

**3D SOUND-SOURCE LOCALIZATION USING TRIANGULATION-BASED METHODS**

by

Alice Lam

B.A.Sc, The University of British Columbia, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Mechanical Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

November 2017

© Alice Lam, 2017

## **Abstract**

The localization of sound sources in a reverberant environment, such as a classroom or industrial workspace, is an essential first step toward noise control in these spaces. Many sound source localization techniques have been developed for use with microphone arrays. A common characteristic of these techniques is that they are able to provide the direction from which the sound is coming, but not the range (i.e. the distance between the source and receiver).

This thesis presents two triangulation-based methods for localizing sound sources in 3D space, including range, using a small hemispherical microphone array. Practical issues with the hemispherical array, such as source resolution and operating frequency limitations, are discussed. The first method - direct triangulation - involves taking multiple sound field measurements at different locations in the room, and then using the combined output of all receivers to triangulate the source. Direct triangulation is conceptually simple and requires no a priori knowledge of the surrounding environment, but proves cumbersome as multiple array measurements are required - this also limits its application to steady-state noise sources. The second method - image source triangulation - requires only one measurement, instead taking into account the early specular reflections from the walls of the room to create "image receivers" from which the source location can be triangulated. Image source triangulation has the advantage of only requiring one measurement and may be more suited to small spaces such as meeting rooms. However, it relies on having accurate pre-knowledge of the room geometry in relation to the microphones. Both triangulation methods are evaluated using simulations and physical in-room measurements, and are shown to be able to localize simple monopole sources in reverberant rooms.

## **Lay Summary**

This thesis presents two methods for locating sound sources inside a room using a microphone array, both being capable of determining not only the direction from which the sound is coming, but also how far the source is from the listener. The first – direct triangulation – involves moving the array to various locations in the room and taking measurements at each location. Direct triangulation is conceptually simple, but the requirement of multiple array measurements becomes cumbersome in more complex situations. The second – image source triangulation – only requires one measurement, instead incorporating information from sound reflected from the room walls to triangulate sound source(s). Image source triangulation only requires one array measurement, but relies on having an accurate knowledge of the array position relative to the walls of the room. Both methods were able to locate simple sound sources in simulations and in physical in-room measurements.

## Preface

The experiment design and data analysis in this thesis are the original work of the author.

The hardware used for the experiments was originally designed and built by Dr. Hedayat Alghassi, with later contributions from Chris Bibby, Mona Nematifar and Glenn Jolly. None of the text of the thesis is taken from previously published or collaborative articles.

## Table of Contents

<b>Abstract .....</b>	<b>ii</b>
<b>Lay Summary .....</b>	<b>iii</b>
<b>Preface .....</b>	<b>iv</b>
<b>Table of Contents .....</b>	<b>v</b>
<b>List of Tables .....</b>	<b>ix</b>
<b>List of Figures.....</b>	<b>x</b>
<b>List of Symbols .....</b>	<b>xvi</b>
<b>Acknowledgements.....</b>	<b>xviii</b>
<b>Chapter 1: Introduction.....</b>	<b>1</b>
1.1 Background and Objective .....	1
1.1.1 Direct Triangulation: Range-finding using Multiple Bearing-only Measurements .....	2
1.1.2 Image Source Triangulation: Using Reverberation to Improve Localization .....	3
1.2 Scope of Thesis.....	4
1.3 Structure of Thesis .....	4
<b>Chapter 2: Microphone Array and Processing Background.....</b>	<b>6</b>
2.1 Construction of Hemispherical Array.....	6
2.2 Free-field Sound Propagation.....	8
2.3 Delay-and-Sum Beamforming .....	9
2.4 Array Properties.....	11
2.4.1 Definition of Azimuth and Elevation Angles .....	11
2.4.2 Beam Width.....	12
2.4.3 Operating Frequency.....	12

2.4.4	Source Resolution .....	15
2.4.5	Beam Width Variation .....	16
<b>Chapter 3: Direct Triangulation - Background .....</b>		<b>18</b>
3.1	Related Work.....	18
3.2	Implementation.....	19
3.3	Practical Considerations.....	21
3.3.1	Multiplicative vs Additive Triangulation.....	21
3.3.2	Spacing of Grid Points .....	22
3.3.3	Receiver Positions.....	22
3.3.4	Averaging Period .....	23
<b>Chapter 4: Direct Triangulation - Experiments .....</b>		<b>25</b>
4.1	Progression of Experiments .....	25
4.2	Experimental Setup.....	25
4.3	Presentation of Results.....	26
4.3.1	Accuracy.....	27
4.3.2	Resolution.....	27
4.4	Free Field Experiments .....	29
4.4.1	Test Environment.....	29
4.4.2	Results .....	31
4.5	Reverberant Room Experiments.....	34
4.5.1	Test Environment.....	34
4.5.2	Receiver Configuration .....	37
4.5.3	Range Variation .....	42

4.5.4 Lateral Position Variation .....	44
4.6 Multiple Sources.....	47
4.6.1 Simulation – Two Sources in Free Field.....	47
4.6.2 Measurement – Image Sources in Reverberant Room.....	52
4.7 Discussion .....	55
4.7.1 Sources of Uncertainty.....	56
<b>Chapter 5: Image Source Triangulation - Background.....</b>	<b>58</b>
5.1 Related Work.....	58
5.2 Implementation.....	59
5.2.1 Equivalence between ISB and Triangulation .....	60
5.3 Practical Considerations.....	62
5.3.1 Multiplicative vs Additive Image Triangulation .....	63
5.3.2 Image Order Estimation .....	64
5.3.3 Edge Artifacts .....	66
5.3.4 Inline Sources .....	68
<b>Chapter 6: Image Source Triangulation - Experiments .....</b>	<b>70</b>
6.1 Progression of Experiments .....	70
6.2 Presentation of Results.....	70
6.3 Experimental Setup.....	70
6.3.1 Source Parameters.....	70
6.3.2 Test Environments .....	71
6.3.3 Background Noise.....	75
6.4 Baseline Case .....	75

6.5	Corner Case .....	78
6.6	Background Noise .....	80
6.7	Inline Case.....	81
6.8	Geometry Variation .....	82
6.9	Discussion .....	83
6.9.1	Sources of Uncertainty.....	84
<b>Chapter 7:</b>	<b>Conclusion .....</b>	<b>85</b>
7.1	Summary .....	85
7.2	Comparison Between Triangulation Methods.....	86
7.3	Future Work .....	87
<b>References.....</b>		<b>89</b>
<b>Appendices .....</b>		<b>93</b>
	Appendix A Microphone Construction and Calibration .....	93
A.1	Detailed Microphone Placement Geometry .....	93
A.2	Microphone Frequency Responses by Model .....	94
A.3	Microphone Circuit Diagram .....	95
A.4	Array Calibration .....	96
	Appendix B Comparison between Array Shapes.....	100

## List of Tables

Table 1. Parameters for direct triangulation experiments .....	26
Table 2. Source and receiver coordinates for free field cases .....	30
Table 3. Error and resolution comparison between free field measured and simulation cases .....	33
Table 4. Error and resolution comparison between all receiver configurations .....	41
Table 5. Error and resolution comparison between Y-offset cases .....	44
Table 6. Error and resolution comparison between X-offset cases .....	46
Table 7. Summary of source conditions for two-source free field simulations.....	49
Table 8. Qualitative summary of results for two-source free field simulations .....	50
Table 9. Comparison between predicted and actual outcomes for source resolution .....	54
Table 10. Comparison between predicted and actual outcomes for relative source level .....	55
Table 11. Experiment parameters for ISB experiments .....	71
Table 12. Error and resolution comparison between baseline shoebox cases .....	78
Table 13. Error and resolution comparison between baseline and corner cases .....	79
Table 14. Error and resolution comparison between baseline and background noise cases .....	80
Table 15. Error and resolution comparison between baseline and inline cases .....	82
Table 16. Error and resolution comparison between baseline and geometry variant cases .....	83
Table 17. Differences between direct and image source triangulation methods .....	86
Table 18. Microphone positions and types by number .....	93

## List of Figures

Figure 1. Distinction between (a) direction-only and (b) range-inclusive localization of a sound source in 2D .....	1
Figure 2. (a) Triangulation technique used by surveyors to locate a point of interest from a known baseline, (b) direct triangulation method to locate a region of interest using two known array positions .....	2
Figure 3. Image source model to determine the shortest reflection path between a source and receiver .....	3
Figure 4. Layout of hemispherical microphone array (a) front view, (b) side view. The center microphone is located at (0,0,0).....	7
Figure 5. Microphone array in test room supported by tripods .....	7
Figure 6. 2D delay-and-sum beamforming output for a simulated 750 Hz pure tone source at (a) 0.3 m and (b) 1.5m distance from receiver.....	11
Figure 7. Definition of azimuth and elevation angles relative to hemispherical array .....	11
Figure 8. Comparison of simulated beamforming output (spherical coordinates) of a pure tone signal at (a) 800 Hz (b) 1150 Hz (c) 2500 Hz. ....	14
Figure 9. Simulated beamforming output of a white noise signal filtered to the [1000 2000] Hz octave band .....	15
Figure 10. Simulated beamforming output for two sources (white noise filtered to [1000 2000] Hz octave band) with (a) 25.7 degree separation (Rayleigh criterion), (b) 17.1 degree separation (2 sources unresolved), (c) 18 degree separation (2 sources minimally resolved) .....	16

Figure 11. Polar plot of angular beam width (FWHM) as a function of azimuthal arrival angle (deg) for hemispherical and spherical arrays.....	17
Figure 12. Comparison between (a) additive and (b) multiplicative direct triangulation .....	22
Figure 13. (a) “Phantom Source” ambiguity with two receiver positions is eliminated in (b) when a third receiver position is added .....	23
Figure 14. Delay-and-sum beamforming outputs at 3 arbitrary start points using (top row) 0.01 s time average and (bottom row) 0.15 s time average. ....	24
Figure 15. Comparison between (a) ineffective point cloud representation and (b) isosurface representation of 3D beamforming output .....	28
Figure 16. Example isosurface view of results with three intersecting Cartesian planes .....	28
Figure 17. Detailed view of X-Y plane blob, with (a) half-maximum contour and (b) equivalent ellipse with major and minor axes .....	29
Figure 18. Free field source-receiver configuration, isometric view .....	30
Figure 19. (a) Microphone array and (b) monopole source in anechoic chamber .....	30
Figure 20. Graphical results for free field simulation, “far” case.....	31
Figure 21. Graphical results for free field simulation, “near” case. ....	32
Figure 22. Graphical results for free field measurement, “far” case.....	32
Figure 23. Graphical results for free field measurement, “near” case .....	33
Figure 24. Illustration of distance skew effect in triangulation of beams .....	34
Figure 25. Plan view of reverberant test room .....	35
Figure 26. Test room as seen from source location .....	36
Figure 27. Test room as seen from receiver location .....	36
Figure 28. Plan view of test room during receiver configuration experiment.....	37

Figure 29. Perspective view of A, B, C, D receiver locations relative to source and room.....	38
Figure 30. Graphical results for ABC receiver configuration. ....	38
Figure 31. Graphical results for ABD receiver configuration. ....	39
Figure 32. Graphical results for ACD receiver configuration .....	39
Figure 33. Graphical results for BCD receiver configuration. ....	40
Figure 34. Graphical results for ABCD receiver configuration.....	40
Figure 35. Plan view of reverberant test room during distance variation experiment .....	42
Figure 36. Graphical results for y - 1.5m source position (ABC receiver configuration). .....	43
Figure 37. Graphical results for y + 1.5m source position (ABC receiver configuration).....	43
Figure 38. Plan view of test room during horizontal variation experiment .....	45
Figure 39. Graphical results for source position L (ABC receiver configuration). .....	45
Figure 40. Graphical results for source position R (ABC receiver configuration). ....	46
Figure 41. Free field two-source source-receiver configuration (a) isometric, (b) top, (c) side views .....	48
Figure 42. Graphical results for two-source Case 4 (“good” localization: 2 blobs resolved and 2 peaks located) .....	50
Figure 43. Graphical results for two-source Case 1 (2 blobs unresolved, peaks not successfully located) .....	51
Figure 44. Graphical results for two-source Case 2 (2 blobs unresolved, but 2 peaks located)....	51
Figure 45. Detailed view of image sources in y+1.5m case (original Figure 36) in (a) X-Z plane, (b) Y-Z plane .....	53
Figure 46. Detailed view of image source in baseline case (original Figure 29) in (a) X-Z plane, (b) Y-Z plane .....	54

Figure 47. (a) Single focus point used in CBF versus (b) focus point and images used in ISB ....	59
Figure 48. (a) Regular ISB interpretation in 2D, (b) equivalent image receiver triangulation for a single focus point $f$ .....	61
Figure 49. Example ISB case to demonstrate triangulation equivalence, with (a) points labelled and (b) vectors labelled .....	61
Figure 50. Image source triangulation with a single image receiver. ....	62
Figure 51. Comparison between additive and multiplicative ISB output .....	64
Figure 52. (a) All image sources up to order 2, (b) $r_{mix}$ extending from receiver to propagation boundary .....	65
Figure 53. ISB output from 0 <sup>th</sup> to 3 <sup>rd</sup> order, no windowing.....	66
Figure 54. Image source triangulation with a single image receiver, edge artifact regions shaded .....	67
Figure 55. Comparison between non-windowed and windowed ISB output (data from “baseline ISB” case – Section 6.3).....	68
Figure 56. (a) Offset and (b) inline image source triangulation .....	69
Figure 57. Top view of shoebox test room with source-receiver locations .....	73
Figure 58. Dodecahedral source in shoebox room, “S1” source position.....	73
Figure 59. Ceiling detail of shoebox room.....	74
Figure 60. Top view of “geometry variation” test room with source-receiver locations.....	74
Figure 61. Octave band noise profile for “background noise” IST case.....	75
Figure 62. Graphical results for ISM simulation, “S1” source position, 1 <sup>st</sup> order IST processing. ....	76

Figure 63. Graphical results for ISM simulation, “S1” source position, 2 <sup>nd</sup> order IST processing.	76
Figure 64. Graphical results for shoebox room measurement, “S1” source position, 1 <sup>st</sup> order IST processing .....	77
Figure 65. Graphical results for shoebox room measurement, “S1” source position, 2 <sup>nd</sup> order IST processing .....	77
Figure 66. Graphical results for shoebox room measurement, “S2” source position, 1 <sup>st</sup> order IST processing .....	79
Figure 67. Graphical results for shoebox room measurement, “S1” source position with background noise, 1 <sup>st</sup> order IST processing.....	80
Figure 68. Graphical results for shoebox room measurement, “S3” source position, 1 <sup>st</sup> order IST processing .....	81
Figure 69. Graphical results for geometry variant measurement, “S1” equivalent source position, 1 <sup>st</sup> order IST processing.....	82
Figure 70. Numbered layout of array microphones, front view .....	93
Figure 71. (Type 1) Typical frequency response for BGO-15L27-C1033 microphone [7].....	94
Figure 72. (Type 2) Typical frequency response for WM61-A microphone [39] .....	94
Figure 73. (Type 3) Typical frequency response for POM-3535L-3-R microphone [40] .....	94
Figure 74. Microphone preamplifier circuit diagram [41] .....	95
Figure 75. (a) Schematic of hemisphere depth layers, (b) Photo of reference microphone (circled in red) at depth 2 .....	96
Figure 76. Sensitivities (V/Pa) for microphones 0 to 8.....	98
Figure 77. Sensitivities (V/Pa) for microphones 9 to 17.....	98

Figure 78. Sensitivities (V/Pa) for microphones 18 to 26.....	99
Figure 79. (a) Spherical, (b) hemispherical and (c) planar microphone array configurations for comparison.....	100
Figure 80. Comparison matrix of beamforming output for spherical, hemispherical and planar arrays at 0, 43.2 and 90 degrees.....	102

## List of Symbols

$a$	Source-receiver path length
$a_{maj}$	Ellipse major axis
$a_{min}$	Ellipse minor axis
$\alpha$	Absorption coefficient
$b$	Source-image path length
$\beta$	Image order
$c$	Speed of sound
$d$	Source-receiver (centroid) distance
$D$	Array aperture
$\Delta$	Imaging grid spacing
$\Delta L_p$	Sound pressure level attenuation between direct and image source
$\Delta t$	Time delay
$e$	Total localization error
$e_{disc}$	Discretization error
$g$	Input signal for free field propagation model
$L$	Average ellipse axis
$\lambda$	Wavelength
$M$	Number of array microphones
$N$	Number of triangulation receivers (direct triangulation), Number of image receivers (image source triangulation)
$r$	Range

$\mathbf{r}_f$	Beamforming focus location vector
$\mathbf{r}_i$	Image source location vector
$\mathbf{r}_{loc}$	True (known) source vector
$\mathbf{r}_m, \mathbf{r}_j, \mathbf{r}_{ij}, \mathbf{r}_r$	Receiver/microphone location vector
$r_{mix}$	Mixing radius
$\mathbf{r}_{pk}$	Found source vector
$\mathbf{r}_s$	Source location vector
$p$	Pressure
$\phi$	Azimuth angle
$s$	Microphone signal
$t$	Time
$t_{mix}$	Mixing time
$T$	Number of time points for RMS averaging
$\theta$	Elevation angle
$\theta_M$	Minimum separation angle between two sources
$\theta_R$	Rayleigh Criterion separation angle between two sources
$w(x)$	Tukey window function
$z_{DAS}$	Conventional (non-triangulated) delay-and-sum beamforming output
$z_{ISB}$	Image source beamforming output
$z_{DT}$	Direct triangulation beamforming output

## Acknowledgements

I thank my supervisor, the late Professor Murray Hodgson. I am grateful to have had his mentorship and support over the past three years. I am better for having known him, and I will miss him. I thank Professors Vincent Valeau and Chris Waltham for their feedback and support throughout the course of this project, and for jointly shouldering the role of supervisor after Murray's passing.

Thank you to Jennifer Pelletier and Scott Yonkman, for facilitating access to the test rooms used in their respective buildings. Thank you to Glenn Jolly, for his patience and thoroughness in explaining to me (repeatedly) the minutiae of microphone circuitry.

Part of this work was made possible due to a partnership with the Institut PPRIME and the Université de Poitiers, with funding from Mitacs Globalink and Campus France. I thank everyone at the Institut PPRIME for welcoming me during my stay in Poitiers.

I thank my parents, my family and friends for their unwavering and unconditional support. Anna Gabaldon volunteered her free time to check over my spelling and verb tenses: an act of true friendship if there ever was one. Any remaining errors are my own.

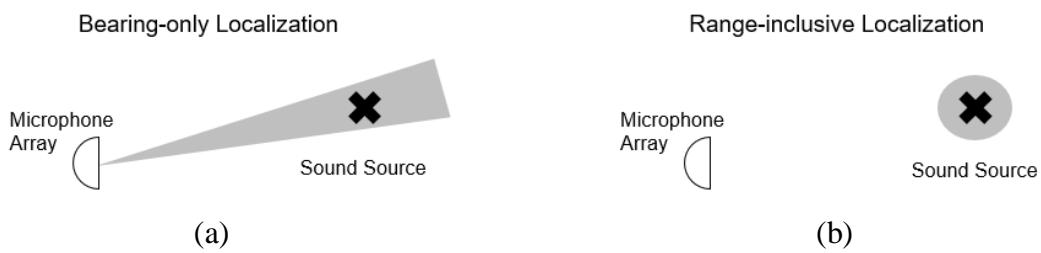
# Chapter 1: Introduction

## 1.1 Background and Objective

Sound source localization (SSL) is the task of determining the spatial location of a sound source, or multiple sound sources, in a given environment. This task has a wide range of applications, including speech enhancement, explosion detection, and industrial noise control.

Microphone arrays are a commonly-used and well-studied tool for SSL. A microphone array is defined as a collection of microphones placed in a known spatial arrangement relative to each other. Much work has been done to develop array processing algorithms of varying complexity and sophistication for the purpose of SSL, as well as other applications such as radar, telecommunications, and medical imaging with sensor arrays in general [1].

In the related literature, the term “localization” can refer to the identification of either a) the direction of arrival (bearing) of a sound source only (e.g. [2], [3]) or b) both the bearing and range (e.g. [4], [5]). The difference between these two definitions is illustrated in Figure 1. A common characteristic of conventional SSL array processing algorithms is that their range localization performance is poor, i.e. only (a) is feasible in most practical situations. However, being able to identify both bearing and range has obvious advantages, especially in situations where the localization task is complex (e.g. speaker identification in a teleconference, noise identification in an industrial room).



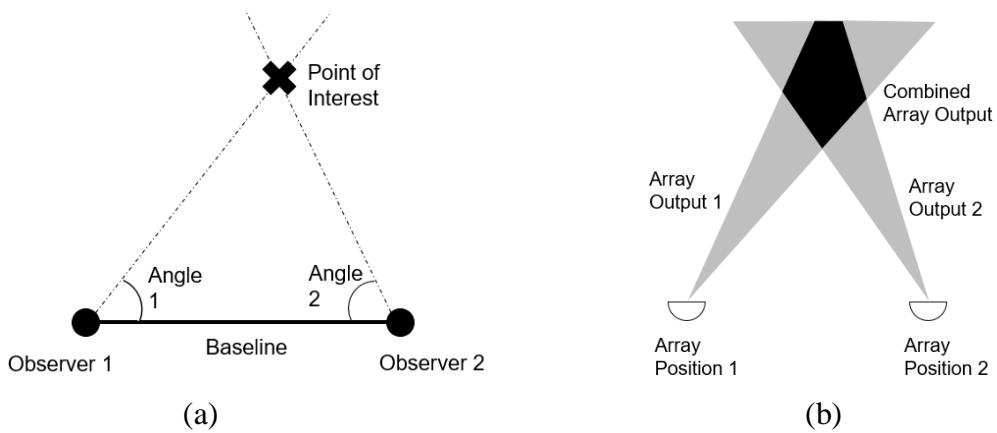
**Figure 1. Distinction between (a) direction-only and (b) range-inclusive localization of a sound source in 2D**

The focus of this thesis is on the localization of sound sources in rooms. For the purposes of this thesis, a room is defined as a space enclosed by walls, a floor and ceiling. In the vast majority of cases, these surfaces are, at least to some extent, acoustically reflective. The reflections caused by these surfaces create a reverberant sound field in the room, which, with conventional SSL algorithms, is interpreted as noise and is detrimental to SSL in rooms.

This thesis presents two methods for SSL using a microphone array: **direct triangulation** and **image source triangulation**. The objective of both of these methods is to locate sound sources in 3D, i.e. in both bearing and range, or in Cartesian coordinates ( $x,y,z$ ). The main concepts behind these two methods will be briefly explained below, with more extended explanations following in the main body chapters of the thesis.

### 1.1.1 Direct Triangulation: Range-finding using Multiple Bearing-only Measurements

**Direct triangulation** is a method to locate a sound source in both bearing and range by combining the output of multiple bearing-only measurements at multiple locations. It takes its inspiration from the technique of triangulation in surveying, where the range of a distant point can be found using multiple angle measurements made from a well-defined baseline, as illustrated in Figure 2.

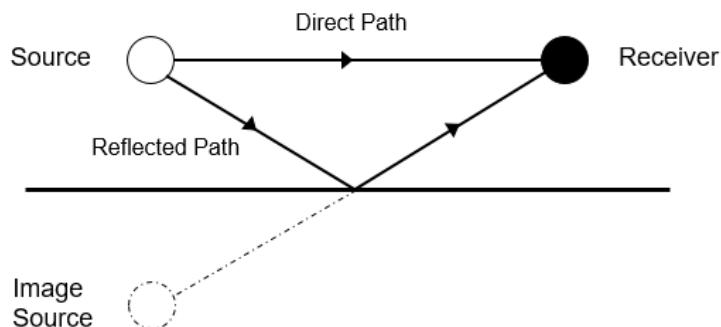


**Figure 2. (a)** Triangulation technique used by surveyors to locate a point of interest from a known baseline,  
**(b)** direct triangulation method to locate a region of interest using two known array positions

The concept of triangulation sound sources using multiple bearing-only measurements has been explored in various applications, which will be elaborated on in Chapter 3. The novelty of the work in this thesis lies in its focus on issues specific to room acoustics. Specifically, the presence of reverberation and image sources may have some impact on the effectiveness of triangulation, and this possible impact has, until now, not been investigated in any meaningful way.

### 1.1.2 Image Source Triangulation: Using Reverberation to Improve Localization

**Image source triangulation** attempts to simultaneously address both the range resolution problem and the problem of room reverberation, both discussed above, by incorporating information from the early specular reflections of a sound source in a room into a conventional localization algorithm. It does this using the image source model (ISM) of sound propagation [6], which posits that the sound field of a room can be modeled using image sources – virtual images of a sound source, mirrored about the reflective surfaces of the room as shown in Figure 3. From the ISM, it follows that, if the geometry of the room is known, the reverberant sound field can provide additional information about the direct source location.



**Figure 3. Image source model to determine the shortest reflection path between a source and receiver**

Reverberation is a persistent problem in SSL in rooms, and there have been many attempts to mitigate its effects, further detailed in Chapter 5. However, there have been fewer attempts to

actually use room reverberation to *improve* localization. The work presented in this thesis is the first attempt to apply the image source model to a beamforming-based SSL algorithm for this purpose.

## 1.2 Scope of Thesis

This thesis focuses on the implementation and evaluation of the direct triangulation and image source triangulation methods in reverberant rooms. Performance evaluation will be based on how accurately and precisely they can locate a single monopole sound source. Though multiple sources are discussed briefly in terms of image sources produced by reflective surfaces, more complex situations with multiple direct noise sources or distributed sources are beyond the scope of this thesis.

It is important to note that both direct triangulation and image source triangulation could be better described as “meta-processing” methods – rather than being an entirely new SSL algorithm, both of these methods start with a pre-existing “base processing” algorithm, on which further developments are made. Though this thesis discusses a specific array shape (hemispherical) and base processing algorithm (delay-and-sum beamforming), the proposed techniques could be applied to any array shape and base processing method available.

## 1.3 Structure of Thesis

This thesis begins with background information relevant to both SSL methods in Chapter 2, covering information about the specific microphone array and processing method used. From here, Chapters 3 and 4 focus on direct triangulation, while Chapters 5 and 6 focus on image source triangulation. Chapters 3 and 5 provide a background for direct triangulation and image source triangulation, respectively, including a review of related work, implementation details of each processing method and the various practical considerations that come with the implementation.

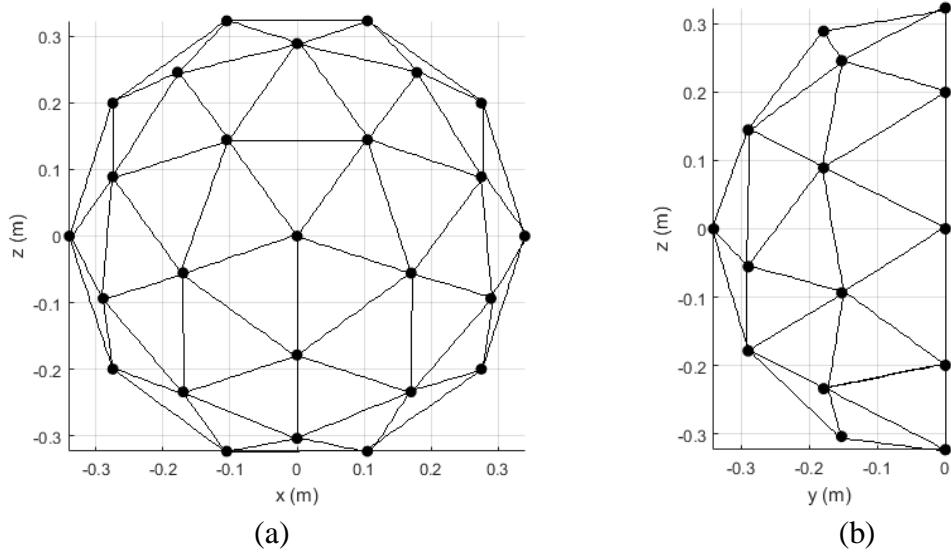
Then, Chapters 4 and 6 describe the experiments performed in order to evaluate the localization performance of each method and their results. Chapter 7 concludes this thesis by summarizing the key findings and recommendations of this research, as well as providing some ideas on future related work.

## **Chapter 2: Microphone Array and Processing Background**

This chapter provides the background information on which both direct triangulation and image source triangulation are based. First, the physical construction of the particular hemispherical microphone array used in this thesis will be described. Then, the base processing algorithm – delay-and-sum beamforming – will be discussed, both in its general conceptual form and in terms of its practical implications for this specific hemispherical microphone array.

### **2.1 Construction of Hemispherical Array**

The hemispherical microphone array was constructed in 2008 for the purpose of sound source localization using a novel algorithm based on the hemisphere’s intended structural similarities to an eye [7]. Since then, a number of projects at UBC (e.g. [8], [9]) involving the array have been carried out over the years. From these the array has been subject to various improvements and structural modifications, consisting of a new welded steel frame, entirely new preamplifier and data acquisition setup, and replacement microphones wherever necessary. In its current incarnation, the hemispherical array consists of 27 electret microphones arranged in a geodesic dome configuration, as shown in Figure 4. A center microphone on the front face (at [0, 0, 0]) is used as the reference point for the array. From here onward, all array processing results, unless otherwise specified, are calculated using this microphone array configuration, and reference to the “array position” or “receiver position” refers to the position of the center microphone.



**Figure 4.** Layout of hemispherical microphone array (a) front view, (b) side view. The center microphone is located at (0,0,0).

The microphones are held in place by 3D-printed plastic brackets, which are attached together using aluminum rods. A welded steel frame provides rigidity to the overall structure. The array can either be hung by the steel frame from the ceiling of a room, or held up using two tripods as shown in Figure 5.



**Figure 5.** Microphone array in test room supported by tripods

Over its years of service, the electret microphones in the original array have been replaced as necessary due to microphone failure. Currently there are three different microphone models (Best Sound Electronics (BSE) BGO-15L27-C1033, Panasonic WM61-A, PUI Audio POM-3535L-3-R) used in the array due to previous microphone models becoming unavailable to purchase – these replacements were selected through the suggestion of the supplier as being the best possible direct substitutes for each other. The microphone output is sent through a preamplifier before being routed to two NI6221 DAQ cards. Data acquisition and processing is performed in MATLAB. Details on the mechanical and electrical construction and calibration of the array are available in Appendix A.

## 2.2 Free-field Sound Propagation

As stated in Section 1.2, all of the sound sources discussed in this thesis (simulated or real) were assumed to be monopole point sources, radiating omnidirectional spherical pressure waves travelling at  $c$ . The propagation of these waves is governed by the spherically symmetric wave equation. In the spherical coordinate system  $(r, \theta, \phi)$ , the propagation of a time-dependent signal  $p(r, \theta, \phi, t)$  can be expressed as [10]:

$$\frac{\partial^2(rp)}{\partial r^2} = \frac{1}{c^2} \frac{\partial^2(rp)}{\partial t^2} \quad (1)$$

Then, it can be shown that the spherically symmetric solution to the wave equation at a distance  $r$  from the source has the form:

$$p(r, t) = \frac{A}{r} p\left(t - \frac{r}{c}\right) \quad (2)$$

A form of Eq. (2) can be used to model the propagation of a monopole source in free field (no reverberation). Given a source signal  $g(t)$ , source position  $r_s$  and receiver (microphone) location  $r_r$ , the signal  $s$  received at the microphone location can be expressed as:

$$s(t, \mathbf{r}_r, \mathbf{r}_s) = \frac{g\left(t - \frac{|\mathbf{r}_r - \mathbf{r}_s|}{c}\right)}{4\pi|\mathbf{r}_r - \mathbf{r}_s|} \quad (3)$$

Eq. (3) can also be considered as the convolution of the Green's function of a monopole source with the original signal,  $g(t)$ . This model is used at various points in this thesis to simulate the response of the hemispherical microphone array to a free field source.

### 2.3 Delay-and-Sum Beamforming

Delay-and-sum beamforming was selected as the base algorithm for both SSL methods in this project due to its conceptual simplicity and relatively low computational cost. The concept behind delay-and-sum beamforming, as with many microphone array processing techniques, is to use the implicit time delay between the microphones in the array in order to “focus” the array output on a certain point of interest [3]. For a microphone located at  $\mathbf{r}_m$  focused on a point  $\mathbf{r}_f$ , this time delay  $\Delta t$  can be written as:

$$\Delta t = \frac{|\mathbf{r}_f - \mathbf{r}_m|}{c} \quad (4)$$

Since each array microphone is located in a different position, each will have its own corresponding  $\Delta t$ . Assuming a microphone signal  $s_j(t)$  at each microphone index  $j$ , the delay-and-sum beamformer output of an array of  $M$  microphones can then be written as:

$$\begin{aligned} z_{DAS}(\mathbf{r}_f, t) &= \sum_{j=1}^M w_j s_j(t + \Delta t_j) \\ &= \sum_{j=1}^M w_j s_j\left(t + \frac{|\mathbf{r}_f - \mathbf{r}_j|}{c}\right) \end{aligned} \quad (5)$$

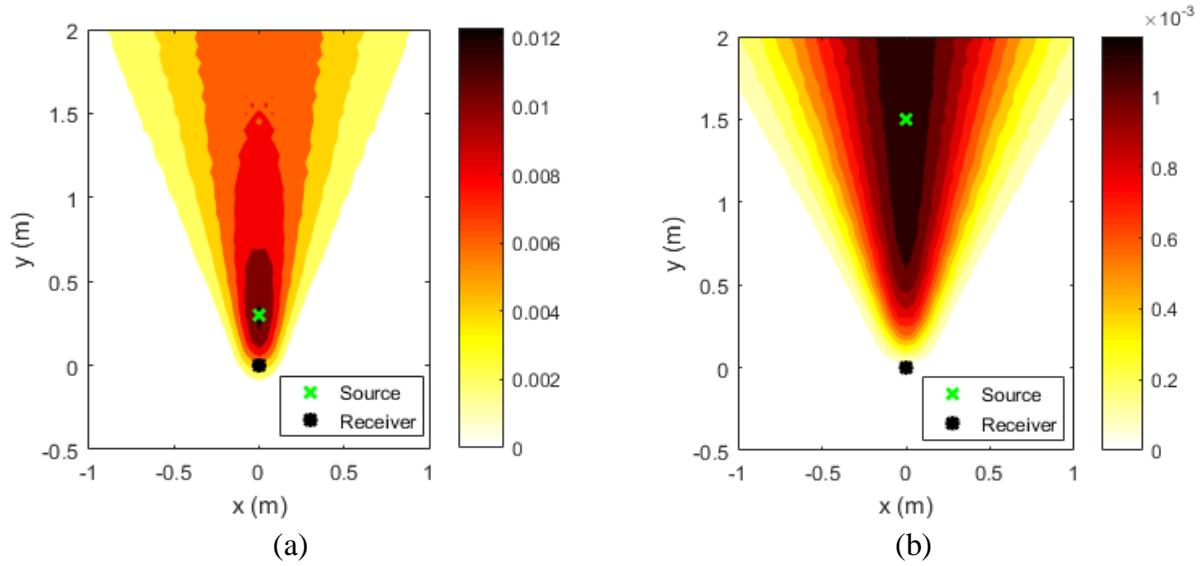
Where  $w_j$  is an amplitude weighting that may be applied to “shade” the array output in order to shape the output beam and reduce side lobe levels (no shading was applied for any of the

calculations in this thesis, i.e.  $w_j = 1$  in all cases). By summing the output signals delayed by the time corresponding to  $\mathbf{r}_f$ , a directional sensitivity or spatial filter is created in that direction: if a strong signal is coming from the location  $\mathbf{r}_f$ , the time delays will align the microphone signal and enhance the output  $z_{DAS}$ . The characteristics of the delay-and-sum beamformer as it relates to the spatial filtering concept will be discussed in Section 2.4.

