Microphone Array Feedback Suppression for Indoor Room Acoustics

by

Tanmay Prakash

Advisor: Dr. Jeffrey Krolik

Department of Electrical and Computer Engineering

Duke University

Abstract

The objective of this project was to use a standoff microphone array to suppress feedback for a microphone-loudspeaker system in an indoor environment. When the gain on the amplifying loudspeaker is too high, the sounds from the loudspeaker that are picked by the microphone keep being re-amplified until there is a loud squeal in a process called feedback. This limits the amount of gain that can be put on the microphone-loudspeaker system. In this project, minimum variance distortionless response (MVDR) beamforming was used to focus a microphone array on only the source and attenuate the sound from the loudspeaker, thereby suppressing the feedback loop. The maximum gain that could be applied to the loudspeaker without inducing feedback was measured experimentally for MVDR and compared to that of conventional beamforming and no beamforming. The experiments were performed on a system that processed the output of 4 microphones from an NIST MK3 array online in Java and output the processed signal to the loudspeaker. The results showed that MVDR allowed for a greater gain than both conventional beamforming and no beamforming.

Table of Contents

1 Introduction

1.1 Problem

A common objective of many sound systems is to amplify the sound from one source using a microphone connected to a loudspeaker with some gain. A common issue seen by these systems is feedback: when the microphone picks up the sound from the loudspeaker, applies gain and outputs to the loudspeaker and then continues to repeat this process until the output exceeds the limits of the loudspeaker. This system can be modeled as below

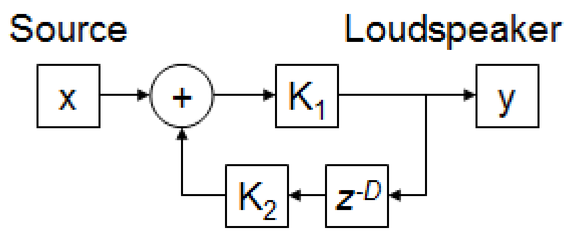


Figure 1: Acoustic Feedback System [3]

where K_1 represents the gain being applied to the microphone array, K_2 represents the attenuation that occurs as the sound travels from the loudspeaker to the microphone, and D represents the delay in samples between the when the sound is first heard from the source and again from the loudspeaker. The transfer function of this system is

$$\frac{Y}{X} = \frac{K_1}{1 - K_1 K_2 z^{-D}}.$$

In the z-plane, this system as a ring of poles spaced evenly around the origin. For this system to remain stable, the poles must remain inside the unit circle, so

$$K_1 K_2 < 1.$$

Because K_2 is a property of the room, this means K_1 must be limited to prevent feedback. This limitation on gain can problematic for auditoriums, hearing aids and public announcement systems, where feedback often occurs, but large gains are desired. One of the earlier methods of suppressing feedback was to shift the frequencies before outputting to the loudspeaker, preventing the sound from building up at any one frequency [5]. This, however, produces a distortion in the signal that may be undesirable. Another common solution is use a filter that places zeros at the same spots as the poles. However, the values of D and K_2 vary with different rooms, arrangements of furniture, and the presence and movement of people among other things. This means the poles move around and this solution cannot account for these changes. As is often the case, any change to the room where this solution is applied requires an audio engineer to readjust the filter. There has been some work into adaptively adjusting a notch filter to the frequencies at which feedback occurs [6], but this method only adapts after the feedback has occured. Multi-channel systems can provide solutions that don't require distorting the desired signal. Rombouts presents the use of generalized sidelobe cancellers (GSC) in the time domain on a microphone array to both suppress feedback and noise and demonstrates their effectiveness in simulation [7]. Odom uses minimum variance distortionless response (MVDR) beamforming in the frequency domain and demonstrates its effectiveness in simulation as well [2]. Under certain conditions, GSC is equivalent to MVDR beamforming [8], though the actual implementation of each would be different. This paper extends the work done by Odom by testing MVDR with an experimental setup in the real world.

