

# **Sound Source Localization and Distance Estimation in Open Environment using Simulation and AI**

**Master Thesis**

Denis Rosset

University of Applied Sciences and Arts Western Switzerland

July 2023

# Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

## Supervisors:

Michael Mäder: Professor in computer science

Beat Wolf Professor: in computer science

## Principals

Marc-Antoine Fénart: Professor in civil engineering

Gabriel Python: Scientific associate at Rosas

# Acknowledgements

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.[2] [1]

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Motivation . . . . .	5
1.1.1	Objectives . . . . .	5
1.2	Challenges . . . . .	5
1.2.1	Challenge n°1: Realistic datasets . . . . .	5
1.2.2	Challenge n°2: Robust models . . . . .	5
1.2.3	Challenge n°3: Adequate evaluation metrics . . . . .	5
1.3	Structure of the thesis . . . . .	6
<b>2</b>	<b>Literature</b>	<b>7</b>
2.0.1	Sound Propagation . . . . .	7
2.0.2	Realistic sound propagation in simulations . . . . .	7
2.0.2.1	Microsoft Project Acoustics . . . . .	7
2.0.3	Sound Propagation in game-engine . . . . .	7
2.1	Sound Source Localization . . . . .	7
2.1.1	Spectrograms for sound source localization . . . . .	8
2.1.2	Convolutional Neural Networks . . . . .	8
2.1.3	Convolutional Neural Networks for source localization . . . . .	8
2.2	Related Work . . . . .	8
2.2.1	. . . . .	8
<b>3</b>	<b>Methodology</b>	<b>10</b>
<b>4</b>	<b>Setup</b>	<b>11</b>
<b>5</b>	<b>Results</b>	<b>12</b>
<b>6</b>	<b>Conclusions and Future Work</b>	<b>13</b>
<b>A</b>	<b>Appendix</b>	<b>14</b>
A.1	. . . . .	15
	<b>List of Tables</b>	<b>16</b>
	<b>List of Figures</b>	<b>16</b>
	<b>Bibliography</b>	<b>17</b>

# 1

## Introduction

Within the framework of the research project "NPR Teleoperation", the engineers of the HEIA-FR have developed the first concept in Switzerland of a remote-controlled automated vehicle. However, teleoperation only makes sense if the vehicle is automated. There can be no teleoperation without automation (economic factors) just as there can be no automation without teleoperation (legal, technical, and social factors). ROSAS then created the Autovete (Automatisation de véhicules téléopérés) project, financed by HEIA-FR, to build up vehicle automation expertise. For a vehicle to be fully autonomous, the detection of other emergency vehicles is mandatory. To solve this issue, V2V (Vehicle-to-Vehicle) communication can be used but is not yet integrated into emergency vehicles. So, to be able to detect such a vehicle, two signals need to be processed: the sound of the emergency siren and the blinking lights of the vehicle. The first use case of this project focuses only on sound source distance estimation and localization.

To understand if sound source estimation and localization could work for emergency detection, a simpler use case has been created for this project. It is the detection of excessively noisy vehicles on the street. The goal is to measure the sound level of the passing vehicles and compare it with the legal limits. If a vehicle exceeds the limit, the system can record its license plate and report it to the authorities. This way, the system can help reduce noise pollution and improve road safety. To implement this use case, the system requires a microphone array, a camera, and a processing unit. The microphone array captures the sound signals from different directions and sends them to the processing unit. The processing unit applies a sound source localization algorithm to estimate the direction and distance of the sound source. The camera captures the image of the vehicle and performs license plate recognition. Big improvements in sound source localization with the help of machine learning are being made<sup>1</sup> and can be used to reliably localize the origin of a sound using one or more microphone arrays (multiple microphones operating in tandem).

A non-negligible problem is that the number of real-world datasets with moving sources in an open environment is limited. A solution is to create the datasets in realistic sound propagation simulation. To validate and use the model, it should also be tested to see how it reacts against adversarial attacks, understand how it can be used in a real environment and limit the attack vector.

