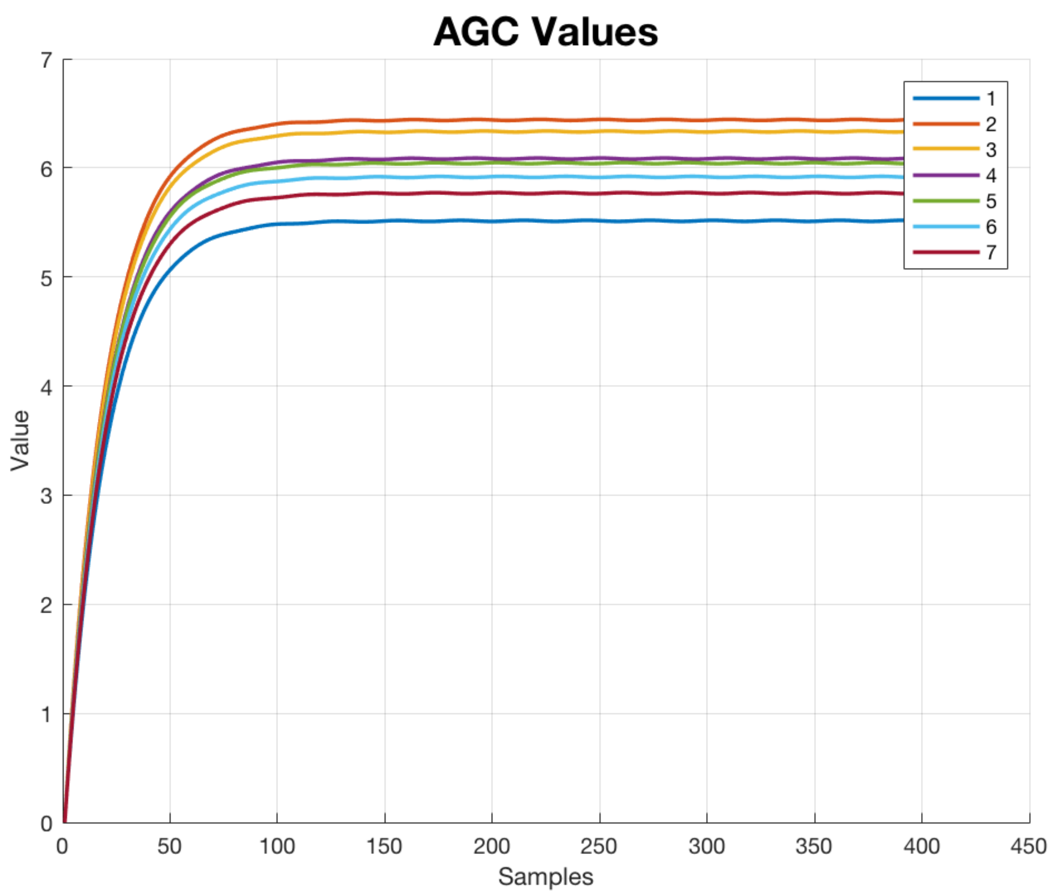


Figure 6 - AGC Coefficient Plot

Delay Calculation

A low frequency sine wave is played from a speaker to calculate the necessary delay times. A speaker is placed where the source of interest will be, and emits a tone. The time between zero crossings of 1 period of audio is used to generate the necessary delay times. Linear approximation is used when a zero occurs between samples. This method is much more precise than calculating delay times using measured distance. After calculating the delay times, the necessary sample and fractional delays are derived.

Low frequencies are necessary to prevent the overlap of signal periods. The delay-time difference determines the maximum frequency that can be used for calibration. The maximum frequency that can be used within vehicles is between 100 Hz and 300 Hz, depending on the vehicle's dimensions.

