SoundSelect Array System

A Directional Hearing Aid Capstone Project

EECE4790

Electrical and Computer Engineering Capstone 1

Professor Charles DiMarzio

Tim Deignan
Keenan Hye
Mark Long
Steve Muscari
Jack Tarricone

1 Contents

2 Abstract

Modern hearing aids improve speech-intelligibility by amplifying sounds around the user in a cardioid fashion. While this is suitable for most near-field situations, far-field intelligibility suffers as a result of lack of microphone directionality. Some devices allow for configuration of directionality, typically offering supercardoid or hypercardioid patterns. Although these may be beneficial in some niche situations, they introduce unwanted regions of sensitivity behind the user which may cause disorientation. In general, directionality capability for hearing aids has been particularly limited due to device size constraints.

The recent integration of Bluetooth technology into hearing aids now enables the creation of external, open source solutions. To this end, we propose a speech-optimized system composed of two beamforming microphone arrays coupled with existing hearing aid technology. This design allows for the extension of directional isolation associated with traditional beamforming to point isolation. We refer to this advanced beamforming concept as “spotlight-beamforming,” on account of its stark resemblance to a stage spotlight system.

3 Problem Formulation

In theory, modern hearing aids are designed to improve speech intelligibility. In practice, however, hearing aids simply amplify all sounds indiscriminately [1]. This elicits a feeling of being closer to all sounds within the general vicinity. In the near-field, this solution typically accomplishes the stated objective of improved speech intelligibility. In the far-field, however, this solution is far less effective. This is due to a lack of microphone directionality and a deterioration of signal to noise ratio (SNR).

For this the scope of this project, we will focus particularly on a traditional classroom setting featuring a professor at the front of the room and students distributed throughout. A hearing-impaired professor may struggle to hear students in the far field (i.e. the back of the classroom), even with the assistance of a modern hearing aid. Our goal is to enable said professor to hear students in the far field without having to physically move toward them. Our system is to be effective in both quiet and noisy classrooms.

4 Impact

Our design will have the most immediate impact on hearing-impaired individuals who frequently work in far-field settings including classrooms, lecture halls, or conference rooms. Although our project focuses specifically on the hearing-impaired, the implications of our speech spatial-isolation techniques can be extended for use in more general situations, including audio surveillance systems or sporting events. Furthermore, the speech band may be extended to include the full audio band, allowing for our isolation techniques to be applied to music-related applications.

5 Introduction

5.1 Human Hearing

The frequency range for healthy human hearing typically spans from 20 Hz to 20 kHz with minimum sensitivity of 0 dB. Certain individuals with exceptional hearing may even be capable of hearing sounds as low as -15 dB. Unfortunately, the opposite scenario is far more common.

5.1.1 Hearing Loss

In order to understand hearing loss, one must first understand how the ear functions. When a sound enters the ear canal, a differential pressure is applied across a taut membrane known as the eardrum. An intricate network of small bones translates this motion to the cochlea, a fluid filled organ known for its characteristic nautilus shape. Pressure waves propagating through the fluid excite tiny hair-like nerves in the cochlea, producing an electrical signal which travels to the brain via a bundle of nerves. Figure 1 shows the different regions of frequency sensitivity within the cochlea.

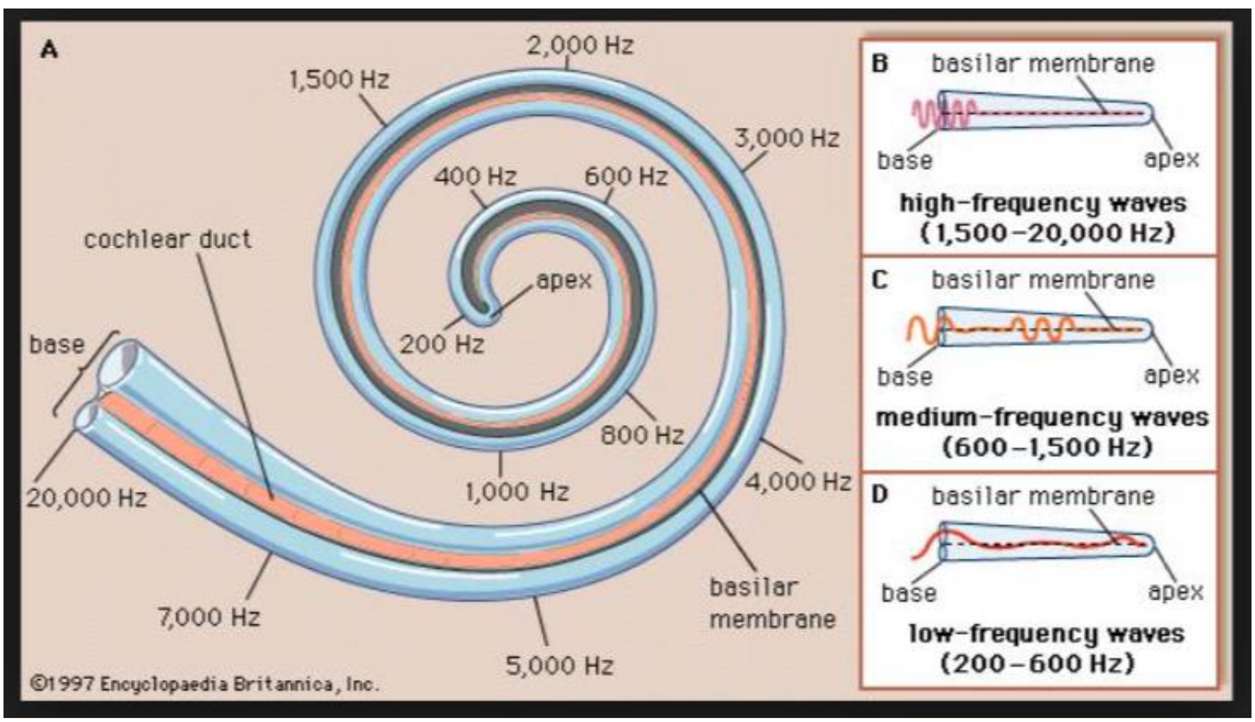

Figure 1: The cochlea [2]

The high-frequency sensitive nerves naturally tend to deteriorate with age. Premature damage to these nerves may also occur as a result of isolated events, including prolonged exposure to sounds over 85 dB, certain diseases, certain developmental problems, and reactions to certain medications.

5.1.2 Speech Perception

The spoken language is composed of building blocks known as “phonemes.” Phonemes include sounds like “th,” “ch,” and “sh.” As shown in Figure 2, phonemes distributed as a function of frequency and loudness form a banana shape. As such, this distribution has been affectionately dubbed, “the Speech Banana” [3].

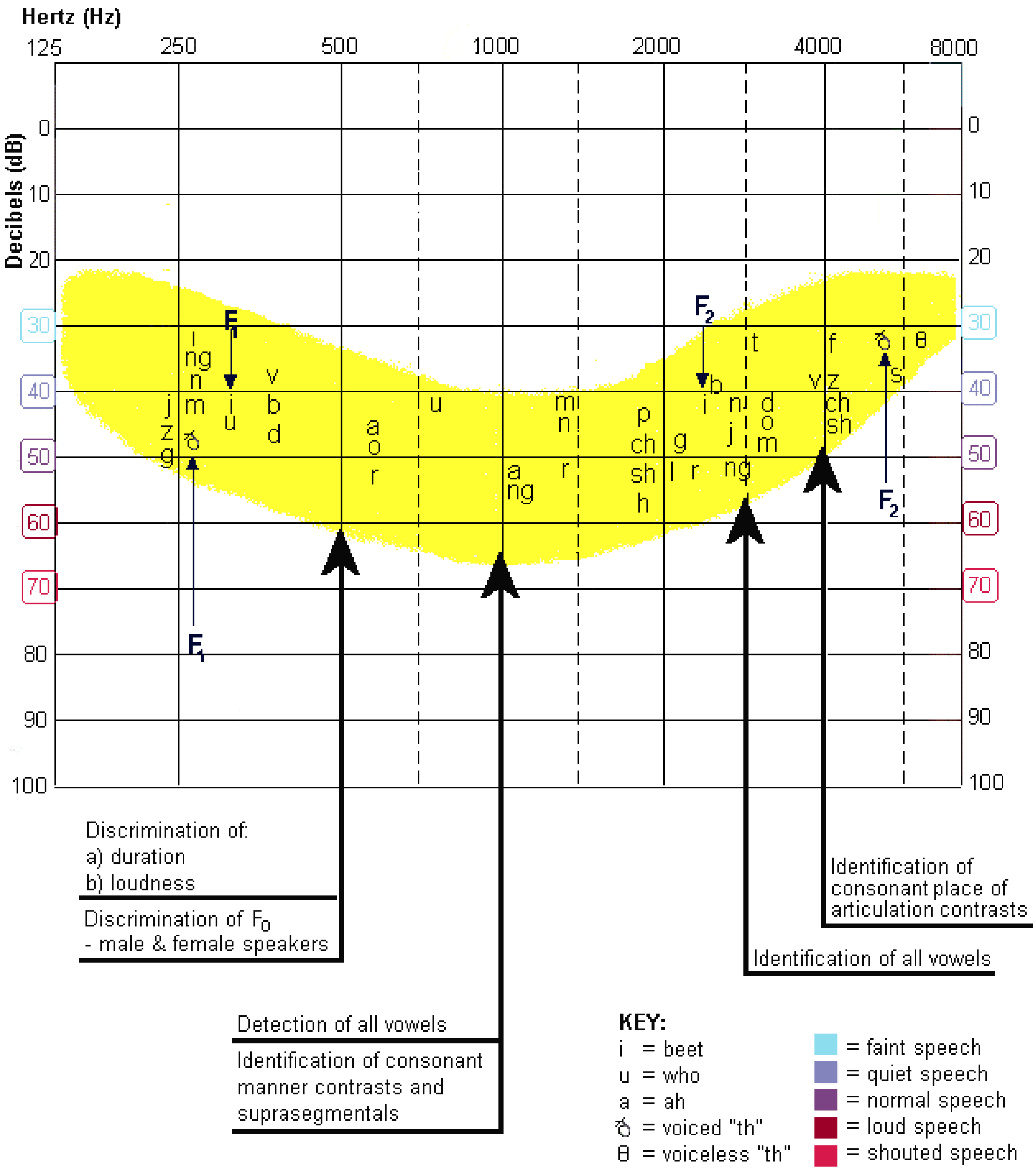

Figure 2: The Speech Banana [4]

Note that, in general, vowel sounds occupy the lower-frequency end of the Speech Banana whereas consonant sounds occupy the higher-frequency end of the Speech Banana. Therefore, hearing-impaired individuals with compromised high-frequency sensitivity typically experience difficulty hearing consonant tones. This is detrimental to overall perception of speech.