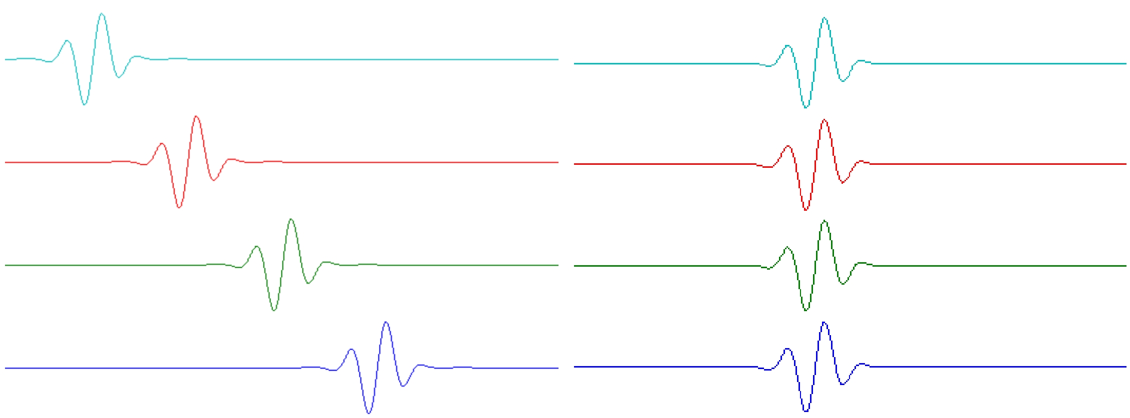
1.2 Conventional Beamforming

The use of a microphone array presents the possibility of spatial filtering to separate the loudspeaker and the targeted sound source, assuming they're located at different directions. Conventional beamforming is one possible solution, since it amplifies only the sound from the steered direction and attenuates sounds from other directions [4]. Sounds from one source arrive at different microphones at different times depending on the location of the microphones and the source. Though the sound propagates away from the source in an expanding circle, if the source is far enough away from the array, the curvature of the circle is flat enough that the sound can be considered a plane wave and any range information is lost. If the microphone array is linear and equally spaced, the difference in the time of arrival of the sound for two microphones is

$$\Delta t = d\sin\theta / c$$

where d is the distance between the microphones, θ is the azimuth of sound source and c is the speed of sound.

Conventional beamforming uses the knowledge of the different times of arrivals to focus on the sound from one direction, acting as a digital, steerable parabolic dish.



(a) (b)

Figure 2: (a) Signal at each microphone (b) Signal from each microphone with beamforming delay

It shifts the signal from each of the microphones to compensate for the delay that would be experienced if the sound were coming from the steer direction and then sums all of the signals together. Because of this, it's often referred to as delay and sum beamforming. Sounds coming from the steer direction would be aligned in all of the microphones and the summing would increase the gain of this signal. Sounds coming from other directions would not align, so they would be attenuated. The amount of attenuation depends on the direction and frequency of the sound. In the frequency domain, the signal at each microphone in a linear array can be represented as

$$v(\theta) = \left[e^{-jd_1 \frac{\omega}{c} \sin\theta} \cdots e^{-jd_m \frac{\omega}{c} \sin\theta} \right]^T$$

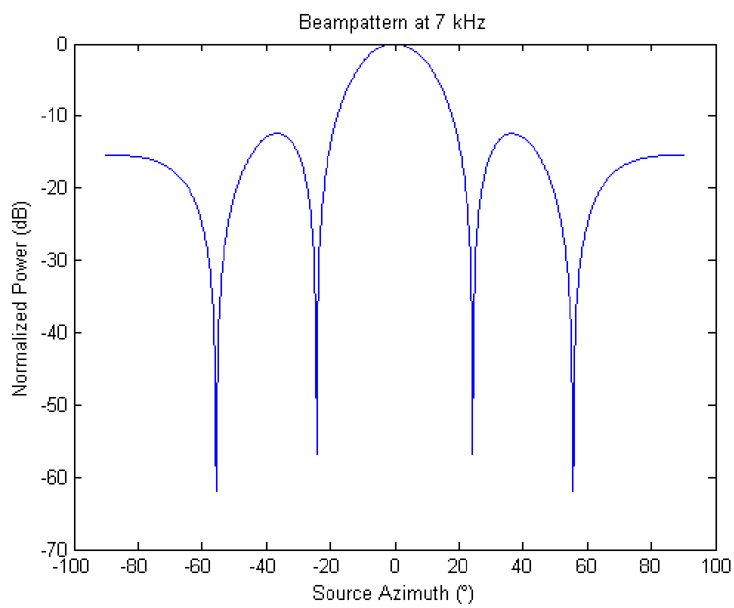
$$x = v(\theta)s(\omega)$$

where θ is the direction of the sound source relative to the array, d_i is the position of the microphone relative to the center of the array, c is the speed of sound, ω is the frequency in radians, $s(\omega)$ is the value of the signal from the source at ω, and x is a column vector containing the value of each microphone at this frequency. The column vector x is often referred to as a snapshot, while v is often called a steering vector. For conventional beamforming, the output, y, at one frequency is

