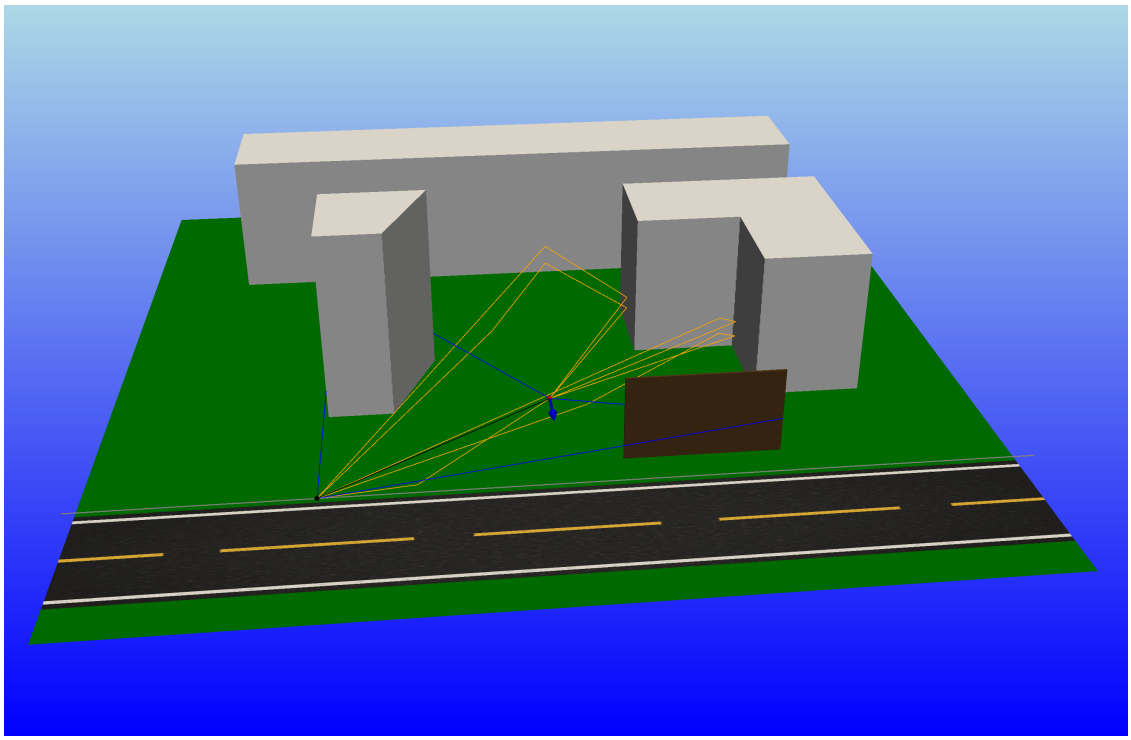


Exploring Auralizations with Multiple Doppler Effects for Urban Sound Planning

Master's thesis in Sound and Vibration

JESPER HOLSTEN

DEPARTMENT OF APPLIED ACOUSTICS

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023
www.chalmers.se

Master's thesis 2023

Exploring Auralizations with Multiple Doppler Effects for Urban Sound Planning

JESPER HOLSTEN

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Architecture and Civil Engineering
Division of Applied Acoustics
Chalmers University of Technology
Gothenburg, Sweden 2023

Exploring Auralizations with Multiple Doppler Effects for Urban Sound Planning
JESPER HOLSTEN

Supervisors:
Jens Forssén, Chalmers University of Technology
Ulf Peter Svensson, Norwegian University of Technology

Examiner:
Jens Forssén, Division of Applied Acoustics,
Chalmers University of Technology

Master's Thesis 2023
Department of Architecture and Civil Engineering
Division of Applied Acoustics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Direct, specular, and diffracted sound paths from a source (black sphere) to a receiver (red sphere) at a given time.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2023

Exploring Auralizations with Multiple Doppler Effects for Urban Sound Planning
JESPER HOLSTEN
Department of Architecture and Civil Engineering
Chalmers University of Technology

Abstract

Increasing noise levels due to urban densification is a major concern in many countries. Long-term noise exposure is associated with a long list of health risks, including heart attacks, diabetes, and sleep disorders, to mention a few. To mitigate such health risks, thorough urban sound planning is a must. Noise prediction models are often used to estimate A-weighted sound pressure levels in a sound planning process. Nevertheless, solitary use of sound pressure levels is not sufficient to evaluate the psychoacoustic aspect of sound. Auralizations were therefore proposed in this thesis as a complement to existing noise prediction models in urban sound planning. Moreover, a specific aim was to validate the plausibility of auralizations with multiple Doppler effects versus a single Doppler effect for pass-by vehicles. This objective was tested in different virtual urban scenes, with and without edge diffraction. A listening experiment was conducted for this purpose, and in the first part, a paired comparison between a single Doppler and multiple Doppler stimuli, the majority of the participants voted for the multiple Doppler stimuli as more plausible. A two-sided Fisher's exact test on these results gave a resultant p-value of 0.0004, indicating a significant difference in preference between single and multiple Doppler effects. Furthermore, an odds ratio of 0.247 indicated that the odds of preferring the multiple Doppler stimuli were higher than for the single Doppler stimuli.
The second part of the listening experiment intended to find an appropriate sample rate of the source position. The results showed that more complex urban scenes, e.g., with diffraction from a noise screen, required higher sampling rates to achieve more plausible auralizations with smoother transitions to shadow zones. Source position sample rates down to 10 Hz seemed sufficient in simple scenes.
Ambient sound was recorded and added to a set of pass-by auralizations in the last part of the listening test. A paired comparison between stimuli with and without ambient sound was carried out to check if the ambient sound made an improvement to the perceived plausibility. Increased plausibility from ambient sound could not be proved due to the small sample size.

Keywords: Auralization, Urban Acoustics, Diffraction, Doppler Effect, Outdoor Sound Propagation

Acknowledgements

This thesis was written at Chalmers University of Technology in Gothenburg as a part of my exchange studies from the Norwegian University of Science and Technology in Trondheim. Through five years of studies at NTNU and Chalmers, I have been given experiences I would never get elsewhere. I have gained valuable knowledge and learned things about myself that have helped shape my perspective and personal growth. But most importantly, I have gained friendships for life.

The thesis would not have been accomplished without the help of countless contributors. I want to begin by thanking the administration of my home university, the Norwegian University of Technology, NTNU, in Trondheim, for making this opportunity for me. Thank you, Ellen Hove, Nina Moxnes, and Kirsti Klemetsaune. I also have to thank the Chalmers administration and admissions office for enrolling me into their exchange program and letting me take courses at the Division of Applied Acoustics. Thank you, Carolina Strandsäter, for all the guidance concerning the exchange, and thank you, Astrid Pieringer and Wolfgang Kropp, for welcoming me as a student in your Division.

While working on this thesis, I conducted a listening test enabled by a resource funded by the HEAD Genuit Foundation under the agreement P-22/01-W. I am very grateful for this and the help from all the participants. Leon Müller and Hannes Helmholz from the acoustic division at Chalmers helped me with the equipment used for the listening experiment and also some background sound recordings that I did. Moreover, Hannes reviewed some of my code implementations with me. Jens Ahrens, also from the acoustic Division at Chalmers, assisted with explanations and discussions related to digital signal processing, which I am very thankful for.

I'd also like to thank Philipp Schäfer, Michael Vorländer, and Jonas Steinen from the Institute for Hearing Technology and Acoustics at the Univerity of Aachen for detailed and comprehensive answers to my numerous emails. I highly recommend checking out their research in auralization and virtual acoustics [1].

My friend and co-student, Halldór Einarsson, was my opponent and gave me a lot of valuable feedback on my report. Additionally, he encouraged a critical approach to my methods and implementations. Another good friend who deserves a special thanks is Martin Krokan Hovden. He has spent much of his time discussing several problems I had in my code implementations.

I cannot express my gratitude fully without thanking my friends, family, and my better half. Your constant support, motivation, and understanding have been invaluable.

A special thanks is reserved for my supervisor from NTNU, Peter Svensson, for

[1] www.virtualacoustics.org [1].

contributing to detailed theoretical discussions. Peter's research and knowledge in edge diffraction, together with his humble personality and humor, have been a huge inspiration.

Lastly, my most sincere gratitude goes to my supervisor from the acoustic Division at Chalmers, Jens Forssén. Jens has consistently been open-minded to my ideas, and his creativity and engagement have made writing this thesis such an enjoyable experience. Jens has been a constant source of knowledge and has helped me through the ups and downs with his expertise in outdoor sound propagation. Without you, it wouldn't have been possible. I cannot thank you enough, Jens!

Jesper Holsten, Gothenburg, June 2023

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

DFT	Discrete Fourier Transform
DSP	Digital Signal Processing
FFT	Fast Fourier Transform
FR	Frequency Response
GA	Geometrical Acoustics
HF	High-Frequency
HRIR	Head Related Impulse Response
IFFT	Inverse Fast Fourier Transform
IR	Impulse Reponse
IS	Image Source
ISM	Image Source Method
ITD	Interaural Time Delay
PDE	Partial Differential Equation
RIR	Room Impulse Response
SIMO VDL	Single-Input Multiple-Output Variable Delay Line
SPL	Sound Pressure Level
UTD	Uniform Theory of Diffraction

Nomenclature

Below is the nomenclature of symbols that have been used throughout this thesis.

Symbols

Symbol	Quantity	Unit
$\vec{R_{\mathrm{N}}}$	Wavefront normal	
θ	Angle between the wavefront normal and the normal incidence	rad
Z_i	Characteristic impedance of medium i	Ω
R	Plane wave reflection factor, length of an acoustic path	, m
p	Sound pressure level	Pa
n	Image source method order, sample index of a signal	
t_{ISM}	Computation time of the image source method	s
$N_{\mathrm{r.p}}$	Number of reflection planes	
A_{diff}	Insertion loss of a sound barrier	dB
N_f	Fresnel number	
h_{eff}	Effective height of a barrier	m
λ	Wavelength	m
t	Time	s
c	Speed of sound	m/s
∇^2	Laplacian operator	
a	Radius of a monopole source	m
$Q(t)$	Amplitude at time t	
$\delta(t)$	Dirac-delta function	
$\vec{x}$	Receiver position	m
$\vec{x_s}$	Source position	m

G_0	The free-field Green's function	
τ	Emission time	s
c_0	Speed of sound in air	m/s
$\vec{v_s}$	Source velocity	m/s
$\vec{v_r}$	Source velocity in the direction of the receiver	m/s
$\|\cdot\|^2$	Norm operator	
M_r	Relative Mach number	
α	Angle between $\vec{v_s}$ and $\vec{v_r}$, air attenuation coefficient	rad, dB/m
f	Frequency	Hz
p_a	Ambient atmospheric pressure	Pa
p_r	Reference atmospheric pressure	Pa
T	Ambient atmospheric temperature	K
T_0	Reference air temperature	K
f_{rO}	Oxygen relaxation frequency	Hz
f_{rN}	Nitrogen relaxation frequency	Hz
h	Molar concentration of water vapour, impulse response	%,
$*$	Convolution operator	
j	Imaginary unit	
k	Wave number, Fourier component number	$\mathrm{rad{\cdot}m^{-1}}$,
N	Number of samples	
M	Length of a delay line in samples	
$X(f)$	Frequency representation of a signal $x(t)$	
Δt	Time difference	s
H_0	Null hypothesis	
H_1	Alternative hypothesis	
$\mathcal{R}$	Set of depropagated recordings	

Contents

List of Figures

List of Tables

1

Introduction

1.1 Motivation

Exposure to noise is a pressing problem in cities worldwide. The main contributors to noise emissions are sound sources such as light or heavy road vehicles, trains, and airplanes, as well as fans, wind turbines, and other industrial sources. As many cities densify at high rates, the volume of traffic flow increases, resulting in overall higher noise levels. The impact of excessive noise on humans has been correlated to a range of medical problems such as myocardial infarction, cardiovascular diseases, increased stress levels, hypertension, and sleep disorders [2]. Continuous sleep disruption can have a detrimental impact on the human body. This includes deteriorated cognitive abilities, increased risk of diabetes, alterations in the immune system, and even increased rates of mortality [3, 4, 5]. In other words, minimizing noise pollution is vital to maintain good public health. Furthermore, noise emission not only affects humans but also negatively affects wildlife. Besides the side effects of noise pollution, such as sleep deprivation or increased stress levels, many animals rely on their auditory cues to gain knowledge about their surrounding environment. Constant noise exposure could potentially undermine this ability which in turn may result in a permanent change of behavior and potential loss of habitat [6]. On the other side, noise is not always entirely unfavorable. In some cases it could add to a vibrant atmosphere, broadening the identity of a place. But with all the health-related concerns that come with noise exposure, together with an increasing urban expansion, it's reasonable to believe that the demand for thorough noise prediction and sound planning is higher than ever.

The Nordic prediction model from 1996, *Nord96*, is the current Swedish standard for road traffic noise mapping. Other models, such as the Nord2000 and CNOSSOS-EU are used in other Nordic countries and Europe respectively [7]. These models serve the purpose of predicting road traffic noise levels at already existing or planned constructions. However, according to [7], they lack the ability to accurately predict noise levels, e.g., for vehicles at low speeds when the traffic flow is dense but uneven, with a varying number of constant-speed and accelerating vehicles. The models could also yield inaccurate results if the infrastructural layout affects how the traffic flows, for instance, in areas with abrupt turns or at traffic lights. Additionally, since vehicles have evolved a lot due to technology in the past decades, the prediction

models must be able to keep up with the sound emission properties of potential new types of vehicles. A common factor for many of the existing noise prediction models is that they are solely based on A-weighted sound pressure levels. However, solitary use of single-number sound pressure levels yields inadequate information about the human perception of sound. While it's assumed that there is a strong connection between noise prediction and sound planning, a psychoacoustic aspect could also be valuable.

In many countries, the only interaction from civilians in urban sound planning processes is based on former citizen complaints about high noise levels or from residents using their right to amend on nearly finalized projects [8]. More recent studies have pointed out that urban planning could profit from civic engagement [9, 10]. They justify this with several arguments. Firstly, it could enhance social cohesion in the community, giving residents a stronger sense of belonging. Secondly, it is economically sustainable since it's expected that locals are more willing to invest in the community. Thirdly, it can strengthen the bond between the local communities and the government since encouraging voluntary work builds trust between them. If planners decide to let citizens participate in a sound planning process, a natural question is how. Since the citizens are not expected to have the acoustic knowledge required for a sustainable sound design, they might need education.

The apparent shortcomings in current noise prediction models and the discussion about citizen inclusion in urban planning suggest exploring alternative approaches to sound planning processes. This thesis proposes auralizations from virtual environments to complement noise mapping from existing noise prediction models. Auralizations, i.e. audio renderings that can be listened to, are easily integrable with noise mapping software as both are heavily reliant on geometrical frameworks. Furthermore, auralizations are portable. Sound files can be easily stored, transferred, and played back on different devices such as computers, phones, and other audio devices. The portability makes it possible to present auralizations e.g. in emails, at conferences, or at public events. This allows for a wider discussion between stakeholders in a sound planning process. Moreover, auralizations are something everyone can have an opinion about regardless of education which makes it highly convenient.

1.2 Thesis

This thesis aims to gain insight into dynamic urban auralizations incorporating multiple Doppler effects. The backend for such auralizations requires a comprehensive geometrical framework for the determination of acoustical propagation paths, including specular reflections and diffraction. Furthermore, the framework should be able to effectively visualize virtual urban scenes to allow for a better understanding of the dynamic relationship between moving sources and their sound characteristics. Auralizations are often implemented by finding the impulse response from geometrical acoustics, followed by a convolution stage. However, such an approach is not possible for simulations with multiple Doppler effects because every transmission path has to be handled separately with interpolation- and filtering techniques.

Therefore an alternative approach must be considered, which is investigated here.

Additionally, the thesis aims to evaluate the perception of auralizations with multiple Doppler effects. Through listening experiments and subjective ratings, psychoacoustic parameters of auralizations can be assessed. These results can reveal potential drawbacks or benefits of auralizations with multiple Doppler effects.

