

Low-Frequency Acoustic Source Detection and Localization

by

Arnav Joshi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Mechanical and Mechatronics Engineering

Waterloo, Ontario, Canada, 2024

© Arnav Joshi 2024

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In aviation, clear-air turbulence (CAT) is a major cause of in-flight injuries. It occurs in cloudless skies and cannot be detected by the onboard weather radar. Studies have predicted the extent of CAT to increase substantially in the next few decades, thus necessitating a method for detecting CAT.

With CAT known to generate low-frequency and infrasonic acoustic emissions, acoustic-based methods can potentially be deployed for detection and localization. This thesis studies low-frequency acoustic source detection and localization in the context of CAT.

Localizing low-frequency acoustic sources is challenging for acoustic beamforming which suffers from poor resolution at low source frequencies. A deep learning-based method is adopted as an alternative. Deep learning models for two-dimensional and three-dimensional acoustic source localization (ASL) have been built using synthetic data and computationally inexpensive neural network architectures. These models are necessary to prove the viability of deep learning for low-frequency ASL. The thesis then explores the potential of a deep learning-enabled, acoustics-based method for CAT detection in the future through a large-scale, virtual flight case, set up for the detection of a representative CAT source. The flight case tries to predict what an in-flight microphone will detect around a CAT source through a technique known as auralization which simulates the acoustic field of a source by modeling the sound propagation and determining what a receiver would hear.

The deep learning models yield promising qualitative and quantitative results that prove the feasibility of using deep learning for low-frequency ASL. Combined with the results from auralization, it can be concluded that there exists considerable scope for deep learning-enabled, acoustics-based detection and localization of CAT. The future work involves expanding the current scale of research with deeper network architectures to process real, in-flight acoustic data.

Acknowledgments

I would like to sincerely thank my supervisor Prof. Jean-Pierre Hickey for his immense contribution not just to this thesis, but to my entire journey as a Master's student at the University of Waterloo. He has been there to guide, support, motivate, and critique my work every step of the way. Working under his tutelage has been an absolute pleasure and a lifelong learning experience.

Dedication

I would like to dedicate this work to all the people close to me back home in India. They are my motivation and I work so that I can make them proud.

I would also like to dedicate this to sport, and in particular, to my football team Manchester City who have given me moments of great joy and something to always look forward to, in what was a bit of a lonely ride.

Table of Contents

Author’s Declaration	ii
Abstract	iii
Acknowledgments	iv
Dedication	v
List of Figures	viii
List of Tables	xi
1 Introduction	1
2 Literature Review and Background	5
2.1 Turbulence-Generated Acoustic Emissions	5
2.2 Acoustic Source Localization (ASL)	6
2.2.1 Acoustic Beamforming	6
2.2.2 Deep Learning-based ASL	7
2.3 Auralization of a Representative CAT Source	8

3	Deep Learning-based Low-Frequency Acoustic Source Localization (ASL)	10
3.1	Method	10
3.1.1	Microphone Array and Acoustic Source Simulation	11
3.1.2	Convolutional Neural Network	14
3.2	Test Case Setups	15
3.2.1	Case 1: Source Localization on a Scanning Plane	15
3.2.2	Case 2: Two-Dimensional ASL on the Horizon	17
3.2.3	Case 3: Three-Dimensional ASL	20
3.3	Results	22
3.3.1	Case 1	22
3.3.2	Case 2	26
3.3.3	Case 3	28
4	Auralization of a Representative CAT Source	31
4.1	Methodology	31
4.1.1	Simulated Flight Case and Atmospheric Properties	32
4.1.2	Curved Ray Tracing Method	34
4.1.3	Straight Ray Tracing	39
4.1.4	Auralization	40
4.2	Results	43
4.2.1	Eigenrays and Range of Detection	43
4.2.2	Propagation Time	47
4.2.3	Spectrograms	52
5	Conclusion and Scope of Future Work	58
	References	60

List of Figures

1.1	Comparison of beamforming result for point sources (located at $(0,0.14,0.3)$, $(0.15,-0.1,0.3)$, and $(-0.12,-0.15,0.3)$) emitting white noise signals over 1/3 octave bands with center frequencies of 8000 Hz and 2000 Hz, obtained using the open-source beamforming framework developed by Sarradj and Herold (2017)	3
3.1	Flow of data in the method. This structure is followed by all the test cases developed in this work.	11
3.2	ASL test cases. All the cases use the framework shown in Figure 3.1. . . .	11
3.3	A simulated 64-channel, logarithmic spiral-shaped microphone array.	12
3.4	Four-microphone array. The three outer microphones form an equilateral triangle while the fourth microphone is at the midpoint of the altitude (origin).	13
3.5	An example showing a generic convolutional neural network and the impact of convolution and pooling layers on the image size and information transfer.	15
3.6	Case 1: Acoustic source localization on a scanning plane. The scanning plane is discretized into a grid of uniform resolution. The finer the grid size, the higher the computational cost.	16
3.7	Case 2: Two-dimensional ASL on the horizon. Two possible source configurations S_1 and S_2 are shown. $\theta < 90^\circ$ for S_1 while $\theta > 90^\circ$ for S_2 . The cartesian coordinates of the source at these two positions are $(r \cos \theta, r \sin \theta)$ and $(-r \cos \theta, r \sin \theta)$ respectively.	18

3.8	GCC pattern for a pair of microphone signals polluted with white noise. The GCC feature map is symmetric about the origin. Therefore, the correlation coefficients can be extracted from either side of the origin.	20
3.9	Case 3: Three-dimensional ASL. The 4-microphone array is shown as a point. The localization domain is: $0^\circ \leq \theta \leq 180^\circ$, $0^\circ \leq \alpha \leq 90^\circ$. The location of the source in cartesian coordinates is $(r \sin(\alpha)\cos(\theta), r \sin(\alpha)\sin(\theta), r \cos(\alpha))$	21
3.10	A multi-task CNN used for three-dimensional ASL. The main branch performed the convolution and pooling operations on the input feature. The two output branches peeled off the main branch, using the flattened feature vector as the input to give separate predictions for θ and α	22
3.11	Configuration 1 (300 Hz)	24
3.12	Configuration 2 (300 Hz)	24
3.13	Sparse sources (100 Hz)	24
3.14	Clustered sources (100 Hz)	24
4.1	Auralization methodology	32
4.2	Setup, sub-case 1: $y = -1$ km	33
4.3	Curved ray at $x = -50,000$ m and $y = -1,000$ m for the simulated flight case	36
4.4	Total frequency-dependant attenuation due to atmospheric absorption shown for the curved ray in Figure 4.3.	38
4.5	Straight ray at $x = -50000$ m and $y = -1000$ m for the simulated flight case	39
4.6	Eigenrays for straight ray tracing	43
4.7	Eigenrays for curved ray tracing	44
4.8	$x - z$ plane projection	44
4.9	$y - z$ plane projection	44
4.10	Example showing shadow zone formation for upward refraction. The shadow zone starts approximately 20 km from the source.	45

4.11 Eigenrays for the longer flight path: $-100 \text{ km} \leq x \leq 100 \text{ km}$.	46
4.12 Shadow zone formation: $y - z$ plane projection	47
4.13 Range of detection: $x - y$ plane projection	48
4.14 Propagation times for straight and curved rays	49
4.15 Difference in propagation times (straight ray - curved ray)	49
4.16 Frequency shift of a unit source frequency (1 Hz) as observed by the receiver.	51
4.17 Range of θ vs $ y $	52
4.18 $ y = 1 \text{ km}$, straight ray tracing (left) and curved ray tracing (right) spectrograms	53
4.19 $ y = 3 \text{ km}$, straight ray tracing (left) and curved ray tracing (right) spectrograms	53
4.20 $ y = 5 \text{ km}$, straight ray tracing (left) and curved ray tracing (right) spectrograms	54
4.21 $ y = 7.5 \text{ km}$, straight ray tracing (left) and curved ray tracing (right) spectrograms	54
4.22 $ y = 10 \text{ km}$, straight ray tracing (left) and curved ray tracing (right) spectrograms	55
4.23 $ y = 15 \text{ km}$, straight ray tracing (left) and curved ray tracing (right) spectrograms	55

List of Tables

3.1	Summary of the convolutional neural network models used in all the cases	23
3.2	Source localization statistics for 1000 test samples. A source is assumed to be located if the predicted source configuration detects a strength of 0.3 Pa or higher at the source location.	25
3.3	Prediction accuracy for azimuth angle (θ) in two-dimensional ASL with a monopole source (S1) at 100 Hz. The accuracy values are for a test set of 10000 samples.	26
3.4	Prediction accuracy for azimuth angle (θ) in two-dimensional ASL with a sinusoidal plane wave source (S2) at 10 Hz. The accuracy values are for a test set of 10000 samples.	27
3.5	Prediction accuracy for θ and α in three-dimensional ASL with a monopole source (S1) at 100 Hz. The accuracy values are for a test set of 10000 samples.	29
3.6	Prediction accuracy for θ and α in three-dimensional ASL with a sinusoidal plane wave source at 10 Hz (S2). The accuracy values are for a test set of 10000 samples.	29
4.1	Peak frequency, peak time, and peak power detected by straight ray tracing	56
4.2	Peak frequency, peak time, and peak power detected by curved ray tracing	56

Chapter 1

Introduction

Acoustic source detection and localization is an innate ability of humans. Our sense of hearing naturally equips us with the ability to “detect” sound, and if possible, “locate” the source of sound. Detection can thus be viewed as the first step of localization. A formal distinction between these two related yet slightly different processes can thus be put forth: acoustic source detection refers to sensing acoustic signals from a source in an environment whereas acoustic source localization (ASL) is the process of determining the location - the exact point, direction, or area in physical space - of the source generating the sound ([Jekaterýńczuk and Piotrowski, 2023](#)). It is a powerful inherent ability, but not powerful enough as the human range of detection and localization is limited to the audible frequency range (20 Hz-20000 Hz) and simple source distributions. And so as humans, we do what we have always done to address our physical limitations: turn to our ally in technology which in this case, consists of complex microphone infrastructure and source localization algorithms.

The famous quote by Albert Einstein, “Look deep into nature and then you will understand everything better”, serves as a reminder of the purpose of science - to try and explain how the natural world works. Viewing nature with our acoustic hat on, nature communicates with us through the lower end of the frequency spectrum. Turbulence, oceans, earthquakes, thunderstorms, and volcanic eruptions, are all known to produce low-frequency (≤ 300 Hz) sound and infrasound (≤ 20 Hz). Thus, the detection and local-

ization of low-frequency and infrasonic acoustic sources is an important topic in the fields of aeroacoustics (Watson et al., 2021), ocean acoustics (Peyvandi et al., 2011; Carter, 1981; Fernandes et al., 2022), and seismology (Tosi et al., 2012; Hill et al., 1976; Sylvander et al., 2007).

As people working in fluids, we are naturally drawn to aeroacoustics in which the theme of source detection and localization is explored through turbulence-generated acoustic emissions. While sound generation from isotropic turbulence is well-studied (Lighthill, 1952), recent works (Daryan et al., 2022) have also looked to understand the fundamental interaction mechanism of vortical structures which is the main source of aeroacoustic noise.

In aviation, a type of turbulence encountered by flights known as clear-air turbulence (CAT) is one of the major causes of in-flight injuries. As recently as May 21, 2024, a Singapore Airlines flight from London to Singapore encountered severe turbulence, resulting in multiple casualties. CAT is encountered in cloudless skies, typically at altitudes of 5.6 km or higher (Ellrod et al., 2003) and occurs when jet streams traveling at different speeds meet, resulting in a strong change in wind direction or wind shear. Research carried out at the University of Reading has found that the turbulence in the North Atlantic jetstream has increased by 50% between 1979 and 2020. Furthermore, climate change and warmer temperatures due to global warming are strengthening the jetstreams and increasing the wind shear, making CAT more frequent and severe. The research projects the amount of CAT in the atmosphere to double or triple in the next few decades and this is likely to present a major challenge to the aviation industry. Typically, an onboard weather radar detects precipitation ahead of the aircraft to estimate the cloud’s size and the extent of turbulence. However, the cloudless nature of CAT makes it invisible to radar. Currently, the only way to detect CAT is through pilot reports.

It is well-documented that CAT emits low-frequency and infrasonic emissions. Acoustics-based methods can potentially be used to detect and locate CAT in the future. These methods are low in cost, robust, and easily deployable (Sutin et al., 2013), thus making them an avenue worth exploring.

While a low-frequency acoustic signal may be detected by a microphone system, the localization of the signal has proved to be challenging with the current source localization

methods such as acoustic beamforming. Beamforming tends to suffer from poor resolution at low source frequencies. Figure 1.1 shows the increasingly poor localization predictions as the source frequency decreases. The resolution is even poorer for the low-frequency range we are interested in. This limitation of beamforming causes us to look for an alternative method for low-frequency source localization.

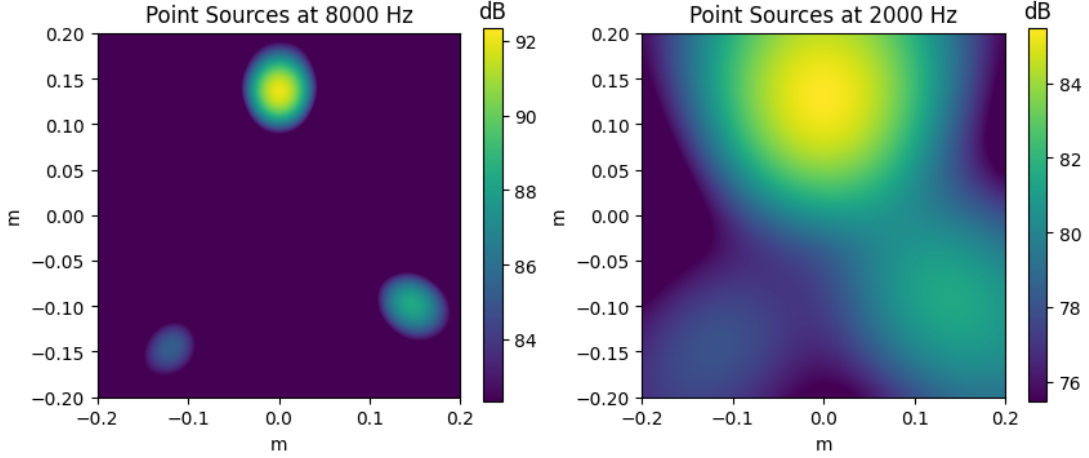


Figure 1.1: Comparison of beamforming result for point sources (located at $(0, 0.14, 0.3)$, $(0.15, -0.1, 0.3)$, and $(-0.12, -0.15, 0.3)$) emitting white noise signals over 1/3 octave bands with center frequencies of 8000 Hz and 2000 Hz, obtained using the open-source beamforming framework developed by [Sarradj and Herold \(2017\)](#).

Due to advancements in computational science and data, machine learning (ML), and deep learning (DL) in particular, have shown great promise in the field of acoustics ([Bianco et al., 2019](#); [Grumiaux et al., 2022](#); [Yalta et al., 2017](#)). Despite the limitations of these methods, such as requiring large amounts of data and not being physically intuitive, the ability of deep learning algorithms to extract features from data in any shape or form provides a lot of scope for their application in acoustic source localization (ASL).

Thus, this thesis builds deep learning models for the localization of low-frequency sources as an alternative to beamforming. These models cover three general cases of two-dimensional and three-dimensional ASL and are built using limited synthetic data and

computationally inexpensive neural network architectures. The models collectively prove the feasibility of using deep learning for low-frequency ASL.

Furthermore, to explore the scope and viability of a deep learning-enabled, acoustics-based method for CAT detection (and localization) in the future, this thesis sets up a large-scale, virtual flight case for the detection of a representative CAT infrasound source. We attempt to predict what an in-flight microphone will detect when approaching the CAT source through a technique known as “auralization”. Auralization is the acoustic counterpart of visualization and refers to simulating the acoustic field of a source by modeling the sound propagation and determining what a receiver would hear.

Modeling the long-range propagation of sound in the atmosphere, which is the most important step in auralization, is done using a method known as *ray tracing* which treats sound waves as rays. The ray tracing method used in this thesis (developed by Schäfer and Vorländer (2021)) tries to model as accurately as possible, the nature of the sound detected by taking into account the turbulence spectrum, wind advection, atmospheric refraction, geometric attenuation, and the Doppler frequency shift experienced by the sound.

As such, the objectives of this thesis are:

- Building deep learning models for localization of low-frequency acoustic sources as an alternative to beamforming using synthetic data and computationally inexpensive networks.
- Auralizing a representative CAT source in a large-scale simulated flight case using ray tracing.

The fulfillment of these objectives will help provide a holistic understanding of low-frequency acoustic source localization in the context of CAT detection.

The rest of this thesis is organized as follows. Chapter 2 gives a literature review and a detailed background of the various aspects of the topic. Chapter 3 then discusses the methodology and results obtained from deep learning-based acoustic source localization, followed by Chapter 4 which does the same for the auralization of a representative CAT source. The conclusion and scope of future work are given in Chapter 5.

Chapter 2

Literature Review and Background

Various aspects collectively make up our understanding of acoustic source localization (ASL) in the context of CAT detection. These aspects have been systematically presented in this chapter to facilitate the reader to get a grasp on the flow of the thesis.

To contextualize our work, a review of sound generation by turbulence is presented. This is followed by acoustic source localization in which we first discuss beamforming and its limitations after which a detailed background of deep learning-based ASL is provided. Finally, auralization and ray tracing in particular, are discussed in depth to round off the chapter.

2.1 Turbulence-Generated Acoustic Emissions

Sound emission from turbulence was studied theoretically by [Lighthill \(1952\)](#) who derived an expression for the acoustic pressure from a patch of isotropic turbulence. [Daryan et al. \(2022, 2020\)](#) explored the principle mechanism behind sound generation by turbulence which is the self and mutual interactions of vortical structures such as vortex reconnection. [Hardin and Wang \(2003\)](#) developed vortex dynamic models to characterize sound generation by aircraft wake vortices and discussed various sources of sound tied to the excitation of the vortex core. Acoustic emissions from vortices have been observed in vari-

ous wake vortex detection experiments carried out at airports ([Michel and Böhning, 2002](#); [Booth and Humphreys, 2005](#); [Dougherty et al., 2004](#); [Wang et al., 2005](#)).

Regarding the spectra of these vortices, [Rubin \(2005\)](#) and [Zuckerwar et al. \(2013\)](#) have both detected infrasonic emissions from wake vortices using highly sensitive infrasonic microphones. [Wescott \(1964\)](#); [Posmentier \(1974\)](#); [Shams et al. \(2013\)](#) have also provided evidence of low-frequency acoustic emissions from atmospheric turbulence. [Wescott \(1964\)](#) captured noise frequencies in the range of 0.2-200 Hz from atmospheric turbulence at high altitudes using free-floating balloons carrying acoustic probes. [Posmentier \(1974\)](#) recorded infrasound in the 1-16 Hz frequency range using a ground-based, four-microphone array for nearly two months and hypothesized, based on spectral analysis and turbulence observations, that the frequency range could be associated with CAT. [Shams et al. \(2013\)](#) also experimentally verified the presence of infrasound from CAT using a ground-based infrasonic microphone array and validated it with pilot reports.

2.2 Acoustic Source Localization (ASL)

2.2.1 Acoustic Beamforming

Beamforming ([Chiariotti et al., 2019](#); [de Santana, 2017](#); [Gombots et al., 2021](#)) uses a microphone array that can focus on different points in space to capture the signal coming from different directions. A delay is added to each microphone signal such that if the array points in the direction of the source, constructive interference occurs upon the addition of all microphone signals, and destructive interference occurs otherwise. By moving the array from point to point, a source distribution map can be obtained. This formulation is normally done in the frequency domain as it enables the application of deconvolution techniques such as DAMAS ([Brooks and Humphreys, 2006](#)) and CLEAN-SC ([de Santana, 2017](#)) to give better results.

Beamforming however suffers from poor resolution at low frequencies. This limitation can be attributed to the Rayleigh resolution criterion, originally given by [Rayleigh \(1879\)](#) for the angular resolution of an optical system but can be applied analogously to an acoustic

system.

According to the criterion, two sources of light can be resolved if the central maxima of the diffraction pattern of one source falls directly on the first minima of the diffraction pattern of the other. The first minima in the diffraction pattern lies at an angle of $\theta = 1.22\lambda/D$ on either side of the maxima, where λ is the wavelength of light and D is the diameter of the circular aperture through which the sources are observed. Applying these analogous principles to acoustic beamforming gives us the Rayleigh Resolution Limit (RRL) which is the minimum required distance between two sources for them to be resolved. This is given as (Xu et al., 2021):

$$\text{RRL} \approx 1.22|z|\frac{c_0}{fD} \quad (2.1)$$

Here $|z|$ is the distance between the source and the microphone, and c_0 and f are the speed and frequency of sound respectively. It can be seen from the expression that the resolution limit is inversely proportional to the frequency. In other words, lower-frequency sources would have to be much farther apart to be resolved by the beamformer. Due to this limitation of beamforming, we turn to deep learning-based approaches.