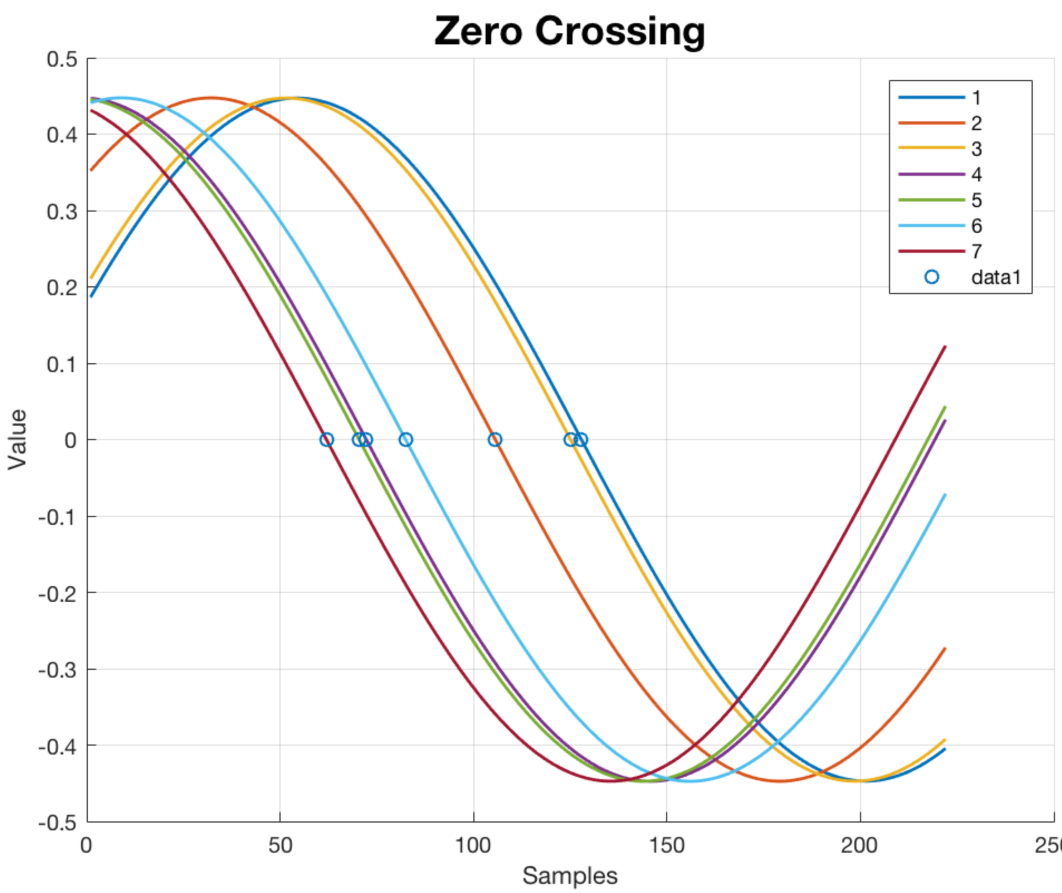


Figure 7 - Zero Crossing Detection for a 150 Hz Calibration Signal

Engineering Efforts

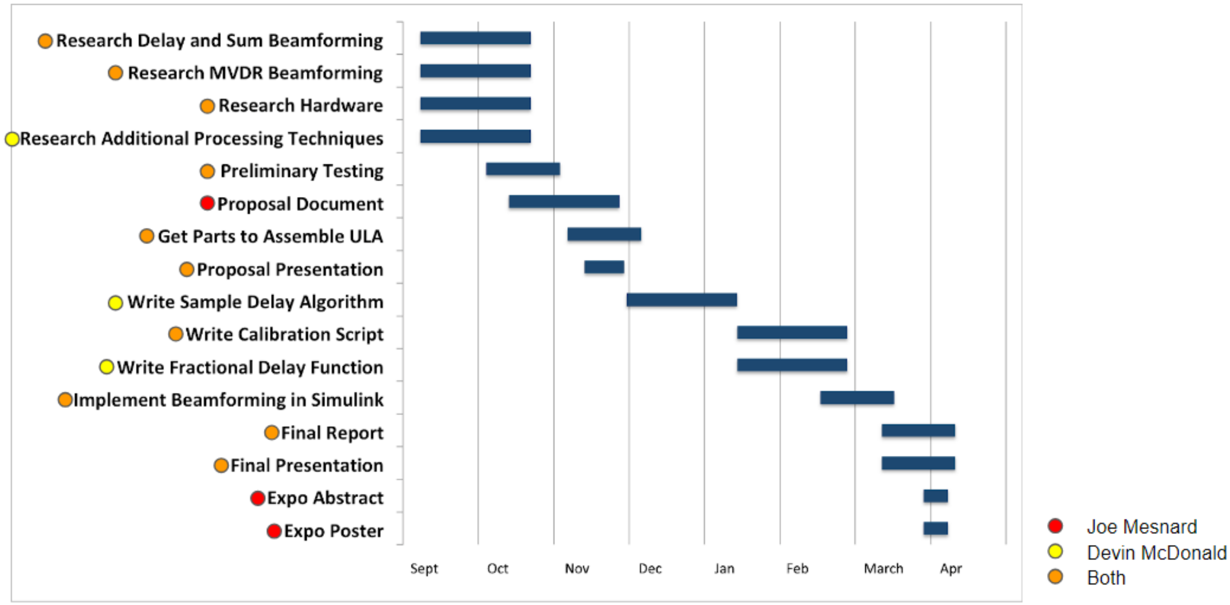


Figure 8 – Gannt Chart of Engineering Efforts

Parts List

Quantity	Description	Price	Ext. Price
1	XLR Patch Cables	$31.75	$31.75
3	Behringer UltraVoice XM1800S Microphones	$39.99	$119.97
5	Pro Black Adjustable Dual Plastic 2pcs Drum Microphone Clip	$7.44	$37.20
1	Scarlett 18i20 Audio Interface	$499.99	$499.99

Deliverables

Date Due	Description
November 9th	Draft proposal report
November 16th	Draft proposal presentation
November 30th	Sign up for proposal presentation
November 30th	Proposal final version
December 7th	Project website

Preliminary Study

The first set up completed is a simple three element microphone array consisting of three Electro-Voice dynamic microphones. There were 9 separate tests to see what the microphone data from three equally spaced microphones looked like.

Test Setup

Three microphones were placed 0.2 meters apart from each other as shown in Figure 9. These microphones were run through a Scarlett Audio Interface. Each microphone was recorded using Logic Pro, a recording software. The software recorded each test at 44.1kHz.

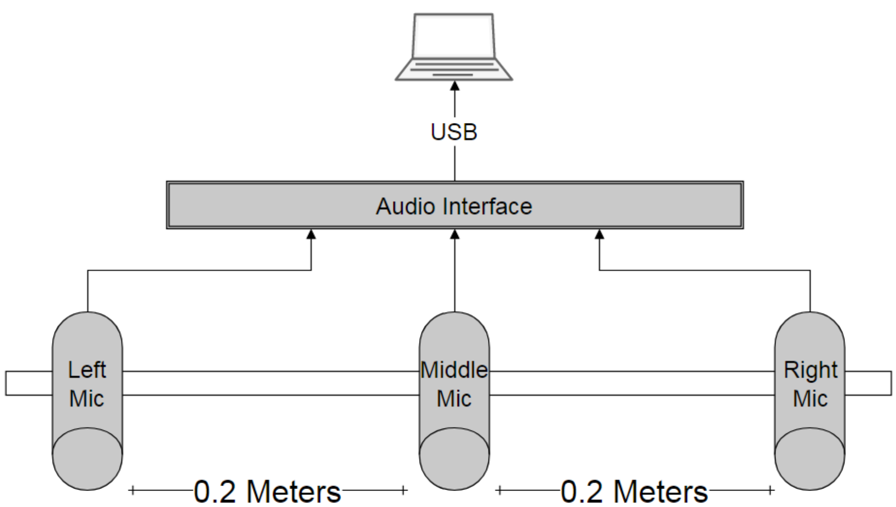


Figure 9 – Preliminary Testing Setup

Test Scenarios

There were two test scenarios that were used. The first was linear translation shown in Figure 10, and the second was spectral sweeping shown in Figure 11.

The linear translation test involved moving a signal source playing a monotonic frequency in front of the microphone array linearly from left to right. The purpose of this test is to be able to see how a delay and sum beamforming algorithm performs with a constant frequency at different angles. For example, it is easy to see which direction the beam is being steered by looking at which angle has the largest amplitude. The angle is determined by the difference between the middle microphone orientation and the source.

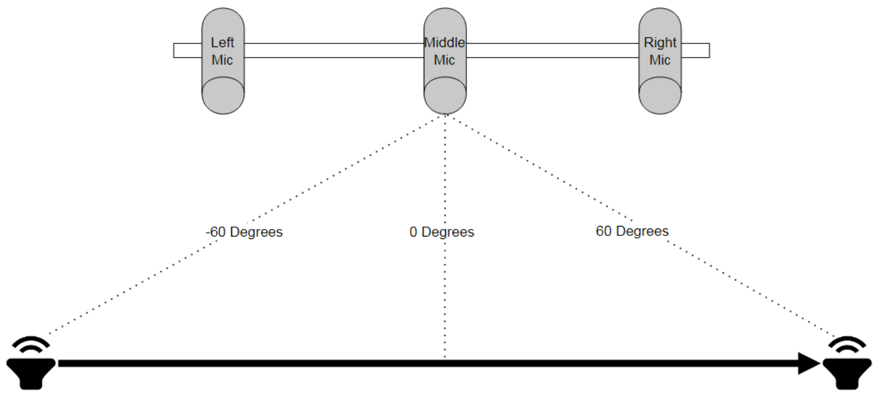


Figure 10 – Linear Translation Test

The spectral sweep test was to see how beamforming performed to different frequencies at discrete angles along the linear translation line. A frequency sweep from 1Hz to 10kHz was done at five discrete points. -60 degrees, -45 degrees, 0 degrees, 45 degrees, and 60 degrees.

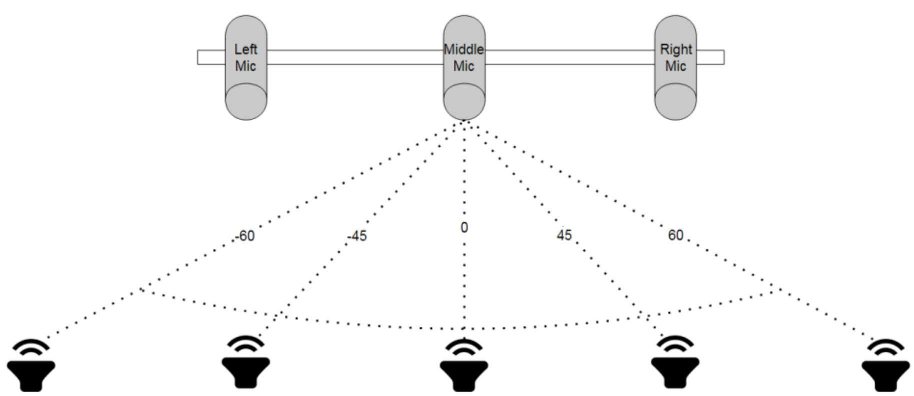


Figure 11 – Spectral Sweep Test

Preliminary Theoretical Results

After the data was recorded, a MATLAB script was written to better understand how the beamforming algorithm performs. Figure 12 shows the beamforming algorithm acting on a sine wave at 1kHz from angles -180 degrees to 180 degrees. The simulation assumes that each mic is placed 0.2 meters apart, a frequency of 1kHz is playing, and the speed of sound in air is 343 m/s.