Note that Eq. (5) is time-dependent, making it suitable for localizing an impulsive source. Many “real-life” acoustic sources (including all of the sources discussed in this thesis) are continuous in nature and thus benefit from time-averaging by taking the mean square (or root mean square) of a number of time points ( $t_k$ ), as follows:

$$\begin{aligned} z_{DAS,MS}(\mathbf{r}_f) &= \frac{1}{T} \sum_{k=1}^T \left( z_{DAS}(\mathbf{r}_f, t_k) \right)^2 \\ &= \frac{1}{T} \sum_{k=1}^T \left( \sum_{j=1}^M s \left( t_k + \frac{|\mathbf{r}_f - \mathbf{r}_j|}{c} \right) \right)^2 \end{aligned} \quad (6)$$

A well-known limitation of delay-and-sum beamforming (and of beamforming techniques in general), alluded to in Section 1.1.1 is its poor ability to locate the range (i.e. distance) of the source, due to the fact that spherical waves lose their curvature and approach a planar wave shape at far distances from a source [10]. Figure 6 (a) and (b) demonstrates this using the example of a simulated 750 Hz pure tone source at 0.3 m and 1.5 m source distances, respectively. At 0.3 m, the source can be localized in both direction and range (the peak output region is centered on the source location). At 1.5 m, the direction information is preserved, but the range is no longer obtainable from the output map.

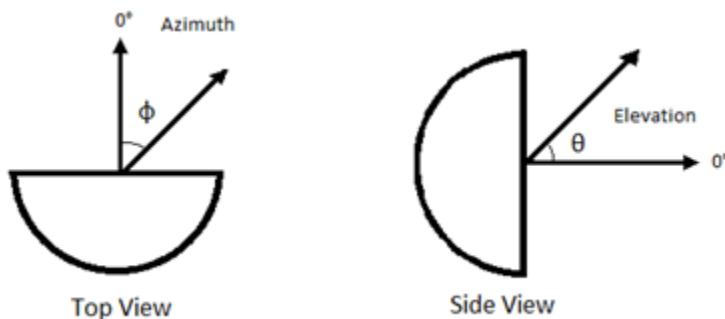


**Figure 6. 2D delay-and-sum beamforming output for a simulated 750 Hz pure tone source at (a) 0.3 m and (b) 1.5m distance from receiver**

## 2.4 Array Properties

### 2.4.1 Definition of Azimuth and Elevation Angles

Since the conventional delay-and-sum beamformer algorithm only produces the arrival directions of a sound field in the far field, it is often more practical to present results in a spherical coordinate system centered on the receiver location (center microphone) instead of Cartesian coordinates. Figure 7 defines the azimuth and elevation angles, relative to the microphone array, that will be used when presenting conventional (i.e. non-triangulated) beamforming results in this thesis.



**Figure 7. Definition of azimuth and elevation angles relative to hemispherical array**

## 2.4.2 Beam Width

The beam width, or source resolution of a beamformer is analogous to the bandwidth of a one-dimensional filter: it is desirable for the filter to only let pass a signal that is coming precisely from the direction of  $\mathbf{r}_f$ , and reject everything else. In practice, as with all filters, the beamfomer has a finite beam width beyond which separate sources cannot be resolved. This width is determined by the array aperture and frequency of interest. The Rayleigh criterion [11], most commonly used in optics, has been shown to be a good reference point to estimate the beam width of an array at a certain frequency [12]. The Rayleigh criterion states that, for a circular aperture of diameter  $D$ , the angle of separation  $\theta_R$  required to resolve two sources of wavelength  $\lambda$  is:

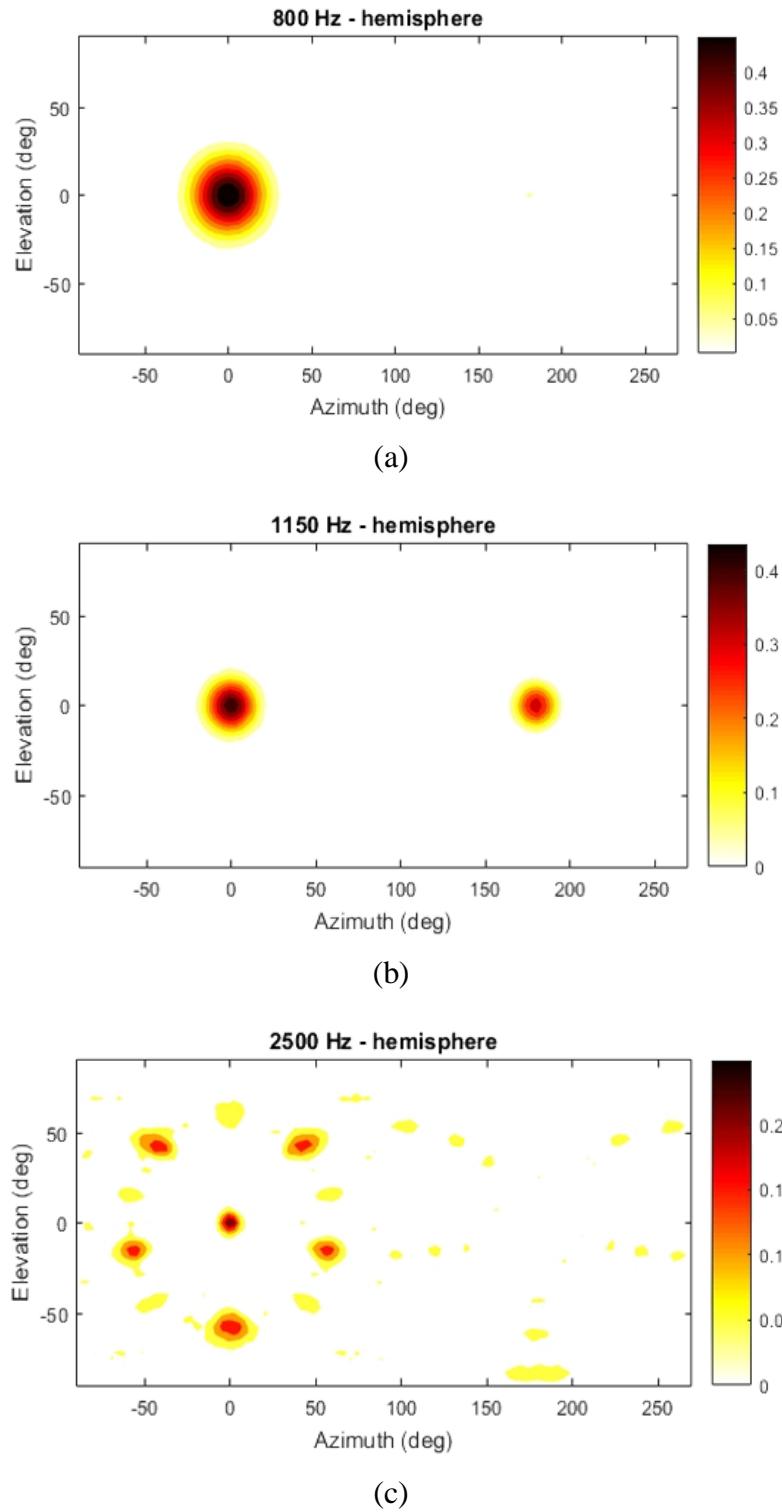
$$\theta_R = 1.22 \frac{\lambda}{D} \quad (7)$$

From the Rayleigh criterion, it follows that high-frequency (i.e. small wavelength) sources can be localized to a greater degree of precision than low-frequency sources. However, the operating frequency of a microphone array is limited by its geometry, as discussed below in Section 2.4.3. In the case of wideband noise (such as most common noise sources in industrial environments), the opportunity to select a frequency band of interest by filtering is available.

## 2.4.3 Operating Frequency

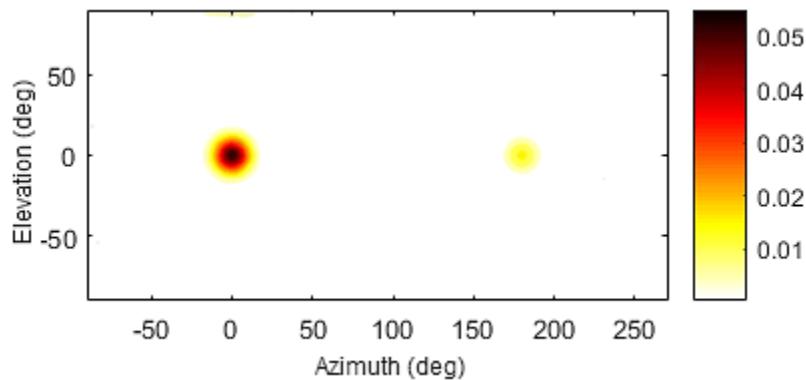
The upper operating frequency limit of a beamformer is determined by the presence (or the lack thereof) of grating lobes. These high-level side lobes (approaching the magnitude of the main lobe) are a symptom of “spatial aliasing” and appear when the spacing of the microphones is sparse compared to the frequency of interest – a more densely-packed array is capable of imaging at higher frequencies. For planar and linear arrays, the threshold at which grating lobes begin to appear is when the microphone spacing exceeds half the wavelength of interest ( $\lambda/2$ ) [13]. For the hemispherical array, the microphone spacing is 0.21 m, which corresponds to the half-wavelength

of an 811 Hz signal. The effect of increasing the frequency beyond this point can be seen is illustrated using a pure tone simulation in Figure 8, below – as frequency increases, the beam (centered at (0,0)) narrows, but grating lobes appear, first as a single back lobe in Figure 8 (b), then as multiple lobes in Figure 8 (c). In both of these cases, the grating lobes make it impossible to determine the actual source location (or if there are in fact multiple sources present) and render the output unusable. Deconvolution methods (e.g. DAMAS [14]) have been developed to address this problem by removing the array-dependent response from the beamforming output, but these methods are, at present, computationally expensive and thus impractical for scanning a large volume such as a room.



**Figure 8. Comparison of simulated beamforming output (spherical coordinates) of a pure tone signal at (a) 800 Hz (b) 1150 Hz (c) 2500 Hz.**

When beamforming with a band-limited signal as opposed to a single pure tone frequency, it appears that grating lobes are to some extent smoothed out – this may be due to the fact that the position and size of the grating lobes varies with frequency while the main lobe is always centered in the direction of the source, so by averaging over a wider frequency range the effects of the grating lobes can be mitigated. Through trial and error, the octave band centered at 1414 Hz (i.e. [1000, 2000] Hz) was selected for the experiments conducted in this thesis. Figure 9 shows the beamforming output produced from a Gaussian white noise signal filtered to this octave band.

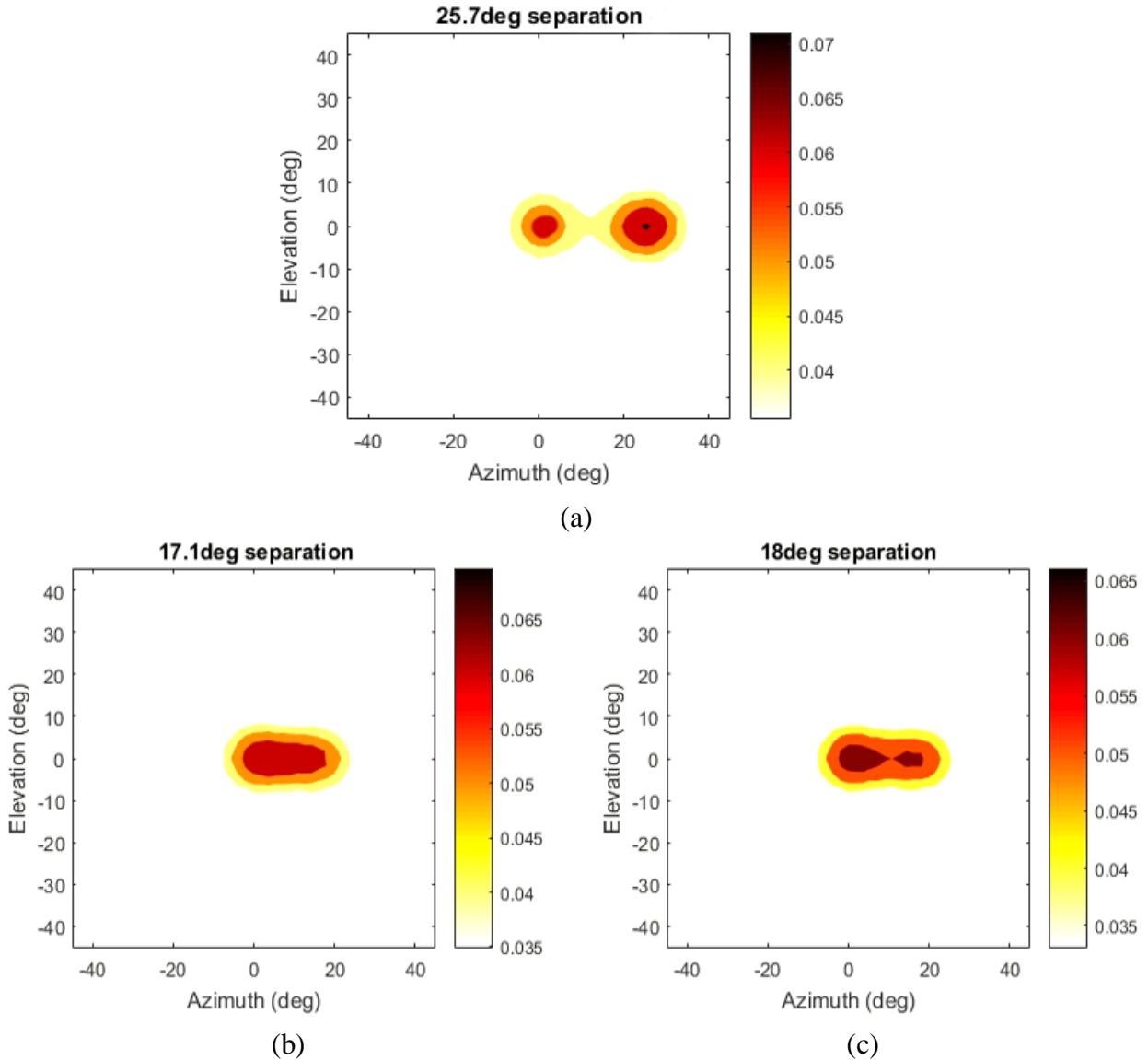


**Figure 9. Simulated beamforming output of a white noise signal filtered to the [1000 2000] Hz octave band**

Note that in this case there still is a small back lobe present (the second spot appearing at 180 degree azimuth) – this was deemed an acceptable tradeoff in exchange for the increased resolution and did not cause any noticeable problems in localization.

#### 2.4.4 Source Resolution

Using the center frequency of 1414 Hz and array radius of 0.34 m, Eq. (7) predicts that two sources will be resolved with a minimum separation angle of 0.448 rad (25.7 degrees) – Figure 10 (a) shows that the sources are indeed well-resolved at this angle. However, Figure 10 (b) and (c) show that, in practice there also exists another, smaller separation angle where the two sources become minimally resolved (i.e. two local maxima can be observed). This “minimum criterion” angle was found to be approximately 0.314 rad (18 degrees).



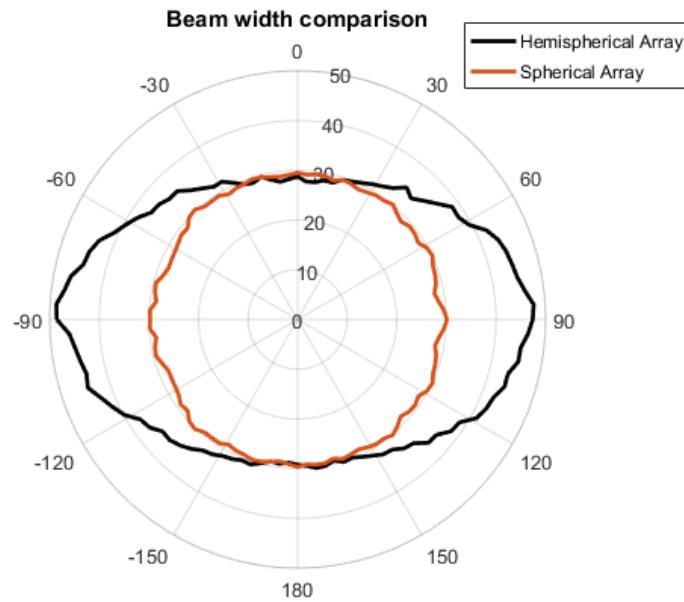
**Figure 10. Simulated beamforming output for two sources (Gaussian white noise filtered to [1000 2000] Hz octave band) with (a) 25.7 degree separation (Rayleigh criterion), (b) 17.1 degree separation (2 sources unresolved), (c) 18 degree separation (2 sources minimally resolved)**

#### 2.4.5 Beam Width Variation

Spherical microphone arrays are considered desirable for source localization due to their omnidirectional response – due to their spherical symmetry, the beam width of a signal is independent of its angle of arrival [15]. On the other hand, planar arrays are simple to construct and analyse, but the beam width of a planar array is heavily dependent of its angle of arrival. With

a hemispherical array, such as the one used in this thesis, one could expect that its directional response lies somewhere in between.

A comparison of the directional response characteristics of a spherical, hemispherical and planar array is presented in Appendix B. Resulting from this analysis, Figure 11 plots the beam width (full width at half maximum) of the same source as used in the previous section, as a function of azimuth arrival angle for the full 360 degree range of the hemispherical array, compared with the same for the spherical array. From this it is apparent that a spherical array would be preferable in order to have an omnidirectional response; however, the hemispherical array has some advantages in being easier to manipulate and position when making measurements.



**Figure 11.** Polar plot of angular beam width (FWHM) as a function of azimuthal arrival angle (degrees) for hemispherical (black) and spherical (orange) arrays. The zero degree line is perpendicular to the flat side of the hemisphere.

## **Chapter 3: Direct Triangulation - Background**

This chapter provides the background information specific to the direct triangulation method. First, an overview of related work will be provided. Then, the practical implementation of direct triangulation will be described, and various practical considerations arising from this implementation will be discussed.

### **3.1 Related Work**

Techniques that involve synthesizing data from multiple array measurements can be divided into two categories – one where data acquisition occurs simultaneously using multiple networked, synchronized arrays, and one where a single array is used and multiple measurements are performed sequentially over a period of time.

Multi-array techniques are generally more robust than single-array techniques – they can be used in situations where the noise is transient, and in applications where the scanning environment is very large (where it would be impractical to move a single array long distances in order to acquire the necessary measurements). These attributes make these techniques attractive for geoacoustic applications: for instance, [16] describes the use of multiple small microphone arrays spread over a 3 km aperture to localize infrasound noise from explosions over a 10 x 10 km<sup>2</sup> area, both by separately calculating the bearings for each array, and by treating entire network as a single large “meta-array”. A similar approach was used in [17], where a network of arrays was used in an urban environment in order to locate sound sources where there may not be a clear line of sight from array to source, due to obstructions from buildings and other urban features. At the room scale, [18] describes a networked series of small microphone arrays placed in a home environment, for the purpose of identifying spoken commands and general room activity level.

Single-array techniques could be considered in some ways more versatile than multi-array techniques – there is no need for synchronization across multiple arrays, and a single array is more portable and may be set up in different environments more conveniently. One approach that has received considerable research attention is to mount an array on a mobile platform and continuously acquire data in order to track sound sources in a given environment. [19] describes a microphone array mounted on a mobile robot to track sources in a 2D horizontal plane. In [20] this idea was extended to 3D space (though the array was moved manually by a human in this case, the authors note that it could have been done with a robot, had resources permitted). In both of these cases, the SSL component of the work was paired with imaging software to provide a real-time map (both visual and aural) of the environment. Perhaps the most similar related work can be found in [21], in which multiple (non-continuous) measurements from a 2D array were used in order to locate sources of sound entering or leaving the boundaries of a room. Since this work specified that the intended task was to search for sources entering a room boundary, it implies that prior knowledge of the room geometry is necessary. In contrast, the direct triangulation method outlined in this thesis requires no prior knowledge of the room geometry, beyond the assumption that all sources are, in fact, inside the room.

### 3.2 Implementation

The direct triangulation process begins by computing the standard delay-and-sum beamforming output (Eq. (5)). The output of all  $N$  receivers is then combined together (with the microphones and focus point being in a global coordinate system), either additively as in (8), or multiplicatively as in (9):

$$z_{DT,add}(\mathbf{r}_f, t) = \sum_{i=1}^N \sum_{j=1}^M s \left( t + \frac{|\mathbf{r}_f - \mathbf{r}_{ij}|}{c} \right) \quad (8)$$

$$z_{DT,mult}(\mathbf{r}_f, t) = \prod_{i=1}^N \sum_{j=1}^M s \left( t + \frac{|\mathbf{r}_f - \mathbf{r}_{ij}|}{c} \right) \quad (9)$$

The above formulation depends on a global time variable,  $t$ . Therefore, it necessitates that all measurements be taken at every position at once, which is possible only if  $N$  duplicates of the microphone array in question are available.

To work around this limitation, triangulation can be performed with a single array moved to each of the  $N$  positions with measurements taken sequentially, provided that the sound field of interest is stationary, ideally both in terms of physical position and power/spectral output. In this case, it is necessary to remove the time dependence by taking the mean square, of  $T$  time points ( $t_k$ ) before combining additively or multiplicatively, as follows:

$$z_{DT,add}(\mathbf{r}_f) = \frac{1}{T} \sum_{i=1}^N \sum_{k=1}^T \left( \sum_{j=1}^M s \left( t_k + \frac{|\mathbf{r}_f - \mathbf{r}_{ij}|}{c} \right) \right)^2 \quad (10)$$

$$z_{DT,mult}(\mathbf{r}_f) = \prod_{i=1}^N \left[ \frac{1}{T} \sum_{k=1}^T \left( \sum_{j=1}^M s \left( t_k + \frac{|\mathbf{r}_f - \mathbf{r}_{ij}|}{c} \right) \right)^2 \right] \quad (11)$$

By computing either (10) or (11) over a three-dimensional grid of focus points  $\mathbf{r}_f$ , a volume of triangulated beamforming output data is produced.

Note that the concept of direct triangulation can be applied to localization algorithms other than the conventional delay-and-sum beamforming algorithm shown here – it is simply a matter of substituting the (time-averaged) localization output of the chosen method into equation (10) or (11).

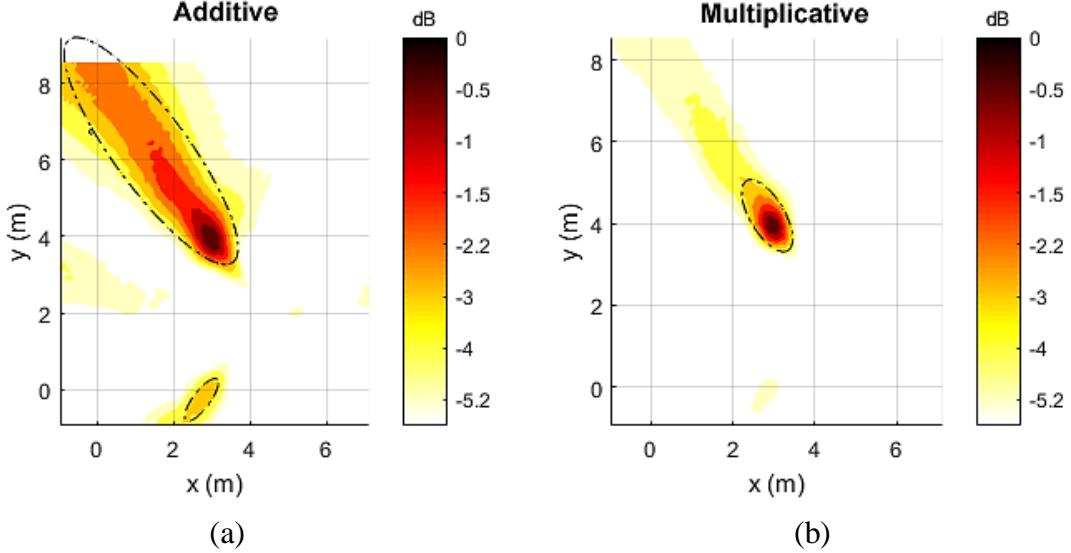
### 3.3 Practical Considerations

#### 3.3.1 Multiplicative vs Additive Triangulation

As shown in the previous section, there are two methods to combine receiver outputs when triangulation – either by adding the outputs at every focus point  $r_f$ , or by multiplying them together.

Considering the properties of multiplication versus addition, multiplicative triangulation is preferable if there is a high degree of certainty that all receivers can “see” the source (i.e. there are no obstructions and the direct sound from the source is sufficiently strong), as it produces better contrast between high- and low-output points, as long as all receivers are in agreement with whether the output of a point is high or low. Conversely, additive triangulation produces better results in situations where it is not certain if all receivers can effectively “see” the source – in this case, the properties of addition will ensure that the output seen at some (but not necessarily all) receivers is at least partially represented in the final output.

Another way to present this difference is to interpret multiplicative triangulation as more prone to “false negative” results, whereas the additive triangulation method avoids this at the expense of contrast and resolution. With direct triangulation in small spaces (such as those used in the tests conducted here), the direct sound from a source is very likely to be “seen” at every receiver location. Therefore, multiplicative triangulation is preferable in this case – an example comparison between the two methods is shown in Figure 12 below. (The input data used for this figure is discussed in Section 4.5: Figure 12 (b) is taken directly from the results shown later in Figure 36.)



**Figure 12. Comparison between (a) additive and (b) multiplicative direct triangulation.** The dashed lines represent the “half-maximum ellipse”, a performance metric discussed in Section 4.3

### 3.3.2 Spacing of Grid Points

When choosing a grid spacing for the scanning volume, it is necessary to consider the tradeoff between computation time and discretization error in peak identification. For a given grid spacing,  $\Delta$ , the maximum (worst-case) discretization error magnitude  $|e_{disc}|$ , is given by:

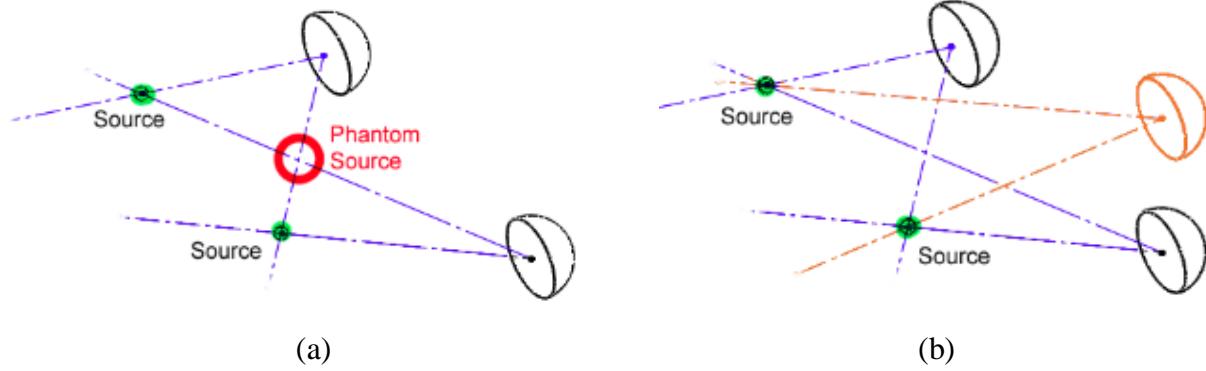
$$|e_{disc}| < \frac{\Delta}{2} \quad (12)$$

For all of the localization results presented in this thesis, the scanning volume had a uniform grid spacing of 0.1 m in all directions (x,y,z), producing a maximum discretization error of 0.05 m in each direction, or 0.0866 m in total.

### 3.3.3 Receiver Positions

In nearly all of the direct triangulation measurements presented in this thesis, three receiver positions are used, with the microphone array moved to three different points on a common plane. The reason for having three receivers in a triangular configuration was in order to resolve an

ambiguity (referred to as a “phantom source” here and as “ghost sources” in [22]) that may occur in a case where there are two sources aligned horizontally, as illustrated in Figure 13.



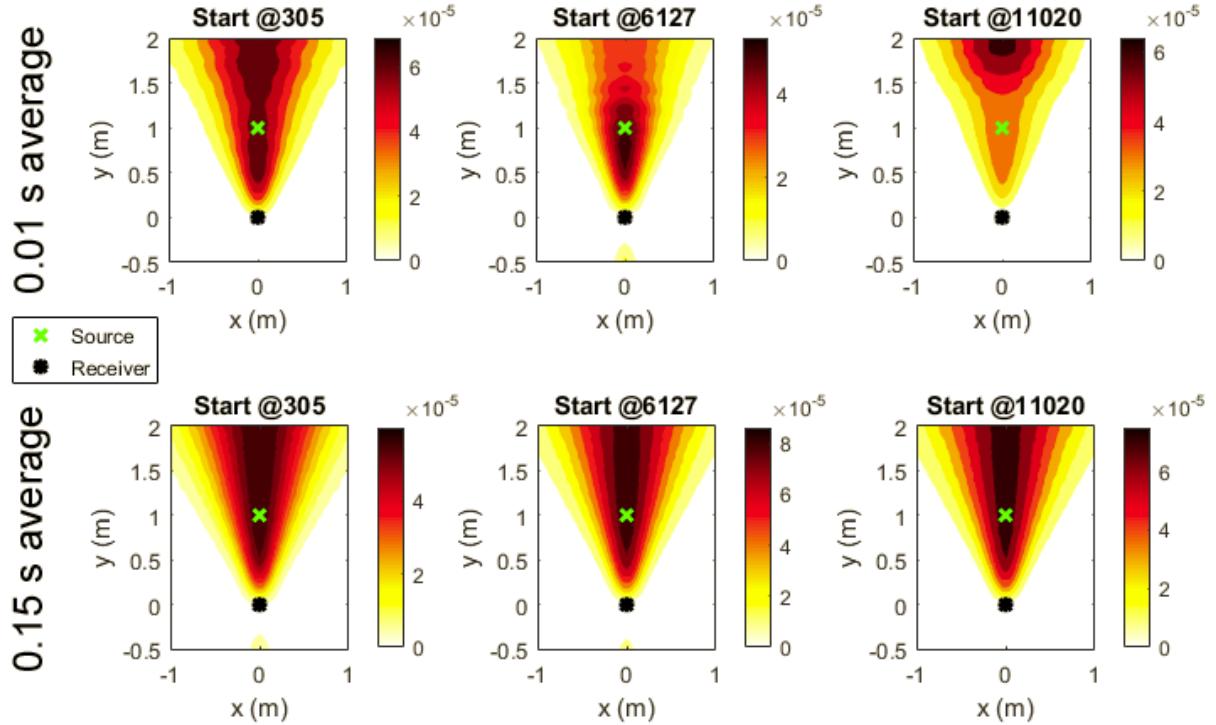
**Figure 13.** (a) “Phantom Source” ambiguity with two receiver positions is eliminated in (b) when a third receiver position is added

The decision to keep all receivers on the same plane was made primarily to simplify the process of physically moving the array around in an accurate manner, as well as for space considerations, since the rooms in which tests were conducted were fairly small. However, this comes with the consequence that the ambiguity seen in Figure 13 above is still possible when there are sources that share the same plane as the receivers – therefore, it is important to select a receiver plane where sources are unlikely to be found, if possible.

### 3.3.4 Averaging Period

This section addresses the period to use when time-averaging the beamforming output (i.e. the selection of  $T$  in Eqs. (6), (10), and (11)). The time-variant nature of a sound pressure field means that, if few points are sampled, the resulting mean square (or root mean square) output may vary widely. However, as  $T$  increases, the output will eventually converge on a steady-state distribution. This is demonstrated in Figure 14 below – conventional (non-triangulated) delay-and-sum beamforming was performed on a simulated Gaussian white noise source filtered to the [1000, 2000] octave band (same as in Section 2.4). The top row of plots shows the beamforming output using a 0.01 s time average at three arbitrary start points in the signal, while the bottom row shows

a 0.15 s time average of the same signal at the same start points. The shape of the 0.01 s results vary depending on the start point and therefore are not ideal for triangulation (or indeed any source localization) while the 0.15 s results have converged to a single solid beam.



**Figure 14. Delay-and-sum beamforming outputs at 3 arbitrary start points using (top row) 0.01 s time average and (bottom row) 0.15 s time average.**

In all cases presented in this thesis, a 5000-point average was used, corresponding to 0.31 s and 0.16 s at 16384 Hz and 31250 Hz (the two sampling rates used in the experiments below) – this was found to be sufficient to produce a solid beam like those found in the bottom row of Figure 14 in all cases.