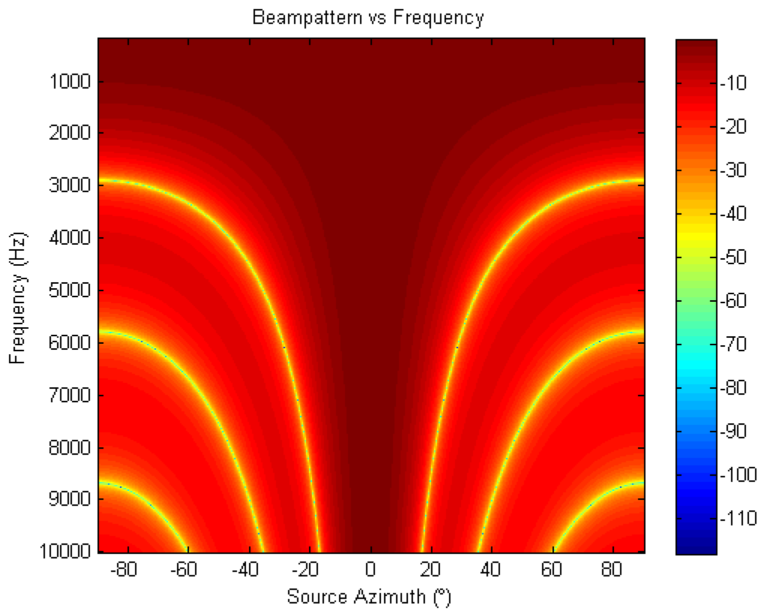
$$w = v$$

$$y = w^H x$$

The column vector w in many methods of beamforming is called the beamforming weights. It is simple to see that if the beamforming weight is the same steering vector as the one used to calculate x, then y is $Ms(\omega)$ where M is the number of microphones.



(a)



(b)

Figure 3: (a) Beam Pattern for 6 microphone linear array at 7 kHz (b) Beam Pattern for this array from 200 Hz to 10 kHz

Figure 3a shows a beam pattern of a linear microphone array with 6 elements spaced 2 cm apart at 7 kHz, demonstrating how the power output by the beamformer steered at 0° varies when the sound source is at different azimuths, while Figure 3b shows how the beam pattern varies with frequency. This shows that conventional beamforming would have some success in filtering out the loudspeaker, so long as the loudspeaker isn't located in the main lobe. However, Figure 3b shows that the main lobe becomes wider at lower frequencies, so a conventional beamformer would be susceptible to feedback for sounds at lower frequencies.

1.3 Minimum Variance Distortionless Response Beamforming

Minimum variance distortionless response (MVDR) beamforming is a form of adaptive beamforming that can suppress sounds from undesired directions and focus on the desired one more effectively than conventional beamforming [1]. The beamforming weights are calculated such that they minimize the power of the output, while still preserving the signal from the steered direction by placing nulls at the directions of interference. The beamforming weights are calculated as

$$w = \frac{R^{-1}v}{v^H R^{-1} v}$$

where v is the steering vector pointed towards the source, and R is the covariance matrix [10]. The covariance matrix is approximated as

$$R = x_{old} x_{old}{}^H$$

where x_{old} is a matrix whose columns are some number of the most recent snapshots. This use of previous data is what allows MVDR to adapt to the interference and focus more effectively on the targeted source. The weights are then applied to the current snapshot, x_{new}, find the beamformer output, y, at this frequency

$$y = w^H x_{new}$$

After these calculations are performed at every frequency, the inverse Fourier transform of the vector of outputs produces the beamformer output in the time domain. As the number of microphones increases, there are more degrees of freedom available to suppress feedback, improving performance, but the processing time increases as well, potentially causing buffering problems.

