Speech Intelligibility Enhancement using Microphone Array via Intra-Vehicular Beamforming

Devin McDonald, Joe Mesnard
Advisors: Dr. In Soo Ahn & Dr. Yufeng Lu

November 9th, 2017

Table of Contents

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 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.

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.
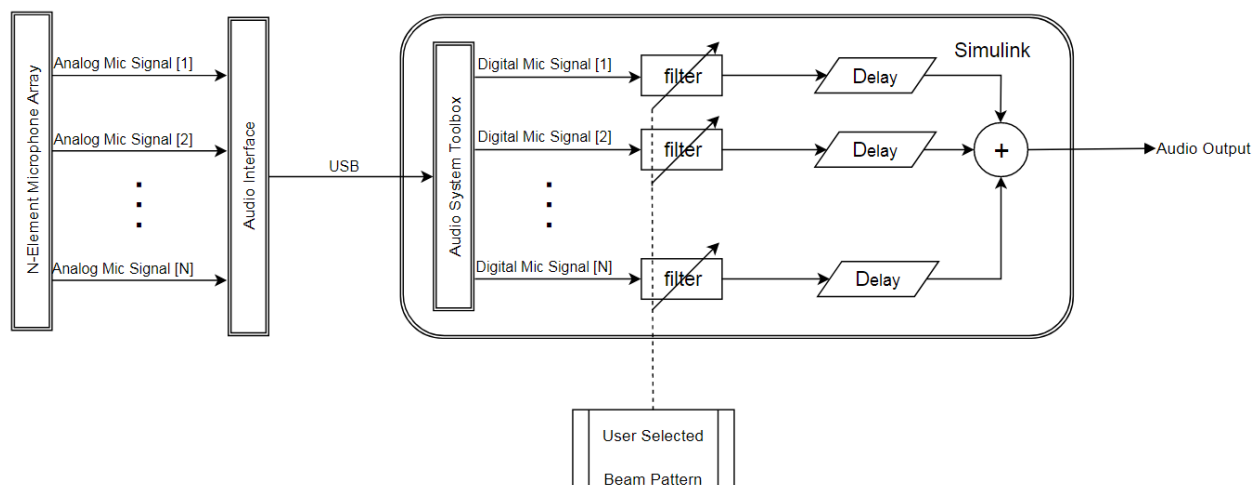
System Block Diagram, Specifications, and Subsystems

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.
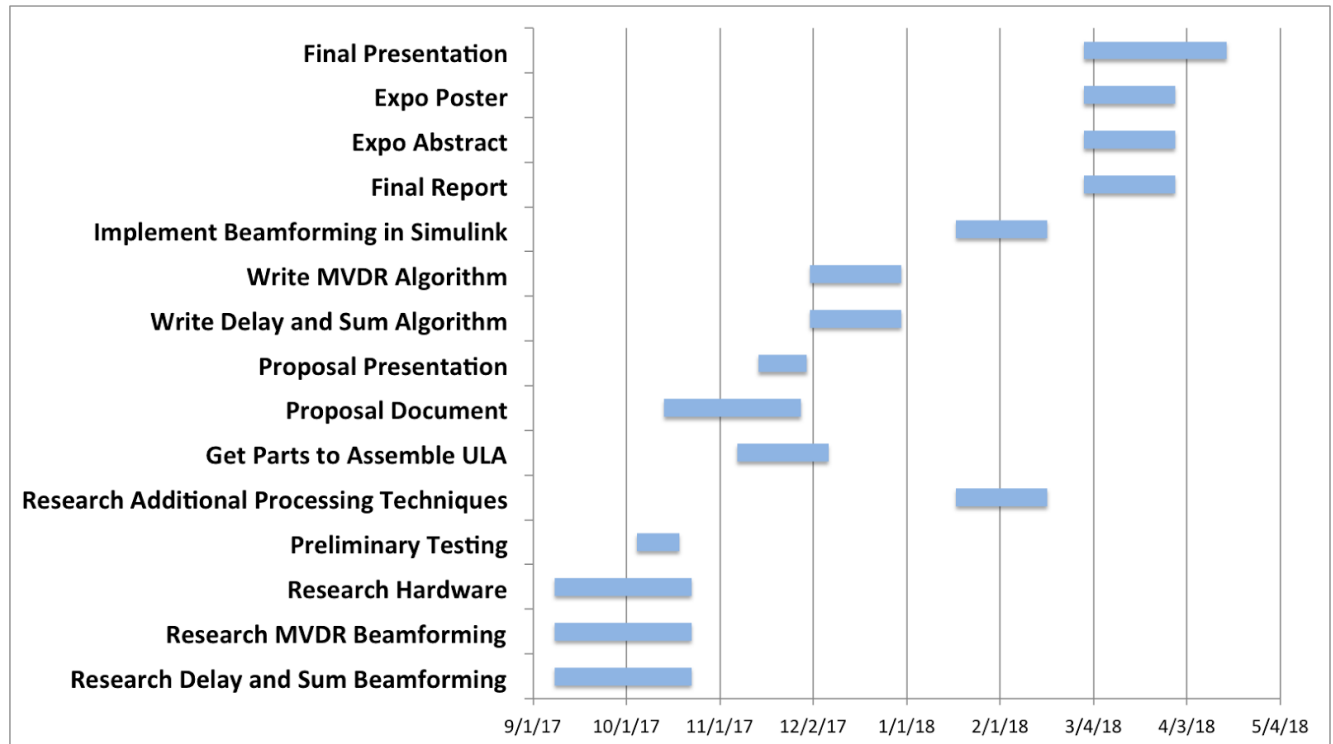
Engineering Efforts