## **Chapter 4: Direct Triangulation - Experiments**

This chapter presents the series of experiments conducted in order to evaluate the direct triangulation method.

### **4.1 Progression of Experiments**

The first round of direct triangulation experiments verified the performance of the direct triangulation method in free field – this was done in simulation (using a free field propagation model) as well as with measurements taken in an anechoic chamber. Next, the triangulation setup was moved to a reverberant room (empty classroom). In this room, experiments were conducted varying the source and receiver positions relative to the room’s various features (walls, geometric irregularities, etc.).

Finally, the effect of multiple sources was briefly investigated in two different ways – first, a free-field simulation of two sources at varying distances was developed to look at the Rayleigh criterion’s applicability to triangulation. Then, some of the reverberant room results obtained in the previous section were re-examined with a focus on the effect of image sources due to the walls and floor of the room.

### **4.2 Experimental Setup**

In the measured cases, the source input signal was broadband noise (10 Hz-100 kHz) generated using an SRS SR770 Network Analyzer. This signal was played through a monopole source consisting of an Altec Lansing 288-16K high frequency driver connected to a thin steel tube [23]. The sound power output of this source was measured to be 0.55 mW. The microphone signal was sampled at 16384 Hz in the free field measurements, and at 31250 Hz for the measurements taken in the reverberant room.

In the simulated cases, the free field sound propagation model (Eq. (3)) was used. The input source signal was Gaussian white noise, with sound power equivalent to that used in the measured case (0.55 mW). Uncorrelated random noise (equivalent to 1% of the source sound power level) was added to each signal in order to simulate microphone line noise.

In all cases (both measured and simulated), the microphone output was filtered using a Butterworth filter (5<sup>th</sup> order) with pass band [1000, 2000] Hz before further post-processing. Table 1 summarizes the key parameters used for the experiments in this chapter.

**Table 1. Parameters for direct triangulation experiments**

Parameter	Value
<b>Triangulation Type</b>	Multiplicative
<b>Sampling Rate</b>	16384 Hz (Free Field) or 31250 Hz (Reverberant Room)
<b>Sampling Time</b>	5000 points (0.31 s in free field, 0.16 s in reverberant room)
<b>Grid Spacing</b>	0.1 m
<b>Source Power</b>	0.55 mW (real or simulated)
<b>Frequency Range</b>	[1000, 2000] Hz

For the convenience of the reader, details regarding the specific test environment and source-receiver positions used in each experimental case will be enumerated in their respective sections.

### 4.3 Presentation of Results

When evaluating the 3D source localization performance of any given measurement or simulation case, there are two characteristics of interest: accuracy in locating the source position, and resolution (width) of the “blob” surrounding the source.

### 4.3.1 Accuracy

In all of the cases to be presented below, the sources are assumed to be acoustic monopoles.

In the case where it is known that there is only one sound source, determining its location is simply a matter of locating the absolute maximum output value in the data volume. In the case where the number of sources is unknown, potential sources can be found by retrieving the local maxima in the data volume, and then removing any maxima that are below a certain threshold of the absolute peak.

The accuracy of source localization can then be evaluated by measuring the distance between the found source point  $\mathbf{r}_{pk}$  and the actual known source location point  $\mathbf{r}_{loc}$  and normalizing by the source-receiver distance,  $d$ , as follows:

$$d = |\mathbf{r}_s - \mathbf{r}_r| \quad (13)$$

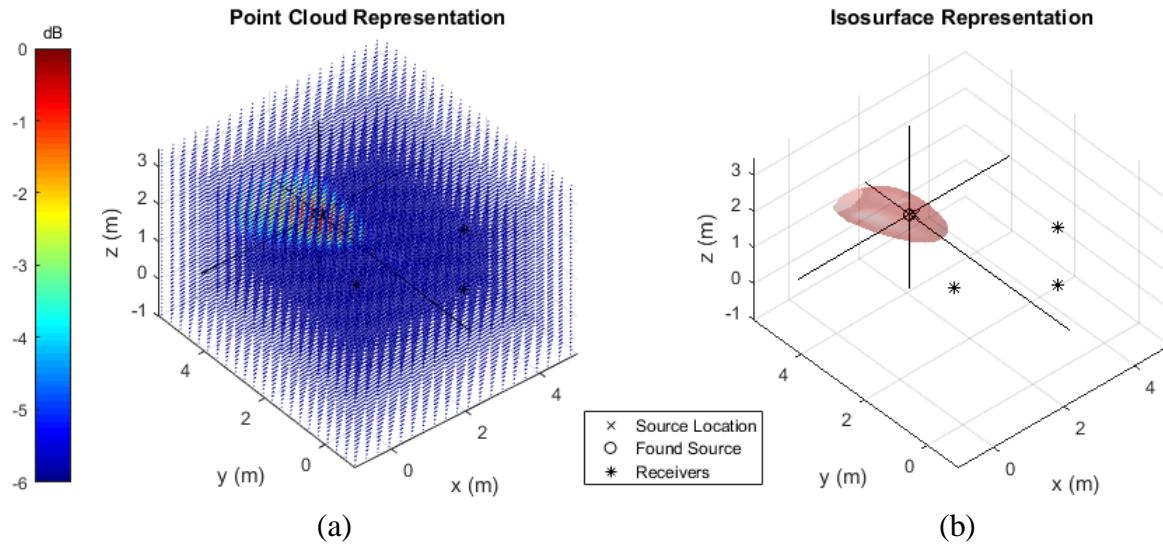
$$e = \frac{|\mathbf{r}_{pk} - \mathbf{r}_{loc}|}{d} \quad (14)$$

When computing the source-receiver distance,  $d$ , it is necessary to clarify what is meant by the “receiver position” since there are multiple receiver locations in question. The single reference “receiver position” is defined as the centroid of the polygon whose vertices are located at the receiver locations.

### 4.3.2 Resolution

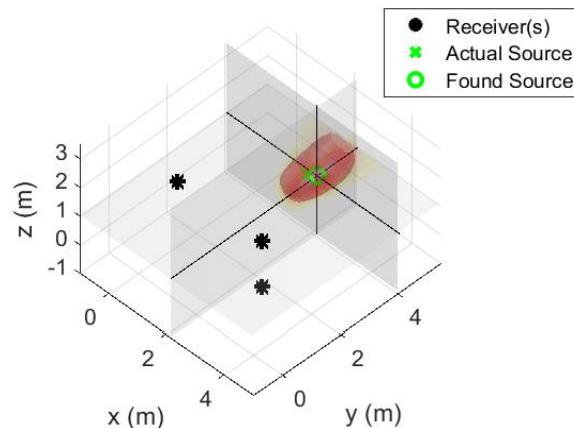
A commonly used (e.g. [12], [21], [24]) measure of resolution in source localization is the half-maximum width around the main source peak. In three-dimensional space, this can be visualized as an isosurface connecting all points with the same value in the same way that a contour line would do so in two dimensions. The isosurface representation, paired with peak point identification, is a quick and useful way to evaluate source localization performance at a glance,

especially compared to the (somewhat cumbersome) point cloud representation of the volume, as shown in Figure 15.



**Figure 15. Comparison between (a) ineffective point cloud representation and (b) isosurface representation of 3D beamforming output**

In order to show further detail in the region surrounding the source, it is also useful to show contour slices of the three Cartesian planes (X-Y, Y-Z and X-Z) intersecting the point of interest (i.e. the assumed source location), as shown in Figure 16. This set of three Cartesian planes, plus the isosurface representation shown in isometric view (Figure 16), constitute the “graphical results” that will be shown for each measurement case in this chapter, as well as in Chapter 6.

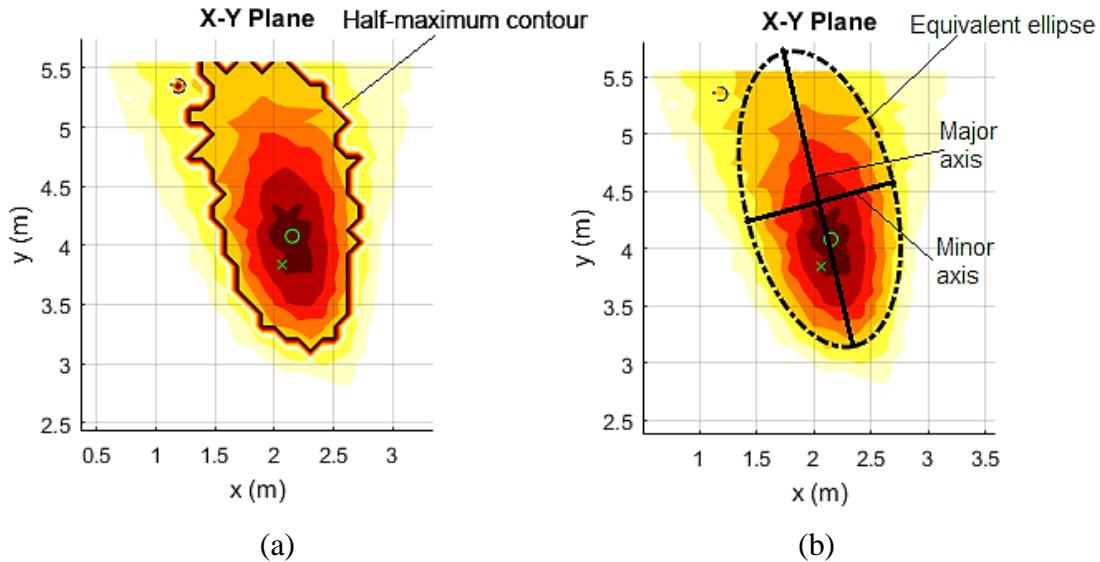


**Figure 16. Example isosurface view of results with three intersecting Cartesian planes**

In order to quantify the size of the half-maximum region surrounding the source, “equivalent ellipses” were drawn that share the same second central moments as the half-maximum contours in each given plane, as demonstrated in Figure 17. Using the average of the major and minor axes ( $a_{maj}$  and  $a_{min}$ , respectively) of each ellipse, we obtain a “blob length” that can be used as a numerical metric for source resolution, normalized by the source-receiver distance:

$$L_{plane} = \frac{1}{2}(a_{maj,plane} + a_{min,plane}) \quad (15)$$

$$Blob\ length = \frac{1}{3d}(L_{xy} + L_{yz} + L_{xz}) \quad (16)$$

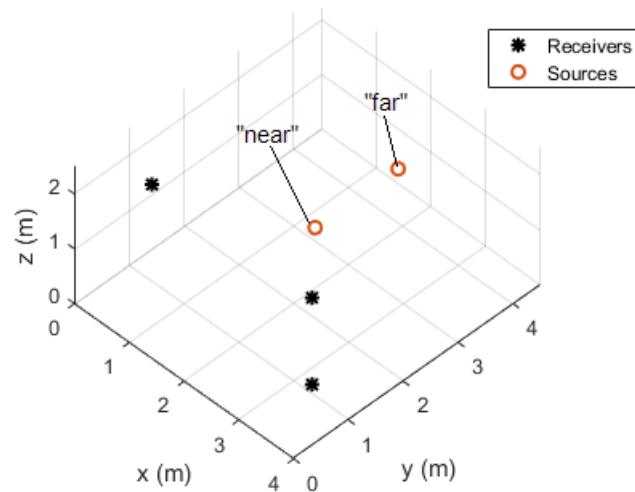


**Figure 17.** Detailed view of X-Y plane blob, with (a) half-maximum contour and (b) equivalent ellipse with major and minor axes

## 4.4 Free Field Experiments

### 4.4.1 Test Environment

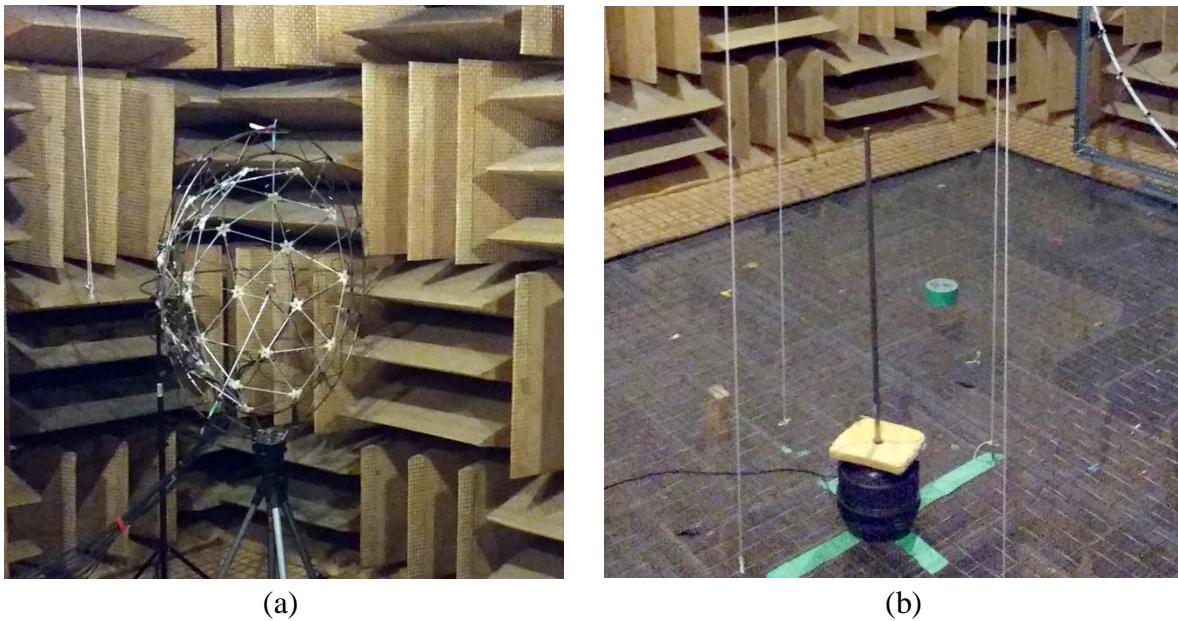
Free field triangulation was performed using a set of three receiver positions in a right-triangle configuration, and two different source positions (“near” and “far”), as shown in Figure 18 and Table 2. In the measured cases, the source and receiver were placed in an anechoic chamber, roughly 4 m x 4.5 m x 2.5 m in size, as shown in Figure 19.



**Figure 18.** Free field source-receiver configuration, isometric view

**Table 2.** Source and receiver coordinates for free field cases

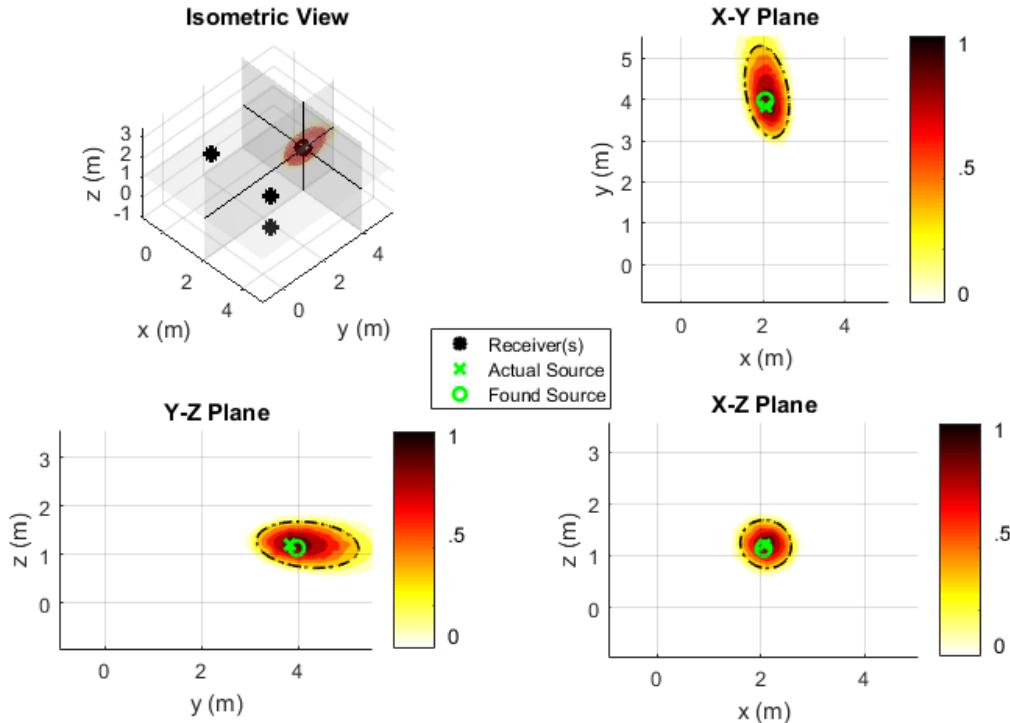
	Coordinates (x,y,z) [m]
<b>Receiver 1</b>	(0.55, 0.865, 1.95)
<b>Receiver 2</b>	(3.47, 0.865, 1.95)
<b>Receiver 3</b>	(3.47, 0.865, 0.37)
<b>Source "near"</b>	(2.07, 2.00, 1.21)
<b>Source "far"</b>	(2.07, 3.835, 1.21)



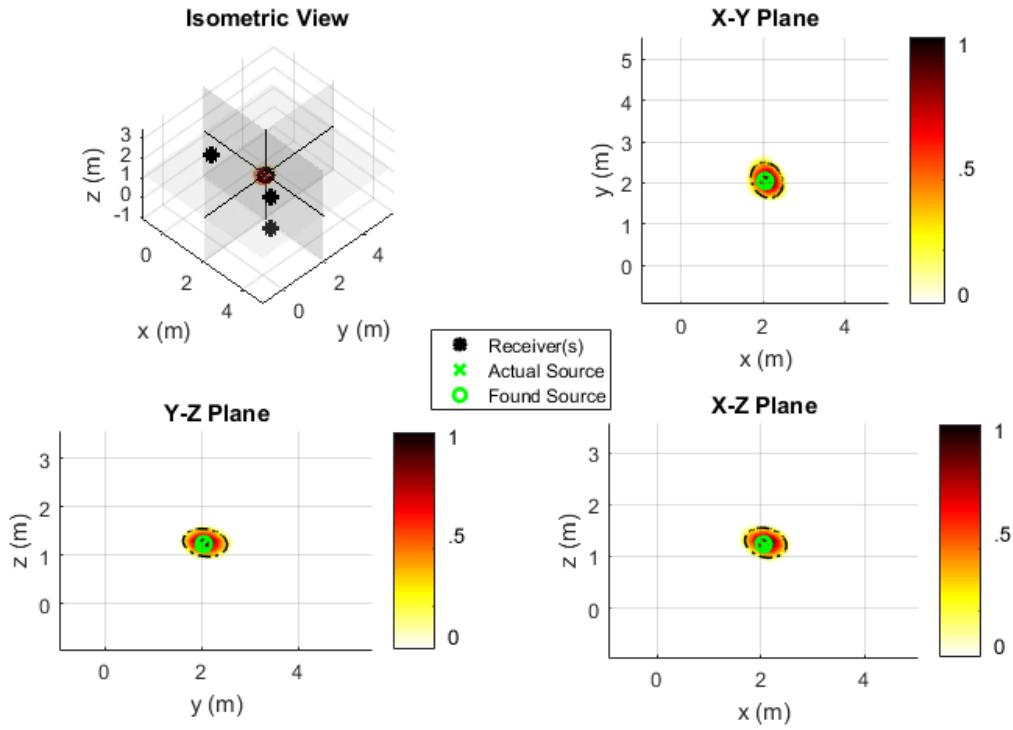
**Figure 19.** (a) Microphone array and (b) monopole source in anechoic chamber

#### 4.4.2 Results

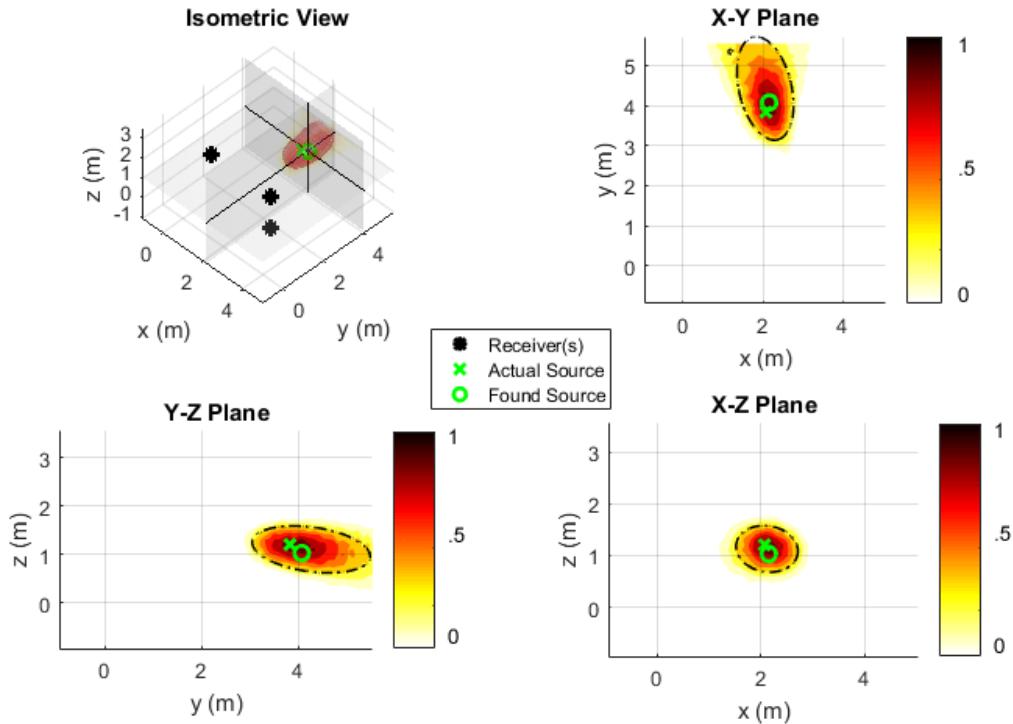
Figures 20 to 23 show the graphical results for these four cases, and Table 3 shows the corresponding error and resolution values obtained from these plots.



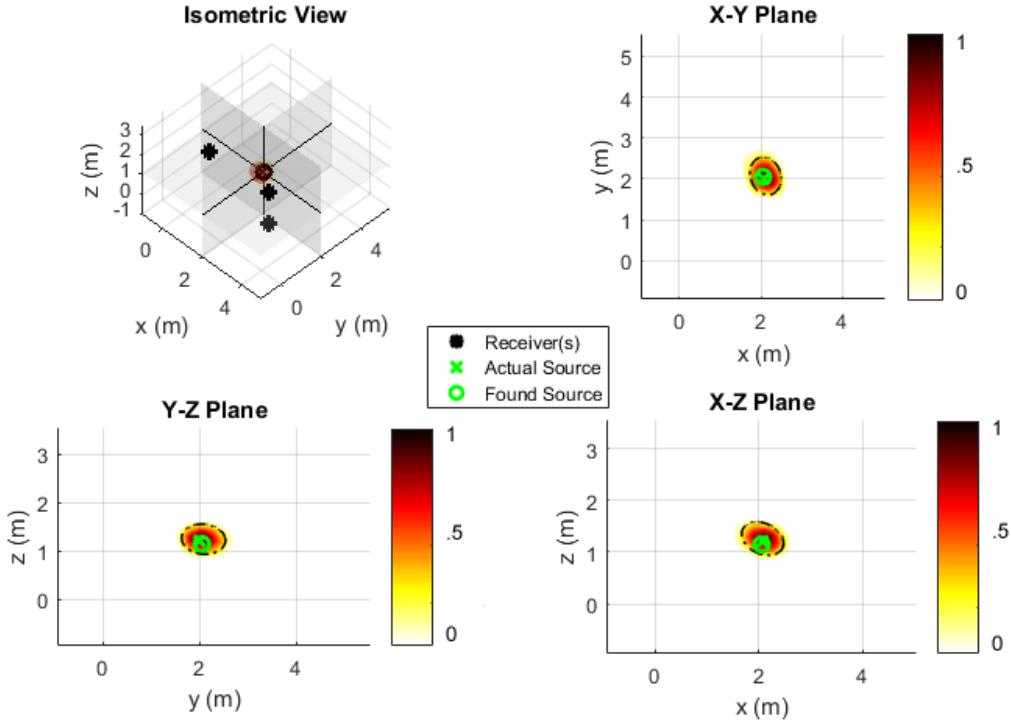
**Figure 20.** Graphical results for free field simulation, “far” case. Peak location was at (2.05, 3.98, 1.13); actual source location was (2.07, 3.835, 1.21)



**Figure 21.** Graphical results for free field simulation, “near” case. Peak location was at (2.05, 2.05, 1.23); actual source location was (2.07, 2.00, 1.21)



**Figure 22.** Graphical results for free field measurement, “far” case. Peak location was at (2.05, 3.98, 1.23); actual source location was (2.15, 4.08, 1.03)



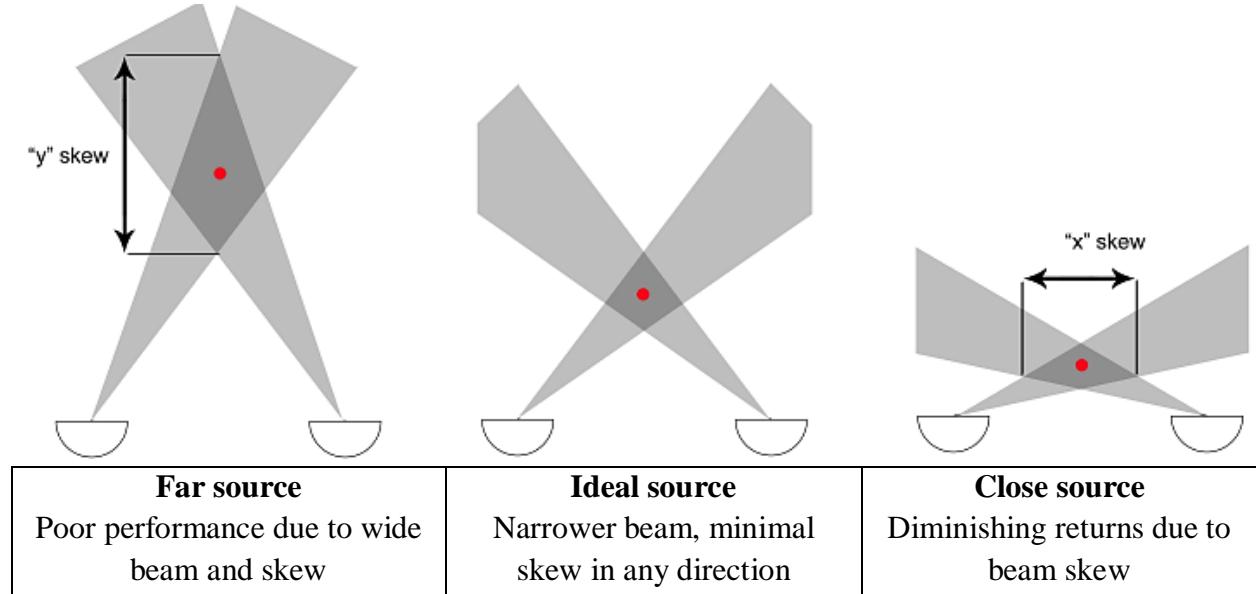
**Figure 23.** Graphical results for free field measurement, “near” case. Peak location was at (2.05, 1.94, 1.23); actual source location was (2.07, 2.00, 1.21)

**Table 3.** Error and resolution comparison between free field measured and simulation cases

Case		S/R distance (m)	Total error (m)	Normalized error	Average blob length (m)	Normalized blob length
<b>Far</b>	Simulation	3.01	0.16	0.054	0.69	0.229
<b>Near</b>	Simulation	1.23	0.06	0.046	0.37	0.301
<b>Far</b>	Measured	3.01	0.31	0.104	0.79	0.261
<b>Near</b>	Measured	1.23	0.09	0.075	0.39	0.315

As expected, non-normalized error and resolution worsen at farther distances between source and receiver – this is due to the beam width increasing as distance increases (angular width remaining the same with increased range). However, attempting to correct for this effect by normalizing using the source-receiver distance does not produce the desired result (i.e. a consistent resolution value – the normalized blob length in the “near” cases are higher than those in the “far”

cases). One possible reason for this is the effect of triangulation at close distances, relative to the receiver spacing, shown in below in Figure 24.



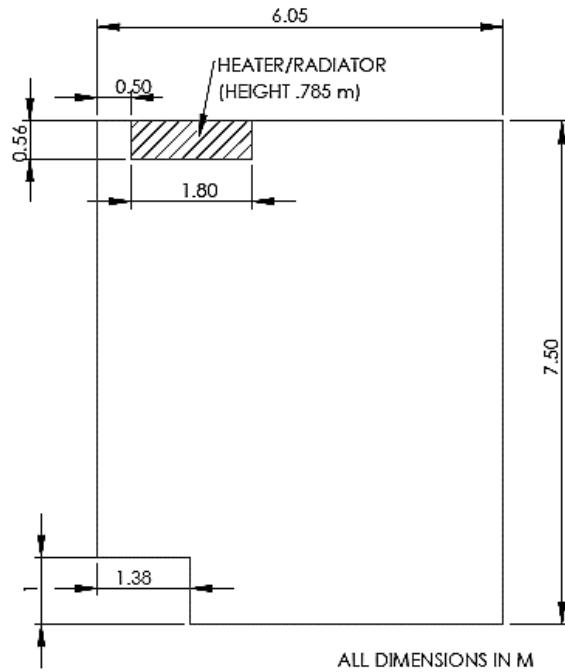
**Figure 24. Illustration of distance skew effect in triangulation of beams**

The normalized error and resolution are poorer in the measured cases compared to the simulated cases – this may be due to error in positioning the source and receivers (in the case of the increased error) and additional noise in the instrumentation chain (in the case of the increased blob length).

## 4.5 Reverberant Room Experiments

### 4.5.1 Test Environment

A more realistic source localization case is presented below, with a single source being localized in an empty classroom. The room is still mostly devoid of furniture and fixtures, but its walls are acoustically reflective, unlike those of the anechoic chamber. Figure 25 shows a plan view of the room only (source and receiver positions for each specific case will be shown in figures accompanying the relevant section).



**Figure 25. Plan view of reverberant test room**

The test room was mostly rectangular, with a slight convex protrusion (bottom left corner of Figure 25). The topmost wall in Figure 25 was composed of painted cinderblock and the remaining walls were single-stud drywall, the floor was vinyl and the ceiling was corrugated steel with truss supports. Its reverberation time ( $T_{60}$ ) was measured to be 1.30 s in the octave band of interest (interpolated from the 1000 Hz and 2000 Hz octave band measurements).

From the room geometry and reverberation time, the Schroeder frequency was found to be 208 Hz and the reverberation radius was 0.57 m. The Schroeder frequency [25] is the approximate frequency threshold above which the room response can be analyzed using statistical means and modal effects are not significant. The reverberation radius [26] is the distance threshold beyond which reverberation (as opposed to direct sound) dominates the sound field. In the case of the experiments conducted in this section, the frequency band of interest is above the Schroeder frequency and the source-receiver distances are beyond the reverberation radius. The latter

property is particularly significant because the direct triangulation method relies on the direct sound field for localization. Reverberation in this case is considered noise and is not desirable, so the reverberant room case is less favourable for localization than the free field case discussed above.

Figures 26 and 27 are photographs of the room, as seen from roughly the source and receiver perspectives, respectively.



**Figure 26. Test room as seen from source location**



**Figure 27. Test room as seen from receiver location**

#### 4.5.2 Receiver Configuration

Due to the asymmetry of the room (both geometrically and in terms of wall composition), a potentially significant factor is the placement of the receivers relative to the various room features. This was previously not an issue in the free field cases, due to the absence of reflective surfaces and features altogether.

In order to investigate the effect of this asymmetry on localization performance, measurements were taken at four different locations (A, B, C and D) as shown in Figures 28 and 29. Then, triangulation was performed using all possible three-source combinations (ABC, ABD, ACD, BCD) as well as all four sources (ABCD). The graphical results are plotted in Figures 30 through 34, and Table 4 shows the corresponding error and resolution values obtained from these plots.