2 Methods

2.1 Experimental Setup

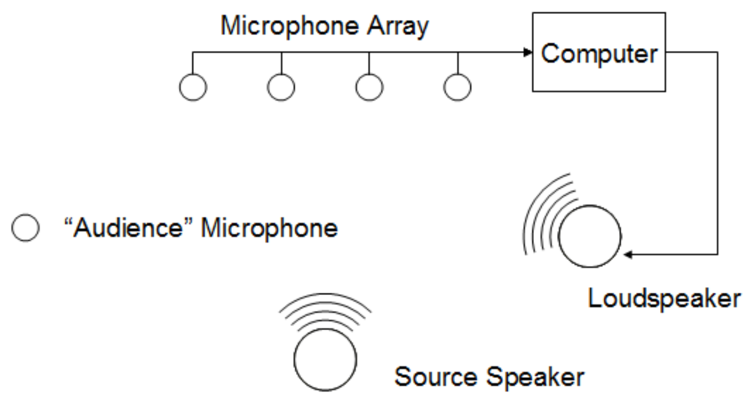


Figure 4: Experimental Setup

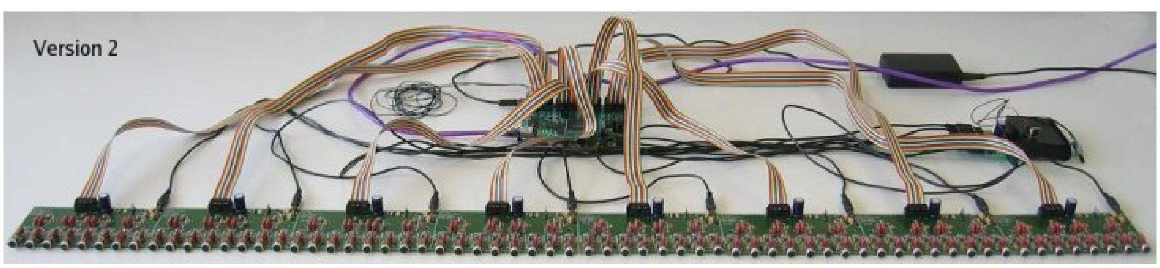


Figure 5: National Institute of Standards and Testing Mark-III Microphone Array

The system uses an NIST MK3 microphone array hooked up to a Linux machine that performs all the processing. This microphone array has 64 microphones spaced about 2 cm apart, though only four were used. The actually microphones that are used can be varied in software, which makes it much simpler to change the number of microphones used and the element spacing. After the processing the output is sent to a loudspeaker placed set at -45° relative to the array and approximately 1.5 m away. The source speaker connected to another computer was placed at 0° relative to the array and approximately 1.5 m away. This was used to play the sound clips necessary to test the system.

2.2 Processing

The software for this system was written in Java, because of its fast development time. Multiple threads are used to run separate processes simultaneously, allowing the program to run online. One thread reads in data from the microphone array at 44.1 kHz using code provided by the NIST, while another downsamples the data from only the desired microphones to 8.82 kHz and queues up windows of 1024 samples to be processed by another thread, which itself uses four threads. Once a window is completely processed, it is passed into one last thread that outputs the sound to the loudspeaker.