Figure 2 – Gannt chart of engineering efforts

Parts List

Quantity	Description	Vendor Part #	Web Link To Item	Price	Ext. Price
1	XLR Patch Cables	B01M0JQX2E	https://www.amazon.com/Pack-Female-Microphone-Extension-Cable/dp/B01M0JQX2E/ref=s	$31.75	$ 31.75
3	BEHRINGER ULTRAVOICE XM1800S	B000NJ2TIE	https://www.amazon.com/Behringer-XM1800S-BEHRINGER-ULTRAVOICE/dp/B000NJ2TIE/r	$ 39.99	$ 119.97
5	Professional Black Adjustable Dual Plastic 2pcs Drum Microphone Mic Clip Clamp Mount KTV Karaoke	B06ZZCMJ26	https://www.amazon.com/Professional-Adjustable-Plastic-Microphone-Karaoke/dp/B06ZZCM	$ 7.44	$ 37.20
1	Scarlett 18i20				

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 Results

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

1. Test Setup

Three microphones were placed 0.2 meters apart from each other as shown in figure X. 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.
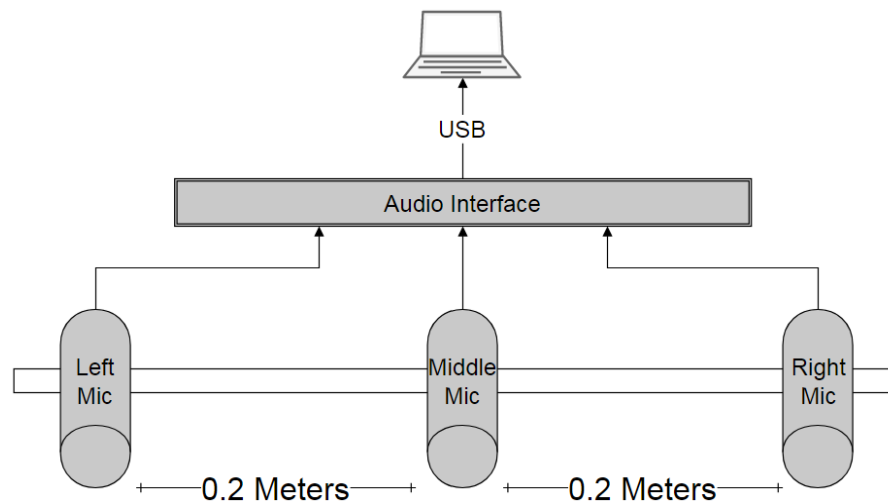

Figure 3 – Preliminary testing setup

2. Test Scenarios

There were two test scenarios that were used. The first was linear translation shown in Figure 4, and the second was spectral sweeping shown in Figure 5.