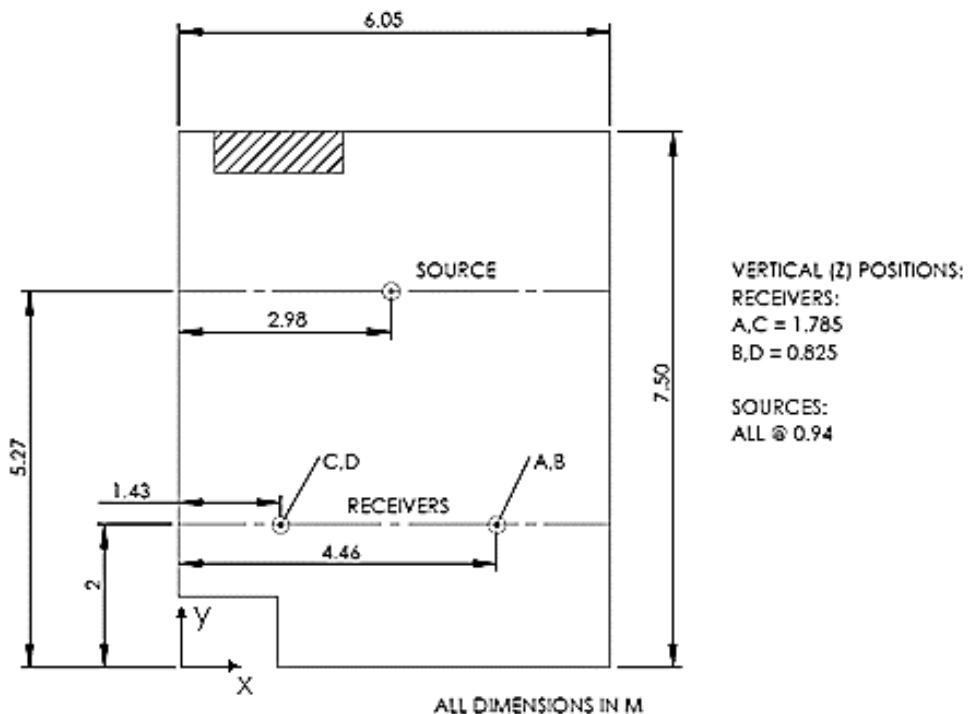
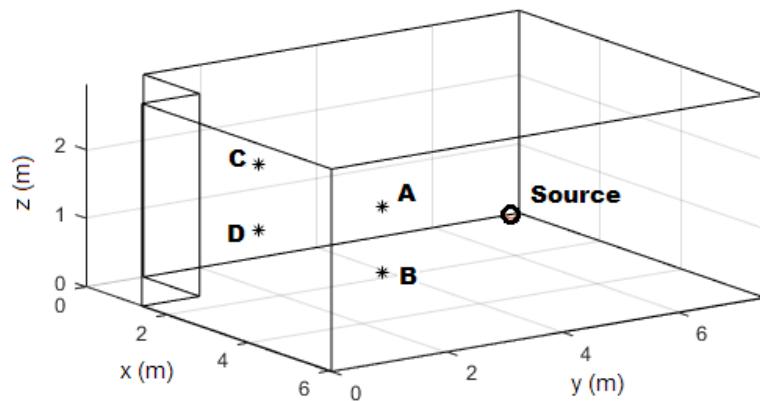
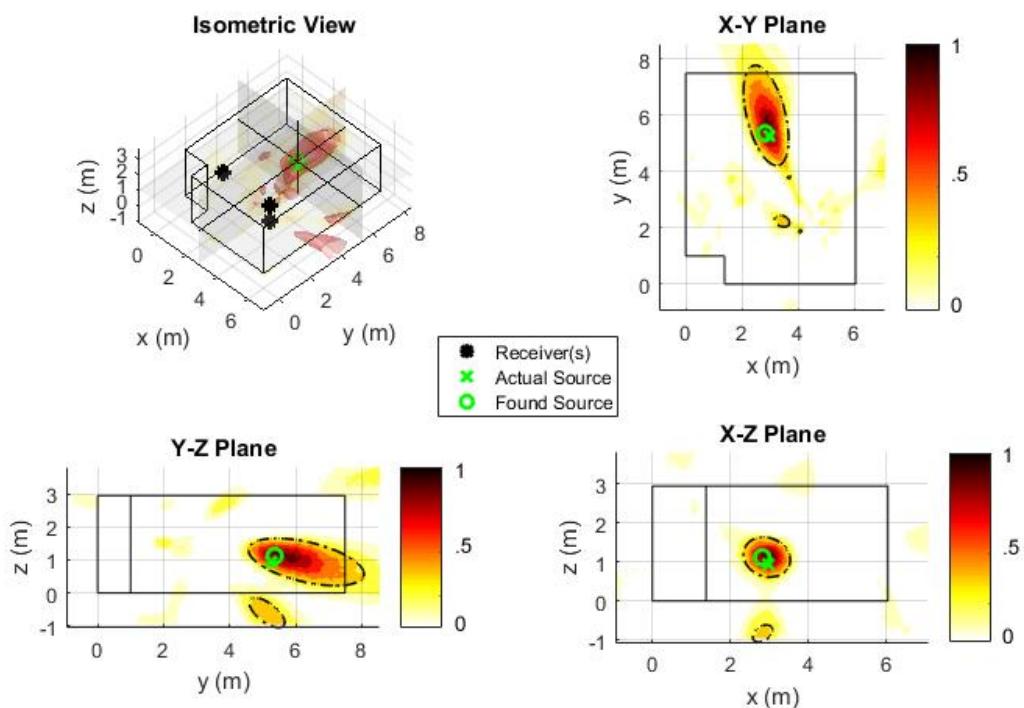


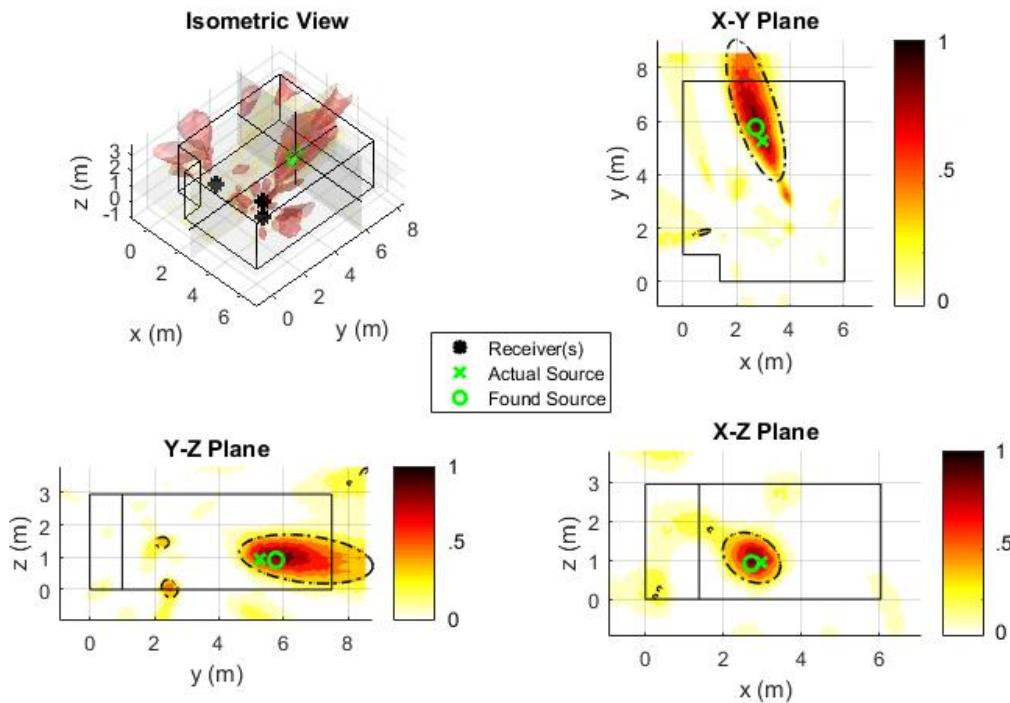
Figure 28. Plan view of test room during receiver configuration experiment



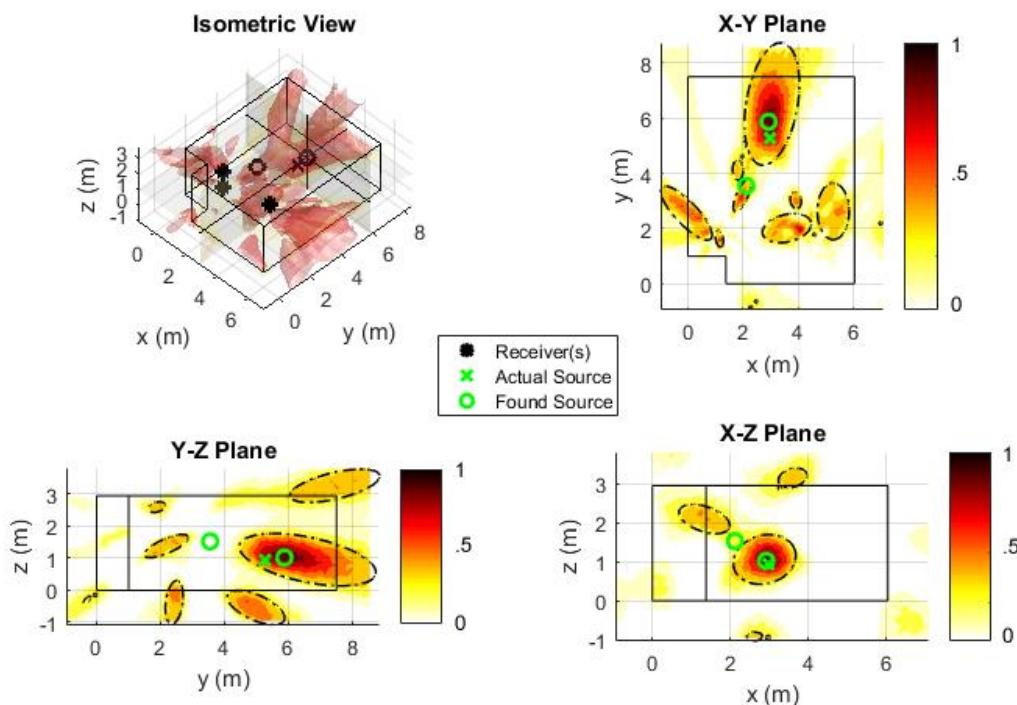
**Figure 29. Perspective view of A, B, C, D receiver locations relative to source and room**



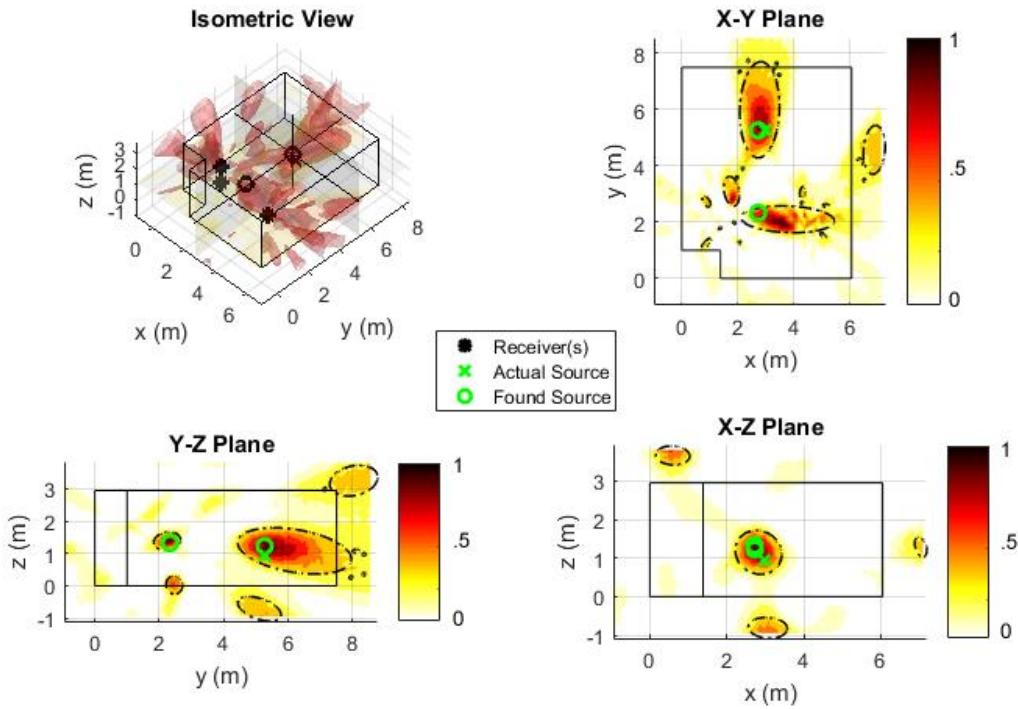
**Figure 30. Graphical results for ABC receiver configuration. Peak location was at (2.82, 5.37, 1.12); actual source location was (2.98, 5.27, 0.94)**



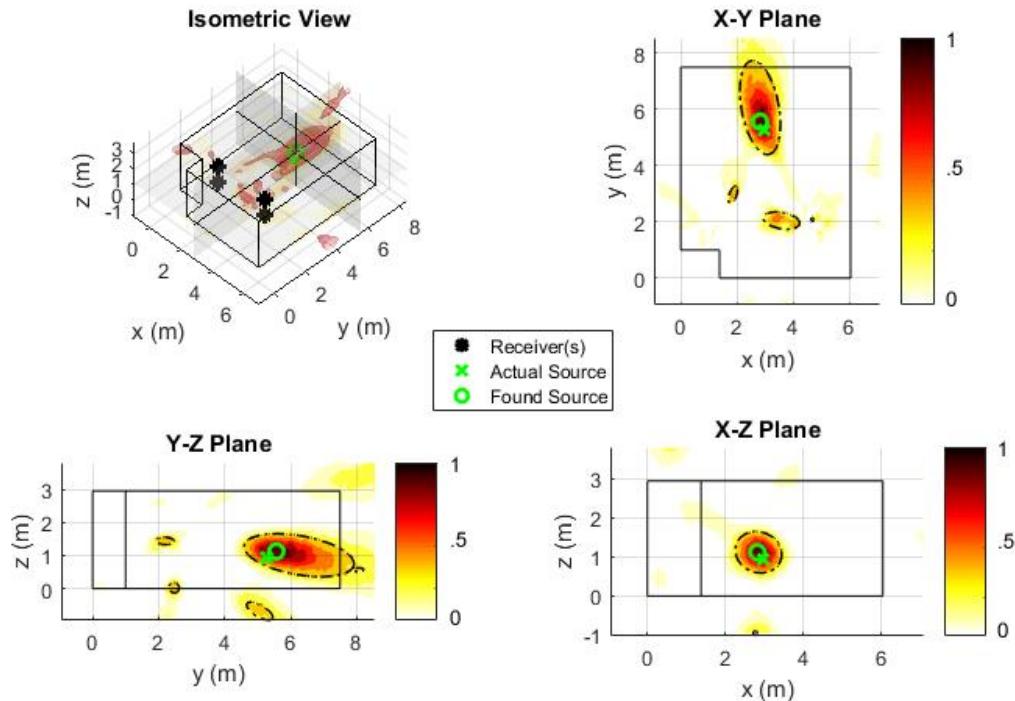
**Figure 31.** Graphical results for ABD receiver configuration. Peak location was at (2.72, 5.77, 0.92); actual source location was (2.98, 5.27, 0.94)



**Figure 32.** Graphical results for ACD receiver configuration. Peak location was at (2.92, 5.87, 1.02); actual source location was (2.98, 5.27, 0.94). An additional (false positive) peak was localized at (2.12, 3.55, 1.53)



**Figure 33.** Graphical results for BCD receiver configuration. Peak location was at (2.72, 5.27, 1.22); actual source location was (2.98, 5.27, 0.94). An additional (false positive) peak was localized at (2.72, 2.34, 1.32)



**Figure 34.** Graphical results for ABCD receiver configuration. Peak location was at (2.82, 5.57, 1.12); actual source location was (2.98, 5.27, 0.94).

**Table 4. Error and resolution comparison between all receiver configurations**

Receiver Configuration	S/R distance (m)	Total error (m)	Normalized error	Average blob length (m)	Normalized blob length
<b>ABC</b>	3.35	0.26	0.077	1.00	0.300
<b>ABD</b>	3.31	0.56	0.170	1.29	0.389
<b>ACD*</b>	3.36	0.61	0.184	1.29	0.383
<b>BCD*</b>	3.32	0.38	0.115	1.01	0.303
<b>ABCD</b>	3.29	0.38	0.117	0.96	0.292

\* False positive source detected

The false positive sources detected in cases ACD and BCD (as seen in Figure 32 and Figure 33, respectively) immediately rule these cases out as effective configurations in this room. Note that both of these cases had two measurements made near the convex section of wall in the room, while the better-performing cases only had one measurement nearby – it is possible that the receivers in closer proximity to the convex feature picked up more interfering reflections, thus lowering localization performance.

Between the two remaining three-receiver cases, ABC (two receivers high, Figure 30) performed better than ABD (two receivers low, Figure 31) in both error and resolution. The difference in performance may be attributed to the difference in material composition between floor and ceiling – the corrugated metal and truss structure on the ceiling would be better at diffusing noise compared with the smooth vinyl flooring, meaning that two measurement positions closer to the ceiling would pick up less specular noise.

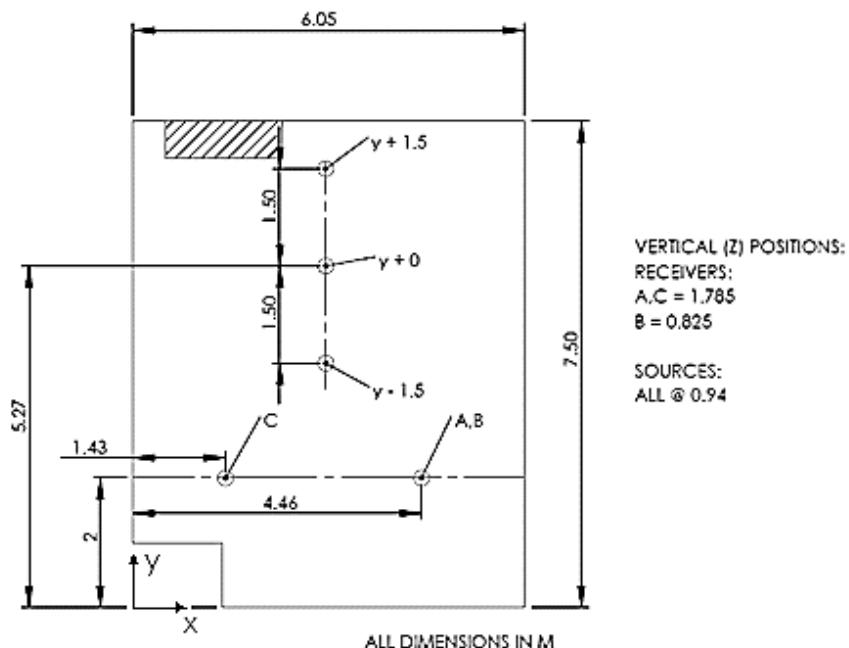
The four-measurement case, ABCD (Figure 34) showed comparable performance to the ABC configuration, but using this configuration would involve an extra measurement location, introducing more room for error in receiver positioning. This is possibly reflected in the higher normalized error in ABCD than in ABC.

From this point onward, the **ABC** configuration will be used as the standard set of receiver locations.

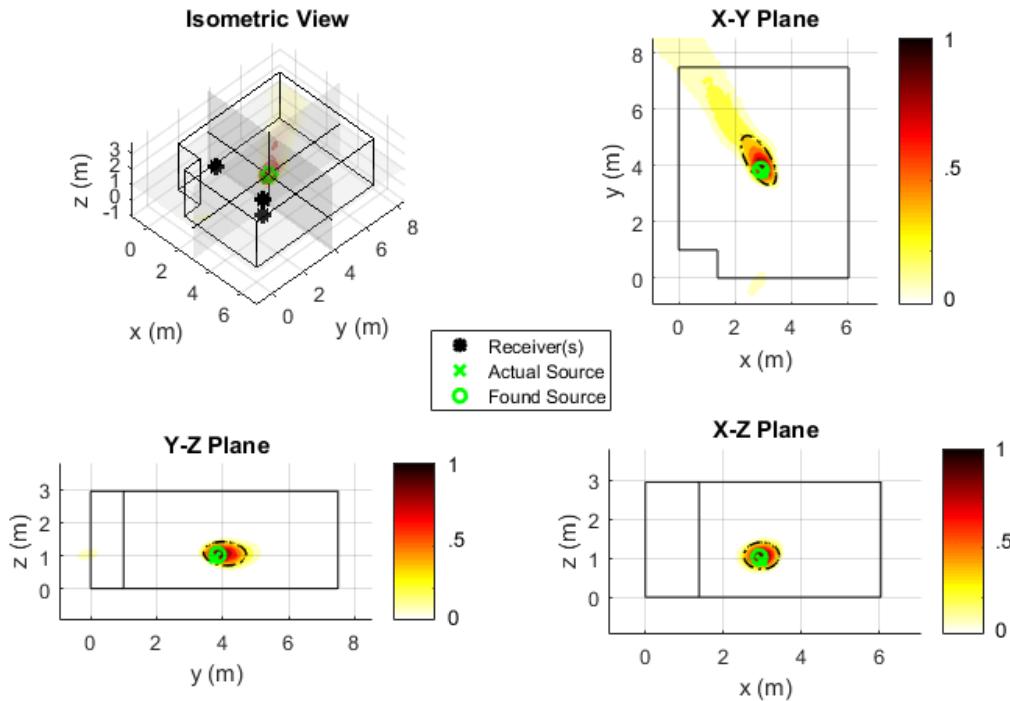
#### 4.5.3 Range Variation

After establishing the standard receiver configuration, a test of source-receiver distance (range) variation was carried out, similar to that done in Section 4.4 comparing the “near” and “far” cases. In the reverberant room, there is the additional variable of the source’s changing proximity to the rear wall when moved in the “Y” direction.

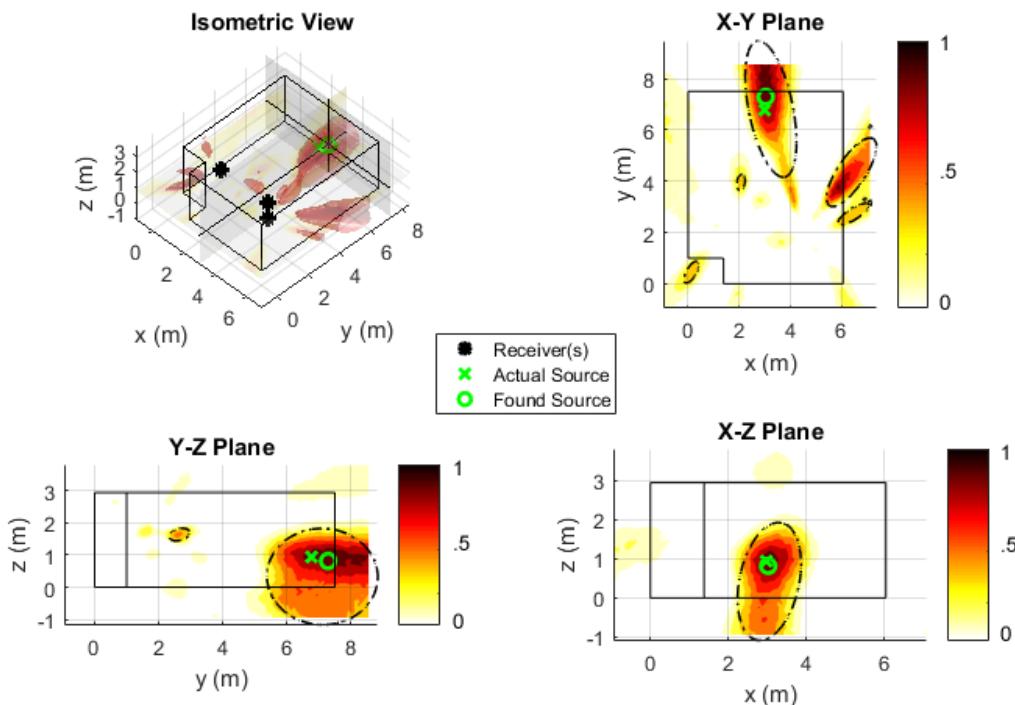
The source location used in the previous experiment was used as the “baseline” position, from which the source was moved forward and backward by 1.5 m in the “Y” direction. Figure 35 shows the source and receiver locations in the test room for this experiment. Figures 36 and 37 show the graphical results for the -1.5m and +1.5m offset cases; the result for the baseline (+0) case was previously shown in Figure 30. The corresponding error and resolution values for all three cases are shown in Table 5.



**Figure 35. Plan view of reverberant test room during distance variation experiment**



**Figure 36.** Graphical results for  $y - 1.5\text{m}$  source position (ABC receiver configuration). Peak location was at  $(2.92, 3.85, 1.02)$ ; actual source location was  $(2.98, 3.77, 0.94)$



**Figure 37.** Graphical results for  $y + 1.5\text{m}$  source position (ABC receiver configuration). Peak location was at  $(3.02, 7.28, 0.82)$ ; actual source location was  $(2.98, 6.77, 0.94)$

**Table 5. Error and resolution comparison between Y-offset cases**

<b>Y offset (m)</b>	<b>S/R distance (m)</b>	<b>Total error (m)</b>	<b>Normalized error</b>	<b>Average blob length (m)</b>	<b>Normalized blob length</b>
<b>-1.5</b>	1.91	0.13	0.067	0.53	0.281
<b>0</b>	3.35	0.26	0.077	1.00	0.300
<b>+1.5</b>	4.82	0.53	0.111	1.52	0.314

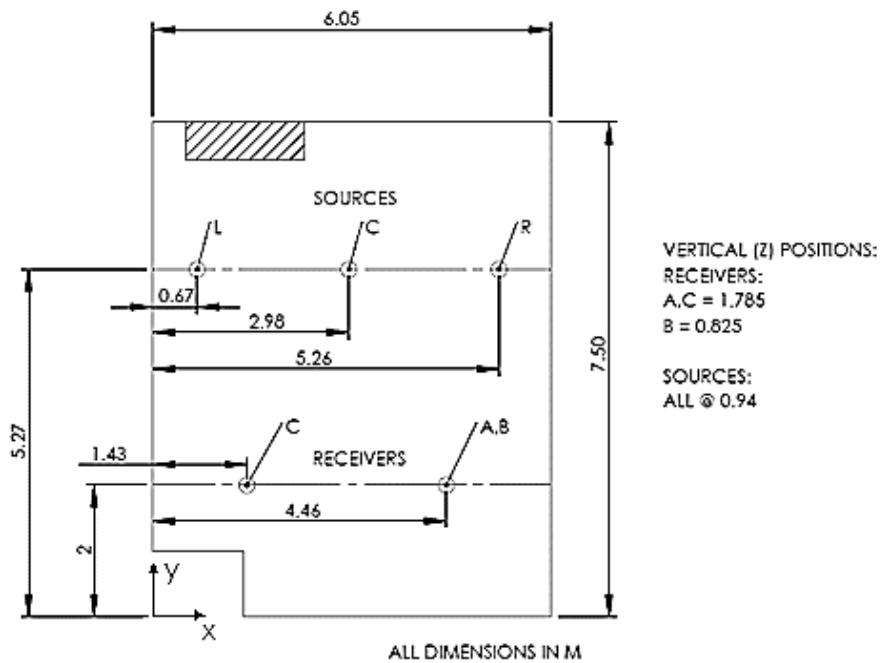
The normalized results for the first two cases in

Table 5 (-1.5 m and 0 m) are comparable in value with the measured free field results (Table 3). The slightly higher normalized error and blob length may be attributed to higher background noise levels and reverberation in the classroom compared to the anechoic chamber.

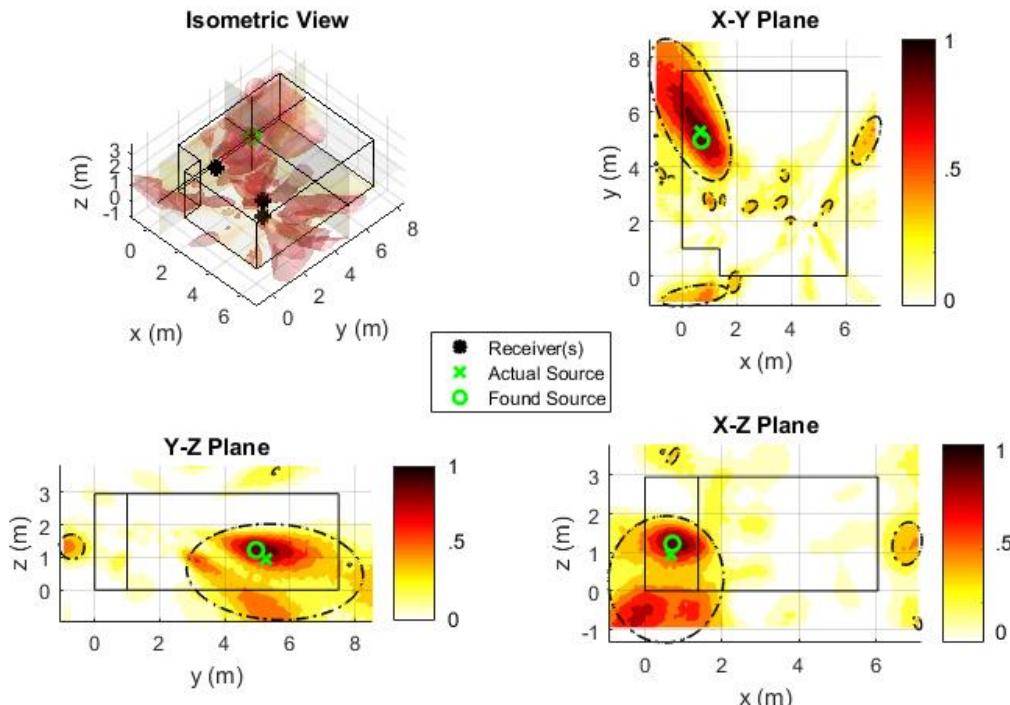
The +1.5m case produced both higher error and blob length – in particular, the blob size in this case as seen in Figure 37 is likely an underestimation, as the blob appears to have been significantly cut off by the bounds of the scanned volume. Examining this figure further, it can be assumed that the larger blob (and, to a lesser extent, the increased peak error) is due to interference from image sources produced by the nearby wall and floor. A more in-depth look at the effect of these image sources can be found below in Section 4.6.

#### **4.5.4 Lateral Position Variation**

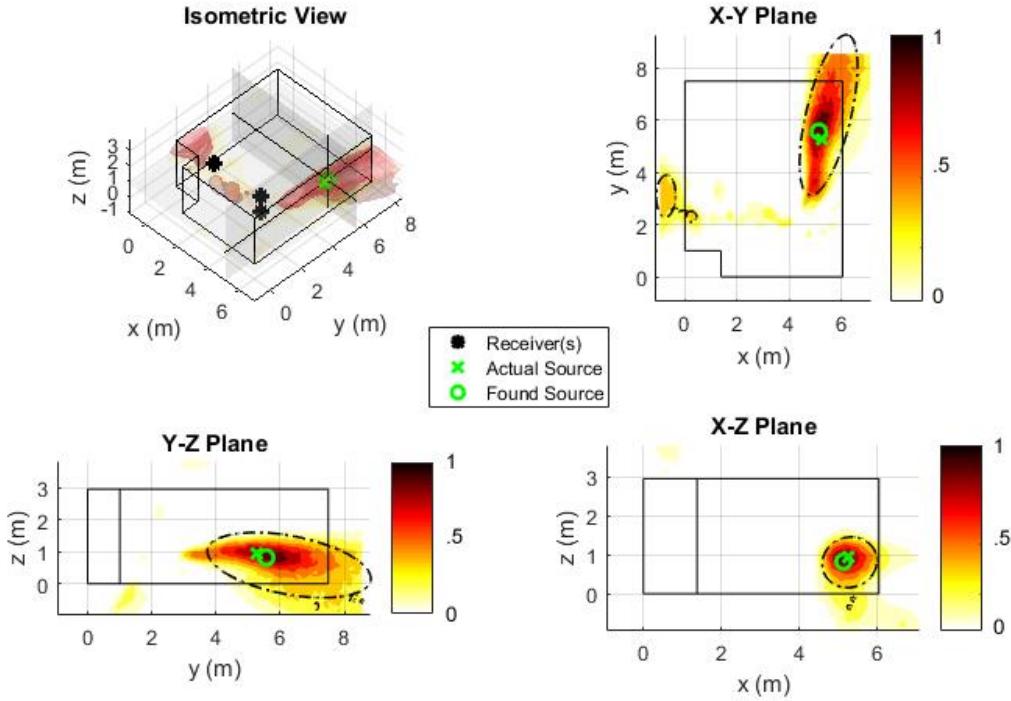
Following the distance variation experiment, a similar experiment varying the lateral (“X”) position was performed. Figure 38 shows the source and receiver locations in the test room for this experiment. Similar to the previous range variation experiment, the central source position (C) is used as the “baseline” position, from which the source was moved approximately 2.3 m to the left (L) and right (R). Unlike the range variation experiment above, both the L and R offset bring the source closer to the side walls of the room. Figures 39 and 40 show the graphical results for the L and R offset cases; the result for the baseline case was previously shown in Figure 30. The corresponding error and resolution values for all three cases are shown in Table 6.



**Figure 38. Plan view of test room during horizontal variation experiment**



**Figure 39. Graphical results for source position L (ABC receiver configuration). Peak location was at (0.71, 4.96, 1.22); actual source location was (0.67, 5.27, 0.94)**



**Figure 40.** Graphical results for source position R (ABC receiver configuration). Peak location was at (5.14, 5.57, 0.82); actual source location was (5.26, 5.27, 0.94)

**Table 6.** Error and resolution comparison between X-offset cases

X position	S/R distance (m)	Total error (m)	Normalized error	Average blob length (m)	Normalized blob length
L	4.32	0.42	0.097	1.87	0.433
C	3.35	0.26	0.077	1.00	0.300
R	3.77	0.34	0.091	1.48	0.391

The results in Table 6 indicate that localization performance is poor at both L and R offsets.

The worsening of resolution may be attributed to the beam widening at oblique angles (a byproduct of the array's hemispherical construction, as discussed in Section 2.4.5). Examining the graphical results further suggests that, similar to in the distance variation case discussed previously, the sources' proximity to the side walls of the room may have resulted in the presence of image sources contributing to the increased error and blob size as well.

## 4.6 Multiple Sources

In section 2.4.4, the issue of resolving multiple sources was discussed in the context of a single receiver position. In the single receiver case, the Rayleigh criterion provides a conservative estimate of the angle,  $\theta_R$  required between two sources in order for both to be well-resolved at a given frequency. It was also shown that, were the requirement to be relaxed to only *minimally* resolve the two sources, a second, lower criterion (the “minimum criterion”),  $\theta_M$  of 18 degrees (as shown in Figure 10 (c)) is defined.

When attempting to apply the same principle to predict the source separation angle required to successfully resolve both sources when triangulating using multiple receivers, the issue is complicated by the fact that there is no single reference point that can unambiguously be called the “receiver position”.