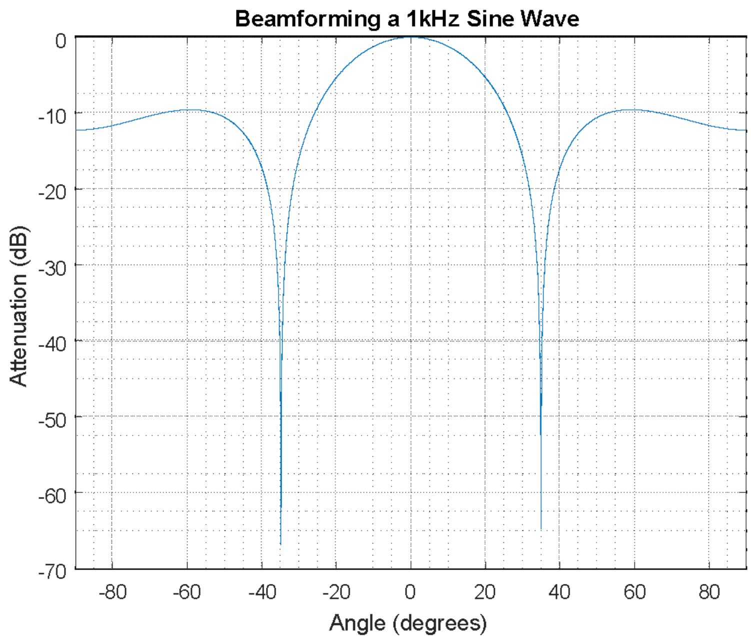


Figure 12 – Beamforming a Simple Sine Wave in MATLAB

Preliminary Experimental Results

The beamforming algorithm was applied in MATLAB to the test data obtained by the three-microphone array. The theoretical data shows that at 1kHz, the signal should have 10 dB of attenuation. Figure 13 shows the power from each microphone over a frame size of 2048 samples. It is apparent that the right microphone has less power for the beginning of the translation since it is farther away from the source at first. This same theory applies to the left microphone towards the end of the translation.

Note that the x axis of these plots is time. However, since the translation is done by moving left to right over time, each time interval represents a different angle.

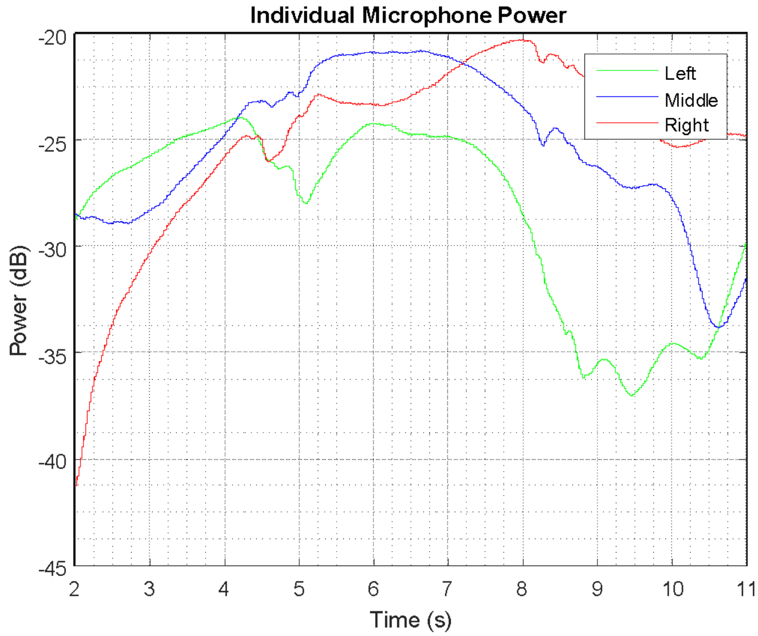


Figure 13 – Three Microphone Instantaneous Power

Figure 14 shows the beamforming results for the 1kHz linear sweep. The microphones are just summed and then normalized. In this case, normalization is dividing by 3 since that is the number of microphones in the array. This experimental data is mimics the theoretical data very closely. The middle of the test (from 5 to 8 seconds) is when the signal source is at 0 degrees. The estimated power around this time is about -25 dB. During the beginning and end of the test, the signal source is off center from 0 degrees. Around those times, the estimated power resides at about -35 dB. This is a 10 dB reduction near the same angles as Figure 12 shows.

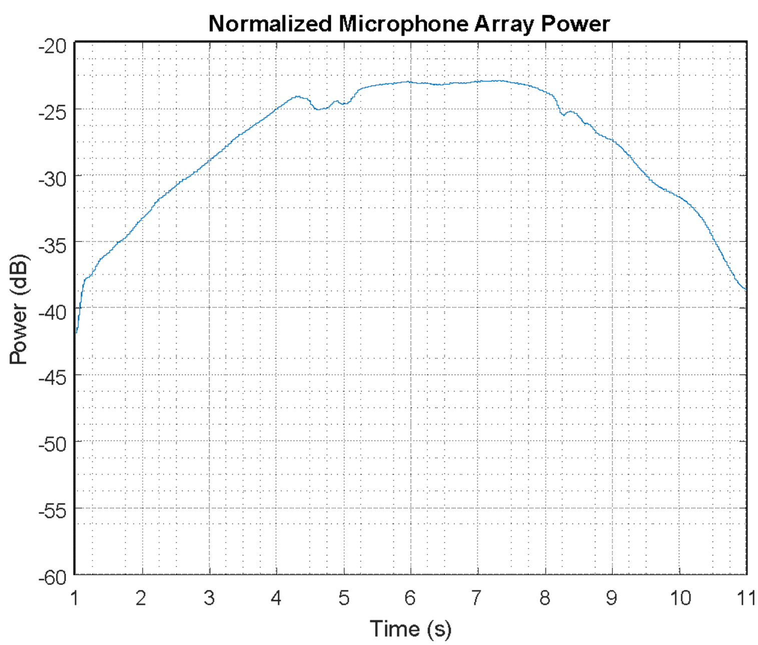


Figure 14 – Three Microphones Summed and Normalized

Preliminary Conclusion

The results from preliminary testing provided an insight into getting the setup and algorithm sufficient. It was determined that 10 dB was not quite enough attenuation to be notice by the human ear. To solve this problem, a 7-microphone array was implemented. By increasing the number of microphones, the side lobes are subject to further attenuation. The microphone spacing was chosen to be 0.17 meters to compensate for the increase microphone number.

Simulation Study

Once the calibration and real-time models are created, they were tested in two ways. The first way was to send simulated sine waves using the DSP sine wave block in simulink. We then manipulated the phase of these signals to simulate the sound waves hitting the microphones at different times. We tested this simulation at multiple frequencies, including 400 Hz, 1 kHz, 3 kHz, and 6 kHz. These tests provides the theoretical maximum improvement our system may create.