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.
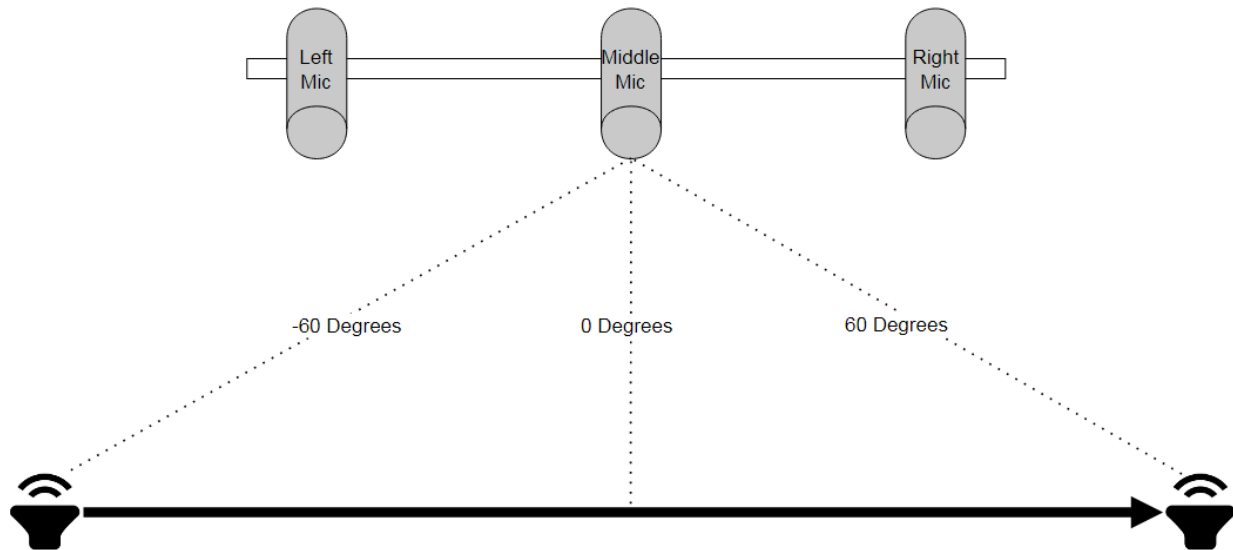
Figure 4 – Linear translation test

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.
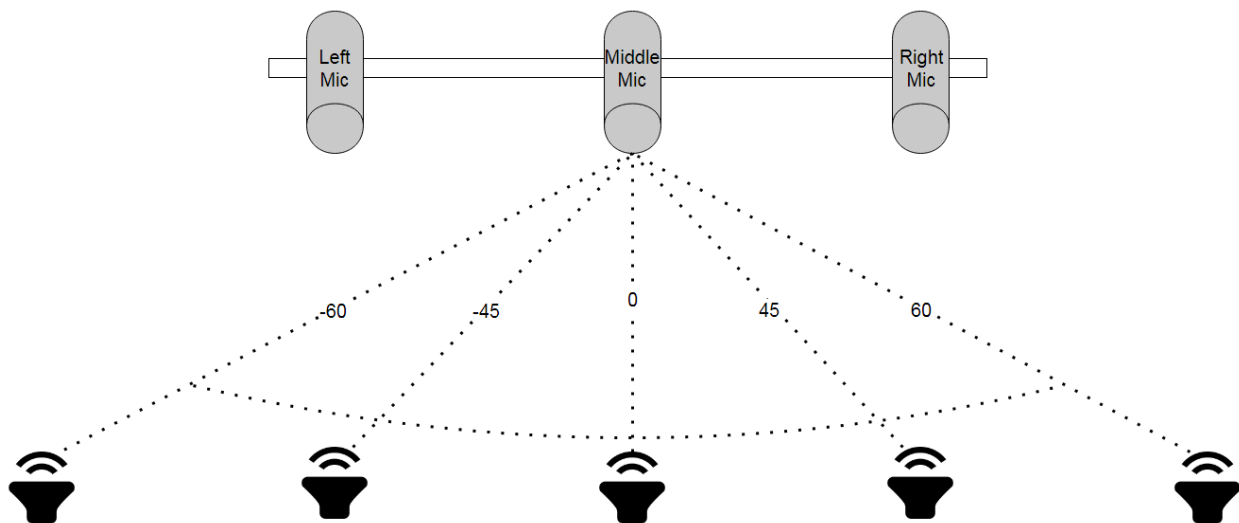

