Sound Localization using the MUSIC algorithm

Abha Pandey

Arizona State University

I. Abstract

Sound localization is a problem with smart devices such as Amazon Alexa, Google Speakers, etc. Beamforming is when one applies an adjusted phased beam in the direction of a signal. One of the most popular algorithm, Multiple Signal Classification or MUSIC, estimates the angle and direction of arrival with these sound systems. For this paper I will only focus on the MUSIC algorithm, due to its robust application to radar systems as well. In this paper, I will conduct a literature review on current academic and industry related research. I will also talk about experiments and analysis from simulations. ***Keywords*—MUSIC,beamforming, direction of arrival, microphone array, emitter location.**

II. Introduction

Multiple Signal Classification algorithm or MUSIC is widely used in the radar world to estimate direction of arrival. For this algorithm, one takes the covariance of noise and check if it is orthogonal to the signal subspace. If the dot product equals to zero, then that angle will be noted as a possible angle of arrival. This is checked with a peakfinding algorithm after taking the magnitude of the matrix. Beamforming is a process where an antenna can adjust the phase and amplitudes of the signal to create a beam towards the object of interest [2]. In radar, one often utilizes beamforming to be able to determine a direction of arrival for the received signal.

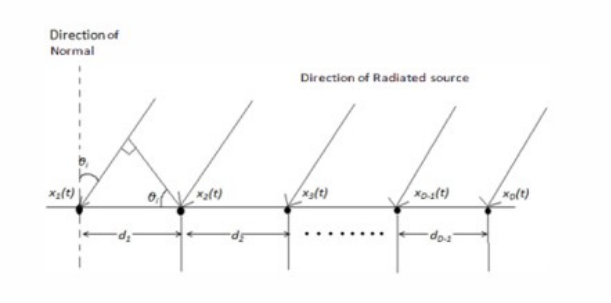


Fig. 1: Direction of Arrival concept

I modeled a sound system to create different beamforming scnearios. More specifically, I modeled the speaker system using a uniform linear array and a rectangular array. The uniform linear array or ULA, is a system where the array elements are placed in such away that they satisfy the Nyquist sampling criteria[5]. To keep the system simple due to design constraints, the ULA and rectangular arrays are optimal systems for MUSIC simulations.

III. Application

In terms of audio signal processing, the MUSIC algorithm and beamforming are used in popular smart speaker systems such as Amazon Alexa, Apple Homepod, and Google Home. Unfortuantely, not much details can be found about what algorithms they use to estimate direction of arrival. However, Amazon does have a white paper on how their beam forming selection works. Zhang, Kristjansoon, and Hilmes at Amazon talk about using signal to interference ratio, or SIR[10]. They look at the instantaneous SIR and feed that into their learning algorithm to be able to determine direction of arrival [10]. In a practical real time scenario, this would be an approach one would use for it as well.

Their inputs are audio signals from Amazon Alexa. The filter banks works as follows: ”The multi-microphone signals are first processed by a AEC block to remove the echoes and then processed by an FBF block. The FBF block employs a subband-based filter-and-sum structure and its weights are specially designed so that the formed beams are able to cover all the look directions of interest”[10].

The authors use adaptive filtering to be able to do clutter suppression, assuming the filtered echoes are unwanted; then passed through a filter bank to prevent aliasing for direction of arrival estimation.

Zhang, Kristjansoon, and Hilmes also apply this beamforming and classify it as a ”front end audio processing”[10]. They take it a step further and use word recognition to determine what words were spoken by the user.

IV. Literature Review

For this project, I consulted a few RF and Audio signal processing papers. Gupta and Kar[5], well explained the basics of the MUSIC algorithm and gave a foundation on how to approach the analysis.