Simulation results from input signals with different frequencies are shown in the following pages. There are two main results shown for each testing scenario. The first plot shows all of the raw data coming in from the DSP blocks. Since the blocks have sine waves of different phases, they are phase shifted to simulate sound waves arriving the microphones at different times. From the first plot which are the signals before beamforming, it clearly shows the signals from different microphone are completely off the synchronization (see Figure 15a, Figure 16a, Figure 17a and Figure 18a). It suggests how poor that would sound to a listener after the audio is reconstructed. The second plot shows the results after system calibration and beamforming are applied (see Figure 15b, Figure 16b, Figure 17b and Figure 18b). In all of these test cases, it can be seen that system calibration and beamforming have greatly improved the quality of overall signal in terms of correlation among signals from different microphones. The third figure is included to numerically shown the system performance (see Figure 15c, Figure 16c, Figure 17c and Figure 18c). The bottom line in these figures is the average power over a test of the unbeamfored signals. Since the signals are summed, when they are not in phase, there is less signal power. The top line in these figures is the average signal power after the signals are beamformed. When the signals become more closely aligned by beamforming, they have a larger average signal power over our tests.

These tests show the largest theoretical improvement in average signal power that can be achieved. These same tests were done with real experimental data, shown later in this document. Our simulated results suggest that there is as high as 20 dB of signal improvement at some frequencies made possible by beamforming.

Simulated 400 Hz Results

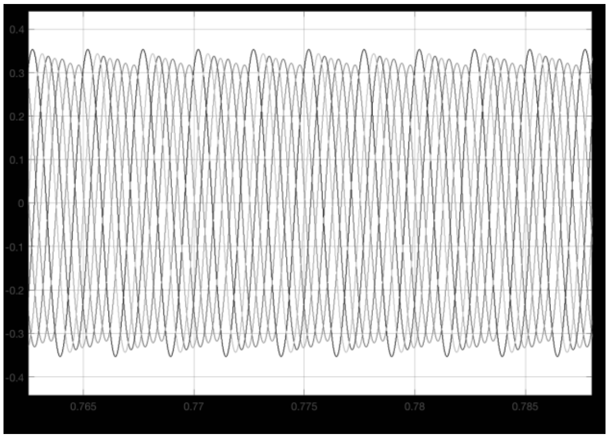


Figure 15a. Simulated 400 Hz (before beamforming)

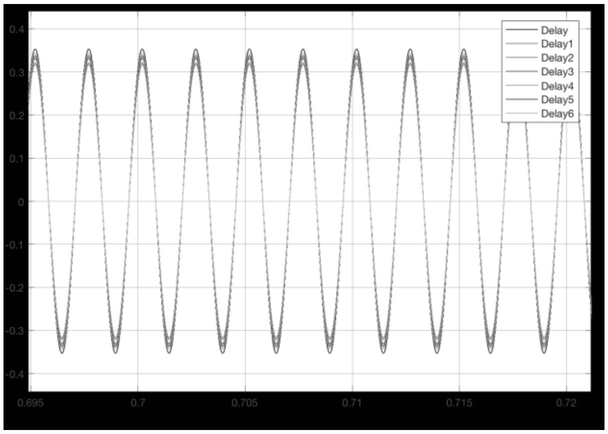


Figure 15b. Simulated 400 Hz (after beamforming)

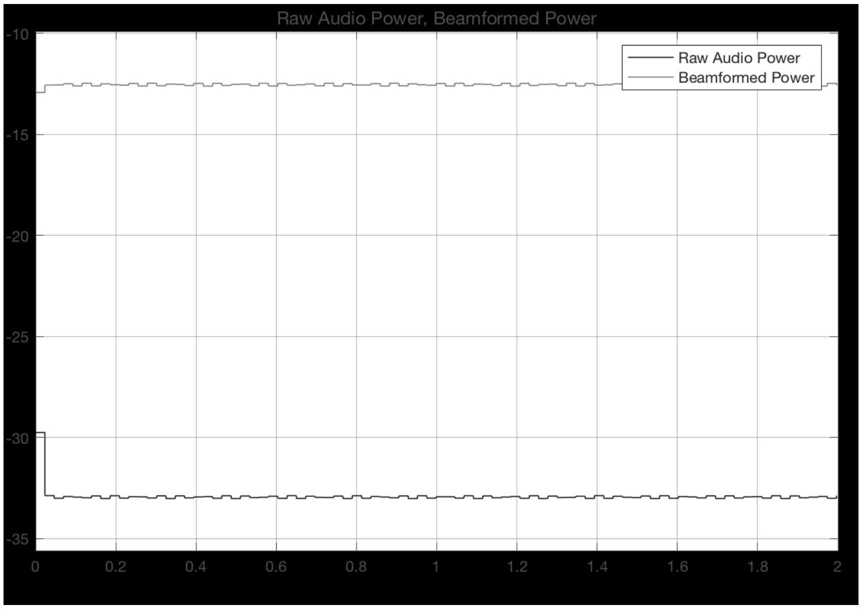


Figure 15c. Simulated 400 Hz power increase of 20.57 dB

Simulated 1 kHz Results

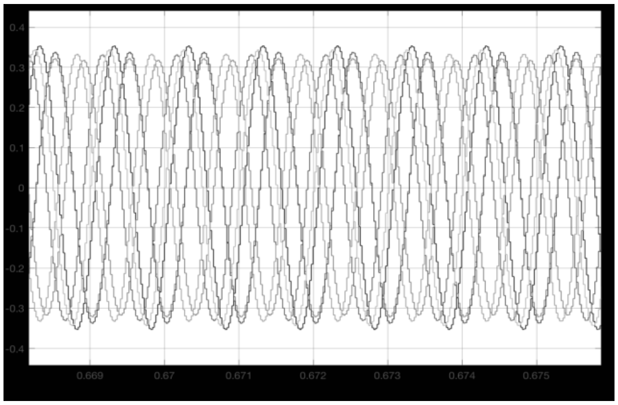


Figure 16a. Simulated 1 kHz (before beamforming)

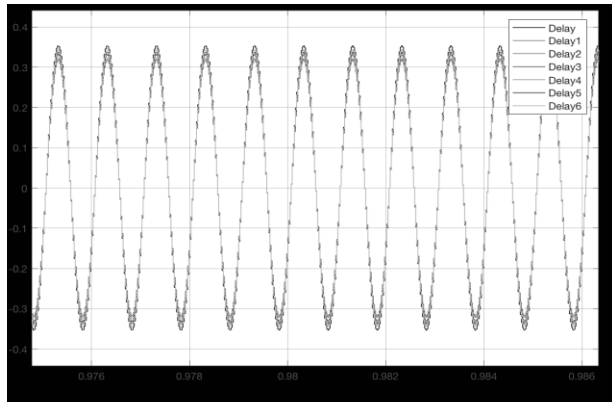


Figure 16b. Simulated 1 kHz (after beamforming)