2.2.2 Deep Learning-based ASL

A general study of the literature on deep learning-based acoustic source localization (ASL) offers insights into the different kinds of input features and network architectures used, which can then be applied to low-frequency ASL. Most of the research in deep learning-based ASL is focused on the localization of speech signals. While most works process the acoustic data before feeding it to a network, Vera-Diaz et al. (2018) were the first to use raw audio signals as input to a convolutional neural network (CNN) (O’Shea and Nash, 2015) to predict the position of the three-dimensional source. Chakrabarty and Habets (2017) and Xiao et al. (2015) used the phase component of the short-time Fourier transform (STFT) and Generalized Cross-Correlation with Phase Transform (GCC-PHAT) respectively, as input for estimating the direction of arrival (DoA) of speech signals.

Studies focusing on a source frequency range such as Ma and Liu (2019) simulated a phased microphone array and monopole sources at high frequencies (3000-8000 Hz) on a

scanning plane and used the cross-spectral matrix (CSM) as input to a CNN for localization of these sources. They were able to demonstrate excellent spatial resolution, comparable to that of DAMAS (Brooks and Humphreys, 2006), and much better than conventional beamforming. However, they did not address the challenges at lower source frequencies. Xu et al. (2021) employed a similar setup and methodology, but focused on a wider source frequency range of 200-20000 Hz and were able to predict the source distribution even for low frequencies. However, they used a densely connected convolutional network (DenseNet) (Huang et al., 2017) that required large amounts of data and memory for training. Niu et al. (2019) carried out ocean source localization in the frequency range of 100-200 Hz using a single hydrophone and the discrete Fourier transform of the raw pressure data. 50-layer residual networks (ResNet-50), which also needed a significant amount of training data, were separately trained to determine the range and depth of the source.

While the literature has touched upon lower source frequencies, there are not many works, to the best of our knowledge, that focus exclusively on low-frequency and infrasound localization. The ones that address it (Xu et al., 2021; He et al., 2016) use very deep and computationally expensive network architectures. Thus, in this thesis, we present deep learning models for three reproducible test cases covering two-dimensional and three-dimensional ASL, to show the viability of deep learning for low-frequency ASL. The models are built using limited synthetic data and computationally inexpensive networks. We adopted a relatively shallow CNN and trained it with an input feature suited to each case. The test cases, which contain key features of applied ASL problems, seek to optimize low-frequency ASL under the imposed computational limits and are intended to serve as benchmark cases for studying and further developing low-frequency ASL using deep learning. The methodology and results are discussed in Chapter 3.

2.3 Auralization of a Representative CAT Source

Modeling the propagation of sound is the most important step in auralization. Sound propagation in the atmosphere is commonly modeled using ray tracing, a method based on ray theory (Schäfer and Vorländer, 2021). Originally derived in optics, ray theory

simplifies the wave equation so that light waves can be assumed to propagate as rays under the condition of the wavelength becoming very small. This assumption gave rise to geometrical optics, a sub-field of optics. These principles can be analogously applied to sound waves that too can be assumed to traverse as rays under the same limiting condition. In our case, although low-frequency and infrasonic wavelengths are not small, the ray theory can still be applied (Groves, 1955). Geometrical methods are applicable as long as the change in the refractive index of the medium is small over the distance of the wavelength and the duration of the period of the wave. As changes in the state of the atmosphere are quite small over infrasonic wavelengths and periods, the geometrical theory remains valid for infrasound propagation (Groves, 1955).

Ray tracing frameworks typically assume the ray path between the source and the receiver, which is known as an eigenray (Ostashev and Wilson, 1997), to be straight (Rizzi et al., 2014), essentially modeling propagation as free-space propagation from a monopole. The monopole has a spherical wavefront. Straight ray tracing thus is efficient and easy to implement. However, eigenrays are rarely straight lines in the atmosphere due to the constantly changing refractive index, governed by wind and temperature gradients (Ingård, 1953). Atmospheric refraction causes the rays to be curved instead of straight and the spherical waveform to be distorted which can have a significant effect on the received sound (Reynolds, 1874).

Schäfer and Vorländer (2021) developed an open-source atmospheric ray tracing (ART) framework that considers the effects of both refraction and advection to provide an efficient way of calculating the eigenrays in a stratified moving medium. We utilized the ART framework to model the sound propagation using curved ray tracing for our flight case consisting of a stationary, broadband, infrasound source (representative of CAT) and a moving receiver (airplane). We also modeled the propagation using straight ray tracing and compared the results with curved ray tracing. From the eigenrays, we derived various acoustic parameters to generate the signal that would be received by an in-flight microphone, thus auralizing the representative CAT source. A spectrogram of this signal was plotted for further analysis. The methodology and results are discussed in Chapter 4.

Chapter 3

Deep Learning-based Low-Frequency Acoustic Source Localization (ASL)

This chapter discusses the development of deep learning models for low-frequency ASL. Section 3.1 describes the methodology of ASL. Section 3.2 provides the details of the three ASL test cases that cover two-dimensional and three-dimensional ASL, followed by Section 3.3 which discusses the results obtained from these cases.

3.1 Method

The methodology used for low-frequency ASL in this work can be described in three steps. The first step involves capturing the synthetic acoustic pressure data using a simulated microphone array. This data is then processed, either in the time or frequency domain, to generate an input feature capable of being fed to a CNN. Finally, the CNN is trained with the input feature against the ground truth and is used to make predictions. The method is summarized in Figure 3.1.

Figure 3.2 shows a flowchart of the three ASL test cases developed in this work. Case 1 (inspired by (Xu et al., 2021; Ma and Liu, 2019)) is two-dimensional ASL on a scanning

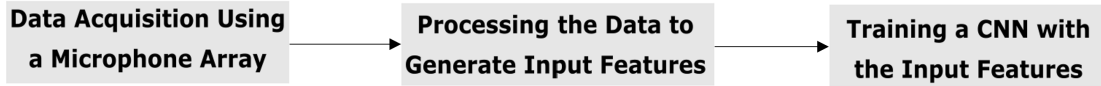


Figure 3.1: Flow of data in the method. This structure is followed by all the test cases developed in this work.

plane parallel to the microphone array plane. This case can be applied to the localization of aeroacoustic sources, such as wake vortices that emit low-frequency noise. Case 2 (source localization on the horizon), also two-dimensional, tackles ASL as a Direction-of-Arrival (DoA) estimation problem. Case 3, which is an extension of case 2, deals with three-dimensional ASL. Cases 2 and 3 are inspired by [Shams et al. \(2013\)](#)’s work on the development of an early-warning CAT detection system using a ground-based microphone array. All the cases are discussed in detail in Section 3.2.

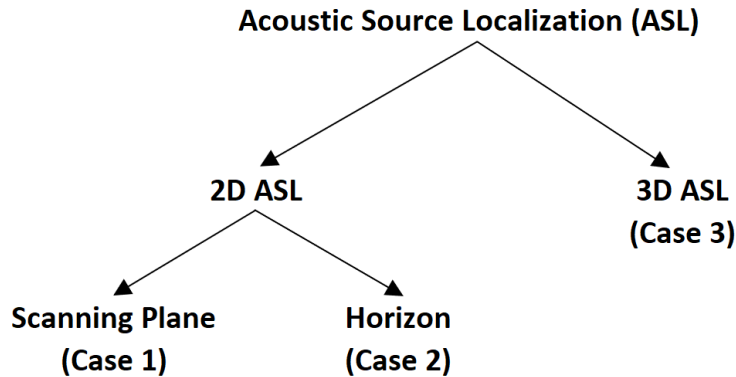


Figure 3.2: ASL test cases. All the cases use the framework shown in Figure 3.1.

3.1.1 Microphone Array and Acoustic Source Simulation

Microphone Array

Two types of microphone arrays are simulated. The first type, used for case 1, is a 64-microphone array in the shape of a logarithmic spiral. A spiral is naturally irregular and has unique inter-microphone spacing which is needed to avoid spatial aliasing, thus making it

a popular architecture for microphone arrays (Prime and Doolan, 2013). The specific type of logarithmic spiral used in this work is called the *Arcondoulis* spiral which performs well over a range of frequencies. The equation of the spiral is $r = ae^{b\theta}$ wherein the parameters a and b were set as $a = 0.018$ and $b = 0.002$ to obtain the microphone array as shown in Figure 3.3.

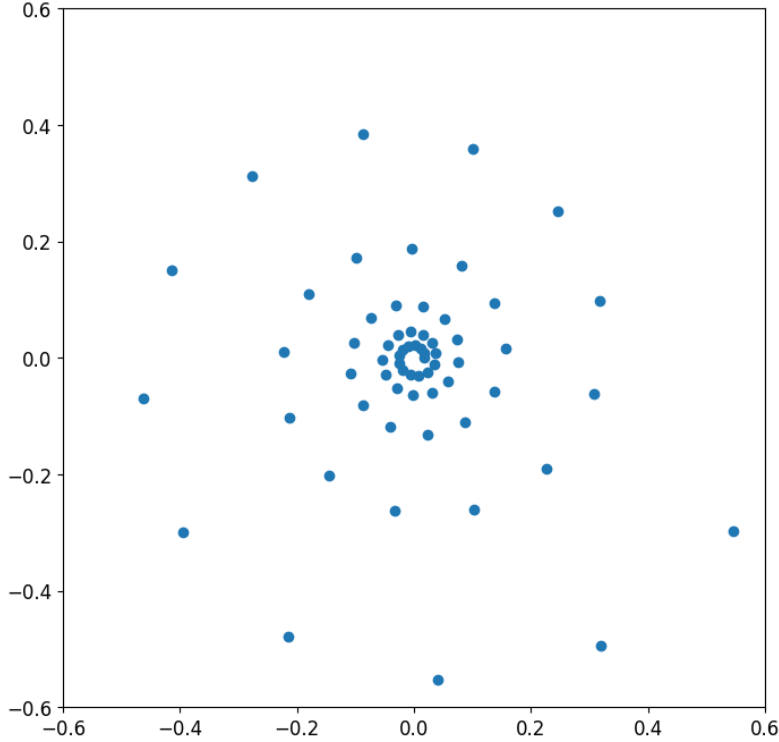


Figure 3.3: A simulated 64-channel, logarithmic spiral-shaped microphone array.

Using a large number of microphones may not always be economically viable nor is it necessary to capture low-frequency content. Thus we also simulated a more minimalistic array. The microphone array used for cases 2 and 3 is a 4-microphone array as shown in Figure 3.4. The type of array chosen depends on whether the case is modeled as a regression or classification task. Case 1 is modeled as a regression problem to accurately predict the location of the sources. Due to its complexity, the array with more microphones is chosen. Cases 2 and 3 are classification tasks that involve estimating the DoA of the

signal. They are relatively less complex and thus use the 4-microphone array.

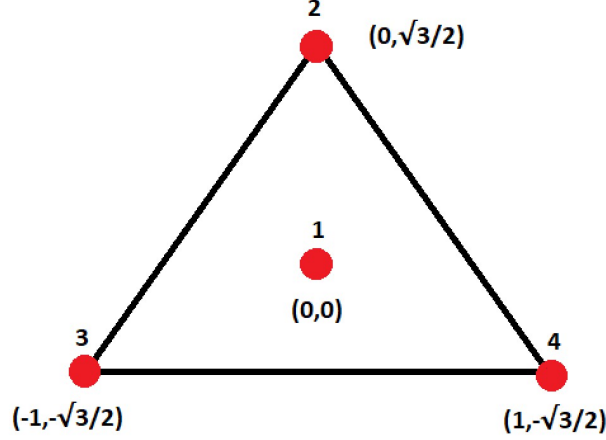


Figure 3.4: Four-microphone array. The three outer microphones form an equilateral triangle while the fourth microphone is at the midpoint of the altitude (origin).

Acoustic Source

We simulated two types of acoustic sources with different levels of complexity. Source 1 (S1) (eq. (3.1)) is an analytically-defined monopole that oscillates at a fixed frequency f . $P_s(m)$ (Xu et al., 2021) shows the complex acoustic pressure due to a monopole s as detected by a microphone m located at a distance r_s from the source. Here, Q is the complex source strength, and c_0 is the speed of sound in the air, both assumed to be constant. S1 exhibits geometric decay through distance-based attenuation of the complex pressure $P_s(m)$. It is computationally inexpensive and allows us to test models rapidly.

$$P_s(m) = Q \frac{e^{-j2\pi f r_s / c_0}}{4\pi |r_s|} \quad (3.1)$$

The second type of source, Source 2 (S2) (eq. (3.2)), is also an analytical source: a time-domain sinusoidal plane wave signal of amplitude A , frequency f , and phase difference ϕ . This signal is assumed to be detected by the microphones. The phase difference for a microphone with respect to another is given as $\phi = 2\pi f \tau$, where τ is the time delay or the

difference in times when the signal is detected by the pair of microphones. In our case, the phase difference for all microphone signals is defined with respect to the microphone at the origin. The signal is also polluted with uncorrelated white noise.

$$P(t) = A \sin(2\pi ft + \phi) + Noise \quad (3.2)$$

Being a plane wave, S2 does not undergo any geometric decay. Unlike S1 which is strictly defined in the frequency domain, S2 can be used for both time- and frequency-domain processing.

3.1.2 Convolutional Neural Network

A shallow convolutional neural network (CNN) similar to the generic CNN shown in Figure 3.5 is adopted for all three cases in this work. A CNN typically finds its application in image classification as it can extract features from images using a combination of convolution and pooling layers.

Figure 3.5 shows the working of a CNN through convolution and pooling operations performed on a $(32 \times 32 \times 1)$ sized image/input feature. The stride values for the convolution and pooling layers are 1 and 2, respectively. Once the image is fed to the CNN, the convolution layer applies a series of filters to help the network capture the various characteristics of the image. The image size thus changes after the convolution operation. For an image of dimensions $I \times I$, filter width $F \times F$, zero-padding P , and stride S , the output dimensions of the image are $\left\lceil \frac{(I-F+2P)}{S} + 1 \right\rceil \times \left\lceil \frac{(I-F+2P)}{S} + 1 \right\rceil$ (Albawi et al., 2017).

This is followed by a pooling layer which downsamples the image, thereby reducing its size while preserving the dominant features. The dimensions of an $I \times I$ image after a $F \times F$ max-pooling filter operation are $\left\lceil \frac{(I-F)}{S} + 1 \right\rceil \times \left\lceil \frac{(I-F)}{S} + 1 \right\rceil$. The resulting image is then flattened to a vector and fed to a fully connected layer as part of a regular feed-forward network. The loss and activation functions in the final layer are chosen depending on the nature of the case i.e. if it is a regression problem or a classification problem. The CNN compares its prediction with the ground truth, calculates the loss/error, and adjusts its learnable parameters to minimize the loss.

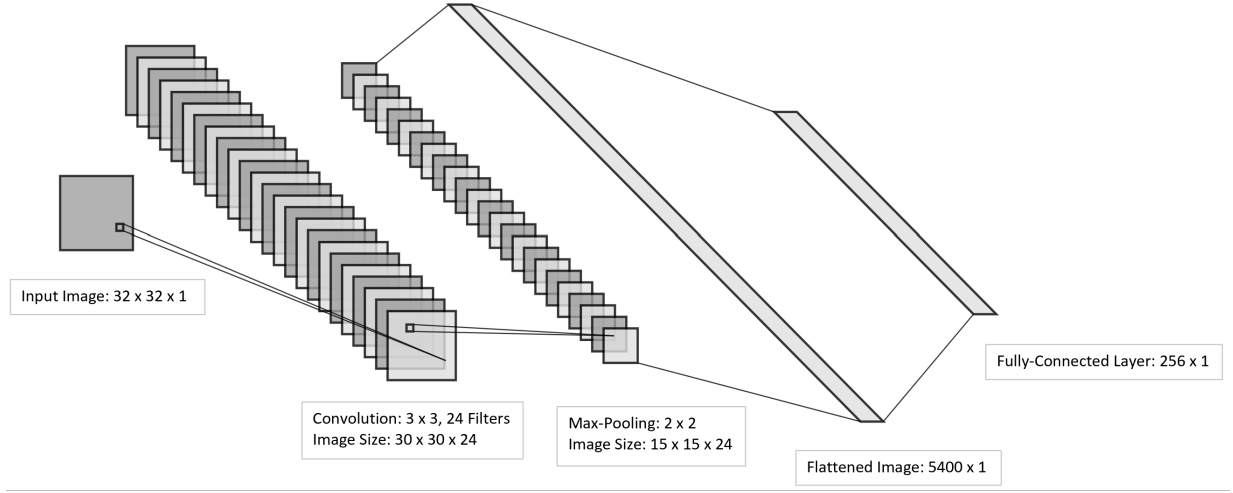


Figure 3.5: An example showing a generic convolutional neural network and the impact of convolution and pooling layers on the image size and information transfer.

3.2 Test Case Setups

3.2.1 Case 1: Source Localization on a Scanning Plane

This case takes inspiration from the work done by (Xu et al., 2021; Ma and Liu, 2019). Xu et al. (2021) in particular, carried out source localization on a scanning plane parallel to the microphone array plane for a wide frequency range of 200-20000 Hz, using a deep neural network called DenseNet-201. They were able to show the superiority of DenseNet-201 over other classic acoustic imaging methods. However, the DenseNet needed a significant amount of simulated data, memory, and processing power to train.

In this work, we focused on ASL at two frequencies in the low-frequency range (100 Hz and 300 Hz) using limited training data and a shallow CNN to prove that ASL is feasible using a computationally less expensive approach. The setup is shown in Figure 3.6.

We used a scanning plane of dimensions $1.2m \times 1.2m$. The scanning plane was located 1.2 m above the plane containing the 64-microphone logarithmic spiral array. To generate the training dataset, the scanning plane was divided into an $N \times N$ scanning grid and S

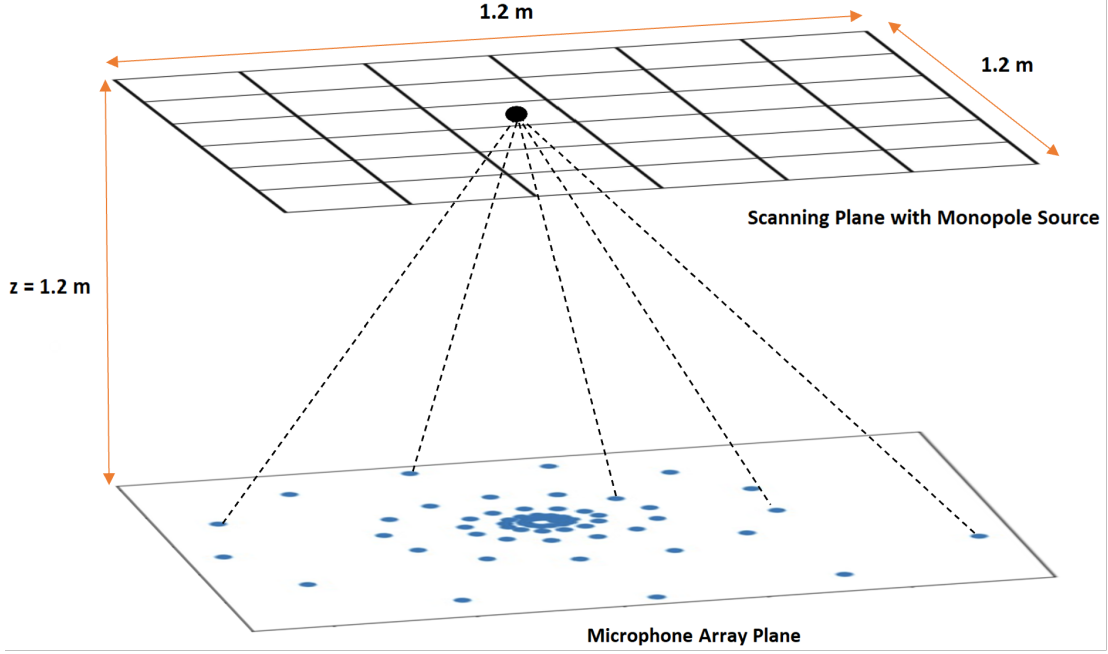


Figure 3.6: Case 1: Acoustic source localization on a scanning plane. The scanning plane is discretized into a grid of uniform resolution. The finer the grid size, the higher the computational cost.

number of monopole sources (S , equation (3.1)) with unit source strength were randomly distributed throughout the grid. The acoustic pressure of all the sources was added at all the microphones (M) to form the combined pressure vector P given as:

$$P = \left[\sum_{s=1}^S P_s(1), \sum_{s=1}^S P_s(2), \sum_{s=1}^S P_s(3), \dots, \sum_{s=1}^S P_s(M) \right] \quad (3.3)$$

The vector P , in turn, was used to calculate the cross-spectral matrix (CSM) as:

$$CSM = PP^H \quad (3.4)$$

where P^H is the conjugate transpose of P . The dimensions of the CSM are $M \times M$ which in this case, is 64×64 . The CSM represents the power distribution over frequency, and we used it as an input feature to the CNN. The CNN was trained against the ground truth *i.e.* the source distribution from which the CSM was derived, to predict the strength and

location of the sources. For the ground truth, the $N \times N$ scanning grid was converted to an $N^2 \times 1$ binary vector in which the elements corresponding to the sources were assigned the value one (equal to the source strength), and the elements without a source were assigned the value zero. As this is a regression problem, we defined the output layer as a dense layer with N^2 nodes and without any activation function. We used the Rectified Linear Unit (ReLU) activation function everywhere else in the network. The network prediction is a vector of output layer dimensions ($N^2 \times 1$) that is compared against the ground truth vector to calculate the loss. The loss function used was the mean squared error (MSE) defined as:

$$MSE = \frac{\sum_{i=1}^n (y_{pred} - y_{gt})^2}{n} \quad (3.5)$$

where, y_{pred} is the predicted vector given by the network, y_{gt} is the ground truth vector, and n is the length of the two vectors (N^2). The optimizer used was *ADAM* (Kingma and Ba, 2014) with the default values of the hyperparameters. Details of the exact CNN model and the training parameters are given in Section 3.3.