Two possible solutions are:

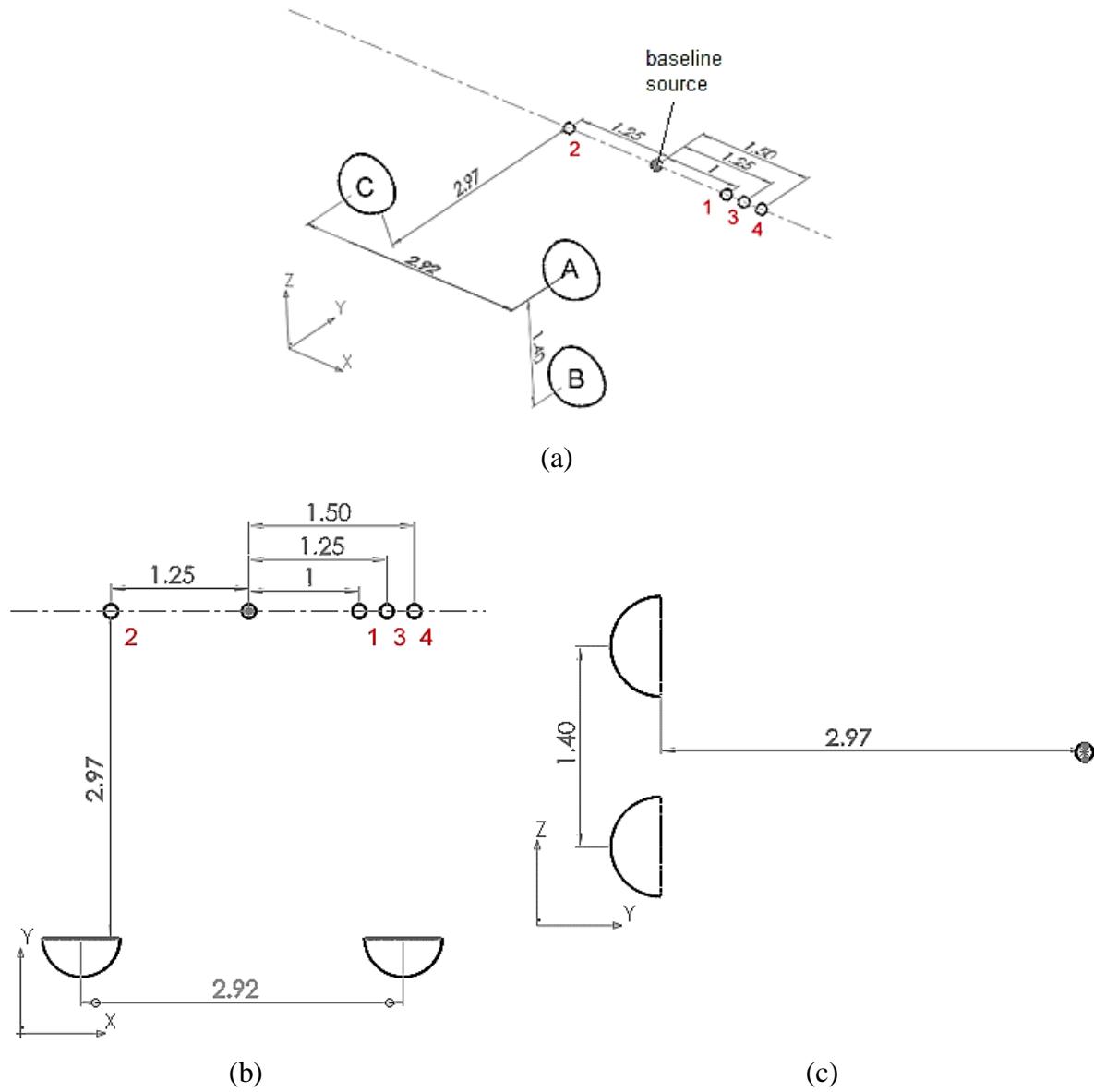
1. To use the centroid as a reference point (as described in Section 3.3.3), or
2. To determine the separation angle at each receiver location individually.

If considering each receiver location individually, there is the possibility that the sources will be able to be resolved at some receiver positions, but not others. The following section aims to investigate what happens in this case.

### 4.6.1 Simulation – Two Sources in Free Field

The setup of the free field simulation is similar to that performed in Section 4.4, with the same receiver configuration and the “far” source position used as one of the two source locations. The second source is then placed at varying distances horizontally to the left and right of the fixed source, to generate four different cases where 0/3, 1/3, 2/3, and 3/3 receivers are predicted to successfully resolve the two sources, as shown in Figure 41. Table 7 summarizes the source

conditions for each case in terms of both individual receiver locations and the centroid position, in terms of “pass/fail” for both the Rayleigh criterion (RC) and the minimum criterion (MC) for resolution.



**Figure 41. Free field two-source source-receiver configuration (a) isometric, (b) top, (c) side views**

**Table 7. Summary of source conditions for two-source free field simulations**

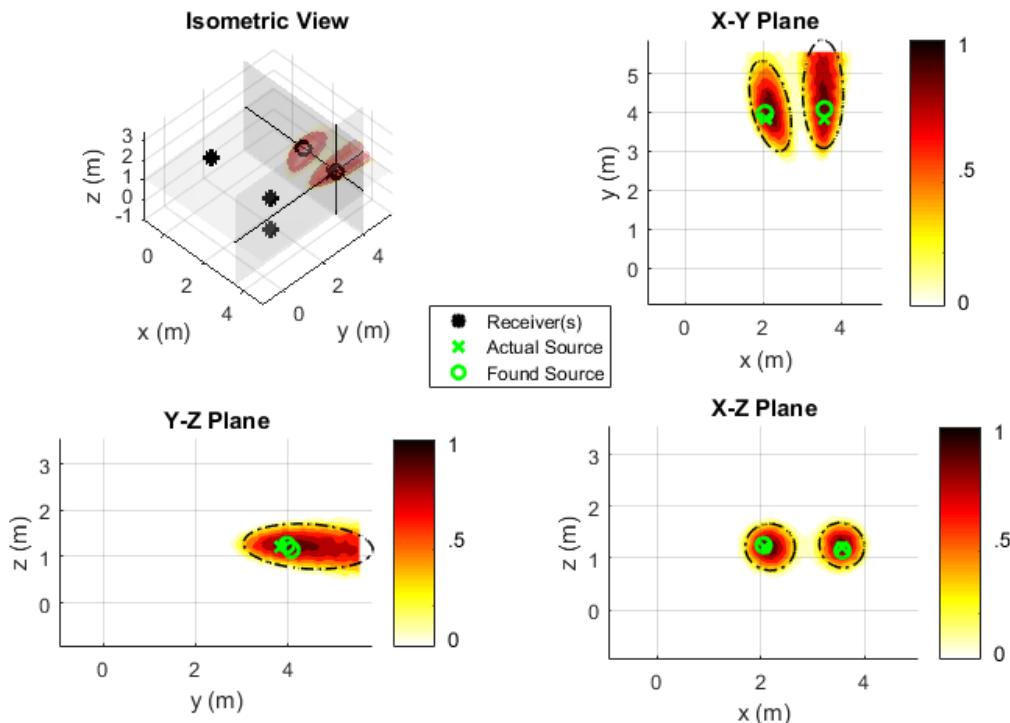
Case	$\theta_M$ (rad)	$\theta_R$ (rad)	Separation angle (rad) from:				# of passing receivers		Centroid pass?	
			A	B	C	Centroid	MC	RC	MC	RC
<b>1</b>	0.314	0.448	0.297	0.299	0.228	0.333	0/3	0/3	Y	N
<b>2</b>	0.314	0.448	0.284	0.285	0.373	0.371	1/3	0/3	Y	N
<b>3</b>	0.314	0.448	0.377	0.380	0.275	0.412	2/3	0/3	Y	N
<b>4</b>	0.314	0.448	0.458	0.462	0.318	0.488	3/3	2/3	Y	Y

Table 8 summarizes the results for all four cases (for the sake of brevity, representative examples of graphical results will be shown below). “Good” localization (i.e. with sources distinct both in terms of separate peaks and blobs, as in Case 4, Figure 42) becomes possible when at least two of three receivers pass the minimum criterion for resolution. Below this point, either both sources are “smeared” into one large blob with a single peak (as in Case 1, Figure 43) or one large blob with two peaks (as in Case 2, Figure 44). Using the centroid position as a reference point for either criterion does not appear to effectively predict whether or not the sources can be resolved – the minimum criterion appears to be too lax (all of the cases pass the prediction test) while the Rayleigh criterion appears to be too conservative (Case 3 fails the prediction test, even though both peaks and blobs are resolved).

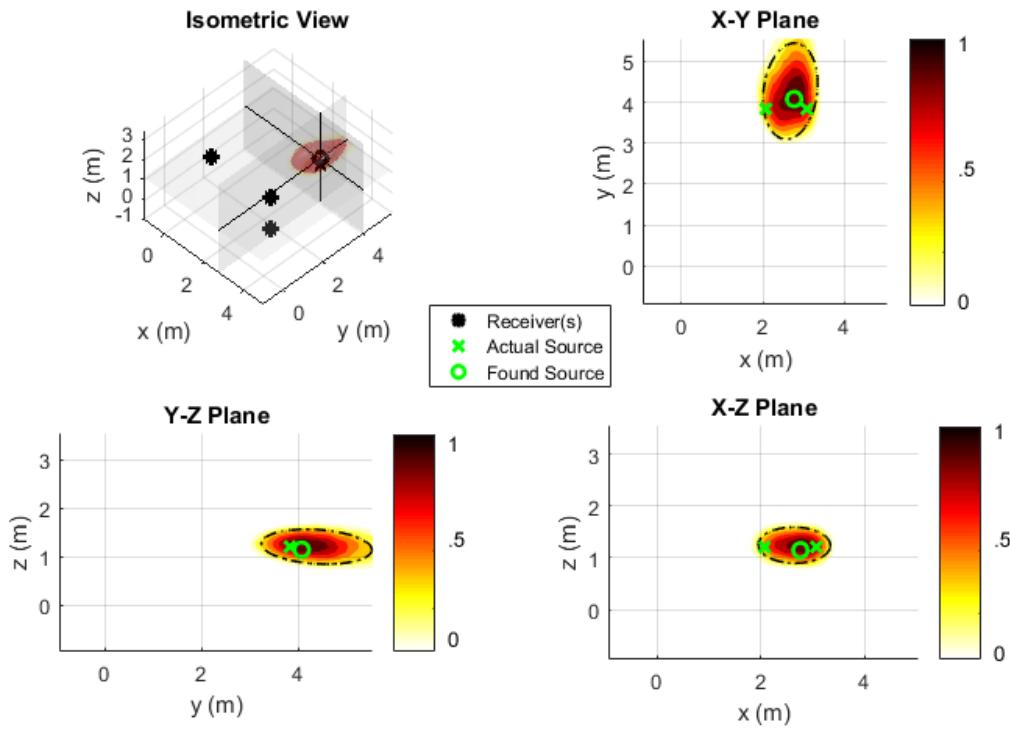
**Table 8. Qualitative summary of results for two-source free field simulations**

Case	# of passing receivers		Centroid pass?		2 peaks found?	2 blobs resolved?
	MC	RC	MC	RC		
<b>1</b>	0/3	0/3	Y	N	N	N
<b>2</b>	1/3	0/3	Y	N	Y*	N
<b>3</b>	2/3	0/3	Y	N	Y	Y
<b>4</b>	3/3	2/3	Y	Y	Y	Y

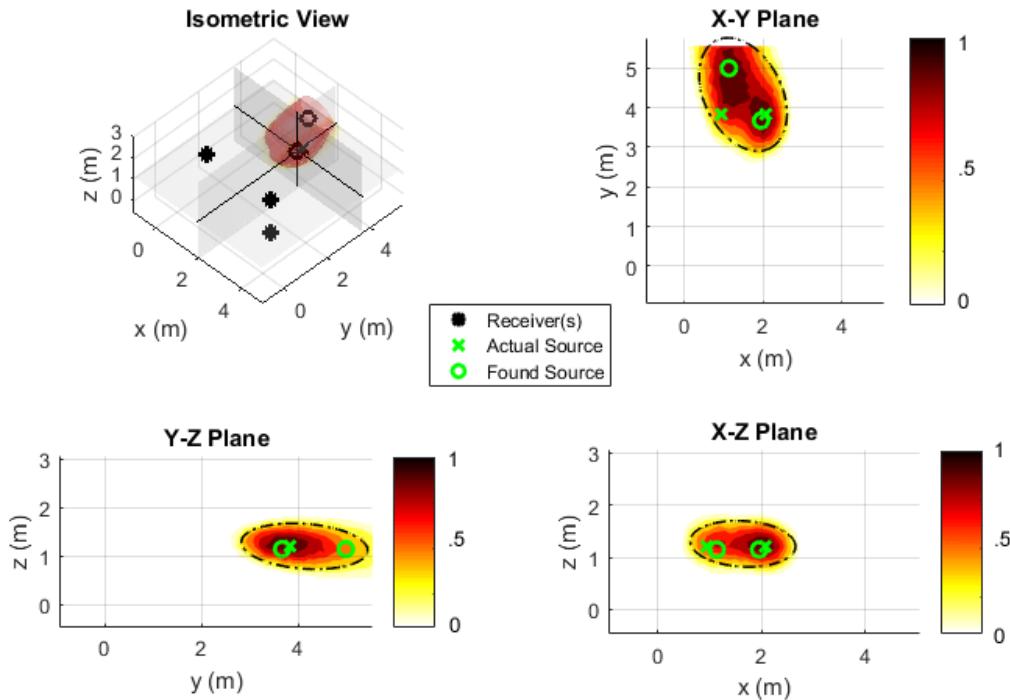
\*High (>1 m) error in locating second source



**Figure 42. Graphical results for two-source Case 4 (“good” localization: 2 blobs resolved and 2 peaks located)**



**Figure 43.** Graphical results for two-source Case 1 (2 blobs unresolved, peaks not successfully located)



**Figure 44.** Graphical results for two-source Case 2 (2 blobs unresolved, but 2 peaks located)

#### 4.6.2 Measurement – Image Sources in Reverberant Room

A common situation where multiple sound sources may be localized in a room – even if there is only one actual sound source present – is when image sources occur, caused by specular reflections from walls or other surfaces of the room. The presence of these image sources was alluded to previously in Sections 4.5.3 and 4.5.4, as a possible cause for decreased localization accuracy and resolution in some cases where the source was positioned close to a wall.

When working with image sources (as opposed to multiple independent sources, as above), there is the additional issue of the relative source levels between the direct and reflected sound. The amount by which the reflected source is attenuated from the direct source is a function of both absorption of the reflection surface, and the difference in free field path length between source/image source and receiver. For first-order reflections, this attenuation can be expressed as:

$$\Delta L_p = 20 \log_{10} \left( \frac{(1 - \alpha)a}{b} \right) \quad (17)$$

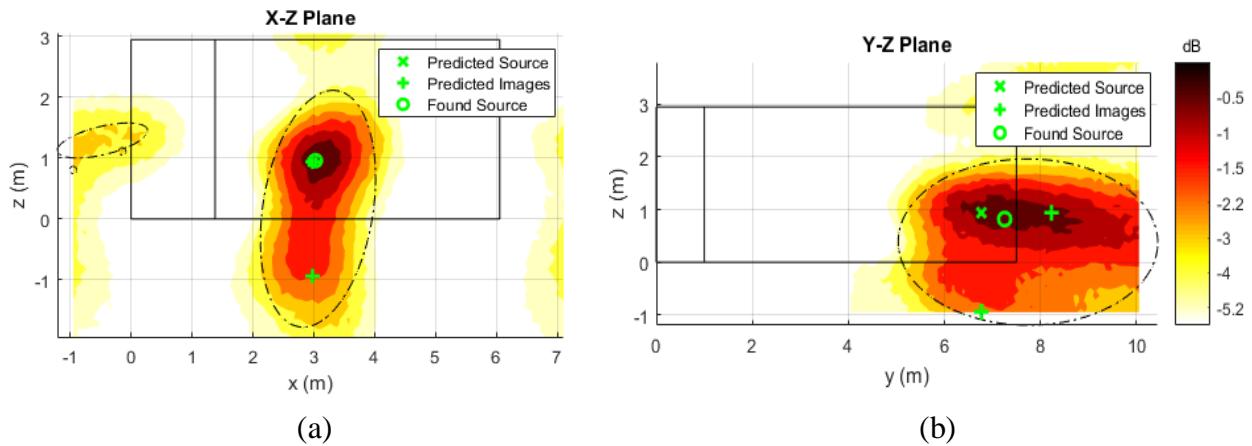
Where  $\Delta L_p$  is the difference in sound pressure level (dB) between the direct and image source,  $a$  is the path length from direct source to receiver,  $b$  is the path length from image source to receiver, and  $\alpha$  is the absorption coefficient of the reflection surface. Therefore, in order for an image source to have an adverse effect on source localization, two conditions must be satisfied:

1. The direct and image sources are close enough together in position such that they fail to meet the minimum/Rayleigh criterion, as discussed in Section 4.6.1, and
2. The direct and image source are close enough together in SPL (i.e.  $\Delta L_p$  is small) such that the image source has a non-negligible effect on the direct source level.

If condition 1 is satisfied but not condition 2, the image source level is low enough that it will cause a negligible effect to the final source localization result. If condition 2 is satisfied but not condition 1, the image source may be visible if the scanning volume contains points outside of

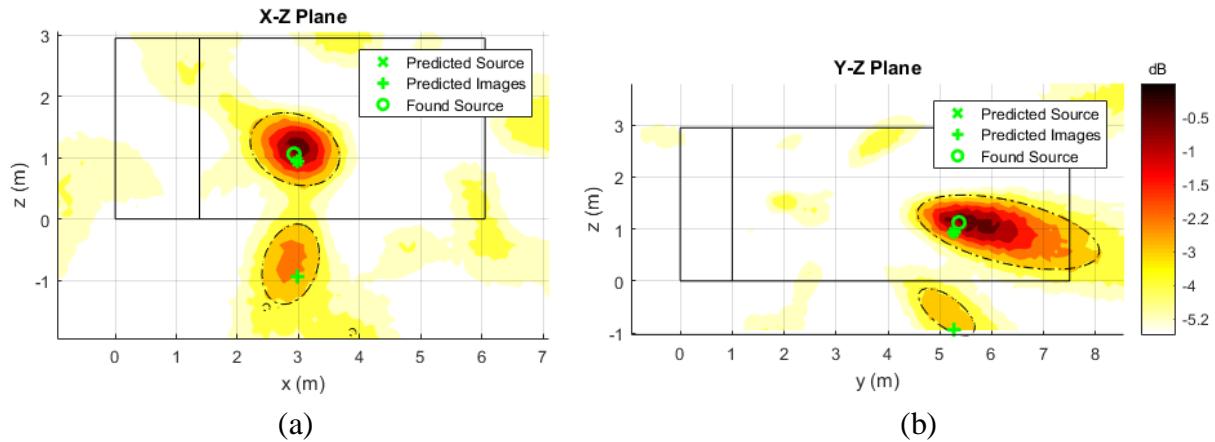
the room volume, but it will appear as an entirely separate source that can then be trivially excluded, since it is not possible for a source to be located outside the bounds of the room (transmission of sound between rooms is outside the scope of this thesis).

An example of a measurement case where image sources interfered with source localization can be found in Section 4.5.3 – in the  $y + 1.5$  m case (Figure 37), this interference can clearly be seen in the X-Z and Y-Z planes, shown in greater detail below (Figure 45). Based on the geometry of the test room, it can be inferred that the interfering image sources are caused by the floor and back wall of the room.



**Figure 45. Detailed view of image sources in  $y + 1.5$  m case (original Figure 37) in (a) X-Z plane, (b) Y-Z plane**

An example of a case where an image source is present, but non-interfering, can be found in the baseline case first introduced in Section 4.5.2. This image is inferred to be caused by a reflection from the floor, and is, again, easily seen in the X-Z and Y-Z planes, and shown in greater detail in Figure 46. Here, the direct and image sources appear as two distinct blobs, with no apparent effect on peak localization accuracy.



**Figure 46. Detailed view of image source in baseline case (original Figure 30) in (a) X-Z plane, (b) Y-Z plane**

The results in this section are summarized by comparing the predicted and actual outcomes for source resolution (condition 1) and relative source level (condition 2) in Tables 9 and 10, respectively. Table 9 shows that minimum/Rayleigh criteria are not necessarily fixed thresholds – in particular, the  $y+1.5m$  floor image (first row of the table) had passed the minimum criterion for resolution, but Figure 45 shows that the floor image blob was not able to be entirely resolved separately from the direct source. In both cases where direct and image source levels were able to be estimated and compared, the predicted (from equation (17)) and measured (from Figures 45 and 46) attenuation was in good agreement.

**Table 9. Comparison between predicted and actual outcomes for source resolution**

Case	Separation angle (rad) from				# of passing receivers		Centroid pass?		2 blobs resolved?
	A	B	C	Centroid	MC	RC	MC	RC	
<b>Y+1.5m: floor</b>	0.332	0.363	0.331	0.356	3/3	0/3	Y	N	N
<b>Y+1.5m: wall</b>	0.076	0.068	0.078	0.034	0/3	0/3	N	N	N
<b>Baseline: floor</b>	0.418	0.489	0.416	0.472	3/3	2/3	Y	Y	Y

**Table 10. Comparison between predicted and actual outcomes for relative source level**

Case	Direct Path (m)	Image Path (m)	Absorption Coefficient	Predicted dB Attenuation	Actual dB Attenuation
<b>Y+1.5m: floor</b>	4.82	5.36	0.03 <sup>1</sup>	-1.06	-1.01
<b>Y+1.5m: wall</b>	4.82	6.27	0.02 <sup>2</sup>	-2.37	N/A
<b>Baseline: floor</b>	3.35	4.09	0.03 <sup>1</sup>	-1.87	-1.84

<sup>1</sup>Linoleum on concrete [25]

<sup>2</sup>Brick, unglazed and painted [25]

#### 4.7 Discussion

As would be reasonably expected, it was found that sources get more difficult to triangulate (i.e. higher error, lower resolution) the farther they are from the receiver plane, both in free field and in a reverberant room. Localization performance also drops (though less dramatically) when the source is very close to the receiver plane, due to the skew effect as explained in Section 4.4.2. In the reverberant room, reverberation in the form of image sources from the walls and floor were occasionally an issue, affecting both source resolution and accuracy in cases where the source is simultaneously far from the receivers and close to a wall (i.e. the direct sound and reflected sound are similarly loud). It is worth noting that, in a practical source localization application, one has no control over the location of the source(s) in a given environment – the only aspect in the operator's control is the positioning of the receivers relative to each other and the room.

Regarding the positioning of receivers: the rationale for keeping all receivers in the same plane for this experiment was twofold: first, it kept the data acquisition procedure (i.e. positioning the arrays relative to the room) straightforward. Secondly, it simulated the feasibility of, for instance, suspending the array from the ceiling of a large workroom. Of interest in the future would be a receiver configuration more evenly distributed around the room (e.g. taking a measurement at each room corner). Distributing the receivers would also help identify sources that would potentially be obstructed from a single vantage point. It may be beneficial to keep the array away

from potentially intrusive features (such as the convex wall section in the reverberant room discussed above).

#### 4.7.1 Sources of Uncertainty

In addition to the discretization error discussed in Section 3.3.2, three additional sources of uncertainty were identified, as follows:

1. **Assumption of the speed of sound:** Throughout the calculations performed in this section, a speed of sound  $c = 341$  m/s was assumed, corresponding to a temperature of approximately  $15^\circ\text{C}$  at sea level. This was judged to be a fairly accurate estimate of temperature conditions in the anechoic chamber and reverberant classroom, but was not actually verified when measurements were being conducted. If the temperature deviated significantly from the assumed conditions, the resulting error in  $\Delta t_j$  would produce an angular error in localization. A straightforward way to eliminate this error would be to record the actual temperature at the time of measurement and use the measured temperature to compute the actual speed of sound for each measurement.
2. **Uncertainty in microphone positions within the array:** the relative positions of the array microphones were assumed to be the idealized positions shown in Appendix A (Table 18). However, a series of measurements performed prior to the beginning of the experiments showed that there was an average discrepancy of 0.86 cm ( $\sigma = 0.37$  cm) between the ideal and measured microphone positions, with the highest error being 1.41 cm (at microphone 20). Due to the slight non-rigidity of the array connectors and microphone fittings, the physical position of the microphones may have shifted between measurements as the array was being moved from location to location, which would again lead to mis-estimation of  $\Delta t_j$ . Two possible ways to mitigate this

uncertainty would be to either improve the rigidity of the array fittings (already attempted once with the addition of the steel frame [23]) or to develop and use some sort of calibration apparatus (e.g. a physical template to verify microphone alignment, or a calibration routine using a source at a known position) that could be used prior to taking a measurement.

3. **Uncertainty in positioning of source and receiver:** the positioning of the source and receiver within the test room (anechoic or reverberant) was done using a tape measure against a constant reference point. Care was taken to make the measurements repeatable, by marking out locations with tape, aligning them with well-defined locating features on the source and array and using repeatable positioning methods (e.g. counting ratchet clicks) whenever possible. However, the possibility of error in the relative positioning of the source and receiver still remains, due to the uncertainty of the tape measure and construction of the room (i.e. unevenness of the floor, walls not being parallel, etc). Related to this error is the uncertainty in angle of orientation of the hemisphere (i.e. whether or not the broad side of the array was parallel to the receiver plane) – note that the “measuring tape error” is an absolute quantity (approximated as  $\pm 5$  cm overall) that does not vary with source-receiver distance, while the angular error will increase with range.

## **Chapter 5: Image Source Triangulation - Background**

This chapter provides the background information specific to the image source triangulation method of source localization. First, an overview of related work will be provided. Then, the practical implementation of image source triangulation will be described, and various practical considerations arising from this implementation will be discussed.

Note that the terms “image source triangulation” (IST) and “image source beamforming” (ISB) will be used interchangeably from here onwards, as the connection between ISB and triangulation is an idea novel to this thesis – previous work has referred to the ISB concept exclusively.

### **5.1 Related Work**

The idea behind the image source beamforming method is to use prior knowledge of the room geometry to predict where image sources would be, and incorporate this information into the beamforming model in order improve performance in the presence of reverberation. This idea has received particular attention in the field of aeroacoustics, primarily for the mitigation of reverberation that occurs in wind tunnels (e.g. [24], [27], [28], [29]) – in general, these works have started with the ISM as a way to predict the impulse response of the wind tunnel, and incorporated this impulse response into various SSL algorithms in the frequency domain (e.g. frequency domain beamforming, CLEAN-SC, DAMAS). In underwater acoustics and geoacoustics, the image source model has been used in [30] to predict the thicknesses of layers of the sea floor, and in [31] to localize sound in shallow water by modeling reflections from the sea floor and water surface.

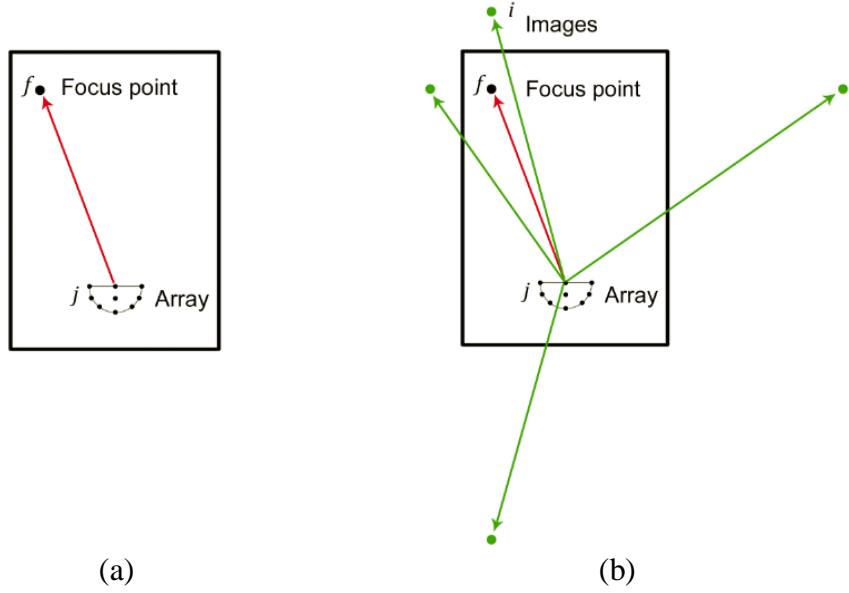
In room acoustics, though the general idea of compensating for reverberation in SSL has been a topic of interest for some time (e.g. [32]), the idea of specifically using the specular reflections predicted by the image source model in a room is a fairly new development. Recently,

[33] has demonstrated that localization with a small array in a conference room using maximum likelihood estimation can be improved by incorporating information from the image source model.

It is also worth noting that the *inverse* task of image source triangulation – estimation of room geometry based on measured acoustical characteristics – has been a subject of some recent research interest as well (e.g. [34], [35]).

## 5.2 Implementation

Like with direct triangulation, the ISB method begins with conventional delay-and-sum beamforming (CBF), the output of which is expressed in Eq. (5). The key aspect of ISB that distinguishes it from CBF is that it also incorporates the potential image sources associated with each  $\mathbf{r}_f$ . This distinction is illustrated in Figure 47. Note that the image source model relies of the assumption that the reflections from the walls of the room are specular, i.e. the walls are rigid and planar.



**Figure 47.** (a) Single focus point used in CBF versus (b) focus point and images used in ISB

For the set of  $N$  image sources associated with  $r_f$ , the additive ISB output can be written as follows:

$$\begin{aligned} z_{ISB,add}(\mathbf{r}_f, t) &= z_{DAS}(\mathbf{r}_f, t) + \sum_{i=1}^N (1 - \alpha_i)^{\beta_i} z_{DAS}(\mathbf{r}_i, t) \\ &= \sum_{j=1}^M s \left( t + \frac{|\mathbf{r}_f - \mathbf{r}_j|}{c} \right) + \sum_{i=1}^N \sum_{j=1}^M (1 - \alpha)^{\beta_i} s \left( t + \frac{|\mathbf{r}_i - \mathbf{r}_j|}{c} \right) \end{aligned} \quad (18)$$

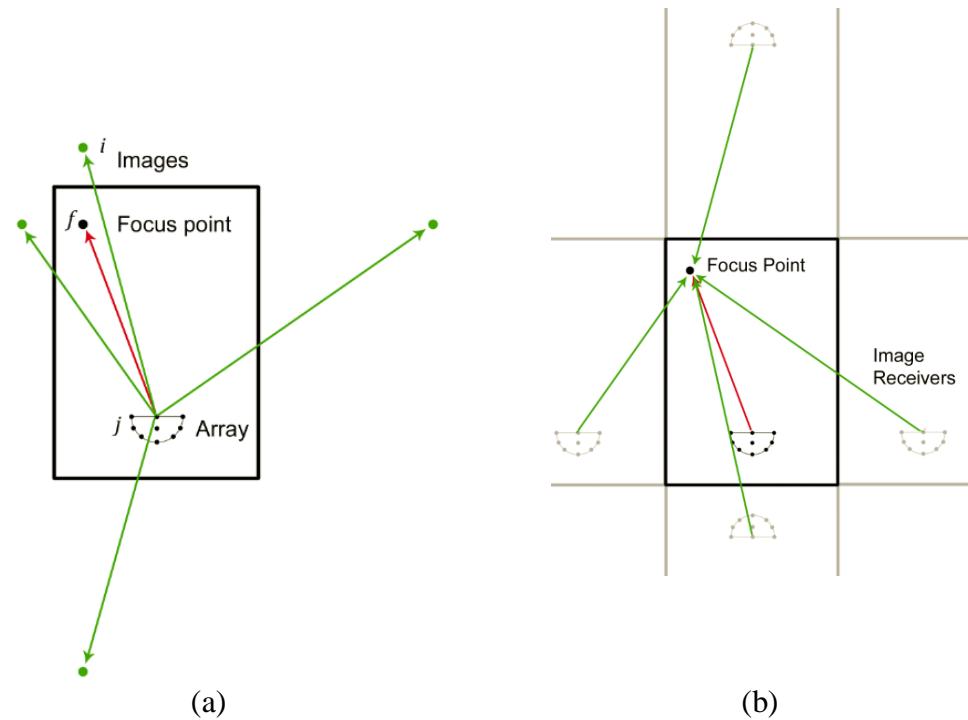
The  $(1 - \alpha)^{\beta_i}$  term is an adjustment to take into account the absorption of the wall from which the  $i$ th image originates:  $\alpha$  is the absorption coefficient of the wall (assumed to be constant for all walls in this case), and  $\beta_i$  is the associated image order, i.e. the number of reflections that have to occur before the image appears (the effect of image order is discussed further below in Section 5.3.2).

As was the case with direct triangulation, the combination of beamformer outputs can be performed either additively, as above, or multiplicatively, as in (19), and time averaging can be performed using the mean square or root mean square as was done in the direct triangulation case; the formulation for this is easily obtained from (10) or (11) and is therefore omitted below.

$$z_{ISB,mult}(\mathbf{r}_f, t) = \sum_{j=1}^M s \left( t + \frac{|\mathbf{r}_f - \mathbf{r}_j|}{c} \right) * \prod_{i=1}^N \sum_{j=1}^M (1 - \alpha_i)^{\beta_i} s \left( t + \frac{|\mathbf{r}_i - \mathbf{r}_j|}{c} \right) \quad (19)$$

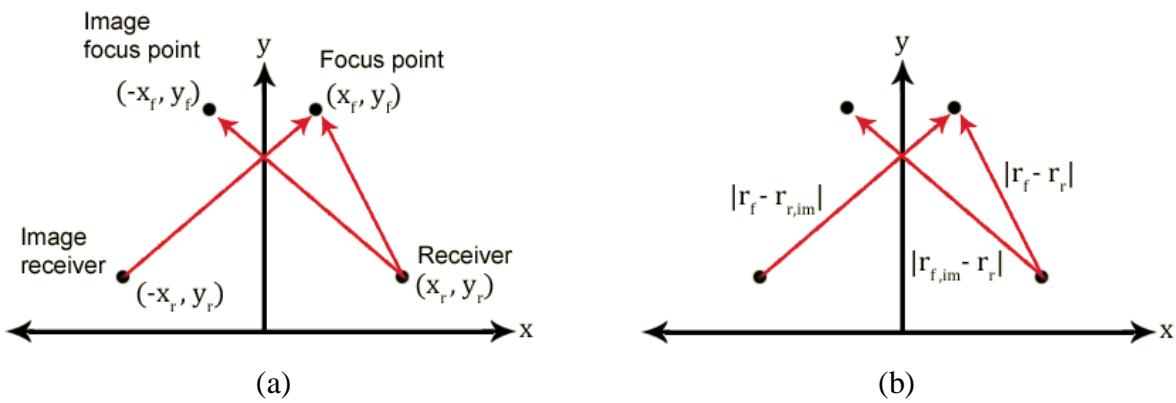
### 5.2.1 Equivalence between ISB and Triangulation

The aim of this section is to show that the classical interpretation of ISB – the combination of  $N$  image source points for a focus point  $r_f$ , as discussed above – is equivalent to triangulation between  $N$  image *receivers* (i.e. receivers mirrored about the walls of the room) at  $r_f$ . This concept is first presented graphically in Figure 48, with a more detailed explanation below.