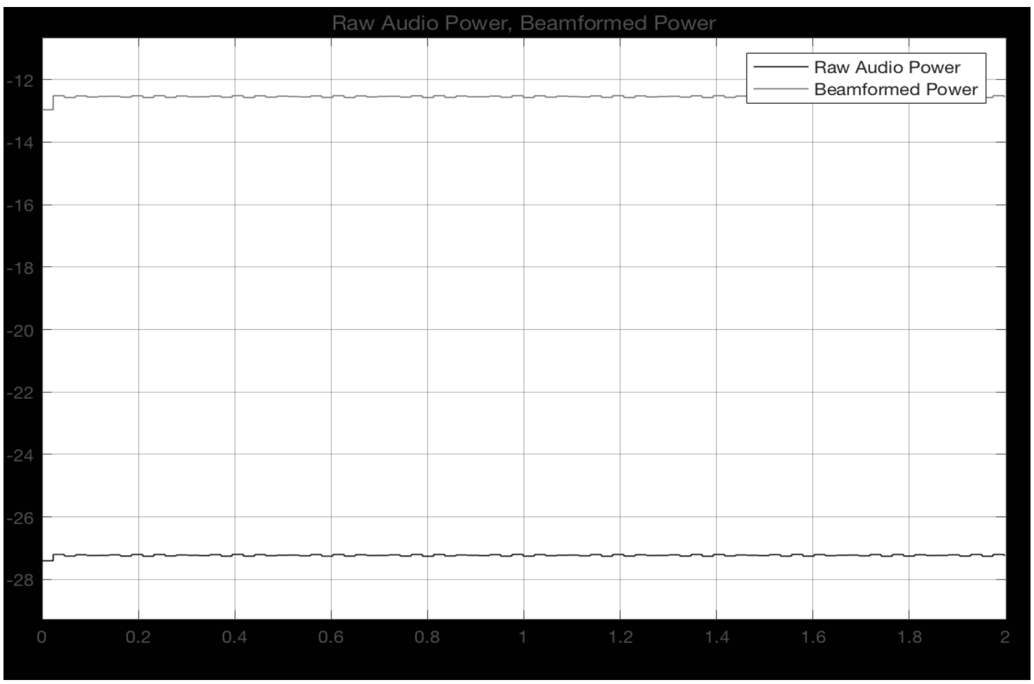


Figure 16c.Simulated 1 kHz power increase of 14.83 dB

Simulated 3 kHz Results

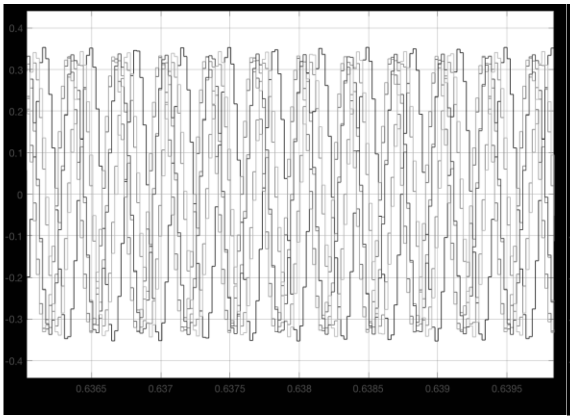


Figure 17a. Simulated 3 kHz (before beamforming)

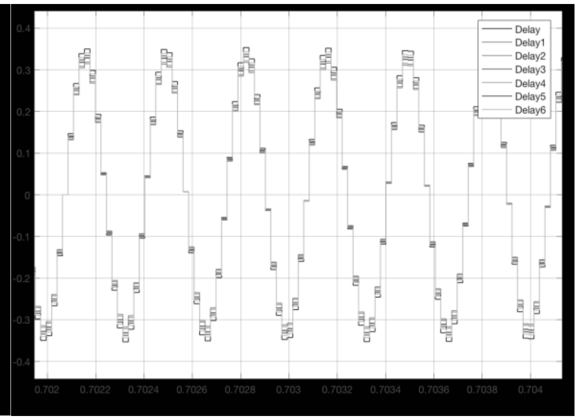


Figure 17b. Simulated 3 kHz (after beamforming)

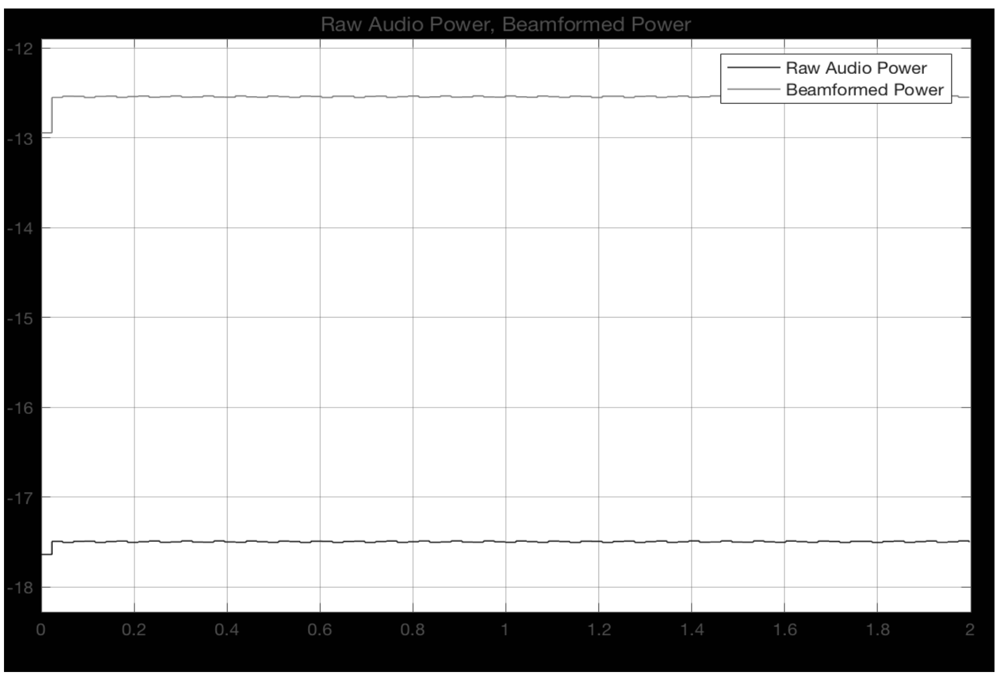


Figure 17c. Simulated 3 kHz power increase of 5.002 dB

Simulated 6 kHz Results

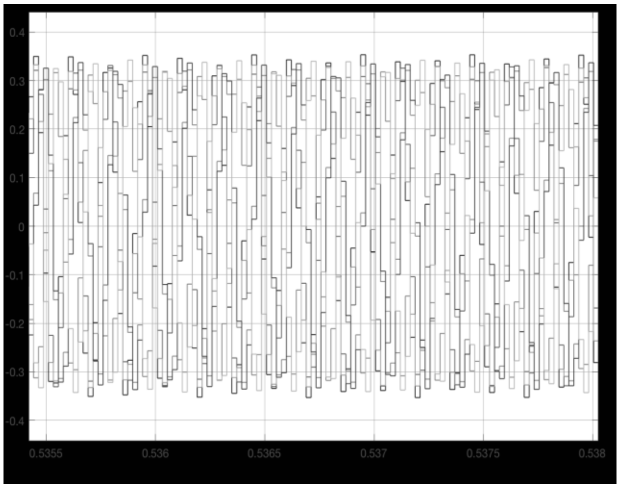


Figure 18a. Simulated 6 kHz (before beamforming)

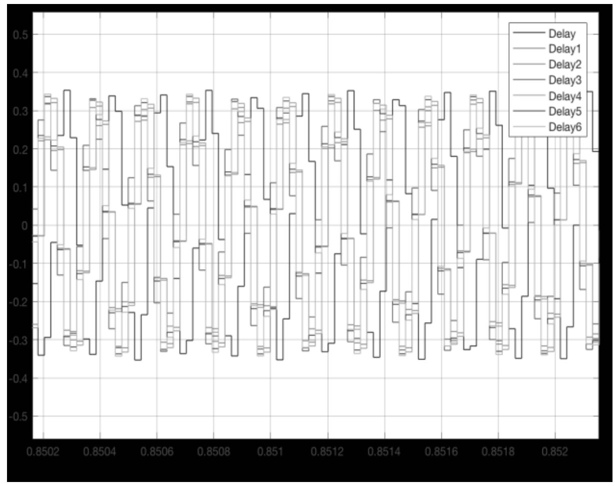


Figure 18b. Simulated 6 kHz (after beamforming)