3.2.2 Case 2: Two-Dimensional ASL on the Horizon

This case used a classification approach to determine the direction of arrival (DoA) of the acoustic signal from a source located on the horizon. The four-microphone array was used for this case. As the DoA, the azimuth angle (θ) had to be determined so the range of possible values of θ (0° - 180°) was discretized into N classes. The output of the network is a probability distribution of all the classes, and the class with the highest probability is the estimated DoA. The setup is shown in Figure 3.7.

Since this case uses a classification approach, the final dense layer of the network had the softmax activation function which normalized the output of the network to give the probability distribution over N classes. ReLU activation is used everywhere else. *ADAM* was used as the optimizer while the loss function was crossentropy which is the standard loss function for classification problems (Ho and Wookey, 2019). We used both types of sources in this case, the monopole and the sinusoidal plane wave, and they have been discussed separately.

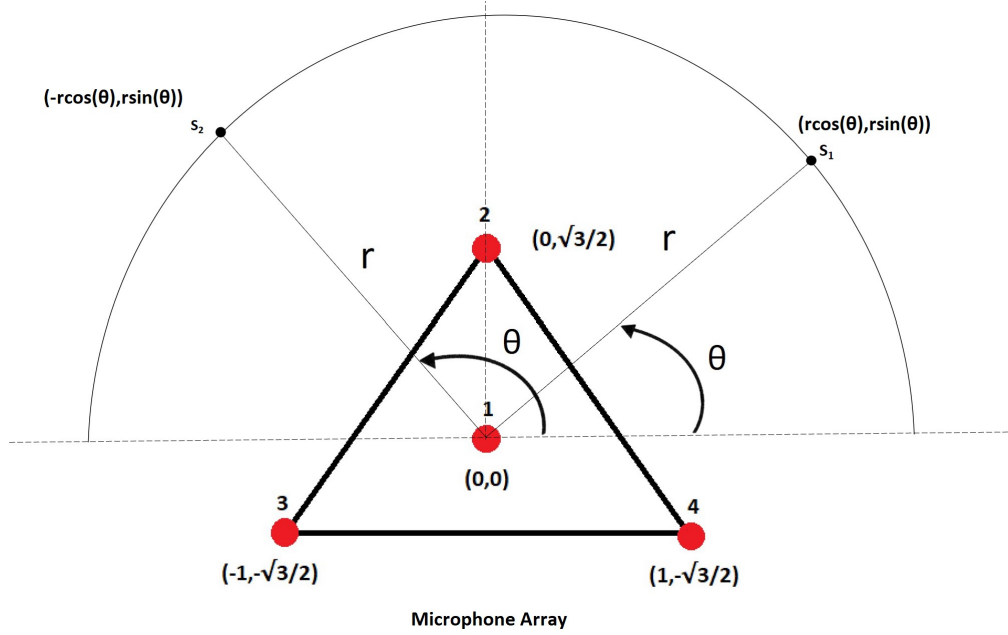


Figure 3.7: Case 2: Two-dimensional ASL on the horizon. Two possible source configurations S_1 and S_2 are shown. $\theta < 90^\circ$ for S_1 while $\theta > 90^\circ$ for S_2 . The cartesian coordinates of the source at these two positions are $(r \cos \theta, r \sin \theta)$ and $(-r \cos \theta, r \sin \theta)$ respectively.

Case 2(i): Monopole (S1)

To generate the training set, a monopole with unit source strength was placed at a real, random value of θ (in the range of θ) and a fixed radial distance $r(= 10 \text{ m})$ from the origin. The cross-spectral matrix was calculated using the same procedure as case 1 and was used as the input to a CNN. The value of θ falls into a certain class and the index of this class was the ground truth against which the network was trained.

Case 2(ii): Sinusoidal Plane Wave (S2)

A sinusoidal plane wave signal (equation (3.2)) was generated at a real, random value of θ and fixed $r(= 10 \text{ m})$. The signal was polluted with white Gaussian noise. The amplitude of the pressure signal and the white noise were equal and set to 10 Pa. We applied the

generalized cross-correlation (GCC) algorithm to all microphone pairs to generate an input feature. GCC is a time delay-based localization approach as opposed to beamforming which is a frequency-domain approach. GCC is known to be robust to noise and reverberation. The generalized cross-correlation of two signals $r_1(t)$ and $r_2(t)$ is defined as the inverse Fourier transform (IFFT) of the cross-power spectrum (Varzandeh et al., 2020) -

$$\text{GCC} = \text{IFFT} (R_1 \odot R_2^*)$$

where, R_1 and R_2 are the short-time Fourier transforms of r_1 and r_2 respectively, $\odot$ represents element-wise multiplication, and $*$ represents the complex conjugate. Figure 3.8 shows an example of a GCC pattern obtained using a pair of microphone signals polluted with white noise. It can be seen that while the signals are distorted, the GCC pattern is still very evident.

The entire GCC vector was not used as input to the CNN. For each microphone pair, there is a certain number of correlation coefficients near the center that contain the necessary information for localization. These coefficients are extracted from the GCC feature map to be used as input to the CNN. The calculation for the number of coefficients to be extracted is as follows: The maximum possible time delay (τ) between two microphones signals in the four-microphone array is the maximum distance between two microphones (2 m) divided by the speed of sound (343 m/s). Therefore $\tau = 2/343 = 5.83$ ms. For a sampling frequency of 16,000 Hz, the maximum delay in terms of samples is $n = 16000 \tau \approx 93$. So, the first 93 coefficients from the center of the GCC map for each microphone pair contain the information needed for localization. The first 93 coefficients from the GCC maps of all microphone pairs (6) were concatenated to form a vector of size (558×1) and then reshaped into a matrix of dimensions (31×18) . This reshaped matrix was used as input to the CNN. Of course, the vector of GCC coefficients can be directly fed to a multi-layer perceptron (MLP) however a CNN is generally more powerful and accurate for classification problems. It also has fewer learnable parameters which reduces the risk of overfitting.

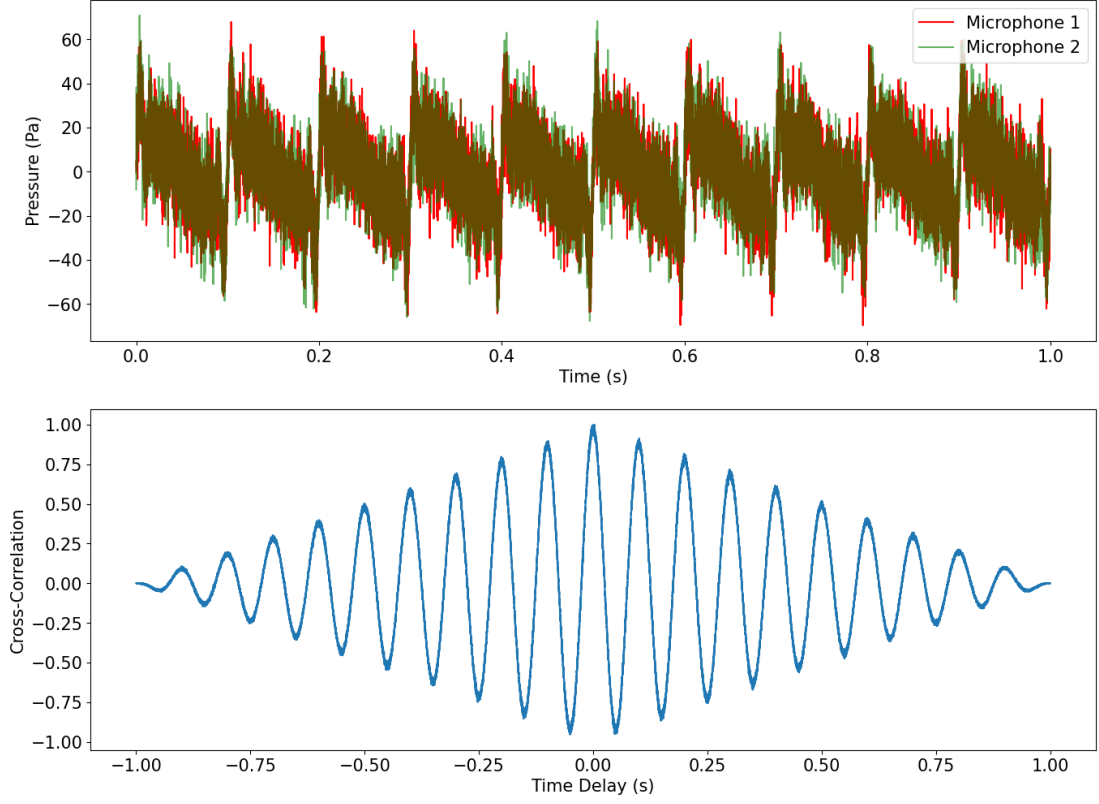


Figure 3.8: GCC pattern for a pair of microphone signals polluted with white noise. The GCC feature map is symmetric about the origin. Therefore, the correlation coefficients can be extracted from either side of the origin.

3.2.3 Case 3: Three-Dimensional ASL

This case is an extension of case 2. Similar to case 2, we used the four-microphone array and a classification approach. Both sources S1 (**Case 3(i)**) and S2 (**Case 3(ii)**) were used in separately.

Along with the azimuth angle (θ), the elevation angle (α) also plays a role in localization in this case. The ranges of θ and α were discretized into N_1 and N_2 classes respectively. To generate the training dataset, we placed the source (S1 and S2 separately) at real, random values of θ and α and fixed $r(= 10 \text{ m})$. Just like case 2, the CSM was calculated for S1,

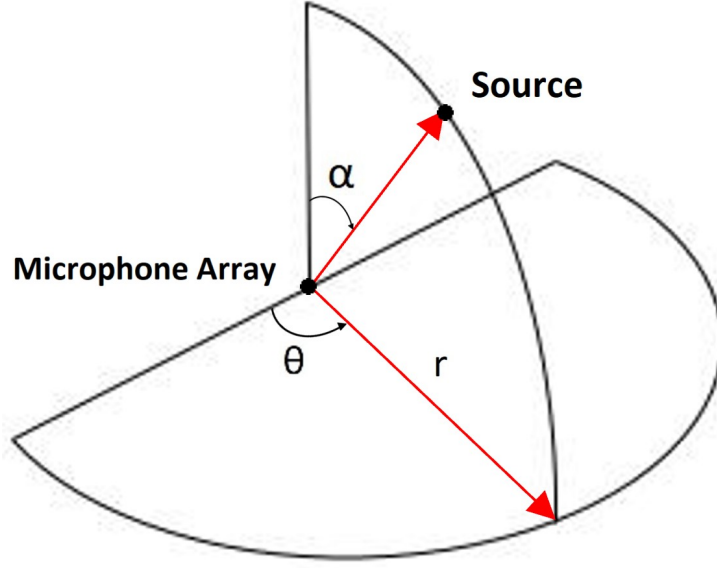


Figure 3.9: Case 3: Three-dimensional ASL. The 4-microphone array is shown as a point. The localization domain is: $0^\circ \leq \theta \leq 180^\circ$, $0^\circ \leq \alpha \leq 90^\circ$. The location of the source in cartesian coordinates is $(r \sin(\alpha)\cos(\theta), r \sin(\alpha)\sin(\theta), r \cos(\alpha))$.

and the GCC matrix was calculated for S2 as an input feature.

However, instead of using a conventional CNN, we used a multi-task CNN in which the separate tasks of determining θ and α were performed simultaneously by a single model. This is made possible because both tasks rely on the same input feature (CSM for S1 and GCC matrix for S2). Separate networks for θ and α are an option (like case 2) however, it is not very efficient as individual datasets have to be generated and each network has to be trained separately. A multi-task network improves data efficiency and reduces the training time and chances of overfitting significantly as the model has to generalize over multiple tasks. Figure 3.10 shows the block diagram of the multi-task CNN used. This is an example of *hard-parameter sharing* wherein the tasks share hidden layers while having separate, task-specific output layers (Amyar et al., 2020). For both output branches (θ and α), the activation function in the last layer was softmax, and ReLU activation was used everywhere else. The optimizer and loss function was *ADAM* and crossentropy respectively for both the branches.

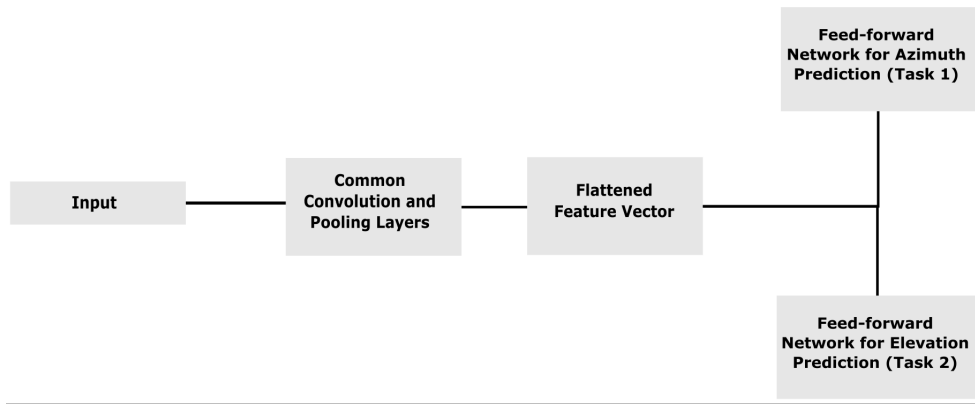


Figure 3.10: A multi-task CNN used for three-dimensional ASL. The main branch performed the convolution and pooling operations on the input feature. The two output branches peeled off the main branch, using the flattened feature vector as the input to give separate predictions for θ and α .

The CNN models for all the cases in this work were built using the deep learning frameworks *Keras* and *Tensorflow*. The details of these models are shown in Table 3.1. The training was done on Google Colaboratory using the Tesla T4 GPU accelerator.

3.3 Results

Table 3.1 gives the details of the CNN model used in all the three cases.

3.3.1 Case 1

For case 1, the model was trained to detect six monopole sources that were spread across a 12×12 scanning grid plane by randomly assigning coordinates. The results are presented and compared for source frequencies of 100 Hz and 300 Hz. Xu et al. (2021), who also presented results for six-monopole source localization using a DenseNet, generated a batch of 512 training samples for each epoch and ran their model for almost 3000 epochs. This means the training was carried out with more than 1.5 million samples. In contrast, our

Cases	Conv. Layers, Dimension	Max-Pool Layers, Dimension	Dense Layers (Nodes), Output Dimension
1	3, (3×3) , Filters: 32,64,64	2, (2×2)	4 (128), Output: 144
2(i): S1	2, (2×2) , Filters: 128, 64	None	1 (128), Output: N
2(ii): S2	3, (2×2) , Filters: 128,64,64	2, (2×2)	1 (128), Output: N
3(i): S1	2, (2×2) , Filters: 128,64	None	$(\theta) : 1(128), N_1$ $(\alpha) : 1(128), N_2$
3(ii):S2	2, (2×2) , Filters: 128,64,64	2, (2×2)	$(\theta) : 1(128), N_1$ $(\alpha) : 1(128), N_2$

Table 3.1: Summary of the convolutional neural network models used in all the cases

shallow CNN model was trained for 600 epochs on 50,000 training samples. The training data is 0.00045% of the total number of possible samples ($\binom{144}{6}$) which is in the order of $\approx 10^{10}$, thus exemplifying the limited nature of the data. The number of learnable parameters was significantly smaller for the CNN (≈ 1 million) compared to [Xu et al. \(2021\)](#)'s DenseNet (≈ 20 million).

The learning rate for the optimizer was fine-tuned to 0.0001. This case took approximately 100 minutes to train. The obtained results were post-processed to highlight the strongest sources (predicted source strengths of 0.3 and higher) and allow us to derive qualitative and quantitative trends. Figures 3.11-3.12 show two source configurations as predicted by the model for monopole sources at 300 Hz. Note that these are representative results at 300 Hz. Figures 3.13-3.14 show similar results for monopole sources at 100 Hz. The red dots represent the ground truth source configuration. These results are followed by Table 3.2 which gives the source localization statistics at the two source frequencies for 1000 test samples.

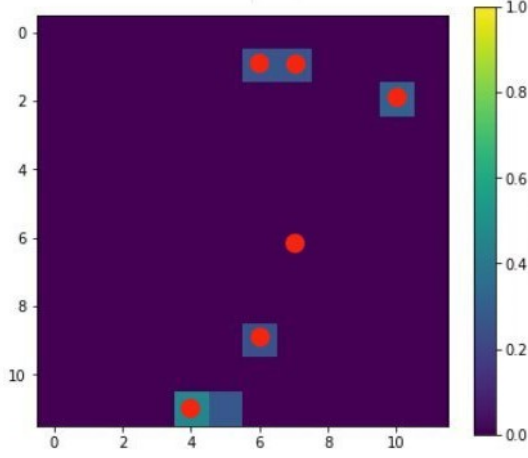


Figure 3.11: Configuration 1 (300 Hz)

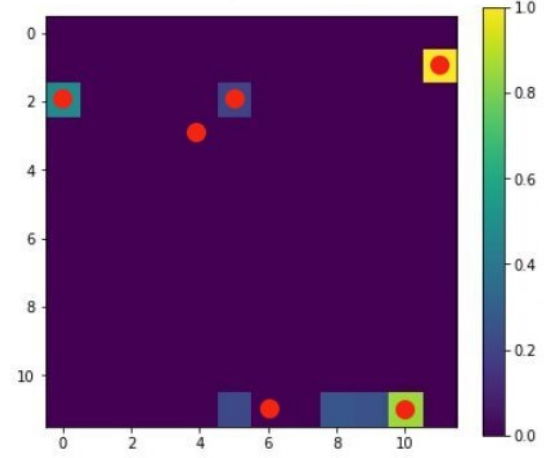


Figure 3.12: Configuration 2 (300 Hz)

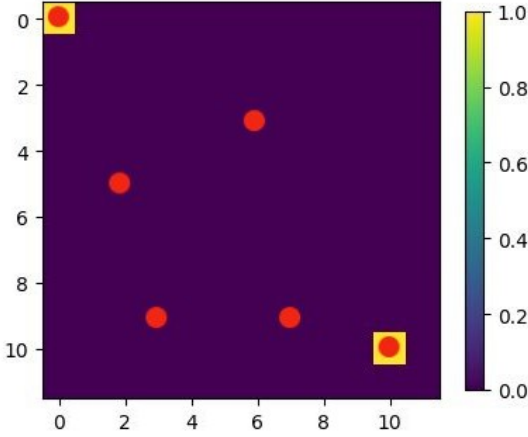


Figure 3.13: Sparse sources (100 Hz)

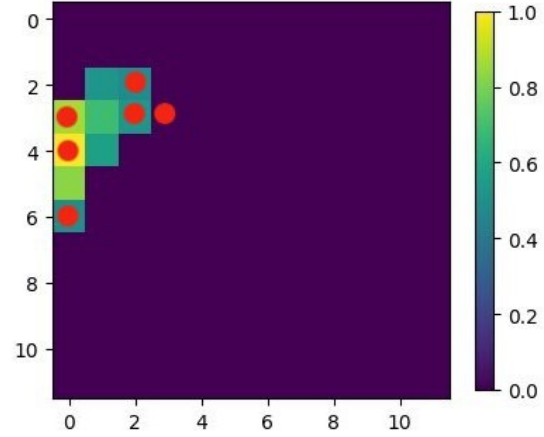


Figure 3.14: Clustered sources (100 Hz)

Discussion

We limited our training data and number of epochs to 50000 samples and 600 epochs respectively and sought to maximize the performance of the model under these constraints. We found that at a relatively higher source frequency of 300 Hz (Figures 3.11-3.12), the model was able to capture the source distribution reasonably well with a limited dataset.

Sources Located	300 Hz	100 Hz
6	9	1
5	70	10
4	272	39
3	397	142
2	210	344
1	42	348
0	0	116
Total	1000	1000

Table 3.2: Source localization statistics for 1000 test samples. A source is assumed to be located if the predicted source configuration detects a strength of 0.3 Pa or higher at the source location.