5.2 Microphone Arrays and Beamforming

In order to fully understand our system, one must first understand microphone arrays and how they relate to beamforming. As the name implies, microphone arrays are composed of microphones. Though there are many microphones available featuring non-uniform polar patterns, we will strictly discuss arrays composed of omnidirectional microphones henceforth, as these types of microphones are used most often in array applications.

A microphone with an omnidirectional response receives sounds of all frequencies equally from all angles in an imaginary circular plane orthogonal to the microphone diaphragm. This polar response, from a vantage point directed into the imaginary plane, is shown in Figure 3.

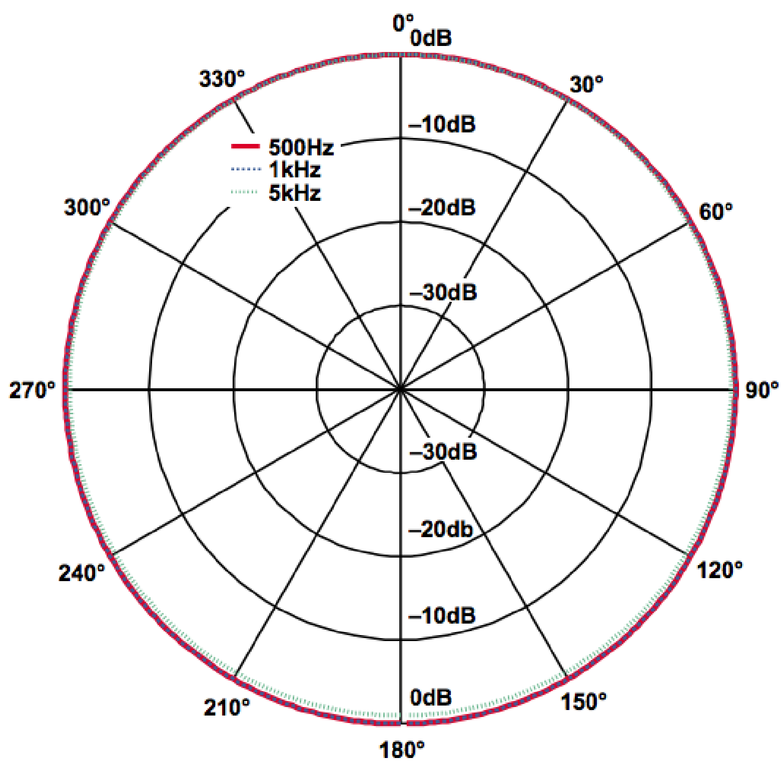

Figure 3: Polar pattern of an omnidirectional microphone [5]

When the outputs of multiple adjacent omnidirectional microphones are summed, their collective polar response deviates from this omnidirectional response. Rather, the directionality of a microphone array becomes a function of geometry, frequency, and angle.

5.2.1 The Simplest Microphone Array

The simplest microphone array is composed of two microphones. This two-element, linear array can be configured as either a broadside or endfire array. When configured as a broadside array, the incoming signals from each microphone are simply summed together. The broadside configuration is shown in Figure 4.

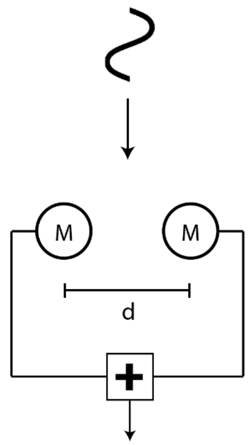

Figure 4: Microphone array in broadside configuration

Figure 5 shows the polar response of this configuration.

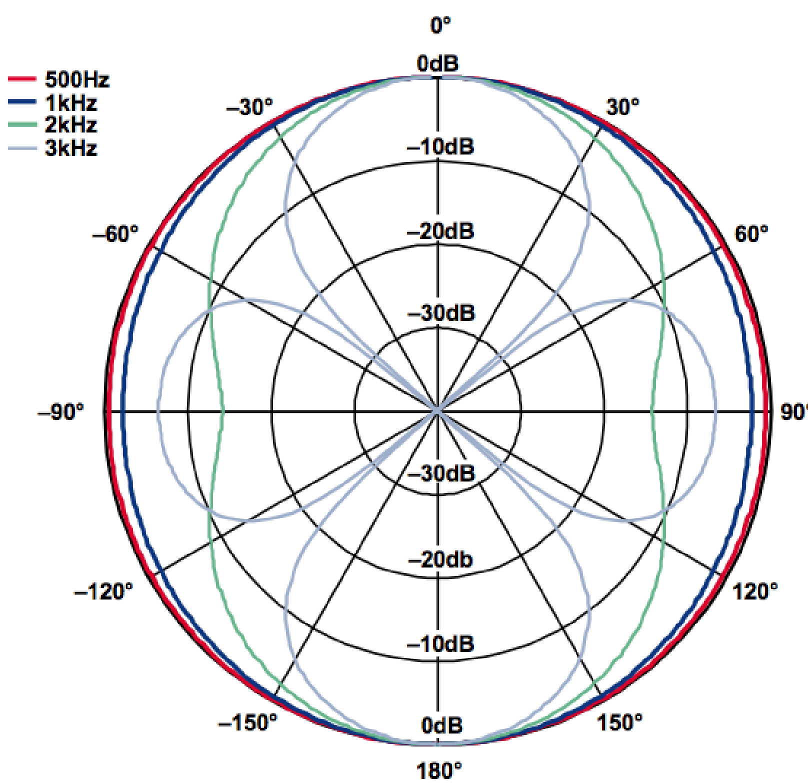

Figure 5: Polar response of a broadside array [5]

Note that the introduction of a second microphone has added directionality and frequency-dependence to the polar pattern of the microphone array. As its name implies, a broadside array features a main lobe emanating from its broadside. That is, signals approaching the array at 0° from normal undergo zero attenuation. As the signal frequency increases, the main lobe grows increasingly narrow, or increasingly directional, while sidelobes of lower magnitude begin to develop. Due to the symmetric nature of the broadside array, a rear-facing lobe exists with magnitude equal to that of the main lobe. This rear-facing lobe is an inherent disadvantage of the broadside microphone array.

The endfire array does not suffer from this disadvantage. An endfire array features a delay network prior to summation. This configuration is shown in Figure 6.

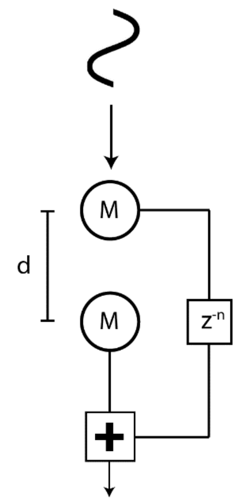

Figure 6: Microphone array in endfire configuration

The purpose of the delay network is to phase align the signals arriving at each microphone. Therefore, the magnitude of the time delay is equal to the time needed for sound to travel distance *d*, given by:

$$t = \frac{d}{v}$$

where v is the speed of sound at standard temperature and pressure. The magnitude of the sample delay n is given by:

$$n = f_s t$$

where f_s is the sample rate of the system. Figure 7 shows the polar response of this configuration.

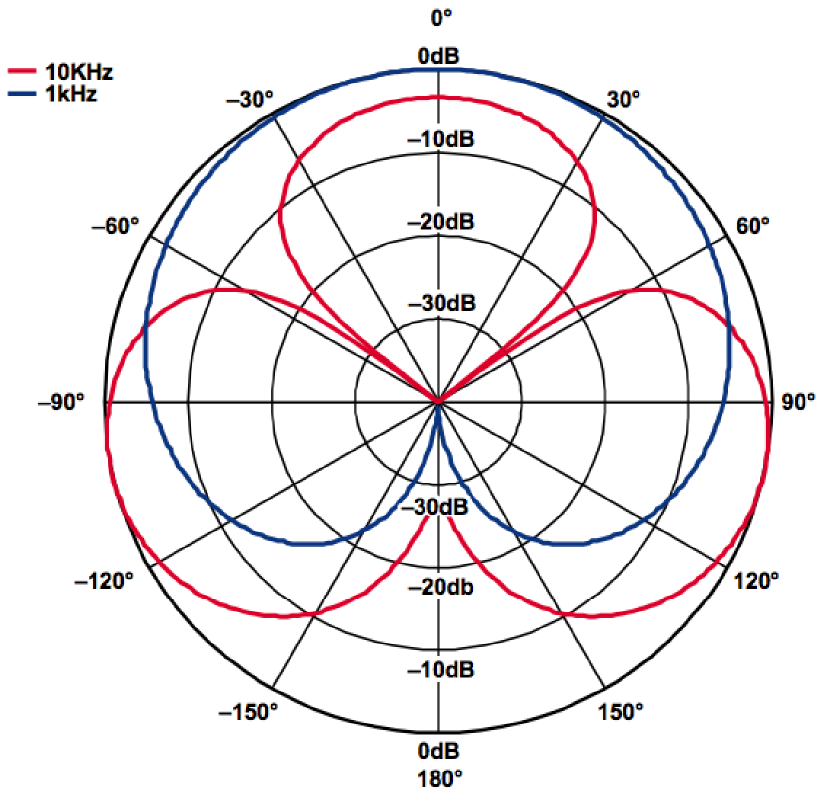

Figure 7: Polar response of an endfire array [5]

Note two important differences: (1) high attenuation in the direction opposite the main lobe for a wide band of frequencies (2) the main lobe of the endfire array resides in line with itself.

5.2.2 Beamforming

The two aforementioned microphone arrays are effective under the condition that desired signals lie within the main lobe. In practice, however, desired signals may approach a microphone array from any direction.