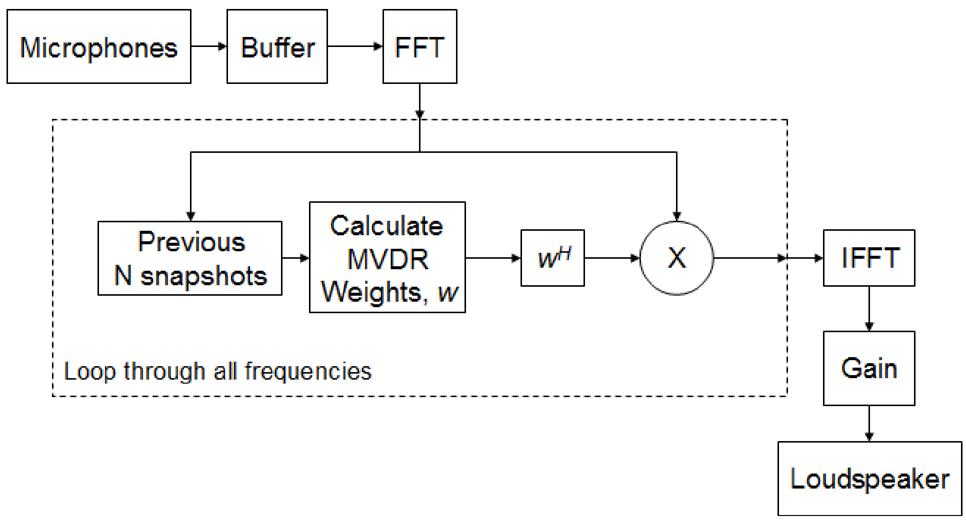


Figure 6: Processing Flowchart

In the processing thread, the Fast Fourier Transform of each microphone is calculated, and the values are arranged into the current snapshots. Then beamforming weight are calculated and applied to each snapshot, which requires time consuming matrix inverses and matrix multiplication. However, because the calculations for each frequency are independent of the one another, they can be split over four separate threads, greatly decreasing the processing time. Once these four threads are finished, the inverse Fast Fourier Transform of the frequency domain beamformer output produces the time domain version that is output to the loudspeaker. Flanagan's Java Scientific Library was used to perform all of the Fourier transforms and complex matrix operations.

2.3 Testing Methods

To test the system, a sound clip (the beginning of a segment on This American Life entitled Squirrel Cop) was played on the source speaker in front of the array. To measure the maximum stable gain for each method of processing, the sound clip was played on the source speaker and the system was run with a set gain on the loudspeaker output. This process was repeated with increasing loudspeaker gain until there was feedback. The maximum gain that didn't produce feedback was then used again to test the system, and data that was sent to the loudspeaker was saved on the computer as it was being played. This saved data was then replayed on the loudspeaker without the source speaker playing and then this sound was recorded with the audience microphone to measure the power. Playing the loudspeaker without the source speaker allowed the measurement of only the loudspeaker power, which was useful since the volume of source speaker was often greater. In a more real-world application, like an auditorium, the loudspeaker would be further away from the array so the loudspeaker could be greater than the source, but in these experiments the spacing was limited by the room size and equipment. This power was used to determine how successful the method of processing was at increasing gain. It was necessary to measure the actual power produced, rather than simply using the value of the gain, because the gain of beamforming alone isn't necessarily constant so the overall gain can't be properly set. Furthermore, the final metric of interest is how loudly a person would experience the loudspeaker's sound given the room's response and this provided a method of directly measuring that.

The beam pattern for the MVDR beamformer was experimentally created by measuring the power output to the loudspeaker with the sound source at different positions. At first the system was run normally with the source speaker playing a 200 Hz to 4 kHz chirp repeated at 1.7 Hz. After 20 seconds the beamforming weights were saved and used for the rest of the experiment. Then the loudspeaker was turned off and the source speaker was moved to different azimuths and the power of the beamformer output was recorded for each azimuth. This process was also repeated with the power measured at only one frequency to get a narrow-band beam pattern.

The beamforming weights from this experiment were also used to create a simulated beam pattern for all the frequencies on which the beamformer operated. This was done by calculating

$$s = 1$$
$$y = w^H v(\boldsymbol{\theta}) s$$
$$P = 10\log_{10}(|y|^2)$$

at each frequency for values of θ spanning -90° to 90°.

3 Results

Table 1: Maximum Stable Gain

Beamforming	Audience Power (dB)	Relative Power (dB)
None	-77.1	0
CBF	-67.9	9.2
MVDR	-60.2	16.9

As Table 1 shows the maximum stable gain that was achieved using the MVDR beamformer was 16.9 dB greater than that achieved without any beamforming and 7.7 dB greater than that achieved with conventional beamforming. One issue with this method of measuring maximum stable gain is that the recorded signal includes the distorted signal from the stable feedback, which will contribute to the power calculated. However, the recorded signals for each processing method will all include these distortions, so this somewhat compensates for the added power. As there is no widely accepted standard for quantifying what qualifies as feedback, there was some subjectivity in deciding when feedback occurred.