Zhang, Kristjansoon and Hilmes wrote an easy to follow paper highlighted in the applications section of this paper. However, seeing their results on using signal to interference ratio for their estimations, about 39 percent overall accuracy, is concerning to apply in real time. It could be the low accuracy does come from non ideal

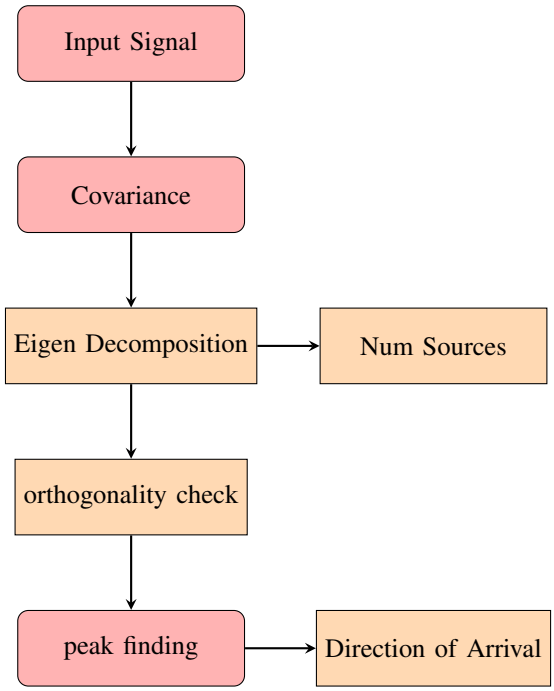
conditions from testing such as environmental factors, noise, etc. However, another design approach I would look into instead of the SIR approach, is look at the Signal to clutter+noise ratio and the clutter to noise ratio. I believe this would improve design for the audio signals as well. Clutter refers to unwanted echos and I strongly believe using environmental 'clutter' in the room would give better results[10].

Although a couple of papers focused on the ESPRIT algorithm, they were still useful on gameplanning the analysis for this project. Muhamed and Rappaport[7] did a very well job of going step by step about the multipath problem for DOA estimations. They looked at how multipath affected performance and thoroughly explained their experimental setup. The biggest takeaway from this paper is the root means square analysis method for DOA estimation is another accurate approach for analysis and external factors such as multipath can present another negative affect on DOA estimations. Another thing to note is this paper looked at RF propgation; however, most of the methodologies is applicable to audio signal processing and speech encoding.

Liu, Lv, Miao, and Shang[6] talk about a high resolution method for the MUSIC algorithm. First I would like to point out they used excellent visual aids to explain the concept of DOA estimation and a very organized block diagram of the MUSIC algorithm. They explained their methodology for applying the two dimensional MUSIC algorithm really well. I used their methodology for my own analysis, particularly the SNR analysis. It was not clear though what types of signals did they use for their analysis and their experimentation. It also was not clear on how they approached calculating the number of samples for their complexity analysis. Though, most papers I found did not go through a complexity or performance analysis for their MUSIC algorithms experimentation.

Guo, Mao, Li, Wang and Wang[4], actually explain how they came up with their computational complexity in rather simple terms. This proved to be helpful to do the complexity analysis for the algorithm. They go into full detail on polarization and its measurements in relation to MUSIC. However, for my case, this part was not in particularly useful as I am working with ideal cases for the audio signals. Another useful thing they looked at was the resolution impact. My professor, Dr. Andreas Spanias, also noted their computational complexity in order to fully take advantage of the resolution benefits of MUSIC. They also did an SNR comparison study on polarization impact on the Polarization MUSIC results, Rank Reduction results and the MUSIC results. From an RF perspective, the polarization aspect of this study and the comparison study would be extremely useful for future research and possibly help more for problems like emitter location. R.O.Schmidt[8], also expands upon the concept of polarization applications to signal of arrival and would be another paper worth looking into for polarization and its application to the MUSIC algorithm.

V. Algorithm Block Diagram



Direction of Arrival block diagram based on algorithm outlined by R.O.Schmidt[8] and Sekiya and Kyobashi[9].

VI. Simulation Methods

I performed simulations using MATLAB's beamforming and MUSIC estimation functions. I propagated the signals with the speed of sound, 340 m/s through an omnidirectional antenna. I created a 3, 5, and 10 source microphone array spaced between 0.01 to 1 cm apart. I varied the noise power with all of these simulations. The input parameters are audio signals consisting of speech, songs, and phone ringtones. All of these audio signals were resampled at 8 kHz and 16kHz respectively to simulate narrowband and wideband data. The beamformer array sampled frequencies between 50 MHz to 500 MHz and operated at 150MHz. The time duration for these signals were set to about 3-10 seconds.