Delay-and-sum beamforming (DASB) addresses this problem. DASB enables the virtual-steering the main lobe of a microphone array by varying the sample delay of the microphones closer to the desired signal. DASB is extendible to a broadside microphone array of N number of microphones, as shown in Figure 9.

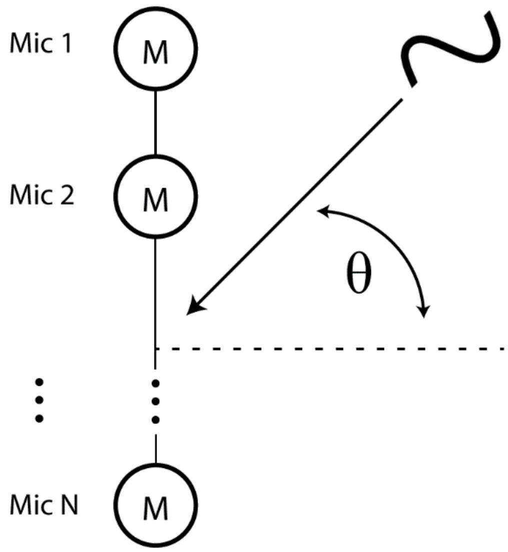

Figure 9: N-element microphone array with signal incident at angle θ

Given an incoming plane wave of direction θ and frequency *f*, each microphone will receive a given wave front with a delay that is a function of its position in the array, the array spacing *d*, and angle θ:

$$\tau_n = \frac{(n-1)d\sin\theta}{c}$$

Note that the delay τ_n is zero for plane waves of normal incidence. For these plane waves, the collective output *y(t)* of the microphone array is given by:

$$y(t) = \frac{1}{N}\sum_{n=1}^{N} x(t)$$

where *x(t)* is the signal seen by each microphone. DASB is implemented by multiplying the spectral content of signals arriving at each microphone by complex weights which are a function of delay τ_n:

$$W_n(f,\theta) = \frac{1}{N}e^{-j2\pi f\tau_n}$$

Note that θ now represents the desired steering angle. The resulting frequency-dependent signal is given by:

$$y(f,\theta) = \frac{1}{N}\sum_{n=1}^{N} x(f,\theta)w_n(f,\theta)$$

The resulting time-dependent signal is given by:

$$y(t) = \frac{1}{N} \sum_{n=1}^{N} x(t - \tau_n)$$

Figure 10 shows the capability of DASB. The polar pattern on the left is that of a non-steered three-element microphone array. The polar pattern on the right is that of the same microphone array steered at an angle of 30°.

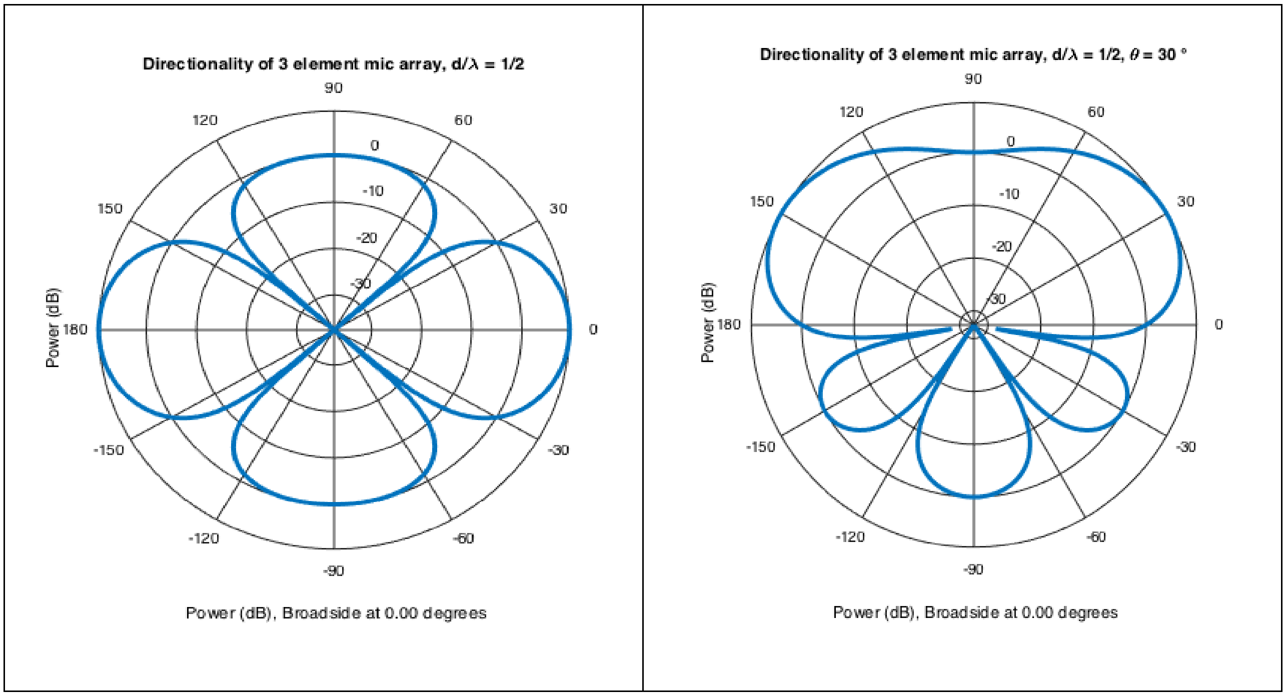

Figure 10: Capability of DASB

5.2.3 Directional Precision

The beam width of an array with microphone spacing d for a signal with frequency f and wavelength λ is given by:

$$\alpha = \frac{2\lambda}{d} = \frac{2c}{fd}$$

Beam width is minimized as the ratio d/λ is maximized. This yields improved directional precision. Figure 11 shows the polar patterns of three-element microphone arrays steered at an angle of 30° with different ratios of d/λ.

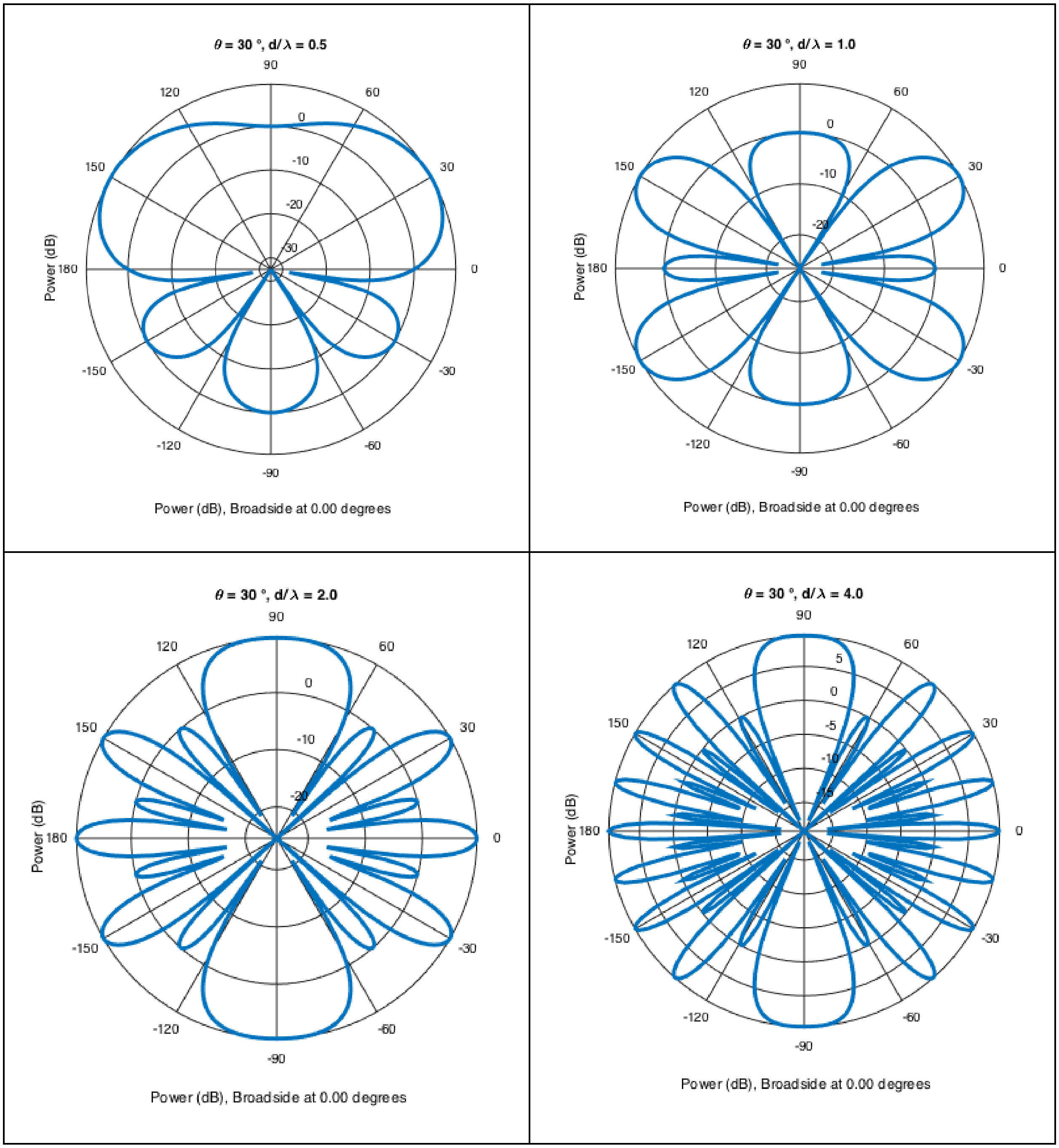

Figure 11: Effect of increased directional precision

Note that improvements in directional precision come at the cost of additional side lobes. These side lobes are result of spatial-aliasing. Spatial aliasing occurs if the following condition is met:

$$d \geq \frac{\lambda_{min}}{2}$$

5.2.4 More Complex Arrays and Geometries

5.2.4.1 Larger Linear Arrays

Figure 12 shows the response of an eight-element linear microphone array steered at 90° as a function of incident angle and frequency.