2

Theory

2.1 Auralization

Auralization is a term within acoustics that describes how a numerical simulation can produce an audible output [11]. For stationary sound sources and receivers, auralization is usually a two-staged process [12]. The first stage involves finding the Impulse-Reponse (IR) for the environment, and the second is a convolution stage combining the source signal with the IR. When dealing with dynamic environments (moving sound sources or receivers), the two-staged process can be repeated over multiple time frames with an overlap and add method to simulate source motion.

2.1.1 Brief History of Auralization

Auralization in itself is nothing new. The idea flourished already in 1929 in Munich, where the German acoustician Friedrich Spandöck and a group of researchers conducted an experiment on sound field modeling in rooms [11]. They used measurements on a scaled model of a room to try recreating the room's response to sound. A special loudspeaker was used to achieve binaural sound, together with a microphone that picked up sound [13]. In 1962 when the development of computers had made significant progress, Schröder, Atal, and Bird presented the general procedure for computer-based auralizations [14]. This was before the term auralization became a thing. Five years later, in 1967, Schröder was asked to assess the initially disastrous acoustics of the Philharmonic Hall, now renamed David Geffen Hall in New York. He came up with the idea to use a ray tracing measurement technique based on digital signal processing to predict the hall's room impulse response (RIR). Unknowingly, such a method was simultaneously under development by Krokstad in Trondheim. These events, amongst others, caused a breakthrough in the modern use of computers in sound field modeling, and at the beginning of the 1990s, the term "auralization" was unveiled [11, 13]. Since then, auralization has become a widely known term in acoustics and has found its application in many places, e.g., in virtual reality.

2.2 Geometrical Acoustics

Geometrical acoustics (GA) assumes that sound propagates in straight lines, often referred to as rays or paths. Reflected rays obey Snell's law, stating that the incident angle θ_i equals the outgoing angle θ_o as shown in figure 2.2. Diffracted sound, however, is neglected in GA since it contradicts the fundamental straight-line constraint. This means that a diffraction model has to be implemented separately. Figure 2.1 shows the color convention for acoustical sound paths that will be used throughout this thesis.

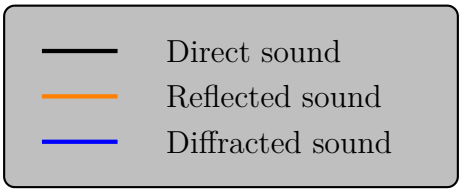


Figure 2.1: Color convention for acoustical sound paths.

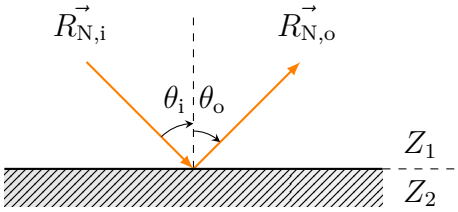


Figure 2.2: Incident and outgoing wavefront normals following Snell's law according to GA.

Geometrical acoustics is applicable when the wavelength of the sound waves is much smaller than the dimensions of the features in the scene (scenes are explained in the following subsection). This implies that GA-based models are high-frequency approximations.

The plane-wave reflection factor determines the ratio of the sound pressure amplitude of the reflected to the incident wave given by

$$R = \frac{p_\mathrm{o}}{p_\mathrm{i}} = \frac{Z_2 \cos\theta_\mathrm{i} - Z_1}{Z_2 \cos\theta_\mathrm{i} - Z_1}, \tag{2.1}$$

where Z_1 and Z_2 are the characteristic impedances of the two media seen in figure 2.2, θ_i is the incident angle, and p_i and p_o are the amplitudes of the incident and outgoing/reflected wave assuming plane waves. Furthermore, R, Z_1 and Z_2 are complex numbers.

The premise for using GA in sound modeling is that the source and receiver are not too close to absorbing surfaces such that plane-wave reflection factors can be assumed. Hence, GA only gives approximate solutions for absorbing walls [11]. However, GA is exact when the reflection factor $|R| = 1$ which is true for hard ($Z = \infty$) or soft ($Z = 0$) walls.

2.2.1 Scenes

A scene is a 3D mesh defining the geometry of an urban environment and contains *features*, which can be anything from buildings to roads and noise barriers. The scene is the fundamental framework to which geometrical acoustics can be applied. In addition to defining the 3D geometry of the urban environment, the scene is also responsible for keeping track of e.g. material properties for the different features in the scene, meteorological conditions such as wind, temperature, and air density, as well as source and receiver movement. The features in the scene can be covered with textures to improve the visual appearance. Figure 2.3 shows an example scene with various features in addition to a source and a receiver. The blue arrow with its origin in the receiver position illustrates the orientation of the receiver for binaural simulations. The gray line just above the road is the source trajectory.

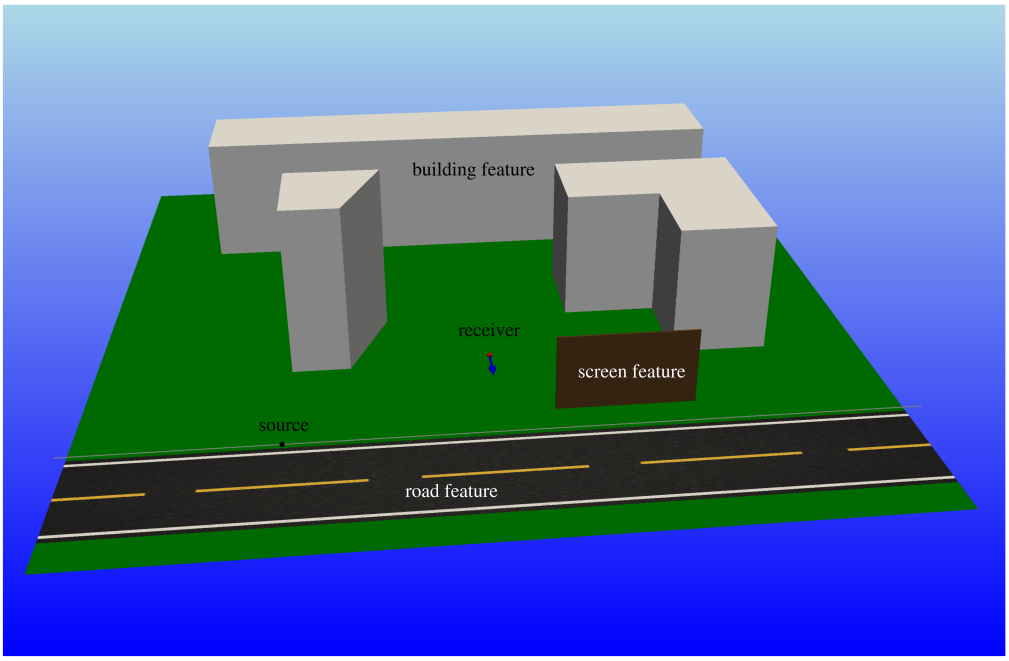


Figure 2.3: An example scene with different features and source and receiver.

2.2.2 Specular Reflection and the Image Source Method

Specular reflections of sound are analogous to how light reflects on a mirror. Such reflections can be acquired with a recursive numerical method called the image source method (ISM), which finds n-th ordered reflected paths from a source to a receiver. The order n of the ISM determines the number of reflection points in the reflected

transmission paths. In the first step of the method, the order is set to $n = 1$, where the primary sound source can be referred to as the image source of order $n = 0$. All image sources in order $n-1$ are then mirrored about all planes in the scene, and the corresponding mirror points are called image source *candidates* as they are not yet valid image sources. To become valid image sources, each image source candidate has to undergo a visibility check and an obstruction check. The visibility can be checked by drawing a straight line from the image source candidate to the receiver. If this line intersects the confined plane that the image source originated from, the image source candidate is visible to the receiver, and a reflection may take place in the plane. Lastly, to check if the path is obstructed, the full path between the source and the receiver has to be constructed to check if it collides with any of the scene's features. If the path is unobstructed, the image source candidate has now become an image source. By increasing the order and repeating the method until the desired order is reached, all specular reflections will be found [15].

The image source method can be quite costly if the image source order is high, especially for complex scenes with many features. If a scene has $N_{\text{r.p.}}$ reflection planes, the original source has to be mirrored about all these planes and checked for visibility and obstruction. Then, this step has to be repeated for every constructed image source of the previous order, even the ones that didn't pass the obstruction and visibility check. That means that the computation time is proportional to

$$t_{\text{ISM}} \propto N_{\text{r.p.}}^{n}, \tag{2.2}$$

where n is the image source order. There are ways to accelerate the ISM significantly, e.g. by beam tracing algorithms or similar, but that is not within the scope of this thesis.

2.3 Diffraction

Since geometrical acoustics is unsatisfactory for computing diffracted paths, the resultant sound field from solely GA will lack continuity near shadow zones. Edge diffraction is, therefore, essential to fill in the gaps where these discontinuities exist. There are many different diffraction models, both wave-based and geometrical methods. Since auralization is reliant on a robust geometrical framework, the geometrical models are the most viable. The following subsections list a few potential diffraction models.

2.3.1 Maekawa

In 1968 Maekawa published a paper on noise reduction by screens where he described a graphical calculation method for shielding effects without the use of a computer [16]. His model is based on Kirschoff's diffraction theory and is nowadays recognized as the *rubber band model* as it is only dependent on the shortest path over a screen

from a source to a receiver. Although the model has been used for noise prediction purposes over the years, the model is not well suited for auralization purposes because it cannot handle mixed problems with diffraction and reflection [17].

2.3.2 UTD

The Uniform Theory of Diffraction (UTD) is a high-frequency approximation model based on the diffraction of light. The main concept of UTD is that edges from geometrical features become sources of diffracted paths. Paths that have been diffracted once can be reflected and diffracted again before they are picked up by the receiver. The UTD follows Fermat's principle, indicating that rays that are diffracted over an infinite wedge take the shortest path of propagation before reaching the receiver. A ray that is diffracted over a wedge results in a cone of new rays, where the angle of the cone is equivalent to the angle between the edge of the wedge and the incident ray. This is what makes UTD slightly challenging to implement because finding all the rays requires an efficient search algorithm. Funkhouser et al. [18] proposed in 2001 a beam-tracing method that can effectively find diffracted paths in complex virtual environments. Another rotation-based algorithm for finding such paths was proposed by Steinen and Vorländer in 2017 [17]. The UTD has thereby been successfully implemented in virtual environments for auralization purposes.

2.3.3 Harmonoise

The Harmonoise model has the simplest diffraction model of the above-mentioned, except for Maekawa's empirical model. Harmonoise is a European outdoor noise prediction method for road traffic and railway noise [19]. The model is a physical source model designed to estimate equivalent continuous A-weighted sound pressure levels in 1/3-octave bands for given outdoor scenarios. Moreover, Harmonoise provides a set of relatively simple equations and methods to predict e.g. attenuation due to diffraction over a screen. As Harmonoise is initially designed for noise mapping, an interesting approach is to see whether Harmonoise can be employed for auralizations or not. The diffraction model is based on the Deygout solution, and the attenuation in dB for first-order diffractions can be calculated as

$$\begin{aligned}
A_{\text{diff}} &= 0, && N_f \leq -0.25 \\
A_{\text{diff}} &= 6 - 12\sqrt{-N_f}, && -0.25 \leq N_f < 0 \\
A_{\text{diff}} &= 6 + 12\sqrt{N_f}, && 0 \leq N_f < 0.25 \\
A_{\text{diff}} &= 8 + 8\sqrt{N_f}, && 0.25 \leq N_f < 1 \\
A_{\text{diff}} &= 16 + 10\log(N_f), && N_f \leq 1,
\end{aligned} \tag{2.3}$$

where N_f is the Fresnel number. The Fresnel number can be computed as

$$N_f = \text{sign}(h_{\text{eff}}) \cdot \frac{2(R_S + R_R - R)}{\lambda}, \tag{2.4}$$

where h_{eff} is the effective height of the barrier (see figure 2.4), R_S is the length of the source to the apex point on the barrier, R_R is the length from the apex point of the barrier to the receiver, R is the direct path between the source and the receiver, and λ is the wavelength of the diffracted wave. The sign function $\text{sign}(h_{\text{eff}})$ is $+1$ when R is obstructed by the barrier and -1 when it is not.

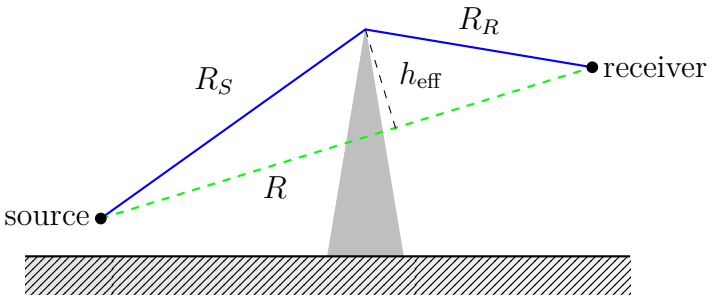


Figure 2.4: Naming conventions for a diffracted path in the Harmonoise model.

For second and higher-order diffractions, the method is similar, but the insertion loss A_{diff} is accumulated with respect to one barrier at a time. This methodology is explained further in section 6.4.3 in [19].

2.3.4 Convex Hull

The convex hull is a geometric concept that can be used to find diffracted paths. The convex hull defines the outer enclosed convex path given a cluster of points or vertices. In figure 2.5, a two-dimensional convex hull computer algorithm is used with 30 random points as input. Vertices in the two-dimensional convex hull are always returned in counterclockwise order.

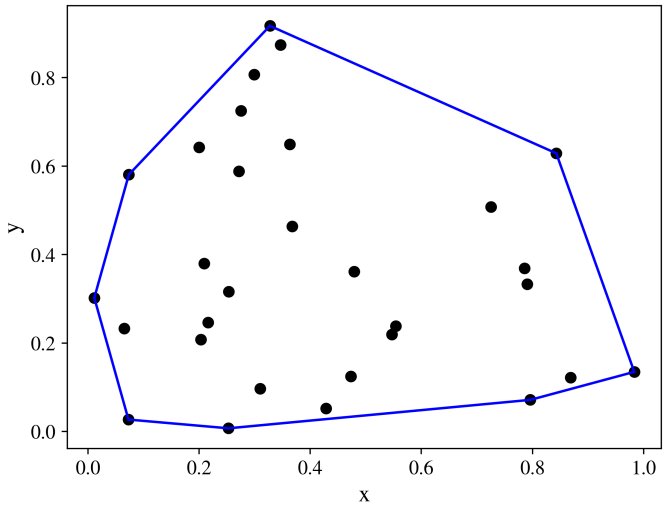


Figure 2.5: Convex hull around a cluster of points.

2.4 Wave Propagation

A wave is a disturbance of some physical quantity that propagates through space without transporting matter. Sound waves are a form of energetic waves that propagate through a medium such as air, water, or solids. Airborne sound waves are non-dispersive, meaning all frequencies propagate at the same speed, roughly 343 m/s, at 20 °C. Since the speed of sound is low compared to other types of waves, propagation delay is a key factor for accurate sound propagation modeling. This section introduces some elementary concepts within wave theory.

2.4.1 The Classical Wave Equation and Green's Function

The rudimentary equation describing the behavior of waves in both space and time is called the classical wave equation and is derived from Maxwell's equations. The classical wave equation is a partial differential equation (PDE) written as

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0, \tag{2.5}$$

where ∇^2 is the laplacian operator, p is the sound pressure, c is the speed of sound and t is time [20]. The solutions to the wave equation are superpositions of waves in 1D, 2D or 3D.

One fundamental solution to the wave equation comes from excitation with a monopole sound source. A monopole is a small spherical radiator where the volume increases and decreases systematically. The sphere has a radius a, where $a \to 0$ represents a point.