**Figure 48. (a) Regular ISB interpretation in 2D, (b) equivalent image receiver triangulation for a single focus point  $f$**

In order to prove the equivalence between Figure 48 (a) and (b), a simplified example is shown below in Figure 49: in this case, there exists a single receiver and reflection axis with no losses due to absorption. The receiver is located at  $\mathbf{r}_r = (x_r, y_r)$ , and the focus point at  $\mathbf{r}_f = (x_f, y_f)$ . Mirrored about  $x = 0$ , the image focus point is  $\mathbf{r}_{f,im} = (-x_f, y_f)$  and the mirrored image receiver is  $\mathbf{r}_{r,im} = (-x_r, y_r)$ .



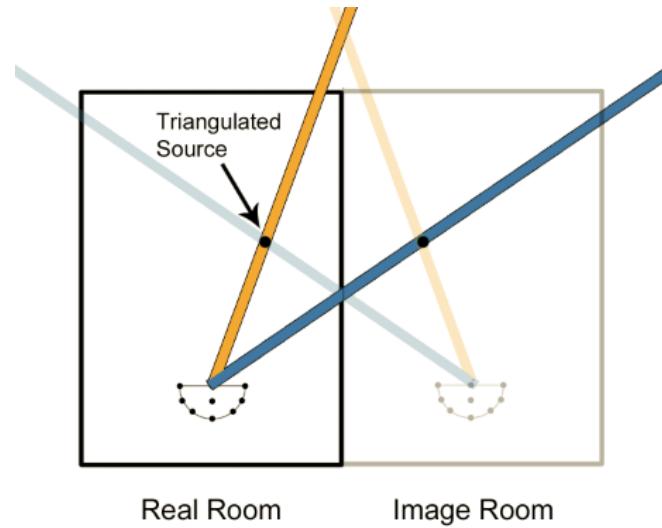
**Figure 49. Example ISB case to demonstrate triangulation equivalence, with (a) points labelled and (b) vectors labelled**

In this case, the ISB and direct triangulation outputs at  $r_f$  are, respectively,

$$z_{ISB}(\mathbf{r}_f, t) = s \left( t + \frac{|\mathbf{r}_f - \mathbf{r}_r|}{c} \right) + s \left( t + \frac{|\mathbf{r}_{f,im} - \mathbf{r}_r|}{c} \right) \quad (20)$$

$$z_{DT}(\mathbf{r}_f, t) = s \left( t + \frac{|\mathbf{r}_f - \mathbf{r}_r|}{c} \right) + s \left( t + \frac{|\mathbf{r}_f - \mathbf{r}_{r,im}|}{c} \right) \quad (21)$$

From the above figures, it is clear that  $|\mathbf{r}_{f,im} - \mathbf{r}_r| = |\mathbf{r}_f - \mathbf{r}_{r,im}|$  (i.e. the magnitude of a vector and its mirror image are equal), and therefore  $z_{ISB} = z_{DT}$ . The image triangulation concept is again illustrated in Figure 50, with a single image source and receiver. Note that successful triangulation of the source is only possible if the image source (blue line) is localized – therefore, the effectiveness of ISB extends only as far as image sources can be localized in the original signal.



**Figure 50.** Image source triangulation with a single image receiver. Orange lines represent the beamforming output due to the direct source and dark blue lines represent the output due to the image source (beams illustrated as straight lines for clarity – in practice, beams would be triangular/conical in shape radiating from the receiver)

### 5.3 Practical Considerations

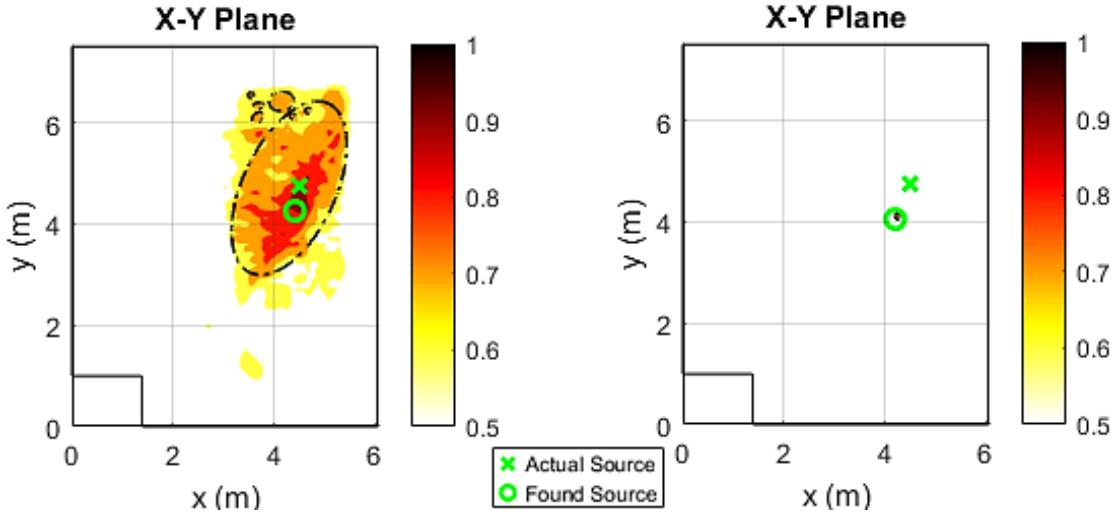
With the equivalence between ISB and triangulation established, many of the practical considerations discussed in Section 3.3 can be directly transferred to the implementation of ISB.

However, there are specific concerns that arise in ISB stemming from the fact that the receivers in this case (aside from the original) are mirrored versions of the original receiver output.

### 5.3.1 Multiplicative vs Additive Image Triangulation

When considering the merits of multiplicative versus additive triangulation in the case of ISB, the distinction between having multiple “real” receivers in direct triangulation, as opposed to mirrored “image” receivers in ISB becomes significant. The effectiveness of ISB extends only as far as image sources can be localized in the original signal – that is, if an image receiver is incorporated into the ISB output sum but its corresponding image source cannot be resolved in the original signal, the contribution from that image receiver will have negligible or negative effect on the final outcome.

Since the problem of predicting which images will or will not be resolved is non-trivial (as discussed further below), in the case of ISB it is necessary to use additive triangulation to combine the signals from the image receivers. Figure 51 shows an example of the ISB output comparing additive and multiplicative triangulation for the same set of data (from the “background noise” case, discussed in Section 6.6). While the multiplicative output indeed produces a very precise peak, this precision is misleading in that the peak is not accurate in terms of locating the actual source position. Therefore, additive combination is preferred for when working with ISB.

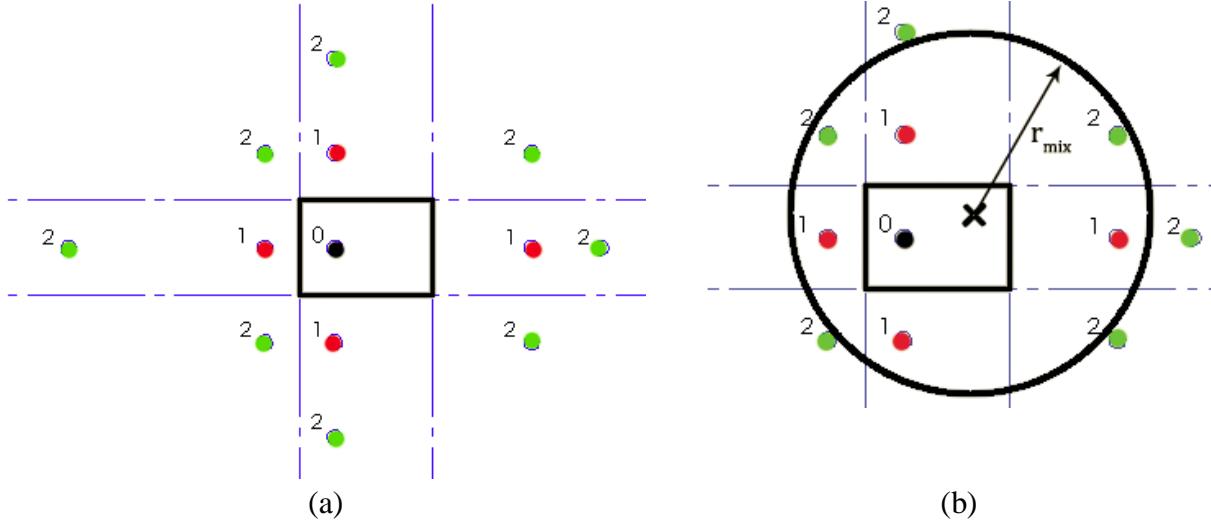


**Figure 51. Comparison between additive and multiplicative ISB output (data from “baseline ISB” case – Section 6.4)**

### 5.3.2 Image Order Estimation

The “order” of an image source model refers to the number of times the sources are reflected about the room’s walls – a first-order model will take the original sound source and reflect it about the room’s walls, a second-order model will treat the first-order reflections as sources and reflect them about the room’s walls, and so on. Figure 52 shows the images of a sound source up to order 2 for a rectangular room. In an image source model, these images are propagated up to a certain radius  $r_{mix}$ , extending from the receiver location. This radius is related by the speed of sound to the mixing time,  $t_{mix}$ , associated with a particular room:

$$r_{mix} = t_{mix} * c \quad (22)$$



**Figure 52.** (a) All image sources up to order 2, (b)  $r_{mix}$  extending from receiver to propagation boundary. The number adjacent to each source indicates its order (order 0 = direct source)

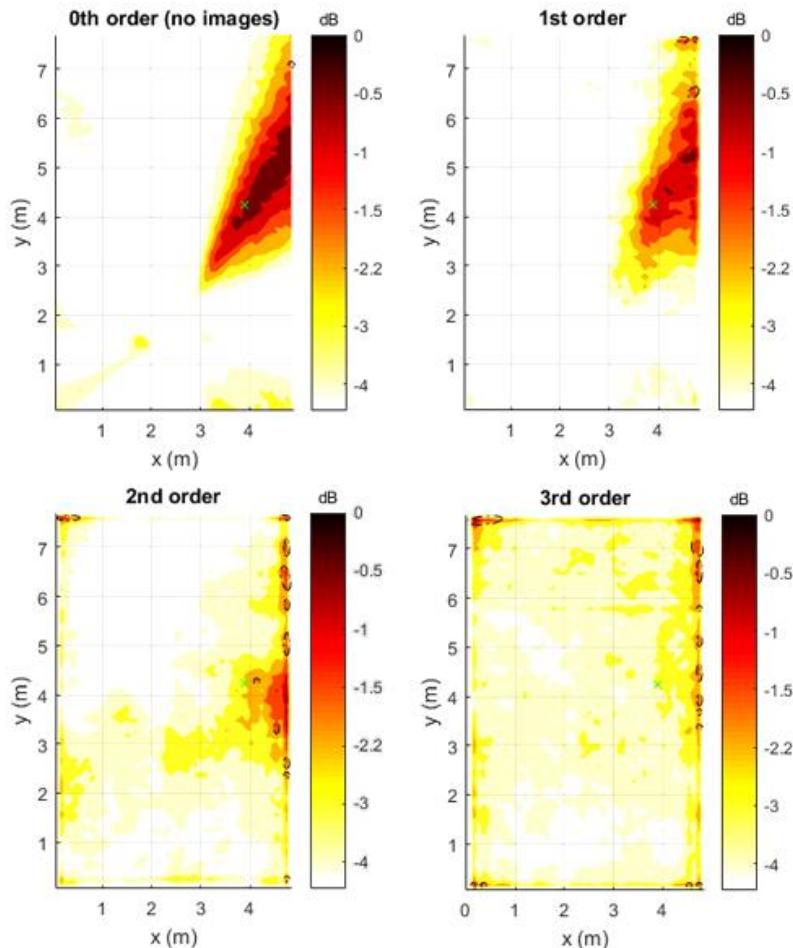
The mixing time is defined as the (estimated) point of transition between a sound field that can be described by a number of specular sources (as in an image source model) and a sound field that is diffuse and can be considered “late reverberation” [36], where the sound energy is equally distributed throughout the room [37].

One way to determine the number of image sources to incorporate in the ISB model would be to use as many as are estimated to occur according to the mixing time/radius – in this way, the model would be maximizing the amount of “useful information” extracted from the original sound field data. However, current mixing time estimation methods place an emphasis on perceptual mixing (i.e. whether something “sounds mixed” to the human auditory system) [26], instead of a true physical definition of diffuseness. Furthermore, this definition of mixing time is only truly applicable to impulsive or transient signals – in the case where a stationary, steady-state sound source is present (such as those presented in this thesis), one can argue that there is always a combination of specular and “mixed” sound present in the room. With this in mind, in conjunction with the fact that edge artifacts (discussed below) become more and more prominent as the number of images incorporated increases, an “ideal” number of images may not exist – instead, a study on

the effect of image order will be conducted in Section 6.4, as part of establishing the “baseline case” image source model.

### 5.3.3 Edge Artifacts

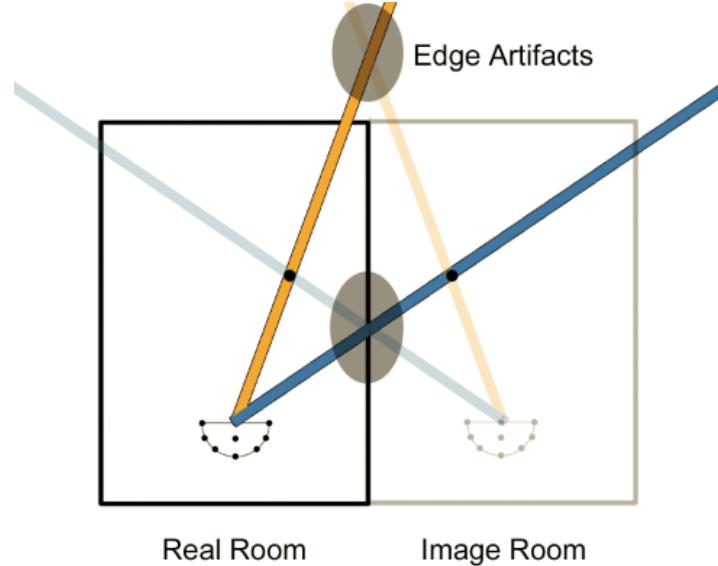
A phenomenon that is observed in the ISB output plots are edge artifacts – areas of high output intensity focused around the edges of the room. These artifacts increase in severity as image order increases, as shown in Figure 53.



**Figure 53. ISB output from 0<sup>th</sup> to 3<sup>rd</sup> order, no windowing (data from “baseline ISB” case – Section 6.4)**

Edge artifacts are an inherent byproduct of the image triangulation method, as the mirror images of a source beam will always intersect at the line about which it is mirrored, creating areas of high intensity as shown in Figure 54. This is similar to the “phantom source” problem in direct

triangulation, discussed in Section 3.3.3. In this case, the mirror lines about which images are created are the room boundaries, thus creating edge artifacts.

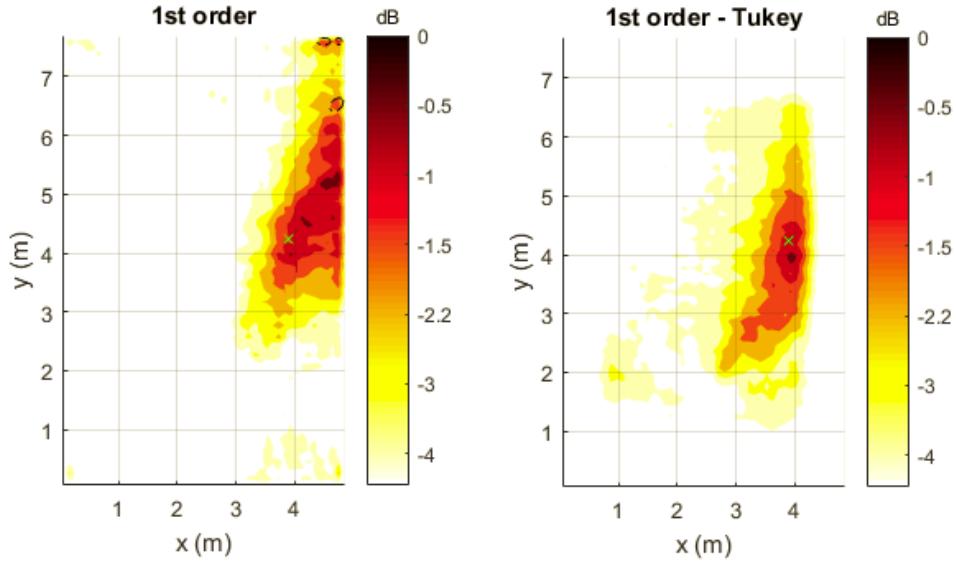


**Figure 54. Image source triangulation with a single image receiver, edge artifact regions shaded**

In order to attenuate these artifacts, a Tukey (tapered cosine) window can be applied to the ISB output, with the tradeoff being decreased ability to identify sources near the edge of the room. A similar approach was taken to reduce edge artifacts in [24]. The Tukey window is defined as:

$$w(x) = \begin{cases} \frac{1}{2} \left[ 1 + \cos \left( \frac{2\pi}{q} \left[ x - \frac{q}{2} \right] \right) \right], & 0 < x < \frac{q}{2} \\ 1, & \frac{q}{2} \leq x < 1 - \frac{q}{2} \\ \frac{1}{2} \left[ 1 + \cos \left( \frac{2\pi}{q} \left[ x - 1 + \frac{q}{2} \right] \right) \right], & x \geq 0 \end{cases} \quad (23)$$

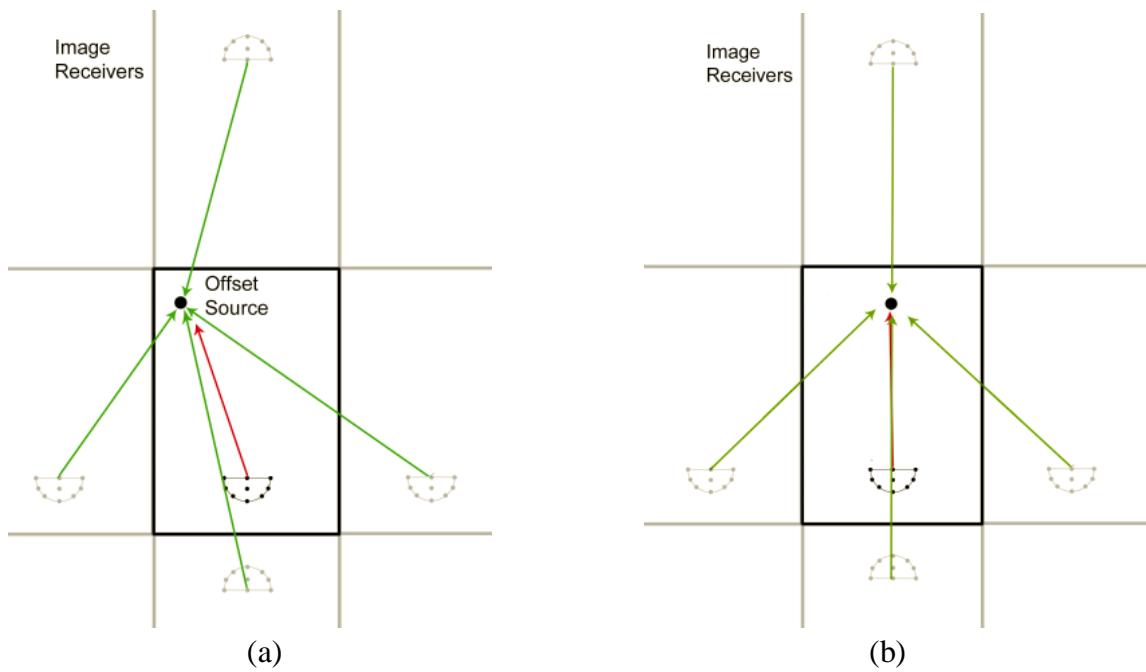
For the purposes of the experiments following, a 3D Tukey window with  $q = 0.3$  was used, with the consequence that 34.3% of the original data volume was untouched, and 61.4% was attenuated by less than 50% of its original value. An example of the ISB output before application of the Tukey window is shown in Figure 55.



**Figure 55. Comparison between non-windowed and windowed ISB output (data from “baseline ISB” case – Section 6.4)**

#### 5.3.4 Inline Sources

An inherent limitation of ISB lies in its handling (or inability to handle) source-receiver combinations that are “inline”, i.e. the source-receiver vector  $r_r - r_s$  is exactly parallel or perpendicular with the walls of the room as shown in Figure 56. The reason for this becomes clear when considering the image triangulation situation in this case: though the side wall images (if present) will contribute some triangulation effect, the output from the front and rear image receivers will form a solid line. This concept is verified experimentally in Section 6.7.



**Figure 56. (a) Offset and (b) inline image source triangulation**

## **Chapter 6: Image Source Triangulation - Experiments**

This chapter presents the series of experiments conducted in order to evaluate the image source triangulation method.

### **6.1 Progression of Experiments**

The experiments presented in this chapter aim to investigate the performance of IST in various situations, starting in a rectangular shoebox room that exemplifies the simplest case in which the IST method can effectively operate.

First, a baseline source-receiver case was tested both in simulation and physical measurement in order to verify the validity of the IST method. The results of this baseline case are then compared with test cases that vary source-receiver distance, background noise, and room geometry (by moving the experimental setup to a different room).

### **6.2 Presentation of Results**

The results in this section are presented in the same manner as those in Chapter 4 (described in Section 4.3). Due to IST's reliance on the reverberant field to improve localization, it was found that the ISB output maps contained more background reverberation than with DT. Therefore, the “half-power” ( $z_{IST}(r_{pk})/\sqrt{2}$ ) contour was used as the resolution performance metric.

### **6.3 Experimental Setup**

#### **6.3.1 Source Parameters**

In the measured cases, the input source signal was, as in the previous experiments, broadband noise (10 Hz – 100 kHz) generated using an SRS SR770 Network Analyzer. This signal was played through a dodecahedral loudspeaker.

In the simulated cases, a validated image source model [38] was used to generate a synthetic room impulse response for each array microphone according to the image source method.

The impulse responses were then convolved with an input signal of Gaussian white noise in order to produce the simulated microphone output signal. The microphone signal was sampled at 31250 Hz in all cases.

Table 11 summarizes the key parameters for the experiments in this chapter.

**Table 11. Experiment parameters for ISB experiments**

Parameter	Value
<b>Triangulation Type</b>	Additive
<b>Sampling Rate</b>	31250 Hz
<b>Sampling Time</b>	5000 points (0.16 s)
<b>Grid Spacing</b>	0.1 m
<b>Source Power</b>	0.90 mW (real or simulated)
<b>Frequency Range</b>	[1000 2000] Hz

### 6.3.2 Test Environments

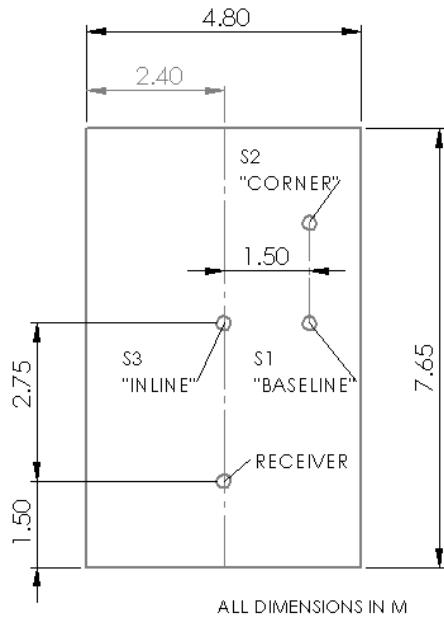
The shoebox test room, used for Sections 6.4 to 6.7, was a 4.8 m x 7.65 m x 3.85 m room with concrete walls, ceiling and floor. The reverberation time ( $T_{60}$ ) was 0.88 s in the octave band of interest (interpolated from the 1000 Hz and 2000 Hz octave band measurements). The Schroeder frequency was found to be 180 Hz and reverberation radius 0.67 m – again, as in Section 4.5, the frequency band of interest is above the Schroeder frequency and source-receiver distances are beyond the reverberation radius. Unlike in the direct triangulation case, reverberation is not necessarily detrimental to localization using IST, since the intention of the IST algorithm is to incorporate information from the early reflections into the beamforming output. Therefore, with ISB, the reverberant room case is favourable to a hypothetical free field case (where IST would not bring any benefit at all).

The receiver remained in the same location for all tests; the three source positions used and referenced in this chapter are shown in Figure 57. The measurement room was mostly devoid of

obstacles and furnishings; several bicycles and bike racks (the room is normally used as a bicycle storage locker) were pushed to the edges of the room, as shown in Figure 58. A notable feature of the shoebox room was that there were many obstructive fixtures (e.g. pipes, ducts, lighting) on the ceiling. This was a concern because the IST model assumes that sound is reflected from planar, rigid walls at the room boundaries – if there are significant obstructions on the ceiling, this assumption is not accurate and the IST output may be skewed as a result. To address this issue, three options were considered:

1. Use the original ceiling height in the IST model
2. Use an intermediate height as the room ceiling in the IST model
3. Remove the ceiling entirely from the IST model (i.e. assume that the ceiling is anechoic and produces no reflections)

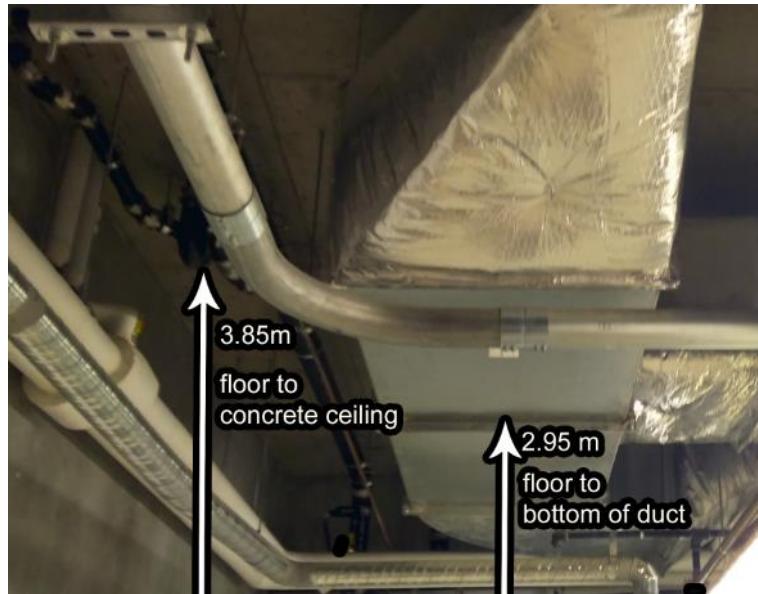
Through trial and error, the best localization results were obtained using an intermediate height (option 2) of 2.95 m, corresponding to the approximate height of the most significant ceiling feature – the network of ventilation ducts suspended from the ceiling, as shown in Figure 59. Therefore, all of the ISB results in this chapter were computed using an assumed room dimension of 4.8 m x 7.65 m x 2.95 m.



**Figure 57.** Top view of shoebox test room with source-receiver locations. Receiver height = 1.02 m, source height = 1.15 m

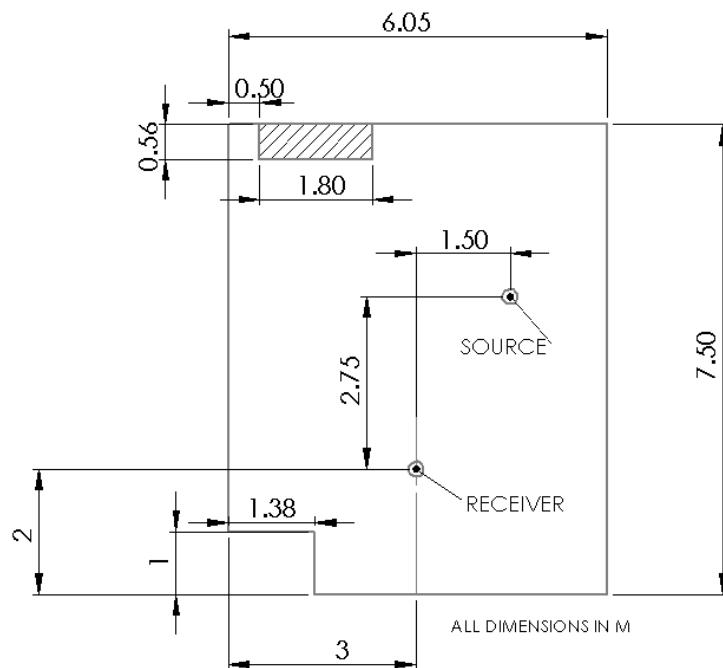


**Figure 58.** Dodecahedral source in shoebox room, "S1" source position



**Figure 59. Ceiling detail of shoebox room**

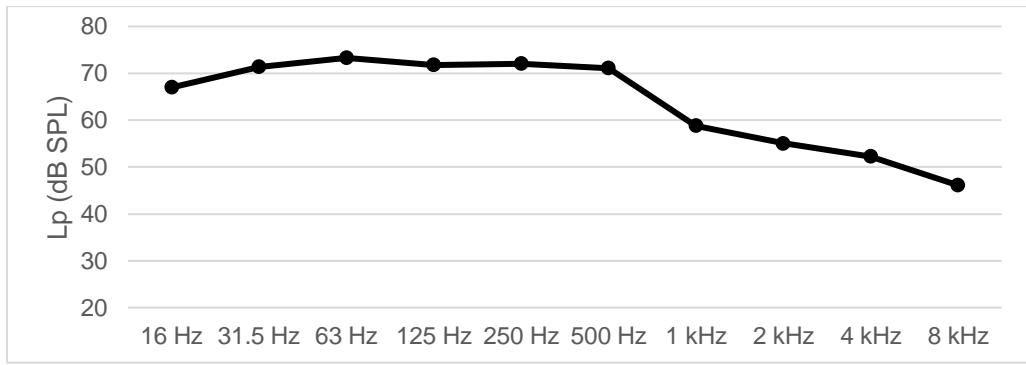
For the “geometry variation” case in Section 6.8, the same test room used for the direct triangulation reverberant room experiments was used – the source-receiver positions for this case are shown in Figure 60, and details of the room were specified in Section 4.5.,



**Figure 60. Top view of “geometry variation” test room with source-receiver locations. Receiver height = 1.02 m, source height = 1.15 m**

### 6.3.3 Background Noise

For the background noise case (Section 6.6), a strong low-frequency noise source (the building HVAC system) was present in the test room. The octave band profile of this noise is shown in Figure 61. Though the total noise level was high, most of the sound energy was concentrated on the 500 Hz octave band and below, meaning that the majority of the noise was filtered out before the localization processing step. The overall SNR was -1.95 dB, while the filtered SNR was 3.53 dB.

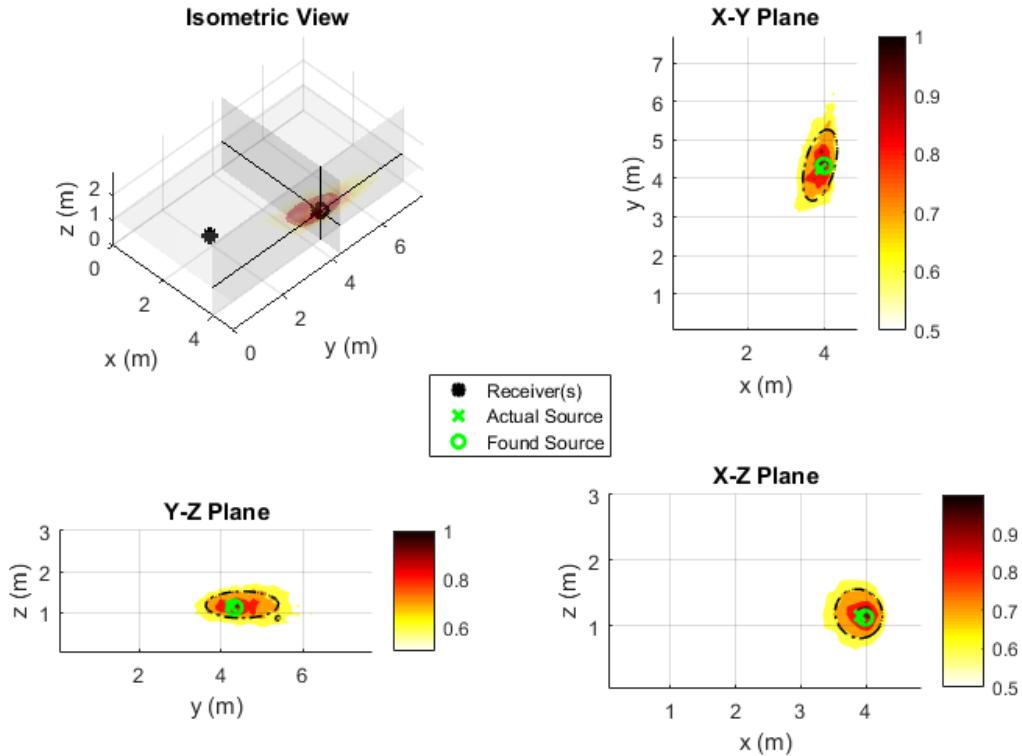


**Figure 61. Octave band noise profile for “background noise” IST case**

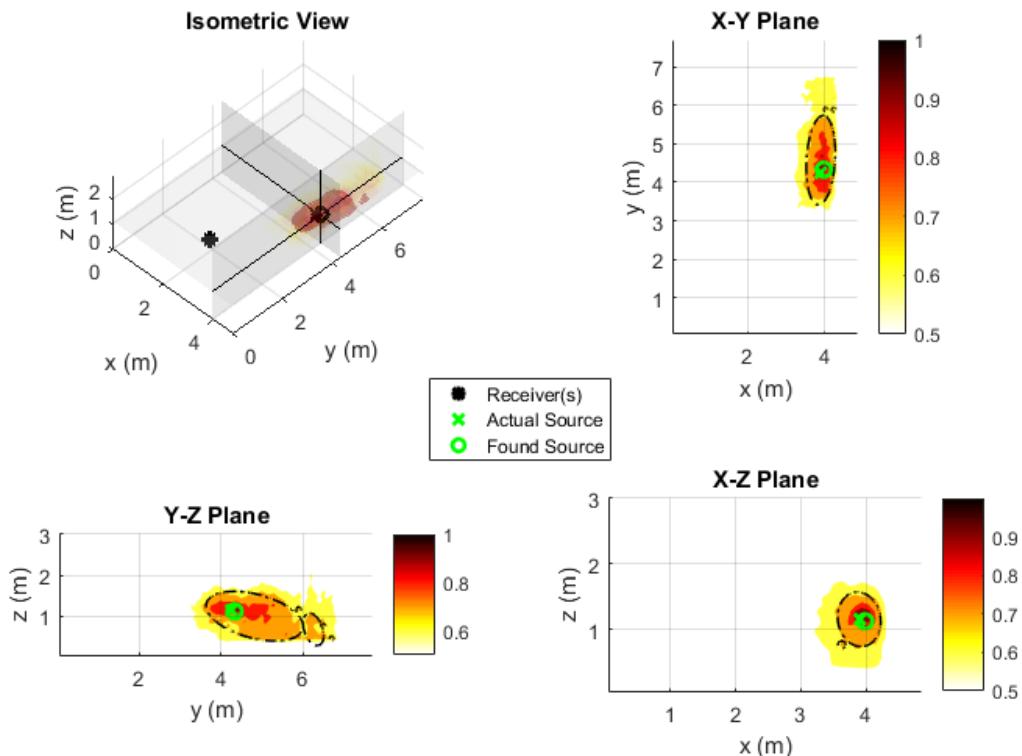
### 6.4 Baseline Case

The purpose of the baseline case is twofold: first, to establish the image order that will be used in subsequent experiments, and also to provide a reference performance point to which subsequent cases can be compared.