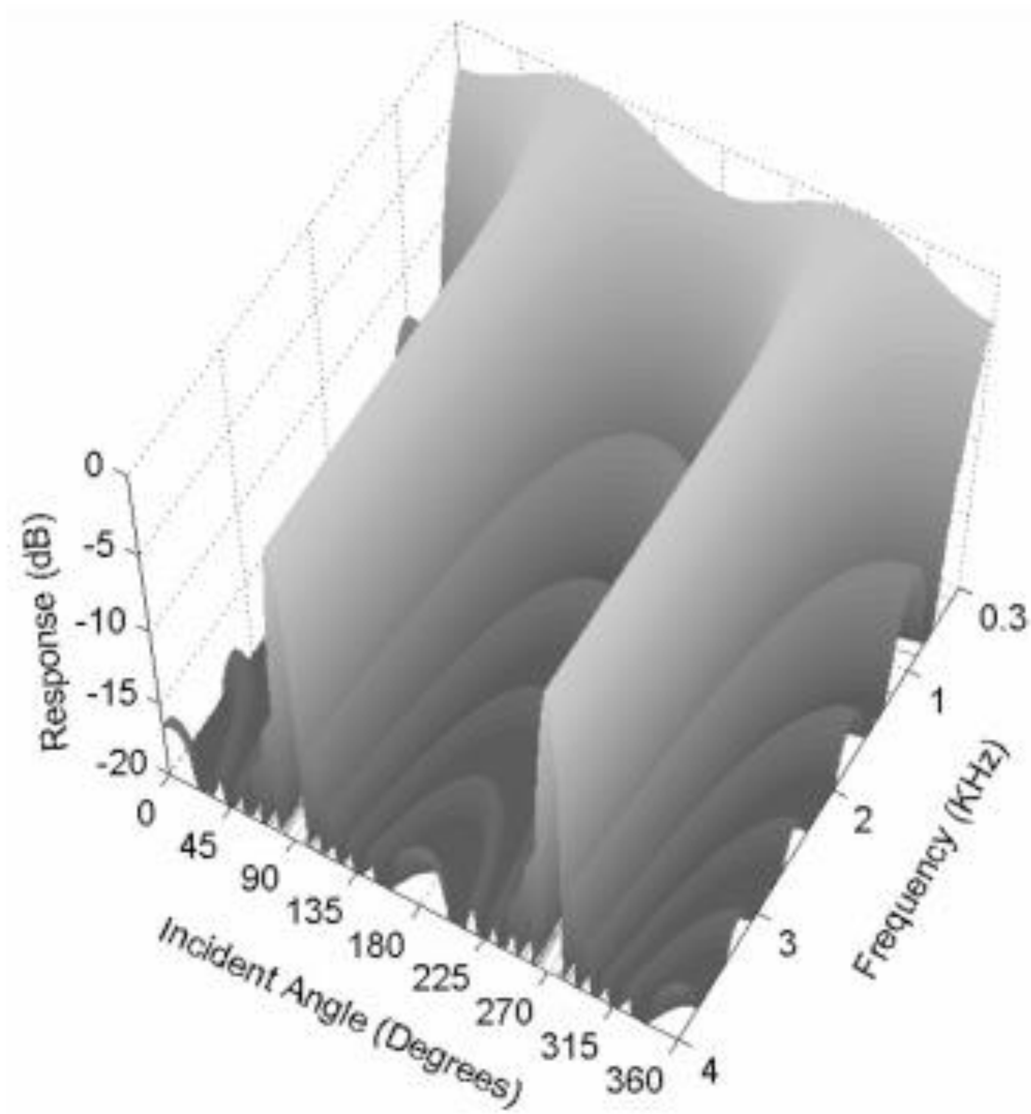

Figure 12: Linear microphone array response [6]

As expected, directional precision improves with higher frequency. Note that the side lobe at 270° is equal in magnitude to the main lobe.

5.2.4.2 Circular Arrays

Figure 13 shows the response of an eight-element circular microphone array steered at 90° as a function of incident angle and frequency.

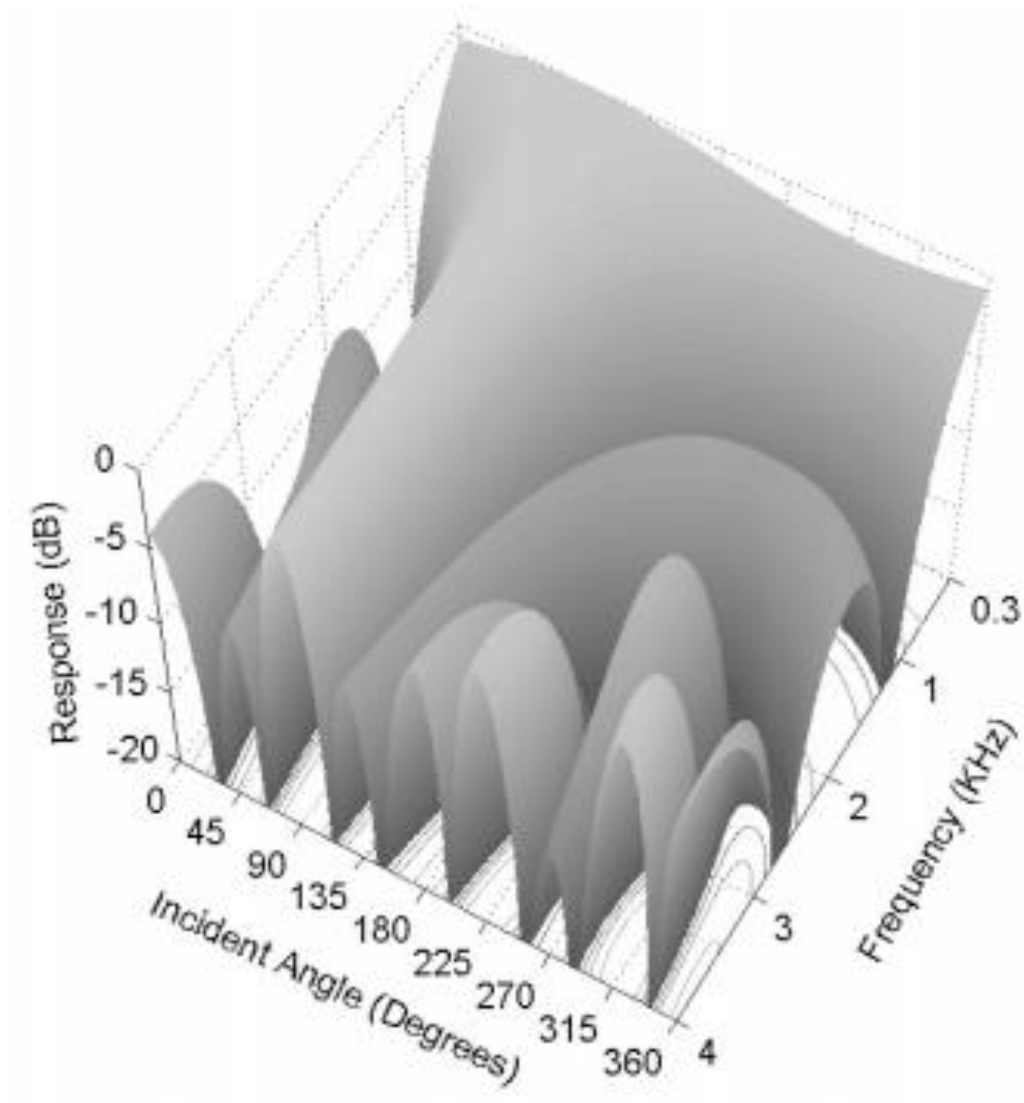

Figure 13: Circular Array Response [6]

While the circular array has significantly more side lobes than the linear array, none are as large in magnitude as the main lobe. This cannot be said of the linear array.

5.3 Basic Room Acoustics

5.3.1 Direct Sound, First Reflections, and Late Reflections

When a real sound emanates from a source within a room, it does not simply stop at the classroom boundaries. A typical classroom features highly reflective walls. This means that when a sound wave hits a wall, it is reflected back with a different amplitude and phase depending on the material from which the wall is made. The first bounce of this sound wave off of any given wall and into a microphone is aptly called a first reflection.

Usually a first reflection will attenuate, to some degree, the high frequency content of a sound, leaving the lower frequency content of the signal rather unaltered. The first reflection also features a time delay relative to the direct sound coming from the source, proportional to the added distance the sound has to travel to hit the wall, and then reach the microphone.

The filtering effect of the first reflection can be modeled by convolving the impulse response of the wall with our desired signal. The delay associated with the first reflection can be modeled by using source-imaging methods to simulate various faux-signals coming from outside of the room.

Unlike early reflections, late reflections refer to any reflected sound that bounces off more than one room boundary before hitting the microphone. Late reflections are also known as reverberation. Reverberation time is proportional to room size. Typically, reverberation will decay as the sound bounces off more and more surfaces until it reaches zero amplitude.

Unfortunately, due to the complexity of room acoustics, there is little that can be done in post-processing to mitigate long-tail reverberation. The advanced de-reverberation algorithms that do exist cannot be implemented in real-time, as they are too latent.

5.3.2 Noise Sources in a Typical Classroom Environment

In addition to long-tail reverberation, typical noise sources in a classroom include computer and projector fans, HVAC systems, electrical noise, and side conversations.

6 Design Strategy

6.1 Top Level Design

At the top level, our design is inspired by a device created by engineers at a university in Belgium known as the SoundCompass [7]. The SoundCompass is a non-real-time beamforming microphone system that is highly scalable, allowing for the integration of N number of separate microphone arrays. Our design aims to implement this concept in real-time.

Our design is composed of two separate linear microphone arrays operating in tandem. Using two separate microphone arrays allows for the extension of directional isolation associated with traditional beamforming to point isolation. We refer to this advanced beamforming concept as "spotlight-beamforming," on account of its stark resemblance to a stage spotlight system. Figure 14 shows a system level visualization of our design.