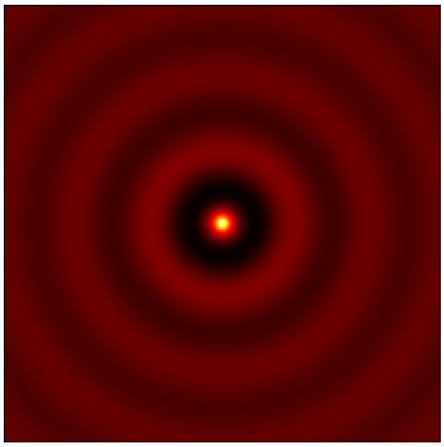


Figure 2.6: Snapshot of the sound field from a monopole in free field.

In the time domain, a monopole can be represented as a Dirac-delta pulse

$$p(\vec{x}, t) = Q(t)\delta(\vec{x} - \vec{x_s}(t)), \tag{2.6}$$

where $\vec{x}$ is the receiver position, $\vec{x_s}(t)$ is the position of the monopole, and $Q(t)$ is the amplitude at time t [21]. The sound pressure at $\vec{x_s}(t)$, i.e. $\vec{x} = \vec{x_s}$, is infinite but decreases radially as $1/R$. The sound field from a monopole at an arbitrary time can be seen in figure 2.6. The solution of the wave equation (eq. 2.5) for such excitations is called a Green's function. The free-field Green's function [21] can be expressed as

$$G_0 = \frac{1}{4\pi R}\delta(t - \tau - R/c_0), \tag{2.7}$$

where c_0 is the speed of sound and R is the distance between the source and receiver at the emission time τ, i.e.

$$\vec{R}(\tau, x) = \vec{x} - \vec{x_s}(\tau), \qquad R = \|\vec{R}\|. \tag{2.8}$$

2.4.2 Doppler Effect

The Doppler effect is the characteristic change of pitch when a sound source is passing by an observer. A signal from a source moving toward an observer would be experienced by the observer as the same signal but with a higher pitch and a lower pitch for a source moving away from the observer. In addition to a change of pitch, the Doppler effect will also alter the amplitude of the sound pressure. Consider the illustration in figure 2.7. The sound source has position $\vec{x_s}$, and the receiver has position $\vec{x}$. The source is moving along the horizontal dashed line with the velocity

$\vec{v_s}$. The receiver is stationary. The relative velocity v_r of the source in the direction of the receiver is given by finding the projection of $\vec{v_s}$ onto $\vec{R}$ as

$$\vec{v_r} = \text{proj}_{\vec{R}}(\vec{v_s}) = \frac{\vec{v_s} \cdot \vec{R}}{\|\vec{R}\|^2}\vec{R}, \tag{2.9}$$

where $\|\cdot\|^2$ is the norm operator.

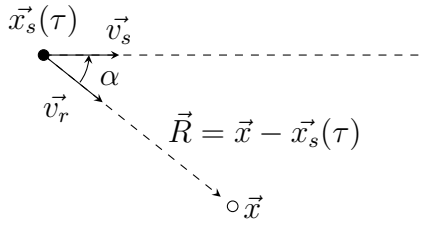


Figure 2.7: A source with position $\vec{x}$ travelling with the speed and direction $\vec{v_s}$. The vector $\vec{R}$ denotes the distance and direction from the source to the receiver at the emission time, τ. The relative source speed in the direction of the receiver is given by $\vec{v_r}$.

A Green's function can be used to determine the resulting sound field including the Doppler effect expressed as

$$p(\vec{x}, t) = \frac{Q(\tau)}{4\pi R(1 - M_r)}, \tag{2.10}$$

where M_r is the relative Mach number of the source and τ is the emission time. The relative Mach number can be interpreted as the source velocity relative to the direction of the receiver, v_r, scaled by the speed of sound in air, c_0. The Doppler factor $(1 - M_r)^{-1}$ can accordingly be expressed as

$$(1 - M_r)^{-1} = \frac{1}{1 - \frac{\vec{R} \cdot \vec{v_s}}{Rc_0}} = \frac{1}{1 - \frac{v_s}{c_0}\cos\alpha}, \tag{2.11}$$

where α is the angle between $\vec{v_s}$ and $\vec{v_r}$ as shown in figure 2.7.

Figure 2.8 shows how the Doppler shift can be modeled in software. In the top subplot, the orange dashed line represents the source trajectory, and the pink vertical lines show the sampled times of the source. The source is moving from left to right with a constant velocity, and as the pink vertical lines are consistent in all subplots, the x-distance in the horizontal axis of the top plot is scaled to the corresponding time axis in the following subplots. In the middle plot, the blue curve represents

the time-varying continuous signal that is sent out by the source while it is moving. To represent the Doppler shift (change in frequency), the continuous signal is interpolated by backtracking the time by $\tau = t - R_0/c$, indicated by the arrow from the first pink dashed line to the first red dot. It looks like the first few samples from time $t = 0.0\,\mathrm{s}$ are aligned with the grid. However, this is not necessarily the case. Moreover, to obtain the new amplitude from the Doppler effect, the interpolated value is multiplied by the Doppler factor $(1 - M_r)^{-1}$. Note that the distance attenuation, $1/4\pi R$, is not modeled here. Each acquired value is written to a new channel at respective times, shown as the green dots in the bottom subplot. The resulting signal has the Doppler effect applied, demonstrated by the black dashed line.

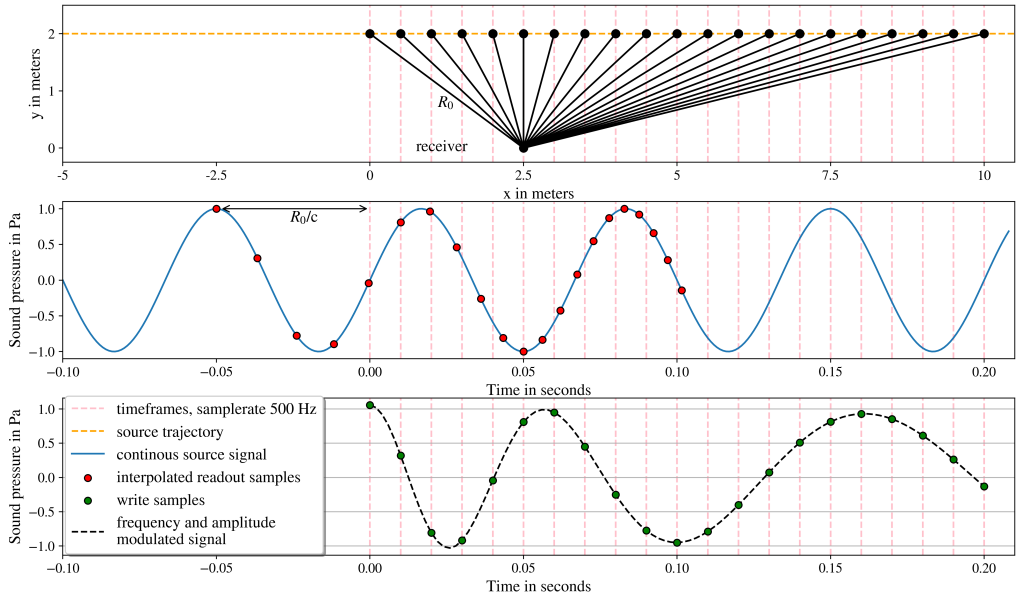


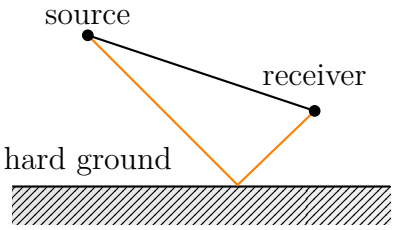
Figure 2.8: The simulated Doppler effect from a moving source and a stationary receiver. The upper plot shows the sampled source positions and respective propagation paths to the receiver. The middle plot shows the continuous signal sent out by the source and the interpolated readout sample points (red dots). The bottom plot shows the resulting signal with the Doppler effect applied.

A drawback of modeling the Doppler effect with this method is that the source position has to be sampled with a very high sampling frequency to avoid aliasing effects (explained in section 2.5.1) because audio signals are often sampled at 44.1 kHz or 48 kHz, which is an overwhelming sampling rate of the source position.

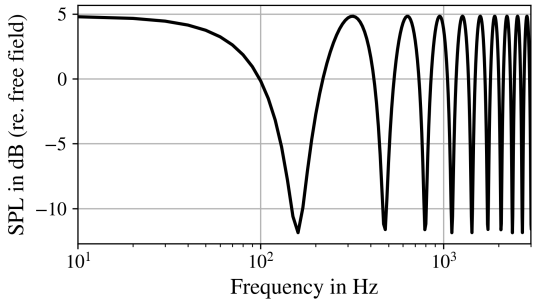
2.4.3 Comb Filter Effects

Comb filter effects appear when two equal signals are combined with a slight relative delay. This causes constructive and destructive interference between the two signals in different frequency bands. Comb filtering is typical when a sound wave is reflected off a hard ground because the reflected wavefronts arrive slightly later than the direct wavefronts. This is illustrated in figure 2.9. Figure 2.9a shows a direct and a reflected

path from a hard ground, and figure 2.9b shows the corresponding sound pressure level (SPL) relative to free-field. The curve has the shape of a comb, hence the name comb filter.



(a) Direct path and ground reflection from a hard ground.



(b) Sound pressure level relative to free-field.

Figure 2.9: The comb filter effect.

2.4.4 Air Absorption

When sound propagates in the air, the air acts as a low-pass filter as compound molecules in the air collide with each other. This is causing some of the energy to dissipate into heat [22]. The international standard ISO9613 [23] provides an analytical method for calculating air attenuation based on meteorological parameters. The attenuation coefficient can be calculated with

$$\alpha = 8.686 f^2 \left(\left[(1.84 \cdot 10^{-11} \left(\frac{p_\mathrm{a}}{p_\mathrm{r}}\right)^{-1} \left(\frac{T}{T_0}\right)^{1/2} \right] + \right.$$
$$+ \left(\frac{T}{T_0}\right)^{-5/2} \cdot \left\{ 0.01275 e^{-2239.1/T} \left[f_\mathrm{rO} + \frac{f^2}{f_\mathrm{rO}} \right]^{-1} + \right.$$
$$\left. \left. + 0.1068 e^{-3352.0/T} \left[f_\mathrm{rN} + \frac{f^2}{f_\mathrm{rN}} \right]^{-1} \right\} \right), \quad (2.12)$$

where f is the frequency in Hz, p_a and p_r is the ambient atmospheric pressure and the reference atmospheric pressure, respectively. The latter is given by $p_\mathrm{r} = 101.325\,\mathrm{kPa}$. The constants T and T_0 are the ambient atmospheric temperature and the reference air temperature in Kelvins, where $T_0 = 293.15\,\mathrm{K}$. Lastly, f_rO and f_rN is the oxygen and nitrogen relaxation frequencies given by

$$f_\mathrm{rO} = \frac{p_\mathrm{a}}{p_\mathrm{r}} \left(24 + 4.04 \cdot 10^4 h \frac{0.02 + h}{0.391 + h} \right) \quad (2.13)$$

and

$$f_{\mathrm{rN}} = \frac{p_{\mathrm{a}}}{p_{\mathrm{r}}} \left(\frac{T}{T_0}\right)^{-1/2} \cdot \left(9 + 280h \exp\left\{-4.170\left[(T/T_0)^{-1/3} - 1\right]\right\}\right), \tag{2.14}$$

where h is the molar concentration of water vapour in percent.

The unit of α is dB/m. Figure 2.10 shows the frequency dependant sound pressure level in dB from a simulation with distance and air attenuation modeled from a pass-by vehicle. The vehicle moves in the positive x-direction and has a constant y-position of 3 m. The receiver is placed in $(0, 0)$m. As seen from the figure, the air attenuation acts as a low pass filter where the slope of the curve is increased with distance relative to the receiver.

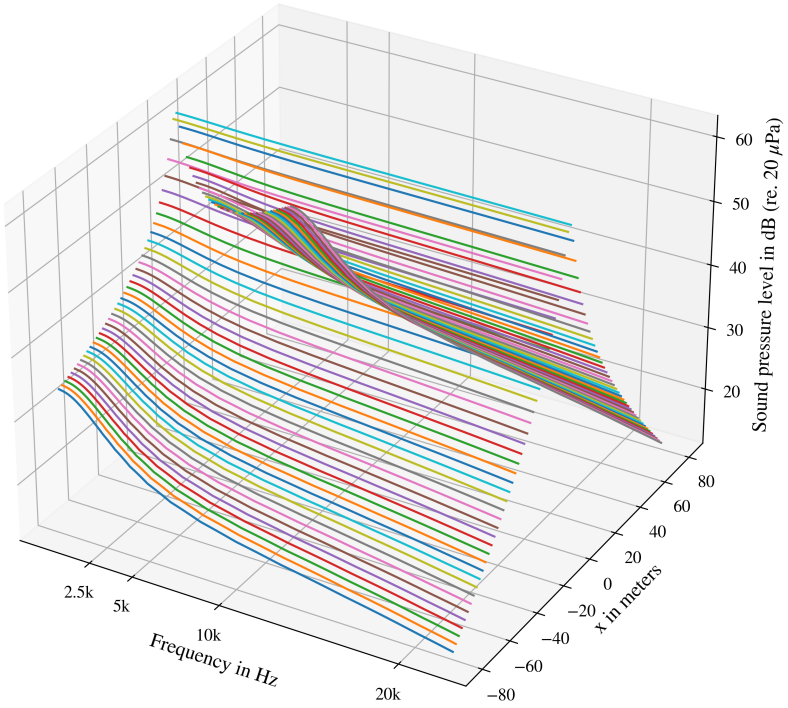


Figure 2.10: The sound pressure level in dB with distance and air attenuation from a pass-by vehicle.

2.5 Digital Signal Processing

Digital signal processing (DSP) is an important part of auralization as it enables the manipulation of audio signals. In the following subsections, some DSP concepts used in auralization are explained.

2.5.1 Aliasing

Aliasing is an artifact that can occur if a signal is sampled too infrequently. This causes a loss of information in the resampled signal. This effect is shown in figure 2.11, where the original signal is resampled with a too-low sampling frequency. The resulting signal has a different frequency than the original signal.

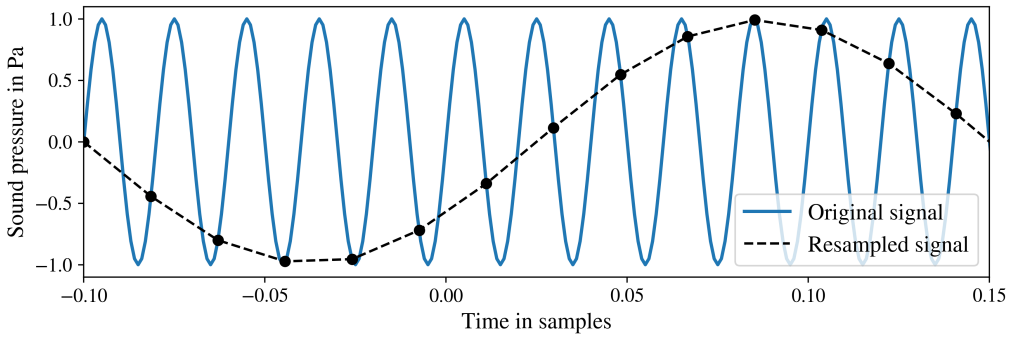


Figure 2.11: Aliasing from resampling with too low sampling frequency.

The Nyquist rule is a theorem to prevent such errors, stating that the resampling frequency should be at least twice the highest frequency represented in the original signal.

2.5.2 Interpolation

In the real world, sound signals are continuous, but in the digital world, signals have to be discretized in samples. In some occasions, it is desirable to get access to values in between adjacent samples. This is the purpose of interpolation. There are several interpolation methods, including linear interpolation, spline interpolation, cubic interpolation, and sinc kernel interpolation, to name a few. The simplest one is the linear interpolator. To find a value $y(x)$ between two adjacent samples, the formula

$$y(x) = y_0 + (x - x_0)\frac{y_1 - y_1}{x_1 - x_0}, \tag{2.15}$$

can be used. In figure 2.12, a linear interpolator is shown.