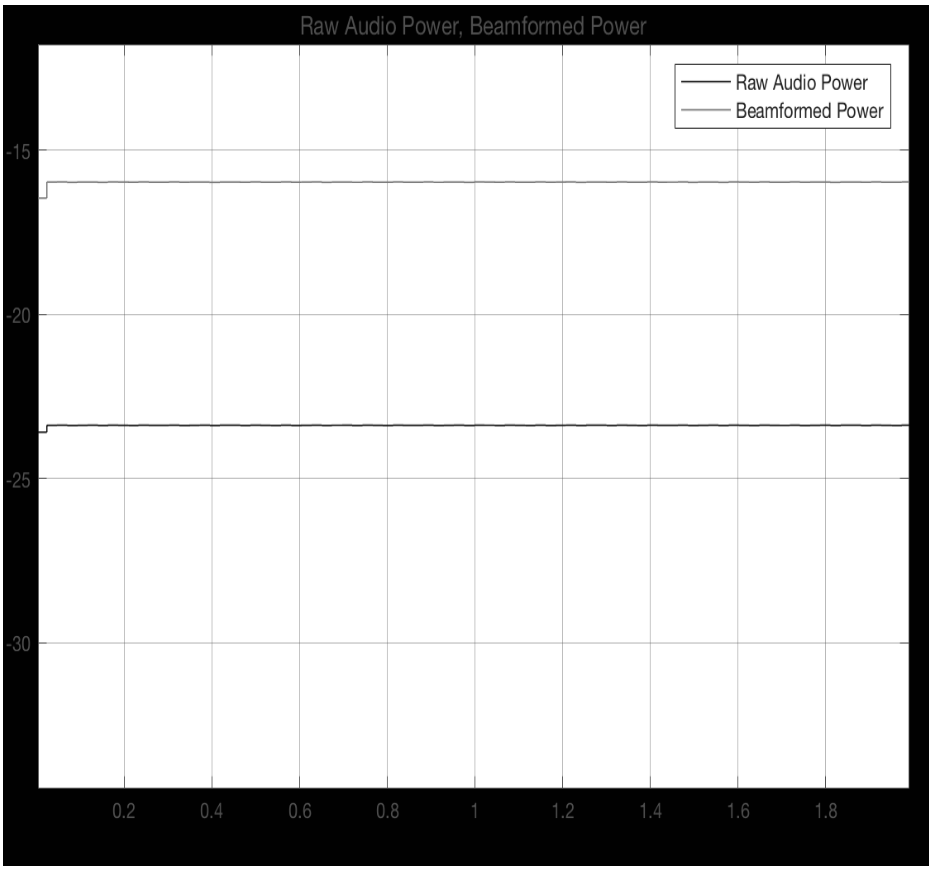


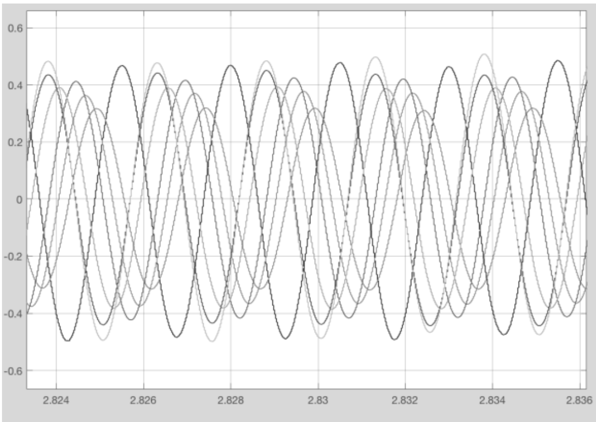
Figure 18c. Simulated 6 kHz power increase of 7.473 dB

Experimental Study

After we used the DSP blocks to simulate data, we took the array to the lab to get real test data. Our live testing data used the same simulink model, but instead of using the DSP sine wave generators, we used the interface and the Audio System Toolbox in Simulink. We then generated the same tones of 400 Hz, 1 kHz, 3 kHz, and 6 kHz and recorded the data. The test setup for the final lab results is shown in Figure 19. The speaker played a tone that was recorded by the microphone array. The microphone data went to simulink through the audio interface. In Simulink we compute the beamforming calculations and output the audio through the audio interface for listening. The audio is also record before and after beamforming to listen to at a later time.

Figure 19. Testing Setup with Computer, Interface, Array, and Speaker

Experimental 400 Hz Results



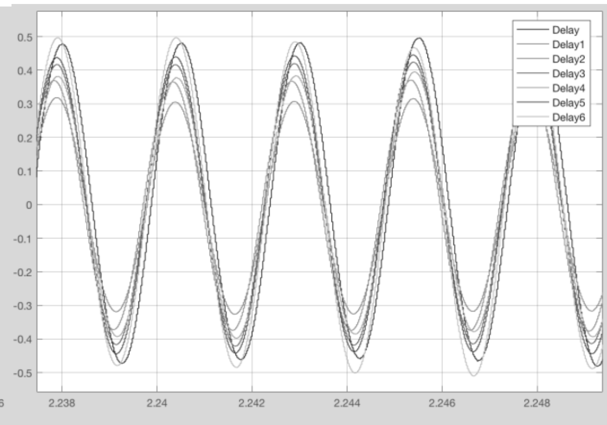


Figure 20a .Experimental 400 Hz Signal (before beamforming)

Figure 20b. Experimental 400 Hz Signal(after beamforming)