The second part of the simulation involves different angles with varying SNRs. For these simulations, I varied the SNR by setting the noise power at 10^{-3}, 10^{-6}, and 10^{-10} respectively. The SNR was added in with the beamforming signal and again at the MUSIC estimated functions. The beamformer created sources sampling from -90 to 90 degrees in elevation and -90 to 90 degrees in azimuth. In order to evaluate the different angles of arrival, I used the 2-dimensional MUSIC estimator in

MATLAB. MATLAB has a function called collect plane wave where each column of the matrix corresponds with a signal. In the collect plane wave function, one would also need to specify the number of sources simulating as well[1].

The amplitudes of the signals were varied as well. Certain speech signals were given a higher amplitude to be able to hear the speech in the final beamformed signal. This was done because certain audio signals, such as theme songs and ringtones often ”overpower” the speech signal; therefore, speech signals need to be scaled to a higher amplitude to distinguish these sources. This was also verified by doing a playback of the 10 channel generated signal. I wanted to see if I could distinguish these sources by ear. I noticed that certain speech signals such as laughter was the dominating speech signal. Speech signals such as the cleanspeech signal from EEE 607 were often the lowest amplitude speech signal. I could not hear that signal unless if I lowered the amplitude of the r2d2 signal and increased the amplitude of the cleanspeech signal.

VII. Results

A. Complexity and Performance Analysis

The angles presented in the simulations are sampled at $1°$. For the computational complexity, I used Guo et al.'s formulas to do my computational complexity. The ranges for the azimuth and elevation were from $-90°$ to $90°$. The 2-D space for MUSIC will consist of 181x181 = 32761 samples. If the values of azimuth and elevation ranges decrease, the computational complexity will decrease as such[6]. As speculated by Dr. Andreas Spanias, this may have been done in order to utilize resolution and improve upon it. For the complexity of eigenvector decompositions in MUSIC, Golab and van Loan[3] mention the costliest multiplication computations are $25*n^3$.

For the performance analysis, I did a timing performance for the different signal sources. For a 10 source MUSIC algorithm, primarily with speech and ringtones, the algorithm runs at 16.025391 seconds. For a 10 source signal with two songs, the timing comes out to: 6.435001 seconds. The table below goes over timing analysis for a 3 source and 5 source.

3 source	5 source	10 source
3.23s	2.61 s	4.37s
5.92s	4.53 s	6.59s

TABLE I: CPU timing of MUSIC for 1D(first row) and 2D(second row)

B. Simulation Results

Another significant impact on source finding is element spacing. Depending on how far I set up the rectangular array elements, the number of sources detected changes. This also held true for the ULA array evaluation.

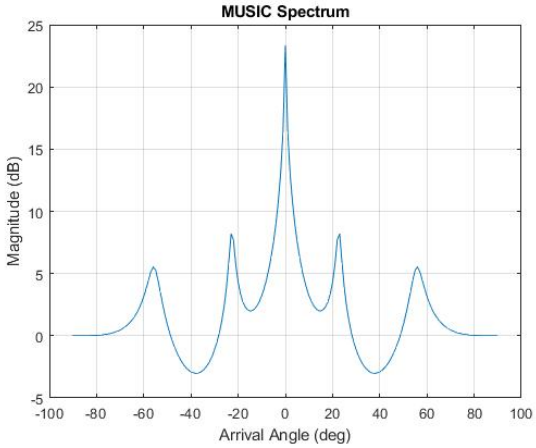


Fig. 2: 10 sources spaced 0.01cm apart

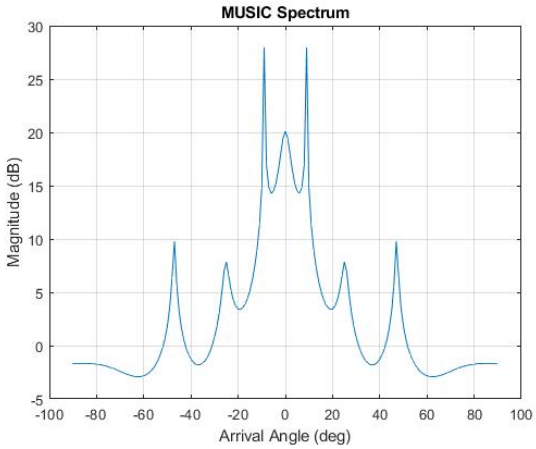


Fig. 3: 10 sources spaced 0.1cm apart

One thing to note, in these figures, I placed two signals at $0°$ azimuth which is why only 6-8 sources are shown on these graphs. In the 0.01 cm element spacing, only 6 sources were identified. In the 0.1 and 1 cm element spacing, only 8 sources were identified. This could be because some sources were placed around $1°$ to $2°$ apart and the algorithm thought only one source are in those areas.