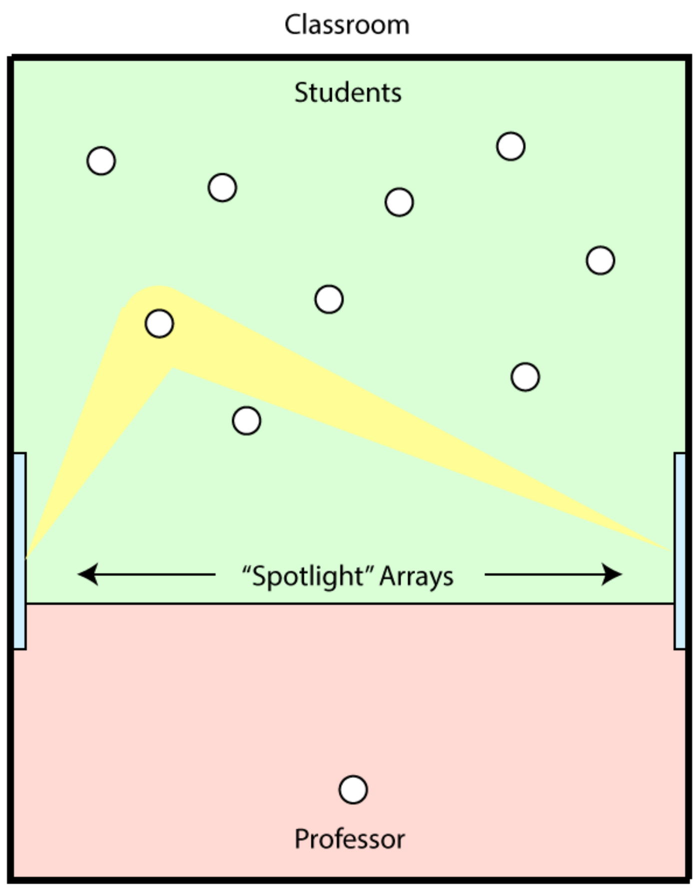

Figure 14: System level visualization

6.2 Informing our Design through Investigative Simulation

Recall the three-element linear microphone array steered at 30^o referenced throughout section 5.2. Figure 15 attempts to visually-project the polar patterns of this microphone array onto a classroom. In this case, colors represent the regions of sensitivity; warmer colors represent regions of signal reception, whereas cooler colors represent regions of signal rejection. Note that the darkest shade of blue saturates at -12 dB.

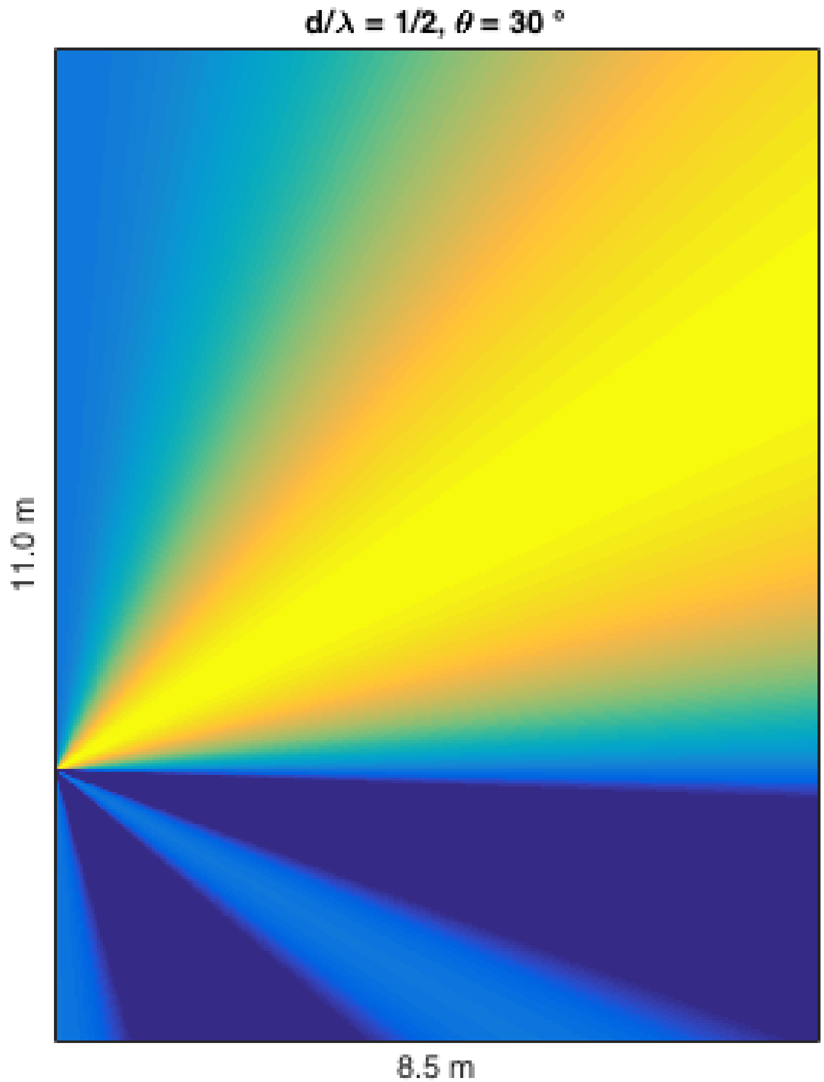

Figure 15: Room attenuation pattern, steered toward 30^o

Spotlight-beamforming is achieved by placing a second, identical microphone array on the opposite side of the room. Generally-speaking, wideband operation of microphone arrays can be simulated by varying the ratio of *d/λ*. Figure 16 shows the wideband operation of spotlight-beamforming.

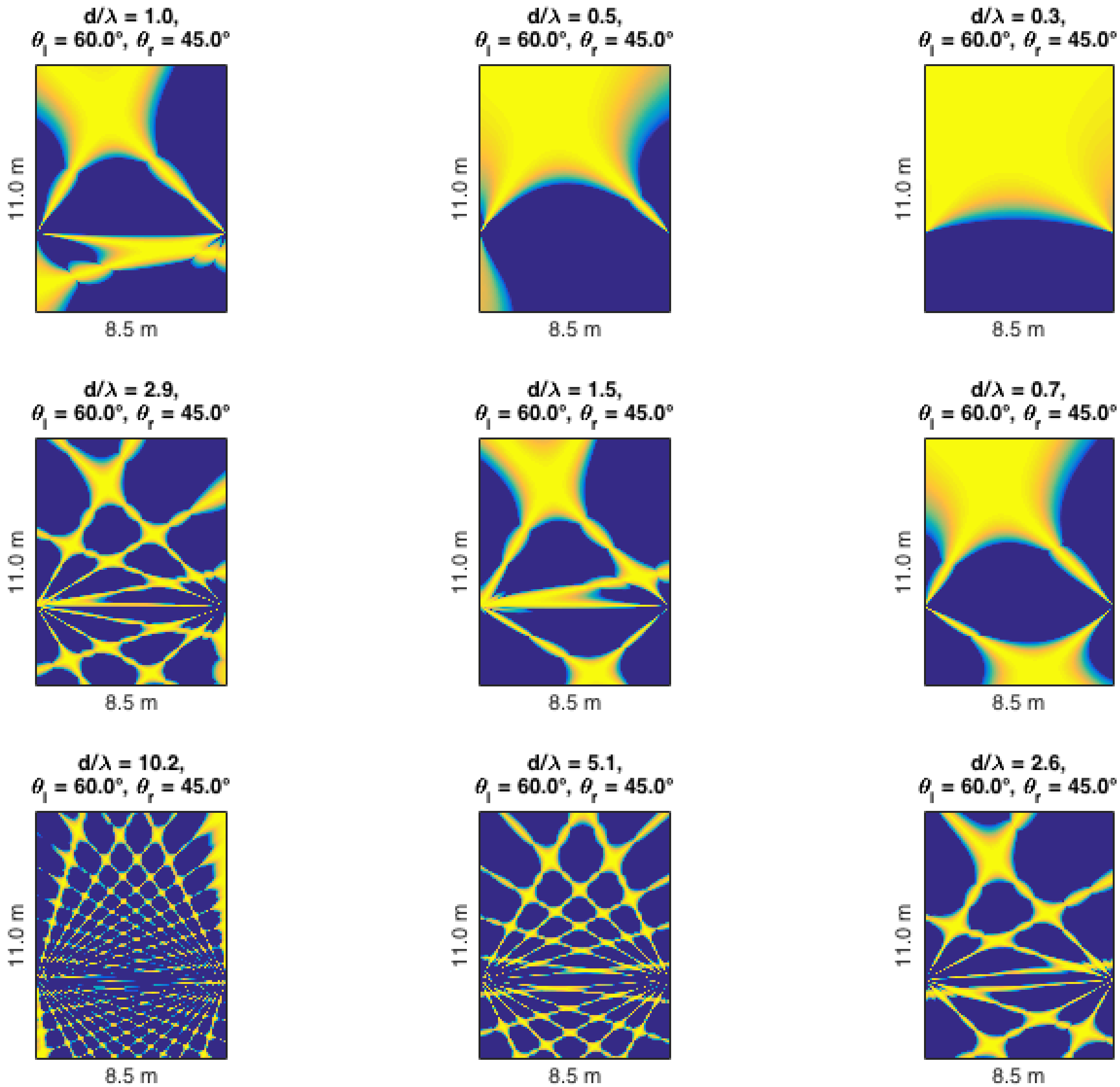

Figure 16: Spotlight-beamforming as a function of d/λ

Note that when the ratio of d/λ is less than one, spatial-aliasing is minimized, but the reception region is broad. Conversely, when the ratio of d/λ greater than one, the reception region grows more precise, but at the cost of creating additional, unwanted reception regions. Given this tradeoff, we believe the best compromise of these factors lies within the range 1.5 < d/λ < 2.9. Figure 17 shows the three attenuation patterns that fall within this range.