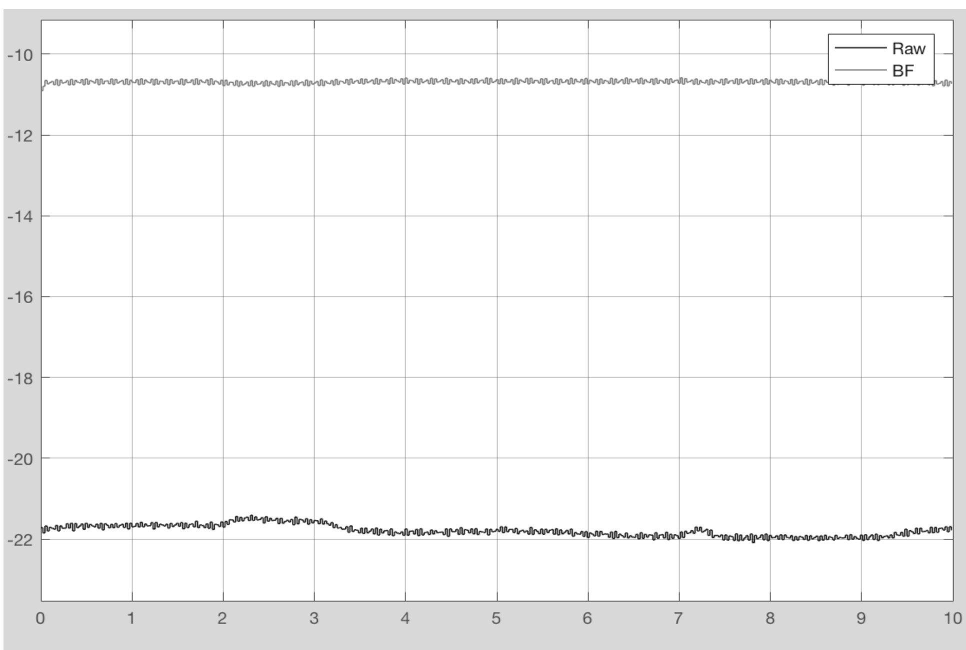


Figure 20c. Experimental 400 Hz power increase of 11.11 dB

Experimental 1 kHz Results

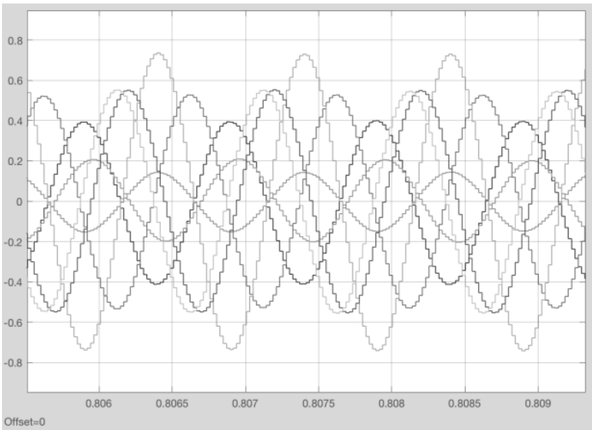


Figure 21a. Experimental 1 kHz (before beamforming)

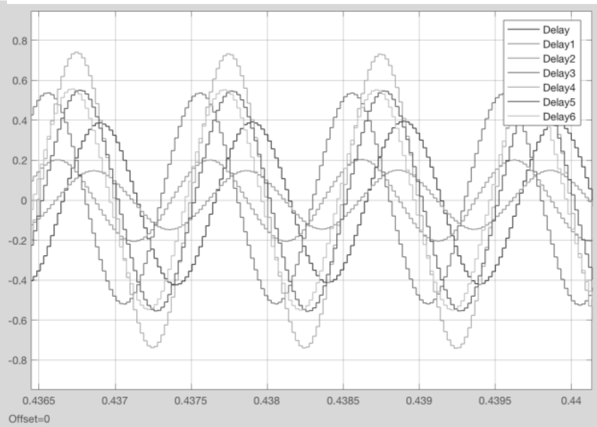


Figure 21b. Experimental 1 kHz (after beamforming)

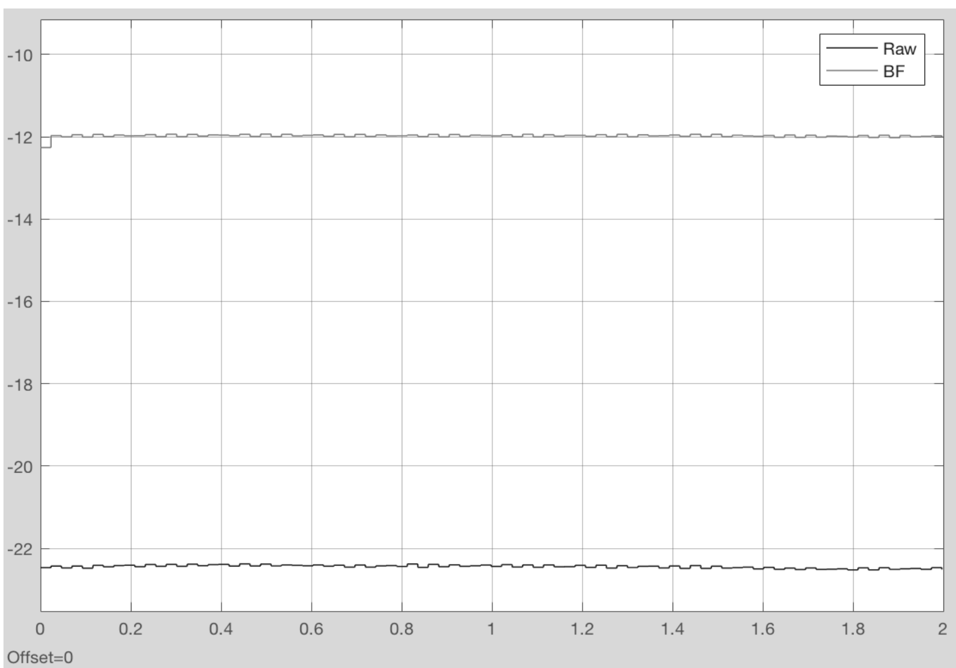
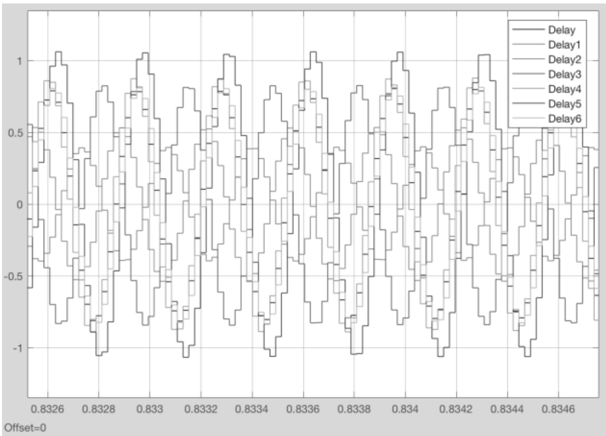


Figure 21c. Experimental 1 kHz power increase of 10.58 dB

Experimental 3 kHz Results



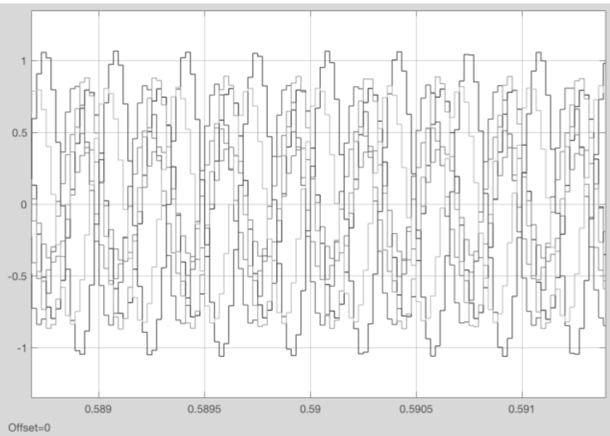


Figure 22a. Experimental 3 kHz (before beamforming)   Figure 22b. Experimental 3 kHz (after beamforming)