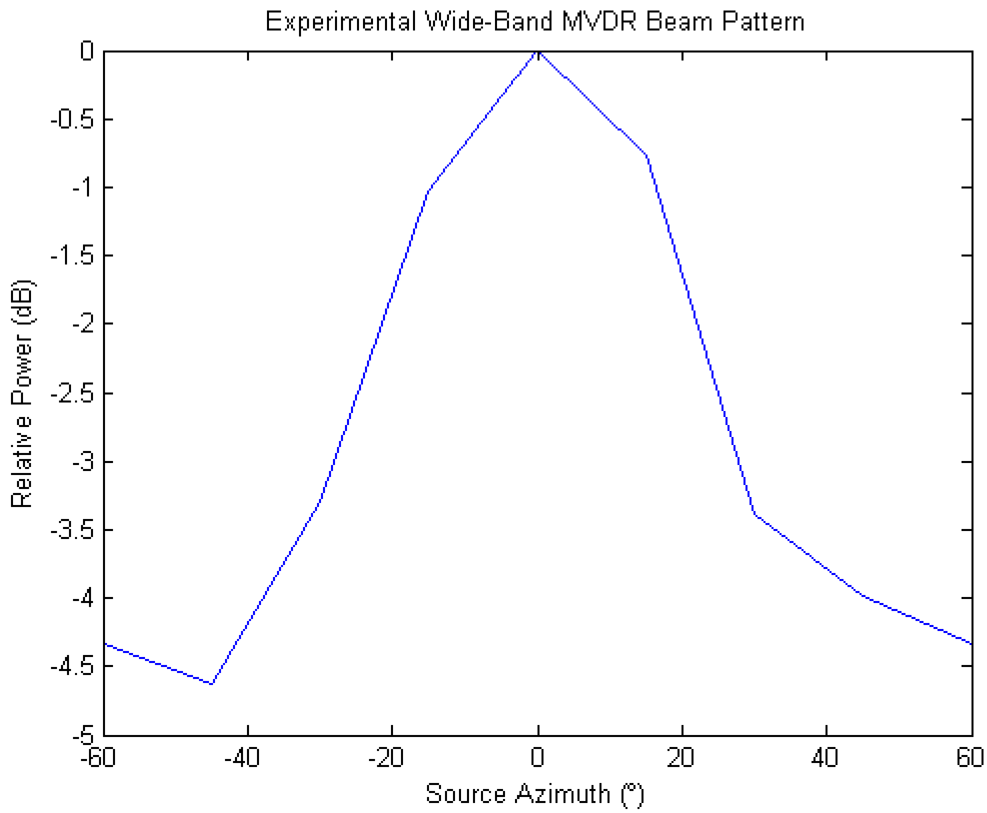


Figure 7: Experimental Wide-Band Beam Pattern

Figure 7 shows the total power output when the source was at various azimuths. The maximum power is at the direction of source and the minimum is at the direction of the loudspeaker, demonstrating that MVDR successfully steers a null toward the interference and preserves the targeted sound source. However, the minimum is not 16.9 dB below the maximum as one might expect from Table 1. One reason may be that different sound sources were played when measuring the maximum stable gain and the beam patterns. The frequencies of the chirp went up to 4 kHz, while those of the sound clip only went up to 2 kHz. When the power at the source direction and loudspeaker direction were experimentally measured for the conventional beamformer, the power at the loudspeaker direction was 3.8 dB less than at the source direction, so MVDR still performed marginally better.

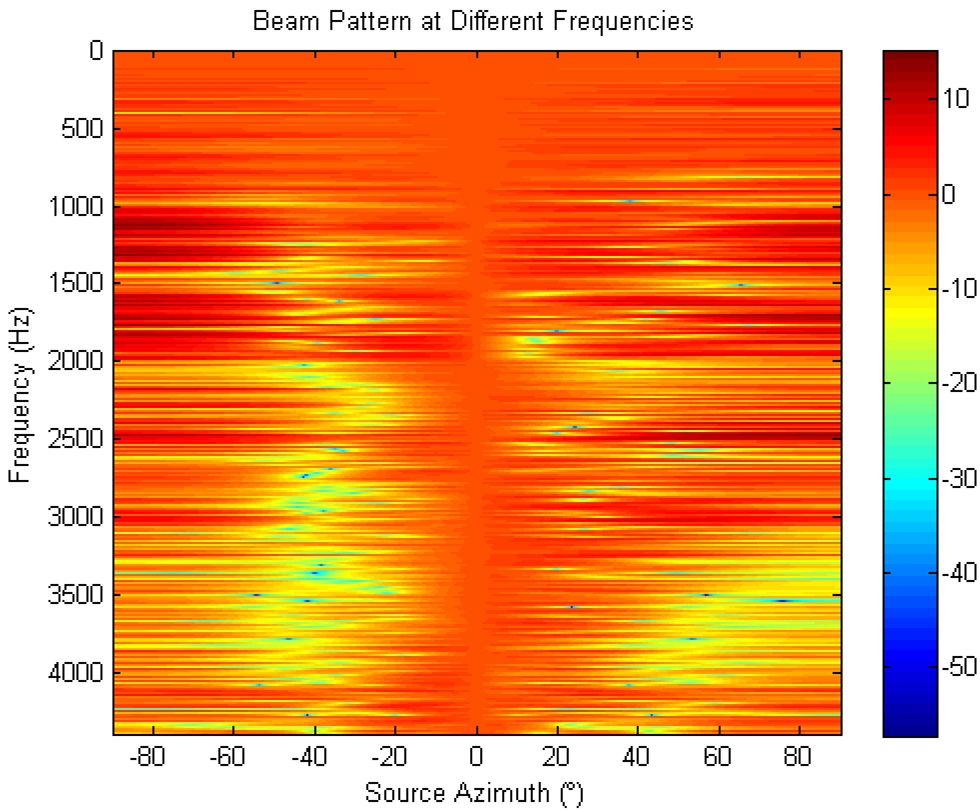


Figure 8: Beam Pattern at Different Frequencies for Experimental MVDR Beamforming Weights

Figure 8 shows the simulated beam pattern for all the frequencies given the same weights as those used for Figure 7. It clearly shows a line of nulls through the frequencies around -45°, the direction of the loudspeaker. There is also a line of 0 dB at the 0° azimuth because one of the constraints of MVDR is that the output be 1 at the steered direction. There are some areas with power greater than that at the source direction. These peaks occur because there was no interference at that direction for that frequency, so the gain in that direction could be increased without increasing the power from interference. These peaks would increase spatially-white noise in that direction, but not contribute to the feedback loop. The beam pattern is flatter in the lower frequencies so lower frequencies are more likely to cause feedback with an MVDR beamformer.

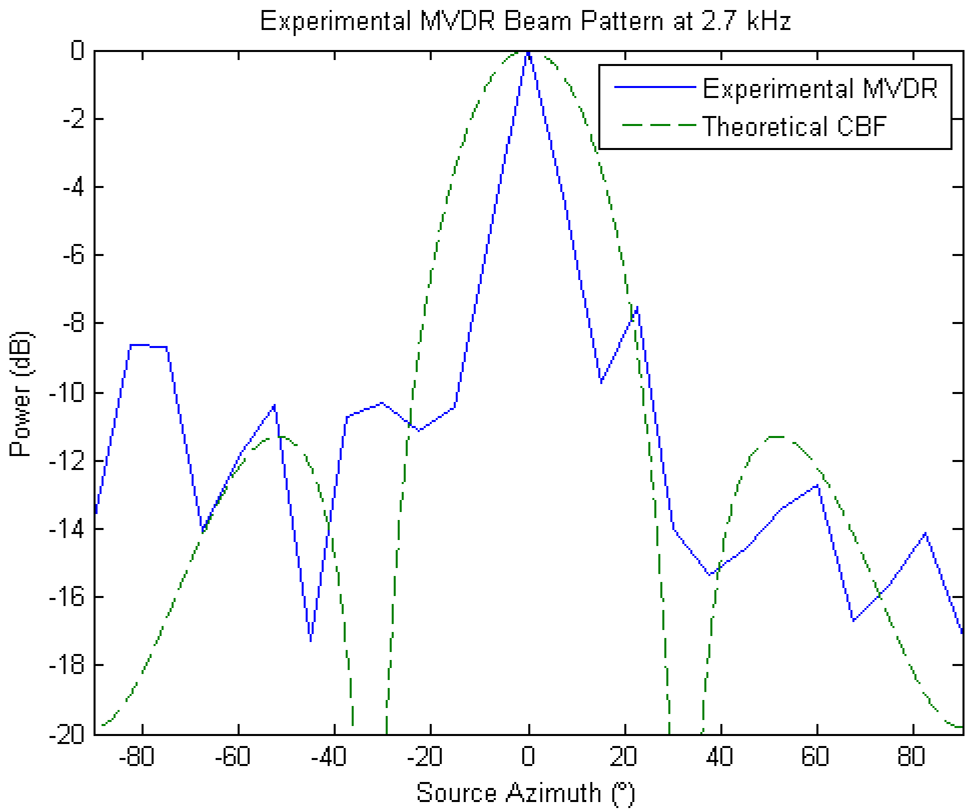


Figure 9: Experimental Narrow-Band Beam Pattern

Figure 9 shows the experimental beam pattern at 2.7 kHz. Like the wide-band beam pattern, it has a maximum at the source direction and minimum at the loudspeaker direction. The value of the minimum is more in line with the 16.9 dB value of the maximum stable gain, and is 5 dB lower than the theoretical power for conventional beamforming at this direction.