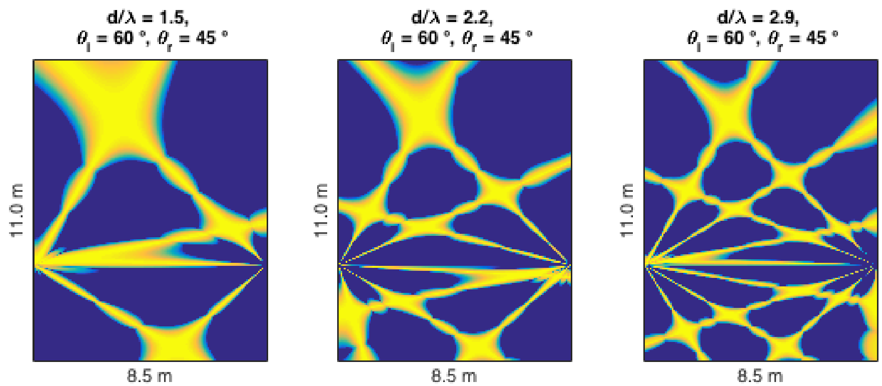

Figure 17: Attenuation patterns where 1.5 < d/λ < 2.9

Note that reception regions resulting from spatial-aliasing are merely narrowband, as their locations vary with frequency. The only wideband, fixed-location reception region is the desired reception region. Furthermore, effects of spatial-aliasing in the near field can be mitigated through the use of a dedicated noise microphone and parallel-processing.

6.3 Composite Microphone Arrays

Our microphone array design is composed of three nested microphone arrays. Figure 18 shows our composite microphone array design.

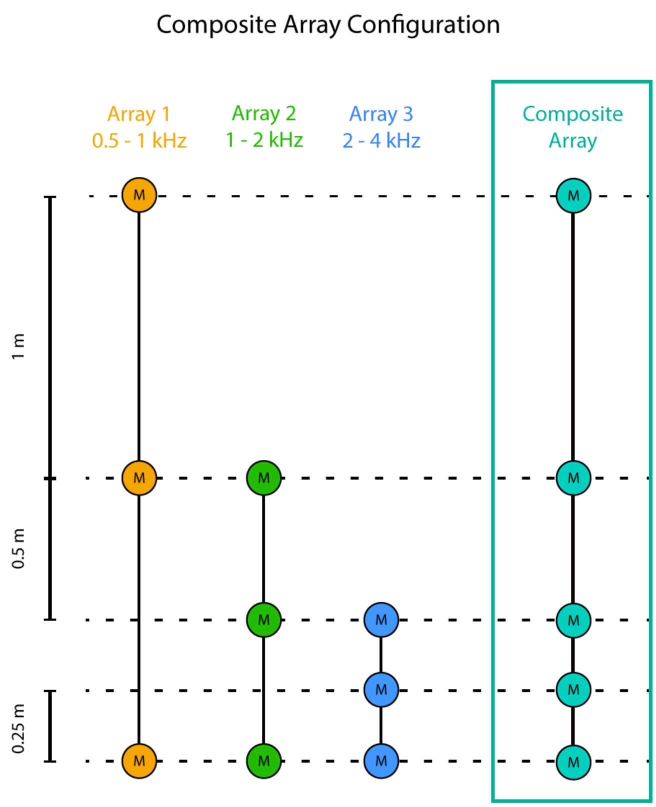

Figure 18: Composite microphone array

The composite array, shown in cyan, results from physically-overlapping the three band-passed arrays, shown in orange, green, and blue, respectively. Informed by the Speech Banana and investigation simulations, our respective microphone array spacing will be 1 meter, 0.5 meters, and 0.25 meters. In the context of sound waves, these respective array dimensions correspond to sub-bands of 500 Hz to 1 kHz, 1 kHz to 2 kHz, and 2 kHz to 4 kHz. Collectively, our composite array will be optimized across the entire speech band.

Note the asymmetrical nature of our design promotes microphone reuse, yielding an additional four virtual microphones for an effective total of nine.

While a full profile of two meters is fairly large, the microphone array itself will be unobtrusive, as it is merely composed of small MEMS microphones and will lie flush against the wall.

6.4 System Level Design

Figure 19 shows own our system design.

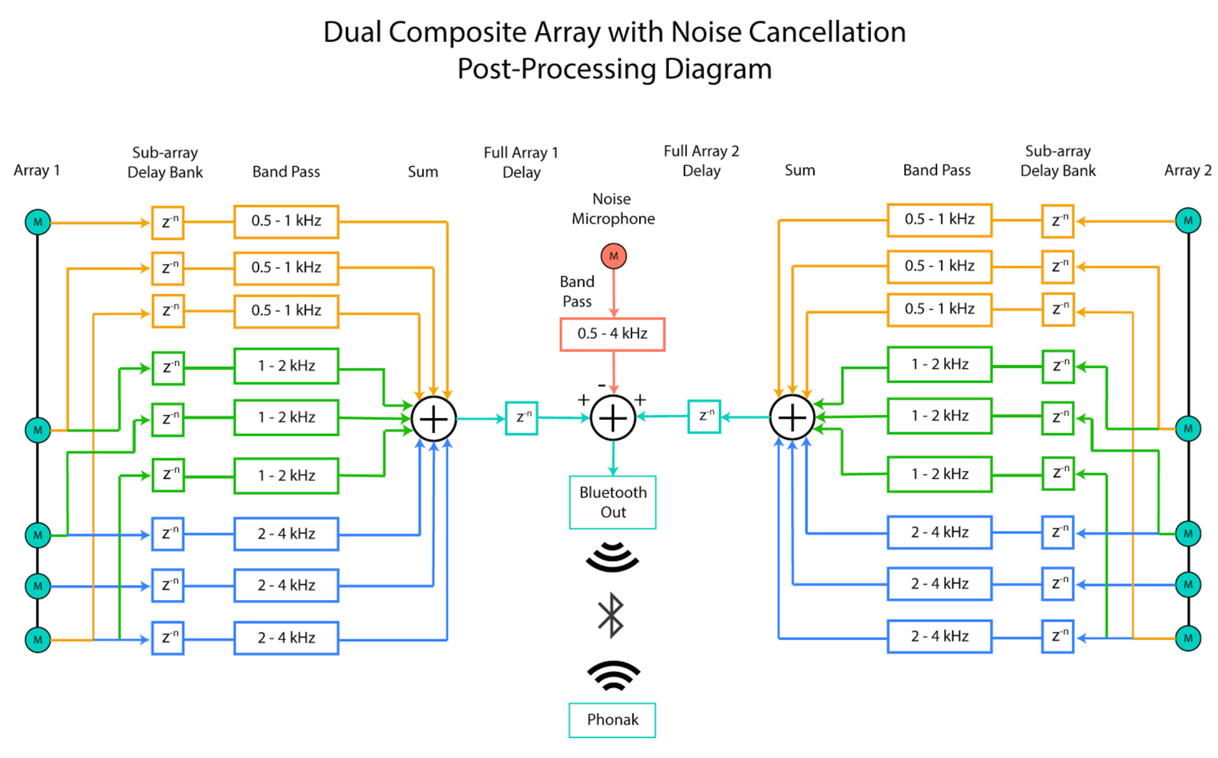

Figure 19: System level design

Each sub-array of the two composite arrays will undergo separate beamforming operations, as shown by the pair of sub-array delay banks. Note that the complex weights used within these delay banks will be a function of user input not shown in this diagram. Theses six beamformed signals will then pass through respective real-time, linear phase band-pass filters prior to summation. The two composite beamformed signals will then be phase-aligned and summed, yielding the spotlight-beamformed signal. Finally, the spotlight-beamformed signal will be sent to the Phonak ComPilot via Bluetooth (shown as mono, but subject to change). Note that the noise microphone represents a stretch goal that will be implemented following implementation of the spotlight-beamforming system.

Compression of the audio signal will be intentionally neglected, as many hearing aids already implement this functional block internally.

6.5 Hardware Specification

6.5.1 Processing

Our design features a broad array of requirements. The beamforming algorithm requires extremely low-latency execution, as received speech must remain in phase with a talker's motions. The Bluetooth communication, however, requires the implementation of a real-time operating system (RTOS). The ZedBoard development kit meets both of these requirements in one convenient package.

The ZedBoard features both a Xilinx Zynq-7000 FPGA and an ARM processor. The FPGA meets our low-latency needs, while the ARM processor supplies our RTOS. The ZedBoard also includes an abundance of I/O, enabling the connection of peripheral devices such as the Bluetooth radio and microphones.

We intend to program the ZedBoard using MathWorks embedded C/C++ and HDL coders.

6.5.2 Microphones

Our design will use analog MEMS microphones. The analog aspect provides two main advantages: (1) ability to transmit signals over non-trivial distances (2) ease of integration through the use of analog to digital converters (ADC). The MEMS aspect also provides two main advantages: (1) high far-field sensitivity (2) small physical package.

7 Performance Simulation

7.1 Performance Simulation Framework

For the purpose of our performance simulations, a typical classroom was generalized to lie flat in a grid 8.5 meters wide in the x-dimension and eleven meters deep in the y-dimension. That is, the front of the classroom (i.e. the whiteboard) falls upon the x-axis while the back of the classroom falls upon the horizontal line y = 11. Two microphone arrays were placed flush with the side walls of the classroom, three meters removed from the front of the room. The room was treated as an anechoic chamber. That is, for the purposes of our performance simulations, reflections were ignored.

Common use cases were composed of a combination of noise source types, signal/noise spatial orientation, and signal/noise separation gap. Noise source types included phonetically-balanced male speech, phonetically-balanced female speech, HVAC rumble, and typical classroom ambiance. Spatial orientations included horizontal orientation, where the signal source lies to the left of the noise source, and vertical orientation, where the signal source lies behind the noise source. Spatial gaps were selected to be one, two, and three meters. Figure 20 shows both spatial orientations, horizontal and vertical, each featuring a three meter gap.