## 1.1 Motivation

### 1.1.1 Objectives

#### 3.1 Objective n°1 Dataset according to the baseline

The first objective is to construct a dataset that is coherent with the project's baseline. The dataset should contain the target variable, features, and necessary pre-processing steps such as missing data imputation, data normalization, and feature engineering. This dataset will help create and understand the problem.

#### 3.2 Objective n°2 Model for better sound source localization and distance estimation

The project should use a neural network model to detect the origin of a sound using a microphone array. The neural network should be trained using the dataset created in objective 3.1 and should be able to accurately localize the sound source. The trained neural network model should be evaluated in a real environment to see how it performs. It should also be evaluated to see how dependable it is in localizing sound sources and how it can be improved.

#### 3.3 Objective n°3 The model should be tested to see how it reacts to attacks

The trained neural network model needs to be evaluated by testing it on data that has been modified in some way, such as by adding or removing noise, or by modifying the sound source. The model could also be tested against various types of attacks, such as masking, time-warping, and frequency-shifting.

## 1.2 Challenges

### 1.2.1 Challenge n°1: Realistic datasets

One of the main challenges in this project is to construct realistic datasets that accurately capture the data in a real-world scenario. The lack of open-source datasets that contain moving sound sources in open-loop environments can make this difficult. To overcome this challenge, the project should use realistic simulations of sound propagation to produce suitable datasets.

### 1.2.2 Challenge n°2: Robust models

Another challenge is to create a neural network model that can accurately detect the origin of a sound using a microphone array. The trained model should be highly accurate and robust enough to resist adversarial attacks. To assess the model's robustness, the model should be tested on data that has been modified in some way, such as by adding or removing noise, or by modifying the sound source. The model should also be able to accurately localize the sound source despite the attack.

### 1.2.3 Challenge n°3: Adequate evaluation metrics

The last challenge is to create an evaluation metric that adequately reflects the model's performance. The evaluation metric should take into account the accuracy of sound source localization in a real environment as well as its ability to resist adversarial attacks. The metric should also be able to capture how well the model can provide reliable results in a variety of environments.

### 1.3 Structure of the thesis

- **Chapter 2: Background:** This chapter provides an overview of the background knowledge necessary to understand the project, such as the theory of machine learning and sound propagation, and a brief introduction of sound source localization.
- **Chapter 3: Methodology:** This chapter explains the methodology used in the project, such as creating a dataset, creating a simulation, building a neural network model, and evaluating it using an appropriate metric.
- **Chapter 4: Setup:** This chapter describes the work done on the project, such as the simulations used to create the dataset, the neural network model used, and the evaluation metrics used.
- **Chapter 5: Results:** This chapter presents the results of the project, such as the performance of the neural network model and the evaluation metrics.
- **Chapter 6: Conclusion and future work:** This chapter concludes this project with a discussion of the main results and a summary of the key findings.

# 2

## Literature

This chapter introduces technical concepts and background used in the conceptualized solution of the thesis.

### **2.0.1 Sound Propagation**

Sound propagation is the physical process by which sound waves propagate in a given environment. The strength of the sound wave is dependent on a variety of factors, including the frequency, environment, and distance from the sound source. This makes it difficult to accurately identify and localize a sound source, and thus a more accurate and robust sound source localization system is needed.

### **2.0.2 Realistic sound propagation in simulations**

#### **2.0.2.1 Microsoft Project Acoustics**

Microsoft Project Acoustics is a sound propagation engine that simulates the propagation of sound waves in a given environment. It is used in a variety of applications, including video games, virtual reality, and augmented reality. It simulates wave effects like obstruction, reverberation and occlusion in complex 3D scenes without requiring zone markup or raytracing. It works similarly to a raytracing engine, but it is precomputed and optimized for real-time performance.