One interesting demonstration of adaptive aspect of MVDR was clapping while the system was running. Because clapping is loud and impulse-like, it contains wide band of frequencies making it more likely to cause a system to go unstable. When clapping while MVDR was running, one could hear the loudspeaker briefly get louder, much as it would during feedback, but then get softer as the system adapts and resumes normal operation.

When the interference and source are correlated, MVDR beamforming results in cancellation of the targeted sound source [9]. The processing delay is enough to prevent the loudspeaker output from being correlated with the input of the microphone. However, when tones are played for a long time, the loudspeaker and the target sound source both output similar sounds, so they become correlated. Because of this the target source is cancelled and the loudspeaker output becomes significantly quieter.

Because the system operates on the data in finite windows there are distortions at the edges of the output windows, and discontinuities between output windows. This occurs because multiplying discrete Fourier transforms is equivalent to circular convolution, rather than linear convolution, and the windows aren't initialized with previous data. In the loudspeaker output, this adds a sound similar to that of a train running over tracks because the distortions repeat with every window. Simple FIR filters avoid this issue using either an overlap and save, or overlap and add algorithm. A similar system is current in development for this system.

Conclusion

A four-element microphone array was used to test, in the real world, the effectiveness of MVDR beamforming for suppressing feedback in a microphone-loudspeaker system. This effectiveness was quantified by measuring the maximum stable gain and the beam pattern provided by MVDR and comparing them to those of conventional beamforming. The maximum stable gain achieved with MVDR was greater than that of conventional beamforming. In the MVDR beam patterns, the nulls can be clearly seen at the direction of the loudspeaker and have lower power than for conventional beamforming. From this it can be concluded that MVDR is an effective method of suppressing feedback.

Future work involves testing the performance with more microphones. The recent drastic improvements in the processing time of the system provided by the threading now allow for up to 12 microphones to be processed. Additional improvements in processing speed may also be possible to allow for even more microphones. Increasing the number of microphones increases the degrees of freedom with which MVDR can steer nulls towards interference, so the additional microphones will likely improve performance. In addition, work on preventing the distortions associated with operating on small windows is currently in progress.

Acknowledgements

I would like to thank my advisor, Dr. Jeffrey Krolik, for allowing me in his lab and guiding my research. I would like to thank the other members of the lab for all of their help, especially Jonathan Odom. Finally, I'd like to thank Duke University for the education and the opportunities it provided.

References

[1] J. Capon "High-resolution frequency-wavenumber spectrum analysis," *Proc. Of the IEEE*, vol. 57, no. 8, pp. 1408-1418, 1969

[2] J. Odom "Standoff Microphone Array with Feedback Suppression," Technical Report. Duke University, Durham, NC

[3] A. Spriet, G. Rombouts, M. Moonen, and J. Wouters, "Adaptive feedback cancellation in hearing aids," *Journal of the Franklin Institute*, vol. 343, no. 6, pp 545-573, 2006

[4] B.D. Van Veen and K.M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2 pp 4-24, Apr. 1988

[5] M.R. Schroeder "Improvement of Acoustic-Feedback Stability by Frequency Shifting," *The Journal of The Acoustical Society of America*, vol. 36, no. 9, pp 1718-1724, 1964
[6] R. Porayath, D.J. Mapes-Riordan, "Acoustic feedback elimination using adaptive notch filter algorithm" U.S. Patent 5,999,631, December 1999
[7] G. Rombouts, A. Spriet, and M. Moonen, "Generalized sidelobe canceller based combined acoustic feedback and noise cancellation," *Signal Processing*, vol. 88 no. 3, pp 571-581, 2008.
[8] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 1, pp 27-34, 1982 vol. 30, no. 3, pp 469-478, May 1982
[9] B. Widrow, K. Duvall, R. Gooch, and W. Newman, "Signal cancellation phenomena in adaptive antennas: Causes and cures," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 3, pp. 469-478, 1982
[10] D.H. Johnson and D. E. Dudgeon, "Array signal processing: concepts and Techniques,".P T R Prentice Hall, Englewood Cliffs, NJ, 1993.