The simulations were able to successfully identify the different sources in 2D. In order to identify the sources, I used two approaches, a 3D graph of the estimator to look at the peaks and contour graphs to confirm location of the peaks. I also did an azimuth estimation of the direction

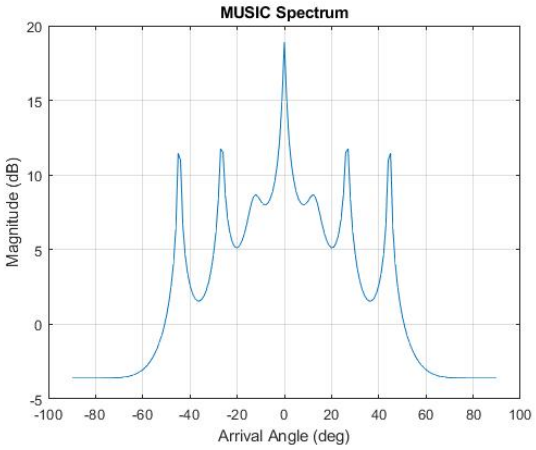


Fig. 4: 10 sources spaced 1cm apart

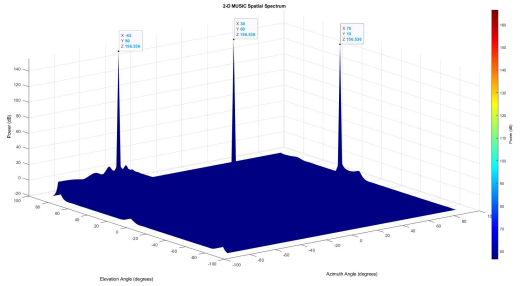


Fig. 5: 3 Sources MUSIC narrowband

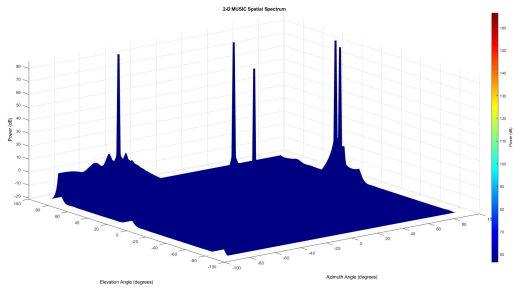


Fig. 6: 5 Sources MUSIC narrowband

of arrivals. I evaluated the peaks based on the power densities as highlighted by R.O.Schmidt[8]. One thing to note is depending on how the amplitude is scaled, the power at certain estimator times would the same all throughout. My initial findings in figures 2 and 3 show the same power all throughout the peaks despite those peaks being labeled as sources of interest. One hypothesis is the amplitudes are not scaled to be higher than the noise floor, which is why the power is shown the same across.