### **2.0.3 Sound Propagation in game-engine**

## **2.1 Sound Source Localization**

Sound Source Localization (SSL) is the process of determining the position of a sound source. This is usually done using a microphone array that captures the sound signals from multiple directions. SSL is used in a variety of applications, such as speech recognition, robot navigation, surveillance, and security. In this thesis, SSL is used to estimate the distance and direction of a sound source to detect excessively noisy vehicles.



### 2.1.1 Spectrograms for sound source localization

Spectrograms are a visual representation of the frequency content of a sound signal. They are often used in sound source localization to identify the direction of a sound source. The spectrogram is a two-dimensional representation of the frequency content of a sound signal.

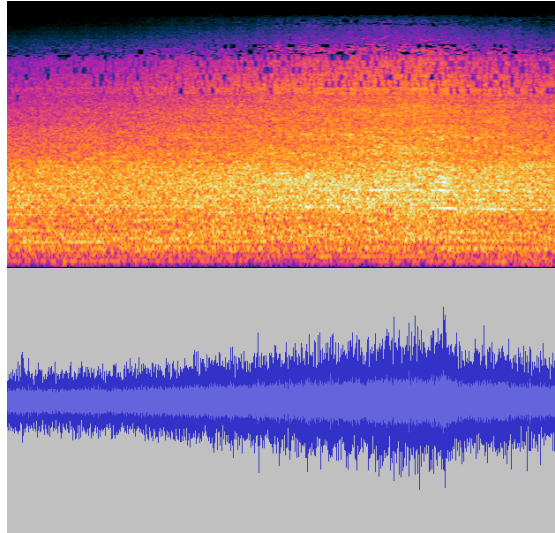


Figure 2.1: Spectrogram of a sound signal

The x-axis represents time and the y-axis represents frequency. The intensity of the color at each point in the spectrogram represents the amplitude of the frequency component then. The spectrogram can be used to identify the direction of a sound source by looking at the frequency content of the sound signal. The frequency content of a sound signal is dependent on the direction of the sound source, so the spectrogram can be used to identify the direction of a sound source.

### 2.1.2 Convolutional Neural Networks

### 2.1.3 Convolutional Neural Networks for source localization

## 2.2 Related Work

### 2.2.1

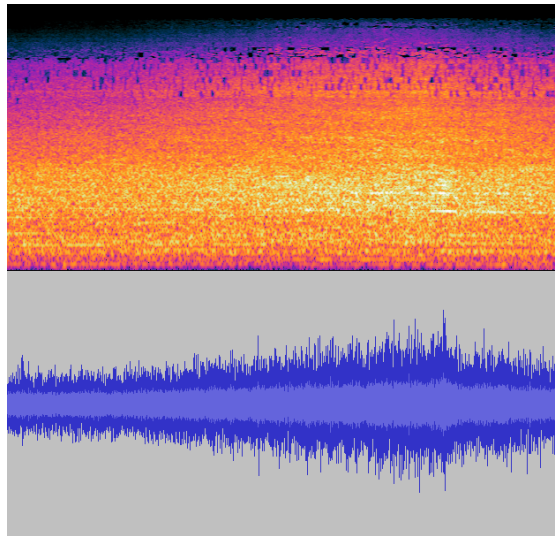


Figure 2.2: CNN sound classification diagram

# 3

## Methodology

# 4

## Setup

# 5

## Results

# 6

## Conclusions and Future Work



## Appendix

## A.1



## List of Tables

# List of Figures

2.1	Spectrogram of a sound signal . . . . .	8
2.2	CNN sound classification diagram . . . . .	9

# Bibliography

- [1] Donald E. Knuth. Literate programming. *The Computer Journal*, 27(2):97–111, 1984.
- [2] Leslie Lamport. *TEX: a Document Preparation System*. Addison Wesley, Massachusetts, 2 edition, 1994.