The model showed good resolution, less spreading, and was able to detect sparsely located sources. The limit of the model’s performance under the computational constraints was identified as 100 Hz since the results at 100 Hz source frequency (Figures 3.13-3.14) with the same amount of training were relatively poor. The model struggled to detect the sources when they were sparsely located (Figure 3.13) and tended to perform better when they were located in a cluster. The increase in spreading is evident from Figure 3.14 but it is still much less than beamforming wherein the poor resolution renders the results unusable at low frequencies. To resolve low-frequency sources using beamforming, the size of the scanning plane and the distance between the sources would have to be much larger than the dimensions of this case. The CNN was able to resolve low-frequency sources on the current scanning plane, thus highlighting its superior resolution and localization capabilities.

Table 3.2 indicates the frequency-dependent performance of the model. For a source frequency of 300 Hz, the model was able to locate a majority of the sources (4 or more) in 35.1% samples (351 out of 1000 samples), and at least half of the sources (3 or more) in 74.8% samples (748 out of 1000 samples). At 100 Hz source frequency, the percentages drop to 5% for locating a majority of the sources (50 out of 1000 samples) and 19.2% for locating at least half of the sources (192 out of 1000 samples). To obtain better results at

Classes (N)	Class Size	Accuracy (θ)
20	9°	$\approx 91\%$
30	6°	$\approx 89\%$
45	4°	$\approx 81\%$
60	3°	$\approx 69\%$

Table 3.3: Prediction accuracy for azimuth angle (θ) in two-dimensional ASL with a monopole source (S1) at 100 Hz. The accuracy values are for a test set of 10000 samples.

100 Hz and lower source frequencies, we have to go beyond the set network architecture and training limits and increase the number of layers/training data samples/epochs.

3.3.2 Case 2

In cases 2 and 3, which are classification problems, the performance was measured in terms of the prediction accuracy of the correct class (and thus the DoA) given by the model. In case 2, the model predicted the class of θ . The training was done using 50,000 training samples and 200 epochs with a fine-tuned learning rate of 0.0001. Since the source can be placed at any real value of θ between 0° and 180° , the training set of 50,000 samples is very limited as the total number of possible cases is infinite. The results for the two types of sources used in case 2 are given below.

Case 2(i): Monopole (S1)

We used a monopole source at 100 Hz. The frequency was once again identified as the limit of the model’s performance under the set training constraints. The source was at a distance of 10 m ($r = 10$ m) away from the origin. Since the dimensions of the cross-spectral matrix (CSM) were small (4×4 for the four-microphone array), only 2 convolutional layers and no pooling layers, were used in the CNN (Table 3.1). The variation in prediction accuracy of θ with the number of classes N is shown in Table 3.3.

Classes (N)	Class Size	Accuracy (θ)
20	9°	$\approx 84\%$
30	6°	$\approx 76\%$
45	4°	$\approx 63\%$
60	3°	$\approx 52\%$
90	2°	$\approx 40\%$
180	1°	$\approx 20\%$

Table 3.4: Prediction accuracy for azimuth angle (θ) in two-dimensional ASL with a sinusoidal plane wave source (S2) at 10 Hz. The accuracy values are for a test set of 10000 samples.

Case 2(i): Sinusoidal Plane Wave (S2)

For S2, we found that the performance of the model did not drop as much as the monopole source at lower frequencies which allowed us to test for source frequencies well into the infrasound range. We generated a sinusoidal plane wave signal at 10 Hz frequency polluted with white noise. The distance of the source from the origin was 10 m ($r = 10$ m). In this case, the amplitude does not undergo any decay and remains constant. The vector formed by concatenating the extracted correlation coefficients from the GCC maps of the microphone pairs (558×1), was reshaped into a matrix (31×18) and fed to the CNN. The larger size of the image allowed us to apply more convolution and pooling layers. Table 3.4 shows the prediction accuracy of the model and its dependence on the number of classes.

Discussion

For the monopole source, the prediction accuracy of the model was very high when the number of classes was less ($N = 20$) and decreased gradually as the number of classes increased. Beyond 60 classes, the model could not give predictions with reasonable accuracy in 200 epochs. The performance of the model for a higher number of classes can be improved with more training. The small size of the input (CSM) significantly reduced the network size hence the model trained very quickly. At the same time, it reduced the scope of

learning as we were very limited in the number of convolution and pooling layers we could apply. This is a clear limitation of the CSM as an input feature with few microphones.

In the case of the sinusoidal plane wave source (Table 3.4), selectively extracting the coefficients containing the essential information from the GCC maps allowed us to reduce the size of the image that has to be processed and save on computational time and resources. The accuracy values for all class sizes were obtained with the same amount of training and thus can be improved for smaller class sizes with more training. An accuracy value of greater than 60% across 45 classes for an infrasound source in the presence of noise shows the robustness and reliability of GCC as an input feature. It also highlights the advantage of using a time-domain input feature (GCC) as opposed to a frequency-domain feature (CSM) for locating low source frequencies with few microphones.

3.3.3 Case 3

As this case is more challenging than case 2, the number of training samples was increased to 80,000 keeping the number of epochs at 200 and the learning rate at 0.0001. The resolution or class size was kept the same for both θ and α , resulting in different number of classes N_1 and N_2 respectively. The performance of the model was measured in terms of the prediction accuracy of the correct class for θ and α .

Case 3(i): Monopole (S1)

Similar to case 2(i), we used a monopole source oscillating at 100 Hz. The CSM (dimensions 4×4) was used as an input to the multi-task CNN. The prediction accuracy of the model is shown in Table 3.5.

Case 3(ii): Sinusoidal Plane Wave (S2)

A sinusoidal plane wave source at 10 Hz was simulated and the GCC coefficient matrix of size (31×18) (same as case 2(ii)) was used as an input to the multi-task CNN. The prediction accuracy of the model is shown in Table 3.6.

Class Size	$N_1 (\theta)$	Accuracy (θ)	$N_2 (\alpha)$	Accuracy (α)
10°	18	$\approx 93\%$	9	$\approx 91\%$
5°	36	$\approx 84\%$	18	$\approx 79\%$
3°	60	$\approx 69\%$	30	$\approx 58\%$
2°	90	$\approx 56\%$	45	$\approx 52\%$
1°	180	$\approx 29\%$	90	$\approx 36\%$

Table 3.5: Prediction accuracy for θ and α in three-dimensional ASL with a monopole source (S1) at 100 Hz. The accuracy values are for a test set of 10000 samples.

Class Size	$N_1 (\theta)$	Accuracy (θ)	$N_2 (\alpha)$	Accuracy (α)
10°	18	$\approx 79\%$	9	$\approx 99\%$
5°	36	$\approx 58\%$	18	$\approx 99\%$
3°	60	$\approx 41\%$	30	$\approx 97\%$
2°	90	$\approx 26\%$	45	$\approx 96\%$
1°	180	$\approx 13\%$	90	$\approx 91\%$

Table 3.6: Prediction accuracy for θ and α in three-dimensional ASL with a sinusoidal plane wave source at 10 Hz (S2). The accuracy values are for a test set of 10000 samples.

Discussion

The multi-task CNN utilizes the same input feature for performing two related tasks: classifying the azimuth angle (θ) and the elevation angle (α). A commonly encountered problem in multi-task learning is “negative transfer” wherein one or more tasks can dominate the training/learning process and hurt the performance of other tasks. The tasks whose performance is hampered tend to perform better in a single-task model compared to a multi-task model (Lakkapragada et al., 2023).

Both tasks are learned reasonably well in the case of a monopole source (Table 3.5) but a clear case of negative transfer is observed for the sinusoidal plane wave source (Table 3.6). The prediction of α dominates the training which is expected, as it is the easier of the two tasks ($N_2 < N_1$). However, the prediction accuracy values for θ are less than what was obtained when a single-task model was used for θ in case 2(ii) (Table 3.4). A suggested way to overcome negative transfer is to weigh the individual losses from the two tasks appropriately such that they are on the same scale (Lakkapragada et al., 2023).

Collectively, the results from the three cases show much promise and it is expected that with more data, training, and deeper networks, deep learning models capable of real-time source localization can be successfully built in the future. Chapter 5 elaborates on this possibility in the context of CAT detection.

Chapter 4

Auralization of a Representative CAT Source

With deep learning proven to be a viable method for low-frequency ASL in the previous chapter, this chapter demonstrates the working of a potential, deep learning-enabled acoustics-based method for CAT detection through a large-scale, simulated flight case. The focus of the chapter is on auralizing a representative CAT source and predicting what an in-flight microphone will detect. The detected signal is then post-processed to obtain spectrograms that reveal characteristics of the signal that could potentially be used to detect and locate CAT. The spectrograms can be further processed using deep learning and the scope of this is discussed at the end of this chapter and also in Chapter 5.

Section 4.1 gives the details of the simulated flight case and curved and straight ray tracing methods. Section 4.2 presents and discusses the obtained results.

4.1 Methodology

Auralization is the mathematical modeling of the acoustic field of a source. It involves simulating the propagation of sound from the source in the given medium and determining what a receiver will hear. A broadband, infrasound monopole source of 1 Pa strength

and emitting 100 equally distributed, discrete frequencies in the range [1 Hz - 20 Hz] is chosen as an abstraction of the realistic acoustic emissions from clear-air turbulence (CAT). Sound propagation in the atmosphere is modeled using ray tracing, a method that treats sound waves as rays. Two types of ray tracing methods - straight and curved ray tracing have been used as mentioned in Chapter 2, to simulate the eigenrays (the ray paths between the source and the receiver). Straight ray tracing represents free space propagation from a monopole that occurs at constant atmospheric conditions while curved ray tracing incorporates atmospheric refraction caused by wind and temperature gradients.

The auralization methodology for the simulated flight case is summarised in Figure 4.1.

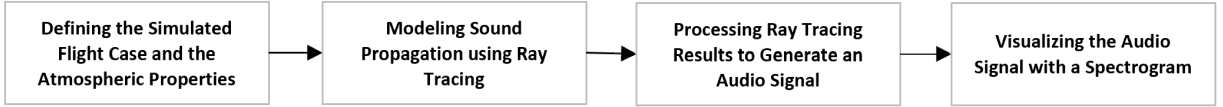


Figure 4.1: Auralization methodology

Each aspect of the methodology has been dealt with separately in the following sections.

4.1.1 Simulated Flight Case and Atmospheric Properties

Simulated Flight Case

The broadband infrasound monopole source is situated at (0, 0, 10 km). We defined a flight path 100 km in length in the x -direction ($-50 \text{ km} \leq x \leq 50 \text{ km}$), at the same altitude as the source ($z = 10 \text{ km}$) which is the typical cruising altitude for commercial flights. The airplane, in this case, is a moving receiver and is flying in a straight line, at a uniform velocity (v_{plane}) of 200 m/s (average cruising speed of the airplane, Mach number = 0.67), covering the flight path in 500 seconds. We placed 5,001 receiver points along the flight path at equal intervals of 20 m. The eigenrays are calculated for all source-receiver pairs, thus yielding a total of 5,001 eigenrays. The lateral distance (y -coordinate) between the source and the receiver was varied ($y = -1, -3, -5, -7.5, -10, -15 \text{ km}$) to obtain six sub-cases. The flight case setup for $y = -1 \text{ km}$ is shown in Figure 4.2.

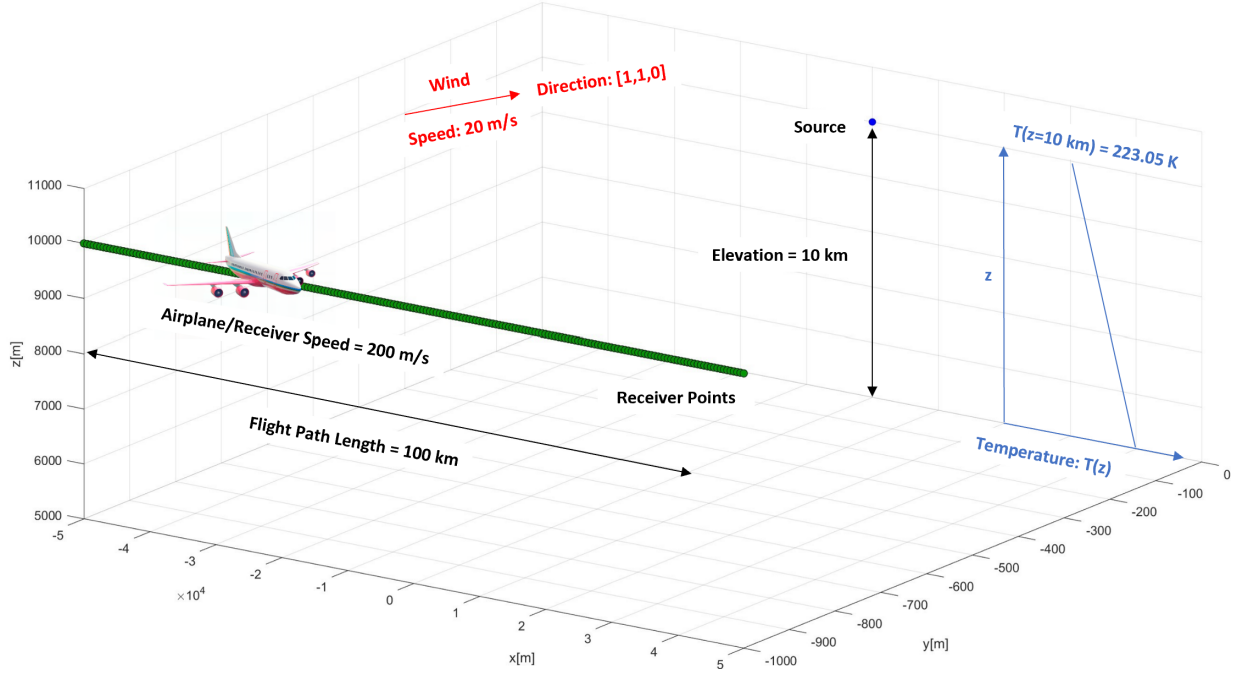


Figure 4.2: Setup, sub-case 1: $y = -1$ km

Atmospheric Properties

The atmosphere is defined using two properties: wind and temperature. Furthermore, it is assumed to be stratified, and thus all the properties remain constant in a particular strata and depend on the altitude z only.

An average, constant wind speed of 20 m/s (Limpinsel et al., 2018) in the direction $(1, 1, 0)$ was assumed throughout the domain for both straight and curved ray tracing methods. While the wind had to be kept constant for straight ray tracing by its very definition, not using a wind speed gradient in curved ray tracing is explained below.

For curved ray tracing, the International Standard Atmosphere (ISA) temperature profile shown in Figure 4.2 was used. This profile gives a linear variation of the temperature in the troposphere ($z < 13$ km) (Schäfer and Vorländer, 2021):

$$T(z) = T_0 - 0.0065 z \quad (4.1)$$

Here, $T_0 = 288.15$ K is the temperature at the mean sea level. The temperature gradient in this case is responsible for atmospheric refraction and curving the rays (Chapter 2). We did not use a wind gradient for curved ray tracing to better understand atmospheric refraction by avoiding the complicated competing effects of wind and temperature gradients. Keeping the wind constant allows us to analyze the effect of advection and refraction separately.

For straight ray tracing, the temperature was constant and equal to 223.05 K at 10 km altitude. As a result, the speed of propagation of straight rays is constant as well. The speed of sound without advection, depends on the temperature as:

$$c(z) = \sqrt{\gamma R T(z)} \quad (4.2)$$

where $\gamma = 1.4$ is the ratio of specific heats and $R = 287$ J/kg.K is the gas constant of air.

Thus, the speed of sound propagation for straight ray tracing at $T = 223.05$ K is equal to 299.47 m/s (speed of sound at 10 km altitude ($c_{z=10 \text{ km}}$)). With advection, the effective speed of sound (c_{eff}) changes, depending on whether the wind is aiding or opposing the propagation.

4.1.2 Curved Ray Tracing Method

The atmospheric ray tracing (ART) framework developed by [Schäfer and Vorländer \(2021\)](#) was used for curved ray tracing. It has a MATLAB interface that allows the users to adapt the framework to their purpose.

Ray Tracing Equations

The acoustic source in the ART framework is modeled as a sphere that emits rays at user-defined emission angles. The ray trajectory is obtained by solving the governing equations of ray tracing consisting of differential equations for two variables: the position of the ray given by a fixed point on a wavefront $\mathbf{r} = (x, y, z)$, and the wavefront normal $\mathbf{n}$ which gives the direction of the ray at that point. The wavefront normal is usually expressed in terms of the slowness vector $\mathbf{s}$ given as:

$$\mathbf{s} = \frac{\mathbf{n}}{c + \mathbf{n} \cdot \mathbf{v}} \quad (4.3)$$

where c is the local speed of sound and $\mathbf{v}$ is the wind vector. The system of equations is given in [Ostashev and Wilson \(1997\)](#) (eqs. (10.72) and (10.73)). In Cartesian coordinates, the equations can be written as:

$$\frac{d\mathbf{r}_i}{dt} = \frac{c^2}{\Omega} \mathbf{s}_i + \mathbf{v}_i \quad (4.4)$$

$$\frac{d\mathbf{s}_i}{dt} = -\frac{\Omega}{c} \frac{\partial c}{\partial x_i} - \mathbf{s} \cdot \frac{\partial \mathbf{v}}{\partial x_i} \quad (4.5)$$

Here, $\Omega = 1 - \mathbf{s} \cdot \mathbf{v}$. Note that $i = 1, 2, 3$ corresponds to the x, y , and z coordinates respectively. The system of equations is solved numerically using the fourth-order Runge-Kutta method (RK4) to give the temporal variation of the position ($\mathbf{r}$) and direction ($\mathbf{n}$) of the ray, therefore obtaining its trajectory.

The framework initially determines the ray closest to the receiver with the user-defined angular resolution and then launches more rays around the closest ray to iteratively get closer to the receiver. This is known as *adaptive ray zooming* ([Schäfer and Vorländer, 2021](#)). It allows the framework to be efficient by starting with a coarse angular resolution for the emission angle and then progressively increasing it. The receiver is modeled as a sphere of 1-meter radius and the ray intercepting any part of the sphere is assumed to be the eigenray.

Acoustic Propagation Parameters

Multiple acoustic parameters can be derived from the eigenrays to quantify the propagation, namely: (1) propagation time and time delay, (2) geometric spreading loss, and (3) analytical Doppler frequency shift. Figure 4.3 shows a curved eigenray at one of the extreme receiver points for the $y = -1$ km sub-case.

The ART framework gives the **propagation times** (pt) of all the eigenrays. The **time delay** (td) in the sound received by two receiver points is the difference in propagation times of the two corresponding eigenrays. We define the time delay at all receiver points with respect to the receiver at $x = -50,000$ m (the first receiver point at which sound is

perceived by the in-flight microphone). The time delay at receiver n $((td)_n)$ is given as:

$$(td)_n = (pt)_{x=-50,000} - (pt)_n \quad (4.6)$$

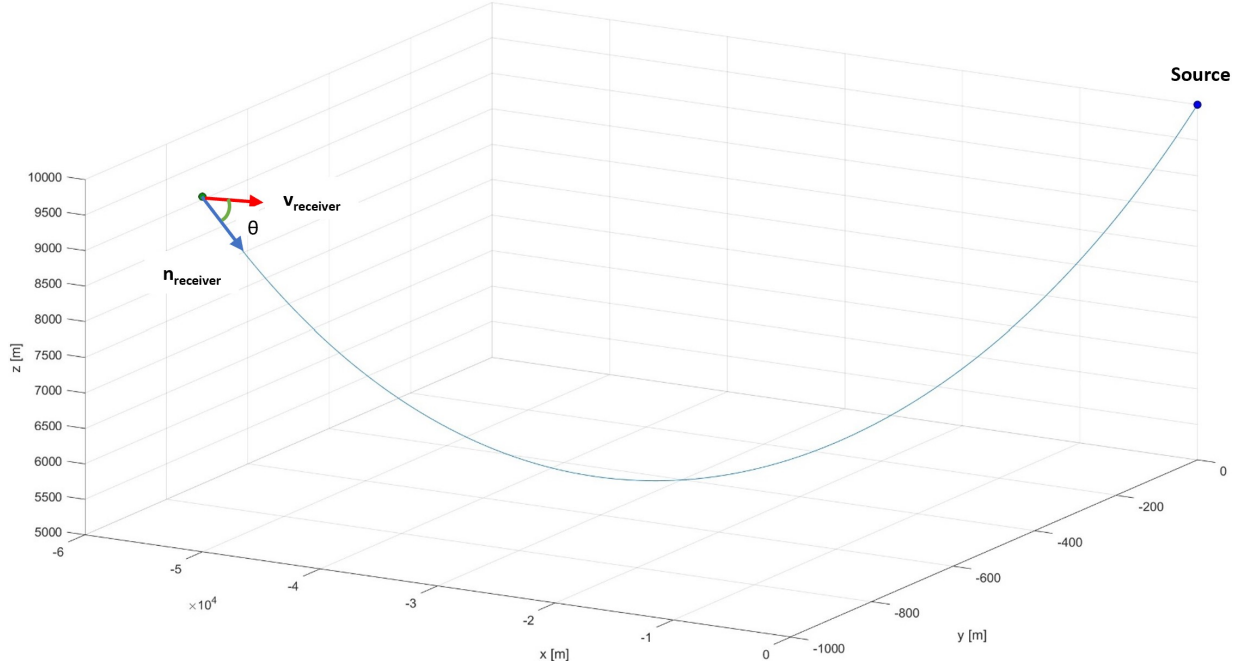


Figure 4.3: Curved ray at $x = -50,000$ m and $y = -1,000$ m for the simulated flight case

The **geometric spreading loss** is the reduction in the intensity of the sound as its energy spreads over a larger area when it travels farther from the source. Sound from a monopole in free space propagates uniformly in all directions and incurs a spherical spreading loss. In the case of a curved ray, the spherical waveform is distorted due to atmospheric refraction hence the distance-based, spreading loss calculations for a point monopole are not applicable. Thus, the ART framework calculates the loss directly along the ray. It uses the Blokhintzev variant ([Ostashev and Wilson, 1997](#)) in which the sound pressure (p) along the ray is inversely proportional to the square root of the cross-sectional area of the ray (a):

$$p^2 \propto \frac{c}{a(1 + \mathbf{n} \cdot \mathbf{v}/c)\sqrt{1 + 2\mathbf{n} \cdot \mathbf{v}/c + v^2/c^2}} \quad (4.7)$$

The framework uses this relation to calculate the spreading loss factor α_{spread} as:

$$\alpha_{spread} = \sqrt{\frac{p_{receiver}}{p_{ref}}} \quad (4.8)$$

where $p_{receiver}$ is the sound pressure at the receiver and p_{ref} is the reference pressure at the source (equal to the source strength (1 Pa)). The spreading loss factor α_{spread} is used to calculate the attenuated sound pressure amplitude (A) at all the receiver points in curved ray tracing as:

$$A = p_{ref} \times \alpha_{spread}^2 \quad (4.9)$$

The other important loss generally responsible for signal attenuation is atmospheric absorption. However, it is found to be more prevalent for higher frequencies. Attenuation of infrasound due to absorption is almost negligible in the lower atmosphere or troposphere (de Groot-Hedlin, 2008) relative to the spreading loss. To show this, the attenuation coefficient (expressed in decibels per meter (dB/m)) given in (for Standardization, 1993) was integrated along the eigenray shown in Figure 4.3 by the framework to obtain the total attenuation due to atmospheric absorption. The attenuation was found to be 3.387 dB for a 20 Hz frequency, which is the highest frequency emitted by our CAT source (and will thus incur the maximum absorption loss), making it negligible relative to the geometric spreading loss which is 190.6 dB. The attenuation due to absorption for other frequencies emitted by the source is even lesser. Figure 4.4 shows the variation of the atmospheric attenuation with frequency for the same eigenray. It can be seen that the attenuation is negligible for infrasound (frequency < 20 Hz) but increases rapidly thereafter.