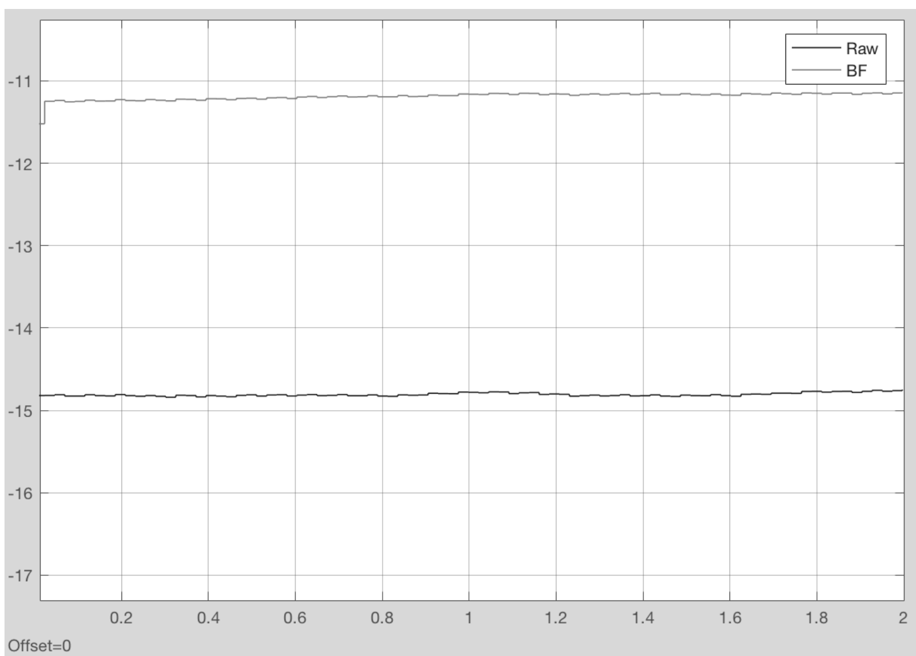


Figure 22c. Experimental 3 kHz power increase of 3.654 dB

Experimental 6 KHz Results

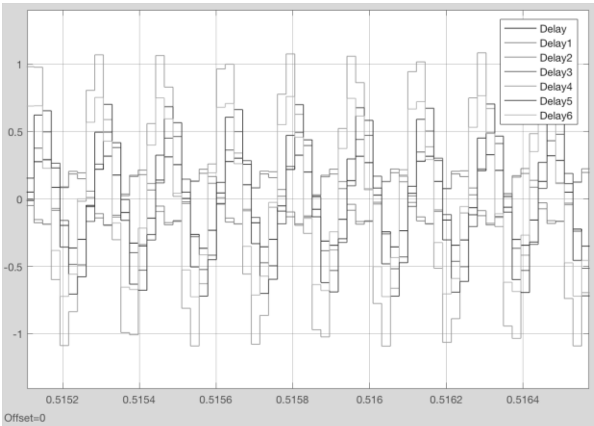


Figure 23a. Experimental 6 kHz (before beamforming)

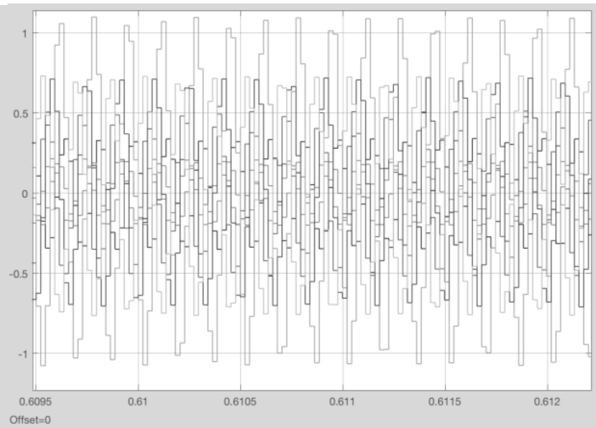


Figure 23b. Experimental 6 kHz (after beamforming)

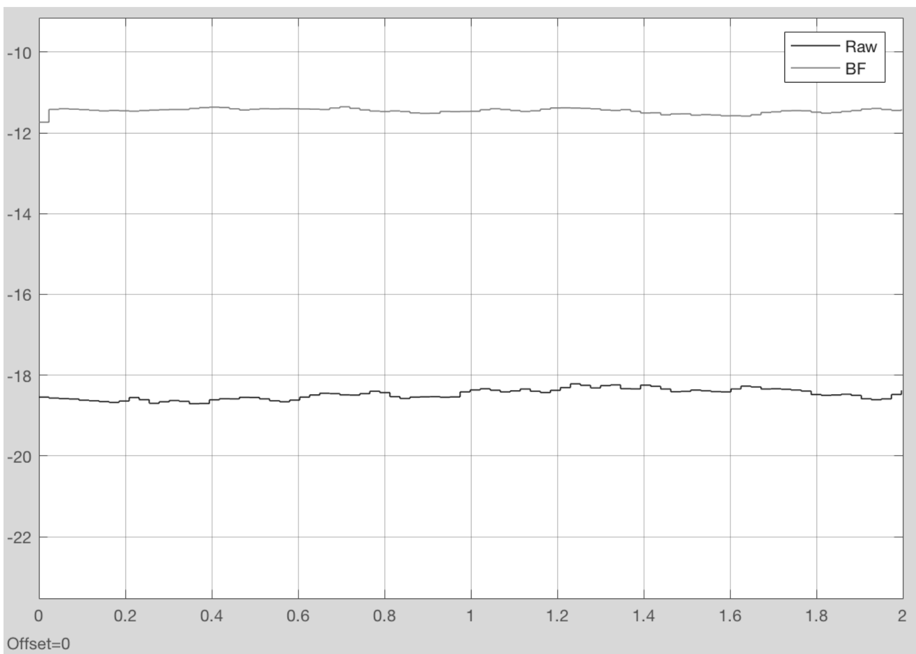


Figure 23c. Experimental 6 kHz power increase of 7.093 dB

The increase in signal power for the experimental results are comparable to the simulated results. The biggest improvements are in the low to mid frequencies, getting over 10 dB of signal improvement in both. This test gives us the confidence that our system performs well at the frequencies that contain the most voice content.

Table 1 - Beamforming Performance Comparison between Simulation Study and Experimental Study

Frequency of test signals	**Signal Power Improvement after Beamforming (Simulation Study)**	**Signal Power Improvement after Beamforming (Experimental Study)**
400 Hz	20.57 dB	11.11 dB
1000 Hz	14.83 dB	10.58 dB
3000 Hz	5.002 dB	3.654 dB
6000 Hz	7.473 dB	7.093 dB

Conclusions

Delay and Sum Beamforming is a simple algorithm that aims at saving lives behind the wheel. Using an array of microphones with this algorithm, far end speech intelligibility is enhanced. This is verified by the quantitative tests that measured the response of sine waves, and qualitatively by listening to recordings of the real-time data. This provides a smooth user experience for both the driver and the far end user, resulting in a completely hands-free conversation where the driver is is less tempted to pick up his or her mobile device during operation of the vehicle.

Calibration Algorithm

The most novel element of this project is the calibration algorithm. In 30 seconds or less, microphones in almost any configuration could be calibrated to successfully perform delay and sum beamforming. This frees up automotive manufacturers to implement the system very easily. Cheap microphones can be mounted to vehicles without modifying the design because they can be put almost anywhere as long as they can sufficiently pick up the driver’s voice.

Future Work

While our project was very successful in implementing delay and sum beamforming using microphones, there are many additional processing techniques that could be used to further enhance intelligibility. Adaptive filtering could reduce engine and road noise, and a voice activity detector (VAD) could get rid of background noise while the person of interest is not talking.

References

[1] "Texting and Driving Accident Statistics - Distracted Driving." *Edgarsnyder.com.* Accessed October 5, 2017. Available: https://www.edgarsnyder.com/car-accident/cause-of-accident/cell-phone/cell-phone-statistics.html

[2] "Phased Array System Toolbox - mvdrweights." (R2017b). *MathWorks.com.* Accessed July 14, 2017.
Available: https://www.mathworks.com/help/phased/ref/mvdrweights.html

[3] "(Ultra) Cheap Microphone Array." *Maxime Ayotte*. Accessed November 28, 2017. Available: http://maximeayotte.wixsite.com/mypage/single-post/2015/06/25/Ultra-Cheap-microphone-array

[4] "Microphone Array Beamforming." *InvenSense.* Accessed November 28, 2017. Available: https://www.invensense.com/wp-content/uploads/2015/02/Microphone-Array-Beamforming.pdf

[5] "Delay Sum Beamforming." *The Lab Book Pages.* Accessed November 28, 2017. Available: http://www.labbookpages.co.uk/audio/beamforming/delaySum.html