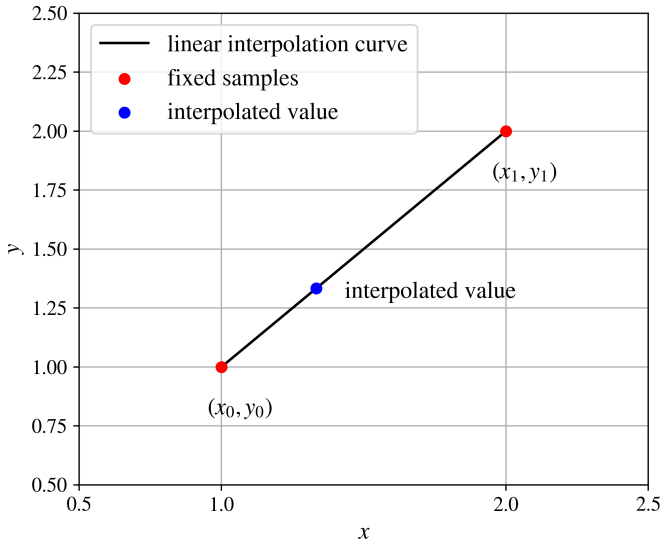


Figure 2.12: Linear interpolation between two samples.

When using interpolation, it is important to proceed with caution. Firstly, interpolation will have a direct influence on the frequency response (FR) of the signal. Interpolation with a fractional delay (such as in figure 2.8) acts as a low-pass filter, attenuating high frequencies. The FR of a white noise signal that has equal power in all frequency bands is shown at the top in figure 2.13. The bottom figure shows the FR after linear interpolation with a varying propagation delay as a fractional delay.

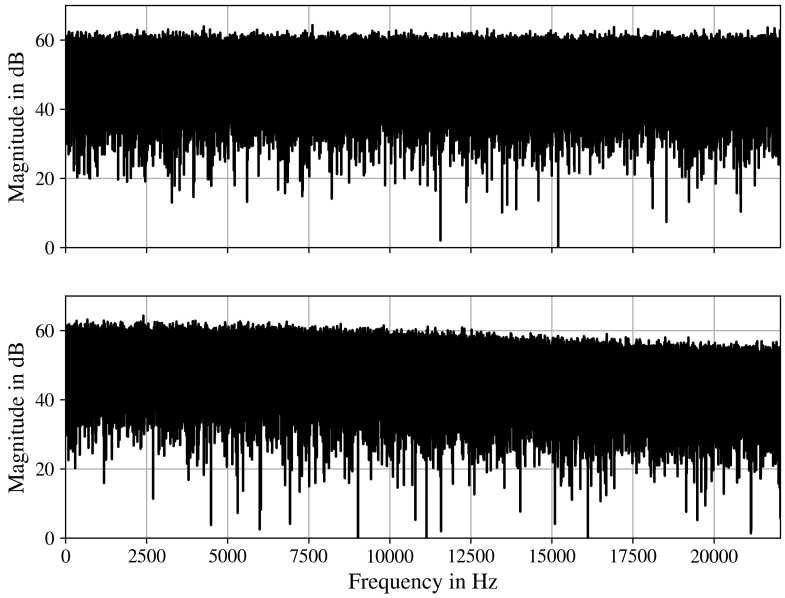
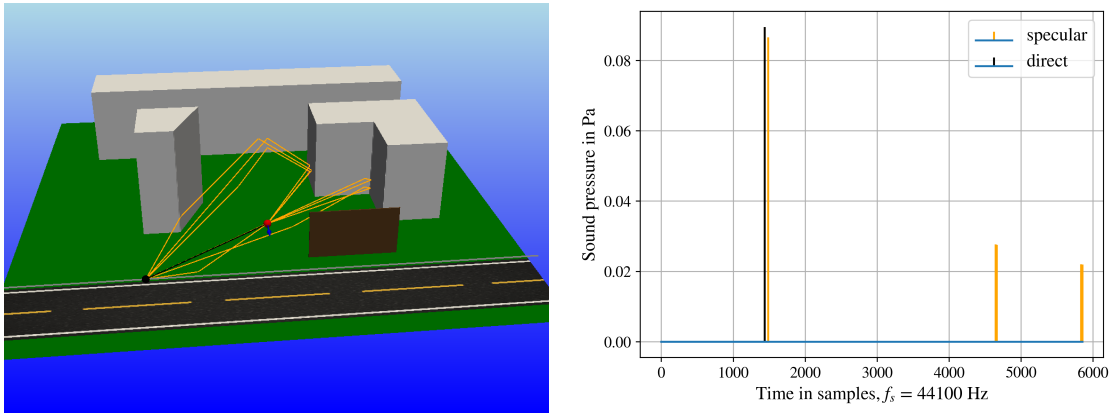


Figure 2.13: Frequency response of a white noise signal before and after linear interpolation with a fractional delay.

Secondly, interpolation can cause aliasing. When resampling with a sampling rate lower than the initial sample rate, the Nyquist rule is possibly violated. To avoid such aliasing, an anti-alias filtering stage should be added at the beginning of the processing chain if the interpolation frequencies violate the Nyquist rule.

2.5.3 Impulse Responses

An impulse response, $h(t)$, describes how a system responds to a unit impulse. Summing up all the contributions from geometrical acoustics and possibly diffraction, and aligning them with the respective time delay and amplitude, yields the discrete-time impulse response, $h[n]$. Figure 2.14a shows the example scene from figure 2.3 with paths from GA added. The black line is the direct path, and the orange lines are specular reflections found from the ISM up to order 3. The respective impulse response from this scene is seen in figure 2.14b.



(a) Example scene with direct sound and specular reflections.

(b) Impulse response.

Figure 2.14: The example scene from figure 2.3 with paths from GA and the corresponding discrete-time impulse response.

2.5.4 Head Related Impulse Responses

The auditory system is dependent on interaural time delays (ITDs) to locate where a sound is coming from. Since sound waves propagate slowly compared to e.g. electromagnetic waves, these ITDs are very important in binaural modeling. For instance, a sound with a slight delay of arrival in the left ear compared to the right ear results in interpreting the sound coming from the right. Head Related Impulse Responses (HRIRs) are two-channel impulse responses that hold information on ITDs and also how the human head shapes the sound, like a filter. The HRIRs are usually measured by placing microphones in the opening of the auditory canal on a human head or a dummy head. Figure 2.15 shows HRIRs for two incident angles, 45 degrees to the left and right for the orientation of the receiver, respectively.

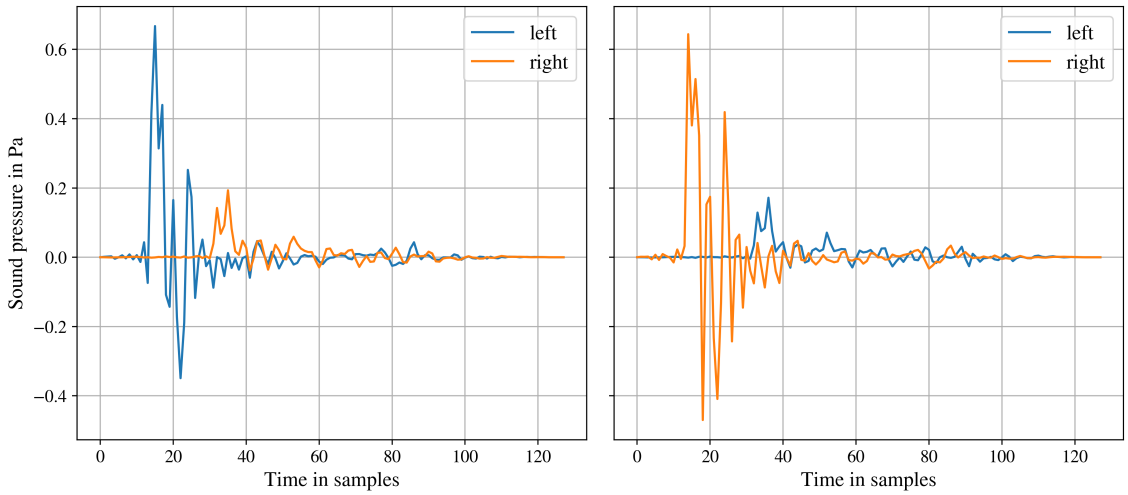


Figure 2.15: HRIRs for two incident wavefront normals, 45 degrees to the left and right from the receiver's neutral axis, in the left and right subplot, respectively. The impulse responses are retrieved from a dataset measured on a Neumann KU100 dummy head with a 2° Gauss-Legendre sampling configuration [24].

2.5.5 Convolution

Convolution is a frequently used method in DSP defined as

$$y(t) = [x * h](t) = \int_{-\infty}^{\infty} x(t) \cdot h(t - \tau) d\tau, \tag{2.16}$$

in the continuous time domain. The equivalent definition in the discrete time domain is given by [25]

$$y[n] = [x * h][n] = \sum_{k=0}^{n} h[k] \cdot x[n - k]. \tag{2.17}$$

Intuitively, a convolution process can be seen as flipping one of the input signals and then multiplying sample by sample with the other input signal. The result of the convolution with the impulse response is a system response to an arbitrary input $x[n]$. If the discrete-time impulse response (like the one in figure 2.14b) is known for a scene environment, convolution can be used to auralize a sound signal $x[n]$ from a stationary source in the receiver position.

2.5.6 Overlap-add and Windowing

To simulate source motion with impulse responses and convolution, an overlap and add method can be used to switch between different impulse responses over time. This commonly includes adding a window function to get the same energy in the overlapping regions between two overlapping blocks. There are a variety of window

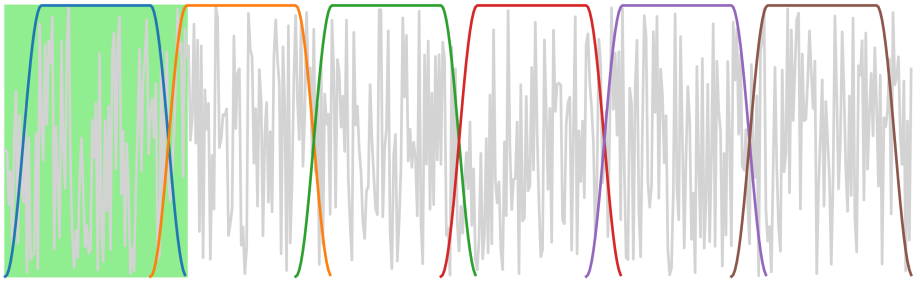


Figure 2.16: Overlap and add method with a Tukey window often referred to as the tapered cosine window. The highlighted green area shows the block size of the first block, and the colored curves are the corresponding windows of each block.

functions that can be used depending on the size of the overlap. Figure 2.16 shows an overlap and add method with a Tukey window, also known as a tapered cosine window. The green box in the left position defines the first signal block. The samples in the block marked as light gray in the figure are multiplied elementwise with the windowing function shown as the blue curve for the first block. This process is repeated for the following blocks, which get accumulated into the total output signal. The resulting signal will have the same energy in the overlapping regions causing a smooth transition between segment effects, e.g. different impulse response filters.

2.5.7 Filtering with FFT/IFFT

The Fast Fourier Transform (FFT) and the Inverse Fast Fourier Transform (IFFT) are computer algorithms to calculate the Fourier transform efficiently. The Fourier Transform takes a signal $x(t)$ in the time domain and transforms it into the frequency domain. In mathematical notation, it can be written as

$$\mathscr{F}\{x(t)\} = X(f), \tag{2.18}$$

where $X(f)$ denotes the frequency representation of the signal. After the signal has been transformed to the frequency domain, a filter can be applied to the signal by multiplying $X(f)$ with the filter's system response, $H(f)$ as

$$Y(f) = X(f)H(f), \tag{2.19}$$

where $Y(f)$ is the filtered signal in the frequency domain. Convolution can be seen in the time domain as the equivalent method to multiplication in the frequency domain. The Inverse Fourier Transform can be used to transfer the signal back from the frequency domain to the time domain as

$$\mathscr{F}^{-1}\{Y(f)\} = y(t), \tag{2.20}$$

where $y(t)$ is the filtered signal in the time domain. The Fourier Transform can be applied to both continuous and discrete signals, but in DSP, the discrete-time Fourier Transform (DFT) is the most relevant. The definition of the DFT is given by

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{j2\pi \frac{n}{N}}, \qquad k = 1, 2, ..., N, \tag{2.21}$$

where n is the index of the sample in the sampled signal x, N is the number of samples in x, k is the Fourier component number, and j is the imaginary unit [26].

There are three things to keep in mind when applying the DFT to a signal. Firstly, the DFT assumes a signal to be periodic with period N meaning $x[n] = x[n+N], \forall n$. However, this is quite seldom the case in reality, as signals will most likely have discontinuities at the edges. The edge discontinuities can be removed by windowing, as discussed in the previous section. Secondly, the transformed spectrum $X[k]$ is periodic and double-sided. Thirdly, $X[k]$ has to be scaled to obtain the amplitude spectra. This can be done by multiplying $1/N_f$ with $X[k]$, where N_f is the number of samples in the signal.

2.5.8 Digital Delay Lines

A delay line is simply a buffering unit that stores samples before they reach the output of the system. Such a delay line can be seen in figure 2.17, where $x[n]$ is the input and $y[n]$ is the output. The length of the delay line in samples is determined by M [27]. If M can be changed during operation, the delay line is called a variable delay line.

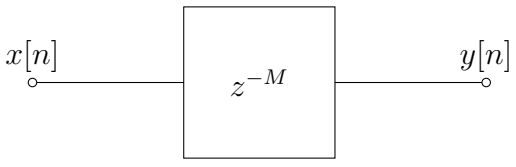


Figure 2.17: A digital delay line with a delay of M samples.

The output of a delay line is determined by

$$y[n] = x[n - M], \tag{2.22}$$

where $x[n] = 0$ for $n < 0$. The delay line in figure 2.17 is time-discrete, but when simulating wave propagation with delay lines, it is likely that the output should have access to values between samples. If that is the case, it is called a fractional delay line, and it can be accomplished by interpolation as described in section 2.5.2 [17]. Furthermore, a delay line can use signal blocks instead of single samples as input.

2.5.9 Multiple Doppler Simulation with Delay Lines

Section 2.4.2 discussed how to model the Doppler effect in software for the direct path. After applying geometrical acoustics and diffraction, acoustic rays will be emitted from the source in multiple directions. The various paths will equivalently have individual Doppler effects. This can be seen in figure 2.18, where the circles around the moving source are analogous to the monopole sound field from figure 2.6, only for a moving source. The black line from the source to the receiver indicates the direct path, and the orange line is a reflected path. From the circles illustrating the sound field, it is apparent that the two have individual Doppler shifts. This means that when modeling multiple Doppler effects, the Doppler shift should always be evaluated from the point at the end of the first path segment from the source. The same goes for diffracted paths. However, the full path length is still needed to calculate the Doppler factor correctly. In figure 2.18, $\vec{R}$ from equation 2.11 would have the direction from the source to the Doppler evaluation point, and R would be the full path length.

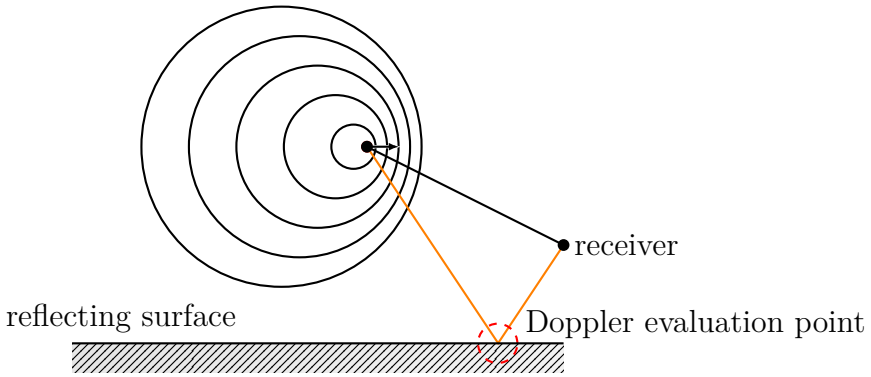


Figure 2.18: A direct and reflected path from a moving source with a stationary receiver. The circles around the source indicate the sound field from the moving source with the Doppler effect. The direct and reflected path has individual Doppler effects.