Figure 5 – Spectral sweep test

3. Theoretical Results

After the data was recorded, a MATLAB script was written to better understand how the beamforming algorithm performs. Figure 6 shows the beamforming algorithm acting on a sine wave at 1kHz from angles -180 degrees to 180 degrees. This figure is obtained assuming 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.
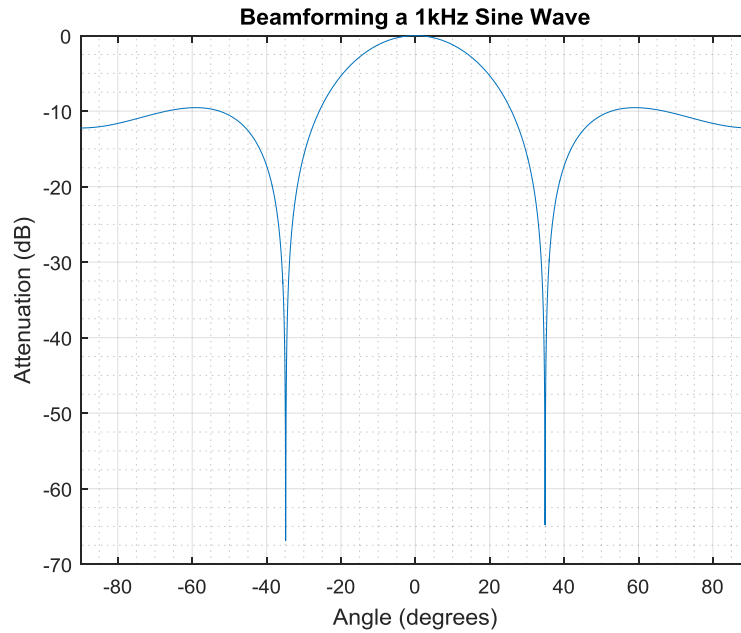

Figure 6 – Beamforming a simple sine wave in MATLAB

4. 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 7 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.
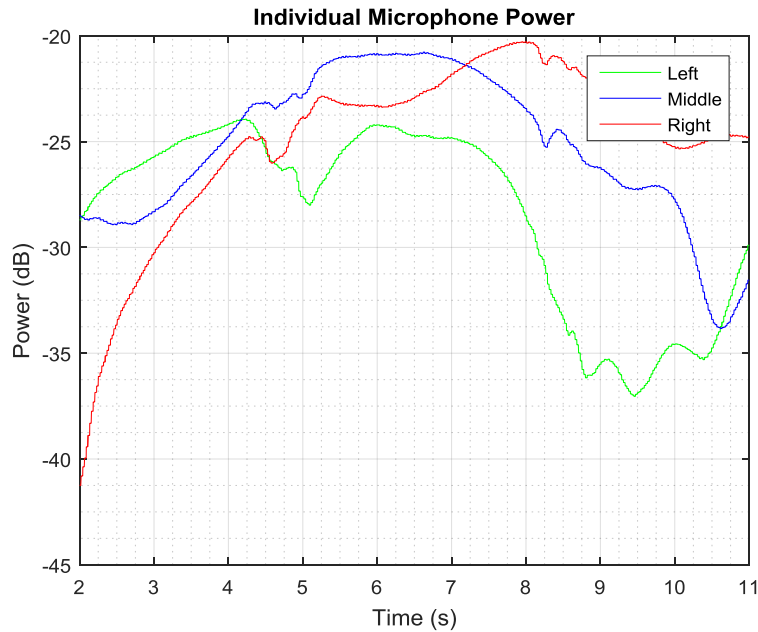
8

Figure 7 – Microphone instantaneous power

Figure 8 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 begging 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 6 shows.
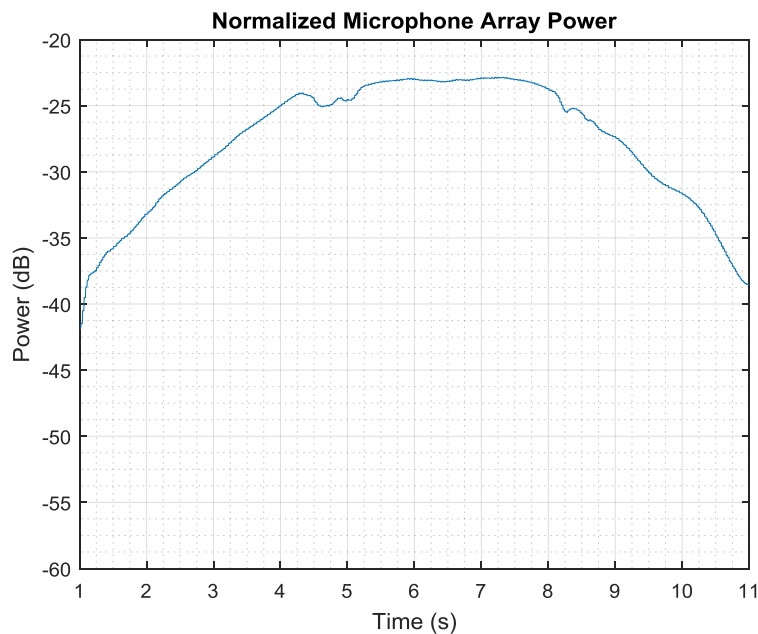

Figure 8 – Microphones summed and normalized

9

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 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 never temped to pick up his or her mobile device during operation of the vehicle.

The results from preliminary testing has 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 will be implemented. By increasing the number of microphones, the side lobes are subject to further attenuation, as seen in Figure 9. The spacing of the microphones will remain at 0.26 meters apart.

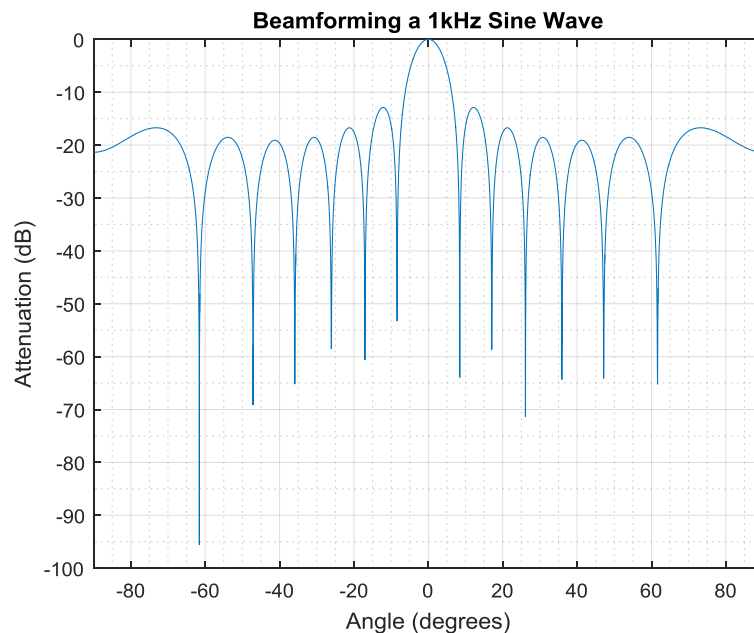This project is justified by the safety issues it solves and has promising preliminary results.

Figure 9 – Beamforming in MATLAB with 9 microphones

References

[1] "Texting and Driving Accident Statistics - Distracted Driving." *Edgarsnyder.com.* Accessed October 5, 2017. 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. https://www.mathworks.com/help/phased/ref/mvdrweights.html