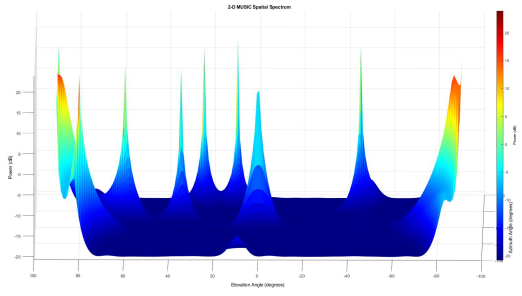


Fig. 7: 10 Sources MUSIC narrowband

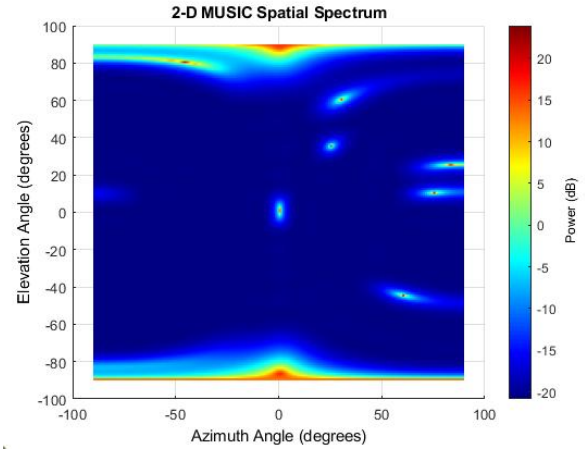


Fig. 8: 10 Sources MUSIC narrowband Contour version

VIII. Conclusion

This paper highlights simulations of microphone beam forming and the application of the MUSIC algorithm to determine sources of arrival. Using different length speech and audio signals alongside other parameters, I have demonstrated that sound localization works well with the MUSIC algorithm. However, since this has been demonstrated in an ideal environment, in an actual experiment environment more parameters would affect accuracy in estimating direction of arrival.

Soon, I plan to expand more research into this topic and look into other signal processing techniques, such as time-frequency techniques to improve direction of arrival and optimization to speed up calculations of the MUSIC algorithm. Another potential topic of interest is to use images as a possibility for direction of arrival estimations and analysis.

IX. Acknowledgments

The author would like to thank her mentors at Naval Surface Warfare Center Dahlgren - Mr. Randy Strock, Mr. Christopher Jahn, Mr. Aaron Cox, Mr. Will Webekind, Mr. Bernie Ulfers, Dr. Wes Hall, Dr. Terry Foreman

and Dr. Marc Salas for their encouragement to pursue this project. The author would also like to thank Mr. Gregory Carter and Dr. Stephane Valladier at Naval Air Weapons Station China Lake for their technical expertise and encouragement for this project.

The author would also like to thank Dr. Antonia Papandreou-Suppapola and Ms. Cindy Fuentes-Munoz for their kind words and for being amazing role models.

X. References

References

[1] Pallavi J Agrawal and Madhu Shandilya. "MATLAB Simulation of Subspace based High Resolution Direction of Arrival Estimation Algorithm". In: *International Journal of Computer Applications* 130 (2015), pp. 22–27.

[2] "Front Matter". In: *Academic Press Library in Signal Processing, Volume 7*. Ed. by Rama Chellappa and Sergios Theodoridis. Academic Press, 2018, p. 403. ISBN: 978-0-12-811887-0. DOI: https://doi.org/10.1016/B978-0-12-811887-0.09991-0. URL: https://www.sciencedirect.com/science/article/pii/B9780128118870099910.

[3] G.H. Golub and C.F. Van Loan. *Matrix Computations*. 2nd. Baltimore: Johns Hopkins University Press, 1989.

[4] Ran Guo et al. "A Fast DOA Estimation Algorithm Based on Polarization MUSIC". In: *Radioengineering* 24 (2015), pp. 214–225.

[5] Pooja Gupta and Sambit Prasad Kar. "MUSIC and improved MUSIC algorithm to estimate direction of arrival". In: *2015 International Conference on Communications and Signal Processing (ICCSP)* (2015), pp. 0757–0761.

[6] Chao Liu et al. "Research on High Resolution Algorithm of Sound Source Localization Based on Microphone Array". In: *2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP)* (2019), pp. 1–6.

[7] Rias Muhamed and Theodore Ted S. Rappaport. "Comparison of conventional subspace based DOA estimation algorithms with those employing property-restoral techniques: simulation and measurements". In: *Proceedings of ICUPC - 5th International Conference on Universal Personal Communications* 2 (1996), 1004–1008 vol.2.

[8] R. Schmidt. "Multiple emitter location and signal parameter estimation". In: *IEEE Transactions on Antennas and Propagation* 34.3 (1986), pp. 276–280. DOI: 10.1109/TAP.1986.1143830.

[9] Hidetomo Tanaka and Tetsunori Kobayashi. *Estimating Positions Of Multiple Adjacent Speakers Based On Music Spectra Correlation Using A Microphone Array*.

[10] Xianxian Zhang, Trausti Kristjansson, and Philip Hilmes. "SIR Beam Selector for Amazon Echo Devices Audio Front-End". In: *2019 IEEE International Workshop on Signal Processing Systems (SiPS)*. 2019, pp. 302–306. DOI: 10.1109/SiPS47522.2019.9020406.