Previous literature has modeled multiple Doppler effects with the use of delay lines. In 2002 Smith et al. used a delay line to simulate the Doppler effects from a rotational loudspeaker called a Leslie [28]. Similarly, in 2018, Steinen and Vorländer proposed a single-input multiple-output variable delay line (SIMO VDL) to simulate multiple Doppler effects in real-time auralization systems [29]. The SIMO VDL is of specific interest as it allows for the treatment of unique transmission paths separately. It has a single write cursor and multiple interpolating read cursors, one for

each transmission path. Figure 2.19 shows a basic SIMO VDL read/write operation for a single transmission path in a given timeframe.

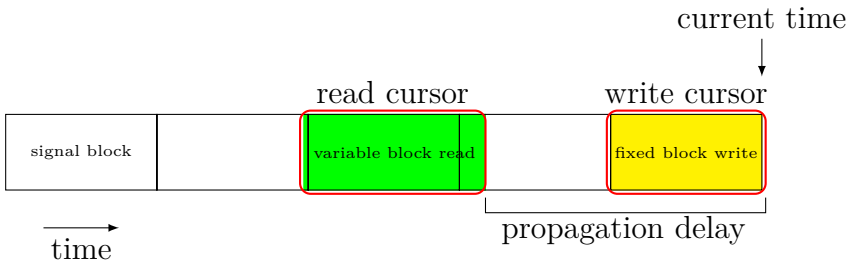


Figure 2.19: Basic read/write operation for one single transmission path from the SIMO VDL.

To use a SIMO VDL, it is necessary to eliminate the constraint of linear and time-invariant filtering, i.e. overlap and add with convolved signals from using impulse responses are not possible. The only convolution process used is when applying the HRIRs. Further information on the SIMO VDL can be found in [29, 17], and a complete open-source implementation is available at www.virtualacoustics.org [1].

3

Methods

This chapter reviews various implementations of the auralization system developed during the course of this thesis work. The final part of the chapter describes how ambient sound was recorded and details about the conducted listening experiment.

3.1 Software Implementations

The software part of the system is implemented in Python, a versatile, object-oriented programming language. Python is an interpreted language, meaning the code is compiled at run-time, which makes Python slower compared to programming languages such as e.g. C++ or Java. However, Python has a simple syntax and numerous external libraries and frameworks that can be integrated easily. The simulations conducted for this thesis were performed offline. Therefore, the computational cost was not the primary limitation, unlike real-time systems. The complete software implementation can be found at https://github.com/gjesshopper/urban_auralization.

3.1.1 Scene Visualization

The 3D visualization of the scenes (see figure 2.3) was implemented with Pyvista (version 0.37.0), an easy-to-use open-source library for data visualization with Python bindings to C++ for faster performance [30]. Pyvista has been used in previous research and has potential within the field of sound planning, as it's capable of processing geospatial meshes. In this implementation, only simple polyhedral objects with rigid faces were used as proof of concept.

3.1.2 Specular Reflections

Specularly reflected paths are acquired from the ISM explained in section 2.2.2. The ISM is a recursive method and finds complete paths for reflecting objects from the source to the receiver by traversing image sources starting from the receiver back to the source. Therefore image sources had to be represented as class objects with attributes such as the source/image source it originated from.

An impedance model for the reflective surfaces was not implemented, and the impedance was assumed to be infinite, yielding a reflection factor of 1.

3.1.3 Diffraction

The primitive implemented pathfinding algorithm for diffracted paths is very similar to Maekawa's rubber band model. To find such paths, the convex hull of all features and all combinations of features is validated. The resulting paths are then checked for obstruction, just as in GA. Since the convex hull is used in 2D (top-down projection with the corners of the features as vertices), the heights of the points have to be calculated separately. The heights of each point in a path are identified by interpolating the extended linear curve from source to receiver to satisfy Fermat's principle. Figure 3.1 shows a top-down and side view of a diffracted path. The top-down view projects the true horizontal length of the path segments. The only two points that have a known height are the source and the receiver, so in the side view, the paths are projected as an extended straight line. This line is interpolated to find the respective heights at the vertices, marked as red dashed circles.

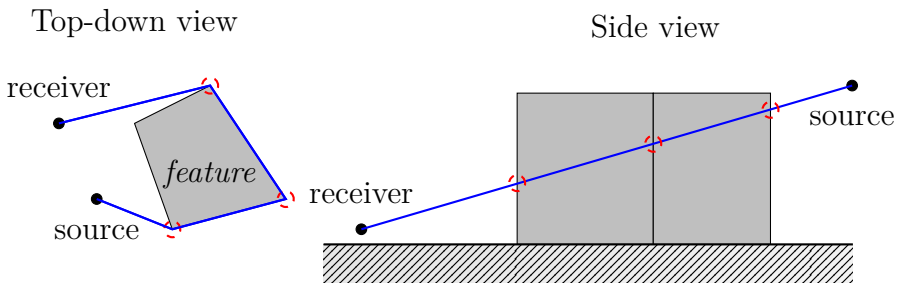


Figure 3.1: Top-down and extended side view for a path diffracted around a feature. The heights of the red dashed circles are found by linear interpolation.

The diffraction pathfinder can similarly find diffracted paths that go above features. The same method is used as in figure 3.1, but the top-down view is the side view and vice versa.

The Harmonoise model was used to calculate the attenuation from diffracted paths as it was the simplest to implement of the diffraction models discussed in section 2.3.

3.1.4 Digital Processing System

The proposed implementation of a DSP system is shown in figure 3.2. The design is strongly based on the SIMO VDL from section 2.5.9. Following the flow in the diagram, starting from the top left, geometrical paths are precomputed for all sampled source positions along the source trajectory in the scene. These paths are obtained from GA and the diffraction pathfinder algorithm discussed in the previous section. In the next step, all geometrical paths are sorted into path groups by a sorting algorithm to ensure that paths with similar properties are processed in order. For instance, the direct sound from the source throughout the entire duration of the simulation makes up one path group. Moreover, reflected paths that have the same order and reflection planes are another path group. The point of sorting the paths

into path groups is to prevent artifacts in the output audio in transition regions between signal blocks. When the paths are processed in groups, the transition region between two signal blocks can be guaranteed to line up perfectly. One by one, a path group is sent into the processing chain, where a single path is extracted and processed separately. The coefficients for filters that account for reflection, diffraction, and air absorption are stored temporarily with a timestamp for future use. A time counter keeps track of where in the timeline the current path belongs. To read out a signal block, the time is rewound from the current time held in the time counter with the propagation delay of the respective path. The resultant time marks the end point of the readout block. The size of the readout block is determined by the Doppler shift. After readout, the signal is interpolated with a fixed number of samples. The interpolated signal block is written from the time indicated by the time counter. The time counter is incremented with Δt (which indicates the fixed block time), so it points to the time of the next path in the path group. When all paths in the given path group have been processed, filters and HRIRs are applied with an overlap and add method before the combined signal is added to an accumulative bus. This process is repeated for all path groups.

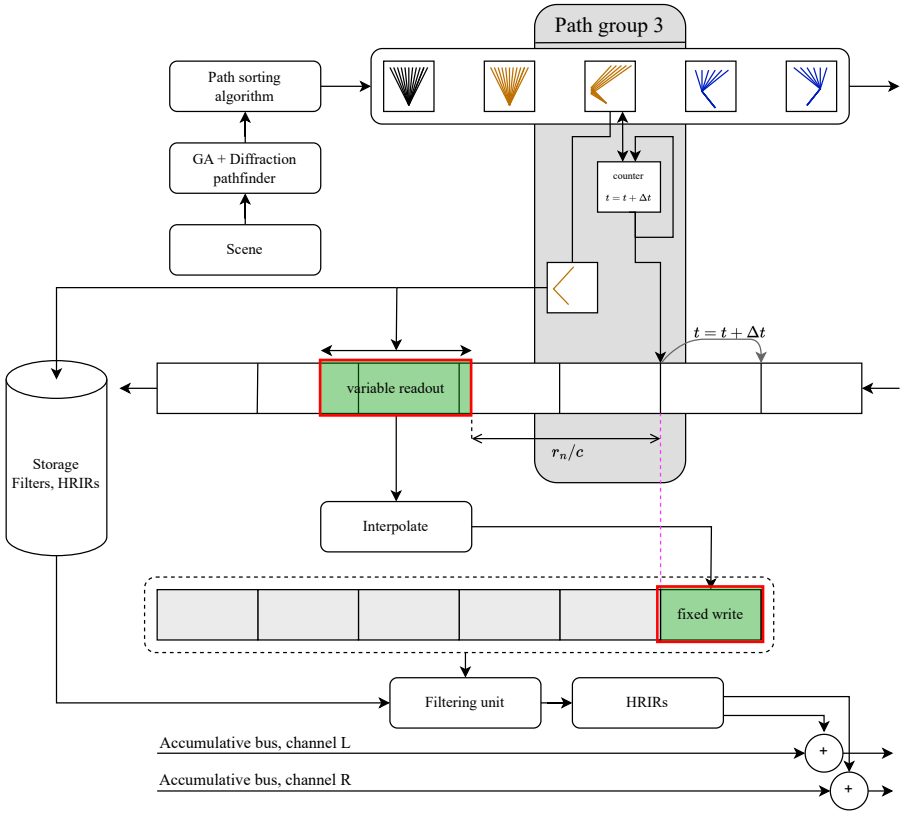


Figure 3.2: The implemented DSP system.

A sample analysis of the read/write operations from figure 3.2 is illustrated in figure 3.3. The path group being processed is the direct paths. In the top subfigure, the sampled source positions are visualized for time t_0 to t_4. The direct paths R_0 to R_4 from the source positions to the stationary receiver are marked as black lines. In

the bottom subfigure, the blue dots mark the sample points of the emitted source signal. Note that the sample rate of the source position and the source signal is not the same. In fact, the ratio between these sampling rates determines the fixed write block size of the DSP system. The green field in the lower plot shows the readout block region, where the end of the readout is found by rewinding the propagation delay from the current time in the time counter (in this case, t_0) as before. The size of the readout is again determined by the Doppler shift, and the readout block is resampled by interpolating at equidistant points with a fixed number of samples. The interpolated block is finally written to the output at the current time, indicated by the gray region.

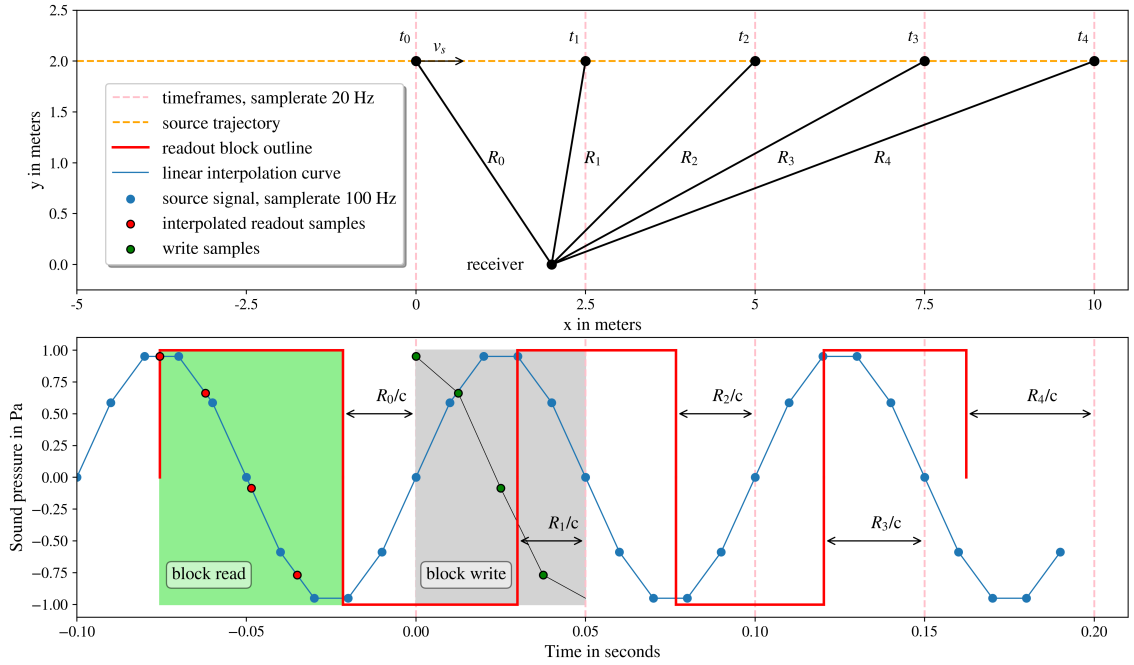


Figure 3.3: A sample analysis of the read/write operations in the proposed DSP system.

3.2 Ambient Sound Recordings

An attempt to further improve the plausibility of the auralizations was by integrating recorded ambient sound i.e. environmental background sound. The ambient sound was recorded at different locations around the Chalmers Johanneberg campus in Gothenburg. Table 3.1 lists the equipment used for these recordings. The portable 4-channel recorder, SQobold, had a built-in calibration wizard, which was used to calibrate the free-field microphone, and the binaural headphones, which were equipped with microphones on the outside of each earcup. A part of the measurement setup is illustrated in figure 3.4. As seen in this picture, the measurement rig had a camera mounted to it to keep track of which recordings belonged to what location. Furthermore, an equipped GPS antenna stored the accurate geographical coordinates for each recording. Typical locations for the recordings included areas

without too much traffic or next to roads but when no vehicles were in the immediate area.

To apply the recordings to the auralizations, the recordings were organized based on the location to match the virtual environment (scene) to the best ability.

Figure 3.4: Setup for the ambiance recordings.

Table 3.1: Equipment used for the ambiance recordings.

Equipment	Model/Type	Manufacturer
4-channel data acquisition system	SQobold	Head Acoustics
Binaural headset/microphone	BHSII	Head Acoustics
Camera	1637175	Microsoft
Free-field microphone	46AE 1/2" CCP	GRAS
GPS Antenna	GCA I.1	Head Acoustics

3.3 Listening Experiment

To validate the auralizations produced by the implemented system, a listening experiment was conducted. The listening experiment consisted of 3 parts, and in total, 11 people participated. A set of recordings $\mathcal{R} = \{r_1, r_2, r_3, r_4\}$ from the Swedish LISTEN-project [31] were used to render the auralizations used in the different parts of the listening experiment. These recordings were pass-by recordings from light vehicles post-processed using a "depropagation" method so that they could be considered stationary sound sources. In each part of the listening experiment, the rendering process was customized to assess different attributes in the auralizations. The single Doppler auralizations were rendered by sending the recordings in $\mathcal{R}$ through the implemented DSP system without a variable-sized readout of the input but instead a fixed block readout. The Doppler effect was applied in the last step of the processing chain.

3.3.1 Part 1 – Paired Comparison

In the first part of the listening experiment, the listeners were given two stimuli, one with a single Doppler effect and the other with multiple Doppler effects. The listeners were asked to rate the stimulus perceived as more plausible. The order of the stimuli was randomized, and each stimuli pair was presented twice, the second time in reversed order. This was done to eliminate the test order bias. A single test scene was set up for this part of the listening experiment, where the sounds in $\mathcal{R}$ were used to render the stimuli. The test scene is illustrated in figure 3.5. The sampling rate of the source position was 20 Hz.