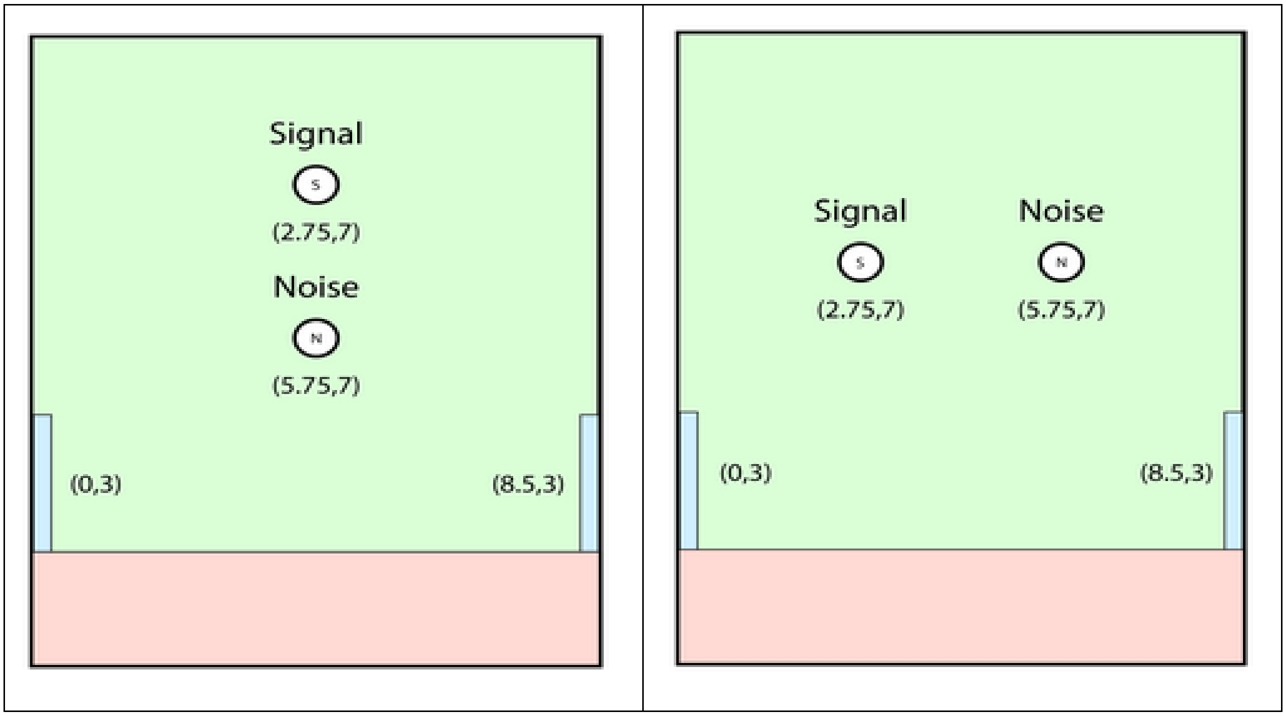

Figure 20: Vertical orientation (left) and horizontal orientation (right)

System performance was measured using a combination of objective and subjective metrics. Our objective metric was normalized signal to noise ratio (SNR). Normalized SNR is defined as the difference between the SNR of the beamformed signal and the SNR of the non-beamformed signal.

$$SNR_{normalized} = SNR_{beamformed} - SNR_{non-beamformed}$$

The non-beamformed signal is simply the signal one would expect from an omnidirectional microphone for the given use case.

Our subjective metric is composed of comparing the audio outputs of the signal, the non-beamformed signal and noise, and the beamformed signal and noise. This purpose of this metric is to ensure that the integrity of the signal is retained.