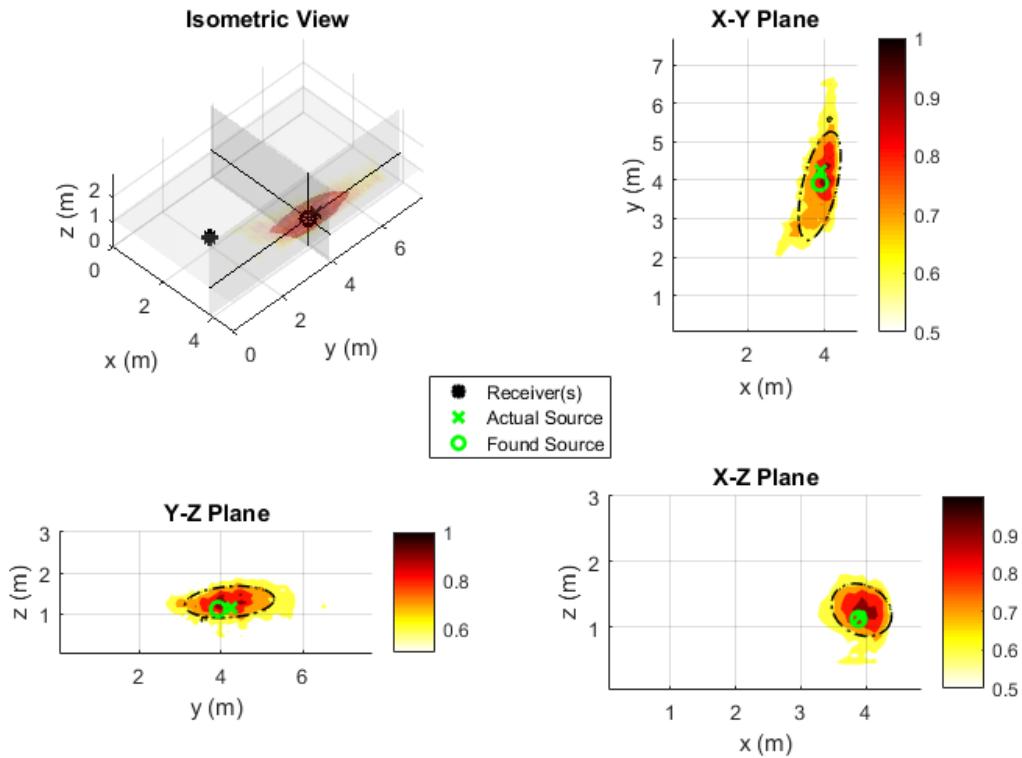
In order to evaluate the effect of image order on localization performance, separate sets of results were generated using first- and second-order sources in the IST model. With third-order sources and beyond, edge artifacts rendered the IST output unusable and thus the results are not shown here. Source position S1 (as marked in Figure 57) was used in both simulation and measurement. The results for these four cases are shown graphically in Figures 62 to 65 and tabulated in Table 12.



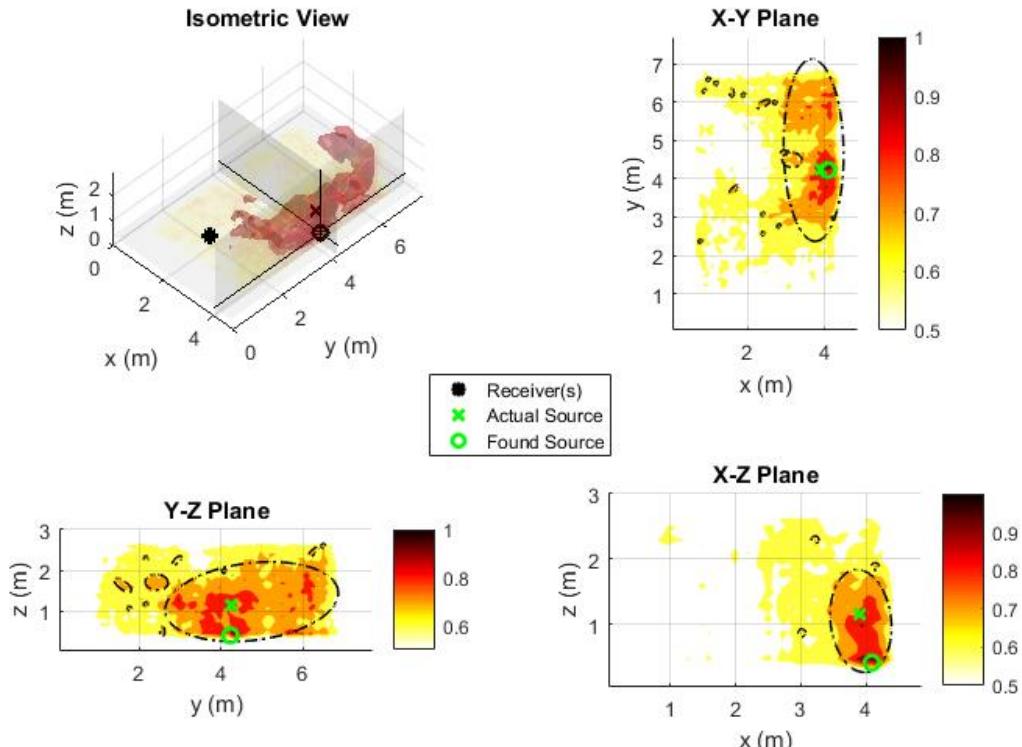
**Figure 62.** Graphical results for ISM simulation, “S1” source position, 1<sup>st</sup> order IST processing. Peak location was at (3.98, 4.33, 1.12); actual source location was (3.90, 4.25, 1.15)



**Figure 63.** Graphical results for ISM simulation, “S1” source position, 2<sup>nd</sup> order IST processing. Peak location was at (3.98, 4.33, 1.12); actual source location was (3.90, 4.25, 1.15)



**Figure 64.** Graphical results for shoebox room measurement, “S1” source position, 1<sup>st</sup> order IST processing  
Peak location was at (3.88, 3.93, 1.12); actual source location was (3.90, 4.25, 1.15)



**Figure 65.** Graphical results for shoebox room measurement, “S1” source position, 2<sup>nd</sup> order IST processing.  
Peak location was at (4.09, 4.23, 0.41); actual source location was (3.90, 4.25, 1.15)

**Table 12. Error and resolution comparison between baseline shoebox cases**

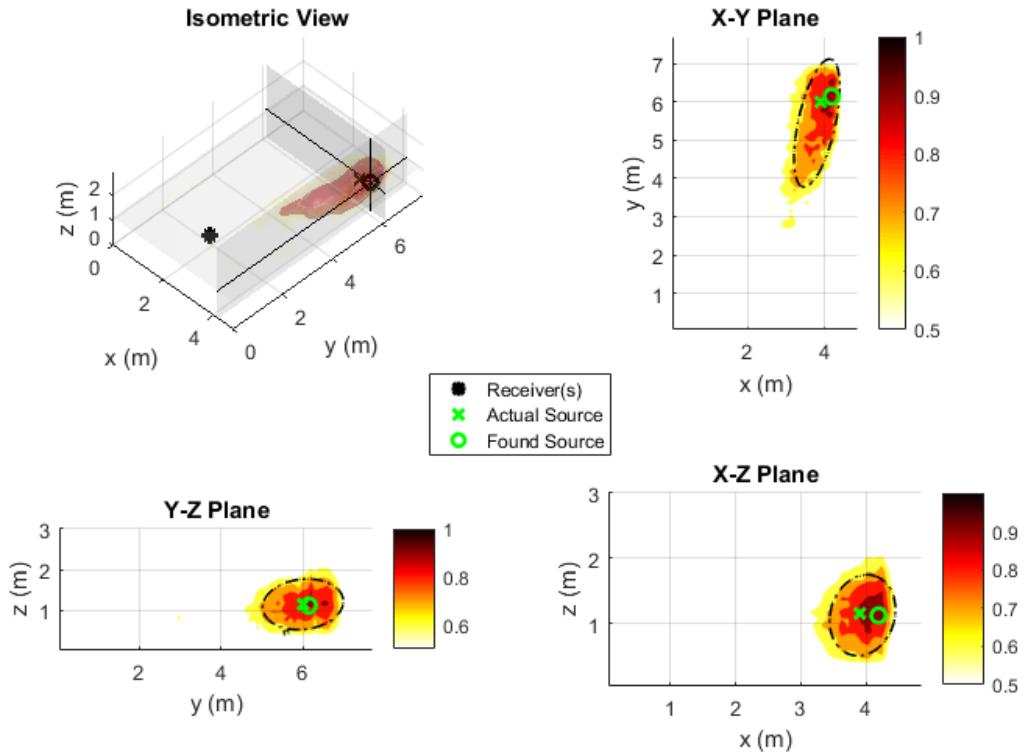
Data Case	Image Order	S/R Distance (m)	Total Error (m)	Normalized Error	Average blob width (m)	Normalized blob width
<b>Simulation</b>	1	3.14	0.12	0.038	0.55	0.175
<b>Simulation</b>	2	3.14	0.12	0.038	0.67	0.214
<b>Measurement</b>	1	3.14	0.33	0.104	0.70	0.224
<b>Measurement</b>	2	3.14	0.77	0.244	1.21	0.387

The image source simulation model can be considered the “ideal case” for IST – reflections are perfectly specular (i.e. no diffusion), and no possibility for error in positioning the source or receiver, either relative to each other or to the walls of the room. For this reason, both simulation cases show better performance in both error and resolution than either of the measured cases.

In both the simulation and measured cases, the first-order IST model performed better than its respective second-order counterpart. In particular, the edge artifacts in the second-order measured case (Figure 65) appear to have already affected the output to the point where the source cannot be accurately located in the “z’ axis. Therefore, the first-order measured case (Figure 64) will be used as the reference measurement for subsequent cases.

## 6.5 Corner Case

A second measurement was taken at position “S2” in Figure 57 in order to investigate the effect of source-receiver distance on the IST method. The results are shown graphically in Figure 66 and compared with the baseline case in Table 13.



**Figure 66.** Graphical results for shoebox room measurement, “S2” source position, 1<sup>st</sup> order IST processing. Peak location was at (4.19, 6.14, 1.12); actual source location was (3.90, 6.00, 1.15)

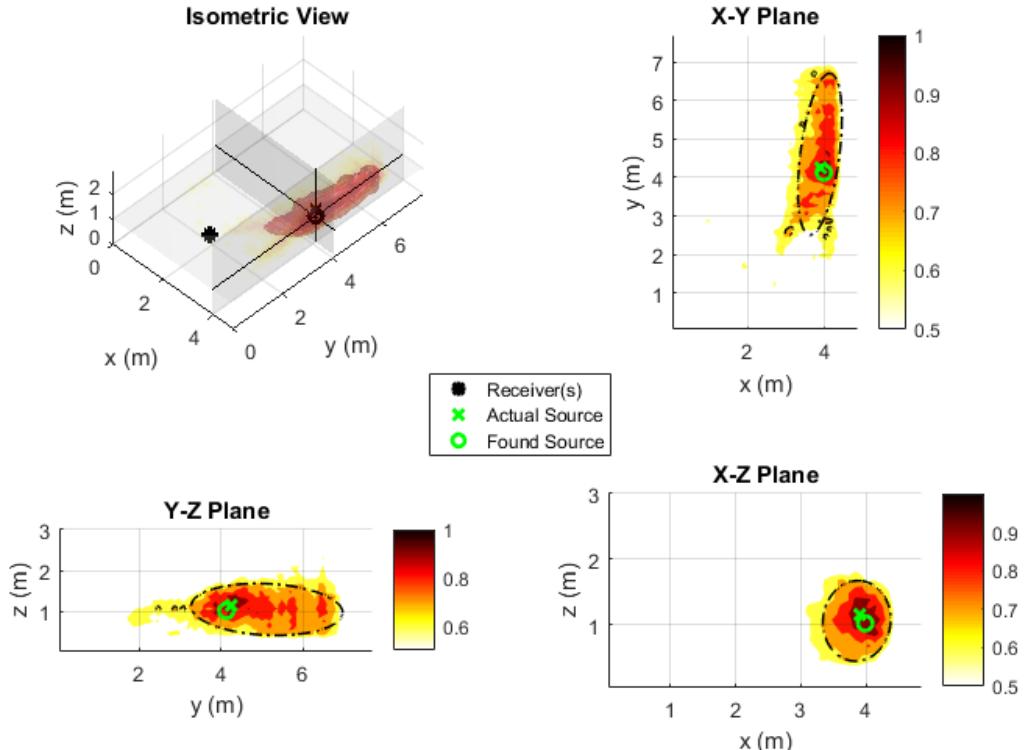
**Table 13. Error and resolution comparison between baseline and corner cases**

Data Case	Image Order	S/R Distance (m)	Total Error (m)	Normalized Error	Average blob width (m)	Normalized blob width
<b>Baseline (reference)</b>	1	3.14	0.33	0.104	0.70	0.224
<b>Corner</b>	1	4.75	0.32	0.068	0.82	0.173

The blob width of the corner case is higher than that of the baseline case, as expected due to the direct sound’s increasing beam width with distance. However, the normalized blob width is in fact lower in the corner case – examining the cross-section contour plots in Figure 66, it appears applying the Tukey window to the edges of the scanning volume may have cut off some of the blob that would otherwise have been present. This was less of an issue in the baseline case, as it was not as close to the far wall of the room as in the current case.

## 6.6 Background Noise

The measurement for the background noise case was taken at position “S1” with the HVAC system running inside the room. The results are shown graphically in Figure 67 and compared with the baseline case in Table 14.



**Figure 67.** Graphical results for shoebox room measurement, “S1” source position with background noise, 1<sup>st</sup> order IST processing. Peak location was at (3.98, 4.13, 1.02); actual source location was (3.90, 4.25, 1.15)

**Table 14.** Error and resolution comparison between baseline and background noise cases

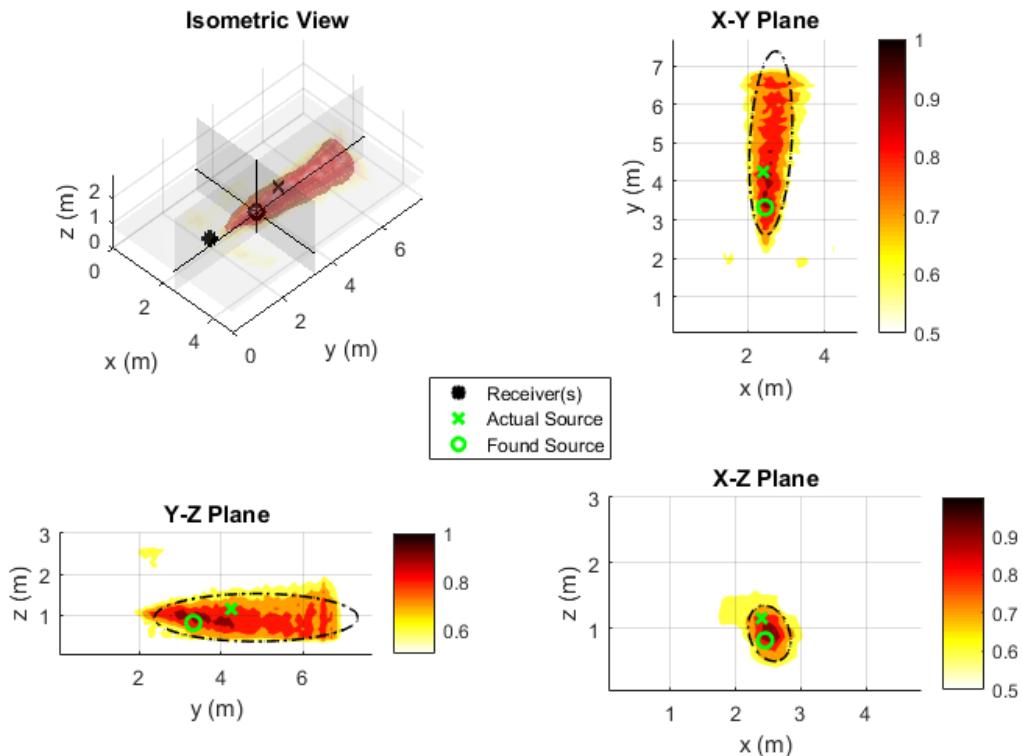
Data Case	Image Order	S/R Distance (m)	Total Error (m)	Normalized Error	Average blob width (m)	Normalized blob width
<b>Baseline (reference)</b>	1	3.14	0.33	0.104	0.70	0.224
<b>Noise</b>	1	3.14	0.20	0.064	1.04	0.331

Compared to either of the previous cases (containing negligible background noise), the background noise case shows worse performance in terms of blob resolution, which is the expected

result as the presence of background noise sources contributes to the overall sound field. However, the background noise does not appear to affect the accuracy of source peak localization – this may be due to the fact that the HVAC noise was distributed across a large surface and largely centered at a low frequency that was mostly filtered out before processing.

## 6.7 Inline Case

The inline case measurement was taken at position “S3” (HVAC system turned off, same as in the baseline case). The results are shown graphically in Figure 68 and compared with the baseline case in Table 15. This measurement serves to verify the limits of IST when attempting to localize inline sources, as discussed in Section 5.3.4 – as expected, the accuracy and resolution in this case are the poorest of the shoebox cases discussed.



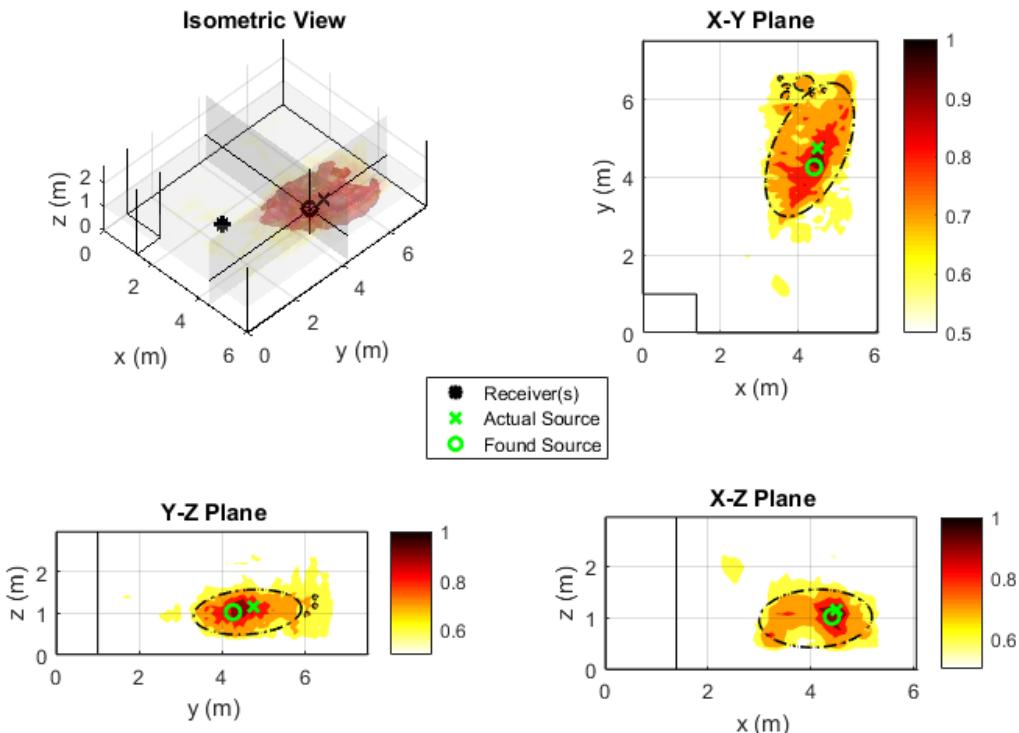
**Figure 68. Graphical results for shoebox room measurement, “S3” source position, 1<sup>st</sup> order IST processing. Peak location was at (2.45, 3.32, 0.81); actual source location was (2.4, 4.25, 1.15)**

**Table 15. Error and resolution comparison between baseline and inline cases**

Data Case	Image Order	S/R Distance (m)	Total Error (m)	Normalized Error	Average blob width (m)	Normalized blob width
<b>Baseline (reference)</b>	1	3.14	0.33	0.104	0.70	0.224
<b>Inline Source</b>	1	2.75	0.99	0.359	1.12	0.409

## 6.8 Geometry Variation

The measurement for the geometry variation case was taken in the initial test room (used for the direct triangulation tests), as shown in Figure 60. Note that the relative source/receiver distance was chosen to be identical to that of the baseline shoebox case. The results are shown graphically in Figure 69 and compared with the baseline case in Table 16.



**Figure 69. Graphical results for geometry variant measurement, “S1” equivalent source position, 1<sup>st</sup> order IST processing. Peak location was at (4.41, 4.26, 1.02); actual source location was (4.5, 4.75, 1.15)**

**Table 16. Error and resolution comparison between baseline and geometry variant cases**

Data Case	Image Order	S/R Distance (m)	Total Error (m)	Normalized Error	Average blob width (m)	Normalized blob width
<b>Baseline (reference)</b>	1	3.14	0.33	0.104	0.70	0.224
<b>Geometry Variation</b>	1	3.14	0.52	0.165	1.04	0.333

The localization performance obtained in this case is worse than either of the two other measurements taken at the same position (“S1”) in the shoebox room. It is possible that this could be due to the irregular geometry of the room, as the current IST model assumes the room shoebox-shaped. In order to conclusively prove or disprove the effect of the model mismatch, it would be necessary to compare the IST results using the current shoebox model with those using a model that takes into account the actual geometry of the room.

## 6.9 Discussion

Overall, image source triangulation offers an improvement over regular delay-and-sum beamforming without any need for additional acoustical measurements – the only extra information necessary is the geometry of the room, which can be obtained in any number of ways (tape measure, range finder, etc) depending on the resources available in a given situation. This also opens up the possibility of locating impulsive or transient noises such as speech in a reverberant room, which would not be possible in the single-array direct triangulation case. However, the poor performance of IST in localizing inline sources, as shown in Section 5.3.4, is a disadvantage. It may be possible to work around this limitation by combining direct triangulation and IST – performing the IST processing on each individual array measurement, then combining them using the direct triangulation method – but in doing so the advantages of IST over direct triangulation are no longer applicable. The presence of edge artifacts, and their increase in severity

with image order, leads to the (somewhat counterintuitive) conclusion that incorporating fewer image sources leads to better localization performance.

The effectiveness of IST depends on the presence of clear, predictable specular reflections.

The current image source model works well for a concrete shoebox room or featureless classroom, but a more complicated room geometry would require more sophisticated modeling tools to accurately predict the location of image sources.

### 6.9.1 Sources of Uncertainty

The sources of uncertainty in the IST experiments are similar to those in direct triangulation, as discussed in Section 4.7.1. However, there are two additional points, as follows:

1. **Assumption of the speed of sound, revisited:** The same constant speed of sound ( $c = 341 \text{ m/s}$ ) as was used for direct triangulation was used in the IST calculations. However, conditions in the shoebox room were different than in the rooms used in the direct triangulation case, it being in a different building and different season – the estimated temperature of  $15^\circ\text{C}$  in the room at the time of measurement was likely too low, which would produce a higher error in  $\Delta t_j$  in the IST cases.
2. **Uncertainty of room dimensions (model error):** In addition to the known model discrepancies such as the ceiling height of the shoebox room discussed in Section 6.3.2, and the obvious error in assuming a shoebox-shaped room for a non-shoebox case as discussed in Section 6.8, there also exists an uncertainty in the measurements of the room itself, which would translate into an error in the location of the predicted image sources and receivers. This error would be compounded with increasing image order, which may be another reason why using fewer images (“quality over quantity”) would be desirable.

## **Chapter 7: Conclusion**

### **7.1 Summary**

This thesis presented two different methods for 3D sound source localization using the principles of triangulation, using a hemispherical microphone array and base processing method of delay-and-sum beamforming.

The hemispherical microphone array used in this project consisted of 27 electret microphones arranged in a geodesic dome configuration. Delay-and-sum beamforming uses the implicit time delay between array microphones in order to “focus” the array on a point of interest – this is effective for determining the direction of arrival of a sound source, but less so for determining the range (distance) between source and receiver. Measurements were conducted in the [1000, 2000] Hz octave band. Due to the hemispherical shape of the array, the output beam width varied based on the sound source direction of arrival, reaching a maximum at  $\pm 90^\circ$ .

The first triangulation method discussed – direct triangulation – involves taking multiple measurements from a single array moved to various positions in a room. The delay-and-sum beamforming outputs of the array at each location are multiplied together in order to obtain the final, triangulated output. This output offers an improvement in range estimation over conventional delay-and-sum beamforming. It is conceptually simple and requires no prior knowledge of the room geometry. The effectiveness of direct triangulation depends on the ability to identify the direct sound field emitted by the source – therefore, reverberation is undesirable.

Using direct triangulation in a reverberant room with 3 co-planar receiver locations, a single monopole source was localized within a location error of 0.16 m to 0.53 m, with error increasing as distance from source to receiver plane increased. Near the walls of the room,

localization performance decreased due to interference from image sources caused by specular reflections from the walls.

The second triangulation method discussed – image source triangulation – uses prior knowledge of the room geometry to predict the locations of the specular reflections of the direct source in order to improve localization. This was shown to be equivalent to triangulation by “image receivers” mirrored about the boundaries of the room. Only one array measurement is required. The effectiveness of the method relies on the assumption that the early specular reflections of the direct source can be successfully localized – therefore, some degree of reverberation is necessary in order for the IST method to function.

The best localization results were obtained when only incorporating first-order images into the beamforming model, with errors ranging from 0.20 m to 0.33 m in a simple shoebox-shaped room. With a slightly more complex room geometry, this error increased to 0.52 m, possibly due to the model’s failure to account for the increase in room complexity. Two disadvantages of image source triangulation are its inability to clearly identify inline sources, and the presence of edge artifacts that require further post processing to remove, which may obscure sources positioned close to the walls of the room.

## 7.2 Comparison Between Triangulation Methods

While both SSL methods presented in this thesis are based on the principle of triangulation, they diverge quite widely in practical implementation, as summarized in Table 17.

**Table 17. Differences between direct and image source triangulation methods**

<b>Direct Triangulation</b>	<b>Image Source Triangulation</b>
Multiple array measurements required	Single array measurement required
Room geometry information not necessary	Room geometry must be known (or estimated)
Uses direct sound only – works best with no reverberation	Uses direct and reverberant sound – works best with reverberation/does not work with no reverberation

The differences outlined above provide some insight into the potential applications of both methods – in particular, the differences dictate when one method would be preferable over another. For instance, direct triangulation would be preferable in an industrial noise control/identification application, where a) many of the sound sources of interest are continuous and therefore can be captured over repeated measurements, and b) the room geometry can be variable, complicated, and difficult to model, especially over large distances (e.g. a factory floor) where it is uncertain if specular image sources are even possible to localize at all. On the other hand, IST would be better suited to a teleconferencing (speech enhancement) application where the signal of interest is transient, and the array is kept in a known, fixed location – obtaining the room geometry could be part of the initial setup of the system, after which the information would be stored for future use.

### 7.3 Future Work

For direct triangulation, the current approach processes all of the array outputs separately before combining them multiplicatively or additively – an interesting further development may be to consider all of the array elements as part of a single “super-array”, which has the benefit of increasing the array aperture and has been shown to improve localization in [16] and [21]. Even without modifying the processing approach, localization results may be improved with some more creative receiver positioning strategies (e.g. in the corners of the room, or angled toward probably source locations).

For image source triangulation, the handling of non-shoebox room shapes is a significant issue, as many “real-life” environments are considerably more complex than a simple 6-sided room. One way to do this would be to extend the image source prediction model to handle arbitrary polyhedra, as in [39]. Other issues that merit consideration are possible methods to mitigate the

inline source and edge artifact problems shown in the results above. It may be possible to improve localization in these cases simply by changing the base SSL algorithm (see below).

For both methods, more tests involving more complex acoustical situations, such as environments with multiple sources and distributed sources would help in validating both SSL methods. Since both direct triangulation and image source triangulation are “meta-processing” methods, they could conceivably work with any array shape or base SSL algorithm. Therefore, another way produce better localization results would be to use a different array (e.g. a spherical array for its direction-invariant response) or more sophisticated base SSL algorithm capable of higher-resolution localization.

## References

- [1] A. L. Swindlehurst, B. D. Jeffs, G. Granados-Seco and J. Li, "Applications of Array Signal Processing," in *Array and Statistical Signal Processing*, Academic Press, Chennai, 2014, pp. 859-953.
- [2] M. Brandstein and D. Ward, Microphone Arrays: Signal Processing Techniques and Applications, Berlin: Springer-Verlag, 2001.
- [3] R. Bader, "Microphone Array," in *Springer Handbook of Acoustics*, New York, Springer, 2014, pp. 1179-1207.
- [4] R. J. Kozick and B. M. Sadler, "Source Localization With Distributed Sensor Arrays and Partial Spatial Coherence," *IEEE Trans. Signal Processing*, vol. 52, no. 3, pp. 601-616, 2004.
- [5] J.-A. Luo, X.-P. Zhang, Z. Wang and X.-P. Lai, "On the Accuracy of Passive Source Localization Using Acoustic Sensor Array Networks," *IEEE Sensors*, vol. 17, no. 6, pp. 1795-1808, 2017.
- [6] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, pp. 943-950, 1979.
- [7] H. Alghassi, "Eye array sound source localization," Ph.D. Dissertation, Dept. Elect. Eng., Univ. British Columbia, Vancouver, Canada, 2008.
- [8] M. Nematifar, "Localization of sound sources using a microphone array," Internal Report, UBC Department of Electrical and Computer Engineering, 2009.
- [9] A. Khaleghi, J. Ryan and M. Pajchel, "3D sound source localization using a hemispherical microphone array," Internal Report, UBC Engineering Physics, 2014.
- [10] D. H. Johnson and D. E. Dudgeon, Array Signal Processing: Concepts and Techniques, Englewood Cliffs: Prentice Hall, 1993.
- [11] M. Born and E. Wolf, Principles of Optics, Cambridge: Cambridge University Press, 1999.
- [12] M. Aldeman, K. Chelliah, H. Patel and G. Raman, "Effects of array scaling and advanced beamforming on the angular resolution of microphone array systems," in *6th Berlin Beamforming Conference*, Berlin, 2016, pp. 1-24.
- [13] D. E. Dudgeon and R. M. Mersereau, Multidimensional Digital Signal Processing, Englewood Cliffs, NJ: Prentice-Hall, Inc., 1984.

- [14] T. F. Brooks and W. M. Humphreys, "A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays," *Journal of Sound and Vibration*, vol. 294, pp. 856-879, 2006.
- [15] B. Rafaely, *Fundamentals of Spherical Array Processing*, Berlin: Springer, 2015.
- [16] K. A. L. Szuberia, J. V. Olson and K. M. Arnoult, "Explosion localization via infrasound," *Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. EL112-EL116, 2009.
- [17] D. Mennitt and M. Johnson, "Multiple-array passive acoustic source localization in urban environments," *Journal of the Acoustical Society of America*, vol. 127, no. 5, pp. 2932-2942, 2010.
- [18] X. Bian and G. Abowd, "Using sound source localization in a home environment," in *Proceedings of 3rd International Conference on Pervasive Computing*, Munich, 2005, pp. 19-36.
- [19] E. Martinson and A. Schultz, "Discovery of sound sources by an autonomous mobile robot," *Autonomous Robots*, vol. 27, pp. 221-237, 2009.
- [20] Y. Sasaki, R. Tanabe and H. Takemura, "Probabilistic 3D Sound Source Mapping using Moving Microphone Array," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Daejon, Korea, 2016, pp. 1293-1298.
- [21] P. Castellini and A. Sassaroli, "Acoustic source localization in a reverberant environment by average beamforming," *Mechanical Systems and Signal Processing*, vol. 24, pp. 796-808, 2010.
- [22] N. Ma, C. S. Chia, R. Yang and G. W. Ng, "Target localization by two passive linear arrays," *IEEE OCEANS*, pp. 1-6, 2006.
- [23] C. Bibby, "Point-Source Design and Performance," Internal Report, UBC Department of Mechanical Engineering, 2009.
- [24] J. Fischer and C. Doolan, "Beamforming in a reverberant environment using numerical and experimental steering vector formulations," *Mechanical Systems and Signal Processing*, vol. 91, pp. 10-22, 2017.
- [25] M. D. Egan, *Architectural Acoustics*, New York: McGraw-Hill, 1988.
- [26] L. Cremer and H. A. Mueller, *Principles and Applications of Room Acoustics*, Vol 1, Essex: Applied Science Publishers, 1978.