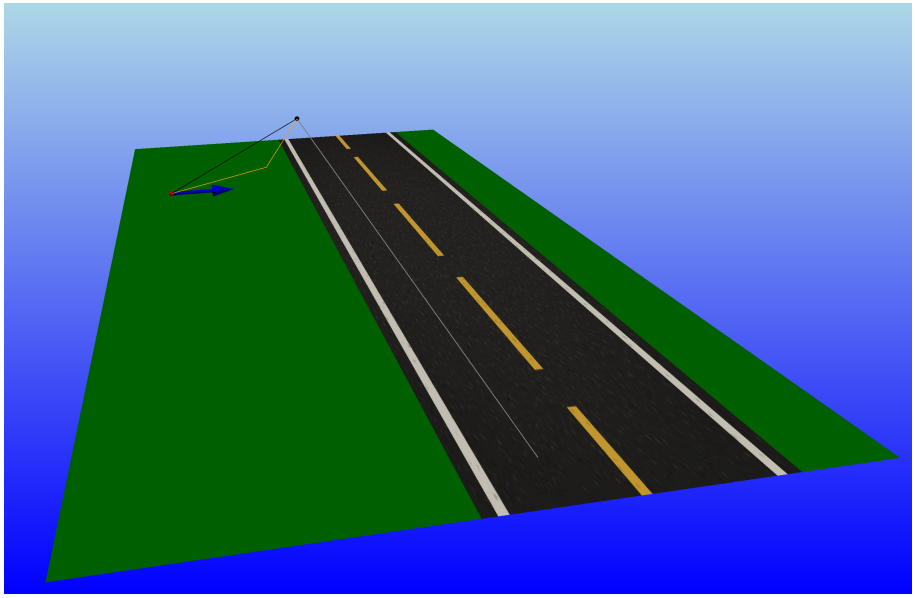


Figure 3.5: The test scene for part 1 (and part 2) in the listening experiment.

3.3.2 Part 2 – Blocksize

The second part of the listening test aimed to find the minimal reasonable sampling rate of the source position. As the sampling rate of the recordings in $\mathcal{R}$ was 44.1 kHz, the source position had to be sampled at a much lower sampling rate, and the simulation was dependent on using signal blocks (see figure 3.2 and 3.3). In this part of the listening experiment, a paired comparison with a reference sound, the listeners were asked to pick one of two sounds that they thought sounded like the reference sound. The reference was equal to one of the two stimuli. The test scene from part 1 was used, but an additional test scene was made to check the minimum block size for the scene with an occluded direct path. This test scene is shown in figure 3.6.

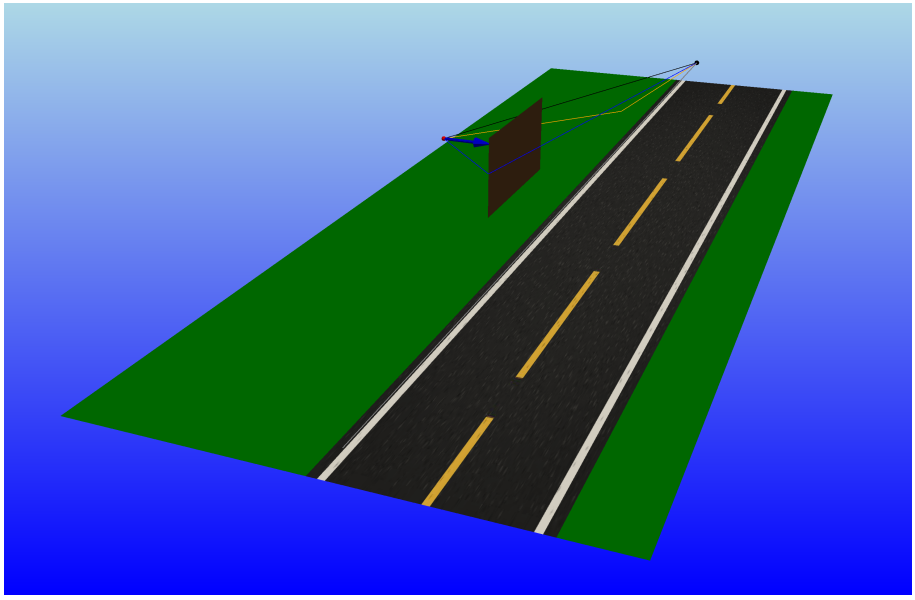


Figure 3.6: Additional test scene for part 2 in the listening experiment.

3.3.3 Part 3 – Ambient Sound

The final part of the listening experiment aimed to see if ambient sound increased the perceived plausibility of the auralizations. The ambiance recordings from section 3.2 were applied to the renderings with multiple Doppler effects from part 1, but in this part, the listeners were asked to rate the plausibility on a scale from 1 to 9, where 1 was *not plausible at all*, and 9 was *very plausible*. Note that the listeners were given both renderings, with and without ambient sound.

3.3.4 Setup

Table 3.2 lists the equipment used for the conducted listening experiment. All listeners were equipped with a pair of calibrated headphones, and the listening experiment was held in a controlled, quiet environment with noise screens between participants not to cause any disturbances. The listening test was executed in SQala, a part of the software Artemis from Head Acoustics used for jury testing [32]. SQala can play back binaural recordings accurately.

Table 3.2: Equipment used for the listening experiment.

Equipment	Model/Type	Manufacturer
Headphones	HD IV.2	Sennheiser
USB adapter for accurate playback with HD IV.2	BPU bundle	Head Acoustics
Computer with Artemis/SQala	Any	Any
Acoustic noise screens	Any	Any

4

Results

This chapter reviews the results from the listening experiment and some of the software-based implementations.

4.1 Listening experiment

Spectrograms for the stimuli used in part 1 and part 3 in the listening experiment are depicted in figure 4.1. The spectrograms show the stimuli without the ambient sound. The first row demonstrates the stimuli with a single Doppler effect, and the second rows show multiple Doppler stimuli.

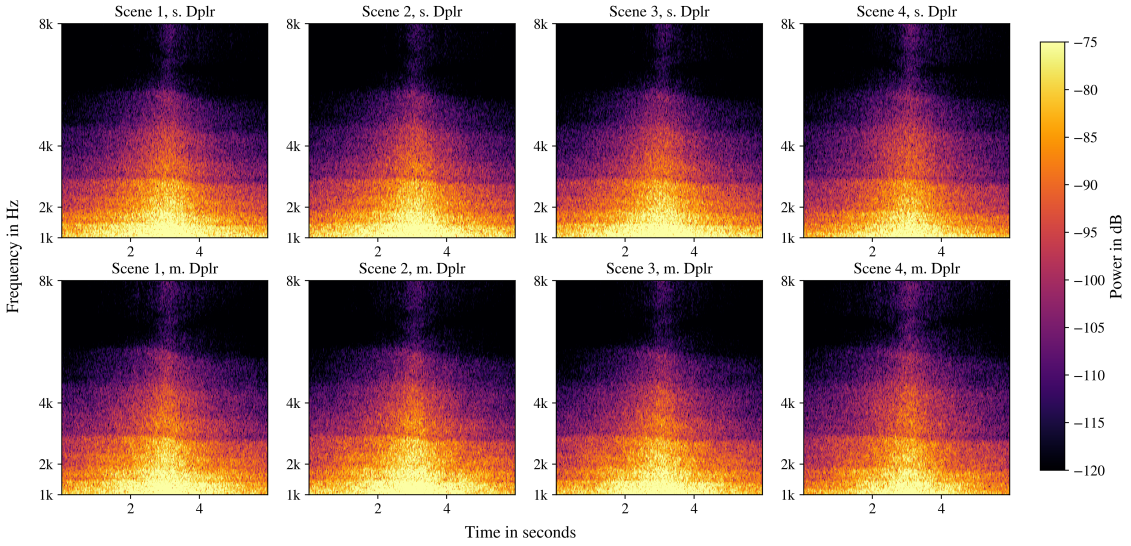


Figure 4.1: Spectrograms used in the first and third part of the listening experiment, without ambient sound.

4.1.1 Paired Comparison

Figure 4.2 shows the results from part 1 of the listening experiment. The green bars in the plot represent the percentage of participants who voted for the multi-

ple Doppler stimuli as more plausible. The red bars represent the percentage of participants who voted for the single-Doppler stimuli as more plausible.

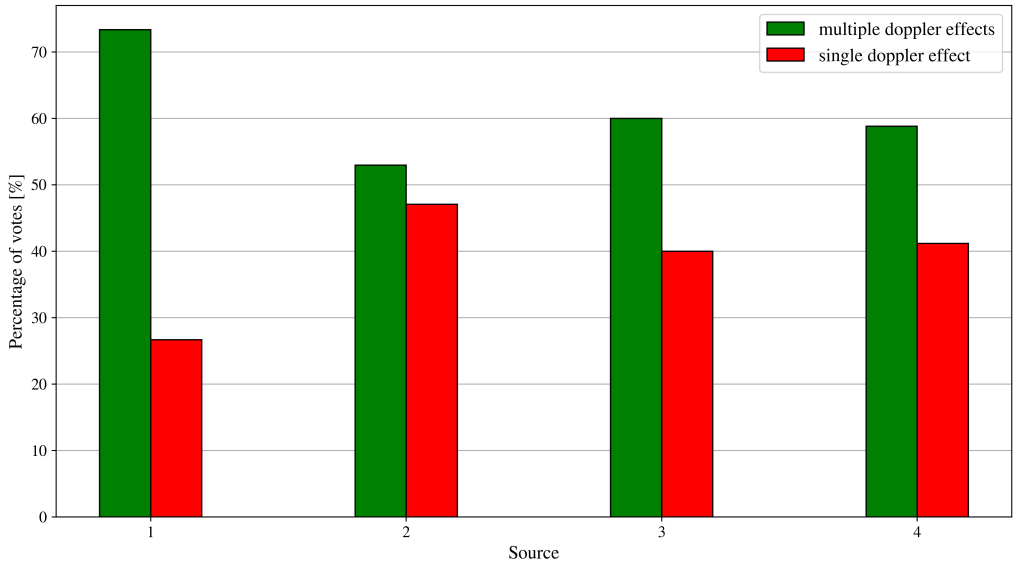


Figure 4.2: Paired comparison between single and Multiple Doppler stimuli.

Fischer's Exact Test

A Fischer's exact test was carried out on the results from figure 4.2. The hypotheses was formulated as

$$H_0\text{: No difference in preferred stimuli} \tag{4.1}$$
$$H_1\text{: Difference in preferred stimuli,} \tag{4.2}$$

where the alternative hypothesis, H_1, was the one attempted to prove with a significance level of 0.05. The contingency matrix and the results from Fischer's exact test are shown in table 4.1 and 4.2, respectively.

Table 4.1: Contingency matrix for Fischer's exact test.

	Single Doppler effect	Multiple Doppler effects
Preferred stimuli	25	39
Non-preferred stimuli	39	25

Table 4.2: Results from the Fischer's exact test.

p-value	0.0004
Odds ratio	0.247

4.1.2 Blocksize

The results from part 2 of the listening experiment are shown in figure 4.3. The green bars show the percentage of participants that guessed the reference correctly. Similarly, the red bars show the percentage of participants that guessed the reference incorrectly. The left and right plots show the results from the scenes depicted in figure 3.5 and 3.6, respectively.

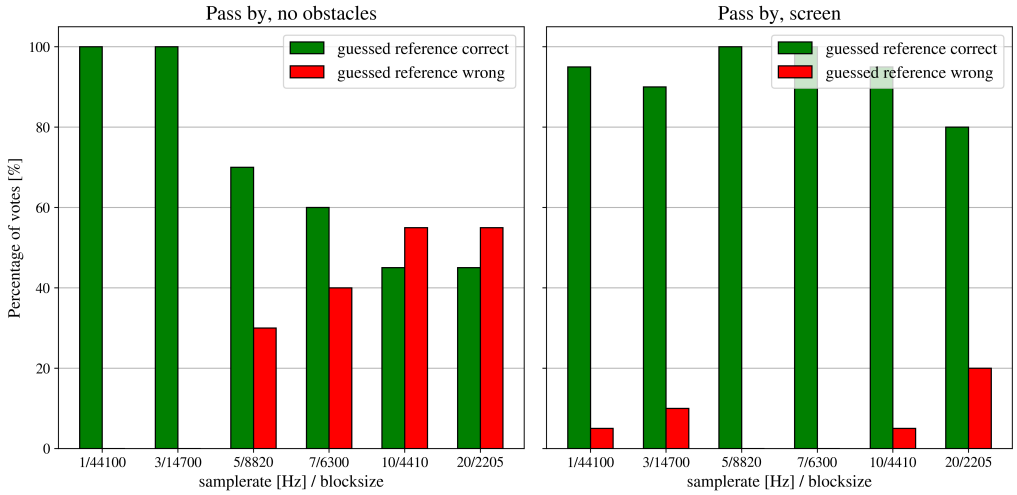


Figure 4.3: The result from part 2 of the listening experiment. The preferred minimum sample rate of the source position was tested for two different test scenes, where the second scene had a diffracting screen.

4.1.3 Ambient Sound

Finally, the mean ratings from part 3 of the listening experiment can be seen in figure 4.4. The mean ratings and 95% confidence intervals for the stimuli with and without ambient sound are shown as blue and green bars, respectively.

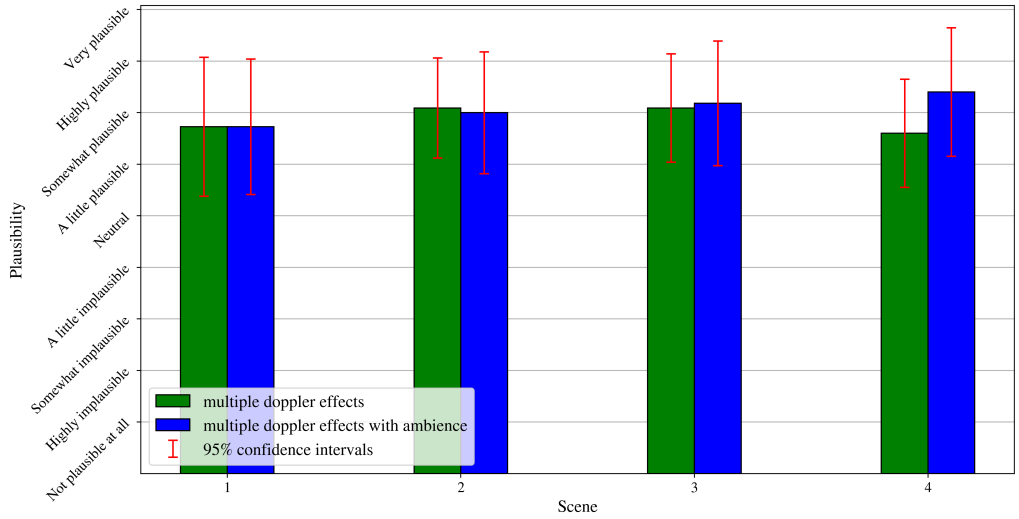


Figure 4.4: Simoultaneous category judgment of perceived plausibility for non-ambient and ambient stimuli.

4.2 ISM Validation

The computational performance of the ISM implementation was tested on the example scene in figure 2.3 with a fixed source and receiver position. The example scene had 25 reflection planes and the computation time for orders 1-4 are shown in table 4.3 below. The 2nd and 3rd columns show the number of constructed image sources (IS) and the number of visible IS, respectively.

Table 4.3: ISM computation time validation on the example scene from figure 2.3.

Order	No. constructed IS	No. visible IS	Computation time
1	25	2	45 ms
2	650	6	129 ms
3	16275	21	2.41 s
4	406900	21	54.95 s

4.3 Frequency Spectra

Figure 4.5 shows the frequency spectra for single and multiple Doppler stimuli from the example scene, where the source signal corresponded to recording r_3. The light gray noisy signal shows the averaged frequency spectrum without smoothing. The left and right channels for the single and multiple Doppler stimuli are shown individually with smoothing as the red/blue and green/yellow curves, respectively.

Similarly, the spectra for the stimuli used in the listening experiment (the two leftmost spectrograms in figure 4.1) for the test scene in figure 3.5 are shown in figure 4.6.