The **frequency shift** in sound that the receiver will experience is calculated analytically by the Doppler effect. The Doppler frequency shift (D) observed at the receiver due to a source frequency f_s is given by the projection of the receiver velocity ($v_{receiver}$) along the ray at the receiver, divided by the effective speed of sound (c_{eff}) at the receiver (Kolano and LAB, 1978):

$$D = f_s \frac{v_{receiver} \cos(\theta)}{c_{eff}} \quad (4.10)$$

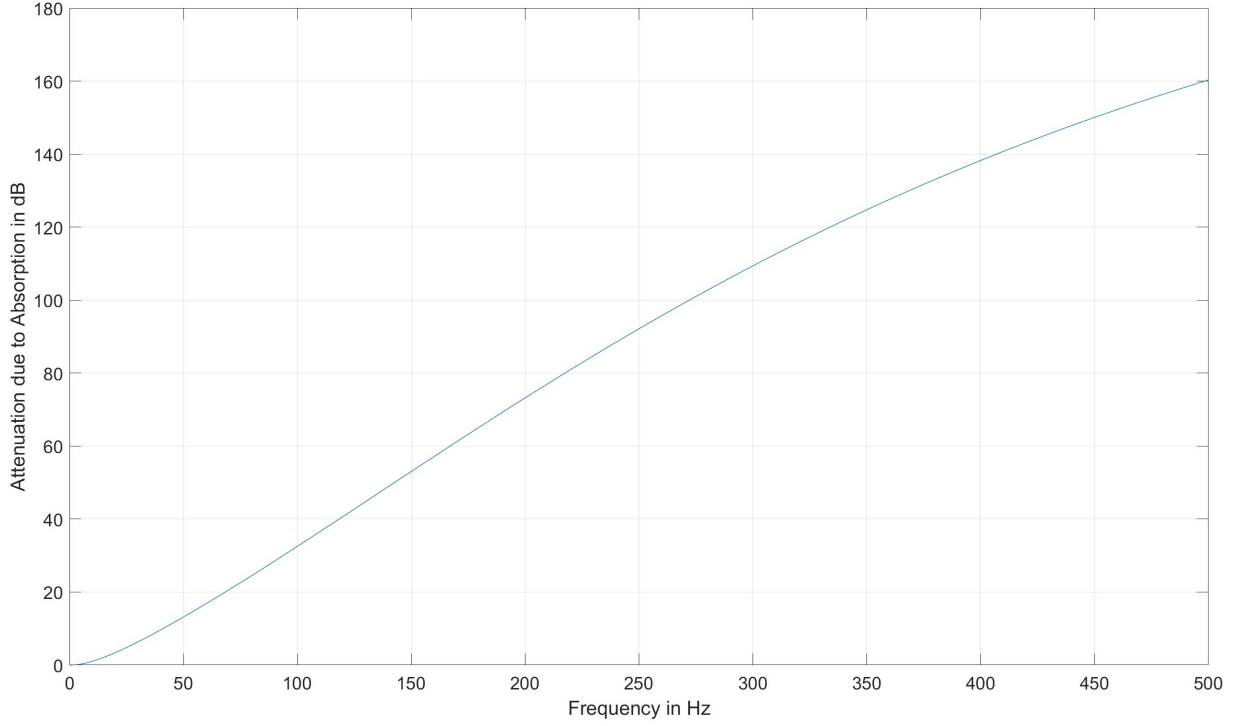


Figure 4.4: Total frequency-dependant attenuation due to atmospheric absorption shown for the curved ray in Figure 4.3.

θ (shown in Figure 4.3) is the angle between the receiver velocity vector and the direction vector of the ray at the receiver ($\mathbf{n}_{receiver}$).

c_{eff} takes into account the effect of wind on sound propagation which is measured by the projection of the wind vector along the direction vector of the ray at the receiver. If the wind aids propagation, $c_{eff} > c_{z=10 \text{ km}}$. If not, $c_{eff} < c_{z=10 \text{ km}}$. Note that $c_{z=10 \text{ km}} = 299.47 \text{ m/s}$.

The perceived frequency (f_o) at the receiver therefore is:

$$f_o = f_s + D = f_s + f_s \frac{v_{receiver} \cos(\theta)}{c_{eff}} \quad (4.11)$$

When the receiver approaches the source, it experiences frequencies higher than the actual source frequency. When the receiver starts to move away from the source, i.e. when $\cos(\theta)$

becomes negative, it perceives frequencies lower than the source frequency. Note that the analytical Doppler shift is only for our understanding. In actual signals, the frequency shift occurs through a varying time delay. This has been shown when we generate an audio signal from the acoustic parameters in the *Auralization* section.

4.1.3 Straight Ray Tracing

Compared to curved ray tracing, straight ray tracing is relatively less complex. The absence of any wind or temperature gradient makes straight ray tracing akin to free space propagation. The eigenrays and acoustic propagation parameters can be derived from the well-established analytical relations for a point monopole. Figure 4.5 shows the example of a straight ray at $x = -50,000$ m and $y = -1,000$ m.

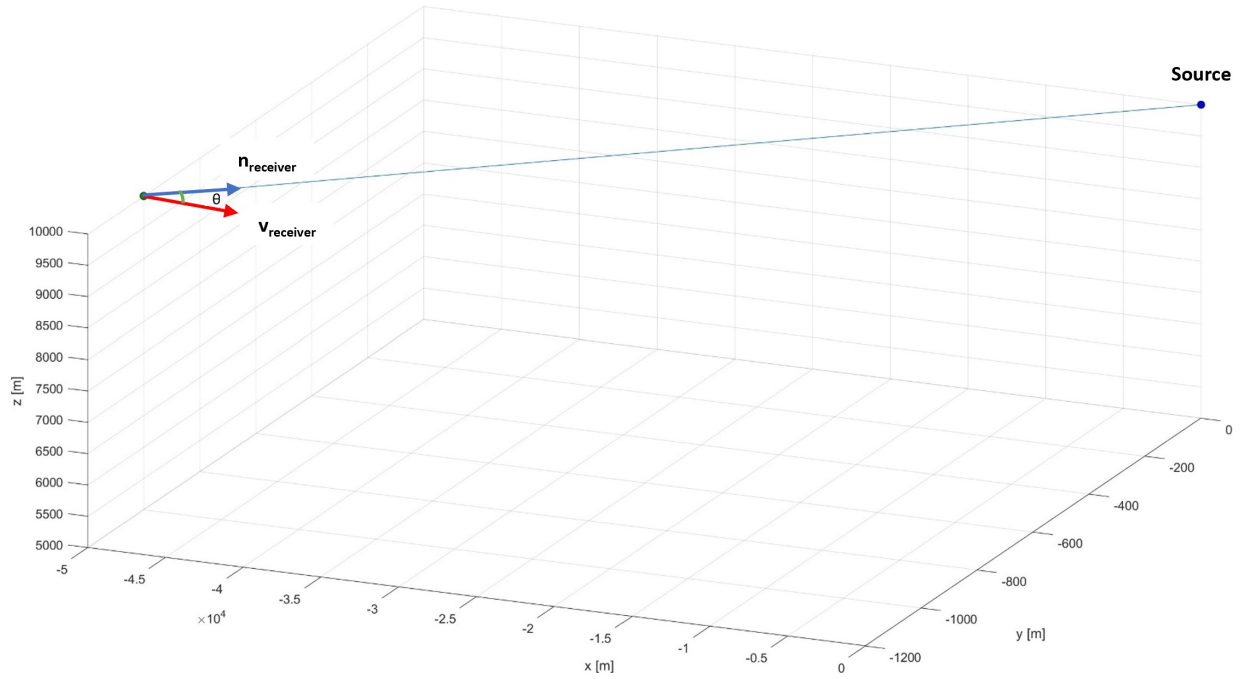


Figure 4.5: Straight ray at $x = -50000$ m and $y = -1000$ m for the simulated flight case

Each eigenray travels at a constant effective speed of sound at 10 km altitude. The **propagation time** of the rays is simply calculated by dividing the ray length by the

effective speed of sound of that ray. The Doppler **frequency shift** is given by the same formulae described in the curved ray tracing section above (eq. (4.10), (4.11)).

The **spreading loss** calculation in straight ray tracing to obtain the sound pressure amplitude at the receiver is done differently from curved ray tracing. First, the sound pressure level (SPL) of the monopole is calculated in decibels (dB) using the general formula for SPL:

$$SPL = 20 \log \left(\frac{p}{20 \times 10^{-6}} \right) \quad (4.12)$$

Here, p is the sound pressure ($p = p_{ref}$ for the monopole). The spherical spreading loss in terms of the attenuation (ATT) in dB is calculated as (for [Standardization, 1996](#)):

$$ATT = 20 \log (d) + 11 \text{ dB} \quad (4.13)$$

where d is the distance between the source and the receiver in meters. The sound pressure level at the receiver ($SPL_{receiver}$) is obtained by subtracting the attenuation (ATT) from the sound pressure level of the monopole (SPL_{source}):

$$SPL_{receiver} = SPL_{source} - ATT \quad (4.14)$$

Using the $SPL_{receiver}$, the sound pressure amplitude (A) at the receiver, which is nothing but the sound pressure term p in equation (4.12), is obtained by inverting the equation. The attenuation due to absorption is once again negligible (2.275 dB for the ray shown in Figure 4.5) in comparison to the geometric spreading loss (105 dB).

4.1.4 Auralization

The final step in auralization is to simulate what an in-flight microphone will hear. This involves creating an audio signal using the parameters obtained (sound pressure amplitude (A) and time delay (td)) from processing the eigenrays (both straight and curved). The total signal heard by the in-flight microphone during the flight duration through linear acoustics is given by the superposition of all the individual signals corresponding to all the frequencies present in the broadband source. Hence, the following procedure is described for a single frequency emitted by the source and is repeated for the rest of the frequencies.

Two things need to be done before creating a signal: increase the number of sampling points to avoid aliasing, and define a frequency-dependent sound pressure amplitude.

Firstly, the current sampling rate, which is tied to the total number of receiver points and the duration of the flight, is approximately 10 samples per second ($\frac{5001 \text{ samples}}{500 \text{ s}}$). According to the Nyquist theorem, the sampling frequency must be at least double the highest frequency present in the signal (20 Hz, in this case) to avoid aliasing. To increase the number of samples (and thus the sampling rate), new samples are generated by linearly interpolating all the acoustic parameters between the current 5001 samples.

A frequency-dependent amplitude is needed as turbulence is made up of eddies of different sizes. The large eddies, which derive kinetic energy from the mean flow, are marked by large length scales and correspondingly lower frequencies. These large, energy-containing eddies are unstable and break down into smaller eddies, transferring their energy through a process known as the “energy cascade”. The smaller eddies are characterized by lower energy but higher frequencies. This difference in energy content between lower frequencies (large eddies) and higher frequencies (small eddies) needs to be accounted for in our broadband CAT source.

To make an educated approximation for the frequency-dependent amplitude, we turn to the turbulence spectrum. The spectrum of pressure fluctuations ($P(f)$) in turbulent flows follows the well-known $(-7/3)$ power law in the inertial range (George et al., 1984; Batchelor, 1951) wherein the energy is transferred from large eddies to smaller eddies. In terms of frequency, $P(f) \propto f^{-7/3}$. The power law characterizes the turbulence-turbulence interaction pressure that is responsible for aeroacoustic noise. Raspet and Webster (2015) found the infrasonic wind spectrum to follow the $(-7/3)$ law in the inertial range. Posmentier (1974), who recorded infrasonic emissions from CAT in the 1-16 Hz range, also found that the power spectral density falls off as $f^{-7/3}$. While there are studies such as Shams et al. (2013) that observed the power spectral density to vary differently, for generating an energy spectrum representative of CAT in this thesis, we assume that the power spectrum of CAT follows the $(-7/3)$ law.

Taking the square root of the power spectrum gives the amplitude spectrum ($amp(f)$)

which varies as:

$$amp(f) \propto f^{-7/6} \quad (4.15)$$

The sound pressure amplitude (A) we have defined thus far has been obtained purely through geometric loss calculation. To model the effect of turbulence, we define the amplitude (A_f) which depends on the source frequency:

$$A_f = A \cdot amp(f) \quad (4.16)$$

A signal can now be created.

For a frequency f_s emitted by the source, we define a sinusoidal acoustic pressure signal as a function of time ($p_{f_s}(t)$) using the frequency-dependent amplitude (A_{f_s}), and the time delay (td) values at all the sampling points (1,000,000) as:

$$p_{f_s}(t) = (A_{f_s}) \cos(2\pi f_s (t + td)) \quad (4.17)$$

Note that the signal experiences frequency shift due to the varying time delay (td).

Similarly, signals are created for all the remaining frequencies present in the source and are added together to obtain the total signal. The spectrogram of this signal is then plotted for analysis.

4.2 Results

The results consist of eigenrays and range of detection, propagation times, analytical frequency shift plots, and spectrograms for both ray tracing methods. All results have been discussed separately.

4.2.1 Eigenrays and Range of Detection

Eigenrays for straight and curved ray tracing for the first sub-case $y = -1$ km ($|y| = 1$ km) are shown in Figures 4.6 and 4.7 respectively. The pattern of eigenrays for the other sub-cases is similar.

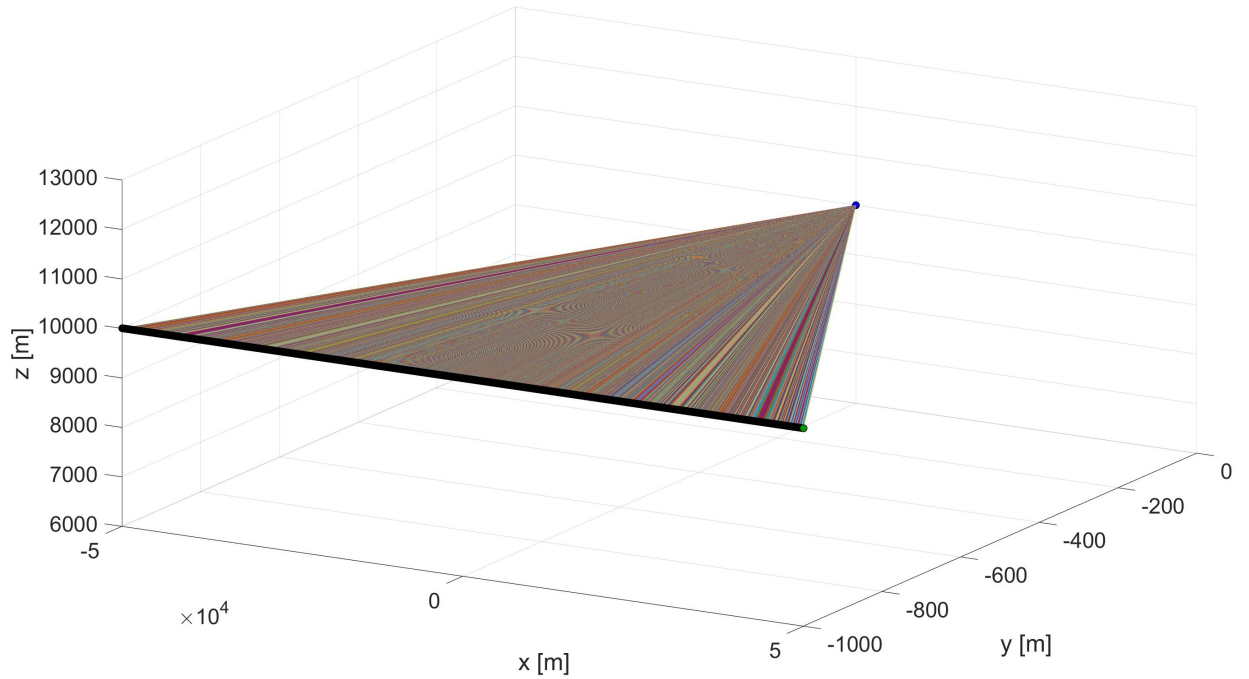


Figure 4.6: Eigenrays for straight ray tracing

For better visualization of the curved rays, their $x - z$ and $y - z$ plane projections are also shown in Figures 4.8 and 4.9.