- [27] S. Guidati, G. Guidati and S. Wagner, "Beamforming in a Reverberating Environment with the use of Measured Steering Vectors," in *7th AIAA/CEAS Aeroacoustics Conference*, Maastricht, 2001, pp. 3-10.
- [28] P. Sijtsma and H. Holthusen, "Corrections for mirror sources in phased array processing techniques," in *9th AIAA/CEAS Aeroacoustics Conference*, Hilton Head, 2003, pp. 1-11.
- [29] B. A. Fenech and K. Takeda, "Towards more accurate beamforming levels in closed-section wind tunnels via de-reverberation," in *13th AIAA/CEAS Aeroacoustics Conference*, Rome, 2007, pp. 1-12.
- [30] L. Guillon, S. E. Dosso, N. R. Chapman and A. Drira, "Bayesian geoacoustic inversion with the image source method," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 4, pp. 1035-1044, 2016.
- [31] X. Wang, S. Khazaie, L. Margheri and P. Sagaut, "Shallow water sound source localization using the iterative beamforming method in an image framework," *Journal of Sound and Vibration*, vol. 395, pp. 354-370, 2017.
- [32] Z. Li, K. F. C. Yiu and S. Nordholm, "On the Indoor Beamformer Design With Reverberation," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 22, no. 8, pp. 1225-1235, 2014.
- [33] F. Ribeiro, C. Zhang, D. A. Florêncio and D. E. Ba, "Using reverberation to improve range and elevation discrimination for small array sound source localization," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 7, pp. 1781-1792, 2010.
- [34] D. Markovica, F. Antonacci, A. Sarti and S. Tubaro, "Estimation of room dimensions from a single impulse response," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, 2013, pp.1-4.
- [35] T. Rajapaksha, X. Qiu, E. Cheng and I. Burnett, "Geometrical room geometry estimation from room impulse responses," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, 2016, pp. 331-335.
- [36] T. Hidaka, Y. Yamada and T. Nakagawa, "A new definition of boundary point between early reflections and late reverberation in room impulse responses," *Journal of the Acoustical Society of America*, vol. 122, no. 326, pp. 326-332, 2007.
- [37] J.-D. Polack, "Playing Billiards in the Concert Hall - the Mathematical Foundations of Geometrical Room Acoustics," *Applied Acoustics*, vol. 28, pp. 235-244, 1993.
- [38] E. Lehmann and A. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 269-277, 2008.

- [39] J. Borish, "Extension of the image model to arbitrary polyhedra," *Journal of the Acoustical Society of America*, vol. 75, no. 6, pp. 1827-1836, 1984.
- [40] Panasonic Corporation, *WM61-A Datasheet*, 2011.
- [41] PUI Audio, Inc., *POM-3535L-3-R Omni-Directional Microphone*, 2015.
- [42] G. Jolly, *Audio Electret Microphone Pre-Amplifier*, UBC Mechanical Engineering, 2012.

## Appendices

### Appendix A Microphone Construction and Calibration

#### A.1 Detailed Microphone Placement Geometry

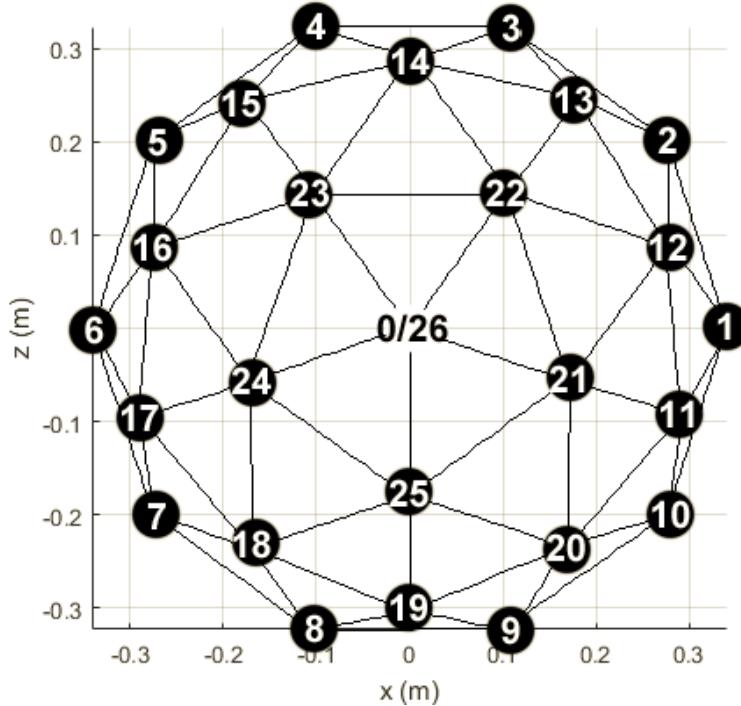


Figure 70. Numbered layout of array microphones, front view

Table 18. Microphone positions and types by number

Mic #	x (m)	y (m)	z (m)	Mic Type	Mic #	x (m)	y (m)	z (m)	Mic Type
0	0.00	0.00	0.00	1	14	0.00	-0.18	0.29	1
1	-0.34	0.00	0.00	1	15	0.18	-0.15	0.25	1
2	-0.28	0.00	0.20	1	16	0.28	-0.18	0.09	1
3	-0.11	0.00	0.32	1	17	0.29	-0.15	-0.09	1
4	0.11	0.00	0.32	1	18	0.17	-0.18	-0.23	1
5	0.28	0.00	0.20	1	19	0.00	-0.15	-0.30	1
6	0.34	0.00	0.00	1	20	-0.17	-0.18	-0.23	2
7	0.28	0.00	-0.20	1	21	-0.17	-0.29	-0.06	1
8	0.11	0.00	-0.32	1	22	-0.11	-0.29	0.14	1
9	-0.11	0.00	-0.32	3	23	0.11	-0.29	0.14	1
10	-0.28	0.00	-0.20	2	24	0.17	-0.29	-0.06	1
11	-0.29	-0.15	-0.09	1	25	0.00	-0.29	-0.18	1
12	-0.28	-0.18	0.09	1	26	0.00	-0.34	0.00	1
13	-0.18	-0.15	0.25	1					

## A.2 Microphone Frequency Responses by Model

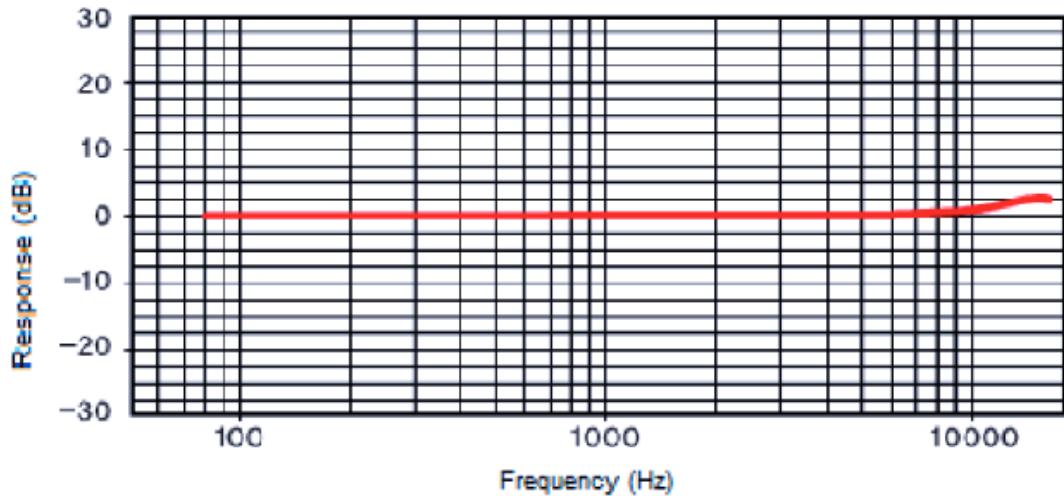


Figure 71. (Type 1) Typical frequency response for BGO-15L27-C1033 microphone [7]

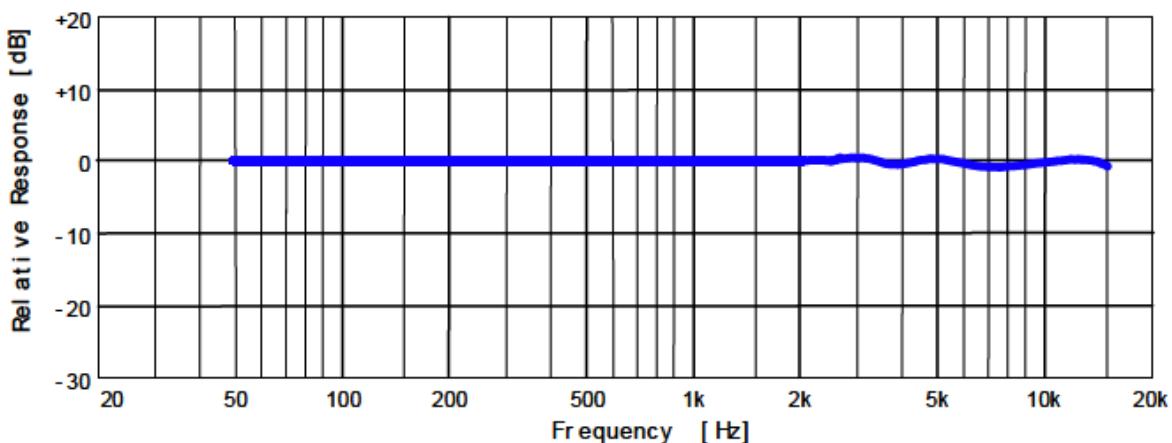


Figure 72. (Type 2) Typical frequency response for WM61-A microphone [40]

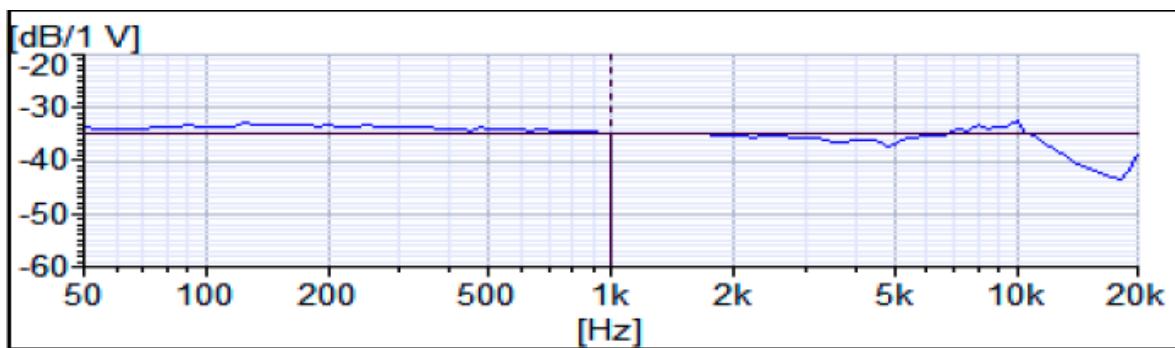


Figure 73. (Type 3) Typical frequency response for POM-3535L-3-R microphone [41]

### A.3 Microphone Circuit Diagram

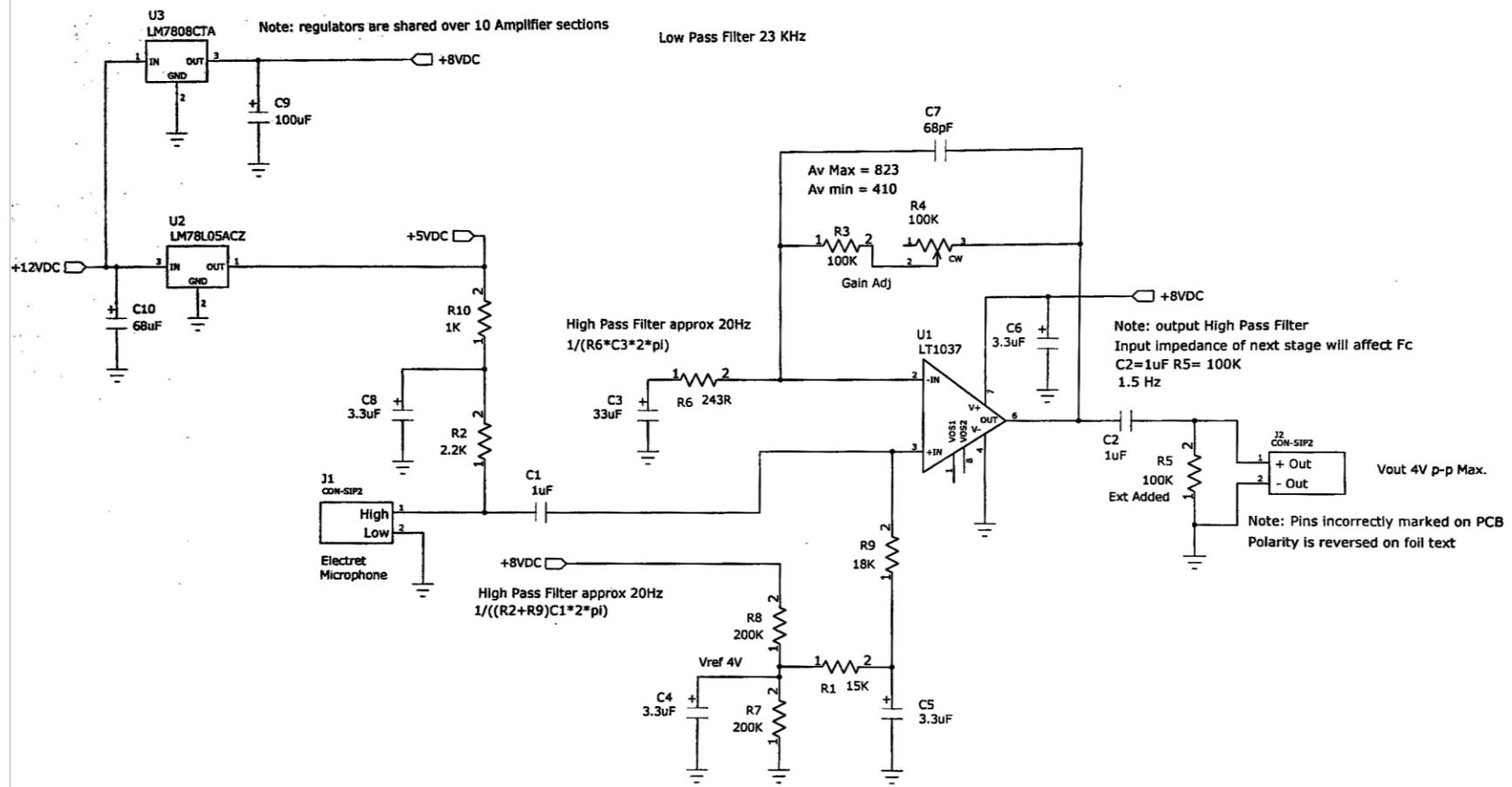
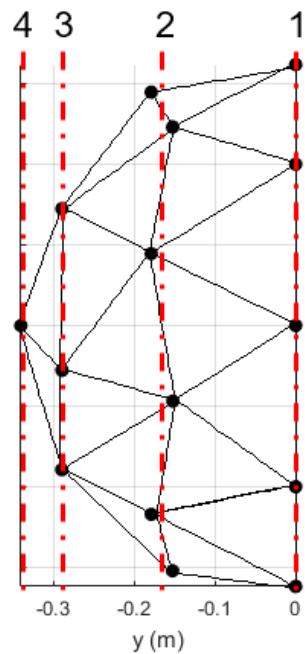


Figure 74. Microphone preamplifier circuit diagram [42]

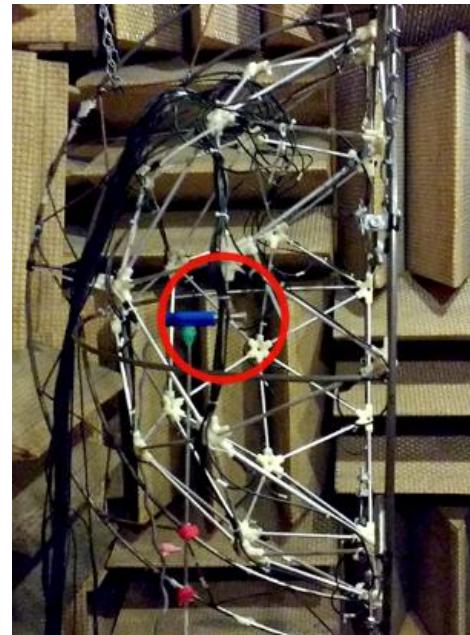
**NOTE:** As of 2 February 2017, resistor R3 was changed from 100K to 4870 Ohms, thus setting the average max gain (Av Max) to 430 and min gain (Av Min) to 20.04.

#### A.4 Array Calibration

Array calibration was performed in an anechoic chamber with a calibrated reference microphone placed at four different “depths” of the hemispherical array, as shown in Figure 75. During calibration, it was assumed that the sound pressure at all of the microphones at a given depth would be the same (i.e. that the arriving waves could be approximated as planar) – therefore, microphone sensitivity (V/Pa) could be obtained for all 27 microphones with only 4 measurements at each octave band. Calibration was performed at 7 octave bands (125 Hz – 8000 Hz) with a signal generator and source placed at a 3 m distance from the array.



(a)



(b)

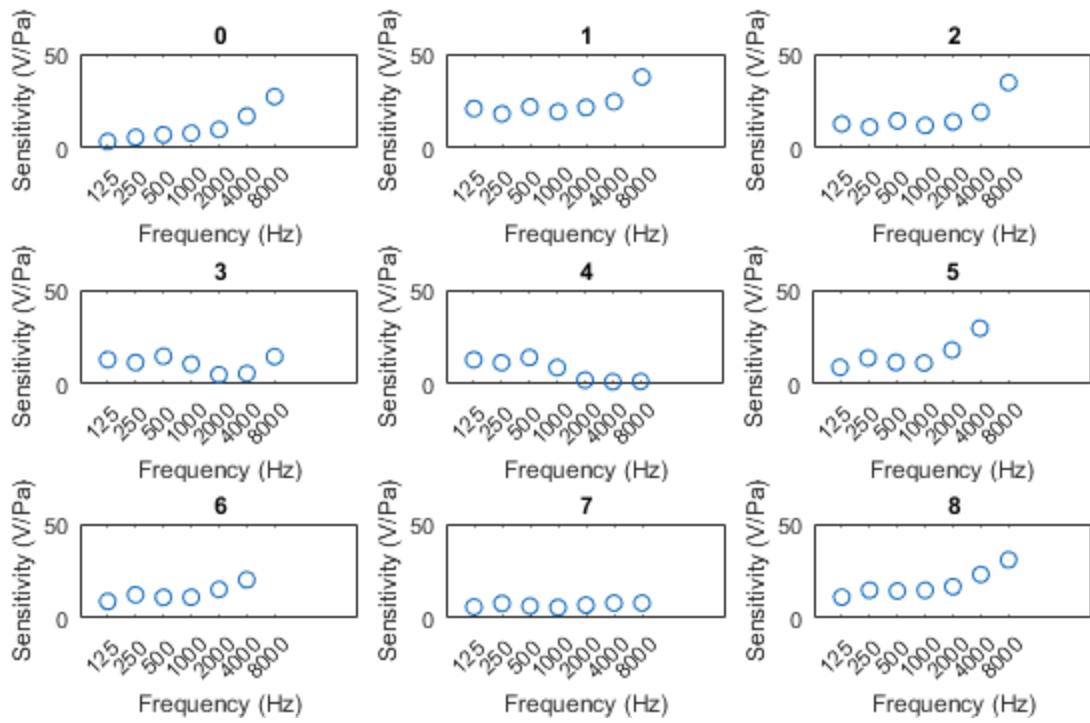
**Figure 75. (a) Schematic of hemisphere depth layers, (b) Photo of reference microphone (circled in red) at depth 2**

**Equipment used:**

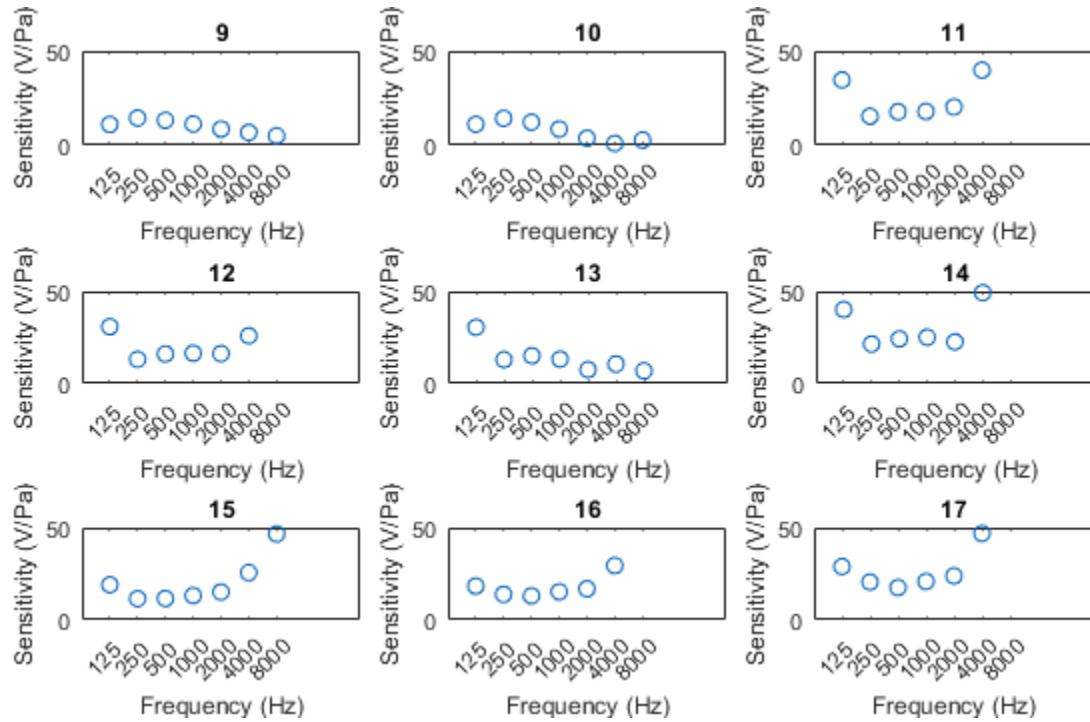
- Reference microphone: B&K Type 4165 Condenser Microphone, calibrated with B&K Type 4231 Calibrator
- Norsonic Real-type Analyser type 830 (to read reference microphone signal)
- Signal generator: SRS770 Network Analyser

**Calibration Procedure:**

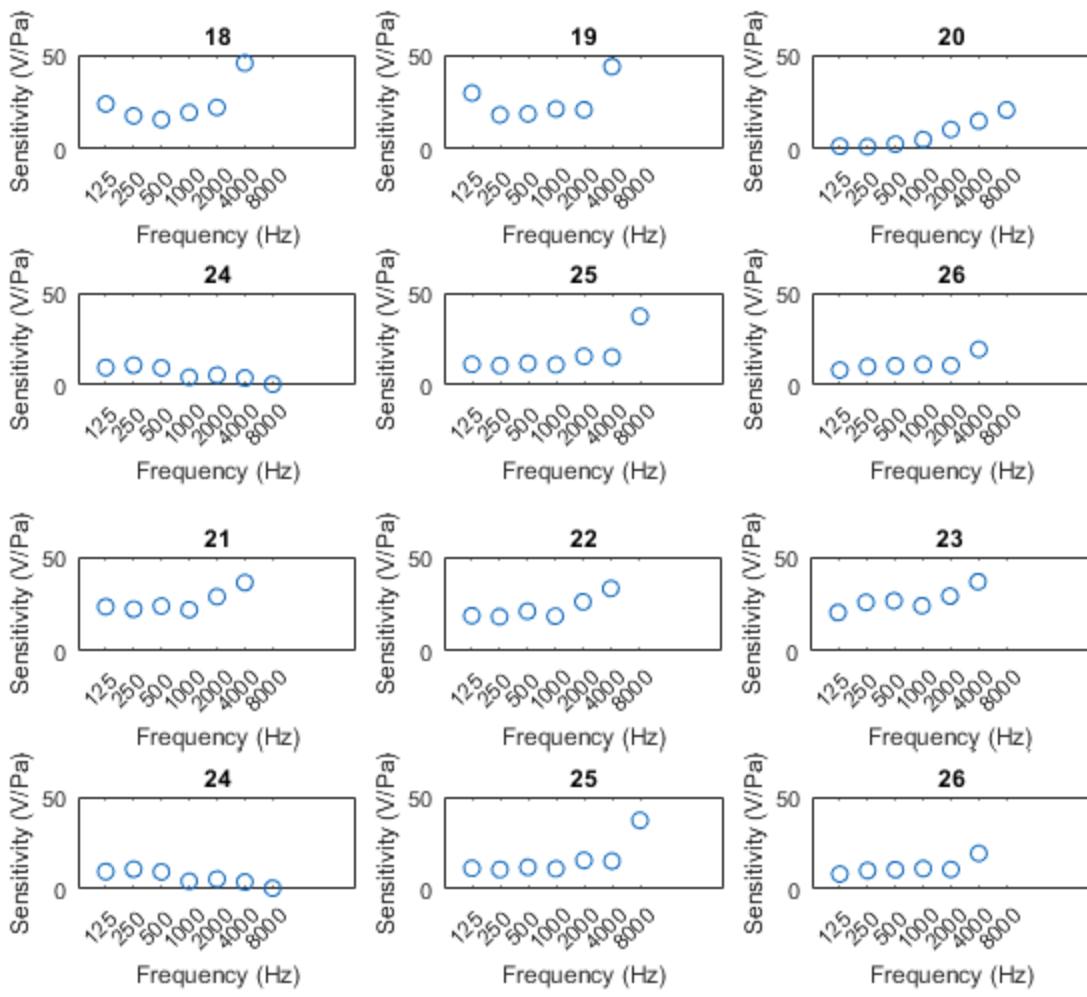
1. Place reference microphone at specified depth
2. Play pure tone from signal generator at specified frequency
3. Acquire array microphone signals through MATLAB
4. Record RMS pressure level at reference microphone
5. Divide RMS voltage by RMS pressure to obtain sensitivity (V/Pa) for microphones at the specified depth
6. Repeat procedure for all depths and frequency bands to obtain sensitivity plots (Figures 76 to 78)



**Figure 76. Sensitivities (V/Pa) for microphones 0 to 8**



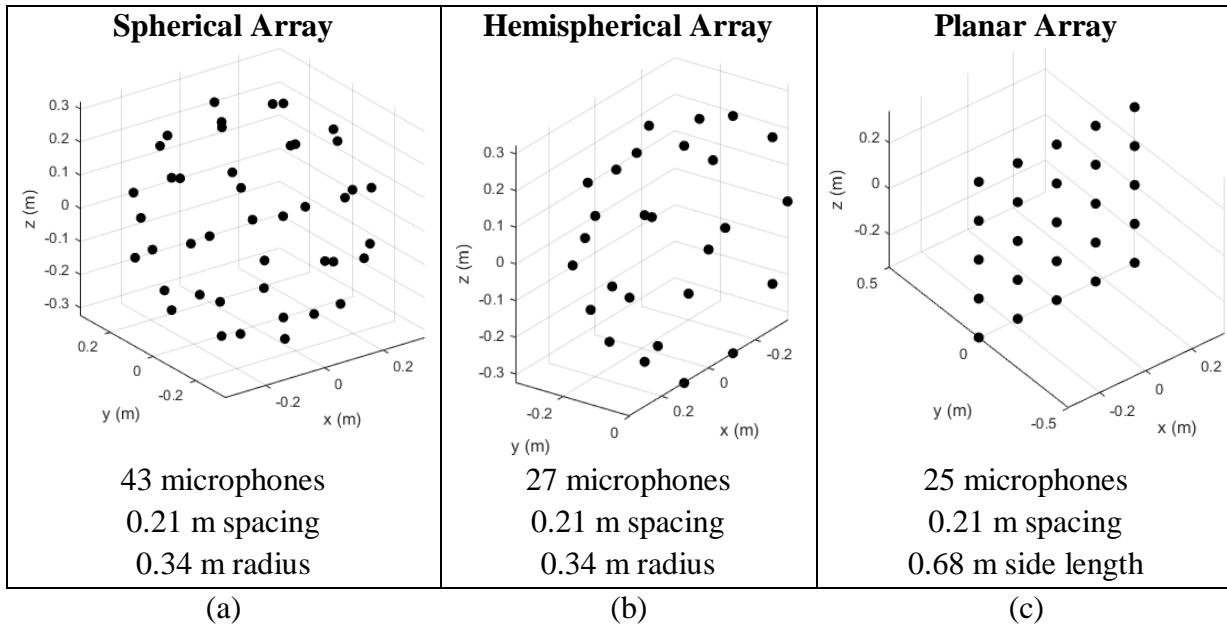
**Figure 77. Sensitivities (V/Pa) for microphones 9 to 17**



**Figure 78. Sensitivities (V/Pa) for microphones 18 to 26**

## Appendix B Comparison between Array Shapes

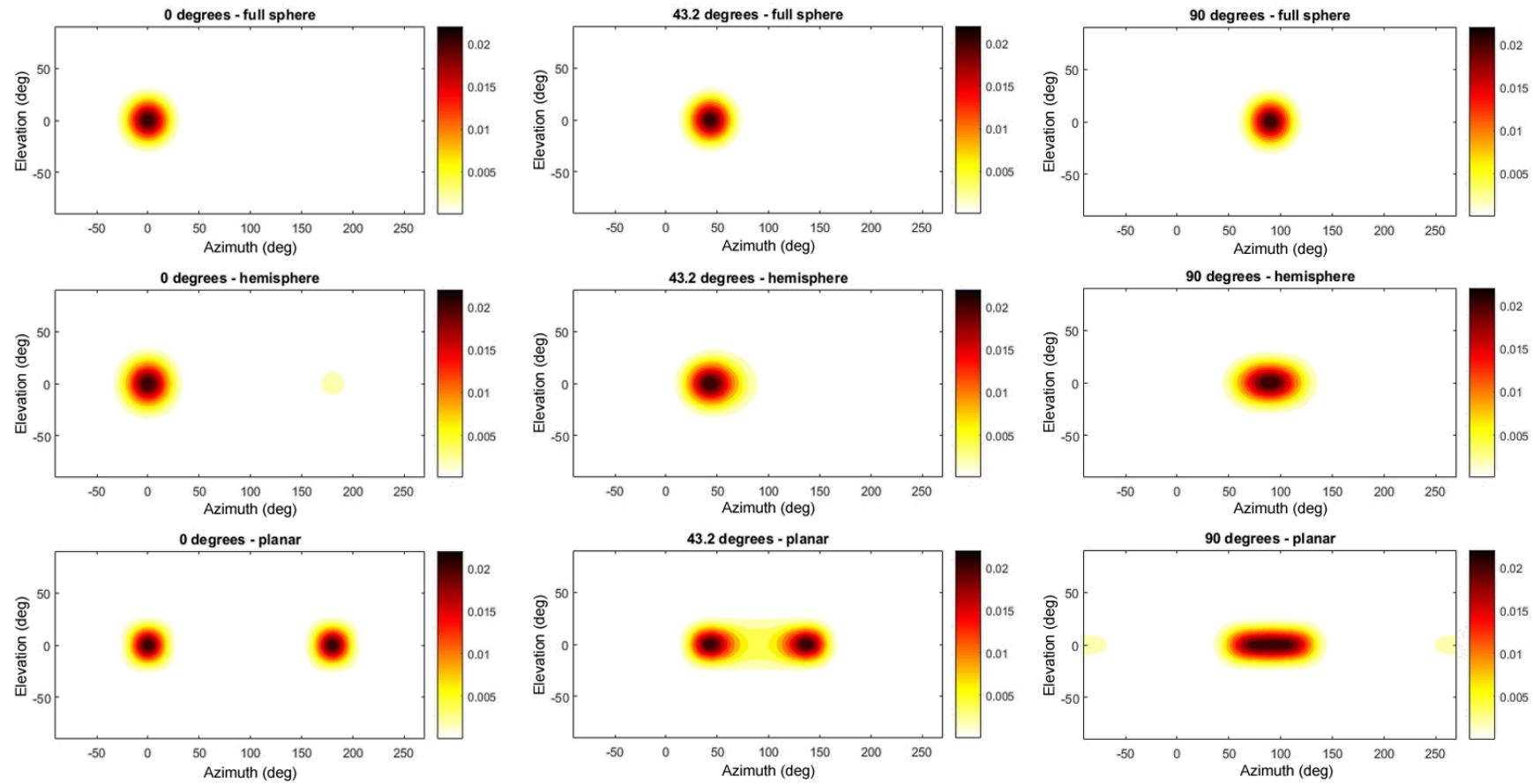
In order to make a comparison of directional response characteristics between a spherical, hemispherical and planar array, a free field propagation simulation was performed using three different microphone configurations, as shown in Figure 79. The hemispherical array (Figure 79 (b)) is identical to that described in Section 2.1. The spherical array has the same geometry as the hemispherical array, mirrored about the front plane (containing the center microphone). The planar array is a  $5 \times 5$  microphone square grid of 0.68 m side length (equivalent to the hemispherical array aperture or spherical array radius). All three arrays had the same microphone spacing.



**Figure 79. (a) Spherical, (b) hemispherical and (c) planar microphone array configurations for comparison**

Figure 80 show the beamforming output results for each array for a source of Gaussian white noise filtered to [1000 2000] Hz at 0 degrees, 43.2 degrees and 90 degrees. As expected, the spherical array showed no variation in beam width in all three cases and the hemispherical array showed a widening of the beam up to 90 degrees. The full extent of this widening as a function of

angle is shown in Section 2.4.5 (Figure 11). In the case of the planar array, the front-back ambiguity is notable – this property (which is common to all arrays arranged on a single 2D plane) limits the usage of planar arrays to applications with a maximum scanning range of  $180^\circ$  ( $\pm 90^\circ$ ).



**Figure 80. Comparison matrix of beamforming output for spherical, hemispherical and planar arrays at 0, 43.2 and 90 degrees**