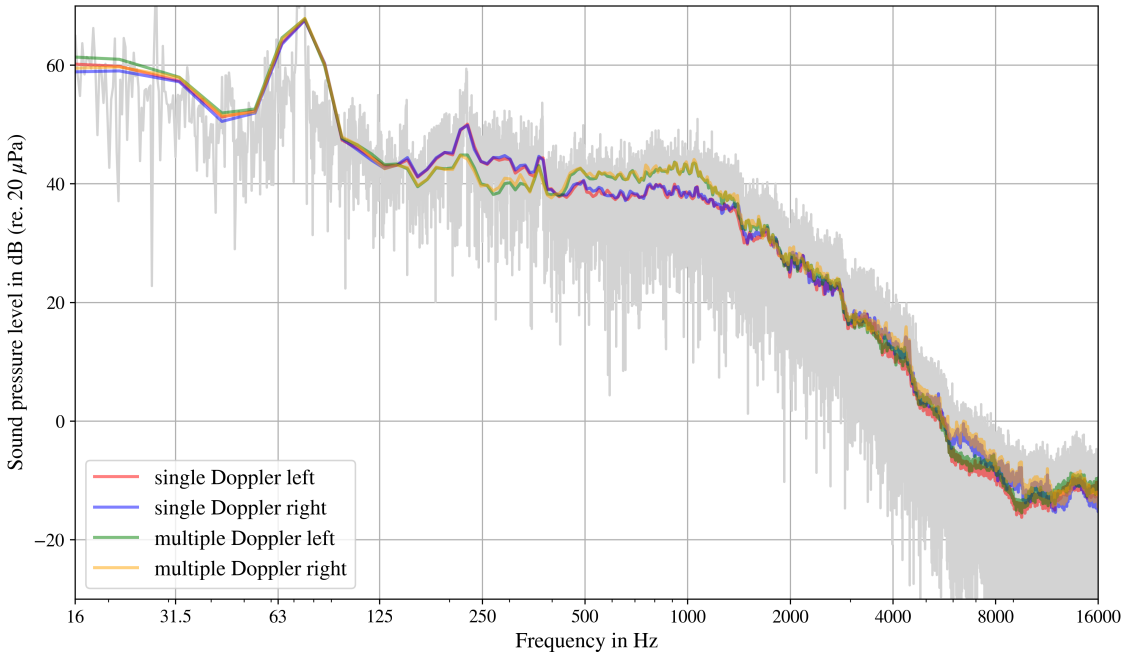


Figure 4.5: Frequency spectra for single and multiple Doppler stimuli for the example scene in figure 2.3, but without the screen.

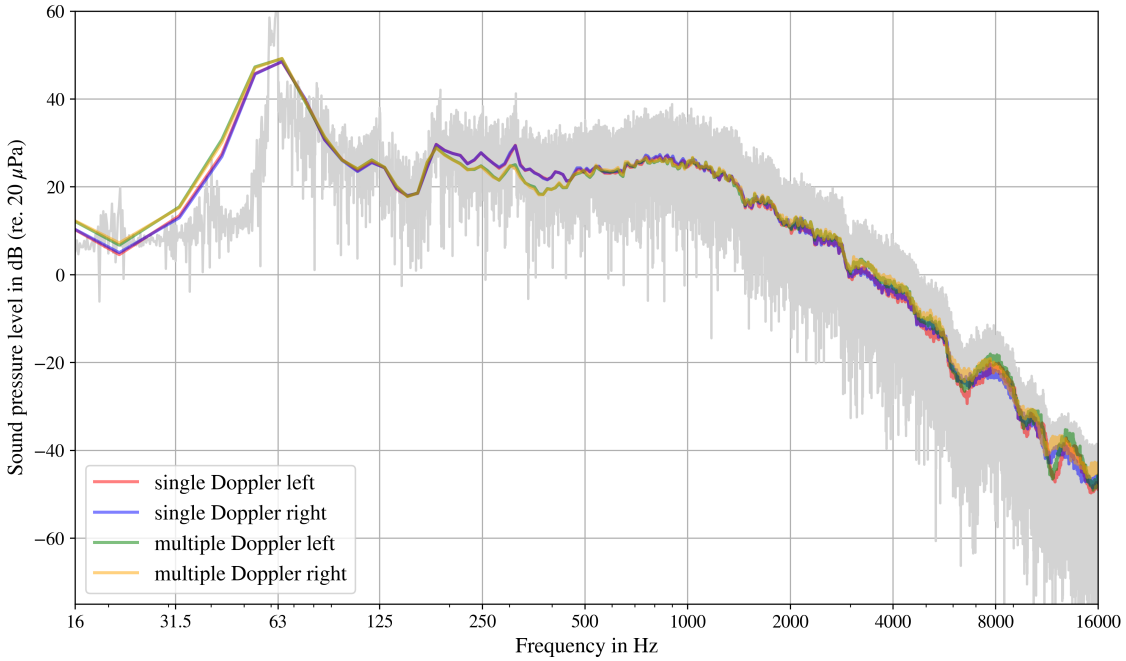


Figure 4.6: Frequency spectra for the single and multiple Doppler stimuli generated from recording r_1 for test scene 1 depicted in figure 3.5.

The spectra for the rest of the stimuli in the spectrograms can be found in appendix A.

5

Discussion

5.1 Listening Experiment

All the stimuli used in the listening experiments depicted in figure 4.1 had a recurring horizontal line at approximately 3 kHz. This line clearly visualized the Doppler shift as it started to sink mid-time in the spectrograms. The power spectral density was also highest at this point when the source was directly in front of the receiver. All stimuli looked very similar with the bright eye, and the only visual difference between the Multiple and single Doppler stimuli was minor spectral deviations in the high-frequency (HF) areas. Since the single Doppler stimuli had the Doppler effect applied from a linear interpolator with a fractional delay, this would attenuate some of the HF content in the frequency response, and this could be the possible explanation. When these stimuli were played back as audio files, the differences became more apparent than the spectrograms could reveal.

The result from the paired comparison in part 1 of the listening test depicted in figure 4.2 revealed that the Multiple Doppler stimuli generally had higher plausibility ratings than the single Doppler stimuli. To determine if there was sufficient evidence to support these observations, a Fischer's exact test was carried out on the data. The Fischer's test was preferred as it's an alternative to the student t-test but intended for small datasets. The results from the Fischer's exact test showed a p-value of 0.0004. Since this value was below the set significance level of 0.05, the H_0 hypothesis could be rejected. Furthermore, the odds ratio showed that participants were more likely to prefer multiple Doppler stimuli over single Doppler stimuli. That said, considering the small sample size, further validation may be necessary.

The data presented in figure 4.3 from part 2 of the listening test revealed that for the first scene (left plot), all listeners could guess the reference up to sampling rates of the source position of 3 Hz. Already at 5 Hz, the answers started to deviate, indicating that low sampling rates seemed sufficient for simple scenes. The results from the stimuli in the second scene (right plot) showed that almost all participants could more or less guess the reference correctly up to about 10 Hz. The sampling rate should have been further increased to find a more exact estimate of where the crossover region is. But the result was clear that more complex geometric scenes seem to require higher sampling rates to follow the sound field change in

the transition from non-shadow zones to shadow zones and vice versa.

The results from the comparison between ambient and non-ambient stimuli in figure 4.4 indicated no immediate difference between perceived plausibility for ambient and non-ambient stimuli. In addition, the 95% confidence intervals had a large span, undermining the certainty of the estimated mean of the perceived plausibility. Moreover, no statistical test was found suitable to further investigate the coherence between the two due to the small sample size, so in conclusion, the perceived plausibility for ambient and non-ambient stimuli needs further validation.

5.2 ISM Validation

The performance of the ISM from table 4.3 showed that the computation time grew immensely with the ISM order, as expected from equation 2.2. The 3rd order ISM found 21 visible image sources in 2.41 s. Moreover, ISM order 4 was completed in 54.95 s but could not find any additional visible image sources compared to order 3 for this particular scene. As a result, specular reflections with order 3 seem like the highest reasonable order of the deterministic ISM in urban scenes, as higher orders cannot find many additional visible image sources. Additionally, for higher orders in urban scenes, a lot of computation time is wasted on evaluating image sources that most likely are not visible. As mentioned in section 2.2.2, beam tracing methods can accelerate the ISM drastically, but for offline simulations where computation time is not the main limitation, the deterministic ISM proved to work well.

5.3 Frequency Spectra

The frequency spectra depicted in figure 4.5 from the generated stimuli for the example scene indicated a significant difference in the mid-frequency bands from approximately 125 Hz to 1.5 kHz. The multiple Doppler stimulus had approximately 3 dB higher SPL around 1 kHz in comparison to the single Doppler stimulus. Since the single Doppler stimulus was generated using linear interpolation with a fractional delay line, this could possibly attenuate certain frequencies and explain the deviations seen in the plot. The plot in figure 4.6 indicated no significant differences except for minor deviations around 250 Hz. An explanation for why the spectra from the example scene showed a bigger difference in SPL than the spectra from the stimuli used in the listening experiment could be that the example scene had a lot more reflected paths. As the reflections for the multiple Doppler stimulus had unique Doppler effects, this would cause the respective signal block for each path to be interpolated differently, causing a wider spread in frequency components.

As discussed, the spectrograms in figure 4.1 showed a little higher spectral density for the multiple Doppler stimuli. This is especially seen from 4 kHz and above. However, from the frequency spectra, there are no visual differences from the single to the multiple Doppler stimuli in the corresponding frequency band. However, the spectra from the multiple Doppler stimuli used in the listening experiment had

a slightly increased sound pressure level between approximately 2 kHz and 5 kHz, which again could be explained by linear interpolation with a fractional delay.

5.4 Diffraction Model

The Harmonoise diffraction model gave reliable results for the produced auralizations in the thesis. However, a drawback of this model was that it could not combine problems with reflection and diffraction. According to [17], the combination of specular reflections and diffraction is highly relevant for urban auralization problems for generally smooth transitions in the auralized sounds from dynamic environments. This suggests the use of other models, such as the UTD. An argument for integrating the UTD is that it incorporates well with the ISM. Additionally, since the UTD has been shown to be able to be combined with beam tracing methods, it could help accelerate the ISM.

5.5 Digital Processing System

The implemented design of the digital system depicted in figure 3.2 was proved to work quite well in the cases studied. Still, some problems were encountered during the implementation process. The initial implementation did not have the path-sorting algorithm, and the readout location of a block was determined solely by the Doppler shift. The result from the initial implementation can be seen in figure 5.1. As seen from the figure, this implementation failed to get a seamless overlap of blocks, causing the output audio to stutter every time the source position was updated. This was particularly noticeable for source signals with distinctive elements, like a siren or speech. The cause of this problem was likely due to a small rounding error in the calculation of the Doppler effect, as the readout size of the blocks was dependent on the Doppler shift.

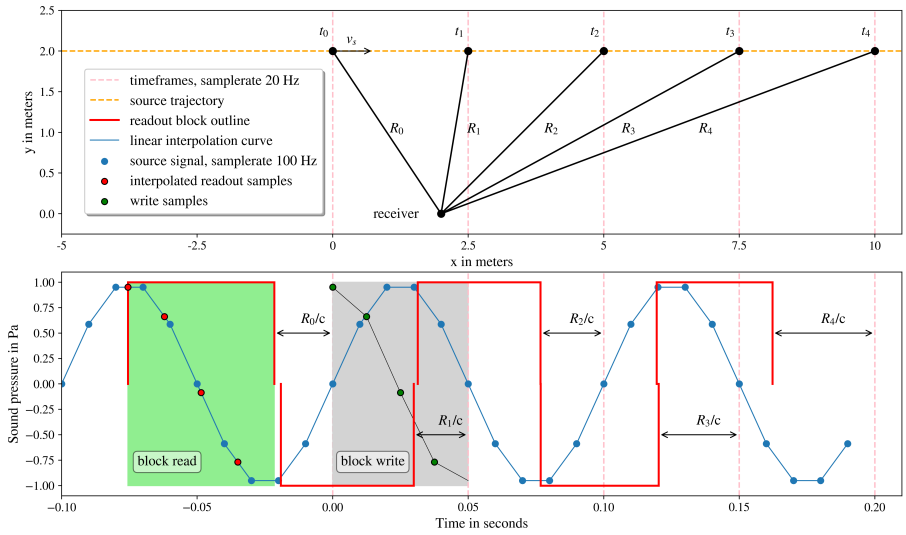


Figure 5.1: Faulty readout of blocks from the initial implementation of the digital processing system.

This eventually led to a revised implementation with a path-sorting algorithm. By sorting the paths, the next start index of a readout block could be guaranteed to be the next after the previous readout block, like in figure 3.3. However, the path-sorting algorithm caused a new set of problems because different types of paths required individual sorting rules. For reflected paths, the paths had to be sorted based on if they shared their reflection planes and had the same order. On the other hand, Diffracted paths could not be sorted based on their order as coherent paths wrapped around one vertex/corner at a time for an occluding object of three dimensions. E.g., a diffracted path with 3rd order in one timeframe could potentially be coherent with a diffracted path of 2nd order in the following timeframe. Instead, diffracted paths had to be sorted based on clockwise/counter-clockwise direction and shared features.

6

Conclusion

This thesis aimed to gain insight into auralizations from virtual urban environments. Moreover, a specific aim was to generate and evaluate auralizations with multiple Dopper effects from moving sound sources. The auralizations were generated with a code implementation in Python, and a listening experiment was conducted to evaluate certain attributes of the auralizations, such as perceived plausibility and suitable sampling rates of the sound source position. Recorded ambient sounds were added to the auralizations to check whether the perceived plausibility was increased compared to auralizations without ambient sound.

The paired comparison in the first part of the listening experiment and the results from Fisher's exact test showed a perceived difference between single Doppler and multiple Doppler stimuli due to a p-value of 0.0004. Furthermore, the odds ratio of 0.247 indicated that participants were more likely to favor the multiple Doppler stimuli over the single Doppler stimuli. However, it is worth noting that the limited sample size suggests further validation. The second part of the listening experiment tested the minimum suitable sampling rate for the position of a moving source. For simple non-diffracting problems, a sampling rate starting at 10 Hz seemed sufficient. For more complex problems, including diffraction, higher sampling rates were favored as the attenuation filters from the diffraction model required frequent updates to follow the change in the sound field in the transition between non-shadow zones and shadow zones and vice versa. The comparison between ambient and non-ambient stimuli did not seem to improve the perceived plausibility. The small sample size and large confidence intervals undermined the estimated mean perceived plausibility, emphasizing the need for further validation.

From geometrical acoustics, the image source method could effectively find specular reflections up to 3rd order without a problem. Furthermore, the revised implementation of the digital signal processing system proved to work well in the cases studied. Despite the difficulties encountered with the readout of the blocks, the point was eventually demonstrated successfully: The auralizations with multiple Doppler effects seemed to increase the perceived plausibility compared to the single Doppler effect auralizations.

6.1 Further Work

In the limited duration of this thesis, only a few topics could be researched more extensively. This section lists some potential future work.

Address problems with the block readout

A good place to start with future work would be to address the issues with the readout of the blocks. This would simplify the system substantially, and the path-sorting algorithm could be dropped.

Implement a ground impedance model

As the implementation did not include an impedance model for reflecting objects, an interesting approach would be to investigate further how material absorption would affect the results as well as the spherical reflection factor for finite-impedance ground surfaces. Only hard and soft surfaces with reflection coefficients $|R| = 1$ could be used in the implemented system.

Provoke audible multiple Doppler effects from measurements

Even though they might be there, multiple Doppler effects are hardly observable in the real world. It would be interesting to try to compose a measurement method that could provoke this effect and possibly compare it with simulations. This could gain valuable insights into the characteristics of this effect.

Explore beam tracing methods

A beam tracing method could be used to reduce the computational cost of the ISM and possibly integrate a more comprehensive diffraction model such as the UTD. UTD has been effectively used before in the auralization of outdoor sound propagation and blends well with the ISM.

Add detail to the scene environment

A geospatial ground could be implemented to further investigate auralization for more specific locations. The detail of the features could also have been extended to include e.g. a gable roof or simple ornaments on the walls.

Sustain the subjective assessment of multiple Doppler auralizations

Since this thesis could only scratch the surface of the subjective assessment of multiple vs. single Doppler effects, this investigation could be continued after improvements in the implementation.