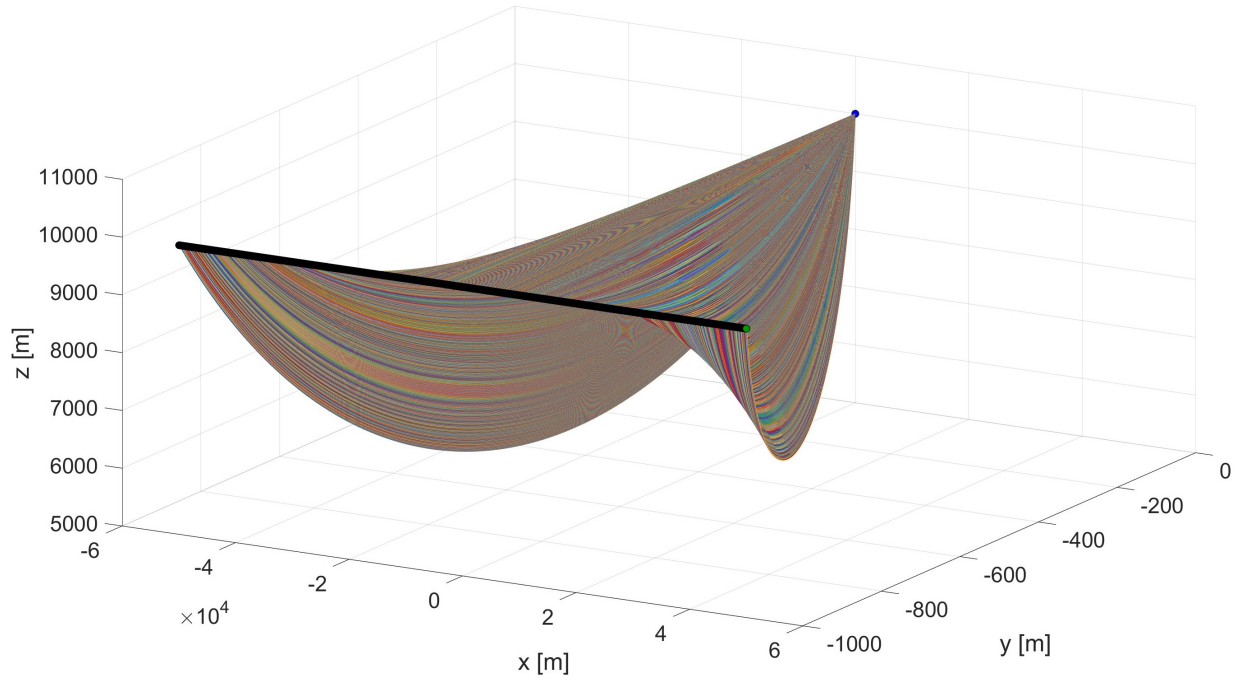


Figure 4.7: Eigenrays for curved ray tracing

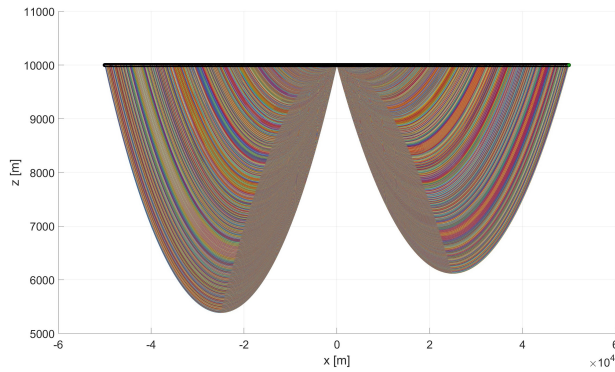


Figure 4.8: $x - z$ plane projection

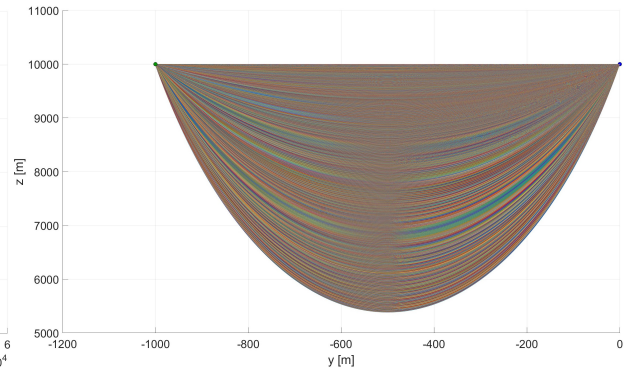


Figure 4.9: $y - z$ plane projection

Sound rays are curved in the atmosphere due to refraction caused by wind and/or temperature gradients. This can be seen in our flight case for curved ray tracing (Figures 4.7, 4.8, and 4.9) in which we have a temperature gradient. As the temperature decreases upward - which is usually the case during the daytime, the rays are refracted upwards. As

a result, an observer at a certain distance from the source may lie in the “shadow zone” of the source, unable to hear anything. The distance to the beginning of the shadow zone from the source is directly proportional to the altitude of the source-receiver pair and inversely proportional to the magnitude of the temperature gradient (Ingård, 1953). Shadow zone formation for upward refraction is illustrated through an example shown in Figure 4.10. The sound rays cannot penetrate the shadow zone region, and an observer standing in the region cannot hear the source.

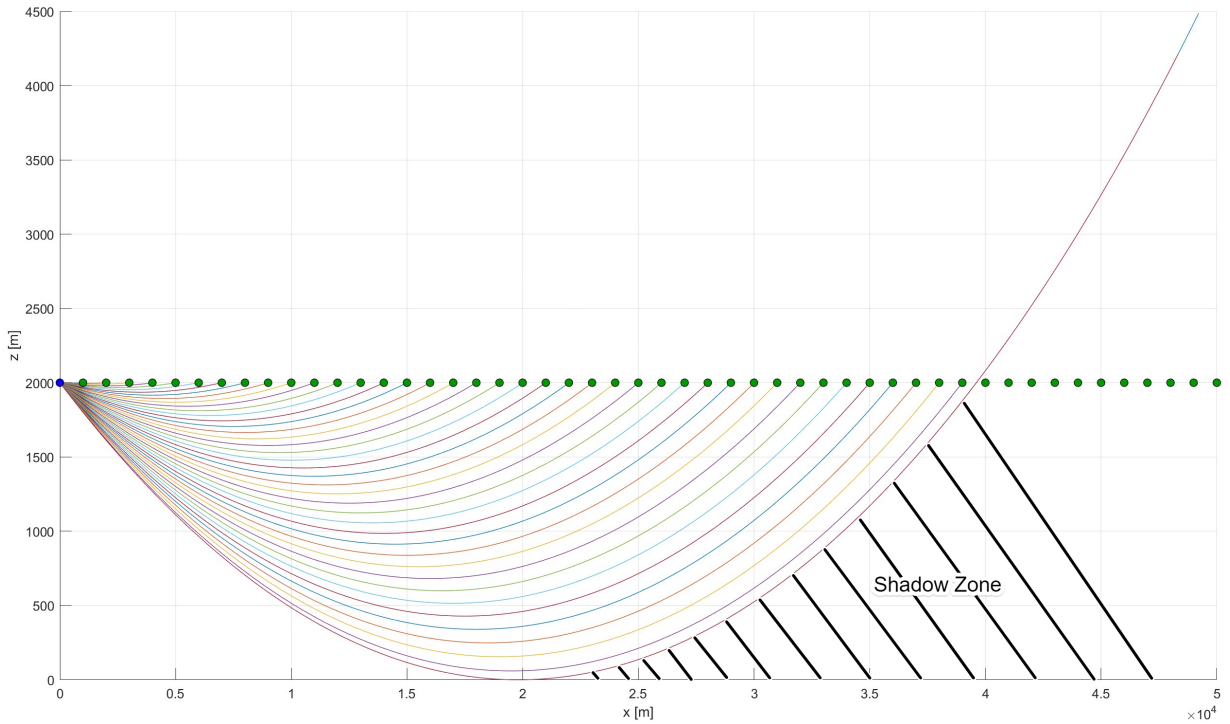


Figure 4.10: Example showing shadow zone formation for upward refraction. The shadow zone starts approximately 20 km from the source.

In the example (Figure 4.10), the receiver points (green dots) after $x = 39$ km cannot hear anything. Thus the range of detection is $0 \leq x \leq 39$ km.

In our simulated flight case, none of the receiver points lie in the shadow zone. However, shadow zone formation and range of detection can be identified by increasing the flight path. We doubled the flight path such that $-100 \text{ km} \leq x \leq 100 \text{ km}$, and calculated eigenrays for

21 receiver points along the flight path at equal intervals of 10 km to roughly estimate the range of detection. Figure 4.11 shows the eigenrays for the sub-case $y = -5$ km. Not all rays can reach the receiver points as shadow zones are formed on either side of the source. Shadow zones can be better seen in the $y - z$ plane projection (Figure 4.12). Figure 4.13, which shows the $x - y$ plane projection, allows us to estimate the range of detection as approximately $-70 \text{ km} \leq x \leq 80 \text{ km}$.

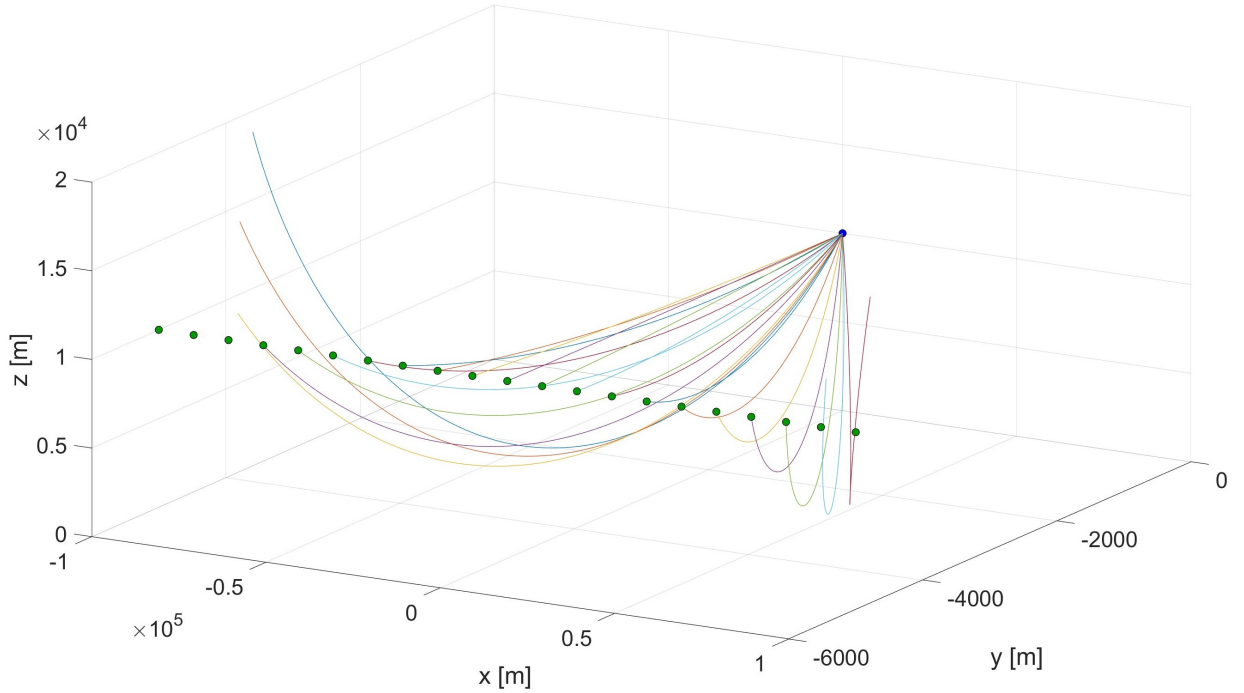


Figure 4.11: Eigenrays for the longer flight path: $-100 \text{ km} \leq x \leq 100 \text{ km}$.

The range of detection is observed to be the same for all the sub-cases. Sound propagation in the negative x -direction is always against the wind hence the range for $x < 0$ (70 km) is smaller than the range in the positive x -direction (80 km). Shadow zone formation is not valid for straight rays; therefore, a limitation of straight ray tracing is that it will theoretically give a signal even outside the range of detection. Note that the range of detection calculated in our case is theoretical and the actual range will be determined by the sensitivity of the in-flight microphone.

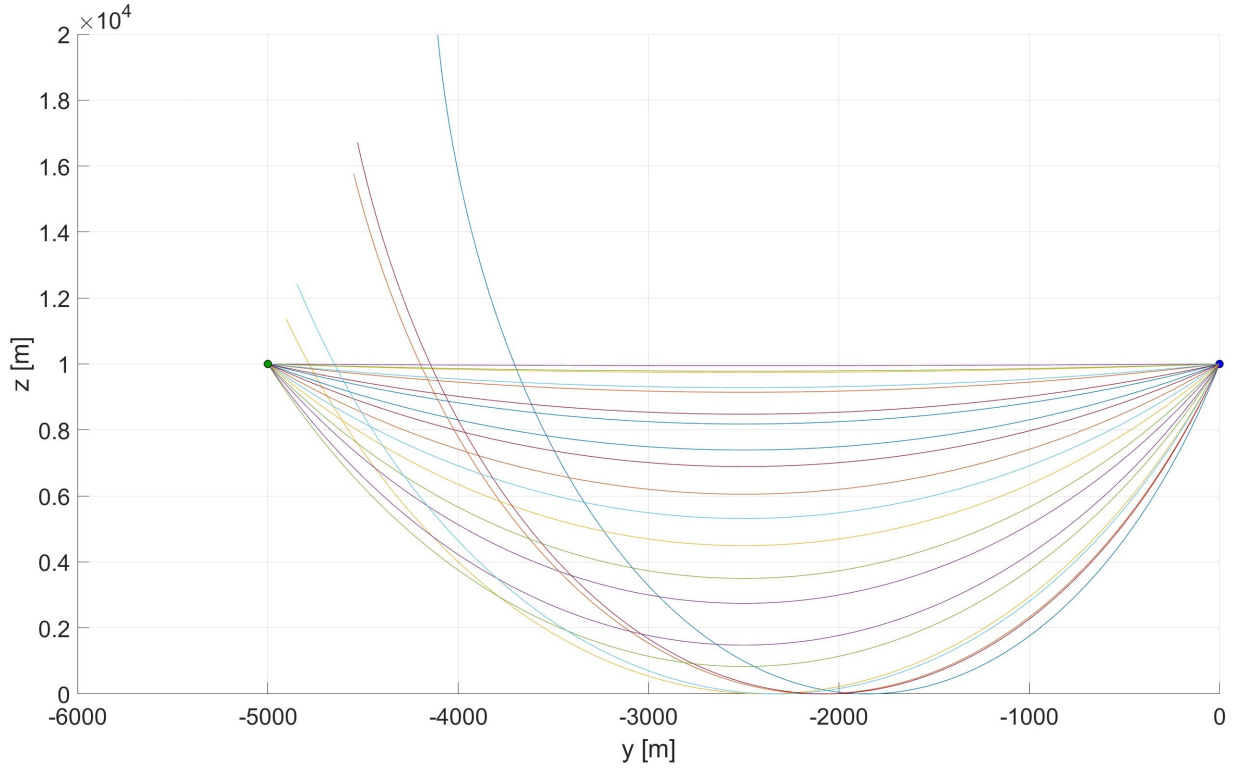


Figure 4.12: Shadow zone formation: $y - z$ plane projection

An inverse temperature gradient exists during the night as the temperature increases with altitude. As a result, rays are refracted downwards and no shadow zones are formed (Ingård, 1953). Thus, the chances of detecting sound at even farther distances are higher at night.

4.2.2 Propagation Time

Curved rays undergo constant refraction due to the temperature gradient and thus travel at different speeds (speed of sound changes with temperature) as opposed to straight rays that travel at a constant speed. The propagation time of a curved ray, when compared with a straight ray, gives valuable insight into the effect of refraction and the wind on sound propagation. Figures 4.14 and 4.15 show the propagation time and the difference in

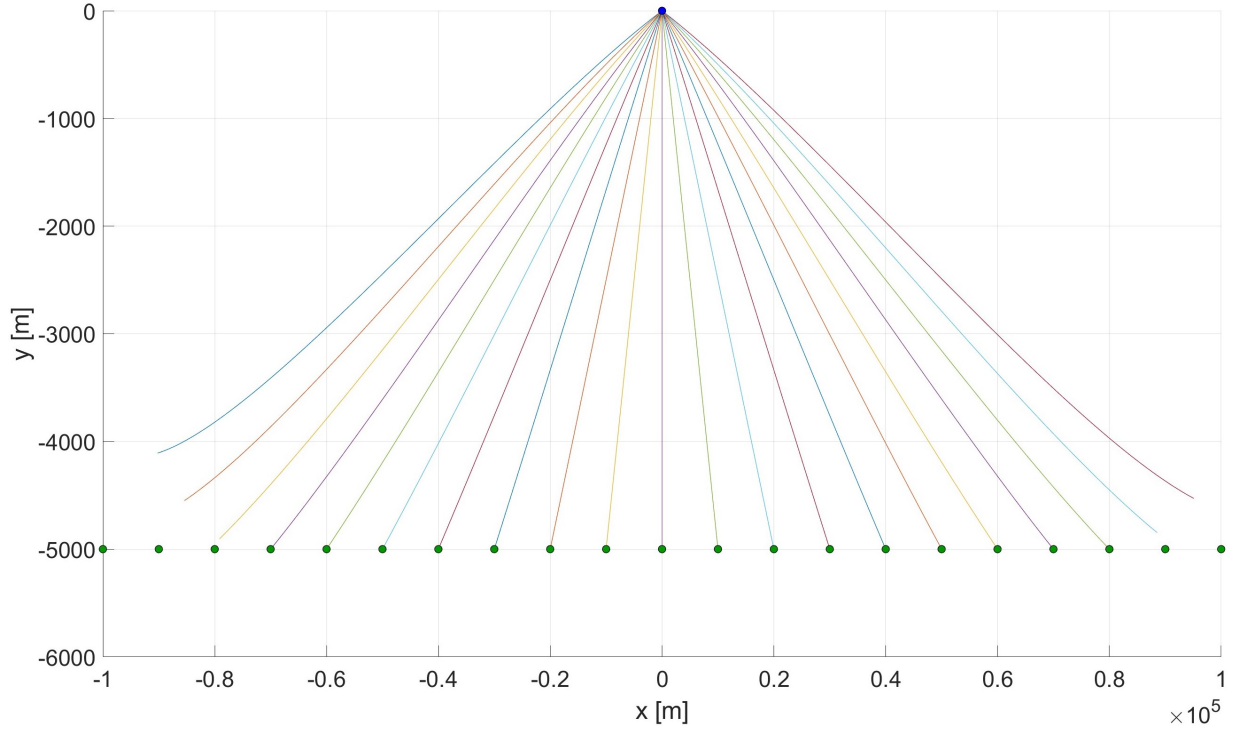


Figure 4.13: Range of detection: $x - y$ plane projection

propagation times between straight and curved rays respectively, for all the receiver points along the flight path at $|y| = 7.5$ km. The trend remains the same for all other values of $|y|$.

From Figure 4.14, we can see that the propagation time for both rays continuously decreases as the source-receiver distance decreases, and then increases when the distance starts to increase after the airplane has passed $x = 0$. The V-shaped plot for both rays is not symmetric about $x = 0$ which can be attributed to the wind. In the case of straight ray tracing, for $-50000 \text{ m} \leq x \leq 7480 \text{ m}$, the component of the wind vector is against the direction of ray propagation (opposes the propagation), thus reducing the effective speed of sound ($c_{eff} < c_{z=10 \text{ km}}$). At $x = 7,500 \text{ m}$, the wind vector is perpendicular to the ray and does not affect the sound speed ($c_{eff} = c_{z=10 \text{ km}}$). For $x > 7,500 \text{ m}$, the wind aids the propagation of rays by increasing the effective sound speed ($c_{eff} > c_{z=10 \text{ km}}$). This causes the rays to travel faster and propagation times to decrease. Similarly, for curved

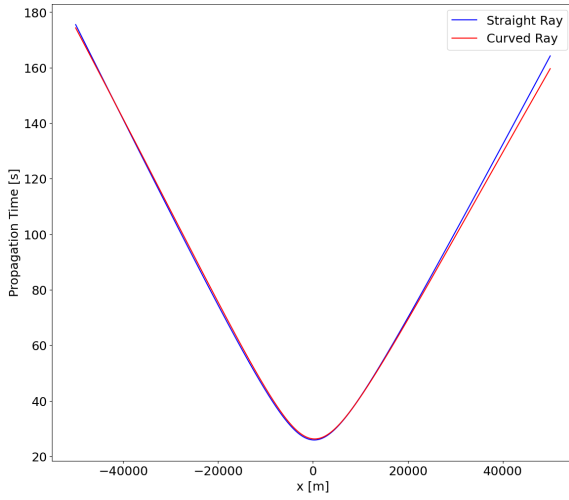


Figure 4.14: Propagation times for straight and curved rays

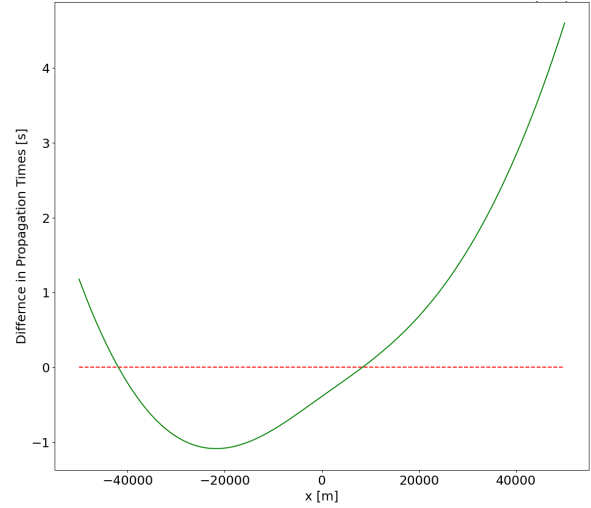


Figure 4.15: Difference in propagation times (straight ray - curved ray)

ray tracing, we calculated that the wind opposes sound propagation till $x = 8,560$ m after which it aids the propagation.

Figure 4.15 shows the difference in propagation times between straight and curved rays. For $x > 8,420$ m, the curved rays propagate faster than the straight rays despite being longer. Straight rays travel at a fixed altitude ($z = 10$ km) and therefore at a fixed sound speed ($c_{z=10 \text{ km}}$). Curved rays, on the other hand, travel through lower altitudes where the temperature is higher, and as a consequence, the speed of sound is greater than $c_{z=10 \text{ km}}$. The wind also aids propagation. As a net effect, the curved ray ends up traveling faster despite being longer for $x > 8,420$ m. For $-41,920 \text{ m} \leq x \leq 8,400$ m, straight rays are faster. For a majority of this interval, the wind opposes sound propagation. After a certain distance against the wind, the higher speed of propagation of the curved ray eventually overcomes the adverse wind to be effectively faster for $x < -41,920$ m.

The difference in propagation times, which can be as high as 4.6 seconds in this case, highlights an important limitation of straight ray tracing which assumes that sound travels at a fixed speed.

Analytical Frequency Shift

Figure 4.16 shows the frequency shift plots for unit source frequency (1 Hz) plotted using eq. 4.10 for all values of $|y|$. A receiver moving along the path will perceive the frequency given by equation 4.11.

The difference in frequency shift plots for straight and curved ray tracing depends on how close the straight ray approximation is to the curved ray. The accuracy of the approximation is inversely proportional to the distance the ray has to travel and therefore, we can see from Figure 4.16 that the difference in the straight and curved ray plots (or the error) naturally increases as $|y|$ increases. The error is largest at the extremes of the flight path for all values of $|y|$. This can be understood from the equation for the frequency shift (eq. 4.10).

$$D = f_s \frac{v_{receiver} \cos(\theta)}{c_{eff}}$$

Theoretically, at $x = \pm\infty$, θ for the straight ray will be ≈ 0 , and thus, straight rays at the extremes of the flight path will experience maximum Doppler shift. For curved rays to be able to travel to $\pm\infty$, the emission angle of the ray will have to be 90° . Since the source and receiver are at the same altitude, by Snell's law, the ray will intersect the receiver path perpendicularly, thus giving zero Doppler shift for the curved ray and therefore maximum error between the two types of rays at the extremes of the flight path.

In all the plots for both types of rays, the steepness of the frequency shift, which can be approximated by the magnitude of the slope at $x = 0$ m and $f = 0$ Hz, decreases as the lateral distance from the source ($|y|$) increases. Returning to the above equation, as $|y|$ increases, the range of θ (measured by the rays at the two extreme receiver points) decreases. This is explained in Figure 4.17 using straight ray tracing. The range of θ for $|y| = 3$ km (θ_2) is smaller than the range of θ for $|y| = 1$ km (θ_1). Therefore, a receiver with a higher value of $|y|$ experiences less frequency shift over the same duration or distance.

For any value of $|y|$, the magnitude of the slope for the straight ray plot is greater than the curved ray plot meaning that a straight ray undergoes more frequency shift in the same duration than a curved ray. This can again be explained by the difference in the range of θ of a straight ray and a curved ray.

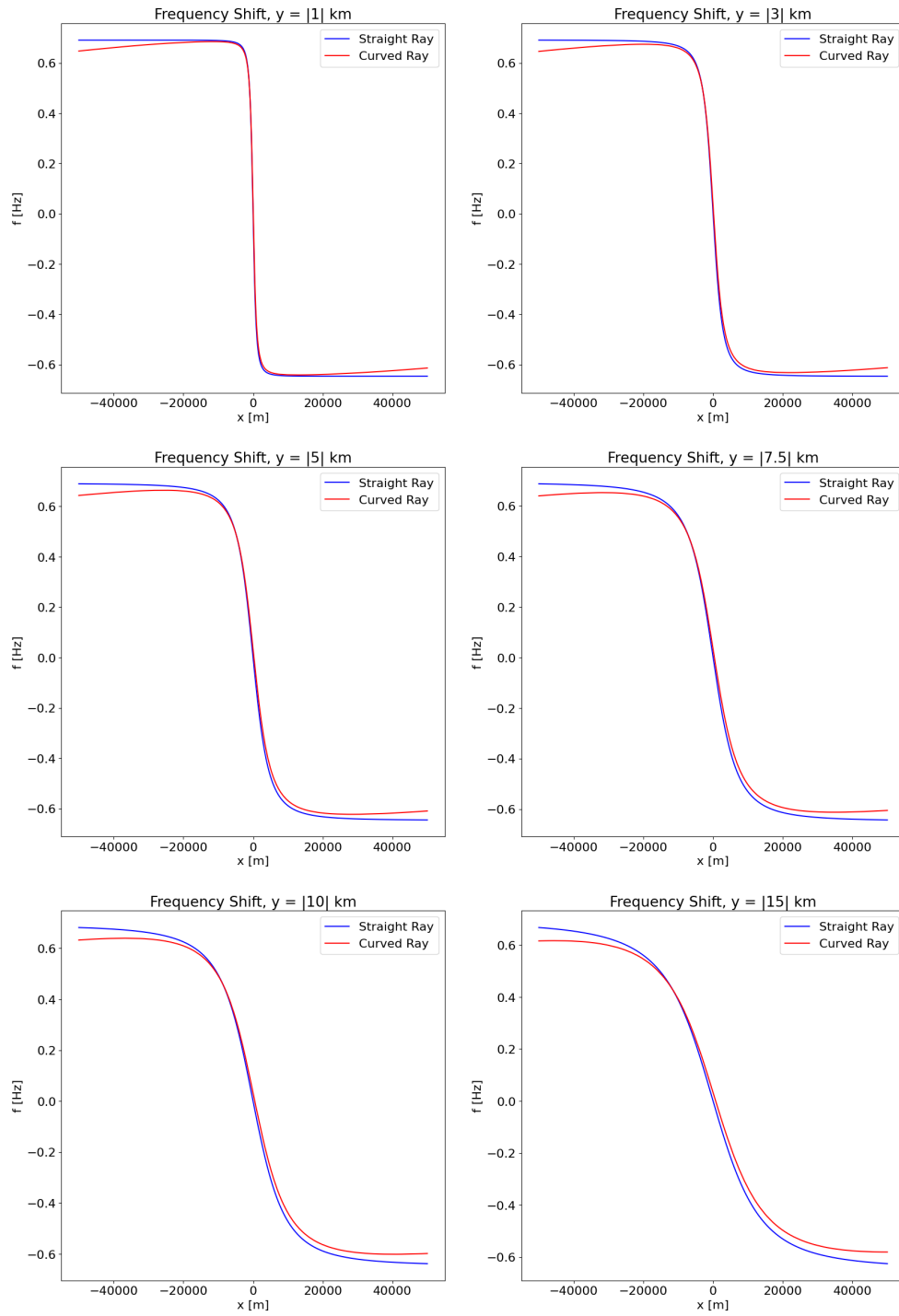


Figure 4.16: Frequency shift of a unit source frequency (1 Hz) as observed by the receiver.

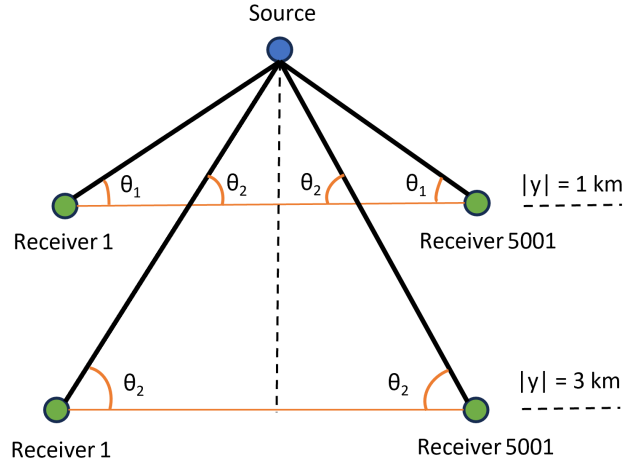


Figure 4.17: Range of θ vs $|y|$

4.2.3 Spectrograms

The spectrogram highlights the spectral and temporal information of the total signal received by the in-flight microphone from the broadband infrasound source, during the entire flight duration. The spectrograms for all values of $|y|$ are presented below (Figures 4.18-4.23). They give the power of the signal at each time and frequency, the power, in this case, being the square of the amplitude (Pa^2) that has been converted to decibels (dB). Power values less than -200 dB have been removed. More than the absolute power values, it is the relative change in the spectrograms that is of interest here.

The black dot in the spectrogram indicates the point of highest power (peak). The blue dotted line shows the time (peak time), and the red dotted line shows the frequency (peak frequency) at which the peak is detected. Tables 4.1 and 4.2 tabulate these values for straight and curved ray tracing respectively.

It is evident from the spectrograms and the tables that the power, or the strength of signal detected using straight ray tracing is significantly greater than the strength detected using curved ray tracing. This is expected as the length of curved rays is greater and therefore they incur more geometric losses.

The curved ray spectrograms show that the in-flight microphone detects higher frequen-

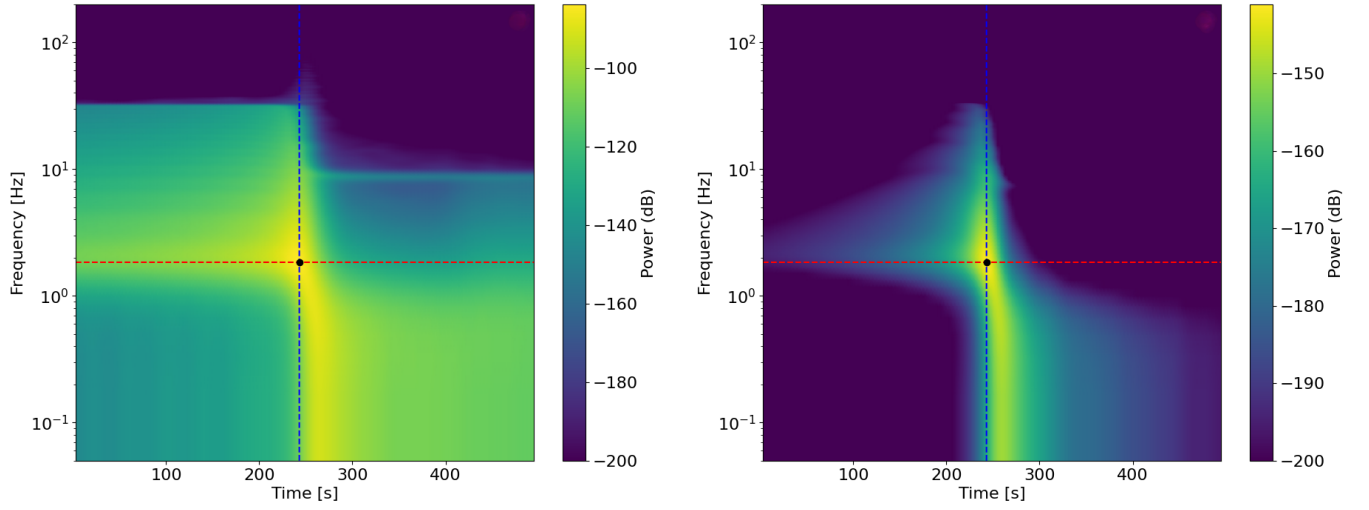


Figure 4.18: $|y| = 1$ km, straight ray tracing (left) and curved ray tracing (right) spectrograms

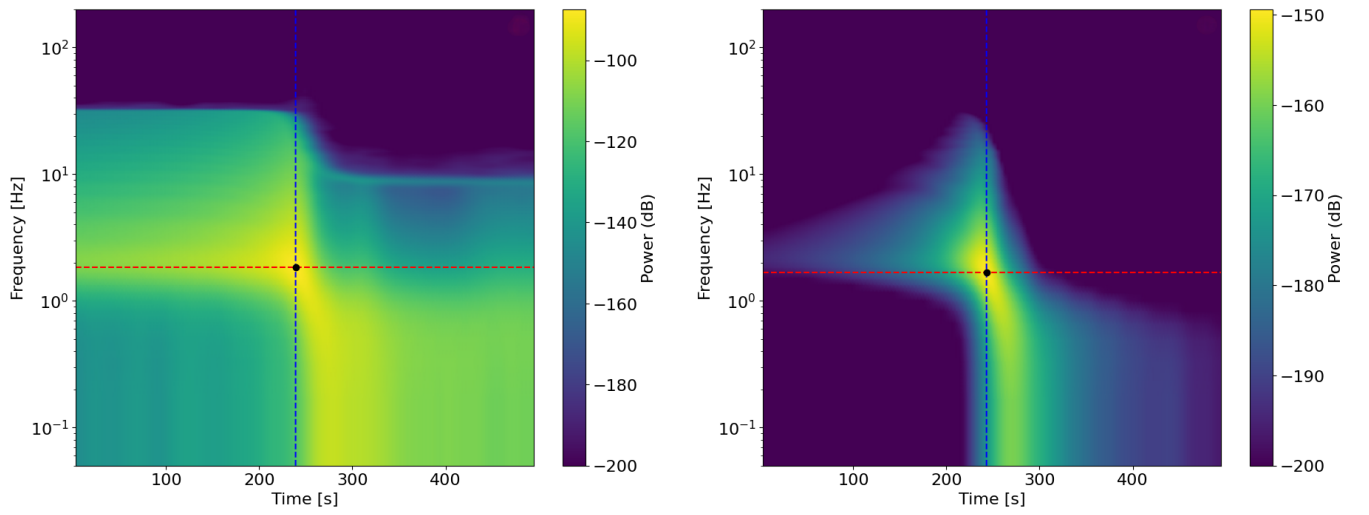


Figure 4.19: $|y| = 3$ km, straight ray tracing (left) and curved ray tracing (right) spectrograms

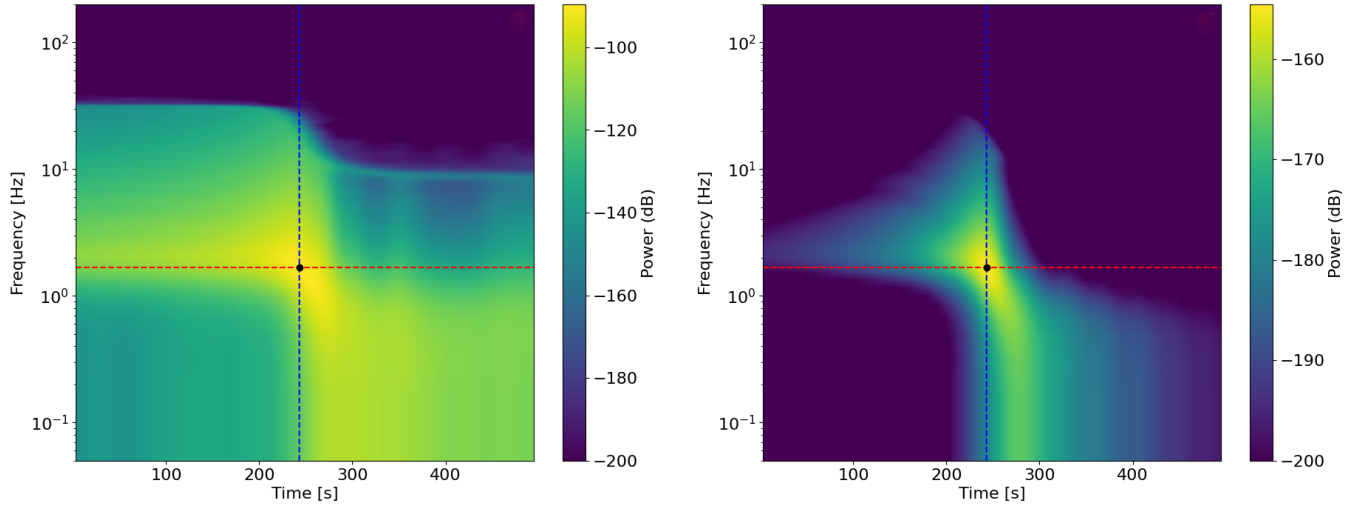


Figure 4.20: $|y| = 5$ km, straight ray tracing (left) and curved ray tracing (right) spectrograms

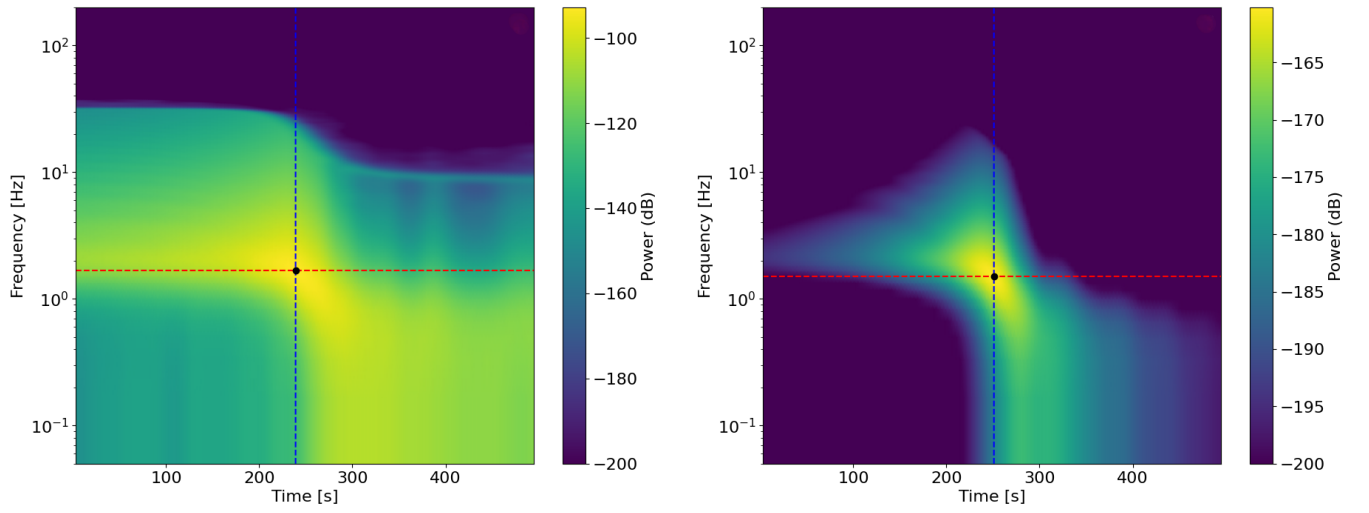


Figure 4.21: $|y| = 7.5$ km, straight ray tracing (left) and curved ray tracing (right) spectrograms

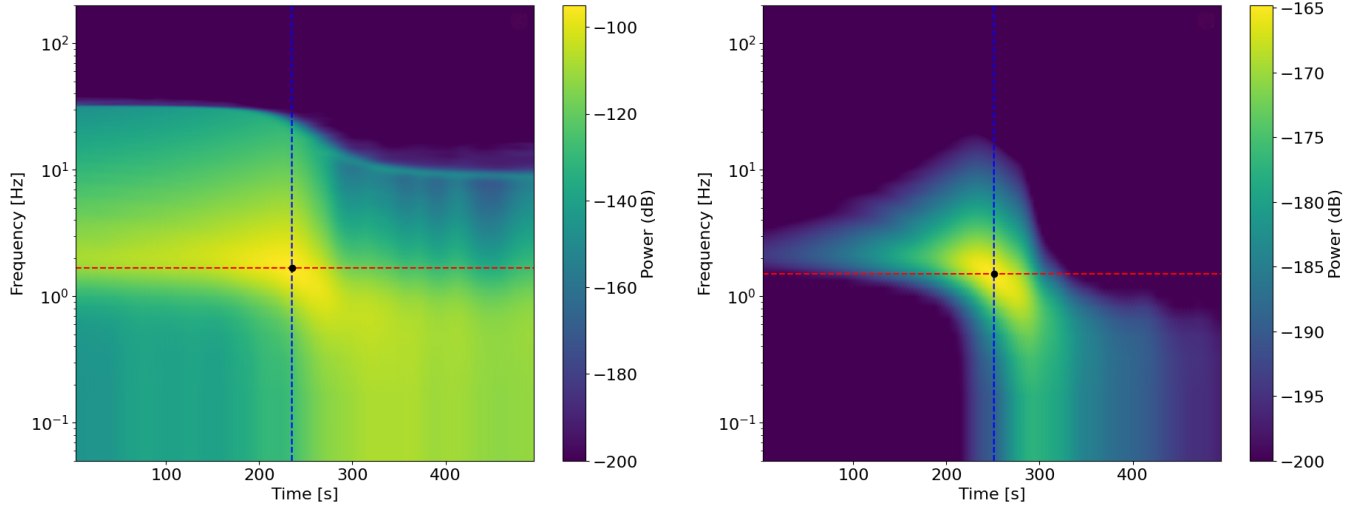


Figure 4.22: $|y| = 10$ km, straight ray tracing (left) and curved ray tracing (right) spectrograms

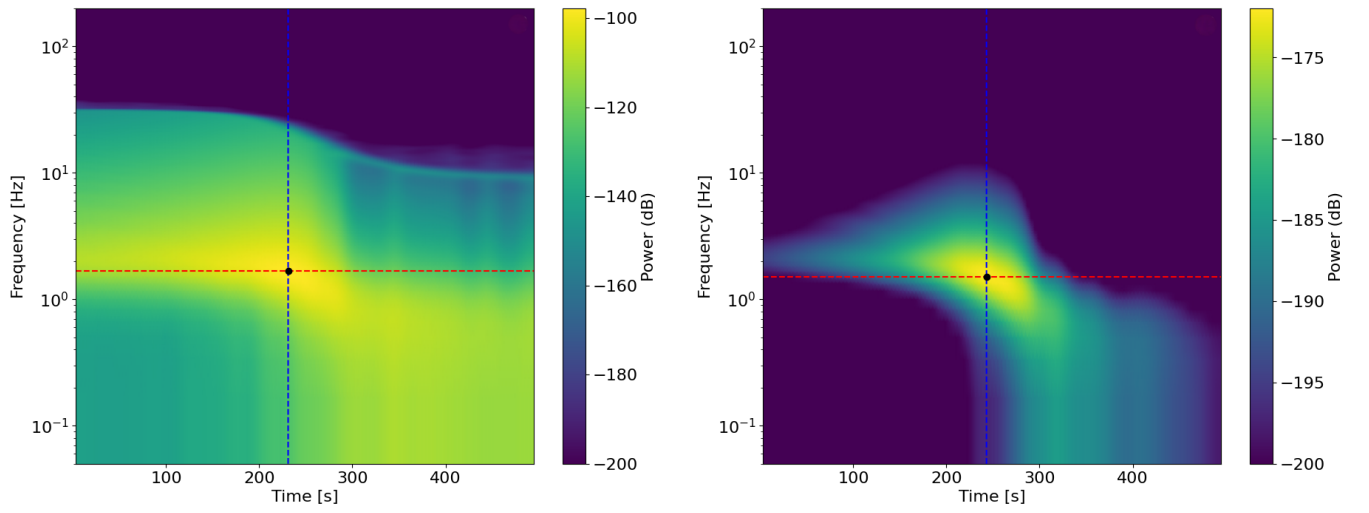


Figure 4.23: $|y| = 15$ km, straight ray tracing (left) and curved ray tracing (right) spectrograms

$ y $	Peak Frequency	Peak Time	Peak Power
1 km	1.83 Hz	243 s	-83.75 dB
3 km	1.83 Hz	239 s	-87.41 dB
5 km	1.67 Hz	243 s	-89.62 dB
7.5 km	1.67 Hz	239 s	-92.73 dB
10 km	1.67 Hz	235 s	-94.95 dB
15 km	1.67 Hz	231 s	-97.77 dB

Table 4.1: Peak frequency, peak time, and peak power detected by straight ray tracing

$ y $	Peak Frequency	Peak Time	Peak Power
1 km	1.83 Hz	243 s	-141.06 dB
3 km	1.67 Hz	243 s	-149.38 dB
5 km	1.67 Hz	243 s	-154.54 dB
7.5 km	1.5 Hz	251 s	-160.18 dB
10 km	1.5 Hz	251 s	-164.8 dB
15 km	1.5 Hz	243 s	-171.97 dB

Table 4.2: Peak frequency, peak time, and peak power detected by curved ray tracing

cies as the airplane approaches the source and lower frequencies when it has passed the source. The highest frequency detected also comes down as $|y|$ increases. This highlights the fact that higher-frequency signals have lower strengths, are unable to travel larger distances, and as a result, are detected predominantly close to the source. The relatively lower-frequency signals, in accordance with the turbulence energy spectrum, have higher strengths - the peak strength detected is always observed at a frequency close to the minimum source frequency of 1 Hz. These lower-frequency signals are thus detected throughout the flight duration as they can travel farther distances. Straight ray spectrograms display the same trend albeit, less evidently due to the set power threshold of -200 dB.

The frequency shift observed due to the variable time delay, and the effect of $|y|$ on the slope of the frequency shift is seen, especially in the straight ray spectrograms where the topmost frequency contour displays the same trend as seen in the analytical frequency

shift plots.

As observed in the curved ray tracing spectrograms, the detection of higher-frequency signals closer to the source is an important, generalizable characteristic that can potentially be used to detect the presence of CAT. Turbulence manifests itself through various frequencies; however, the good thing about it is that the band of frequencies is universally known. Real-time spectrograms can be analyzed specifically for low-frequency patterns. Sudden and abnormal spikes in the detected frequencies amidst clear skies can serve as a cue to alert passengers of possible turbulence. The highest frequency detected decreases as $|y|$ increases. Thus, the difference between the highest frequency and the frequency detected at the outset can be used to estimate the extent of turbulence and its distance from the airplane. Along with the spectral analysis, the temporal information in the spectrogram can identify the onset and duration of turbulence.

The spectrogram contains valuable acoustic data and can be further processed by a deep learning model. Applying the concepts developed in Chapter 3, the spectrogram can be used as an input feature to train a deep learning model to detect CAT. In a regression model, the spectrogram features such as the change in frequency can be used to predict the proximity and intensity of CAT. A classification approach can also be used. By training the model with spectrograms obtained in normal flight conditions, CAT, and in other known and visible types of turbulence (clouds), the model can specifically learn to recognize the patterns of CAT and classify whether a given spectrogram corresponds to CAT or not. The spectrogram can be complemented with information from other onboard sensors that measure temperature and wind speed in a multi-input deep learning model to give more accurate predictions for CAT localization.

The future scope of this deep learning-enabled, acoustics-based method is discussed in the next chapter.

Chapter 5

Conclusion and Scope of Future Work

Clear-air turbulence (CAT) encounters are projected to increase significantly in the future due to climate change and global warming. The inability of the current, onboard, radar systems to detect CAT presents a major challenge to the aviation industry. With acoustics-based methods, there lies scope for detecting the low-frequency acoustic emissions generated by CAT. Thus, this thesis explored low-frequency acoustic source localization (ASL) in the context of CAT detection.

The limitations of beamforming in localizing low-frequency sources necessitated a deep learning approach. Deep learning models for two- and three-dimensional ASL were developed. The models used a shallow convolutional neural network (CNN) with limited training data and sought to optimize the localization performance at low source frequencies under the training constraints. Promising qualitative and quantitative results were obtained which collectively prove the viability of deep learning for low-frequency ASL.

To get a grasp on how a deep learning-enabled, acoustics-based method could potentially work for CAT detection in the future, a large-scale flight case was simulated in this thesis for the auralization of a representative CAT source. Two methods - straight and curved ray tracing - were used and the differences between them were analyzed. While straight ray tracing assumes sound rays travel in straight lines at constant speeds, curved

ray tracing takes into account the atmospheric refraction caused by wind and temperature gradients. Eigenrays were obtained from which, various acoustic parameters were derived to construct the audio signal that an in-flight microphone would detect. The spectrogram of this signal reflected the turbulence spectrum - with higher frequency signals, being weaker in strength and detected only as the airplane got close to the source, while lower frequency signals, due to their higher strengths were detected even at farther distances. This is an important, generalizable trait that can be used to detect and locate CAT.

The scope of future work is ambitious and involves expanding the scale of the current work. The results from the deep learning models provide insight into the parameters that can be changed to expand the scope of these ASL models further and obtain better results. With appropriate resources, deeper networks like ResNets and DenseNets can be trained however, the dataset must be commensurate with the size of the network otherwise, the network will tend to overfit the data and generalize poorly. Studies employing a ResNet-50 used 200-800 million training samples to prevent overfitting. The size of the input feature is also an aspect to consider when using deep neural networks. Smaller input sizes would be reduced by the network to the point that learning the features of the input would become very difficult.

The spectrogram lends itself conveniently for further processing. There exists a lot of scope for using deep learning models to learn the features of the spectrogram and give quick and accurate predictions for CAT localization. By increasing the complexity of auralization through better atmosphere modeling and more realistic source distribution, synthetic spectrogram data can be generated by simulating a large number of flight cases with different flight trajectories. A deep neural network can then be trained on this synthetic database and fine-tuned with relatively small amounts of real, in-flight acoustic data, to give predictions that can potentially be used to detect and locate CAT. Even if accurate localization may not be possible given the complexity of the problem, with the current scarcity of information on CAT, the slightest of useful information from the deep learning model can be a valuable addition to pilot reports and potentially be used to alert passengers, other flights, and airports.

References

- Albawi, S., Mohammed, T. A., and Al-Zawi, S. (2017). Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)*, pages 1–6. Ieee.
- Amyar, A., Modzelewski, R., Li, H., and Ruan, S. (2020). Multi-task deep learning based ct imaging analysis for covid-19 pneumonia: Classification and segmentation. *Computers in biology and medicine*, 126:104037.
- Atlas, D. and Srivastava, R. (1971). A method for radar turbulence detection. *IEEE Transactions on Aerospace and Electronic Systems*, (1):179–187.
- Azaria, M. and Hertz, D. (1984). Time delay estimation by generalized cross correlation methods. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(2):280–285.
- Batchelor, G. K. (1951). Pressure fluctuations in isotropic turbulence. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 47, pages 359–374. Cambridge University Press.
- Beghi, M. G. (2013). *Modeling and Measurement Methods for Acoustic Waves and for Acoustic Microdevices*. BoD–Books on Demand.
- Bianco, M. J., Gerstoft, P., Traer, J., Ozanich, E., Roch, M. A., Gannot, S., and Deledalle, C.-A. (2019). Machine learning in acoustics: Theory and applications. *The Journal of the Acoustical Society of America*, 146(5):3590–3628.

- Blanchard, T., Thomas, J.-H., and Raoof, K. (2020). Acoustic localization and tracking of a multi-rotor unmanned aerial vehicle using an array with few microphones. *The Journal of the Acoustical Society of America*, 148(3):1456–1467.
- Booth, E. and Humphreys, W. (2005). Tracking and characterization of aircraft wakes using acoustic and lidar measurements. In *11th AIAA/CEAS Aeroacoustics Conference*, page 2964.
- Brooks, T. F. and Humphreys, W. M. (2006). A deconvolution approach for the mapping of acoustic sources (damas) determined from phased microphone arrays. *Journal of sound and vibration*, 294(4-5):856–879.
- Cakir, E., Parascandolo, G., Heittola, T., Huttunen, H., and Virtanen, T. (2017). Convolutional recurrent neural networks for polyphonic sound event detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(6):1291–1303.
- Carter, G. (1981). Time delay estimation for passive sonar signal processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(3):463–470.
- Chakrabarty, S. and Habets, E. A. (2017). Broadband doa estimation using convolutional neural networks trained with noise signals. In *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 136–140. IEEE.
- Chiariotti, P., Martarelli, M., and Castellini, P. (2019). Acoustic beamforming for noise source localization—reviews, methodology and applications. *Mechanical Systems and Signal Processing*, 120:422–448.
- Daryan, H., Hussain, F., and Hickey, J.-P. (2020). Aeroacoustic noise generation due to vortex reconnection. *Physical Review Fluids*, 5(6):062702.
- Daryan, H., Hussain, F., and Hickey, J.-P. (2022). Sound generation mechanism of compressible vortex reconnection. *Journal of Fluid Mechanics*, 933:A34.
- de Groot-Hedlin, C. (2008). Finite-difference time-domain synthesis of infrasound propagation through an absorbing atmosphere. *The Journal of the Acoustical Society of America*, 124(3):1430–1441.

- de Santana, L. (2017). Fundamentals of acoustic beamforming. *NATO Educational Notes EN-AVT-287*, 4.
- Dougherty, R., Wang, F., Booth, E., Watts, M., Fenichel, N., and D’Errico, R. (2004). Aircraft wake vortex measurements at denver international airport. In *10th AIAA/CEAS Aeroacoustics Conference*, page 2880.
- Dowling, A. and Hynes, T. (2004). Sound generation by turbulence. *European Journal of Mechanics-B/Fluids*, 23(3):491–500.
- Ellrod, G., Lester, P., and Ehernberger, L. (2003). Clear air turbulence. In Holton, J. R., editor, *Encyclopedia of Atmospheric Sciences*, pages 393–403. Academic Press, Oxford.
- Fernandes, J. d. C. V., de Moura Junior, N. N., and de Seixas, J. M. (2022). Deep learning models for passive sonar signal classification of military data. *Remote Sensing*, 14(11):2648.
- for Standardization, I. O. (1993). *Acoustics: Attenuation of Sound During Propagation Outdoors*. International Organization for Standardization.
- for Standardization, I. O. (1996). *Acoustics-Attenuation of Sound During Propagation Outdoors: Part 2: General Method of Calculation*.
- George, W. K., Beuther, P. D., and Arndt, R. E. (1984). Pressure spectra in turbulent free shear flows. *Journal of Fluid Mechanics*, 148:155–191.
- Gombots, S., Nowak, J. J., and Kaltenbacher, M. (2021). Sound source localization—state of the art and new inverse scheme. *Elektrotechnik und Informationstechnik: e & i*, 138(3):229–243.
- Groves, G. (1955). Geometrical theory of sound propagation in the atmosphere. *Journal of Atmospheric and Terrestrial Physics*, 7:113–127.
- Grumiaux, P.-A., Kitić, S., Girin, L., and Guérin, A. (2022). A survey of sound source localization with deep learning methods. *The Journal of the Acoustical Society of America*, 152(1):107–151.

- Hardin, J. C. and Wang, F. Y. (2003). Sound generation by aircraft wake vortices. Technical report.
- Harvey, B. and O’Young, S. (2018). Acoustic detection of a fixed-wing uav. *Drones*, 2(1):4.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- Hill, D. P., Fischer, F. G., Lahr, K. M., and Coakley, J. M. (1976). Earthquake sounds generated by body-wave ground motion. *Bulletin of the Seismological Society of America*, 66(4):1159–1172.
- Ho, Y. and Wookey, S. (2019). The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. *IEEE access*, 8:4806–4813.
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708.
- Ingård, U. (1953). A review of the influence of meteorological conditions on sound propagation. *The Journal of the Acoustical Society of America*, 25(3):405–411.
- Jain, A. K., Mao, J., and Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer*, 29(3):31–44.
- Jekaterýńczuk, G. and Piotrowski, Z. (2023). A survey of sound source localization and detection methods and their applications. *Sensors*, 24(1):68.
- Joshi, A., Rahman, M. M., and Hickey, J.-P. (2022). Recent advances in passive acoustic localization methods via aircraft and wake vortex aeroacoustics. *Fluids*, 7(7):218.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kolano, R. A. and LAB, P. S. U. U. P. A. R. (1978). Doppler shift in layered media.

- Lacanna, G., Ichihara, M., Iwakuni, M., Takeo, M., Iguchi, M., and Ripepe, M. (2014). Influence of atmospheric structure and topography on infrasonic wave propagation. *Journal of Geophysical Research: Solid Earth*, 119(4):2988–3005.
- Lakkapragada, A., Sleiman, E., Surabhi, S., and Wall, D. P. (2023). Mitigating negative transfer in multi-task learning with exponential moving average loss weighting strategies (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 16246–16247.
- Lighthill, M. J. (1952). On sound generated aerodynamically i. general theory. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 211(1107):564–587.
- Limpinsel, M., Kuo, D., and Vijh, A. (2018). Smarts modeling of solar spectra at stratospheric altitude and influence on performance of selected iii-v solar cells. In *2018 IEEE 7th world conference on photovoltaic energy conversion (WCPEC)(A joint conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC)*, pages 3367–3373. IEEE.
- Ma, W. and Liu, X. (2019). Phased microphone array for sound source localization with deep learning. *Aerospace Systems*, 2(2):71–81.
- Mesaros, A., Heittola, T., Diment, A., Elizalde, B., Shah, A., Vincent, E., Raj, B., and Virtanen, T. (2017). Dcase 2017 challenge setup: Tasks, datasets and baseline system. In *DCASE 2017-Workshop on Detection and Classification of Acoustic Scenes and Events*.
- Michel, U. and Böhning, P. (2002). Investigation of aircraft wake vortices with phased microphone arrays. In *8th AIAA/CEAS Aeroacoustics Conference & Exhibit*, page 2501.
- Niu, H., Gong, Z., Ozanich, E., Gerstoft, P., Wang, H., and Li, Z. (2019). Deep learning for ocean acoustic source localization using one sensor. *arXiv preprint arXiv:1903.12319*.
- O’Shea, K. and Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.
- Ostashev, V. E. and Wilson, D. K. (1997). *Acoustics in moving inhomogeneous media*. E & FN Spon London.

- Peyvandi, H., Farrokhrooz, M., Roufarshbaf, H., and Park, S.-J. (2011). Sonar systems and underwater signal processing: classic and modern approaches. *SONAR systems*, pages 173–206.
- Posmentier, E. S. (1974). 1-to 16-hz infrasound associated with clear air turbulence predictors. *Journal of Geophysical research*, 79(12):1755–1760.
- Prime, Z. and Doolan, C. (2013). A comparison of popular beamforming arrays. *Proceedings of the Australian Acoustical Society AAS2013 Victor Harbor*, 1:5.
- Raspet, R. and Webster, J. (2015). Wind noise under a pine tree canopy. *The Journal of the Acoustical Society of America*, 137(2):651–659.
- Rayleigh (1879). Xxi. investigations in optics, with special reference to the spectroscope. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 8(49):261–274.
- Reynolds, O. (1874). On the refraction of sound by the atmosphere. *Proceedings of the Royal Society of London*, 22(148-155):531–548.
- Rizzi, S. A., Aumann, A. R., Lopes, L. V., and Burley, C. L. (2014). Auralization of hybrid wing–body aircraft flyover noise from system noise predictions. *Journal of Aircraft*, 51(6):1914–1926.
- Rojas, R. and Rojas, R. (1996). The backpropagation algorithm. *Neural networks: a systematic introduction*, pages 149–182.
- Rubin, W. L. (2005). The generation and detection of sound emitted by aircraft wake vortices in ground effect. *Journal of Atmospheric and Oceanic Technology*, 22(5):543–554.
- Sarradj, E. and Herold, G. (2017). A python framework for microphone array data processing. *Applied Acoustics*, 116:50–58.
- Schäfer, P. and Vorländer, M. (2021). Atmospheric ray tracing: An efficient, open-source framework for finding eigenrays in a stratified, moving medium. *Acta Acustica*, 5:26.

- Schechter, D. A. (2011). A method for diagnosing the sources of infrasound in convective storm simulations. *Journal of applied meteorology and climatology*, 50(12):2526–2542.
- Schönhals, S., Steen, M., and Hecker, P. (2013). Towards wake vortex safety and capacity increase: the integrated fusion approach and its demands on prediction models and detection sensors. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 227(1):199–208.
- Sedunov, A., Sutin, A., Sedunov, N., Salloum, H., Yakubovskiy, A., and Masters, D. (2016). Passive acoustic system for tracking low-flying aircraft. *IET Radar, Sonar & Navigation*, 10(9):1561–1568.
- Shams, Q. A., Zuckerwar, A. J., Burkett, C. G., Weistroffer, G. R., and Hugo, D. R. (2013). Experimental investigation into infrasonic emissions from atmospheric turbulence. *The Journal of the Acoustical Society of America*, 133(3):1269–1280.
- Sutin, A., Salloum, H., Sedunov, A., and Sedunov, N. (2013). Acoustic detection, tracking and classification of low flying aircraft. In *2013 IEEE International Conference on Technologies for Homeland Security (HST)*, pages 141–146. IEEE.
- Sylvander, M., Ponsolles, C., Benahmed, S., and Fels, J.-F. (2007). Seismoacoustic recordings of small earthquakes in the pyrenees: Experimental results. *Bulletin of the Seismological Society of America*, 97(1B):294–304.
- Tosi, P., Sbarra, P., and De Rubeis, V. (2012). Earthquake sound perception. *Geophysical research letters*, 39(24).
- Varzandeh, R., Adiloğlu, K., Doclo, S., and Hohmann, V. (2020). Exploiting periodicity features for joint detection and doa estimation of speech sources using convolutional neural networks. In *ICASSP 2020-2020 IEEE International conference on acoustics, speech and signal processing (ICASSP)*, pages 566–570. IEEE.
- Vecchiotti, P., Principi, E., Squartini, S., and Piazza, F. (2018). Deep neural networks for joint voice activity detection and speaker localization. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 1567–1571. IEEE.

- Vera-Diaz, J. M., Pizarro, D., and Macias-Guarasa, J. (2018). Towards end-to-end acoustic localization using deep learning: From audio signals to source position coordinates. *Sensors*, 18(10):3418.
- Wang, F., Wassaf, H., Gulrud, A., Delisi, D., and Rudis, R. (2005). Acoustic imaging of aircraft wake vortex dynamics. In *23rd AIAA Applied Aerodynamics Conference*, page 4849.
- Watson, L. M., Dunham, E. M., Mohaddes, D., Labahn, J., Jaravel, T., and Ihme, M. (2021). Infrasound radiation from impulsive volcanic eruptions: nonlinear aeroacoustic 2d simulations. *Journal of Geophysical Research: Solid Earth*, 126(9):e2021JB021940.
- Watson, L. M., Iezzi, A. M., Toney, L., Maher, S. P., Fee, D., McKee, K., Ortiz, H. D., Matoza, R. S., Gestrich, J. E., Bishop, J. W., et al. (2022). Volcano infrasound: progress and future directions. *Bulletin of Volcanology*, 84(5):44.
- Wescott, J. W. (1964). *Acoustic detection of high-altitude turbulence*. University of Michigan, Institute of Science and Technology, Acoustics and
- Xiao, X., Zhao, S., Zhong, X., Jones, D. L., Chng, E. S., and Li, H. (2015). A learning-based approach to direction of arrival estimation in noisy and reverberant environments. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2814–2818. IEEE.
- Xu, P., Arcondoulis, E. J., and Liu, Y. (2021). Acoustic source imaging using densely connected convolutional networks. *Mechanical Systems and Signal Processing*, 151:107370.
- Yalta, N., Nakadai, K., and Ogata, T. (2017). Sound source localization using deep learning models. *Journal of Robotics and Mechatronics*, 29(1):37–48.
- Zhao, J., Luo, X., and Liu, H. (2023). An airborne visible light lidar’s methodology for clear air turbulence detection based on weak optical signal. In *Photonics*, volume 10, page 1185. MDPI.

Zuckerwar, A. J., Shams, Q., and Knight, H. K. (2013). Infrasonic emissions from aircraft wake vortices: Experimental results. *The Journal of the Acoustical Society of America*, 134(5_Supplement):4128–4128.