7.2 Performance Simulation Results

7.2.1 Individual Use Cases

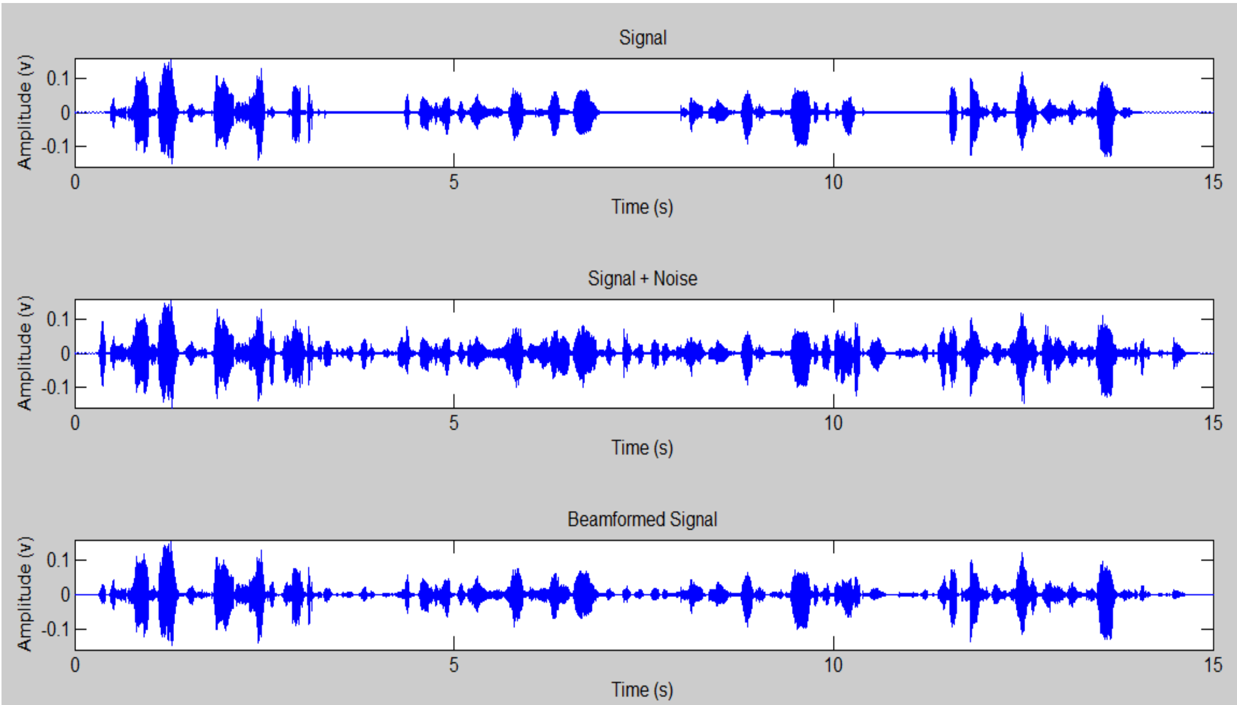

Figure 21: Male speech, horizontal orientation, three meter gap

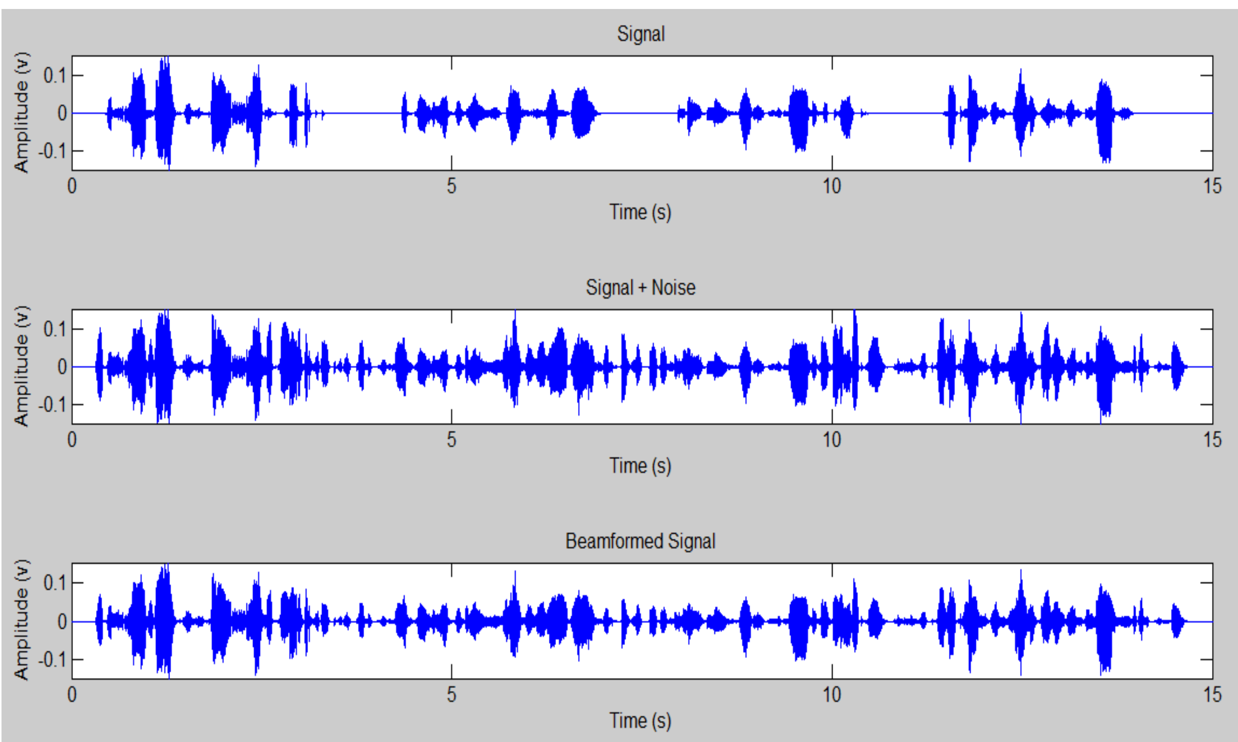

Figure 22: Male speech, horizontal orientation, one meter gap

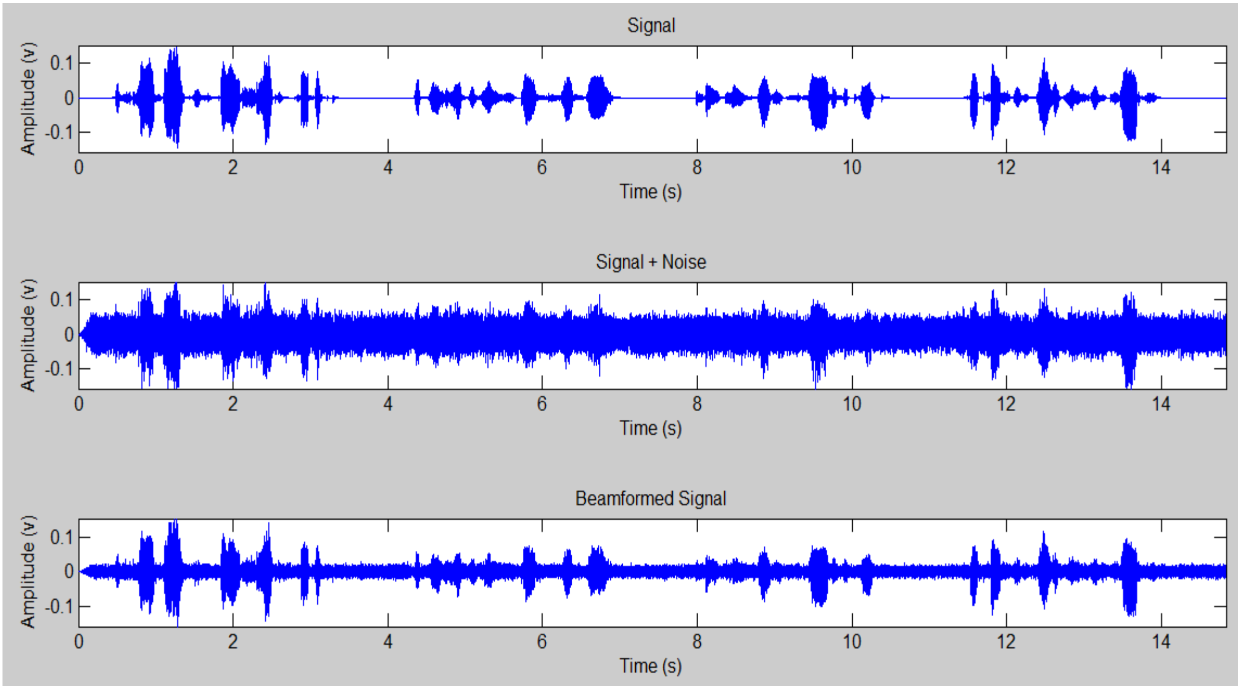

Figure 23: Industrial HVAC, vertical orientation, three meter gap

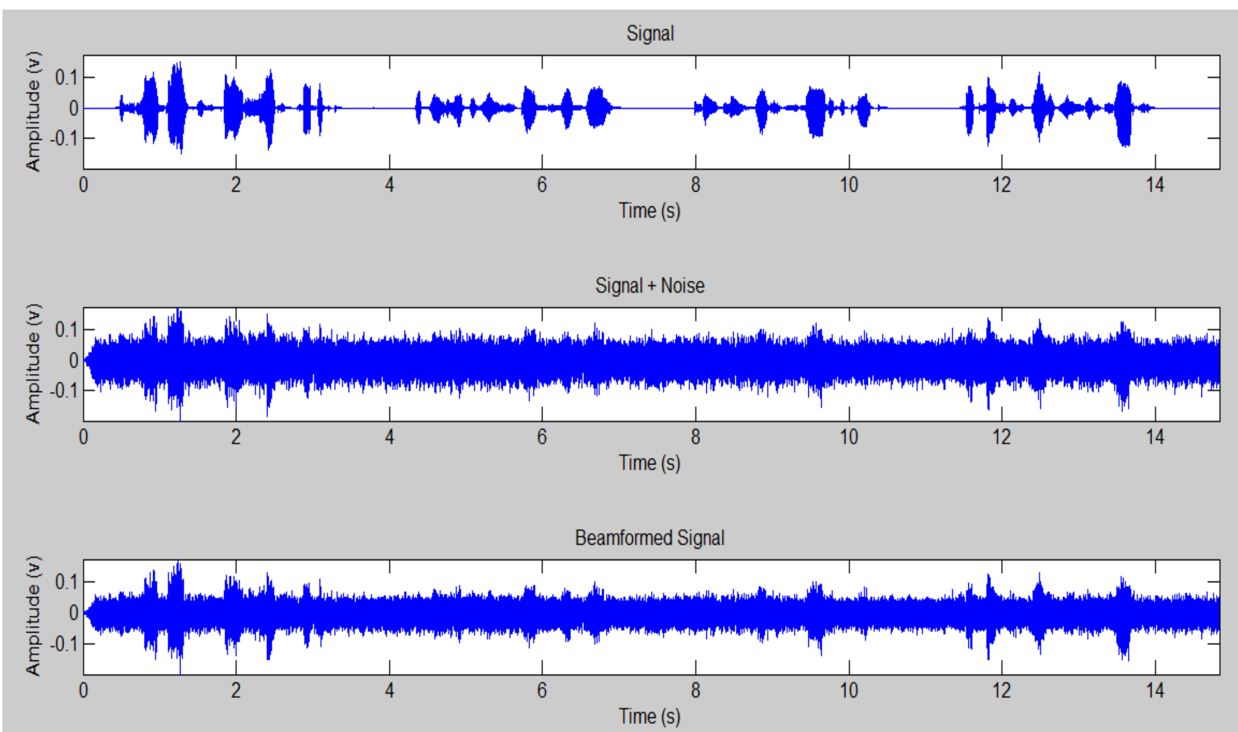

Figure 24: Industrial HVAC, vertical orientation, one meter gap

7.3 Use Case Summary

	Horizontal			Vertical		
	3	2	1	3	2	1
Male Speech	7.79	9.5	2.98	8.22	9.14	3.02
Female Speech	8.58	9.09	3.07	10.43	10.09	3.64
Industrial HVAC	12.75	6.2	2.71	10.46	8.62	2.29
Classroom Ambiance	7.27	11.38	4.02	6.59	10.18	3.84

Table 1: Normalized SNR (dB) as a function noise source type, spatial orientation, and gap (m)

In general, our system offers significant SNR benefits outside of two meters, and moderate SNR benefits outside of one meter. For the most part, normalized SNR for a given noise source and spatial orientation varied positively with gap size. However, in certain isolated instances, this relationship was exactly the opposite. Consider the noise source of classroom ambiance; when in the horizontal orientation, the normalized SNR increased from 7.27 dB to 11.38 dB as the gap size fell from three meters to two meters. This was due to the presence of side lobe intersection in the directivity pattern as a function of frequency and steering angle.

8 Resources

8.1 Team Roles

Team Member	Role
Tim Deignan	Software support for microphone A/D interface
Mark Long	Software support for ComPilot Bluetooth interface
Keenan Hye	Audio CODEC implementation for FPGA ARM interface, beamforming algorithm design support
Steve Muscari	Beamforming algorithm design and implementation
Jack Tarricone	Hardware support for microphone arrays and microphone A/D interface

Table 2: Team roles

8.2 Bill of Materials

Part No.	Description	Unit Price	Quantity	Donated	Cost
ZEDBOARD	ZedBoard Development Kit	319	1	Yes	0
SPU0414HR5H-SB	Analog MEMS Microphones	1.25	10	No	12.5
WL1835MODCOM8	WiLink 8 Bluetooth Module	65.55	1	No	65.55
AES-PMOD-WILINK8-G	WiLink 8 Bluetooth Pmod Adapter	34	1	No	34
ADS7813	Single Channel, 16-Bit, 40 ksps ADC with SPI	39.19	10	No	391.9
EPSA033180U-P6P-SZ	3.3 AC/DC Adapter	9.42	1	No	9.42

Table 3: Bill of materials

Total Cost: $513.37

9 Project Execution

Task	Resources	Start Date	End Date
Software installation - includes installing Xilinx Vivado Design Suite, upgrading to MATLAB 2015a, and installing ZedBoard Support Packages	Tim Deignan, Mark Long, Steve Muscari, Keenan Hye	8/25	9/9
Asses feasibility of hand soldering MEMS microphones	Jack Tarricone	8/25	9/9
Specify hardware interfaces for FPGA, ADCs, and microphones	Jack Tarricone	9/10	9/25
Interface an FPGA with a single ADC - log test audio streams to MAT file	Tim Deignan	9/10	9/25
Interface ARM processor with Bluetooth device - send a prerecorded audio stream to a Bluetooth speaker	Mark Long	9/10	9/25
Interface FPGA with ARM processor through implementing an audio CODEC	Keenan Hye	9/10	9/25
Design and implement beamforming algorithm	Steve Muscari, Keenan Hye	9/10	10/2
Partial system integration - send a real time audio stream from a microphone/microphone array to a Bluetooth Speaker	Tim Deignan, Mark Long, Jack Tarricone	9/25	10/2
Full System integration	All	10/3	10/16
System Tuning	All	10/17	10/30
Polish UX	Tim Deignan, Mark Long	10/31	11/20
Implement noise-cancelling microphone	Keenan Hye, Steve Muscari, Jack Tarricone	10/31	11/24
Compile Documentation	Tim Deignan, Mark Long	11/21	11/24

Table 4: Project execution schedule

10 References

[1] N. Disarno, "What Should be done About Hearing Loss?" *The New York Times*, Feb. 27, 2013. [Online]. Available: http://www.nytimes.com. [Accessed Aug. 16, 2015].

[2] "Electrodes and Channels," *Cochlear Implant HELP.* [Online]. Available: http://cochlearimplanthelp.com/journey/choosing-a-cochlear-implant/electrodes-and-channels/. [Accessed: Aug. 16, 2015]

[3] "The Speech Banana," *Listening and Spoken Language Knowledge Center.* [Online]. Available: http://www.listeningandspokenlanguage.org/SpeechBanana/. [Accessed: Aug. 16, 2015]

[4] "Speech Banana: Interpretation of Acoustics of Speech," *First Years Development Through Distance Education.* [Online]. Available: http://firstyears.org/c1/u2/banacoustics.htm. [Accessed: Aug. 19, 2015]

[5] "Microphone Array Beamforming," *InvenSense.* Dec. 31, 2013. [Online]. Available: http://www.invensense.com/wp-content/uploads/2015/02/Microphone-Array-Beamforming.pdf. [Accessed: Aug. 20, 2015].

[6] A. AlShehhi *et al.*, "Linear and Circular Microphone Array for Remote Surveillance: Simulated Performance Analysis," Dept. of Electrical and Computer Eng., Khalifa Univ. of Science, Sharjah, UAE. [Online]. Available: http://www.academia.edu/3154683/Linear_and_Circular_Microphone_Array_for_Remote_Surveillance_Simulated_Performance_Analysis. [Accessed: Aug. 20, 2015].

[7] J. Tiete *et al.*, "SoundCompass: A distributed MEMS Microphone Array-Based Sensor for Sound Source Locatization." Dept. of Electronics and Informatics, Vrje Univ., Brussels, Belgium. Jan. 23, 2014.