Bibliography

[1] Institute for Hearing Technology and Acoustics, RWTH Aachen University. *Virtual Acoustics – A real-time auralization framework for scientific research.* URL: https://www.virtualacoustics.org/ (visited on 06/07/2023).

[2] Sahar Geravandi et al. "Noise Pollution and Health Effects". en. In: *Jundishapur Journal of Health Sciences* 7.1 (Jan. 2015). ISSN: 2252-021X, 2252-0627. DOI: 10.5812/jjhs.25357. URL: https://brieflands.com/articles/jjhs-60312.html (visited on 05/21/2023).

[3] Mohammad Javad Zare Sakhvidi et al. "Association between noise exposure and diabetes: A systematic review and meta-analysis". en. In: *Environmental Research* 166 (Oct. 2018), pp. 647–657. ISSN: 00139351. DOI: 10.1016/j.envres.2018.05.011. URL: https://linkinghub.elsevier.com/retrieve/pii/S0013935118302512 (visited on 05/21/2023).

[4] Penelope A. Bryant, John Trinder, and Nigel Curtis. "Sick and tired: does sleep have a vital role in the immune system?" en. In: *Nature Reviews Immunology* 4.6 (June 2004), pp. 457–467. ISSN: 1474-1733, 1474-1741. DOI: 10.1038/nri1369. URL: https://www.nature.com/articles/nri1369 (visited on 05/21/2023).

[5] Tai Sekine et al. *The impact of sleep deprivation on the coronary circulation | Elsevier Enhanced Reader*. en. DOI: 10.1016/j.ijcard.2009.01.013. URL: https://www.sciencedirect.com/science/article/abs/pii/S0167527309000539 (visited on 05/21/2023).

[6] Hans Slabbekoorn et al. *Effects of Anthropogenic Noise on Animals*. en. Google-Books-ID: jl5qDwAAQBAJ. Springer, Aug. 2018. ISBN: 978-1-4939-8574-6.

[7] Krister Larsson and Hans Jonasson. *Uppdaterade beräkningsmodeller för vägtrafikbuller*. URL: https://bransch.trafikverket.se/contentassets/773857bcf506430a880a79f76195a080/forskningsresultat/sp_uppdaterade_bullerberakningar.pdf.

[8] Timothy Van Renterghem. *Multi-stage sound planning methodology for urban redevelopment | Elsevier Enhanced Reader*. en. DOI: 10.1016/j.scs.2020.102362. URL: https://www.sciencedirect.com/science/article/abs/pii/S2210670720305837 (visited on 05/21/2023).

[9] Beitske Boonstra and Luuk Boelens. "Self-organization in urban development: towards a new perspective on spatial planning". en. In: *Urban Research & Practice* 4.2 (July 2011), pp. 99–122. ISSN: 1753-5069, 1753-5077. DOI: 10.

1080/17535069.2011.579767. URL: http://www.tandfonline.com/doi/abs/10.1080/17535069.2011.579767 (visited on 05/22/2023).

[10] Samuel D. Brody, David R. Godschalk, and Raymond J. Burby. "Mandating Citizen Participation in Plan Making: *Six Strategic Planning Choices*". en. In: *Journal of the American Planning Association* 69.3 (Sept. 2003), pp. 245–264. ISSN: 0194-4363, 1939-0130. DOI: 10.1080/01944360308978018. URL: https://www.tandfonline.com/doi/full/10.1080/01944360308978018 (visited on 05/22/2023).

[11] Michael Vorländer. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. en. Google-Books-ID: jf75DwAAQBAJ. Springer Nature, Aug. 2020. ISBN: 978-3-030-51202-6.

[12] U Peter Svensson. "MODELLING ACOUSTIC SPACES FOR AUDIO VIRTUAL REALITY". en. In: ().

[13] Mendel Kleiner, Bengt-Inge Dalenbäck, and Peter Svensson. "Auralization-An Overview". English. In: *Journal of the Audio Engineering Society* 41.11 (Nov. 1993). Publisher: Audio Engineering Society, pp. 861–875. URL: https://www.aes.org/e-lib/browse.cfm?elib=6976 (visited on 05/22/2023).

[14] Ning Xiang and Gerhard M. Sessler, eds. *Acoustics, Information, and Communication: Memorial Volume in Honor of Manfred R. Schroeder*. en. Cham: Springer International Publishing, 2015. ISBN: 978-3-319-05659-3 978-3-319-05660-9. DOI: 10.1007/978-3-319-05660-9. URL: https://link.springer.com/10.1007/978-3-319-05660-9 (visited on 05/22/2023).

[15] Ulf Peter Svensson. "L2 NUMERICAL METHODS TTT12 - Geometrical Acoustics". en. Unpublished document. Norwegian University of Science and Technology.

[16] Z. Maekawa. "Noise reduction by screens". en. In: *Applied Acoustics* 1.3 (July 1968), pp. 157–173. ISSN: 0003682X. DOI: 10.1016/0003-682X(68)90020-0. URL: https://linkinghub.elsevier.com/retrieve/pii/0003682X68900200 (visited on 05/07/2023).

[17] Jonas Stienen. *Real-Time Auralisation of Outdoor Sound Propagation*. Logos Verlag Berlin, Mar. 2023. ISBN: 978-3-8325-5629-7. DOI: 10.30819/5629. URL: https://www.logos-verlag.de/cgi-bin/engbuchmid?isbn=5629&lng=eng&id= (visited on 06/05/2023).

[18] Nicolas Tsingos et al. "Modeling acoustics in virtual environments using the uniform theory of diffraction". en. In: *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, Aug. 2001, pp. 545–552. ISBN: 978-1-58113-374-5. DOI: 10.1145/383259.383323. URL: https://dl.acm.org/doi/10.1145/383259.383323 (visited on 06/05/2023).

[19] Renez Nota, Robert Barelds, and Dirk van Maercke. *Harmonoise WP 3 Engineering method for road traffic and railway noise after validation and fine-tuning*. Tech. rep.

[20] Allan D. Pierce. *Acoustics: An Introduction to Its Physical Principles and Applications*. en. Cham: Springer International Publishing, 2019. ISBN: 978-3-030-11213-4 978-3-030-11214-1. DOI: 10.1007/978-3-030-11214-1. URL: http://link.springer.com/10.1007/978-3-030-11214-1 (visited on 01/20/2023).

[21] A Hirschberg and S W Rienstra. “An introduction to aeroacoustics”. en. In: (2004).

[22] Daniel A. Russell. *Absorption and Attenuation of Sound in Air*. URL: https://www.acs.psu.edu/drussell/Demos/Absorption/Absorption.html (visited on 05/31/2023).

[23] *ISO9613-1 Acoustics - Attenuation of Sound during propagation outdoors*.

[24] Benjamin Bernschütz. *Spherical Far-Field HRIR Compilation of the Neumann KU100*. Publisher: Zenodo. Merano, Italy, July 2020. DOI: 10.5281/zenodo.3928297. URL: https://zenodo.org/record/3928297 (visited on 01/20/2023).

[25] John G. Proakis and Dimitril K. Manolakis. *Digital Signal Processing*. Fourth Edition. Pearson. ISBN: 978-1-292-02573-5.

[26] Patrik Andersson et al. “Sound and Vibration Measurements. Lab 4: System Analysis. Basic steps and measures in practical frequency response measurements.” en. Unpublished document. Chalmers University of Technology, Jan. 2023.

[27] Gary P. Scavone. *Delay Lines*. URL: https://www.music.mcgill.ca/~gary/307/week2/delaylines.html (visited on 06/06/2023).

[28] Julius O Smith et al. “Doppler Simulation and the Leslie”. en. In: (2002).

[29] J Stienen and M Vorlã. “Real-time auralization of propagation paths with reflection, diffraction and the Doppler shift”. en. In: (2018).

[30] Bane Sullivan and Alexander Kaszynski. “PyVista: 3D plotting and mesh analysis through a streamlined interface for the Visualization Toolkit (VTK)”. In: *Journal of Open Source Software* 4.37 (May 2019), p. 1450. DOI: 10.21105/joss.01450. URL: https://doi.org/10.21105/joss.01450.

[31] Mats E Nilsson et al. “Perceptual validation of auralized road traffic noise”. en. In: ().

[32] *SQala – Software for Jury Testing*. en. URL: https://www.head-acoustics.com/products/jury-testing/sqala (visited on 06/09/2023).

A

Frequency Spectra

The following figures depict the frequency spectra for the single and multiple Doppler stimuli used in the listening experiment (the six rightmost spectrograms in figure 4.1).

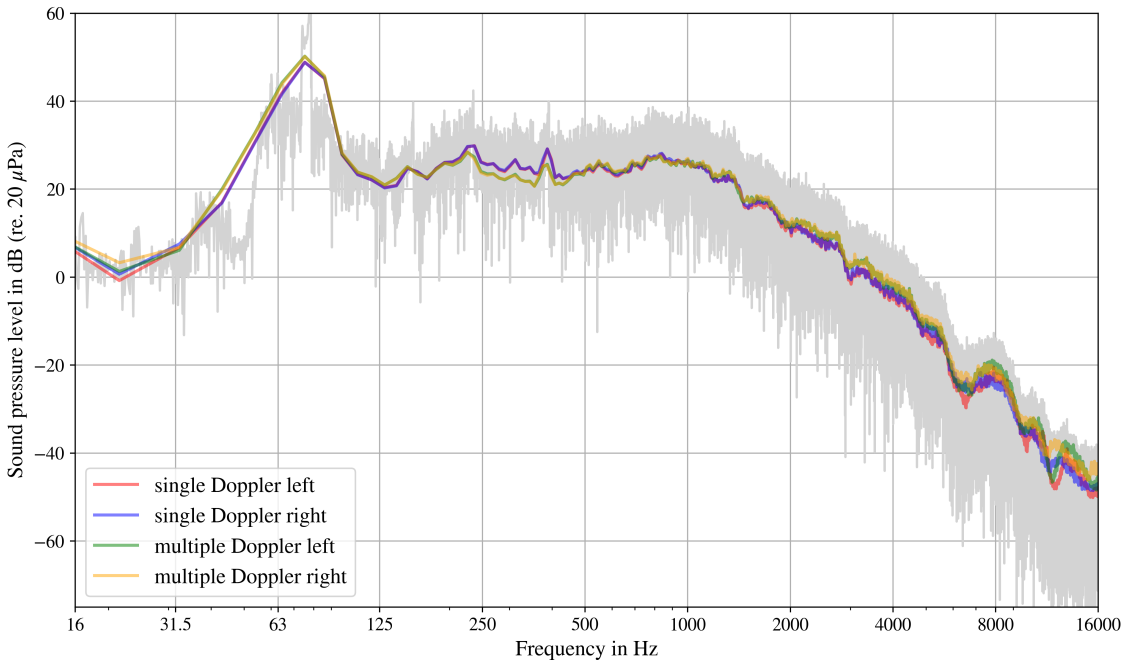


Figure A.1: Frequency spectra for the single and multiple Doppler stimuli generated from recording r_2 for test scene 1 depicted in figure 3.5.

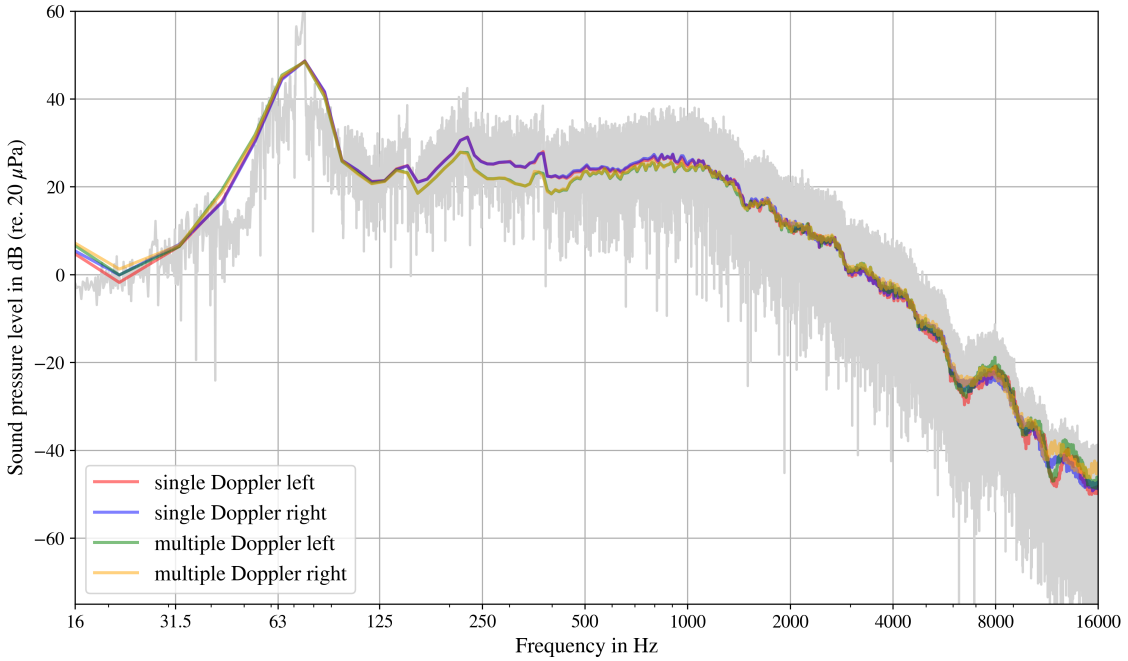


Figure A.2: Frequency spectra for the single and multiple Doppler stimuli generated from recording r_3 for test scene 1 depicted in figure 3.5.

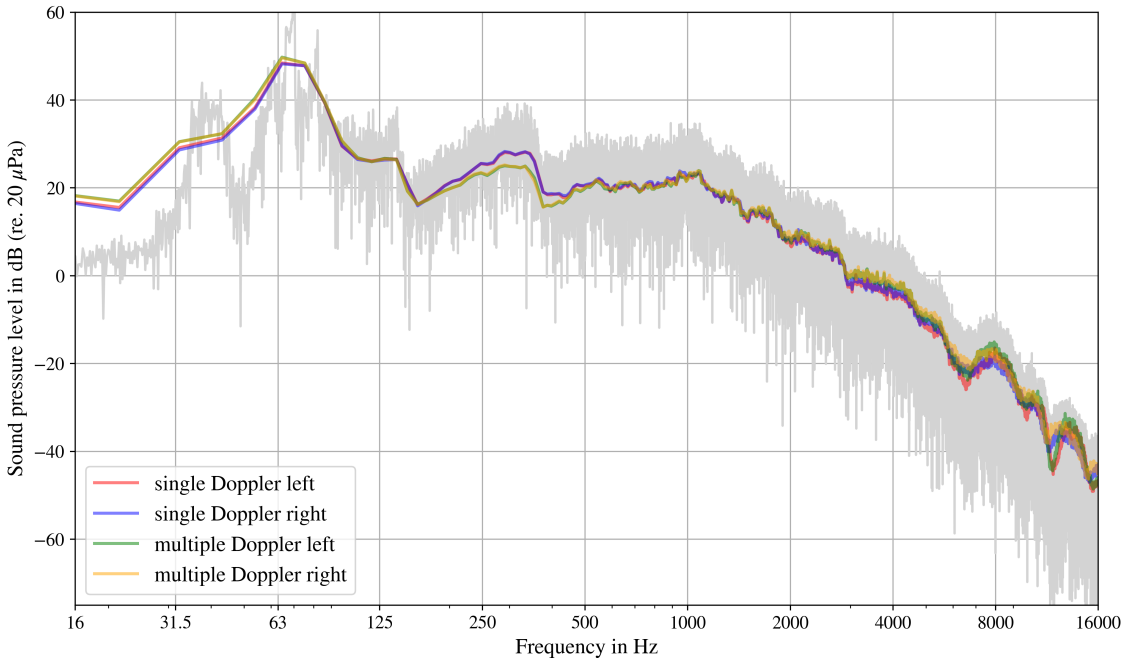


Figure A.3: Frequency spectra for the single and multiple Doppler stimuli generated from recording r_4 for test scene 1 depicted in figure 3.5.

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY