THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN APPLIED ACOUSTICS

Listening Experiments in Virtual Acoustic Environments: Framework and Application to Electric Vehicle Noise

Leon Müller

Department of Architecture and Civil Engineering
Division of Applied Acoustics
Chalmers University of Technology
Gothenburg, Sweden, 2026

Listening Experiments in Virtual Acoustic Environments: Framework and Application to Electric Vehicle Noise

LEON MÜLLER
ISBN 978-91-8103-402-8

Acknowledgements, dedications, and similar personal statements in this thesis reflect the author's own views.

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5859
ISSN 0346-718X
https://doi.org/10.63959/chalmers.dt/5859

Department of Architecture and Civil Engineering
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
Phone: +46 (0)31 772 1000

Cover: Participant performing the electric vehicle localization experiment described in Section 4.3.2.

All photos and illustrations by Leon Müller, unless otherwise stated.
Figure 2.5 illustrated by Janneke van der Leer.

Printed by Chalmers Digital Printing
Gothenburg, Sweden, 2026

Listening Experiments in Virtual Acoustic Environments:
Framework and Application to Electric Vehicle Noise
LEON MÜLLER
Department of Architecture and Civil Engineering
Chalmers University of Technology

Abstract

Environmental noise, and in particular road traffic noise, has a well-documented impact on human health, well-being, and quality of life. Understanding these effects and developing acoustic solutions that improve everyday sound environments requires controlled human-subject experiments that capture both acoustic and perceptual complexity. However, reproducing realistic acoustic environments in controlled laboratory settings is a methodological challenge.

This thesis develops and demonstrates a methodological framework for studying human responses to sound using controlled virtual acoustic environments. The framework combines modular, physically informed, and perceptually validated auralization with experimental paradigms for assessing subjective, physiological, and behavioral responses. Electric vehicle (EV) noise, and in particular acoustic vehicle alerting systems (AVAS), serve as a concrete application through which the framework is implemented, tested, and refined.

The framework and its applications are demonstrated through six studies addressing complementary aspects of EV noise perception. The first two establish and validate the auralization model for outdoor and indoor environments. Subsequent studies build on this foundation to examine how AVAS directivity influences perceived vehicle speed, how different AVAS designs affect auditory localization accuracy, and how low-level EV and traffic noise influence attention, workload, electrodermal activity, and annoyance in residential settings.

The results show that AVAS design choices can substantially affect localization performance as well as subjective and physiological responses, even when signals comply with current regulations. Strongly tonal two-tone designs performed poorly across several response domains, while broadband noise-based and multi-tone designs showed more balanced performance under the tested conditions. Beyond these application-specific findings, the thesis contributes a generalizable methodological approach for integrating auralization-based virtual acoustic environments with human response evaluation. It emphasizes the importance of perceptual validity, reproducibility, and interdisciplinary integration between acoustical simulation methods and empirical human-response research.

Keywords: Auralization, virtual acoustic environments, electric vehicles, acoustic vehicle alerting system (AVAS), human response.

Preface

This thesis is based on research performed between June 2021 and April 2026 at the Division of Applied Acoustics, Department of Architecture and Civil Engineering, Chalmers University of Technology. It was funded primarily by FORMAS – a Swedish research council for sustainable development – under grant FR-2020-01931, with additional support from the HEAD-Genuit-Foundation under grant P-22101-W.

First, I want to thank my supervisor, Wolfgang Kropp, for his continuous support and for our discussions on all aspects of acoustics over the years. The freedom and trust you gave me to pursue my own research interests – and plenty of other activities along the way – meant a lot. I also want to thank my examiner, Jens Forssén, and co-supervisor, Jens Ahrens, for always being available when I needed input.

A special mention goes to Thomas Deppisch for endless conversations on everything from sourdough to spherical harmonics, for companionship through many travel adventures, and for being a good friend. I would also like to thank my current and former colleagues and visitors at Applied Acoustics, especially Elin, Hannes, Jannik, Samanta, Xavi, Carl, Fati, and Krister, for exchange, fika, and good company. Thanks as well to Roland Sottek for years of collaboration on the HRSV course and for the trust and freedom to teach in my own way.

This thesis would look quite different if I hadn't met Marlene Wessels at the DAGA conference in 2022. Thank you, Marlene, for introducing me to the world of statistics and experimental psychology, and for a genuinely fun and productive collaboration since then.

During my time at Chalmers, I was lucky to be part of Dr.GENIE, GENIE@ACE, the ACE PhD Council, and the DS. I want to thank all members of these groups for volunteering their time and energy working toward a better academic community – and for the sense of purpose it gave me beyond day-to-day work.

I would like to thank all the other people who have accompanied me throughout these five years, both inside and outside of Chalmers, and have been there for me, even in hard times. Especially Janneke, for countless running, climbing, and coffee sessions – and for being there when it mattered. I also want to thank Staffan, Angelica, Anna, and the climbing gang for everything outside the office.

My family has always supported me, unconditionally and without question, and that has meant more than I can put into words here. Danke für alles! Lukas – thanks for staying close despite the distance. And Alicia – thank you for your support, your patience, and for making everything better.

List of Publications

This thesis is based on the following publications, which are the result of collaborative research. For all papers, the author of this thesis was responsible for the primary scientific contributions, including study conception, method development, implementation of the auralization framework, experimental design and data acquisition, statistical analysis, and manuscript preparation. Co-authors contributed through supervision, methodological discussions, and constructive feedback during the research and publication process.

[A] Leon Müller, Wolfgang Kropp, "Auralization of Electric Vehicles for the Perceptual Evaluation of Acoustic Vehicle Alerting Systems". *Acta Acustica*, 2024, 8, 27.

[B] Leon Müller, Jens Ahrens, Wolfgang Kropp, "Loudspeaker Array-Based Auralization of Electric Vehicle Noise in Living Environments". *Proceedings of Forum Acusticum*, 11th Convention of the European Acoustics Association, Forum Acusticum 2025, Málaga, Spain.

[C] Leon Müller, Wolfgang Kropp, "On the Influence of AVAS Directivity on Electric Vehicle Speed Perception". *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, INTER-NOISE 2024, Nantes, France.

[D] Leon Müller, Jens Forssén, Wolfgang Kropp, "Auditory Localization of Multiple Stationary Electric Vehicles". *Journal of the Acoustical Society of America*, 157, 2025.

[E] Leon Müller, Jens Forssén, Wolfgang Kropp, "Effects of Low-Level Electric Vehicle Noise on Attention, Electrodermal Activity, Workload, and Annoyance". *Journal of the Acoustical Society of America*, 159, 2026.

[F] Leon Müller, Jens Forssén, Wolfgang Kropp, "Traffic Noise at Moderate Levels Affects Cognitive Performance: Do Distance-Induced Temporal Changes Matter?". *International Journal of Environmental Research and Public Health*, 20(5), 2023.

The following publications are not included in this thesis due to an overlap in content or content going beyond the scope of this thesis:

[1] Jens Ahrens, Leon Müller, "Perceptually Transparent Binaural Auralization of Simulated Sound Fields". *Journal of the Audio Engineering Society*, 74(3), 2026.

[2] Jens Forssén, Leon Müller, Elin Hedlund, Wolfgang Kropp, "Hybrid prediction tool implemented for acoustic design studies of open plan office spaces". *Applied Acoustics*, 248, 2026.

[3] Marlene Wessels⋆, Leon Müller⋆, Anna Luisa Maier, Johannes Kraus, "Auditory localization and subjective assessment of autonomous cleaning robot sounds: A VR experiment on speed, operating mode and alerting signals". Under review in *Acta Acustica*, ⋆These two authors contributed equally to this work and share the first-authorship.

[4] Marlene Wessels, Jorge de Heuvel, Leon Müller, Anna Luisa Maier, Maren Bennewitz, Johannes Kraus, "Auditory Localization and Assessment of Consequential Robot Sounds: A Multi-Method Study in Virtual Reality". *34th IEEE International Conference on Robot and Human Interactive Communication*, IEEE RO-MAN 2025, Eindhoven, The Netherlands, 2025.

[5] Elin Hedlund, Leon Müller, Wolfgang Kropp, "Influence of Speech Exposure on Working Memory: Exploring Reverberation Time in Open Plan Office Environments". *Fortschritte der Akustik – DAGA 2024*, Hannover, Germany, 2024.

[6] Wolfgang Kropp, Leon Müller, Jannik Theyssen, "An auralisation approach for tyre/road noise". *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, INTER-NOISE24, Nantes, France, 2024.

[7] Felicitas Bederna, Leon Müller, Jens Ahrens, "Perceptual Detection Thresholds for Alterations of the Azimuth of Early Room Reflections". *Fortschritte der Akustik – DAGA 2023*, Hamburg, Germany, 2023.

[8] Leon Müller, Jens Ahrens, "Perceptual Differences for Modifications of the Elevation of Early Room Reflections". *International Conference on Audio for Virtual and Augmented Reality 2022*, Redmond, USA, 2022.

[9] Leon Müller, Wolfgang Kropp, Jens Forssén, "Measurement, Simulation and Auralization of Indoor Road Traffic Noise". *Fortschritte der Akustik – DAGA 2022*, Stuttgart, Germany, 2022.

[10] Wolfgang Kropp, Leon Müller, "The Description of Sound Sources by a Set of Equivalent Sources Using Impulse Response Functions and the Least Mean Square Algorithm". *Fortschritte der Akustik – DAGA 2022*, Stuttgart, Germany, 2022.

[11] Leon Müller, Wolfgang Kropp, Georgios Zachos, Jens Forssén, "Investigating Low Frequency Sound from Traffic in a Living Room Lab". *Fortschritte der Akustik – DAGA 2021*, Vienna, Austria, 2021.

Contents

List of Figures

Acronyms

AVAS: Acoustic Vehicle Alerting System

ANOVA: Analysis of Variance

BEM: Boundary Element Method

BRIR: Binaural Room Impulse Response

BTL: Bradley–Terry–Luce

CI: Confidence Interval

CNOSSOS-EU: Common Noise Assessment Methods in Europe

CPX: Closed Proximity Method

CPT: Continuous Performance Test

EDA: Electrodermal Activity

EEA: European Environment Agency

EEG: Electroencephalography

EU: European Union

EV:	Electric Vehicle
FEM:	Finite Element Method
FMVSS:	Federal Motor Vehicle Safety Standard
GPS:	Global Positioning System
HEV:	Hybrid Electric Vehicle
HOA:	Higher-Order Ambisonics
HRTF:	Head-Related Transfer Function
ICBEN:	International Commission on Biological Effects of Noise
ICEV:	Internal Combustion Engine Vehicle
ILD:	Interaural Level Difference
ITD:	Interaural Time Difference
LRL:	Living Room Lab
LTI:	Linear and Time-Invariant
MAC:	Modal Assurance Criterion
MIMO:	Multiple Input Multiple Output
NASA:	National Aeronautics and Space Administration
NHTSA:	National Highway Traffic Safety Administration
RANSAC:	Random Sample Consensus
SCL:	Skin Conductance Level
SCR:	Skin Conductance Response
SDG:	Sustainable Development Goal
SH:	Spherical Harmonics
UN:	United Nations
UNECE:	United Nations Economic Commission for Europe
VBAP:	Vector Base Amplitude Panning
WFS:	Wave Field Synthesis
WHO:	World Health Organization

CHAPTER 1

Introduction

1.1 Road Traffic Noise and Public Health

Transportation noise, and *road traffic noise* in particular, is a global public health problem. According to the latest environmental noise report of the *European Environment Agency (EEA)*, 24% of the total European population is exposed to long-term harmful transportation noise levels [1]. Of these, approximately 112 million people are subject to road traffic noise, mostly in densely populated urban areas. Such widespread exposure not only causes annoyance and discomfort but also has serious consequences for health and well-being. So-called *burden of disease* studies have shown links between chronic noise exposure and outcomes such as elevated blood pressure, cardiovascular disease, sleep disorders, and premature mortality [2]. Certain groups, including children, older adults, and individuals with preexisting health conditions, are particularly vulnerable to these effects, and the EEA report further highlights that noise exposure also entails broader societal consequences, including increased healthcare costs, reduced productivity, and lower quality of life.

While the *European Union (EU)* has the monitoring and research infrastructure to quantify these effects, comparable studies are scarce in many less-developed countries. This disparity highlights not only differences in research capacity but also global inequalities in exposure and protection, as communities in rapidly urbanizing regions may face higher noise levels with fewer mitigation measures in place.

However, even in industrialized nations, reducing road traffic noise remains a considerable challenge. Much of the exposure occurs in densely populated urban areas where mobility demand is high, and traffic is closely linked to economic activity and accessibility. Beyond structural measures such as the careful planning of new neighborhoods, mitigation on existing roads often requires restricting or redistributing traffic flows, which can be politically and socially sensitive.

These challenges, both in terms of exposure and mitigation, are directly linked to the United Nations *Sustainable Development Goals (SDGs)* [3], where reducing the health impacts of noise contributes directly to SDG 3 (Good Health and Well-Being), improving urban soundscapes supports SDG 11 (Sustainable Cities and Communities), and ensuring protection from harmful noise for all contributes to SDG 10 (Reduced Inequalities). Additionally, strategies to mitigate noise exposure frequently intersect with broader sustainability goals. For example, the electrification of transport is primarily motivated by the need to reduce greenhouse gas emissions in line with SDG 13 (Climate Action). Still, it also has far-reaching consequences for the urban acoustic environment. However, the specific implications of this transition to electromobility from a noise perspective are not yet well understood, as discussed in the following.

1.2 Noise Implications of Electromobility

1.2.1 Acoustic Characteristics of Electric Vehicles

The ongoing transition to electromobility [4] may alter the urban acoustic environment. For any road vehicle, the level and characteristics of the radiated noise depend on driving speed. For *Internal Combustion Engine Vehicles (ICEVs)*, *propulsion noise* from the engine and exhaust is most significant at low speeds. As speed increases, *rolling noise* from tire-road interaction becomes more prominent, and at high speeds, *aerodynamic noise* due to wind dominates the overall sound level [5]. *Electric Vehicles (EVs)* have a different acoustic profile, as electric motors radiate very little noise compared to combustion engines, so propulsion noise is largely absent at low speeds. This reduced propulsion noise at low speeds makes tire noise relatively more noticeable. However, at medium and high speeds, tire and aerodynamic noise remain the most important contributors for both EVs and ICEVs. Additional factors, such as vehicle weight and torque characteristics, further complicate direct comparisons of overall noise levels between ICEVs and EVs.

However, most harmful road traffic noise exposure originates from major urban roads and highways [6], where driving speeds are typically moderate to high and propulsion noise plays a minor role compared to tire-road noise. The transition to electromobility alone is therefore unlikely to fully resolve the public health issues associated with road traffic noise, as discussed in Section 1.1.

At the same time, it is evident that there are low-speed scenarios where electric vehicles emit significantly less noise than ICEVs [7, 8]. While this reduction has the potential to decrease noise exposure in residential areas with low driving speeds, it also diminishes the *acoustic cues* that pedestrians, cyclists, and other vulnerable road users rely on to detect and localize approaching vehicles – cues that are essential for maintaining situational awareness [9] in traffic. This issue is particularly critical for people with visual impairments, a concern highlighted by organizations such as the *European Blind Union* [10]. Accident data from several countries further suggest a higher likelihood of collisions involving slow-moving EVs and *Hybrid Electric Vehicles (HEVs)* in urban areas [11–15].

1.2.2 Acoustic Vehicle Alerting Systems (AVAS)

To address the traffic safety concerns associated with quiet EVs, many countries require new electric vehicles to be equipped with an *Acoustic Vehicle Alerting System (AVAS)*. These systems emit artificial warning sounds that signal the vehicle's presence and driving behavior. In practice, AVAS solutions typically consist of a small loudspeaker mounted in the vehicle's front bumper, driven by a microcontroller that generates sound as a function of vehicle speed. The specific legal requirements for AVAS differ between regions and therefore directly shape how manufacturers implement these systems.

In the European Union, Japan, China, and other countries, Regulation No. 138 of the *United Nations Economic Commission for Europe (UNECE)* applies [16]. This regulation requires that AVAS sounds:

1. Meet specific minimum overall *sound pressure levels*.

2. Cover at least two separate *third-octave bands*, with level requirements varying by frequency band, and at least one of these bands located at or below 1600 Hz.

3. Include a speed-dependent *frequency shift* to indicate acceleration and deceleration.

4. Not exceed a defined maximum overall sound pressure level.

The exact sound pressure level requirements are specified for 10 km/h and 20 km/h passages as well as reversing maneuvers.

By contrast, the U.S. *Federal Motor Vehicle Safety Standard (FMVSS)* No. 141 defines minimum levels for two or four nonadjacent third-octave bands up to 30 km/h, as well as for stationary and reversing conditions [17]. Unlike the UNECE regulation, FMVSS 141 does not require a speed-dependent frequency shift and imposes no maximum sound pressure level for AVAS signals.

Both regulations are relatively unspecific and leave manufacturers broad freedom in how AVAS sounds are designed. Current implementations, therefore, range from engine-like sounds to highly tonal or "futuristic" designs that sound noticeably different from the sounds we have traditionally associated with cars.

This variety raises a central question: which types of AVAS signals best balance traffic safety with their impact on the acoustic environment? Some sounds may be highly effective at warning pedestrians, but at the same time, they may become a source of annoyance for residents living near busy, low-speed streets. Moreover, people have spent a lifetime learning to interpret the sounds of combustion engine vehicles as cues for judging proximity and danger. Replacing these familiar cues with unfamiliar sounds may require adaptation, and it is uncertain whether such adaptation is realistic when different manufacturers adopt very different sound designs. On the other hand, this challenge also offers an opportunity: instead of being constrained by propulsion noise as a byproduct, AVAS sounds can be intentionally designed to be effective in signaling danger while disturbing the surroundings as little as possible.

The only way to determine which AVAS designs best balance safety and acoustic quality is through evidence-based methods that quantify human responses to sound. Such methods are critical for developing systems that protect vulnerable road users without unnecessarily disturbing others. Although developing such quantitative methods is challenging in itself, they provide the foundation for informed policy and sound design. The following section outlines approaches that are commonly used to address these challenges in contemporary noise research.

1.3 Quantifying the Effect of Noise on Humans

Research on the impact of noise on humans dates back to the early 20th century [18]. Early studies, often motivated by industrialization, examined workers exposed to high levels of machine noise and demonstrated that noise can affect humans both physiologically and cognitively [19]. Since then, extensive research has shown that noise exposure is not limited to short-term disturbances [20] but can also have long-term consequences. These include damage to the auditory system as well as broader impacts on health and well-being, such as hypertension [21], cardiovascular disease [22], sleep disorders [23], learning impairments [24], annoyance [6], and other adverse health outcomes [25–27].

To investigate these effects, researchers generally rely on two complementary approaches. The first is *observational research*, in which noise exposure is investigated in natural environments, often using surveys, long-term cohort studies, or by linking noise maps to health databases [28]. These methods have the advantage of high *ecological validity*, i.e., they capture how noise affects people in real life and can involve very large populations. However, they also face challenges. One important issue is the presence of *confounding variables*, meaning factors other than noise that can influence health outcomes. For example, people living in noisy areas may also have higher exposure to air pollution, different socioeconomic conditions, or lifestyle factors such as smoking rates, which makes it difficult to isolate the effects of noise itself.

Another challenge is the limited accuracy of exposure estimates. Noise maps, often used in observational studies, provide simulated outdoor levels that may not reflect what people actually hear indoors, and they typically cannot distinguish between different types of noise, such as low-frequency truck noise versus passenger-car noise. Finally, observational research is not well-suited to investigate novel mechanisms or interventions that are not yet widely implemented (e.g., electric vehicle noise) or to systematically prototype acoustic solutions before deployment.

The second approach, *experimental research*, addresses many of these limitations by studying human responses under controlled laboratory conditions [29]. In such experiments, the acoustic environment and potential confounders can be carefully controlled, enabling precise manipulations and a wide range of outcome measures, from physiological indicators such as heart rate and blood pressure to cognitive performance and subjective ratings. Besides applications in noise research, laboratory *listening experiments* have also gained popularity for subjective ratings of sound quality in product development, for example, in the automotive or home appliances industry, where participants evaluate the perceived quality of sounds [30]. In this context, these subjective experiments are also referred to as *jury tests*. The main drawback of experimental research, however, is that laboratory results may not always generalize to real-world situations, and experiments with sufficient statistical power, i.e., a large number of participants, require substantial time and resources.

A central challenge in laboratory noise research lies in reproducing realistic acoustic environments. While simple setups for fundamental noise exposure research may be conducted using broadband noise presented over a single loudspeaker, more sophisticated studies aim to accurately recreate complex acoustic scenarios such as road traffic noise in residential settings. The sound presented to participants during an experiment is referred to as *stimulus* and can either be based on real-life recordings of the acoustic scenario of interest, or be the result of a computational simulation that has been made audible, a process referred to as *auralization* [31]. Advances in computational power have made it possible to auralize simulated complex acoustic environments as stimuli that can be reproduced using a large number of loudspeakers or 3D-tracked headphones, in some cases even in real time. These approaches create new opportunities to study human responses to noise under controlled yet realistic conditions. At the same time, they raise important methodological questions: Which acoustic features must be reproduced to ensure *ecological validity*, i.e., that the laboratory conditions realistically reflect everyday listening situations? Is it sufficient to capture overall sound levels, or must spatial characteristics, such as the perception of vehicle movement, also be rendered accurately? How can we validate that a given reproduction method is "good enough" for the purpose of studying noise effects?

These questions form the foundation of the present thesis. The following section outlines how this work addresses the challenges of designing, validating, and applying controlled yet ecologically meaningful listening experiments to evaluate human responses to sound, specifically in the context of electric vehicle noise.

1.4 Scope of the Thesis

Conducting laboratory listening experiments to capture human responses in complex acoustic environments with high ecological validity requires expertise in two distinct fields. On the one hand, the creation and validation of virtual acoustic environments demands deep knowledge in acoustic engineering, particularly when the scenarios of interest cannot be addressed with off-the-shelf solutions. On the other hand, designing human subject studies, measuring behavioral and physiological responses, and conducting robust statistical analyses often require expertise that researchers primarily trained in engineering acoustics may lack.

As a consequence, many existing studies focus methodologically on one side of this spectrum. Some noise exposure studies apply advanced study designs with large participant samples and sophisticated response measures, yet rely on relatively basic acoustic reproductions. Others employ highly refined auralization techniques, but combine them with simplified study designs. Both approaches are valid and can provide valuable insights when interpreted carefully, yet their limitations are evident. Integrating both fields promises higher overall quality and potentially more sensitive outcomes, which form the central motivation for the present work.

The scope of this thesis is therefore to establish a general methodological framework for conducting listening experiments in virtual acoustic environments and to demonstrate its potential through a focused application to electric vehicle noise. This work integrates both ends of the spectrum, from advanced auralization methods to the quantification of human responses, thereby linking precise acoustic modeling with experimentally validated perceptual, behavioral, and physiological outcomes.

The methodological focus lies on three areas. First, it addresses the creation of custom auralization solutions tailored to specific experimental requirements. Second, it examines how such virtual environments can be validated and what level of accuracy is required to obtain meaningful results. Third, it integrates a broad range of outcome measures, including subjective assessments, physiological indicators, and behavioral responses. Together, these elements define a methodological framework for evaluating human responses to sound across different areas of noise research.

Electric vehicle noise serves as the central case study through which the framework is applied and tested. This application not only yields new insights into how humans perceive, localize, and respond to different AVAS signals but also demonstrates how auralization can be used as a tool for evidence-based design. By enabling controlled investigations before new acoustic solutions are introduced into everyday environments, the application highlights the role of virtual acoustics as both a research method and a prototyping tool. In addition to its scientific contribution, the methodology offers practical value to policymakers and industry stakeholders by providing a systematic approach to assessing acoustic interventions before large-scale implementation. The acoustic perspective adopted throughout this application is that of external observers, such as pedestrians and nearby residents. The interior soundscape experienced by vehicle occupants is not addressed.

Although the general framework outlined in this thesis can be extended to other contexts, the present work focuses on a specific application. It does not provide a universal manual for listening experiments nor a comprehensive study of all possible electric vehicle noise effects. Instead, it presents a focused research program that develops and validates a specific auralization implementation and applies it across a set of targeted studies. Finally, the scope of this work does not extend to long-term epidemiological studies, comprehensive reviews of all available auralization methods, or transport noise sources beyond road traffic. These areas are acknowledged but lie outside the present focus.

1.5 Thesis Outline

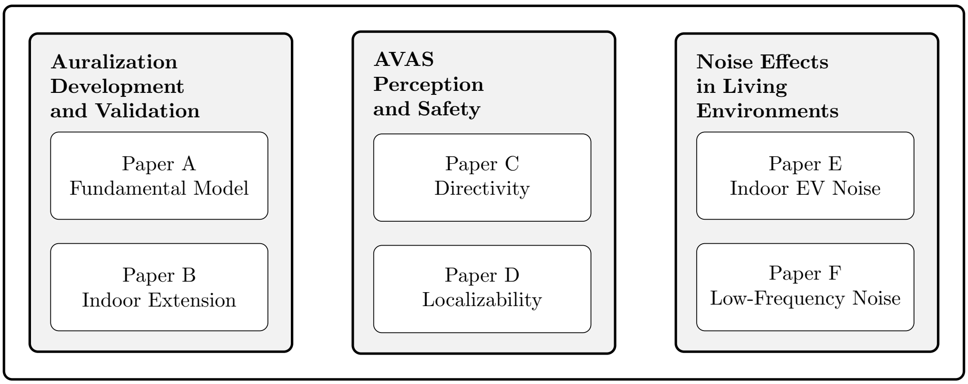


Figure 1.1: Thematic clusters of the included papers.

This thesis is organized into seven chapters. After introducing the motivation, context, and research objectives in Chapter 1, the thesis proceeds from a broad methodological perspective on virtual acoustic environments and human response research to a focused application of these methods to electric vehicle noise.

Chapter 2 introduces the principles of creating virtual acoustic environments through auralization. It reviews the main stages of the auralization process, including source signal generation, radiation, propagation, encoding, and reproduction, as well as requirements for numerical and perceptual validation. Together, these elements define a methodological framework in a broad sense, that is, a structured set of concepts, models, and techniques that can be combined to address diverse research questions.

Chapter 3 complements this by reviewing methods for measuring the human response to sound. It covers subjective, physiological, and behavioral response measures, along with general considerations on study design and statistical evaluation. Like the preceding chapter, it contributes to the overall methodological framework by outlining established approaches and constraints rather than describing a specific implementation.

Building on this foundation, Chapter 4 presents a concrete realization of the framework outlined in Chapters 2 and 3, tailored to the study of electric vehicle noise. The chapter describes the implementation of an electric vehicle auralization toolchain, including reference measurements, source modeling, radiation directivity, propagation, and reproduction. It further details the applied validation strategies and the design of the listening experiments used to evaluate human responses to EV noise under controlled conditions.

Chapter 5 provides an overview of the six papers included in the thesis, situating each contribution within the overall research objectives. The included papers can be grouped into three thematic clusters, as illustrated in Figure 1.1. The first cluster concerns model development and validation and comprises a general auralization model for electric vehicle noise in outdoor conditions (Paper A), and its extension to indoor scenarios (Paper B). The second cluster addresses perception and safety and includes studies on vehicle localizability (Paper D) and the influence of AVAS directivity on speed perception (Paper C). The third cluster focuses on broader human response evaluation and covers the effects of electric vehicle noise in living environments on attention, annoyance, workload, and electrodermal activity (Paper E), as well as a related experiment on low-frequency noise with different temporal structures (Paper F). While the latter is not specific to electric vehicles, it strengthens the general methodology presented in this thesis.

Chapter 6 synthesizes the findings across the different studies and reflects on methodological insights gained throughout the work. It discusses implications for electric vehicle noise research, broader applications of virtual acoustic environments in human response studies, and remaining limitations and open questions.

Finally, Chapter 7 summarizes the main conclusions of the thesis and outlines their implications for electric vehicle sound design, methodological practice, and future research directions.

CHAPTER 2

Creating Virtual Acoustic Environments Through Auralization

This chapter introduces auralization as a broad methodological framework for creating virtual acoustic environments. Rather than describing a single implementation, it summarizes the main building blocks that can be combined and adapted for applications ranging from transportation noise to room acoustics and product sound design. The chapter first clarifies the term auralization and its scope (Section 2.1), then outlines the typical stages of an auralization chain (Section 2.2). The subsequent sections review common approaches for source signal generation (Section 2.3), radiation (Section 2.4), propagation (Section 2.5), encoding (Section 2.6), and reproduction (Section 2.7). Finally, general requirements and validation strategies are discussed (Section 2.8), and existing auralization tools are briefly reviewed (Section 2.9). Together, these elements define the methodological space from which the later thesis chapters select and combine methods for the specific application considered in this work.

2.1 The Term Auralization

The process of creating virtual representations of acoustic scenes is often referred to as auralization. The term is often understood as an acoustic parallel to visualization, in which data are made perceptible through visual representation, and was first coined by Kleiner *et al.* [32]. At that time, it was defined as: "Auralization is the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space." In other words, the term originally described the process of making room acoustic simulations audible. Behind this concept lies the fundamental insight that numerical acoustic parameters alone, such as *clarity*, *definition*, or *gain* in room acoustics, may not fully capture how humans perceive the quality of an acoustic environment. Auralization was therefore introduced as a complementary approach: instead of reducing the acoustic experience to a set of numbers, it provides a perceptually relevant, audible impression of a simulated space or scenario.

The analogy to visualization is therefore only partly accurate. Whereas visualization can be applied to virtually any data type (e.g., temperature, velocity, or pressure), auralization refers specifically to the auditory rendering of acoustic models. The broader counterpart to visualization is sonification, which refers to the mapping of arbitrary data into sound for analysis, monitoring, or communication, such as Geiger counter clicks [33]. Another term that has gained prominence in recent engineering and simulation discourse is the digital twin, which is a virtual replica of a physical system that mirrors its state and behavior. First introduced in aerospace engineering [34], digital twins typically cover multiple dimensions, such as structural, thermal, acoustic, and environmental aspects. The concept applies to a wide range of complex processes. For example, a city's digital twin could be a comprehensive virtual model that integrates traffic flow, pedestrian movement, air quality, and noise pollution in real time [35]. Within this framework, auralization can be conceived as a sensory interface of the digital twin, providing a physically and perceptually accurate rendering of its acoustic dimension, e.g., simulating traffic noise or ambient soundscapes under different urban scenarios.

Today, the concept of auralization has been adopted in multiple acoustic disciplines [31], but its exact scope varies. In some contexts, it denotes only the final step of making a simulation audible [36], while in others, it encompasses the entire toolchain from source signal generation through propagation modeling to the audible rendering [37]. The latter view has gained prominence, especially in applied fields such as drone, railway, or tire noise auralization [38–40], where source signal generation often plays a more critical role than the final rendering. In this thesis, auralization is understood in this broader sense, covering the full process from signal generation, via sound propagation modeling, to the final auditory reproduction.

2.2 Overview of the Auralization Process

To introduce the concept of auralization, it is helpful to recall the basic chain of physical events that occur whenever we hear a sound. A source – for example, a loudspeaker diaphragm or a rolling tire – is excited by a force, causing it to vibrate. This vibration sets the surrounding air particles into motion, producing alternating regions of high and low pressure. The process by which a vibrating source

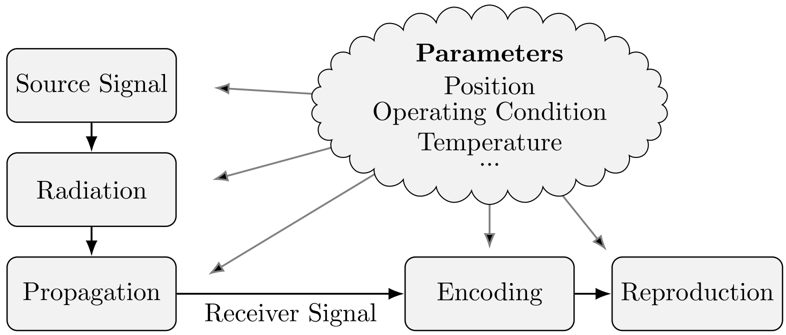


Figure 2.1: Stages of auralization from source to reproduction.

couples energy into the surrounding medium is called *sound radiation* and can be both frequency-dependent and directional. For instance, a loudspeaker may radiate nearly omnidirectionally at low frequencies, whereas at high frequencies it radiates more strongly forward than to the sides.

Once radiated, the resulting pressure fluctuations propagate through the air as sound waves at approximately 343 m/s. Along their path, the waves may interact with surfaces and objects and, depending on their geometry and acoustic impedance, be reflected, diffracted, scattered, or absorbed. Atmospheric conditions such as wind or temperature gradients can further affect propagation. In addition, the sound energy spreads geometrically with distance, causing the intensity at a given point to decrease.

Eventually, the sound waves reach the listener. Depending on the source position, they may reach the two ears at different times and with different amplitudes, providing interaural cues for spatial localization. The complex shape of the outer ear further filters the signal in a direction-dependent way. Together, these *binaural cues* enable the auditory system to estimate both the direction and distance of the source and provide the listener with a spatial impression.

If the sound field is enclosed by reflecting surfaces, as when listening inside a room, the listener perceives not only the direct sound but also a dense sequence of reflections, which eventually merge into a *diffuse field*. By contrast, in the absence of reflections, as in an idealized anechoic environment, one speaks of a *free field*.

Auralization, understood as the process of re-creating this entire chain of events in a virtual acoustic environment, can be structured into the following stages: source signal generation, sound radiation, sound propagation, receiver encoding, and sound reproduction. As shown in Figure 2.1, all of these stages may be influenced by environmental or simulation parameters. Some parameters, such as source position, can be time-varying, while others, such as air temperature, may remain static. Whether parameters are predefined or interactively modified in real time depends on the application. Some frameworks support real-time rendering of the entire acoustic scene, others update only specific stages (e.g., the encoding stage to account for head movements), while still others rely solely on precomputed signals with no real-time interaction. Section 2.8 further discusses requirements for real-time computation and interactivity.

While all processes in Figure 2.1 are included in some form in any auralization framework, their categorization can vary. Some authors treat encoding and reproduction as a single stage, while others distinguish between "source-based" and "receiver-based" auralization [41]. Source-based auralization synthesizes sound at the source and simulates its propagation through the environment, whereas receiver-based auralization synthesizes the sound directly at the receiver position. Even in the latter case, however, the acoustic effects of source signal, radiation, and propagation are implicitly contained in the synthesized signal, though not explicitly modeled as separate stages.

The following sections provide an overview of common implementations of these stages (Sections 2.3 to 2.7), discuss framework requirements and validation strategies (Section 2.8), and review existing end-to-end auralization solutions (Section 2.9).

2.3 Source Signal Generation

In a strict physical sense, the *source signal* describes how a sound source excites the surrounding medium, for example, in the form of a surface velocity or volume velocity distribution that drives the subsequent radiation of sound pressure. In the context of auralization frameworks, however, the term is more commonly used in a pragmatic sense to denote an audio signal representing the source prior to radiation and propagation modeling. Such a signal may already be a pressure waveform obtained from recordings or synthesis, which is then modified in the radiation stage to account for directivity or operating conditions. In the following, the term *source signal* is therefore used in this broader, pragmatic sense, referring to a time-domain signal that characterizes the source independently of the propagation environment.

The nature of such a source signal can vary widely, from a person speaking in a room to a rolling tire or an aircraft engine, and the complexity of source signal modeling reflects this diversity. In room acoustic simulations, simple anechoic speech or instrument recordings are often sufficient, since the primary interest is the perception of the room rather than the source itself. In noise-related applications, by contrast, generating the source signal often represents the primary challenge in the auralization framework, as it must capture non-stationary and operating-condition-dependent sound [42].

2.3.1 Recording-based Models

A straightforward method for generating the source signal is to, if possible, record clean, anechoic signals of the source under the conditions of interest. For example, when studying the human response to heat pump noise, one could place a heat pump in an anechoic chamber, record its radiated sound under different operating conditions, and use those recordings as input to subsequent stages of the auralization chain [43]. Such recordings typically yield highly plausible results for stationary conditions, but they are limited to the specific source type and operating conditions captured. Because the remainder of the virtual environment is flexible, a lack of adaptability at the source stage can be restrictive, particularly for time-varying scenarios such as accelerating vehicles. Moreover, it is not possible to evaluate acoustic designs that have not yet been implemented in real-world settings using such a purely recording-based approach.

2.3.2 Physically Detailed Models

At the other extreme, acoustic source signal generation can rely on physically detailed models, for example, numerical simulations of tire–road interaction [44] or turbulent flow in jet engines [45]. Such approaches are well established in engineering research and are particularly valuable when investigating design solutions that do not yet exist or when isolating the influence of specific physical parameters.

The applicability of these models to auralization-based listening experiments, however, depends on the intended use case. Physically detailed simulations are often computationally demanding, which can limit their practicality when many source variations, operating conditions, or dynamic scenarios are required. In the context of road traffic noise auralization, modeling fine-grained mechanical details may not contribute proportionally to perceptual outcomes unless the research question explicitly targets those components, such as comparisons between different motor designs.

While it is in principle possible to precompute source signals for specific operating conditions, this approach can reduce flexibility when adapting the auralization to varying velocities, trajectories, or environments. In addition, the accuracy of highly detailed simulations depends on the availability of precise geometric and material input data, which is often difficult to obtain for real-world sources. As a result, even physically detailed models may not always yield a perceptual match to recorded references without further calibration.

2.3.3 Signal-based Synthesis

Within the spectrum of source signal generation methods, *signal-based synthesis* occupies the middle ground between purely recording-based approaches and computationally demanding physical models. These approaches use empirical reference data – either from recordings or high-resolution simulations – and aim to resynthesize signals under different operating conditions at a lower computational cost than full physical modeling. The term "signal-based" indicates that the final audio signal itself, rather than the underlying generation mechanisms, serves as the reference to be modeled.

For example, a reference might consist of a recording of an accelerating vehicle, together with accurate measurements of the vehicle's speed. A synthesis model can then establish a functional relationship between the recorded sound pressure and vehicle speed, enabling the resynthesis of new passages with arbitrary velocity profiles. Adapted from the field of computer music, several synthesis techniques have found widespread application in auralization:

- *Additive synthesis* superimposes tonal components [46]. For instance, spectral peaks in an in situ recording can be identified and mapped to speed-dependent functions. Although not exact in a physical sense, this method enables flexible resynthesis of tonal signals and can be extended to include amplitude or frequency modulation. Its main limitation is the difficulty of reproducing broadband noise components.

- *Subtractive synthesis* starts from a broadband noise signal and applies filters derived from recordings, which can be interpolated to represent different operating conditions. This approach is efficient for broadband noise, but less suited for tonal or transient components. Moreover, filter functions are often less intuitive to manipulate than the simple frequency mappings used in additive synthesis.

- *Granular* or *sample-based synthesis* reassembles short excerpts of recorded sound (grains or longer segments) into new sequences, making it possible to capture non-stationary signals. These grains can also be physically motivated, for example, by extracting individual strokes from a combustion engine and reassembling them at different velocities [47, 48]. Granular synthesis can reproduce tonal, noisy, and impulsive components, but constructing flexible models is often less straightforward than with additive or subtractive synthesis.

- *Spectral modeling synthesis* decomposes a signal into deterministic sinusoidal components and a stochastic residual [37]. This separation enables independent manipulation of tonal and noisy components, making spectral modeling synthesis particularly suitable for sources that exhibit both harmonic and broadband features, such as engines. However, its ability to capture strongly transient signals is limited.

Signal-based synthesis, therefore, provides a flexible and computationally efficient alternative to recordings or detailed physical models. In environmental acoustics, this principle is further extended in standardized parametric approaches, as described below.

2.3.4 Standardized Parametric Models

In environmental noise research, several *standardized parametric models* are in widespread use, including Harmonoise [49], Nord2000 [50], and the more recent CNOSSOS-EU framework [5]. These models were originally developed for regulatory and engineering-oriented noise prediction rather than for auralization. They typically describe sources in terms of sound power levels or third-octave band emission spectra as functions of source category, speed, and operating condition.

Although they do not produce time-domain audio signals directly, their spectral output can serve as a foundation for synthesizing source signals. A common strategy is to use predicted spectra to filter broadband noise, thereby generating signals with the correct spectral characteristics. If the models specify tonal components, these can be added explicitly. In this way, standardized models provide physically justified and widely accepted references for source strength and spectral shape, ensuring comparability across studies and consistency with regulatory frameworks.

However, because these models were not designed for perceptual evaluation, they have inherent limitations. In particular, they lack temporal fine structure and transient behavior, and they are restricted to a limited set of standard source categories rather than enabling detailed modeling of individual vehicle types. For this reason, standardized models are best regarded as complementary to the signal-based synthesis methods discussed above: they provide reliable spectral data that can guide or constrain the generation of more plausible and flexible source signals.

2.3.5 Machine Learning Approaches

In recent years, *machine learning* methods have gained increasing attention in acoustics [51], with applications ranging from environmental noise prediction [52, 53] to room acoustic auralization [54, 55]. Given the success of neural networks in speech and music synthesis, where neural vocoders, autoregressive models, and more recently diffusion-based methods have achieved remarkable audio quality [56], their use for source signal generation in auralization appears promising.

So far, however, applications in noise-related auralization remain sparse. This is likely due to several challenges: suitable datasets with sufficient coverage of operating conditions are rarely available, large amounts of labeled training data are typically required, and trained models often struggle to generalize beyond the conditions seen during training.

Despite these obstacles, data-driven methods offer attractive opportunities. They can capture complex, nonlinear source characteristics that are difficult to describe analytically and may enable end-to-end frameworks in which physical parameters, such as vehicle speed, are directly mapped to synthesized audio. Important open questions include the controllability of such models, their interpretability, and their robustness across different scenarios. With ongoing progress in generative audio modeling [57, 58], machine learning approaches are likely to evolve into an increasingly relevant complement to the recording-based, parametric, and standardized methods discussed above.

2.3.6 Summary

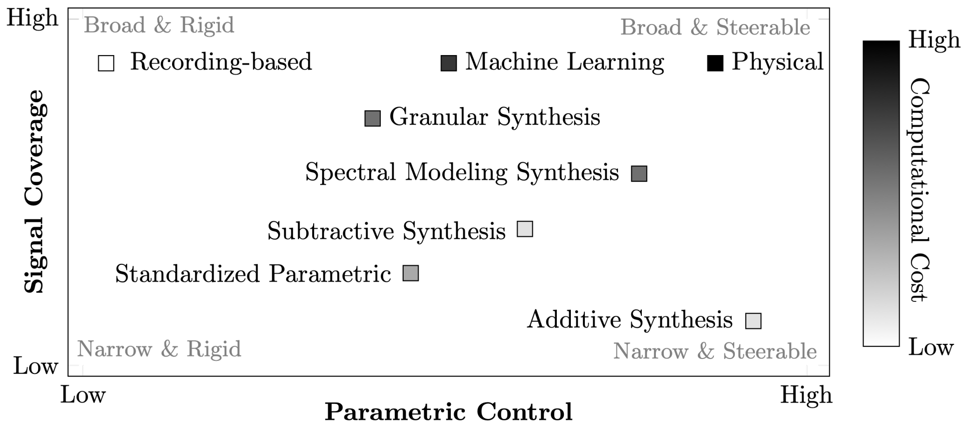


Figure 2.2: Source signal generation methods positioned by parametric control and signal coverage. Marker shading indicates approximate computational cost. Note that the values are only illustrative, as the exact levels of parametric control, signal coverage, and computational cost depend on the implementation and application of each method.

In summary, the choice of source signal generation method depends strongly on the application: simple recordings may suffice for room acoustics, while engineering or environmental studies often require more flexible or parametric synthesis methods. Figure 2.2 summarizes the methods discussed above in terms of three dimensions: *parametric control*, i.e., how easily a model can be tuned to new operating conditions; *signal coverage*, i.e., the range of signal types that can be represented; and estimated *computational cost*. These values should be understood as illustrative rather than as a universal ranking, since the exact position of each method depends on its implementation and the intended application.

A further aspect, not explicitly shown in the figure, is the demand for reference data, which tends to correlate with signal coverage. Methods such as standardized parametric models or signal-based synthesis require only limited reference data for calibration, whereas recording-based approaches, machine learning models, or detailed physical simulations demand extensive input data: either recordings of all relevant operating conditions, large labeled training sets, or detailed material and environmental parameters.

Regardless of the chosen method, the generated source signal serves as input to the next stage of the auralization process, sound radiation, in which its spatially and frequency-dependent emission into the environment is determined, as described in the following section.

2.4 Radiation

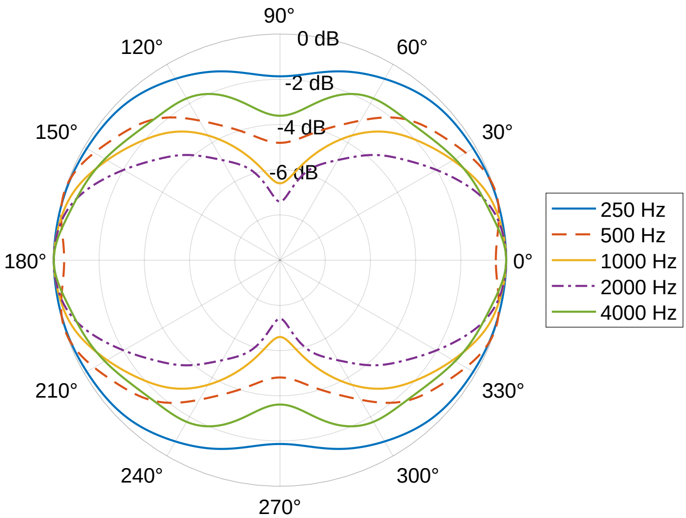


Figure 2.3: Exemplary normalized horizontal tire radiation directivity, based on measurements performed in [59] and interpolated using spherical harmonics.

Once a source signal is defined, its radiation into the acoustic environment must be modeled. Thereby, the *radiation directivity* describes the frequency- and direction-dependent distribution of acoustic energy, which may also vary with the operating condition of the source. For instance, a rolling tire may radiate low-frequency sound almost omnidirectionally, while higher-frequency components are concentrated in the forward driving direction (see Figure 2.3).

As with source signal generation, the complexity of radiation modeling varies widely. At the simplest level, sources may be approximated as omnidirectional radiators (monopoles), which enables straightforward propagation modeling but neglects angular dependence. At the other extreme, radiation characteristics can be obtained from numerical simulations, for example, using the *Boundary Element Method (BEM)* [60]. Such simulations provide high spatial resolution but are computationally demanding, particularly at mid- to high frequencies.

Experimental approaches represent a common middle ground. Directivity can be measured using microphone arrays in anechoic chambers [61], or, where controlled environments are not available, through in-situ setups [62] or measurements with limited angular resolution [59]. The resulting data can be interpolated to cover missing directions or operating conditions.

Between purely measurement-based and purely model-based strategies, hybrid approaches have emerged. These combine partial experimental data with computational models to extend spatial or frequency coverage. For example, a measured directivity pattern at a limited number of angles may be expanded using a low-order analytical model [63]. Such hybridization balances accuracy with feasibility, and is increasingly relevant when full high-resolution measurements or simulations are impractical. Finally, standardized noise assessment frameworks such as CNOSSOS-EU [5] define simplified radiation models, often based on angular correction terms, that enable consistent predictions across large-scale environmental studies.

The perceptual relevance of directivity is often lower than that of the source signal itself, and the necessity of including it in an auralization depends strongly on the source type and application. The required complexity is also frequency-dependent: most sources exhibit relatively simple low-frequency radiation patterns, whereas high-frequency radiation tends to be much more complex. In free-field simulations with stationary sources and receivers, one could even argue that directivity acts only as a static filter, which can be incorporated directly into the source signal. By contrast, when either the source or receiver is in motion, radiation directivity becomes more critical for realistic rendering [64, 65]. In room environments, directivity can further influence the reverberant field, which may become particularly relevant for highly directional sources such as certain musical instruments [66, 67].

In addition to directivity, the apparent spatial extent of a source may be relevant. A monopole model not only assumes omnidirectional radiation but also a point-like source. In reality, extended sources such as a car chassis may scatter sound so that the listener perceives energy arriving from a broader region. This aspect of radiation is not yet fully understood and is rarely addressed in practical auralization frameworks.

2.4.1 Spherical Harmonic Representation

In addition to the computational and experimental challenges discussed above, incorporating complex radiation directivity into auralization may also pose difficulties for data handling. If directivity is defined as complex pressure values on an arbitrary measurement or simulation grid, the dataset can become large, irregular, and cumbersome to interpolate.

A practical way to address this is through a *Spherical Harmonic (SH)* expansion. In this approach, which is used later in this thesis and hence covered in greater detail here, pressure data measured or simulated on a sphere around the source are

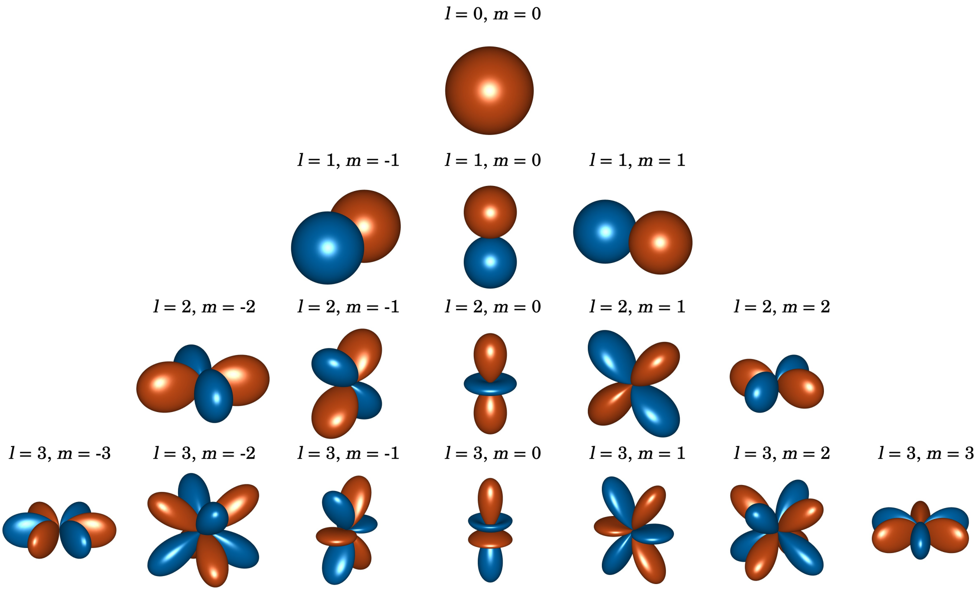


Figure 2.4: First few real spherical harmonics Y_l^m of order l and degree m. The surface radius encodes magnitude, and color indicates positive (orange) and negative (blue) values.

decomposed into spherical harmonic coefficients, $W_l^m(f)$. These coefficients describe how the frequency-dependent pressure distribution, $p(\phi, \theta, f)$, can be reconstructed as a weighted sum of spherical harmonics basis functions, $Y_l^m(\phi, \theta)$, of order l and degree m, defined over azimuth ϕ and colatitude θ (Figure 2.4) following [68]:

$$p(\phi, \theta, f) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} W_l^m(f)\, Y_l^m(\phi, \theta). \tag{2.1}$$

For spatially discrete pressure observation points, Equation (2.1) can be written in vector-matrix form as $\mathbf{p} = \mathbf{Y}\,\mathbf{W}$, where $\mathbf{p}$ contains the measured pressures and $\mathbf{Y}$ the evaluated basis functions. The SH coefficient matrix $\mathbf{W}$ can then be estimated in the frequency domain using a least-squares solution [69].

The main benefit of this representation is that complex directivities can be approximated with a limited SH order, reducing data size while preserving essential features. At the same time, the SH basis inherently provides interpolation, allowing the pressure to be evaluated at arbitrary angles rather than only at the measured or simulated grid points. In this way, radiation patterns can be described using a compact, universal set of coefficients rather than irregular pressure maps.

While other multipole-based representations, such as expansions in monopole, dipole, or plane-wave bases, have also been used for acoustic applications [68, 70], spherical harmonics have become the most common framework for radiation directivities because they provide a complete basis for functions defined on the sphere and, in practice, allow many natural sound sources to be represented accurately with only a few low-order terms.

Nevertheless, SH representations also have limitations. In particular, they assume that the source is located at the center of the expansion sphere, which may not be valid for extended or distributed sources. This issue becomes relevant in the application discussed in Section 4.1.3.

2.5 Propagation

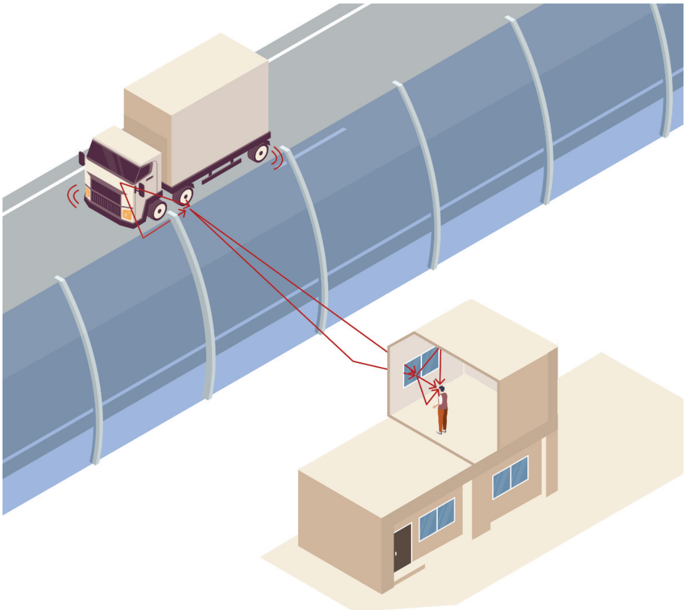


Figure 2.5: Illustration of typical road-traffic noise propagation path including ground reflections, noise barrier diffraction, window transmission, and room reflections.

Propagation describes how sound travels from a source to a receiver through the surrounding medium and environment. In the context of auralization, this process is often represented by an impulse response, $h(t)$, which specifies how a short impulse emitted at the source arrives at the receiver. Its frequency-domain equivalent is the transfer function, $H(f)$. Under the assumption of a *Linear and Time-Invariant (LTI)* system, either of these functions fully characterizes the propagation path, meaning that once the transfer function is known, the received signal can be obtained by filtering the source signal accordingly. Figure 2.5 illustrates a typical road-traffic noise propagation path, from an outdoor source via ground reflections, diffraction over a noise barrier, transmission through a window, and reflections within an indoor environment.

In the simplest case of free-field propagation, the transfer function reduces to a time delay, a geometric distance attenuation, and, at longer distances, frequency-dependent air absorption. In more realistic scenarios, however, additional phenomena such as reflections, diffraction, and scattering strongly shape the received signal.

The relative importance of these effects depends on the acoustic environment. In enclosed spaces, reflections and reverberation dominate the auditory impression and are the central focus of room acoustic auralizations. In outdoor environments, long propagation paths and inhomogeneous atmospheric conditions pose unique modeling challenges. Finally, when either source or receiver is in motion, the propagation path becomes time-varying and introduces additional perceptual cues such as *Doppler shifts*.

The following sections introduce common modeling approaches for indoor and outdoor sound propagation, as well as the effects of source and receiver motion.

2.5.1 Indoor Sound Propagation

Indoor sound propagation includes both airborne sound within enclosed spaces, which is covered by the field of *room acoustics*, and transmission through building structures, as described by *building acoustics*.

Room Acoustics

Modeling sound propagation in rooms has been a central focus of acoustics research for decades, from the design of concert halls to the evaluation of classrooms and offices [71]. In this context, a wide range of simulation methods has been developed, each with its own balance of accuracy and computational efficiency [72]. These can be broadly grouped into *geometrical*, *energy-based*, and *wave-based* approaches.

Geometrical approaches explicitly follow the propagation paths between the source and the receiver. The simplest example is the *image source method*, which represents reflections by introducing virtual sources. This allows straightforward inclusion of source and receiver directivity as well as frequency-dependent absorption coefficients. However, scattering is usually neglected, and the computational effort increases rapidly with the number of reflections. More scalable are *ray-based methods*, such as ray tracing or beam tracing, where a large number of rays are emitted from the source and tracked through the room as they reflect, scatter, or are absorbed at surfaces. Modern Graphics Processing Units (GPUs) can accelerate these simulations considerably. Nevertheless, the resulting impulse responses may appear sparse, since reflections are only captured if a sufficient number of rays intersect with the receiver. This stochastic sampling makes early reflections particularly difficult to model reliably, as, unless extremely high ray densities are used, the direct sound and first-order reflections may be missing or inaccurately represented, leading to audible artefacts [31].

Energy-based methods compute the exchange of acoustic energy, for example, between surface elements. The most common is the *radiosity method*, in which each surface patch reflects and redistributes energy to all others in an iterative process until equilibrium is reached [73]. This efficiently models diffuse reflections and the build-up of late reverberation, producing smoother decay characteristics than ray tracing. However, because radiosity operates on energy rather than phase, it cannot reproduce interference effects or resolve precise early reflection paths, and is therefore often combined with geometrical methods.

Wave-based methods, such as the *Finite Element Method (FEM)*, *Boundary Element Method (BEM)*, or *Finite-Difference Time-Domain (FDTD)*, simulate sound propagation by directly solving the acoustic wave equation. For very simple geometries, analytical solutions can be derived using separation of variables with appropriate boundary conditions (e.g., Neumann for rigid walls or Dirichlet for pressure-release surfaces) [70]. These approaches provide high physical accuracy, particularly at low frequencies, but are computationally demanding and are typically restricted to simple geometries or narrow frequency ranges.

In practice, *hybrid approaches* are often employed, where wave-based methods are used at low frequencies and combined with geometrical or surface-based methods at mid-to-high frequencies to achieve broadband coverage at a feasible cost.

Building Acoustics

A special case of indoor sound propagation arises when part of the sound field travels through building structures rather than solely through air. Examples include environmental noise penetrating through facades into dwellings, or noise from neighboring apartments being transmitted through walls or floors.

Modeling such phenomena requires tools from building acoustics [74]. Detailed physical simulations of structure-borne transmission are possible but computationally expensive for full buildings. More commonly, *analytical and semi-empirical models*, such as those described in the ISO 12354 series [75], are applied to estimate sound insulation based on material properties and construction details. In auralization frameworks, these effects are often simplified to frequency-dependent filter functions representing typical wall or facade constructions. However, there are also mechanisms, such as the *coincidence effect* (i.e., the fact that sound transmission through a facade or window varies with the angle of incidence), that may be perceptually relevant in some scenarios but cannot be represented by a simple static filter function.

In addition to direct transmission, *flanking paths*, where sound travels indirectly through structural connections, can contribute significantly to the overall propagation path for some applications. However, flanking is often neglected in simplified models, and accounting for these effects remains an important challenge for perceptually accurate auralizations of indoor environments.

2.5.2 Outdoor Sound Propagation

Modeling outdoor sound propagation poses distinct challenges compared with indoor environments. In the absence of enclosing boundaries, long source–receiver distances and atmospheric inhomogeneity become important. Thereby, the effects relevant for an auralization framework depend strongly on the propagation distance. While the sound from a nearby e-scooter may primarily be influenced by ground reflections and simple distance attenuation, an aircraft at several kilometers distance is subject to pronounced atmospheric absorption, refraction, and turbulence effects that need to be included in the propagation model.

At the most basic level, outdoor propagation can be represented by geometric spreading combined with frequency-dependent *air absorption*, often modeled according to ISO 9613-1 [76] as a distance-dependent filter. For moving sources, the filter parameters may be adapted dynamically as the source–receiver distance changes. In addition, interaction with the ground plays an important role: depending on source and receiver heights, ground impedance, and frequency, interference between direct and reflected waves can cause spectral coloration.

At longer propagation distances, *atmospheric effects* become increasingly relevant. Gradients in temperature or wind velocity can refract sound rays, leading to focusing or shadowing effects. Small-scale turbulence causes stochastic fluctuations in amplitude and phase, which may be described using models such as the von Kármán spectrum [77]. These effects are particularly important in the auralization of aircraft or other long-range sources, where they strongly contribute to perceived realism.

Another important case in outdoor environments is the presence of obstacles such as *noise barriers* or building edges. Here, diffraction effects dominate and can be modeled analytically, e.g., using Fresnel approximations [78], or numerically, for example, using the boundary element method. In urban environments, additional complexity arises from multiple reflections in street canyons or between building facades.

As in indoor acoustics, a range of outdoor sound propagation modeling approaches exists, from simplified engineering standards (ISO 9613-2 [78], CNOSSOS-EU [5], Nord2000 [50]) to computationally intensive numerical methods. For auralization purposes, simplified filter-bank approaches or hybrid models are often employed to capture the dominant perceptual effects while maintaining computational efficiency.

2.5.3 Movement

When either the source or the receiver is in motion, the sound propagation path becomes time-varying and therefore more complex to model. From the receiver's perspective, movement could in principle be treated only at the reproduction or encoding stage, for example, by adjusting the intensity between the left and right headphones to create a movement impression. However, important physical effects occur before the sound wave reaches the receiver, making source movement an essential part of propagation modeling. The most prominent example is the *Doppler effect*, where relative motion leads to a compression or expansion of the wavelength and a corresponding shift in perceived frequency. In addition, movement modifies the overall intensity through distance-dependent attenuation and alters the balance between direct and reverberant sound fields.

Analytical treatments of moving sound sources have been studied extensively [70, 79]. For simple monopole sources, the Doppler shift can be described in closed form. More complex sources, however, pose greater challenges since their directivity interacts with motion [80, 81], and multiple radiation components may be shifted differently.

In practical auralization frameworks, simpler engineering solutions are often employed. One approach is to compute transfer functions for a sequence of discrete source positions and apply these to the corresponding source signal segments. This produces a *time-varying impulse response* (sometimes also referred to as a *moving Green's function* [82]), which naturally captures Doppler shifts and distance-dependent attenuation, but at the expense of increased computational cost. Efficient implementations typically rely on interpolation between precomputed impulse responses or on partitioned convolution to manage the computational load.

Simpler approximations treat movement only at the reproduction stage. For example, in binaural or loudspeaker-based auralization, source motion can be simulated by interpolating between panning directions without explicitly modeling the underlying propagation. Such approaches omit Doppler shifts but may be perceptually sufficient in many room-acoustic applications where sources and listeners move slowly.

Listener motion introduces similar challenges. A moving receiver experiences Doppler shifts relative to each source and a continuously changing balance between direct and reflected sound. These effects can be incorporated using the same methods as for source motion, but at higher computational cost because the entire sound field must be updated.

Overall, the importance of including movement effects depends strongly on the application. For environmental and transportation noise, Doppler shifts and time-varying propagation are often critical perceptual cues, whereas in indoor simulations with slow-moving sources, simplified rendering strategies may be adequate.

2.6 Encoding

In the context of auralization frameworks, *encoding* refers to the step in which all source contributions arriving at the receiver position after propagation modeling are transformed into a signal format that represents the virtual sound field. Encoding does not itself produce audible sound, but defines the representation that the reproduction stage will later render to the listener. This encoded representation determines how spatial information is preserved and made available for playback. Depending on the application, encoding may range from a single summed signal to highly flexible spatial formats.

The simplest case is *monaural encoding*, in which all source contributions are added to form one single audio channel. This may be sufficient when spatial impressions are irrelevant or when reproduction is later performed with a single loudspeaker.

For most auralization applications, however, spatial cues are required, leading to *binaural encoding*. In this case, direction-dependent cues such as *Interaural Time Differences (ITDs)*, *Interaural Level Differences (ILDs)*, and source-position-dependent spectral colorations caused by the listener's head, torso, and pinnae are incorporated into the signal. In its basic form, the sound arriving at the receiver is filtered with a *Head-Related Transfer Function (HRTF)*, producing left- and right-ear signals. When room acoustics are included, this can be extended to *Binaural Room Impulse Responses (BRIRs)*, which combine source, room, and listener characteristics. Since HRTFs and BRIRs are direction-dependent, this approach assumes knowledge of the incidence direction of each sound component. For simulated or measured multichannel room responses, methods such as the *Spatial Decomposition Method (SDM)* [83] decompose the response into reflections from specific directions, which can then be processed individually using the corresponding HRTFs.

Beyond direct HRTF or BRIR filtering, more advanced approaches treat binaural synthesis as a *Multiple-Input Multiple-Output (MIMO)* problem. In this case, a measured or simulated sound field is first sampled on a surface or volumetric grid around the receiver position. The transfer functions between these sampled points (inputs) and the listener's left and right ear signals (outputs) are then estimated, typically via a least-squares fit to HRTF measurements [36]. This produces binaural signals that can flexibly incorporate spatial detail from the sampled field. While powerful, such methods are computationally demanding and less common in real-time auralization frameworks.

A more reproduction-independent option is provided by *Ambisonics*, or its extension to *Higher-Order Ambisonics (HOA)* [84]. Here, the sound field at the receiver is encoded as a set of spherical harmonics coefficients in a standardized format. This representation is not tied to any specific playback system and can later be decoded for headphones, loudspeaker arrays, or other setups. A key advantage is that head movements can be applied flexibly at the reproduction stage, since HRTFs are not embedded in the encoded signal. The achievable spatial resolution depends on the Ambisonics order.

Another family of approaches is *channel-based encoding*, which directly assigns audio signals to a fixed number of channels corresponding to loudspeakers. In the simplest case, two loudspeakers enable stereo encoding, in which sources are positioned by intensity or time differences within the stereo panorama. Surround formats such as 5.1 or 7.1 extend this to more loudspeakers. With larger arrays, methods such as *Vector Base Amplitude Panning (VBAP)* [85] allow three-dimensional source positioning by distributing energy across nearby speakers. At the most complex end, *Wave Field Synthesis (WFS)* [86] aims at the physical reconstruction of the sound field using large loudspeaker arrays. Since WFS specifies reproduction signals directly rather than providing an intermediate encoding, it partly blurs the distinction between encoding and reproduction, but is often categorized as a reproduction technique and will therefore be discussed further in the following Section 2.7.

2.7 Reproduction

The final step in an auralization framework is the acoustic reproduction of the encoded signals through a specific set of transducers, either headphones or loudspeakers. Reproduction does not alter the encoded representation itself but renders it as audible sound to the listener. In this sense, reproduction must be matched to the chosen encoding format. For example, binaural signals are typically rendered over headphones, while Ambisonics or channel-based encodings require decoding for loudspeaker arrays. Reproduction also determines how many listeners can be addressed simultaneously, how robust spatial cues remain during head movement, and how much calibration effort is required.

2.7.1 Headphone-Based Reproduction

The simplest case of headphone-based reproduction is the direct playback of binaural signals over a pair of headphones. This approach is widely used in room-acoustic auralizations, where the absolute sound pressure level may be of secondary importance, and the main concern is to convey spatial cues. In such cases, the choice of headphone type (in-ear, on-ear, or over-ear) may be the only relevant factor. A common perceptual issue in binaural reproduction is *internalization*, where sounds are perceived “inside the head” rather than externalized in space. To mitigate this effect, headphone equalization is commonly applied to avoid additional spectral coloration that would otherwise distort the HRTF cues encoded in the binaural signal. Individualized HRTFs can further improve externalization and spatial realism.

For applications where level accuracy is critical, such as evaluations of environmental noise effects, headphone reproduction must be carefully calibrated. In principle, calibration ensures that the sound pressure level at the entrance of the listener’s ear canal matches the intended simulation and can be performed using artificial heads or ear simulators. In practice, however, headphone equalization is considerably more difficult than for loudspeakers [31] since the relevant radiation impedance is that of the ear canal rather than a free-field condition, and large inter-individual differences in ear canal geometry lead to resonances that vary across listeners. Even with digital equalization and artificial ears, uncertainties remain due to mounting conditions, leakage in closed headphones, and other listener-specific factors. As a result, overall level calibration is often considered sufficient for many applications, while more demanding experiments face inherent limitations in achieving a flat or standardized response across listeners.

Headphone reproduction can further incorporate *head tracking*, which dynamically updates the binaural signals according to the listener's orientation. This substantially improves externalization and spatial stability in HRTF-based rendering by preserving the natural relationship between head motion and auditory cues. However, head-tracked playback requires low end-to-end latency, that is, minimal delay between head motion and the corresponding update of the auditory scene. Achieving such responsiveness demands high processing performance and careful real-time implementation.

Despite these possibilities, headphone-based reproduction inevitably reduces the ecological validity of virtual environments, since listeners must wear transducers on or in their ears. In addition, spatial accuracy relies on either generic HRTFs, which may limit localization performance for some individuals, or on individualized HRTF measurements, which require considerable effort. Headphones also only address a single listener at a time. Loudspeaker-based reproduction, by contrast, can offer higher ecological validity, supports multiple listeners, and avoids the need for HRTFs altogether, since participants listen with their own ears.

2.7.2 Loudspeaker-Based Reproduction

Loudspeakers allow listeners to experience sound without wearing transducers, which can enhance the perception of naturalness and ecological validity [31, 87]. This is particularly important in applications that aim to simulate everyday acoustic environments, such as evaluating noise exposure during sleep, where the use of headphones is not feasible. Loudspeakers also permit the reproduction of very low frequencies that extend below the range of most headphones, where the perception involves not only the ears but also the whole body [88].

Depending on the chosen encoding format, the encoded signals must be adapted to the loudspeaker array before reproduction. For example, channel-based encodings, such as stereo or multichannel surround formats, can be directly assigned to the corresponding loudspeakers, while Ambisonic signals require decoding to the specific loudspeaker geometry [89]. The resulting loudspeaker signals are then played back over the array to reconstruct the desired spatial impression at the listener's position.

A special case is binaural reproduction over loudspeakers using crosstalk cancellation [90]. In this approach, two or more loudspeakers are used to deliver binaural signals such that the left signal reaches only the left ear and the right signal only the right ear. This requires cancelling the unwanted contributions of each signal at the contralateral ear. While the method can work well under controlled conditions, such as with two loudspeakers in an anechoic chamber and a fixed listener position, it becomes increasingly difficult to maintain with multiple loudspeakers, in reverberant environments, or when the listener moves.

More advanced reproduction techniques aim not only to reconstruct the pressure at a single listening position but to reproduce an entire sound field over a listening area. The most prominent example is Wave Field Synthesis (WFS), which uses a distribution of secondary loudspeakers to synthesize the desired sound field according to the Kirchhoff–Helmholtz integral [86]. Ideally, this allows listeners to move freely while maintaining an authentic spatial impression and enables simultaneous reproduction for multiple listeners. However, practical limitations arise from the spatial sampling theorem, which requires at least two loudspeakers per wavelength and often restricts the aliasing-free frequency range of most arrays to below approximately 2 kHz. Furthermore, loudspeaker arrays that do not form a closed geometry suffer from truncation effects, which further reduce reproduction accuracy [91].

While equalization of individual loudspeakers is comparatively straightforward in anechoic conditions, the calibration of large arrays is more complex. Each loudspeaker must be matched in frequency response and level, and the entire system must be adjusted to the intended reference. This process is relatively simple in controlled environments but becomes increasingly challenging in reflective rooms, where loudspeaker-room interactions need to be compensated. Another limitation of loudspeaker-based approaches is the existence of a "sweet spot", since most reproduction techniques are optimized for a specific listening area. Ambisonics and WFS can extend this region, but only within limits imposed by array geometry and spatial aliasing.

While the best results for loudspeaker-based reproduction are obtained in anechoic or otherwise controlled laboratory environments, some approaches place loudspeakers in existing rooms. In such cases, the room itself provides the propagation and reverberation characteristics, so that the reproduced signals represent only the direct contributions of the sources. This can be advantageous when the goal is to evaluate auralization in ecologically valid everyday settings, for example, by reproducing traffic noise in a domestic environment.

Overall, headphone and loudspeaker reproduction each offer distinct advantages. Headphones are portable, inexpensive, and easy to deploy, providing consistent spatial cues for a single listener, but they entail inherent calibration difficulties and reduced ecological validity. Loudspeakers, on the other hand, provide more natural listening conditions, support multiple listeners, extend reproduction into the very low-frequency range, and allow for full-body perception, but at the cost of greater technical complexity in large-array calibration and a stronger dependence on the acoustic environment.

2.8 Requirements and Validation

Auralization is always used with the purpose of subjectively or objectively evaluating the human response to an acoustic environment. This could happen informally by listening to simulated room models to judge the perceived quality of a design, or more formally through controlled listening experiments. What distinguishes auralization from other virtual acoustic environments, such as those used in computer games, is the ambition to accurately reflect the underlying simulation. While for entertainment purposes it may be sufficient that a simulated room sounds realistic and matches the visuals, for auralization it is not enough that it merely "sounds good". Instead, it should ideally reproduce the relevant parameters of the simulated environment, ensuring both *technical validity*, in the sense of being numerically correct and free of artifacts, and *perceptual validity*, in the sense that listeners experience the acoustic environment in a way consistent with reality.

Technical and perceptual validity, however, are not necessarily binary attributes. Whether a technical inaccuracy matters depends on its perceptual consequences, and conversely, perceptual validity does not always guarantee sufficient technical correctness. While the ideal outcome of an auralization would be sample-accurate agreement with a real-world reference, this level of accuracy is generally unattainable for complex acoustic environments. In practice, auralization necessarily involves approximations, and the key question becomes which deviations are acceptable for a given application. The tolerable degree of inaccuracy, therefore, depends on the specific purpose and outcome measures of the auralization.

Even though the concrete requirements placed on an auralization framework depend on its intended application, several general aspects can be formulated. Depending on the use case, these may include sufficient numerical accuracy, robust implementation, absence of artifacts, and perceptual fidelity appropriate to the task. In addition, some applications demand real-time operation with interactive listeners or source movements, or flexibility to handle a wide range of source types, scenarios, and operating conditions. Different domains also prioritize requirements differently: in architectural acoustics, perceptual fidelity of room parameters such as reverberation time may be central, whereas in environmental noise studies, correct sound levels and spectral content may be more critical.

To ensure that such application-specific requirements are met, an auralization approach must be *validated* before it can be reliably used for perceptual experiments or design evaluation. Validation can occur at multiple levels. *Component-level validation* examines individual parts of the chain, for example, by comparing a propagation model to reference measurements or analytical solutions. *System-level validation*, in contrast, assesses the complete auralization chain by comparing its output to reference recordings or real-world listening situations. In practice, validation is typically approached from two complementary perspectives: *numerical validation*, which evaluates the technical accuracy of the implementation, and *perceptual validation*, which assesses whether listeners experience the rendered environment as intended. The following two subsections discuss these perspectives in more detail.

2.8.1 Numerical Validation

Numerical validation generally involves comparing the output of an auralization framework, or of individual components within the processing chain, to a *ground truth*. This ground truth can be obtained from real-life measurements, for instance, recordings of a vehicle pass-by, or from analytically computed references, such as a high-resolution numerical model used to validate a simplified parametric approach. In many cases, however, obtaining such a ground truth is difficult or even impossible, for instance, when a design does not yet exist or when extreme operating conditions cannot be measured. Moreover, reference recordings are rarely perfectly defined, as environmental parameters such as reflection coefficients or exact geometry are often only approximated. In such cases, step-by-step or component-level validation can be a practical alternative, allowing individual parts of the auralization chain to be assessed independently. Sensitivity analyses, which examine how variations in input parameters propagate through a simulation, can further help to identify the parameters with the greatest influence on the results [92].

A central question in numerical validation is the choice of appropriate metrics for comparing an auralization against its ground truth. Frequency-domain metrics may include overall sound pressure levels, third-octave band levels, or narrowband spectra. Time-domain metrics can capture temporal variations in level, waveform similarity, or reverberation characteristics. Psychoacoustic metrics such as loudness, sharpness, tonality, or spatial cues [93] are particularly relevant for estimating perceptual differences between an auralization and binaural reference recordings. For binaural signals, additional considerations arise, such as whether to average metrics across channels or to apply auditory models of spatial hearing to better capture perceived spatial differences [94]. Importantly, signals may match in conventional metrics or psychoacoustic measures yet still sound perceptually distinct. Furthermore, a limited set of reference signals can lead to overfitting, i.e., an auralization model may reproduce a specific scenario very accurately but degrade in quality when parameters such as vehicle speed or source position change.

The requirements for numerical validation are highly application-dependent. For example, in transportation noise studies, matching third-octave band levels and capturing dynamic spectral changes may be critical, whereas for office acoustics or architectural spaces, accurate reproduction of time structure and reverberation characteristics may be more important. However, determining which metrics are most relevant to perception cannot be achieved through numerical analysis alone. Informal listening to the generated signals often provides more immediate insight into auralization quality than numerical metrics in isolation.

Despite its central role in method development, model tuning, and the identification of trade-offs between accuracy and computational cost, numerical validation cannot replace perceptual evaluation. Perceptual validation is always required to confirm that listeners experience the rendered signals in a manner consistent with the intended acoustic environment.

2.8.2 Perceptual Validation

Perceptual validation relies on listening experiments and therefore overlaps methodologically with human-response research. In this thesis, however, it is treated as part of the auralization development process, as a prerequisite for subsequent experiments rather than an outcome-focused investigation of human responses, and is therefore described here rather than in Chapter 3. In the context of perceptual validation, several terms are commonly used to describe the perceived quality of auralizations, including *plausible*, *authentic*, *realistic*, and *perceptually transparent*. Clarifying their meaning is important, since they set different expectations for what an auralization should achieve.

The notion of *authenticity* refers to perceptual indistinguishability from a real acoustic event when compared directly to an external reference [87, 95]. In other words, listeners are unable to perceive a difference between the auralization and a reference signal, even when hearing them back-to-back. By contrast, *plausibility* describes whether a simulation corresponds to a listener's expectation of a real event, based on internal references formed through everyday experience [96, 97]. Plausibility thus relies on subjective judgment without an external reference and depends strongly on the listener's prior exposure. In this sense, authenticity implies that all perceptually relevant features of an acoustic environment are reproduced [98], while plausibility requires only those features necessary for a given purpose. The term *realistic* is often used more loosely in the literature and in practice, usually denoting that an auralization creates a convincing impression of reality without distinguishing whether it is authentic or merely plausible. Finally, the concept of *perceptual transparency*, introduced by analogy to audio coding, can be seen as a special case of authenticity: both imply that an auralization is indistinguishable from a reference, but transparency emphasizes the absence of audible artifacts introduced by processing, whereas authenticity refers more broadly to the faithful reproduction of all perceptual features of the real event.

Evaluating plausibility can be relatively straightforward. Informally, it can be assessed by listening to an auralization and judging whether it sounds as expected. More formally, it can be evaluated in listening experiments in which participants rate plausibility on a scale, for instance, from "not plausible at all" to "highly plausible". Such ratings require that listeners have an adequate internal reference. This is often unproblematic for sounds such as combustion engine traffic noise, which most people have experienced extensively, but more difficult for less familiar cases, such as drone noise. In these cases, recruiting expert listeners or providing participants with reference recordings during a familiarization phase can help establish a consistent baseline.

Authenticity, in contrast, always requires external reference recordings for direct comparison. The strictness of this comparison depends on the chosen test method. Sensitive paradigms such as ABX tests, where listeners must identify whether a stimulus (X) matches one of two references (A or B), are designed to detect even subtle differences between signals [99, 100]. However, for many applications, less conservative approaches, such as similarity ratings on a scale, may be sufficient, and established paradigms in audio quality evaluation, such as MUSHRA tests, can also be adapted to compare auralizations with reference signals [101].

Both plausibility and authenticity evaluations face the challenge of defining thresholds for success. For example, should an auralization only be considered authentic if no listener can detect differences in a direct comparison, or is a method still acceptable if a small percentage of expert listeners consistently perceive deviations? Proving similarity is generally more difficult than proving difference, which is why sensory evaluation often uses criteria such as “30% of subjects do not detect a difference” as a practical threshold. Statistical methods, such as significance or equivalence testing, can support these judgments, but the chosen threshold always depends on the intended application.

In practice, achieving perfect authenticity for complex auralizations is rarely possible. Instead, perceptual validation is often conducted at the component level, with different aspects of the auralization targeted for different levels of fidelity. For example, when auralizing a train passage to study the perceptual influence of different track types, it may be critical that the wheel–rail noise is reproduced as authentically as possible, while aerodynamic noise from the wagon, though relevant for overall realism, could be represented only plausibly. This, however, raises its own challenges, since isolating reference recordings for individual components is often impractical—in the train example, obtaining in situ recordings of wheel noise without also capturing track noise would be extremely difficult.

Overall, there is no universal method for perceptual validation of auralizations. Outcomes depend strongly on the application, the availability of reference data, and the listener population. For this reason, it is essential to transparently report the limitations of any auralization framework and to account for these limitations when designing listening experiments and interpreting their results. In combination with numerical validation, which ensures technical correctness, perceptual validation provides the complementary perspective required to confirm that listeners actually experience the rendered signals as intended. Together, the two approaches form a robust validation framework for assessing the quality of auralizations.

2.9 Existing Auralization Tools and Software

A variety of software solutions exist that enable auralization in different contexts. These range from commercial room acoustics packages, through general-purpose multi-physics solvers, to dedicated auralization frameworks developed within the research community. Each of these categories provides valuable functionality, but they differ widely in scope, computational requirements, and suitability for perceptual experiments. The following subsections provide an overview of these three groups, highlighting their typical applications and main limitations in the context of this thesis.

2.9.1 Room-Acoustic Software

For room-acoustic applications, practically all modern commercial tools (e.g., CATT-Acoustic [102], ODEON [103], Treble [104] or RAVEN [105]) include an auralization framework. These typically rely on precomputed or recorded anechoic source signals (e.g., speech or instruments) and allow the definition or import of source radiation directivity. Propagation is then solved via geometrical acoustics, ray tracing, or wave-based approaches. The signal at the receiver is encoded as a two-channel binaural signal, an Ambisonics signal, or is already prepared for a multi-speaker loudspeaker setup. These encoded signals can then either be directly played back in the simulation software, where some products even allow for 360-degree head rotation, while others render static sound files that can be listened to. Some solutions, such as RAVEN, also support real-time rendering and the definition of movement trajectories for the source and receiver. These solutions are very powerful for all room acoustics applications. However, they rely on predefined sound-source signals, which, for applications that deviate from the standard use case, must be precomputed. For example, the sound emitted by a cleaning robot may depend on its speed and operating mode. Using any of these existing software solutions would be efficient for calculating propagation, but the source signal would need to be precomputed in separate software for each trajectory. Additionally, these tools are limited to room acoustics applications, so, for example, the sound transmission from an outdoor source through a closed window cannot be modeled easily.

2.9.2 Multiphysics Numerical Tools

For applications beyond room acoustics, commercial multiphysics modeling tools such as COMSOL Multiphysics, Ansys, or Simcenter can be employed. These environments are highly flexible and allow the coupling of different physical domains, such as structural vibrations, aeroacoustics, and sound transmission through inhomogeneous media. Some already provide rudimentary methods to auralize simulated pressures, enabling a first impression of the sound field. In theory, one could set up a fully physical model of the entire acoustic chain, for example, starting with the turbulent flow and sound emission of a drone flying outdoors, simulating propagation through an inhomogeneous medium, modeling the vibration response of a window, and finally calculating the sound radiation into a room as received by a listener.

While attractive in principle, such an approach faces severe challenges. Solving the full chain over the entire audible frequency range would be computationally prohibitive and extremely time-consuming. Moreover, the accuracy would strongly depend on knowing all boundary conditions and material parameters in detail, which is rarely feasible in practice. As outlined in Section 2.8, such exhaustive accuracy is often unnecessary for perceptual evaluations. Nevertheless, multiphysics tools are valuable for modeling individual components of the auralization process (e.g., radiation or transmission), but they are not yet practical as complete end-to-end auralization frameworks.

2.9.3 Auralization-Specific Software

A third group of solutions consists of software developed specifically for creating auralizations and virtual acoustic environments. Examples include Virtual Acoustics [106], EVERTims [107], TASCAR [108], liveRAZR [109], or the NASA Auralization Framework [110]. These systems are explicitly designed to render audio signals for perceptual studies and often support real-time operation, head tracking, synchronization with other modalities, and a range of source models. Virtual Acoustics, for instance, provides modules tailored to outdoor noise propagation, aircraft noise, and room acoustics. Such frameworks are very powerful and under active development, but their breadth of functionality can also make them overwhelming to configure. Moreover, integrating external models or custom source descriptions is not always straightforward, and some frameworks impose specific technical requirements (e.g., server-based execution).

Game engines such as Unity have also become increasingly popular for immersive audio-visual rendering, especially when combined with virtual reality. With suitable plugins, accurate acoustic simulations can be integrated into interactive virtual environments. This is particularly useful for perceptual studies where the listener's position or head orientation must be updated in real time. However, experimental procedures such as randomized source appearances or synchronization with non-audio events may require substantial additional scripting. For certain types of listening experiments, it may be more efficient to implement the full procedure directly within an environment such as MATLAB.

In addition to these "complete" frameworks, there is a large body of research on auralization techniques for highly specific applications. While these approaches demonstrate the diversity of possible solutions, each of the three main categories reviewed here has inherent limitations: room acoustics tools are restricted in scope and in source flexibility, multiphysics solvers are computationally prohibitive for perceptual use, and dedicated auralization frameworks are often specialized or difficult to adapt. The present thesis does not aim to develop yet another general-purpose framework, but rather to demonstrate a methodology for combining existing tools with tailored implementations to address the specific requirements of human subject studies.

CHAPTER 3

Measuring the Human Response to Sound

This chapter provides a broad methodological overview of how human responses to sound are assessed in experimental research. It focuses on the approaches and measures most relevant to controlled laboratory studies, and complements the auralization framework introduced in Chapter 2 by describing how subjective experience, physiological reactions, and behavioral performance can be quantified. After a brief background section (Section 3.1), the chapter reviews subjective (Section 3.2), physiological (Section 3.3), and behavioral responses (Section 3.4), and then summarizes key considerations in study design (Section 3.5) and statistical evaluation (Section 3.6).

3.1 Background and Evolution of Human Response Research

Measuring the human response to environmental stressors such as sound has been a research topic for centuries. While early studies in occupational health primarily focused on general environmental factors such as air quality, nutrition, and toxic exposure [111], reports of hearing loss among artisans and mill workers appeared as early as the 1700s [112]. As experimental approaches to human physiology developed in the nineteenth century, occupational exposure became one of the first contexts for studying how sound affects the body. With the rise of industrial machinery in the early twentieth century, large populations were chronically exposed to high noise levels, and research began to link excessive exposure to auditory damage [113]. Soon after, studies extended this perspective to other physiological and psychological domains, including motor performance, respiration, cardiovascular activity, and cognitive function [18].

Building on these early investigations, research in the mid-twentieth century expanded beyond a primarily physiological focus to include perceptual and cognitive aspects of noise effects. Behavioral observations such as the involuntary increase of vocal effort in noisy environments, the so-called *Lombard effect* [114], illustrated that sound not only affects the body but also influences behavior and communication. At the same time, advances in psychoacoustics and experimental psychology enabled systematic studies of auditory attention and information processing. Classic experiments by Cherry [115], and Broadbent [116] established paradigms for studying selective listening and distraction, revealing how competing sound sources can interfere with perception and task performance. Together, these developments marked a transition from viewing noise solely as a cause of hearing damage to understanding it as a multifaceted environmental stressor interacting with cognitive, emotional, and physiological processes.

In recent decades, research on the human response to noise has evolved into a broad, interdisciplinary field. Whereas early investigations mainly examined acute effects, it is now well established that noise exposure can have both short- and long-term consequences for health and well-being. Beyond auditory damage, studies have shown associations with hypertension [21], cardiovascular disease [22], sleep disturbance [23], learning impairments [24], annoyance [6], and other adverse outcomes [25–27]. This broader perspective integrates physiological mechanisms, behavioral adaptations, and subjective experience within a unified conceptual framework for understanding the human response to noise. Against this background, the following sections summarize the main response domains and methodological tools used in experimental noise research.

3.2 Subjective Response

Subjective responses describe how people consciously experience and evaluate sound. They can reflect how noise is perceived with respect to, e.g., effort, disturbance, or pleasantness, thereby complementing physiological and behavioral measures that capture less conscious responses. Because they rely on conscious reporting, subjective measures can be affected by factors such as context, expectations, or individual interpretation. Subjective responses are typically assessed through questionnaires or rating scales during or after an experiment. In noise research, common examples include ratings of perceived workload, annoyance, and sound quality. The following sections give an overview of typical measures used for this purpose.

3.2.1 Noise Annoyance

Noise annoyance is one of the most common subjective response measures in noise research. It has been applied across many fields because it is simple to administer and easy for participants to understand. While early studies often used their own rating scales, noise annoyance is now standardized by the *International Commission on Biological Effects of Noise (ICBEN)* scales, as described in ISO/TS 15666 [117]. The ISO specification defines two recommended question formats: a five-point verbal scale and an eleven-point numerical scale. Both are unipolar *Likert scales*, meaning that respondents rate the degree of annoyance along an ordered scale from a neutral ("not at all annoyed") to a negative extreme ("extremely annoyed").

In English, the standard wording of the corresponding question is: "Thinking about the last [12 months or so], when you are here at home, how much does noise from [noise source] bother, disturb, or annoy you?". Responses are provided either on the five verbal categories ("not at all," "slightly," "moderately," "very," "extremely") or on a 0–10 numerical scale anchored by "not at all annoyed" and "extremely annoyed." The method was originally developed for socio-acoustic and community surveys, but it is also frequently adapted for use in experimental studies, where the time frame and context are modified to fit short-term exposures (for example, "How annoying did you find the sound you just heard?").

While the ICBEN scale provides a robust and comparable measure of perceived annoyance, it does not reveal why a sound is judged as annoying. Numerous studies have therefore tried to predict annoyance from psychoacoustic parameters such as loudness, sharpness, fluctuation strength, or tonality [93]. However, such models are typically limited to specific noise types (e.g., road traffic, aircraft, or wind turbines) and cannot yet reliably predict annoyance for arbitrary sounds. This reflects the inherently subjective nature of annoyance, which is influenced not only by acoustic properties but also by individual differences such as age, noise sensitivity, expectations, and social context. Section 3.5.3 discusses in more detail how such between-subject factors can be accounted for in experimental design.

In experimental research, simplified or alternative methods are often used to assess short-term annoyance responses. These include direct numerical ratings of single sound exposures, paired-comparison or ranking tasks to compare several stimuli, and continuous response techniques for time-varying or dynamic sounds [93]. Broader perceptual scales with bipolar adjectives (e.g., pleasant–unpleasant or calm–disturbing) are also common when annoyance is evaluated alongside other perceptual attributes, such as comfort or sound quality [118].

3.2.2 Perceived Workload

In experimental psychology and human factors research, a commonly used subjective measure to evaluate responses to environmental stressors is *perceived workload*. In such experiments, participants typically perform a defined task and are then asked to rate how demanding they experienced it. This provides valuable insight into how external factors, such as noise, can increase cognitive effort, even when the sound itself may not be perceived as particularly annoying.

A standardized and widely used method for assessing perceived workload is the *NASA Task Load Index (NASA-TLX)*. Developed by the U.S. *National Aeronautics and Space Administration (NASA)* [119], it was originally designed to evaluate the mental and physical demands experienced by operators in complex human–machine systems, such as aviation and air traffic control. Since then, it has become one of the most established tools for workload assessment across many research areas, including ergonomics, human–computer interaction, and environmental psychology [120].

The NASA-TLX consists of six subscales, each rated on a 100-point Likert scale with 5-point increments:

1. *Mental Demand* – how much mental effort was required.
2. *Physical Demand* – the amount of physical effort involved.
3. *Temporal Demand* – the feeling of time pressure or pace of the task.
4. *Performance* – how successful the participant felt in accomplishing the task.
5. *Effort* – how hard the participant had to work to achieve their performance.
6. *Frustration* – how insecure, irritated, or stressed the participant felt.

In the original version, participants first rate each dimension individually and then perform a pairwise weighting procedure to indicate which dimensions contributed most to their overall workload. The final workload score is calculated as a weighted average across the six dimensions. However, many studies omit this step and instead use the so-called Raw TLX (RTLX) modification, which averages the six ratings. This simplified version correlates strongly with the weighted form and is easier to administer, especially in short experimental sessions [120]. Another common variation is to omit calculating an overall workload score altogether and instead analyze the individual subscales separately.

Although the NASA-TLX remains the most widely used method, several alternative tools exist. For example, the *Rating Scale for Mental Effort (RSME)* [121] uses a single continuous scale to estimate perceived effort, thereby reducing completion time while maintaining sensitivity to mental demand. Other approaches, such as visual analog or task-specific category scales, are also used when experimental conditions require minimal interruption. Overall, including a workload questionnaire alongside physiological or behavioral measures provides an efficient way to add a subjective dimension to experimental research.

3.2.3 Perceived Sound Quality

Another subjective measure commonly used in acoustic research, and particularly in product design and acoustic engineering, is *perceived sound quality*. This type of assessment focuses on how suitable or pleasant a sound is for its intended function. Sound quality evaluations can refer to an overall judgment, such as the general pleasantness or acceptability of a sound, or be divided into perceptual attributes describing specific aspects of the auditory impression. Typical descriptors include *powerful*, *smooth*, *sharp*, or *harsh*, which can be rated individually or combined into broader perceptual dimensions.

For example, the sound quality of air-conditioning units can be assessed using semantic differential scales, where listeners rate recordings along bipolar adjective pairs (e.g., *pleasant–unpleasant*, *soft–harsh*) [122]. Statistical analysis of such data can reveal perceptual dimensions such as *noisiness*, *tonal balance*, or *pleasantness*, which can then be related to measurable acoustic features, such as loudness, sharpness, or spectral balance. Similar approaches are widely used in product-sound engineering to optimize the auditory impression of domestic appliances, tools, or vehicle interiors [93].

In contrast to annoyance, which usually reflects a negative evaluation, sound quality also includes positive attributes such as comfort, smoothness, and luxury, thereby offering a more comprehensive view of how a sound is perceived and accepted. This broader perspective makes sound-quality evaluation an important tool in modern product sound design. While such assessments are a central focus of industry-oriented psychoacoustics, academic research often places greater emphasis on analytic or diagnostic measures aimed at explaining underlying perceptual mechanisms. In the context of electric vehicle noise, sound quality is nonetheless relevant from a manufacturer's perspective, since commercially deployed AVAS signals will ultimately need to meet standards of perceived quality and brand identity – considerations that fall outside the scope of this thesis but should not be overlooked in applied AVAS design.

3.3 Physiological Response

Measuring the physiological response to noise has long been a central topic in human noise research, most prominently in the context of noise-induced hearing loss. However, beyond its effects on the auditory system, a wide range of physiological parameters have been shown to be sensitive to noise exposure. Some of these responses occur during specific physiological states such as sleep, where nocturnal noise can trigger autonomic arousals and EEG awakenings [123]. Other physiological effects are associated with long-term health outcomes, such as cardiovascular disease, and can therefore only be examined through observational or epidemiological studies [124]. In contrast, this thesis focuses on short-term physiological effects that can be elicited and measured under controlled laboratory conditions.

Physiological responses to noise can broadly be grouped into those reflecting neural activity, which indicate auditory and cognitive processing in the central nervous system, and those reflecting autonomic and cardiovascular activity, which capture changes in electrodermal, cardiac, or vascular function related to stress and arousal. Additional indicators, such as pupil dilation, respiratory rate, or hormonal activity, can further complement these measures by providing insight into autonomic activation and overall physiological state. The following sections provide a broad overview of these categories and summarize measures commonly employed in noise-related research.

3.3.1 Neural Activity

Neural activity measures can provide direct insight into how the brain processes sound and how noise influences auditory and cognitive functions. Among the available techniques, *Electroencephalography (EEG)* and *Event-Related Potentials (ERPs)* are the most common tools for assessing neural responses to noise. EEG records continuous electrical activity from the scalp, whereas ERPs are time-locked responses derived from the same signal that reflect neural responses to specific acoustic or cognitive events. These measures provide millisecond temporal resolution, allowing researchers to investigate auditory perception, attention, and cortical arousal in response to acoustic stimulation [125, 126]. Recent studies, for example, have used EEG to examine how different noise levels and spectral content affect brain activity during cognitive tasks, illustrating the method's potential to quantify neural responses to environmental sound [127]. Specific ERP components, such as the P300 [128], have been used to quantify how background noise and other types of sound affect sensory processing and selective attention [129–131]. EEG-based analyses of spectral power or coherence can provide indicators of mental effort or fatigue [132] and are also commonly employed in sleep research to classify sleep stages and detect noise-induced arousals [123, 124].

Despite their high temporal resolution and direct access to neural processes, EEG and ERP methods have several practical and interpretative limitations. Brain signals measured at the scalp are small and highly sensitive to artifacts from eye and muscle activity, making both recording and analysis technically demanding. Interpreting ERP components is often nontrivial, as different cognitive processes can contribute to the same response [128], making interpretation particularly challenging for researchers without a neuroscience background. Moreover, experiments require participants to have EEG electrodes attached to the scalp. This is often done using EEG caps, i.e., arrays of typically 8 to 64 electrodes embedded in a flexible fabric, which can be time-consuming to apply and may cause discomfort. In studies aiming for realistic listening conditions, this setup can substantially diminish ecological validity. Consequently, EEG and ERP methods are most suitable when the research focuses on auditory attention or neural mechanisms rather than on more applied research that requires a naturalistic listening experience.

Other neurophysiological techniques, such as *magnetoencephalography (MEG)* or *functional near-infrared spectroscopy (fNIRS)*, have also been applied to study cortical responses to sound [133–135]. While they offer complementary advantages, with MEG providing improved spatial resolution and fNIRS enabling more freedom of movement, such methods are less common in laboratory noise research due to cost, complexity, and limited availability.

Overall, neural activity measures can provide valuable insights into how the brain processes noise. While they are powerful tools for investigating auditory neural mechanisms, their use in more applied noise-related laboratory experiments remains limited by methodological and practical challenges.

3.3.2 Autonomic and Cardiovascular Activity

Autonomic and cardiovascular measures describe how the body responds to environmental stressors, such as noise, through changes in physiological functions, including heart activity, blood circulation, and skin conductance. Unlike neural recordings, which reflect brain activity, these measures indicate the body's overall state of arousal or relaxation and can therefore reveal stress reactions to noise exposure.

Typical parameters include *heart rate (HR)*, *heart rate variability (HRV)*, *blood pressure*, *respiration rate*, and *electrodermal activity (EDA)* [136]. These can all be measured non-invasively and continuously, for example, using *electrocardiography (ECG)*, which records the heart's electrical activity to measure HR and HRV, or small electrodes attached to the fingers for EDA. Heart rate and blood pressure tend to increase when a person is exposed to unexpected or unpleasant sounds, reflecting a short-term activation of the body's stress response [124, 137, 138]. HRV, in contrast, often decreases under stress and is therefore considered a marker of physiological strain or reduced recovery capacity [139].

Electrodermal activity reflects small changes in the electrical conductance of the skin that occur when sweat gland activity increases [140]. It is one of the most sensitive indicators of momentary arousal and has been shown to respond even to relatively low-level noise if it is perceived as disturbing or attention-demanding [141]. Respiratory rate may also increase during noise exposure, though this effect is generally small and depends on the task performed. Beyond these immediate physiological responses, noise exposure may also trigger a slower hormonal stress response, leading to elevated cortisol or adrenaline levels. Such hormonal responses develop more slowly but indicate similar stress pathways that, if triggered repeatedly, may contribute to long-term health effects [137, 142].

These measures are particularly relevant from a health perspective, since changes in heart rate, HRV, or blood pressure often point to mechanisms that, if activated repeatedly over long periods, could contribute to cardiovascular disease or other chronic stress-related outcomes [124, 137]. Hence, even subtle short-term changes observed in laboratory studies can provide important mechanistic evidence linking noise exposure to long-term health effects.

However, such measures are also sensitive to numerous unrelated influences, such as breathing patterns, body posture, and individual fitness. As a result, observed effects are often small and require careful experimental control and averaging over many participants. In practice, autonomic and cardiovascular indicators are most informative when interpreted alongside subjective and behavioral measures, providing a more comprehensive picture of how noise affects both perception and physiology.

3.4 Behavioral Response

Behavioral responses describe how environmental factors, such as sound, influence observable human actions and performance. They reflect the combined outcome of perceptual, cognitive, and motor processes and therefore complement subjective and physiological measures. Rooted in experimental psychology and human factors research, behavioral assessment quantifies performance, attention, and decision making through objective measures such as reaction time, accuracy, and error rate in tasks probing attention, memory, or spatial hearing [143]. These measures provide direct evidence of how environmental conditions affect information processing and interaction with the surroundings. In the context of this thesis, the most relevant behavioral measures are cognitive performance, perceptual performance, and motor and behavioral adaptation, as described below.

3.4.1 Cognitive Performance

Cognitive performance encompasses mental functions such as attention, working memory, and problem-solving that determine how efficiently individuals perceive, process, and respond to information. In experimental noise research, cognitive performance is often evaluated using tasks that require sustained concentration or rapid decision-making, such as reaction-time tests, continuous performance tests, or memory-based paradigms [144]. Changes in these measures under different sound conditions can reveal whether noise interferes with or facilitates mental processing.

Several mechanisms have been proposed to explain how noise influences cognitive performance. At a basic level, noise can act as a source of distraction that competes for attentional resources, particularly when it contains meaningful or varying content [116, 145]. This is well illustrated by the *irrelevant speech effect*, where background speech disrupts short-term memory tasks such as serial recall or mental arithmetic [146]. Other theories emphasize arousal and mental effort as mediating factors. According to the *Yerkes–Dodson law*, performance improves with increasing arousal or stress only up to an optimal level, beyond which excessive activation reduces efficiency and increases errors [147]. Moderate levels of stimulation may therefore enhance alertness and simple-task performance. In contrast, high noise levels or complex tasks can push activation beyond the optimal range, resulting in distraction or fatigue. Consequently, the relationship between noise and performance depends strongly on task type, sound characteristics, and the listener's state of alertness or fatigue [148–150].

3.4.2 Perceptual Performance

Perceptual performance describes how efficiently sensory information is detected, identified, and interpreted [143, 151], reflecting the accuracy and speed with which people perceive their environment through different sensory channels. In the context of human response to sound, it refers to how effectively listeners can detect, identify, and localize sounds in their surroundings, which is essential for communication, orientation, and safety in everyday life. These measures capture the perceptual aspect of behavior and indicate how acoustic conditions influence awareness of the auditory environment.

The most relevant perceptual performance measures for this thesis are detection and localization tasks. Detection tasks typically assess how background noise and sound type influence the ability to detect or identify a target sound, such as an approaching vehicle or a warning signal. Such tasks can reveal both energetic masking, where noise physically masks the signal, and informational masking, where complex or fluctuating backgrounds cause perceptual confusion [93, 152]. Response accuracy or reaction time in detection experiments, therefore, provides an indicator of how well important sounds can be perceived in realistic acoustic scenes.

Localization tasks, in contrast, evaluate spatial hearing performance by measuring how accurately and quickly listeners can indicate the direction of a sound source [87]. Research on spatial hearing has a long tradition, and already in the 1930s it was shown that localization accuracy depends on the signal type [153]. Yet, it is still not fully understood how localization operates in complex auditory scenes. Although computational models can predict binaural sound localization under controlled conditions, their accuracy decreases in reverberant or multi-source environments, making reliable prediction of real-world localization challenging [154]. Noise can further impair localization by obscuring spatial cues or diverting attention, particularly when multiple sound sources are present [155, 156]. The resulting localization errors or increased response times provide behavioral evidence of how acoustic interference affects spatial perception. In applied contexts, such as vehicle warning sounds or alarm design, these measures are crucial for evaluating the perceptual effectiveness and potential disturbance of sound signals, and they are well-suited to virtual acoustic environments in laboratory research.

3.4.3 Motor and Behavioral Adaptation

Beyond changes in cognitive or perceptual performance, noise can also affect people's actions and physical movements, collectively referred to as *motor behavior* [157].

Such adaptations often occur unconsciously and help maintain communication or task performance in noisy situations. A well-known example is the *Lombard effect*, the involuntary increase in vocal effort when speaking against background noise [114]. Similar adjustments include changes in speech rate, articulation, or body posture, which help preserve intelligibility and effectiveness despite acoustic interference. In occupational and environmental contexts, noise can also lead to behavioral changes such as avoidance, where individuals move away from the noise source, or strategy adjustments, where they work faster but less accurately, or take short breaks to recover from exposure [158]. These adaptations are an important component of the human response to noise, as they demonstrate how people actively compensate for acoustic challenges. Studying such behavioral and motor adjustments complements physiological and subjective data by revealing how individuals manage environmental stressors in a practical, goal-directed manner.

3.5 Study Design

A well-designed experiment is essential for obtaining reliable and interpretable results. In studies of human responses to sound, the experimental design determines whether observed differences can be attributed to the manipulated acoustic conditions or are caused by uncontrolled influences. Careful planning of variables, participant selection, and procedural control is, therefore, a crucial part of conducting listening experiments. Designing such studies can be challenging, as human perception and behavior are influenced by numerous internal and external factors that vary across individuals and over time.

The following section introduces the fundamental concepts of experimental study design relevant to the research presented in this thesis. It is not intended as a comprehensive methodological guide, and readers seeking a more detailed treatment are referred to standard textbooks such as [159] or [143].

3.5.1 Trials, Variables, and Factors

Every human subject experiment consists of a sequence of *trials*, each representing one presentation of a stimulus under defined conditions, followed by a participant's response. Trials are often grouped into *blocks* that share similar properties, such as a common task or stimulus type, and are separated by short breaks to reduce fatigue. Randomizing or counterbalancing the order of trials and blocks helps minimize order effects and maintain experimental validity. The set of all trials and blocks completed by one participant in a sitting constitutes a *session*.

Each trial manipulates one or more *independent variables* (or *factors*) that the researcher intentionally controls, such as sound level, signal type, or the presence of background noise. The resulting participant responses define the *dependent variables*, for instance, annoyance ratings, reaction time, or physiological measures. When multiple factors are included in a study, their effects can be analyzed individually as main effects or jointly as interactions, showing how the influence of one factor depends on the level of another. Such *factorial designs* are widely used in behavioral and human factors research, where multiple experimental and contextual parameters often interact [143, 159]. For example, a 3×2 factorial design might combine three sound pressure levels with two signal types to examine both their individual and combined effects on perceived annoyance or electrodermal activity.

3.5.2 Design Types

Experimental designs are generally categorized as *between-subject* or *within-subject* (also referred to as *repeated-measures*) designs. In between-subject designs, different groups of participants experience different conditions, which avoids potential learning or fatigue effects but requires a larger sample size to compensate for individual variability. In within-subject designs, the same participants are exposed to several conditions, allowing more sensitive comparisons because each person serves as their own control. This approach is widely used in psychoacoustics, since exposure to one sound typically does not produce lasting effects and, with sufficient time between trials, only marginally affects responses to subsequent sounds. In contrast, in areas such as pharmacology or clinical research, repeated-measures designs are often infeasible because exposure to one condition (for example, a drug) may permanently alter the response to another.

Repeated-measures designs require special care to prevent *order effects* such as learning, fatigue, or adaptation. To minimize these influences, conditions are often *counterbalanced*, meaning that the order of presentation varies systematically across participants, for example, through randomizing of trial sequences. Randomization can also be applied to other aspects of an experiment, such as the allocation of participants to groups or the selection of stimulus examples, reducing the risk that uncontrolled factors bias the results. In some contexts, *blinding* is used to prevent either the participant or the experimenter from knowing which condition is currently being tested, thereby reducing expectation-related biases. While blinding is less common in auditory perception studies, the underlying principle of minimizing expectancy effects remains relevant, for instance, when participants are informed about the task's purpose or difficulty.

3.5.3 Participants and Confounding Factors

Individual differences between participants represent another major source of variability in human experiments. Factors such as age, hearing ability, or prior experience can influence responses to sound [93, 160]. To reduce such *between-subject variability*, participants are often screened for relevant characteristics and randomly assigned to conditions. However, practical limitations frequently restrict participant diversity. For instance, studies conducted at universities often rely on student volunteers, which limits the range of ages, education levels, and socioeconomic backgrounds represented in the sample. One way to manage such differences is to include them as additional factors or covariates in the statistical analysis, as described in Section 3.6. These so-called *nuisance variables* are not of primary interest but can help account for systematic variability in the data [161]. However, this approach requires that these characteristics are sufficiently balanced across participants, e.g., ensuring that the sample is approximately gender-balanced or covers a representative age range.

In addition to individual variability, *confounding factors* can arise from the experimental environment or procedure. These are uncontrolled influences that vary systematically with the experimental conditions and can distort the results. For instance, air quality or temperature in a laboratory may gradually change over the course of a session, or background noise from a ventilation system may vary across experiment days. Participant fatigue, familiarity with the task, or expectations about the study's purpose can also bias responses. Standard procedures to minimize such effects include maintaining consistent environmental conditions, providing clear and identical instructions, allowing practice trials, and scheduling sufficient breaks. Through such measures, potential confounders can be reduced, improving the reliability and interpretability of the findings [159].

3.6 Statistical Evaluation

Once an experiment is conducted, the data obtained must be statistically analyzed to determine whether the factors under investigation have significant effects on the dependent variables. Statistical analysis, therefore, serves two purposes: to determine whether an observed effect is *systematic* rather than random, and to quantify its magnitude.

A wide range of statistical approaches exists, differing in their assumptions and suitability for different types of data [161]. A comprehensive treatment of all possible methods is beyond the scope of this thesis. Instead, the following sections focus on the specific techniques applied in the studies conducted here, namely, parametric tests such as the *Analysis of Variance (ANOVA)* and nonparametric tests such as the *Friedman* and *Wilcoxon signed-rank* tests. Before introducing these methods, it is useful to review several fundamental statistical concepts that underlie all subsequent analyses.

3.6.1 Fundamental Concepts

In classical hypothesis testing, statistical evaluation begins by formulating a *null hypothesis* (H_0) that assumes no effect or difference, and an *alternative hypothesis* (H_1) that assumes the presence of an effect.[1] A statistical test then estimates the probability of obtaining results at least as extreme as the observed ones under the assumption that H_0 is true. This probability is the *p-value*. A small p-value indicates that the observed data would be unlikely if the null hypothesis were true.

To decide when a p-value is considered small enough, a threshold α (the *significance level*) is chosen in advance. A significance level of $\alpha = 0.05$ is commonly used in most scientific fields, meaning that we accept a 5% risk of drawing a false-positive conclusion in the long run. If $p < \alpha$, the result is said to be statistically significant, and the null hypothesis is rejected in favor of the alternative. The remaining 5% probability of incorrectly rejecting H_0 is known as the *Type I error* rate.

A related concept is the *confidence interval* (CI), which provides a plausible range of values for an estimated effect or parameter. For example, a 95% CI means that if the same experiment were repeated many times and a new interval were calculated each time, about 95% of those intervals would contain the true parameter value [162]. In practice, researchers often interpret this as being "95% confident" that the true value lies within the observed interval. Confidence intervals convey both the direction and precision of an effect and are a useful complement to p-values, as they express the uncertainty of estimation rather than a binary significant or non-significant outcome.

If a significant effect is found, its magnitude can be quantified using an *effect size* measure. Effect sizes quantify the magnitude of a difference or relationship in standardized units, making them more interpretable across studies. Different tests use different effect size metrics, of which some range from 0 to 1, others from -1 to 1, and others are unbounded [163]. In practice, effect sizes are often more informative than p-values, since with enough data, even trivial effects can become statistically significant.

While p-values and Type I errors are most prominently reported, it is equally important to consider *Type II errors*, which occur when an existing effect fails to be detected. The probability of making a Type II error is denoted by β, and the complement $(1-\beta)$ is called the *statistical power* of a test. Statistical power describes the probability of detecting an effect if it truly exists. It depends on four main factors: the true effect size, the chosen significance level α, the sample size N, and the variability in the data. Larger effects, larger samples, lower variance, and higher α values all increase statistical power. To estimate in advance which sample size is needed to achieve adequate power, pre-study *power analysis* can be conducted using

[1] An alternative to classical (frequentist) inference is the Bayesian framework, which estimates the probability of a hypothesis given the observed data and prior information. While not applied in this thesis, Bayesian approaches represent a useful complementary perspective.

expected effect sizes, variability, significance level, and target power [163].

All these fundamental concepts can be illustrated with a simple example. Suppose we suspect that a coin is biased, so we flip it 100 times and observe 65 heads and 35 tails. The null hypothesis H_0 states that the coin is fair ($p_{\text{Heads}} = 0.5$). Using basic binomial statistics, we can calculate the probability of obtaining a result this extreme or more under H_0, i.e., if the coin were not biased, which is approximately $p = 0.004$ (0.4%). Since this is smaller than the common significance level $\alpha = 0.05$, we reject H_0 and conclude that the coin is likely biased. The corresponding effect size (Cohen's $h \approx 0.31$) indicates a small to medium bias that is both statistically and practically meaningful.[2]

If the true bias were smaller, the probability of detecting it with only 100 tosses would decrease, increasing the risk of a Type II error. For instance, with an effect half as large, our test with 100 coin flips might have only 40% power, i.e., a 40% chance of detecting the bias even if it exists. Increasing the number of coin flips to, say, a million would greatly increase the test's power, making even tiny deviations from a 50/50 probability statistically significant. However, those differences would be so small as to be meaningless in practice. This illustrates why p-values should always be interpreted in conjunction with effect sizes and confidence intervals, since statistical significance alone does not necessarily imply practical importance.

While calculating probabilities for a simple coin flip is relatively straightforward, statistical testing becomes more complex when multiple factors, participants, or repeated measures are involved, as is typically the case in human response studies. In such situations, *parametric approaches* are commonly applied to analyze the influence of experimental factors, as described in the following section.

3.6.2 Parametric Approaches

Parametric statistical tests are among the most widely used methods for analyzing experimental data and for estimating effects and their statistical significance. They are called *parametric* because they assume that the data can be described by a specific probability distribution characterized by a small number of parameters, typically

[2]For $n = 100$ trials with $x = 65$ observed heads, and the hypothesized probability of heads under H_0 (the value specified by the null hypothesis) $p_0 = 0.5$, a two-sided exact binomial p-value can be computed as

$$p = 2 \sum_{i=k}^{n} \binom{n}{i} p_0^i (1 - p_0)^{n-i}, \quad k = \max(x, n - x),$$

which yields $p \approx 0.00352$ (*Note:* this specific doubling-of-one-tail formula is only valid in this specific case where $p_0 = 0.5$). For this application, the effect size can be estimated with Cohen's h,

$$h = 2 \arcsin \sqrt{x/n} \; - \; 2 \arcsin \sqrt{p_0} \approx 0.305.$$

the mean and variance. Consequently, these methods rely on certain assumptions about the underlying population and data distribution – most importantly, that the *residuals*, i.e., the deviations of individual observations from their predicted or mean values, are approximately normally distributed, and that the *variances*, i.e., the spread of observations within each group, are roughly equal across conditions.

The simplest parametric comparison involves two groups and can be performed using a *t-test*. In its independent-samples form, it tests whether the means of two unrelated groups differ significantly, whereas the paired-samples version accounts for within-subject designs in which each participant experiences both conditions.

When more than two conditions or factors are involved, the *analysis of variance (ANOVA)* provides a general framework to test for overall differences between group means. ANOVA partitions the total variance in the data into components attributable to systematic effects of the experimental factors and to random residuals. The ratio of these components forms the F-statistic, whose significance indicates whether at least one group mean differs from the others. A one-way ANOVA tests a single factor with multiple levels, while multifactor designs allow the evaluation of main effects and interactions between factors. The strength of an effect in ANOVA is often reported using *partial eta squared* (η_p^2), which expresses how much of the variance a given factor explains relative to what that factor and the unexplained residual variance account for together.

In experiments where the same participants are exposed to several conditions – a common situation in perceptual or human-response research – a *repeated-measures ANOVA* is often used. This variant explicitly models the correlation between repeated observations from the same participant, thereby increasing sensitivity and reducing the influence of inter-individual variability. In such within-subject designs, an additional assumption called *sphericity* must be satisfied, requiring the variances of all possible pairwise condition differences to be equal. Violations of sphericity can be corrected, for instance, using the *Greenhouse–Geisser* adjustment of degrees of freedom.

If an ANOVA reveals a significant main effect, it only indicates that at least one group mean differs, but not which one. To identify specific differences between pairs of conditions, *post hoc* comparisons are conducted, commonly using paired or independent-samples t-tests. However, when multiple pairwise comparisons are performed, the probability of obtaining at least one false-positive result (Type I error) increases rapidly. For instance, if one were to compare one hundred different conditions, purely by chance, several pairs could appear significantly different even if no true effects existed. To control this cumulative error rate, it is important to adjust the significance level accordingly. A simple and conservative method is the *Bonferroni correction*, which controls the overall error rate by dividing the significance level α by the number of comparisons, or equivalently, by multiplying each obtained p-value by that number.

It is generally not appropriate to perform multiple pairwise t-tests without first establishing whether any systematic effect is present at all. When more than two conditions are compared, a global test such as ANOVA should always precede pairwise comparisons, as testing many pairs directly without this step would greatly inflate the false-positive rate.

Beyond t-tests and ANOVA, several other methods fall under the category of parametric approaches. These include the *analysis of covariance* (ANCOVA), which controls for continuous covariates, *multiple linear regression*, which models relationships between continuous predictors and outcomes, and *linear mixed-effects models*, which generalize ANOVA to handle unbalanced or hierarchical data. Although these techniques are not used in the present work, they share the same theoretical foundation and assumptions as the simpler parametric tests described above.

The main advantages of parametric approaches are their statistical power, interpretability, and the ability to model complex factorial designs with interactions. They are particularly efficient when their assumptions are approximately met, and moderate deviations from normality often have little practical impact, especially with larger sample sizes. However, parametric tests can become unreliable when applied to ordinal data, i.e., data based on ordered categories without equal spacing between levels, heavily skewed distributions, or small samples that do not justify the normality assumption. In such cases, nonparametric methods provide more robust alternatives at the cost of reduced power, as described in the following section.

3.6.3 Nonparametric Approaches

In contrast to the previously described parametric approaches, nonparametric methods do not assume that the data follow a specific probability distribution or that variances are homogeneous across conditions. Instead, they rely on the relative ordering (*ranks*) of observations rather than their absolute values [164]. This makes them particularly suitable for ordinal data, such as rating scales where the order of responses matters but the distance between them is undefined, as well as for skewed distributions or small sample sizes in which normality cannot be assumed. Nonparametric tests are therefore often used in perceptual and human-response experiments that employ rating scales or bounded scores.

The nonparametric counterpart to the repeated-measures ANOVA is the *Friedman test*. It is used when several related samples or experimental conditions are compared within the same participants. The test ranks the data within each participant and then assesses whether the rank distribution differs systematically across conditions. The Friedman test yields a chi-square statistic (χ^2), which can be complemented by *Kendall's W* as a measure of effect size, indicating how consistently the conditions are ranked across participants. A significant result indicates that at least one condition differs from the others, but it does not indicate which one. As with ANOVA, further *post hoc* tests are required to identify specific pairwise differences.

These pairwise comparisons are commonly performed using the *Wilcoxon signed-rank test*, which serves as the nonparametric alternative to the paired t-test. The Wilcoxon test considers the signed differences between paired observations, ranks them by absolute magnitude, and then assesses whether the positive and negative ranks are symmetrically distributed around zero. A significant result indicates that the median difference between the two conditions is not zero. For Wilcoxon tests, the standardized test statistic Z is typically reported along with the corresponding p-value. Like its parametric counterpart, the Wilcoxon test assumes symmetric distributions of differences, but it does not require normality. When multiple Wilcoxon comparisons are performed following a Friedman test, the same issue of inflated Type I error arises as in parametric post hoc testing. Consequently, the same correction method, e.g., the Bonferroni correction, should be applied to maintain a controlled overall significance level.

The main advantages of nonparametric approaches are their robustness and general applicability. They can be used for virtually any ordered type of data, regardless of scale, distribution, or variance characteristics, and thus provide a safe and reliable option when parametric assumptions are uncertain or clearly violated. However, this robustness comes at the cost of statistical power: because nonparametric tests use ranks instead of raw data, they are generally less sensitive to small effects. As a result, nonparametric methods may fail to detect subtle but genuine effects that a parametric analysis could identify under ideal conditions.

CHAPTER 4

Application to Electric Vehicle Noise

This chapter presents a concrete application of the methodological framework introduced in Chapters 2 and 3, focusing on the auralization and experimental evaluation of electric vehicle noise. The chapter details the specific auralization components, reproduction strategies, validation procedures, and listening experiments employed in this thesis, together with their assumptions and limitations. The listening experiments are combined in the thematic clusters of this thesis that were introduced in Chapter 1, namely Section 4.3 (AVAS Perception and Safety) and Section 4.4 (Noise Effects in Living Environments).

4.1 Electric Vehicle Auralization Implementation

The following introduces the specific implementation of the general auralization framework for electric vehicles. This application aimed to develop perceptually validated outdoor and indoor auralizations of electric vehicle passages with flexible velocity profiles, multiple AVAS designs, and representative tire-road noise, to enable controlled investigations of localization, perceived speed, annoyance, workload, and physiological response. The section starts with a description of the performed reference measurements in Section 4.1.1 and then follows the structure of the auralization framework introduced in Chapter 2, i.e., source signal synthesis (Section 4.1.2), radiation (Section 4.1.3), propagation (Section 4.1.4), and encoding and reproduction (Section 4.1.5). Figure 4.1 provides an overview of the specific methods from the broader framework that were chosen for this application. The code for a large part

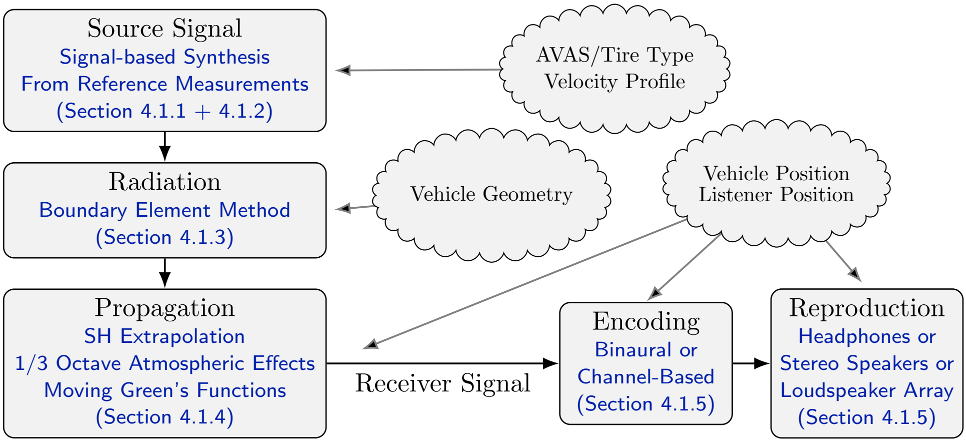


Figure 4.1: Stages of auralization from source to reproduction (see Figure 2.1) with specific methods implemented for EV application (blue) and corresponding simulation parameters (clouds).

of these methods is published in the electric vehicle auralization toolbox[1], and all relevant reference data is openly published in an online repository[2].

4.1.1 Reference Measurements

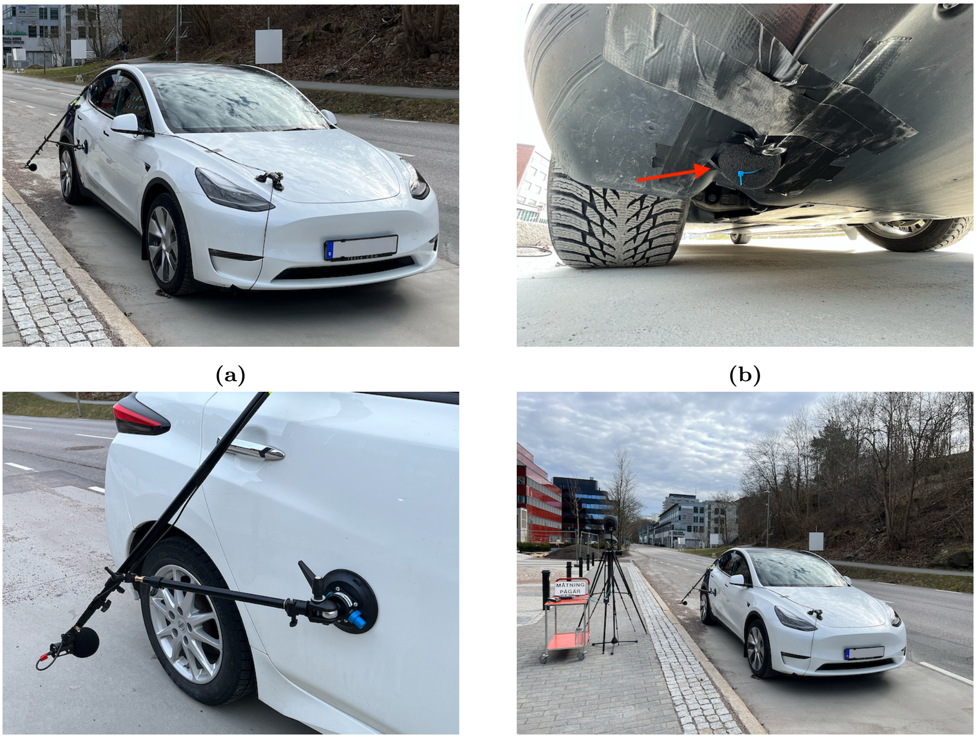
The first step in applying the auralization framework to EV noise was to perform controlled reference measurements on three representative, regulation-compliant electric vehicles. The measurement procedure and the full set of results are described in detail in Paper A. The following text summarizes the most relevant aspects in the context of this thesis and discusses possible improvements and limitations.

Setup

Measurements were performed using calibrated, omnidirectional, free field microphones mounted approximately 10 cm in front of the AVAS loudspeaker embedded in the front bumper (Figure 4.2b), and 40 cm from the tire, perpendicular to the wheel plane, as shown in Figure 4.2c. The microphone signals were recorded synchronously with the vehicle position using GPS, with the antenna mounted on the hood (Figure 4.2a).

At the roadside, a calibrated artificial head and an omnidirectional reference microphone were placed at a fixed observer position, whose coordinates were also determined via GPS. The roadside and on-vehicle recording systems were synchronized,

[1]https://github.com/leonpaulmueller/evat
[2]https://doi.org/10.5281/zenodo.10610490



(a) (b)

(c) (d)

Figure 4.2: Electric vehicle reference measurement setup with GPS antenna on front hood (a), microphones mounted in front of AVAS speaker (b) and tire (c), as well as roadside artificial head and reference microphone position (d).

resulting in a dataset comprising AVAS and tire source signals, time-aligned position-time data for the vehicle and observer, and binaural and omnidirectional pressure signals at the roadside position.

For each vehicle, forward and reverse passages were recorded for constant-speed and accelerating velocity profiles up to 30 km/h, at a sampling rate of 48 kHz. The distance between the vehicle path and the observer position ranged from 5 m to 10 m. The road surface was dense asphalt concrete, and all recordings were conducted under dry, windless conditions.

Results and Post-processing

Figure 4.3 shows examples of AVAS signals measured in front of the AVAS loudspeaker, together with the recorded vehicle velocity, for the AVAS types later referred to as noise AVAS (Figure 4.3a) and multi-tone AVAS (Figure 4.3b). Visually,

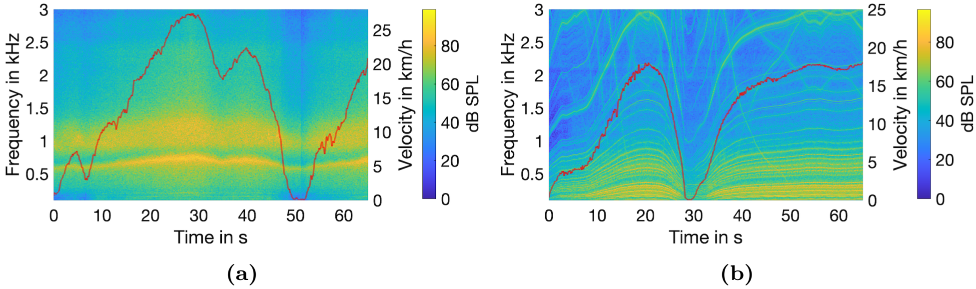


Figure 4.3: Excerpt of recorded isolated noise AVAS (a) and multi-tone AVAS (b). The red line shows the recorded vehicle velocity.

these recordings show a clear correlation between the measured velocity and the pitch of both signals. However, this continuous-time representation does not permit a velocity-dependent analysis at discrete speeds.

For further processing, all measurements were downsampled to a sampling rate of 12 kHz and transformed into velocity-dependent magnitude spectra, as shown in Figure 4.4. The spectra were obtained by dividing the signals into blocks of 512 samples ($\approx$ 43 ms), assigning each block a mean velocity based on the GPS data, and averaging the magnitude spectra of blocks with similar velocity in steps of 0.1 km/h. Even though the GPS-based velocity estimates contain some uncertainty, this procedure resulted in more than 200 blocks per discrete velocity value on average. After linear averaging of the magnitude spectra and applying a two-dimensional smoothing filter to reduce jumps between neighboring velocity values, the variance was substantially reduced. The top part of Figure 4.5 illustrates this spectral averaging procedure.

These velocity-dependent magnitude spectra form the basis for the source signal synthesis of both AVAS and tire-road noise described in Section 4.1.2. Across forward and reverse driving, the three measured vehicles implemented five different AVAS designs, yielding three slightly different tire-road noise components. While all of these measurements are described in more detail in Paper A, this thesis mainly focuses on the three AVAS designs that were used in the subsequent experiments, namely the noise AVAS (Figure 4.4a), the two-tone AVAS (Figure 4.4b), and the multi-tone AVAS (Figure 4.4c). As illustrated in Figure 4.4, the spectra capture the characteristic behavior of the different AVAS and tire signals and make the velocity dependence of their spectral content explicit. They also show that, apart from some higher frequency components in the multi-tone AVAS, downsampling the recordings to a sampling rate of 12 kHz, that is, limiting the frequency range to 6 kHz, still captures all relevant source signal components.

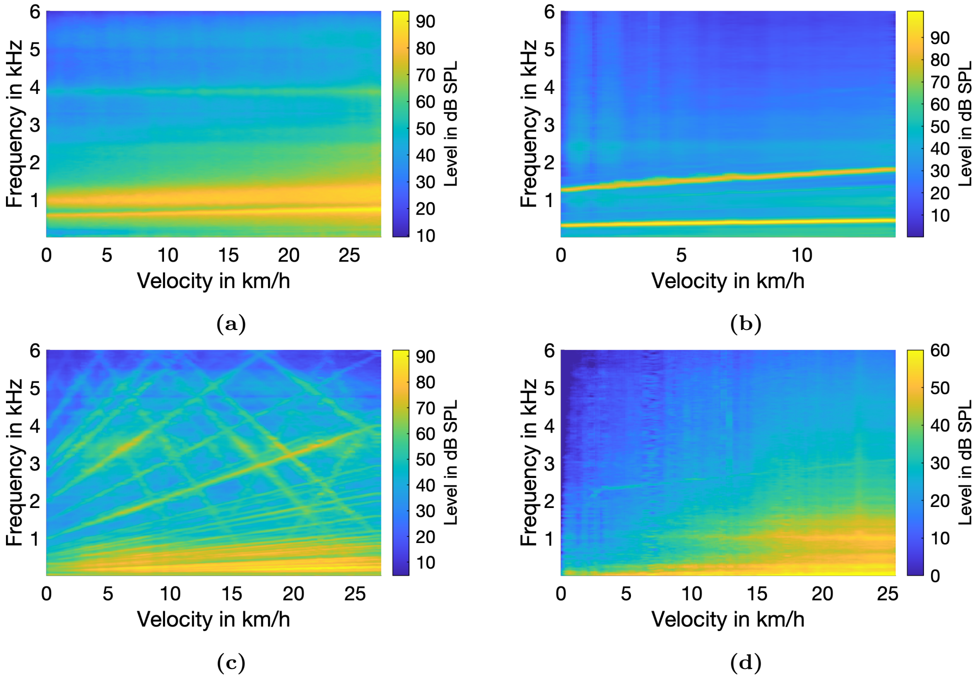


Figure 4.4: Velocity-dependent magnitude spectra for noise AVAS (a), two-tone AVAS (b), multi-tone AVAS (c), and tire-road noise (d).

Limitations

A first limitation of the reference measurements is the use of a single GPS receiver to estimate vehicle velocity. While the accuracy was sufficient for this context and the velocity-dependent spectral averaging substantially reduced the variance, a more advanced position measurement system could have decreased the uncertainty further.[3] An alternative would have been to record the vehicles' controller area network (CAN) signals, that is, the digital data containing the vehicles' own velocity estimate. Since this internal estimate drives AVAS generation, recording it would have provided more directly relevant information, at least for the AVAS components.

A second limitation concerns the recording of tire road noise. The collected data were adequate for plausible auralizations, but they were not fully controlled. A more rigorous approach would require knowledge of the exact tire type and the road surface characteristics, and, ideally, measurements performed in accordance

[3] Under ideal outdoor conditions, the employed Doppler-based speed estimates from standard GPS typically achieve an accuracy of about 0.1 km/h to 0.3 km/h. More advanced methods, such as differential GPS, reduce this to approximately 0.05 km/h to 0.15 km/h, and Real-Time Kinematic GNSS can achieve 0.01 km/h to 0.05 km/h. [165]

with the *Close Proximity Method (CPX)* [166] using a dedicated trailer. When working with a prototype vehicle, turning the AVAS system off would at least avoid crosstalk between AVAS and tire noise. An alternative would be to use existing datasets from previous CPX studies. However, although many publications report CPX measurements for different tire-road surface combinations, the corresponding raw multichannel time-domain data are, to the author's knowledge, not publicly available. More open science practices in this field would enable applications such as this thesis to incorporate high-quality reference data without having to repeatedly conduct similar measurements.

A third limitation is that the recordings were performed on a regular road segment without a standardized reference track or a professional test driver. A controlled track with reproducible driving conditions would have reduced variability in vehicle trajectories, road surface conditions, and background noise. Despite these limitations, the measurements enabled reliable reconstruction of the reference AVAS signals and plausible tire-road noise and are therefore considered sufficient for the purposes of this thesis.

4.1.2 Source Signal Generation

Based on the reference measurements, three source signal synthesis approaches were developed that, collectively, are expected to cover the full range of observed AVAS and tire-road noise signals: *additive synthesis*, *subtractive synthesis*, and *sample-based synthesis*. The three approaches were chosen because subtractive synthesis models broadband components, additive synthesis models tonal components, and sample-based synthesis models periodic signals that are not well represented by simple oscillators. These methods all fall into the category of signal-based synthesis methods (see Section 2.3) and are briefly described below. For a detailed derivation and evaluation, the reader is referred to Paper A. In all cases, the goal is to generate a time domain pressure signal from an arbitrary input velocity profile that reproduces the velocity-dependent spectral and temporal characteristics of the corresponding reference signals as closely as possible.

Subtractive Synthesis

The subtractive synthesis approach is based on the idea that a broadband noise signal is shaped by velocity-dependent filtering so that the resulting output matches the measured reference signals. This approach is particularly suitable for signals whose spectral structure is broadband and noise-dominated, such as the measured noise AVAS and most tire road noise components.

The implementation used in this thesis is based on the velocity-dependent magnitude spectra discussed in Section 4.1.1. The top part of Figure 4.5 again illustrates how these spectra are obtained by dividing the recorded pressure and velocity signals

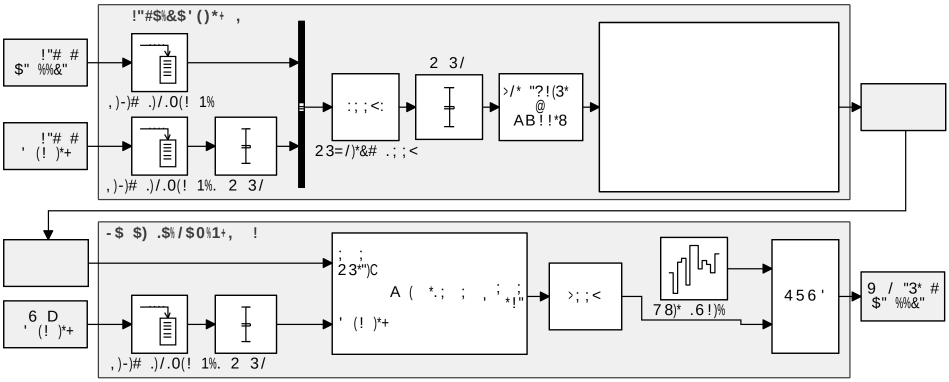


Figure 4.5: Subtractive synthesis approach including velocity-dependent spectral averaging of reference recordings.

into short blocks, averaging the magnitude spectra of blocks with similar velocity, and applying a two-dimensional smoothing filter. The outcome of this procedure is a frequency and velocity-dependent magnitude matrix, $|H(f, v)|$.

For signal generation at arbitrary velocities, the input velocity profile is first segmented into blocks of N samples. Since the magnitude spectra $|H(f, v)|$ do not contain phase information, a minimum phase approximation is reconstructed using the real cepstrum [167], which yields a set of velocity-dependent impulse responses $h(n, v)$. Each output block is then generated by convolving the corresponding noise block with the impulse response associated with the block's mean velocity. To ensure smooth transitions between consecutive blocks, the same noise block is also filtered with the impulse response associated with the previous block's velocity, and the two resulting signals are cross-faded using a raised cosine window. This avoids discontinuities and suppresses audible clicking artifacts during rapid changes in velocity.

Additive Synthesis

The additive synthesis approach assumes that the signal can be represented as the sum of several simple harmonic oscillators whose properties vary with vehicle velocity. For each oscillator, both the instantaneous frequency and the amplitude are expressed as functions of the current velocity, and each oscillator is further shaped by an additional amplitude modulation whose frequency and depth also depend on velocity. In practice, this means that every oscillator is defined by four velocity-dependent parameters, namely its carrier frequency, carrier amplitude, modulation frequency, and modulation depth.

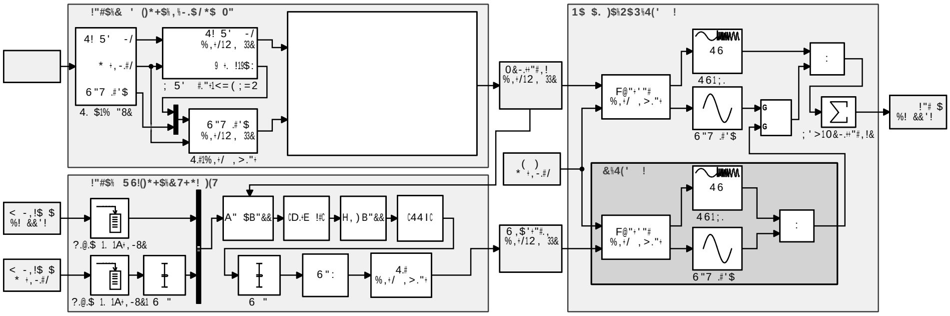


Figure 4.6: Block diagram of additive synthesis analysis and generation algorithm.

To estimate these parameter trajectories from the reference recordings, the velocity-dependent magnitude spectra introduced in Section 4.1.1 are analyzed to identify the dominant tonal components. These components appear as frequency ridges that vary with velocity. However, overlapping partials and missing peaks make direct tracking unreliable. To obtain stable trajectories of the tonal components over velocity, the extracted peak positions are therefore processed using a sequential *Random Sample Consensus (RANSAC)* procedure. RANSAC repeatedly fits simple polynomial models to small random subsets of the data and selects the model that is supported by the largest number of inliers. Applied here, the method yields one polynomial curve for each oscillator's carrier frequency and another polynomial describing its velocity-dependent amplitude. Once the strongest oscillator has been identified, its inlier points are removed, and the procedure is repeated to extract the remaining tonal components.

The modulation characteristics are determined by isolating each tonal component through band-pass filtering, computing the Hilbert envelope, and analyzing its spectrum. The dominant modulation frequency and depth are then fitted with velocity-dependent polynomials in the same way. Together, these polynomial functions describe how each oscillator changes with velocity and form the parameter basis for the additive synthesis model. Figure 4.6 visualizes the structure of the entire additive synthesis model, including both analysis and generation stages.

During synthesis, the frequency trajectories are integrated to obtain phase, the modulation signals are constructed from their instantaneous properties, and the final output is generated as the sum of all oscillators. This approach is particularly suited for AVAS signals dominated by distinct tonal components with clear velocity-dependent behavior. Details on the RANSAC-based parameter estimation and a full mathematical description of the additive synthesis model are given in Paper A.

Sample-based Synthesis

A third source signal generation method, implemented and evaluated in Paper A, is sample-based synthesis. This model was developed to recreate the reversing "pling-ing" AVAS of one of the reference vehicles. Since this AVAS design was not used in the subsequent experiments, this method is of limited relevance to the main results of this thesis. However, it is included briefly here because it demonstrates how the framework can accommodate atypical AVAS designs that do not fit well within subtractive or additive models.

In contrast to subtractive and additive synthesis, sample-based synthesis uses a prerecorded sound sample as the basis for all output signals. Variants of such approaches have been applied previously, for example, to auralize combustion engine noise [47, 48], often using granular techniques or pitch-synchronous overlap-add to modify pitch and time scale. In Paper A, the focus is restricted to the reversing AVAS of one of the reference vehicles, which consists of a simple "pling" sound repeated at constant pitch. In this case, a fully granular framework was not required, and the implemented model modulates only the sound pressure level and repetition rate as functions of vehicle velocity.

The synthesis model generates the output as a sequence of repeated copies of a prerecorded sample, with both the equivalent continuous sound pressure level and the repetition rate modeled as polynomial functions of vehicle velocity, analogous to the additive synthesis approach. For a given velocity profile, these polynomials define the instantaneous repetition rate and target level, which are used to place and scale the sample instances to achieve the desired RMS values. The polynomial descriptions are derived from recorded AVAS signals by manually selecting one period of the "pling" sound as a reference, detecting repetitions via correlation with time-shifted versions of the recording, and assigning the resulting RMS values and inter-event intervals to the corresponding vehicle velocities. These velocity-dependent data are then fit with polynomials, and the repetitions that correlate most strongly with the reference are averaged in the time domain to obtain a clean sample with reduced background noise, which serves as the basis for the synthesis model.

Results

Figure 4.7 shows an exemplary comparison between recorded noise and multi-tone AVAS signals and their re-synthesized counterparts, using the same velocity profile as input to the subtractive synthesis (for the noise AVAS) and the additive synthesis (for the multi-tone AVAS). The noise AVAS is reproduced very accurately, and an

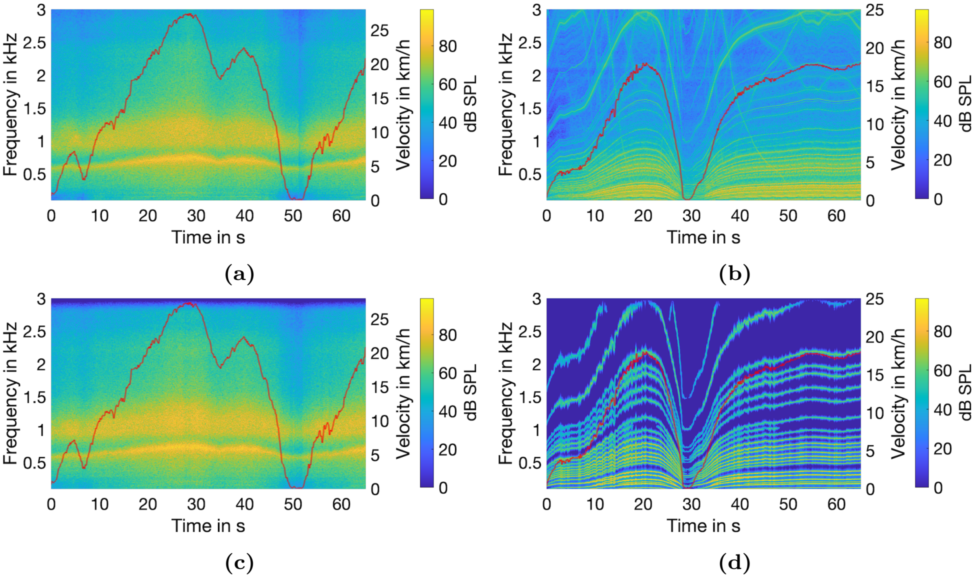


Figure 4.7: Recorded noise AVAS (a) and multi-tone AVAS (b), and corresponding re-synthesized signals using subtractive synthesis for the noise AVAS (c) and additive synthesis for the multi-tone AVAS (d). The red line shows the recorded vehicle velocity, which was used as input for the re-synthesis.

informal perceptual check confirms a high level of plausibility.[4]

For the multi-tone AVAS, the differences between the recorded and resynthesized signals are more pronounced. The most obvious difference is the absence of background noise in the synthesized version, as expected under the additive synthesis model. Although this difference is clearly visible in the spectrograms and audible when listening to the isolated source signals, it is of limited relevance in the full auralization. The background noise present in the reference recording mainly originates from tire-road noise cross-talk, which is undesirable in the isolated AVAS component. In the final auralization, where tire-road noise is added separately, a cleaner synthesized AVAS is therefore preferable. This distinction between validating individual source components and validating the complete auralization system is discussed further in Section 4.2.

A more relevant difference is that the re-synthesized multi-tone AVAS contains fewer tonal components than the reference. This is because the sequential RANSAC procedure could not reliably track more than 27 distinct tones in the velocity-dependent magnitude spectra of this particular recording. While this number could

[4] Audio examples of measured and re-synthesized AVAS signals are available at https://doi.org/10.5281/zenodo.10610490

likely be increased by adjusting the RANSAC peak detection parameters, the perceptual validation in Section 4.2 shows that, once the AVAS is combined with tire-road noise, ambient noise, and the complete propagation model, the reduced number of tones becomes much less perceptually important. In other words, differences that are clearly audible in isolated source signals may have a limited impact on the plausibility of the full auralization.

Limitations and Alternative Approaches

One fundamental limitation of the implemented subtractive and additive synthesis models is their reliance on a single velocity-dependent magnitude spectrum. This implicitly assumes linear behavior, meaning that for a given velocity, the vehicle always radiates the same sound, regardless of whether it is accelerating or decelerating. For the measured vehicles, this assumption holds reasonably well. Still, one can imagine AVAS implementations that use different sounds when starting from rest or braking to a stop. Similar limitations occur in environmental noise applications. For example, a more detailed tire-road noise model would depend not only on velocity but also on acceleration, torque, surface conditions, and vehicle load. For such cases, the current synthesis models could be extended to operate on multidimensional magnitude spectra or on parameter sets that explicitly incorporate additional physical quantities.

A further limitation of the additive synthesis approach is that the sequential RANSAC-based tracking of tonal components can become unreliable for AVAS signals with very dense or strongly overlapping partials. In such cases, the number of oscillators that can be extracted in a stable way depends critically on user-selected thresholds, peak prominence criteria, and the inlier distance used in the RANSAC procedure. This may require careful tuning and does not guarantee that all relevant tonal components are captured. For very complex tonal structures, a different tracking strategy or a more constrained parametric model might be preferable.

Computational aspects also play a role. Additive synthesis with a large number of oscillators is more demanding than subtractive synthesis and may become costly for real-time rendering, especially when instantaneous frequency integration and amplitude modulation are used. Likewise, the current framework does not explicitly model phase relationships or coherence between oscillators, which may be relevant for highly structured tonal AVAS but was not critical for the signals analyzed in this work.

A synthesis approach that was not explored at the time, but may have been a suitable alternative, is spectral modeling synthesis (see Section 2.3). Unlike the separate additive and subtractive methods used here, a spectral modeling framework can represent signals that contain both tonal and noise-like components and could therefore have unified the two methods into a single model. On the other hand, a practical advantage of the implemented additive synthesis is that it yields a compact set of polynomial functions for each oscillator, which can be manually tuned and used to straightforwardly design new AVAS signals. This simplicity and interpretability would be more difficult to maintain in a more complex spectral modeling framework.

4.1.3 Radiation

For the next step in the auralization implementation, the radiation directivity of both AVAS and tire noise needed to be determined. The tire directivity was based on previously published measurements [59] which, although of relatively low angular resolution, were considered sufficient for the purposes of this work. For AVAS directivity, the radiation pattern was expected to be more complex and may vary across vehicle models. To examine this assumption and obtain representative data, pilot measurements were conducted on one electric vehicle. The outcomes of these measurements then motivated the development of a more detailed numerical directivity model, as described in the following subsections.

Pilot Measurements

Figure 4.8: AVAS directivity measurement setup.

To better understand AVAS radiation behavior, pilot measurements were conducted on an exemplary electric vehicle. Although these measurements were not included in Paper A, they are presented here because they form a natural part of the overall methodological workflow, that is, first establishing reference measurements, then determining the level of modeling accuracy required for the intended human response evaluations.

Setup The tested vehicle, a Tesla Model Y 2019, had the unique feature that it allowed playback of arbitrary signals through its AVAS loudspeaker when station-

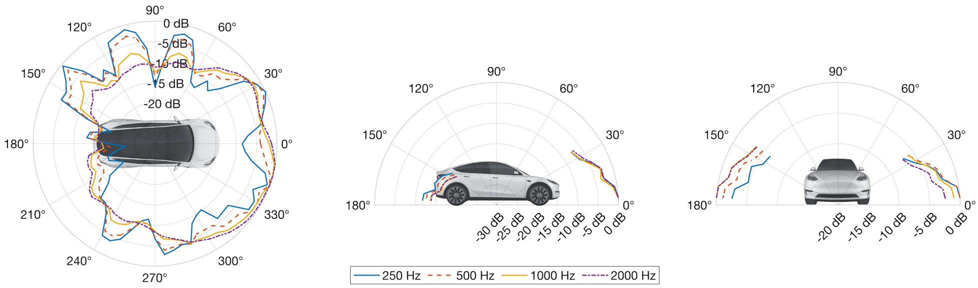


Figure 4.9: Magnitude of measured AVAS radiation directivity in octave bands.

ary.[5] This enabled the playback of a sweep test signal through the AVAS speaker while recording the radiated sound using one microphone positioned directly in front of the loudspeaker and seven additional microphones arranged at 4 m distance in a vertical line, with a spacing of 38.5 cm between adjacent microphones, as shown in Figure 4.8. Transfer functions from the speaker microphone to each remote microphone were computed to estimate the directional radiation behavior. The vertical microphone array was then rotated around the vehicle in 5° increments, yielding full 360° horizontal coverage and approximately 35° elevation coverage.

Results Figure 4.9 shows the magnitude of the measured AVAS radiation directivity pattern for different octave bands. The polar plots indicate that the AVAS exhibits pronounced directivity, with differences of up to 20 dB between the front and rear radiation. This confirmed that AVAS directivity is relevant to include in an auralization model. However, the patterns also revealed inconsistencies. For example, the 250 Hz band appeared more directional than the 2 kHz band, which is physically counterintuitive, since higher frequencies typically show stronger directivity. These irregularities are likely due to reflections and background noise, as the measurements were conducted in a parking lot rather than in an anechoic chamber.

Given the considerable measurement effort per vehicle and the limited angular resolution achievable with this setup, these findings motivated the decision to employ numerical methods rather than additional measurements. The *Boundary Element Method* was therefore used to determine AVAS radiation directivity for the three reference vehicles, as described in the following subsection.

Boundary Element Model

To obtain numerical estimates of the AVAS radiation directivity, all three reference vehicles were modeled using the *Boundary Element Method (BEM)* in COMSOL MULTIPHYSICS 6.1.

[5] Tesla refers to this feature as “Boombox Mode”.

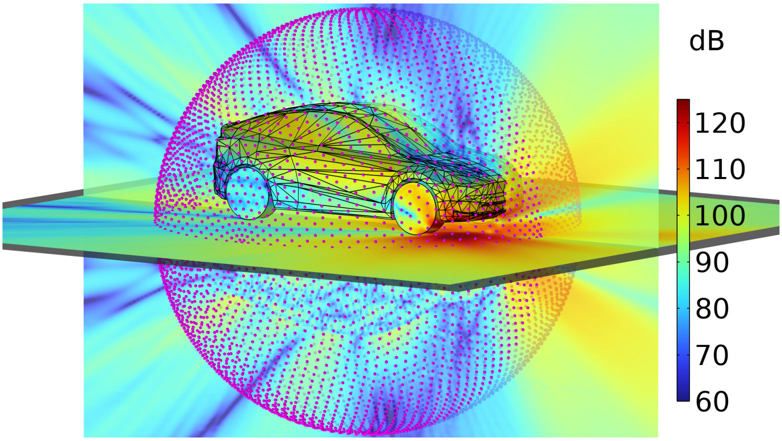


Figure 4.10: Simplified 3D model of electric vehicle with BEM results for radiated sound pressure at $f = 2\,\text{kHz}$ and evaluation points on Lebedev grid. The mirrored pressure below the ground plane is due to the symmetry boundary condition used to model an infinite sound-hard ground.

Setup The simulated vehicle geometries were based on simplified commercially available 3D models of the three reference vehicles, in which the AVAS loudspeaker was represented as a circular disk with a radius of 5 cm embedded in the vehicle chassis. The disk was driven with a velocity proportional to $1/(j\omega)$, which corresponds to a monopole source with constant sound pressure level in free field. A sound hard ground plane was included using a symmetry boundary condition, and a simple porous impedance model was assigned to the vehicle floor to suppress resonances between the car body and the ground. The radiated pressure field was evaluated on a 131st order *Lebedev grid* [168] with 5810 points distributed on a sphere of 3 m radius around the vehicle, as illustrated in Figure 4.10. Since the complex pressure on this enclosing surface is known, the field can be extrapolated to arbitrary receiver positions using *spherical harmonic expansion*, as described in the following subsection. The symmetry boundary condition yields a mirrored pressure field beneath the ground, which naturally produces correct ground reflections when extrapolating the pressure beyond the sphere.

The simulations were solved up to 3 kHz in 30 Hz steps. Transformed to the time domain, this yields impulse responses from the AVAS source to points on the evaluation sphere with a sampling rate of 6 kHz and a duration of 33.3 ms. The upper frequency limit of 3 kHz was chosen due to computational constraints and is acceptable for the present work, as the measured AVAS signals (except for the multi-tone AVAS) contained no energy above this range. Higher-frequency patterns could be obtained by increasing computational resources or by combining the BEM model with a high-frequency method, such as ray tracing. Further details on the numerical setup, discretization choices, and convergence considerations are provided in Paper A.

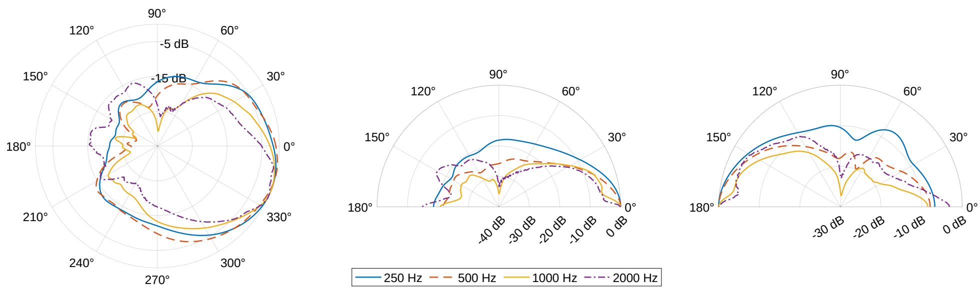


Figure 4.11: BEM results for AVAS directivity magnitude of Tesla Model Y 2019 in octave bands.

Results Figure 4.11 shows polar plots of the calculated AVAS radiation directivity for one of the evaluated electric vehicles, namely the same vehicle used in the pilot directivity measurements described in the previous subsection. The results indicate that, in the horizontal plane, the radiation is concentrated towards the front right of the vehicle, and that the frontal radiation is consistently skewed in this direction across all evaluated frequency bands. This is consistent with the physical configuration, in which the AVAS loudspeaker is mounted on the right side of the front bumper, and it aligns with the overall trend observed in the pilot measurements.

There are, however, noticeable differences in the finer details of the radiation patterns obtained from the boundary element model compared to the pilot measurements. These discrepancies are likely due to simplifications in the vehicle geometry, uncertainties in material properties, and the idealized boundary conditions used in the simulations. Despite these limitations, the results clearly show that AVAS radiation in the relevant frequency range is not omnidirectional but is predominantly directed toward the front right. For the purposes of this thesis, it was therefore deemed appropriate to include AVAS directivity in the auralization model, and later-described listening experiments confirm that AVAS directivity can significantly influence pass-by perception.

Spherical Harmonic Encoding

To represent the radiation characteristics obtained from BEM simulations and tire measurements in a compact, flexible manner, the pressure values on the evaluation sphere were encoded using *Spherical Harmonics (SH)*. This representation not only simplifies the handling of radiation directivity, but also replaces the physical AVAS speaker and tire sources with an *equivalent source description* that can be used ef-

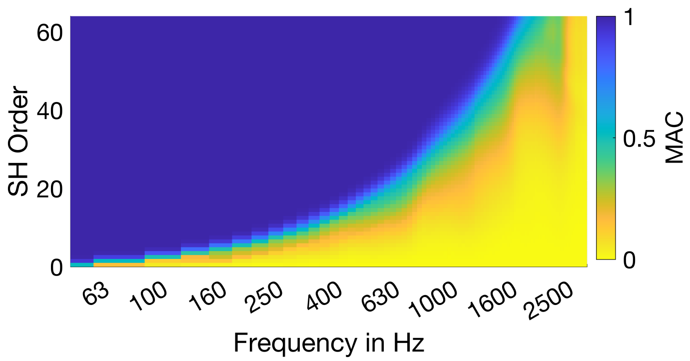


Figure 4.12: Modal assurance criterion between extrapolated SH directivities and corresponding BEM results as a function of SH order and frequency.

ficiently in subsequent propagation calculations via SH extrapolation. The basic principles of spherical harmonics and their use for representing directional sound fields were introduced in Section 2.4. In this context, the pressure values simulated on the sphere surrounding the vehicle are expressed as a weighted sum of spherical harmonic basis functions, with the weights given by the spherical harmonic coefficients.

A practical advantage of this representation is that the spatial resolution of the directivity can be controlled simply by choosing the maximum spherical harmonic order. Higher orders allow for finer spatial detail, while lower orders produce a smoother, more diffuse pattern. Importantly, reducing the spatial resolution does not require re-running the BEM model or re-interpolating the measured data. It is achieved by truncating the spherical harmonic expansion, which makes the representation both compact and computationally efficient.

The number and distribution of pressure samples on the evaluation sphere determine the maximum order that can be stably encoded. Based on the dense Lebedev grid used in the simulations, the AVAS directivities were encoded up to spherical harmonic order 64. The tire directivity measurements, which had a lower spatial sampling density, were encoded up to order 16. Previous studies [169, 170] indicate that these orders provide sufficient spatial resolution for auralization purposes. A full mathematical description of the spherical harmonic encoding procedure, including the least squares solution used to estimate the coefficients, is provided in Paper A.

One way to assess how well the spherical harmonic representation reproduces the simulated sound field is to calculate the modal assurance criterion (MAC) between the BEM ground truth and the spherical harmonic reconstruction. This measure is similar to a correlation coefficient, but it captures both magnitude and spatial similarity. Figure 4.12 shows the MAC for one of the evaluated vehicles as a function of frequency and spherical harmonic order. Values close to 1 indicate a very good match, while values near 0 indicate low similarity. The results show that relatively high spherical harmonic orders are required to represent the high-frequency range accurately for this particular directivity.

This behavior can be explained by the acoustic center of the source, i.e., the AVAS loudspeaker, not coinciding with the center of the evaluation sphere [171] (see Figure 4.10). In that case, the pressure pattern on the sphere becomes more complex, requiring a higher spherical harmonic order to describe it. To quantify this effect, an additional set of directivities was computed using the Lebedev evaluation grid, centered on the AVAS loudspeaker rather than the vehicle center. This shift reduced the required spherical harmonic order considerably, to about 10 to 24, depending on the vehicle. While the resulting auralizations did not differ perceptually in the scenarios considered here, the speaker-centered encoding is more efficient, since it achieves the same perceptual quality with fewer coefficients and lower computational cost.

However, because scattering from the vehicle body is part of the effective sound source, the evaluation sphere must still enclose the entire car. Centering the sphere on the front-mounted AVAS loudspeaker, therefore, requires a larger radius, which in this case increases the minimum auralization distance from approximately 3 m to 5 m. Speaker-centered directivities are thus less suitable for very close vehicle passages but attractive when only larger source receiver distances are of interest. In this thesis, both options are employed, depending on the specific human response experiment. For applications that focus on near passages, the vehicle-centered encoding is preferable, whereas studies that only require larger distances can benefit from the more efficient speaker-centered representation.

Limitations

The boundary element models used to estimate AVAS radiation directivity are based on simplified vehicle geometries and do not include detailed material impedance properties. This means that, although the method, in principle, allows highly specific predictions, the underlying simplifications limit the physical accuracy of the resulting patterns. The directivities should therefore be regarded as plausible approximations of the main radiation trends rather than exact replicas of the real vehicles.

A further simplification is that the AVAS loudspeaker itself is not modeled with its detailed electroacoustic characteristics. This was a deliberate choice. Since the reference AVAS signals in Section 4.1.1 were obtained by placing a microphone directly in front of the loudspeaker, the measured source signals already include the influence of the specific driver and enclosure. For the present work, this is advantageous because it allows the numerical model to focus on the scattering and radiation of the vehicle body. However, this assumption would no longer hold if manufacturers' digital AVAS signals were used as input rather than measured loudspeaker signals.

Finally, the ground in the BEM simulations is assumed to be perfectly sound hard.

Given the AVAS loudspeaker's position in the front bumper, ground reflections are expected to contribute substantially to the sound at typical receiver positions, and different road surfaces could modify the combined directivity in a non-negligible way. A more flexible approach would be to omit the ground in the BEM model, compute free field directivities, and then include ground effects using an image-source method with frequency-dependent reflection coefficients. Such an extension would be appropriate for studies that specifically target the influence of ground surfaces on perceived AVAS noise. For the experiments in this thesis, which focus on relative differences between AVAS designs under controlled conditions, the hard ground assumption is considered sufficient, and including reflections in the spherical harmonic representation substantially simplifies implementation compared with an image-source approach.

4.1.4 Propagation

In the context of electric vehicle auralizations, the propagation model must account for distance-dependent attenuation, atmospheric effects such as air absorption and turbulence, and the source's motion relative to the listener. These aspects are described in the following subsections. The scope of this thesis is limited to free field outdoor propagation at distances relevant for EV applications, that is, without reflections from buildings, diffraction over noise barriers, and without strong wind or temperature gradient effects. Propagation through building facades, as used in Paper E and Paper F, is treated as part of the reproduction stage rather than the propagation model itself and is therefore discussed in Section 4.1.5 rather than in this section.

Spherical Harmonic Extrapolation

A useful property of the spherical harmonic representation is that, once the sound field is known on a closed surface surrounding the source, the pressure at any point outside this surface can be obtained by extrapolation. This follows from the general solution of the Helmholtz equation, which states that the exterior sound field is uniquely determined by the pressure or normal velocity on an enclosing surface. In practice, this means that the spherical harmonic coefficients obtained from the BEM evaluation sphere can be scaled with spherical Hankel functions to compute the pressure at arbitrary receiver positions. The underlying mathematics of this scaling is described in Paper A.

An important advantage of this approach is that all scattering effects included within the evaluation sphere, including the ground reflection in the BEM simulations, are automatically preserved during extrapolation. This behavior is illustrated in Figure 4.12, which compares the pressure obtained from spherical harmonic extrapolation at a distance of 6 m with the corresponding reference values computed directly by the BEM model. Apart from the expected limitations at high frequencies due to the finite spherical harmonic order, the agreement at lower frequencies is, within numerical precision, perfect. This confirms that geometric spreading and ground reflection are correctly represented within the spherical harmonic framework.

The result of the extrapolation step is a transfer function that describes the propagation from the source position, as represented in the BEM model, to a desired receiver point. After transforming this transfer function into the time domain, the resulting impulse response can be convolved with the synthesized source signal. This single convolution step therefore applies directivity, ground reflection, and distance-dependent attenuation in a unified manner.

Atmospheric Effects

The purpose of including air absorption and atmospheric turbulence in the auralization was to enhance overall plausibility, rather than to enable detailed studies of atmospheric parameters or to reproduce specific meteorological conditions. The implementations were therefore kept deliberately simple.

Atmospheric absorption was modeled according to ISO 9613-1:1993 [76] by attenuating individual third octave bands of the source signal as a function of the instantaneous distance between source and receiver. For Paper A, Paper C, Paper E, and Paper F, only this distance-dependent air attenuation was applied, since the scenarios mainly involved close-proximity vehicle passages.

For Paper B and Paper E, the auralization was extended to also account for atmospheric turbulence. This was done by applying a time-variant filter whose amplitude and phase fluctuations were modeled using a *von Kármán turbulence model.* Random amplitude and phase perturbations with a spatial correlation consistent with the von Kármán spectrum were generated and mapped onto the moving source trajectory based on its instantaneous position. The fluctuations increase with frequency and propagation distance, and their variance was scaled to physically plausible values.

The resulting frequency-dependent perturbations were converted into a set of time-varying impulse responses, which were applied to the source signal sample by sample. This introduces small, slowly varying changes that mimic the random scattering and phase wandering caused by atmospheric turbulence in outdoor sound propagation. The underlying model and its relation to von Kármán statistics follow the derivations in [172] and [173] and are not further detailed here.

In principle, the time-varying turbulence filters could have been merged with the propagation impulse responses into a single processing stage. However, this would have reduced modularity and increased implementation complexity. Instead, atmospheric effects were implemented as a separate block in the auralization chain, thereby making it straightforward to isolate, disable, or replace this component as required for a specific study.

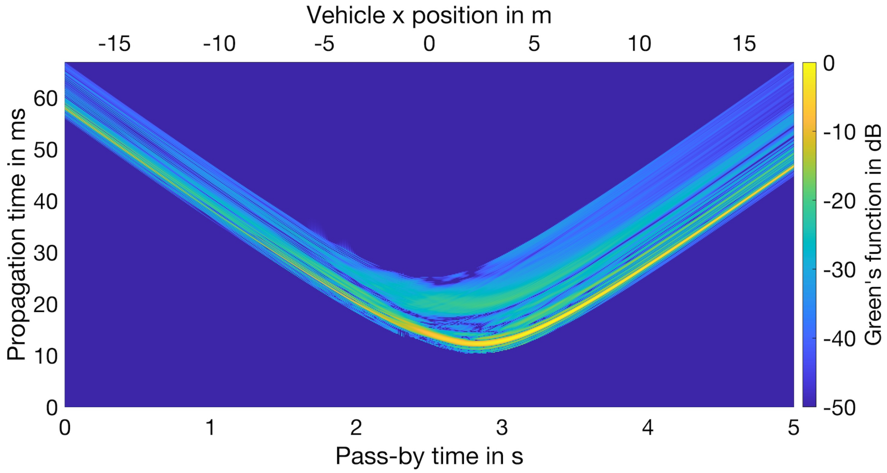


Figure 4.13: Green's functions obtained from spherical harmonic extrapolation, describing source–receiver propagation for each discrete source position of a linear 25 km/h pass-by.

Source Movement

The final component of the propagation model is the inclusion of source movement. This requires that the changing source-receiver geometry is represented correctly and that the *Doppler effect*, that is, the frequency shift caused by relative motion [79], is reproduced. Although Doppler shifts are relatively small for the velocities relevant to EV and AVAS applications, they can still be perceptually noticeable, particularly for tonal components.[6]

A practical way to model a moving source with an arbitrary radiation pattern is to compute an individual *Green's function*, that is, the impulse response describing the propagation from source to receiver, for each discrete source position along the trajectory, as illustrated in Figure 4.13. This approach, also referred to as the moving Green's functions method, assumes that at each instant the source behaves like a stationary radiator at its instantaneous position, avoiding the need for analytical solutions for moving sources [174]. Since the Green's function is updated for every discrete source position, the method is not restricted to constant-velocity motion but can also handle arbitrary, accelerating trajectories. In this formulation, the Doppler effect does not need to be implemented explicitly as it emerges naturally from the changing source–receiver distance and the corresponding sequence of Green's functions. In this thesis, the Green's functions are obtained from the spherical harmonic extrapolation described earlier in this subsection.

[6]For instance, a 1 kHz tone emitted by a vehicle moving at 30 km/h is shifted to $f' = f_0 \frac{c}{c-v} = 1000\,\mathrm{Hz} \cdot \frac{343\,\mathrm{m/s}}{343\,\mathrm{m/s} - 8.33\,\mathrm{m/s}} \approx 1025\,\mathrm{Hz}$, which corresponds to roughly a quarter-tone shift. This is clearly audible, especially for a pure tone signal.

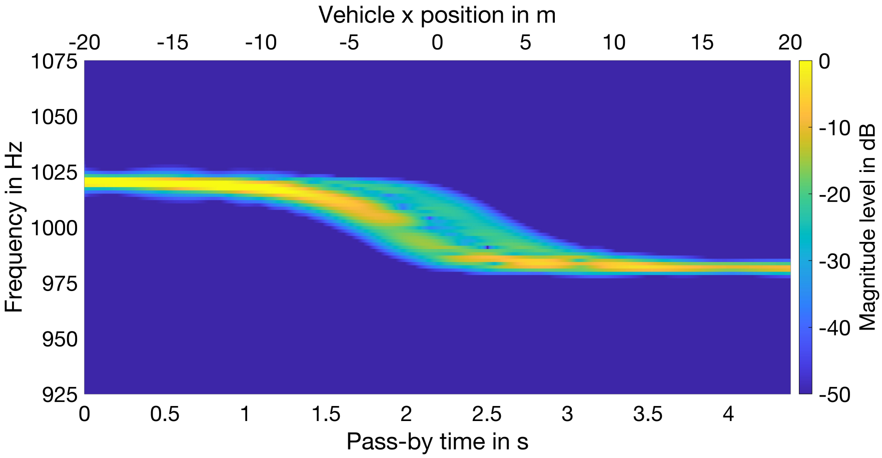


Figure 4.14: Spectrogram of sound pressure at the receiver position for a linear 25 km/h pass-by with a 1 kHz source signal, simulated using moving Green's functions from spherical harmonic extrapolation of the BEM radiation results.

The source trajectory is discretized at the temporal resolution of the auralization. For example, simulating a 5 s pass-by at 25 km/h with a sampling rate of 6 kHz results in 30,000 discrete source positions. The fundamental requirement for the validity of this approach is that the movement during a single time step must remain small compared to the wavelength of the signal [79–81]. This condition is fulfilled for all velocities relevant to EV applications, where $v \ll c$. For each discrete position, the corresponding Green's function describes the propagation from source to receiver. Figure 4.13 visualizes such a set of Green's functions for a 25 km/h pass-by at a lateral distance of 3 m.

To obtain the sound pressure at the receiver, each sample of the source signal is convolved with the Green's function corresponding to its source position. The convolution results are then combined sequentially, sample by sample, which can be interpreted as time-variant filtering. A full mathematical description is provided in Paper A. Figure 4.14 illustrates the result of this procedure for a 1 kHz pure tone. The expected Doppler shift is clearly visible, as is the amplitude variation caused by the source directivity and the changing source-receiver geometry.

Limitations

The propagation model based on spherical harmonic extrapolation, simple atmospheric modeling, and moving Green's functions has several limitations. A first constraint is that spherical harmonic extrapolation can only provide sound pressure outside the evaluation sphere used in the BEM calculations. When the sphere is centered on the vehicle and scaled to enclose the entire body, for example, with a radius of 3 m, auralizations at smaller source-receiver distances are not possible. This is generally acceptable for EV pass-by scenarios but may be restrictive for very close-range situations, such as parking-lot maneuvers.

A second limitation concerns computational effort. The combination of spherical harmonic extrapolation and moving Green's functions is demanding, particularly for higher spherical harmonic orders or long source-receiver distances. Although this prevents real-time rendering, the approach has the advantage that all Green's functions for a given trajectory can be precomputed once and then reused for multiple AVAS signals under identical propagation conditions.

For large distances, the required Green's functions become very long because they must contain the full propagation delay. Directly extending the BEM bandwidth to support such long impulse responses would be computationally prohibitive. In this thesis, this was mitigated by removing the propagation delay in the frequency domain, zero-padding the time-domain Green's functions to the required length, and then reintroducing the delay afterward. This allows short BEM-derived impulse responses to be used for longer source-receiver distances, although the final Green's functions still need to be long enough to represent the full delay. While manageable for the distances and sampling rates relevant to EV applications, this becomes impractical for scenarios such as aircraft flyovers, where Green's functions would span several seconds.[7] In such cases, interpolation schemes or a separate treatment of the propagation delay would be required.

Finally, the moving Green's functions approach is restricted to subsonic source velocities. At $v = c$, the Doppler formula exhibits a singularity. For $v > c$, the sound field forms a Mach cone with multiple retarded times, which cannot be represented by a simple sequence of stationary Green's functions [79, 81]. For $v < c$, the method is in principle exact in the continuous-time limit. However, at finite sampling rates, the increasingly compressed wavefronts ahead of the source require sufficiently fine temporal discretization. The practical requirement that the source displacement per time step remains small compared with the wavelength is comfortably fulfilled for all velocities relevant to road-traffic noise, where $v \ll c$. It is also worth noting that, as implemented, the method assumes a moving source and a stationary receiver. For the reversed case, the receiver position at the time of sound arrival rather than at the time of emission would need to be used, which introduces an implicit retarded-time problem that is not present in the moving-source formulation.

4.1.5 Encoding and Reproduction

The final element of the auralization implementation is the encoding and reproduction stage. In this step, the different source contributions are combined, ambient

[7] Auralizing a 30 s aircraft flyover with up to 2 km distance at a sampling rate of 10 kHz would result in a matrix of Green's functions with $30\,\mathrm{s} \cdot 10\,\mathrm{kHz} \times 10\,\mathrm{kHz} \cdot \frac{2000\,\mathrm{m}}{343\,\mathrm{m/s}} \approx 17.5 \times 10^9$ values, which, using Matlab's `double` data type, requires about 130 GB of memory.

noise is added, and the resulting signals are encoded in a format suitable for the chosen reproduction system. The reproduction setup is adapted to each listening experiment, depending on the research question and practical constraints.

In the broader methodological framework described in Chapter 2, encoding and reproduction are treated as conceptually distinct stages. In the present application, however, they are closely interlinked, and for clarity and continuity, they are therefore described together in a single section. The following subsections outline the approaches used in this thesis, namely reproduction via static headphones, binaural crosstalk cancellation, point sources in a circular loudspeaker array, and wave field synthesis in the *Living Room Lab*.

Level Calibration and Mixing

For the final auralization of an electric vehicle pass-by, the AVAS signal and four separate tire-noise receiver signals (i.e., the outputs of the previous processing stages) are mixed with ambient background noise at realistic levels. In principle, the source signal synthesis is based on calibrated reference measurements and should therefore produce plausible sound pressure levels. However, radiation directivity models, which link synthesized source signals to the velocity of a vibrating disk, cannot be assumed to preserve the absolute level. The same applies to the tire directivities, which are based on measurements that are not absolutely calibrated. As a result, the relative level of tire noise relative to AVAS must be adjusted.

In this thesis, this adjustment was performed using a recursive procedure based on roadside reference measurements described in Section 4.1.1. The recorded vehicle velocity profile was used as input to the full auralization chain, re-synthesizing the AVAS and tire noise components. In an ideal case, their sum would exactly match the measured roadside pressure signal. To determine suitable tire levels, auralization results for different tire-to-AVAS gain settings were compared with the reference recording in third-octave bands, and the gain combination that minimized spectral error was selected. This calibration step ensures that all subsequent listening experiments are based on a consistent, physically plausible spectral balance among AVAS, tire noise, and ambient background. To further refine the spectral balance, an optional octave-band equalization filter can be applied to ensure the resulting spectrum matches the reference as closely as possible. This equalization is determined once for a given vehicle and then reused for other velocities, compensating for systematic spectral imbalances in the synthesis and radiation models. The ambient background noise recorded at the roadside is already calibrated and was therefore added at its original level.

Figure 4.15 shows an exemplary comparison between a measured vehicle pass-by and its corresponding resynthesized version. Both the spectrograms and the numerical comparisons reported in Paper A indicate that the auralizations agree well with the reference on a numerical level. A system-wide perceptual validation of the

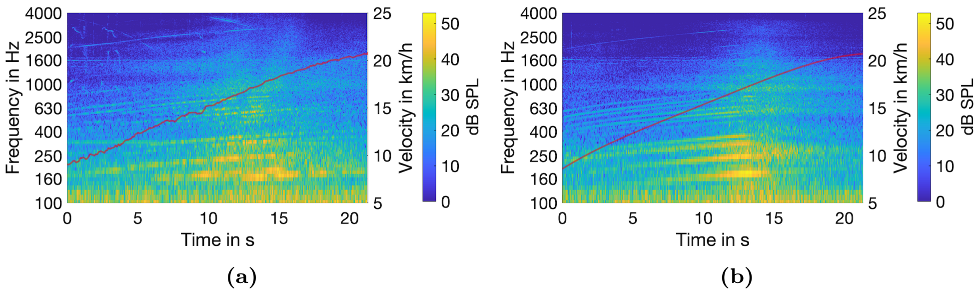


Figure 4.15: Recorded (a) and auralized (b) sound pressure at roadside observer position, including multi-tone AVAS, tire-road noise, and ambient background noise.

complete auralization chain is presented in Section 4.2.1. Informal listening suggests an overall good similarity between recording and reproduction, with the exception that some fine temporal details of the tire noise, such as crackling components, are not fully captured by the subtractive synthesis model.[8]

Headphone-based Reproduction

For static headphone-based reproduction, the mono pressure signal at the receiver position is converted to a binaural signal by convolution with head-related transfer functions (HRTFs) that depend on the instantaneous source direction. A simple nearest neighbor strategy is used to select the most suitable HRTF for each source angle. In practice, the synthesized receiver signal is first resampled to match the HRTF sampling rate, which is 48 kHz in this case. The signal and the corresponding source position vectors are then divided into short blocks, for example, of 512 samples, and each block is convolved with the HRTF corresponding to the mean source direction within that block. This yields two ear signals per source. To avoid clicking artifacts when the selected HRTF changes from block to block, the previous HRTF is also applied to each block, and the two results are crossfaded using overlapping windows.

In the current implementation of the auralization framework, AVAS and each of the four tire-noise contributions are treated as separate sources at distinct positions. Each of these signals is therefore convolved with its own direction-dependent HRTF before the level calibration and mixing described above. This HRTF-per-source approach is adequate in the far field, where each physical source can be approximated as a point source in a single direction relative to the listener. For very close distances, however, this simplification becomes less accurate. A more precise approach would be to reconstruct the full sound field at the listener's head position, for example, on a spherical grid. The resulting set of pressure signals can then be treated as the inputs

[8]The reader can listen to these comparisons at https://doi.org/10.5281/zenodo.10610490

to a multiple-input, multiple-output system, with the left- and right-ear signals as the two outputs. The mapping from the sampled sound field to the ear signals is obtained by determining a set of transfer functions that best reproduce measured HRTFs in a least squares sense [36]. This would ensure that scattered contributions from the vehicle body are represented as incident from multiple directions rather than being collapsed to a single angle per physical source. For the studies in this thesis, these near-field effects on apparent source width and detailed scattering perception were neglected. Given the focus on relative differences between AVAS designs and the moderate source distances used, this simplification was considered acceptable. For applications that explicitly target the perceptual effects of vehicle scattering or the optimization of AVAS loudspeaker placement within the chassis, a more advanced binaural rendering based on sound field reconstruction would be preferable.

In principle, the same framework could be extended to dynamic headphone reproduction by tracking the listener's head orientation and updating the source directions accordingly when selecting HRTFs. In this thesis, the only experiment that employed headphone-based reproduction was the perceptual validation in Paper A, where the auralizations were compared to static binaural reference recordings. In that context, introducing head tracking would have asymmetrically altered the reproduction conditions and was therefore undesirable.

Binaural Crosstalk Cancellation

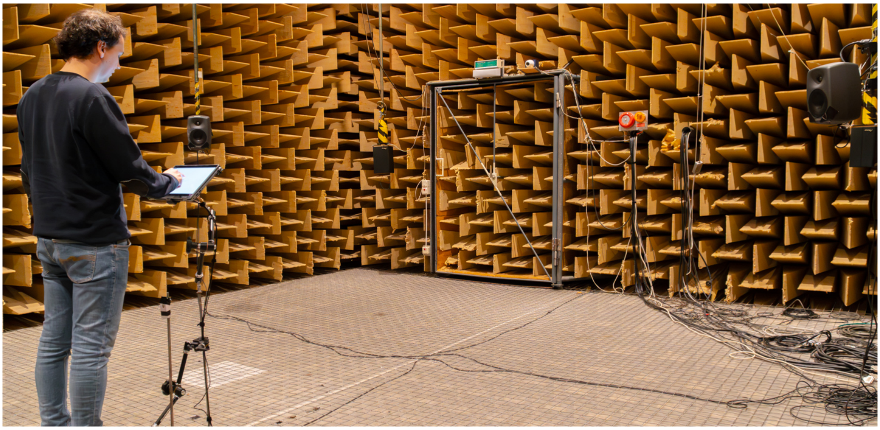


Figure 4.16: Binaural crosstalk cancellation setup with Paper C listening experiment interface in anechoic chamber.

A closely related reproduction method to the headphone-based setup described above is binaural crosstalk cancellation. In this approach, the same binaural signals that would normally be presented over headphones are reproduced via a pair of

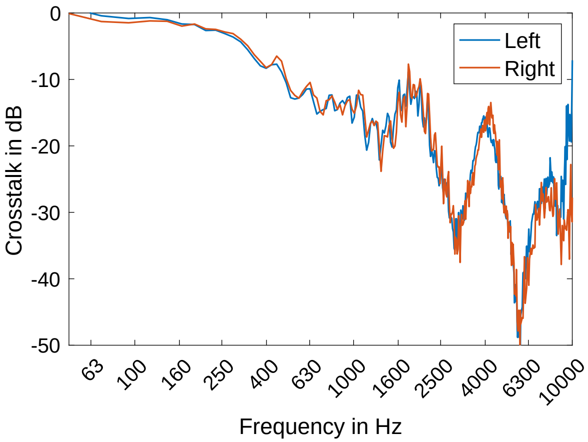


Figure 4.17: Inter-channel crosstalk level measured for setup shown in Figure 4.16.

loudspeakers, while inter-channel crosstalk is reduced by appropriate filtering. The main advantage is that listeners do not need to wear headphones, which can increase comfort and perceived immersion, at the cost of requiring a relatively fixed head position and a more constrained listening geometry. However, this restriction matches the stationary roadside listener scenario in Paper C, which is why it was employed there.

For the implementation used in this thesis, two loudspeakers were mounted in an anechoic chamber at a distance of 3.1 m from the listening position and at azimuth angles of $\pm 50°$, as shown in Figure 4.16. To evoke the impression of standing at the side of a road, participants stood throughout the experiment and were instructed to face the virtual road straight ahead. The transfer functions from each loudspeaker to each ear were measured using an artificial head placed at the listener position. These four transfer functions form a 2×2 matrix that describes how the two loudspeaker signals map to the two ears.

Crosstalk cancellation filters were obtained by inverting this loudspeaker-to-ear transfer matrix in the frequency domain using a least squares solution with Tikhonov regularization [90, 175]. Applying these filters to the binaural signals reduces the unwanted contribution of each loudspeaker to the contralateral ear and approximates the ear signals that would be obtained with ideal headphone reproduction. Exact implementation details are given in Paper C.

To assess the technical performance of the implementation, single-channel white noise was passed through the crosstalk cancellation filters and reproduced, such that, in the ideal case, the signal would reach only one ear. The level of the residual signal measured at the opposite ear was interpreted as crosstalk and expressed relative to the desired signal, as shown in Figure 4.17. The results indicate that channel separation generally increases with frequency and that crosstalk levels below -20 dB were achieved over most of the frequency range of interest. These measurements, however, do not fully determine the method's perceptual accuracy, as additional factors, such as individual HRTF differences, may also influence perceived localization and externalization [176, 177]. Therefore, a dedicated perceptual validation of the reproduction method was conducted, as described in Section 4.2.3.

Free-field Circular Loudspeaker Array

Figure 4.18: Circular loudspeaker array setup in anechoic chamber.

Both the headphone-based reproduction and the binaural crosstalk cancellation methods described above rely on generic HRTFs or require individualized HRTF measurements. Since individual HRTFs could not be obtained for each study participant with the available resources, and since generic HRTFs inevitably come with the risk of reduced localization accuracy for some listeners [87], these approaches were unsuitable for one of the central studies of this thesis, namely the listening experiment on AVAS localization. For that experiment, the reproduction method needed to resolve the source direction as accurately as possible without introducing additional artifacts. Because the study focused solely on horizontal localization, a dedicated circular loudspeaker array was chosen as the reproduction method.

A ring of 24 studio loudspeakers with a radius of 3 m was installed in an anechoic chamber and hidden behind an acoustically transparent curtain, as shown in Figure 4.18. In theory, such an array can reproduce two-dimensional Ambisonics up to order $L = 11$, which corresponds to a main lobe width of approximately

$\Delta\varphi \approx 360°/(2L+1) \approx 15.7°$ for a synthesized beam [84]. For a listener at the array center, the corresponding spatial aliasing frequency is on the order of $f_{\text{alias}} \approx \frac{L\,c}{2\pi r}$, which, for $L = 11$, $r = 3\,\text{m}$, and $c = 343\,\text{m/s}$, yields approximately 200 Hz. Above this frequency, the reproduced sound field is increasingly affected by spatial aliasing, so the effective usable order at audio frequencies is significantly lower than the theoretical maximum. Although the nominal angular resolution is high, the combination of finite order and aliasing means that the reproduced sources would still be broader and less stable than a single physical AVAS loudspeaker, making this approach unsuitable for the intended AVAS localization study.

An alternative would have been to use *Vector Base Amplitude Panning (VBAP)* [85] to place virtual sources between loudspeakers. While VBAP provides continuous azimuth control, the perceived source width varies depending on whether the virtual source lies close to or between loudspeakers. For the localization study, which aimed to measure very small differences in localization accuracy, such order-dependent changes in apparent width were considered undesirable.

For these reasons, a simpler and more robust approach was chosen: each virtual vehicle was reproduced by a single physical loudspeaker. This method avoids Ambisonics-related artifacts entirely and provides point-like sources whose spatial accuracy is not limited by modal order or aliasing frequency, aside from physical loudspeaker characteristics. It restricts reproduction to 24 discrete azimuths and static vehicle positions, which was sufficient for the experimental design. The loudspeakers were concealed behind a curtain, preventing participants from perceiving or inferring these discrete positions. One remaining approximation is that a loudspeaker at 3 m does not reproduce the exact curvature of a spherical wave that a source at 7 m would produce. Consequently, distance cues related to listener movement are not reproduced accurately. However, listeners were physically restrained to the center of the array, so this limitation did not affect the intended perceptual task.

In addition to the vehicle sounds rendered as point sources, a fourth-order Ambisonics ambiance recording from a quiet parking lot was reproduced over the full loudspeaker array. The combination of high-precision point-source rendering for the vehicles and a spatially diffuse Ambisonics ambiance created an immersive and realistic acoustic scene. Although this reproduction method was not formally validated, the fact that study participants achieved mean localization errors below 3° for several stimuli demonstrates that the setup supported localization performance near the limits of human auditory resolution [87]. Further details on this implementation and the experimental results are provided in Paper D.

Wave Field Synthesis in *Living Room Lab*

Investigating the human response to indoor road traffic noise poses a specific challenge for auralization and sound reproduction. A key difficulty is that sound trans-

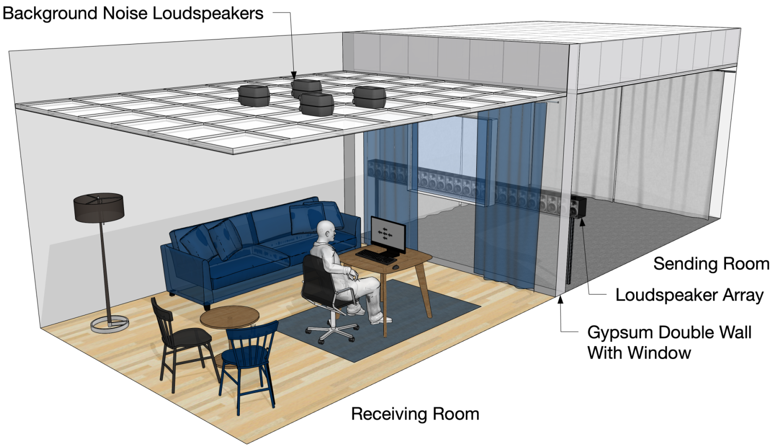


Figure 4.19: Visualization of Chalmers *Living Room Lab*.

mission through a facade and window depends strongly on the angle of incidence, particularly around coincidence [74]. Although this behavior can be simulated numerically, for example, by combining facade transmission models with a room acoustic simulation and applying the result to a headphone-based reproduction, such an approach requires extensive modeling effort and substantial computational resources. In addition, headphone-based playback may reduce immersion. Simpler filter-based approximations of facade transmission would be computationally easier, but would not recreate the spatial characteristics of a radiating window, such as the perception of a moving vehicle indoors. Using only a small number of loudspeakers inside the listening room, as done in related studies, may be adequate for some metrics, but it does not reproduce the directional characteristics of an outdoor pass-by or the correct transmission through a real facade.

This illustrates a general challenge in auralization for human response research. Without dedicated studies, it is difficult to predict whether spatial aspects of the indoor sound field influence annoyance, attention, or physiological response. It is also difficult to obtain controlled in situ recordings that include the exact facade, vehicle, and driving trajectory as reference material. The reproduction method used in this thesis was therefore designed not for a single, narrow response measure, but to support a broad range of perceptual and physiological experiments. The aim was to reproduce the indoor sound field as accurately as possible, so that potential influences on spatial fidelity could be investigated explicitly rather than assumed negligible.

For this purpose, the present implementation adopts a hybrid approach that combines wave field synthesis (see Section 2.7) with a physical facade and window. This is implemented in the Chalmers *Living Room Lab*, a transmission suite comprising

(a) (b)

Figure 4.20: Receiving (a) and sending room (b) of *Living Room Lab*.

two rooms separated by a double gypsum wall with a real window. A linear 24-channel loudspeaker array in the sending room synthesizes the outdoor sound field of a passing vehicle in front of the window. The resulting sound interacts with the real facade and window, meaning that the indoor field includes the true angle-dependent transmission, the physical vibration behavior of the window, and the spatial distribution of sound inside the receiving room. Figure 4.19 and Figure 4.20 illustrate this setup. Because the receiving room is very quiet, with background noise levels below 12 dBA, hidden loudspeakers in the ceiling introduce a controlled background noise. The detailed implementation of the wave field synthesis, including outdoor propagation and AVAS radiation directivity, is described in Paper B.

A fundamental limitation of the array is its finite length, which leads to deviations from the desired wave field, particularly amplitude errors at large incidence angles. Additional limitations arise from reflections in the sending room and from the array's limited spatial resolution. To analyze the impact of these limitations and explore possible improvements, Paper B includes a numerical model of the *Living Room Lab*. This model uses a modified modal superposition approach to compute Green's functions from the loudspeakers to the window and into the receiving room. The window is modeled as a simply supported plate, whose velocity response and radiation into the receiving room are computed. Figure 4.21 shows an example in which the array synthesizes a 600 Hz plane wave. The model calculates the resulting pressure in the sending room, the window vibration, and the transmitted pressure in the receiving room.

This simulation framework enables a systematic evaluation of different auralization strategies. Paper B uses it to examine the role of sending room reflections and to compare alternative array geometries by propagating the simulated loudspeaker output through the window to the listener position and generating binaural signals for both numerical and perceptual validation. The perceptual validation results are summarised in Section 4.2.3.

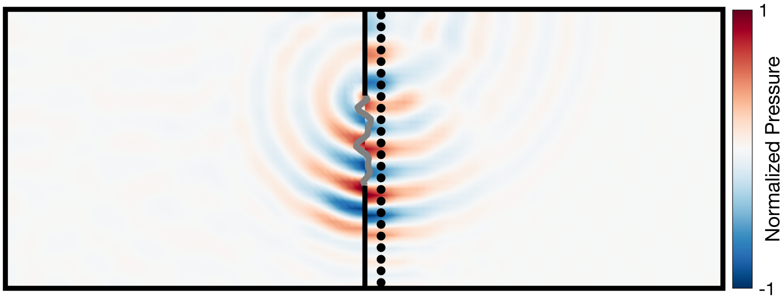


Figure 4.21: Simulation result for exemplary sound field in *Living Room Lab*.

Limitations

The encoding and reproduction stage described in this section is subject to several general limitations in addition to the method-specific aspects already discussed. For the headphone-based reproduction, the binaural crosstalk cancellation, and the circular loudspeaker array, the reproduction is optimized for a single, well-controlled listener position. This sweet-spot assumption is appropriate for laboratory experiments with seated or standing participants, but it does not capture the range of listener positions and movements observed in real traffic environments. In contrast, the wave field synthesis-based reproduction in the *Living Room Lab* aims to provide a plausible sound field throughout the receiving room. However, it remains limited to a single facade and room geometry and does not capture the full variability of real dwellings.

All binaural reproduction stages in this thesis rely on generic HRTFs and do not account for individual anatomical differences. This choice was dictated by practical constraints and is common in spatial audio research. However, it may introduce additional variability in localization and externalization that is not directly related to the auralization methods themselves. Similarly, the loudspeaker-based methods assume ideal loudspeaker behavior within the calibrated frequency range and do not explicitly model residual loudspeaker directivity or small deviations from the intended room conditions.

Finally, the level calibration and mixing rely on a limited set of reference recordings for specific vehicles, background noise, and facade conditions. This ensures a consistent and plausible spectral balance across the experiments in this thesis, but it does not encompass the full range of real-world traffic conditions and building configurations.

These limitations are acceptable for the present work, which focuses on controlled comparisons between AVAS designs under well-defined conditions. Their implications for generalizing the results to more complex and dynamic real-world scenarios are discussed further in Chapter 6.

4.2 Perceptual Validation Strategies

A fundamental requirement for using an auralization framework to evaluate human response is an appropriate validation strategy. The previous section introduced numerical methods for comparing individual stages and the overall auralization output with reference measurements. Such numerical comparisons are useful for diagnosing errors and tuning the system, but they do not necessarily indicate whether remaining discrepancies are perceptually relevant. Even when reference and auralized signals closely match in third-octave spectra or in psychoacoustic metrics such as loudness, tonality, or fluctuation strength, they may still be perceived as different due to differences in temporal structure or spatial cues. Conversely, numerical deviations that appear large in certain metrics may have little perceptual consequence.

Because no single numerical measure captures the perceptual similarity of complex, time-varying signals such as vehicle pass-bys, perceptual validation is essential. This can be done informally, for example, by listening to intermediate results during model development, or formally through dedicated listening experiments. Perceptual validation can target the entire auralization chain, either directly by asking for plausibility or authenticity, or indirectly by evaluating perceptual attributes that are sensitive to specific modeling choices (see Section 2.8.2). In addition, individual modules, such as reproduction methods, can be validated independently of the rest of the chain.

This thesis employs all three validation approaches. First, a system-wide plausibility assessment of the full auralization chain was conducted using headphone-based reproduction. Second, indirect validation was achieved by evaluating perceptual attributes such as perceived speed and annoyance. Third, independent validations of the reproduction methods were conducted to ensure that the selected approaches did not introduce unintended artifacts. All validation results presented here apply to the specific implementation described in this chapter and should not be interpreted as general properties of the underlying framework. The following subsections summarize these validation strategies.

4.2.1 System-wide Plausibility Validation

A system-wide plausibility validation is perhaps the most direct way to assess whether an auralization chain produces perceptually convincing results. In such validations, the final auralized stimuli are judged against reference recordings (*authenticity*) or against the listener's internal reference for how the scenario should sound (*plausibility*). For scenarios listeners are not routinely exposed to in everyday life, providing reference recordings during the training phase can be essential, as participants may otherwise lack a stable internal reference, and plausibility ratings may become unreliable.

This general challenge is particularly relevant for electric vehicle sounds, which are still unfamiliar to many listeners, and it becomes even more pronounced in indoor noise scenarios, where obtaining controlled in situ recordings that match the simulated laboratory environment is difficult. For this reason, the system-wide validation in this thesis focused on outdoor pass-bys, for which binaural reference recordings were obtained as described in Section 4.1.1.

In addition to reference sounds, it is often beneficial to include low-quality anchor stimuli in a plausibility test. Anchors help stabilize the response scale and reduce inter-subject variability by providing a clear perceptual lower bound. This is particularly important when several stimuli are already reasonably realistic, since listeners might otherwise compress their ratings toward the upper end or apply inconsistent interpretations of the scale. Including explicit poor-quality examples ensures that plausibility judgments for the high-quality auralizations remain interpretable and comparable across participants. The following briefly summarizes the plausibility validation setup and results relevant to this thesis. A detailed description of the experiment and a statistical analysis of the results are provided in Paper A.

Experiment Setup

The system-wide plausibility experiment consisted of two parts. In the first part, participants listened to 10 binaural in situ recordings, two per measured vehicle and driving direction. These recordings served two purposes. First, they familiarized participants with the sound of real electric vehicle passages and helped establish an internal reference for plausibility judgments. Second, they provided ground truth ratings for other perceptual attributes, which are discussed in Section 4.2.2.

In the second part, participants listened again to a subset of the real recordings together with twenty auralized stimuli. These stimuli were generated from the measured vehicle velocity profiles using the complete auralization chain and were reproduced over headphones. Ten auralizations were generated using spherical harmonic order $L = 64$, five were generated with $L = 16$, and five were intentionally degraded low-quality anchor stimuli. The anchors consisted of amplitude-panned white noise, white noise combined with a binaural ambiance, auralizations without ambiance noise, and auralizations without source movement. For all stimuli in this second part, participants rated plausibility on an 11-point numerical scale ranging from 0 (*not at all plausible*) to 10 (*extremely plausible*).

All stimuli were presented via calibrated Sennheiser HD 650 headphones using HRTFs measured for a HEAD acoustics HMS II.3 artificial head, which matches the geometry of the artificial head used for the reference recordings. Twenty participants with self-reported normal hearing completed the experiment.

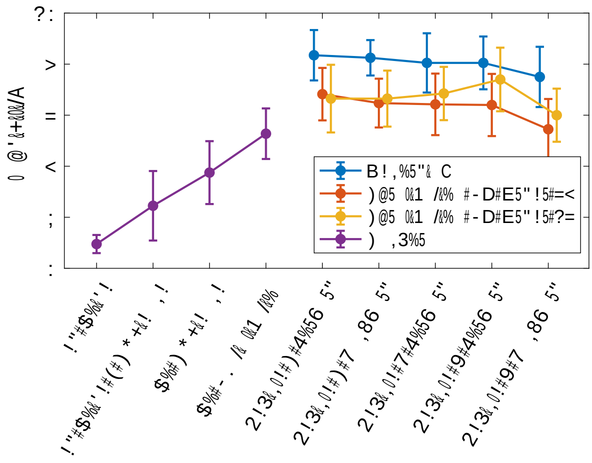


Figure 4.22: System-wide plausibility ratings of auralization, recording, and anchor, plotted as arithmetic mean and 95% confidence interval ($N = 20$).

Results

The obtained plausibility ratings are summarized in Figure 4.22. Across all vehicles, the real in situ recordings received the highest plausibility scores, confirming that participants established a clear internal reference during the first part of the experiment. As expected, the anchor stimuli were rated as the least plausible, with amplitude-panned noise receiving the lowest scores. The two evaluated spherical harmonic orders did not show a consistent difference in plausibility.

Overall, the results suggest that, although the auralizations are not indistinguishable from the reference recordings, they achieve plausibility ratings that are sufficiently high for the intended listening studies. At the same time, it is important to acknowledge that the notion of an "acceptable" level of plausibility is not well defined and depends on the specific perceptual task. While Paper A reports statistically significant differences between recordings and auralizations when averaging across vehicle types[9], the overlapping confidence intervals for almost all individual comparisons in Figure 4.22 indicate that the synthesized stimuli are at least close to the reference in perceived plausibility.

Regardless of how one interprets the absolute level of plausibility, this validation step provides an important contextual basis for the subsequent listening experiments. Rather than assuming plausibility, as is often done in applied auralization studies, the present framework includes an explicit perceptual verification. This enables

[9] The grand mean plausibility for all recordings is 8.0, compared to 6.3 for auralizations with $L = 64$ and 6.7 for auralizations with $L = 16$.

interpretation of subsequent perceptual results with greater confidence, knowing that the stimuli are plausible but not perfect and that any residual discrepancies are at least quantified rather than unknown.

4.2.2 Indirect Validation via Perceptual Attributes

Instead of directly validating plausibility or authenticity, an auralization can be assessed indirectly through other perceptual attributes. The choice of attributes should be motivated by the intended response measure of the final listening experiment. For example, if an auralization is developed to study loudness and annoyance differences between heat pumps, one could argue that the most relevant validation step is to demonstrate that the auralized sounds yield the same loudness and annoyance ratings as corresponding real-life references.

However, indirect validation alone is not necessarily sufficient. An auralization that reproduces loudness judgments accurately but is perceived as implausible overall may still limit the ecological validity and real-world transferability of the experimental results. Indirect validation via perceptual attributes should therefore be viewed as a complement to, rather than a replacement for, system-wide plausibility validation. When used together, the two approaches offer a more nuanced understanding of where and how an auralization deviates from reality.

For instance, in a concert hall auralization, one could first evaluate overall plausibility, then assess specific perceptual dimensions such as timbre, perceived room size, or spatial envelopment. This would refine an overall statement like "the auralization is moderately plausible" into a more informative description such as "the auralization is moderately plausible, consistent in timbre and room size, but underestimates spatial envelopment". In this thesis, such indirect validation was achieved by analyzing two perceptual attributes relevant to electric vehicle pass-bys: perceived vehicle speed and perceived annoyance. The experimental setup and results are described below.

Experiment Setup

The indirect validation via perceptual attributes was integrated into the same listening experiment used for plausibility validation, in which participants also rated perceived annoyance and perceived pass-by speed for all stimuli. In the first part of the experiment, participants evaluated 10 binaural in situ recordings, two per measured vehicle and driving direction. These ratings served as ground-truth reference data and simultaneously as training data, allowing participants to form a stable internal reference for subsequent comparisons. In the second part, five of these recordings were repeated, and twenty auralized pass-bys synthesized from the corresponding velocity profiles were presented. Participants again rated annoyance and perceived vehicle speed for each stimulus. This design enables two comparisons. First, it

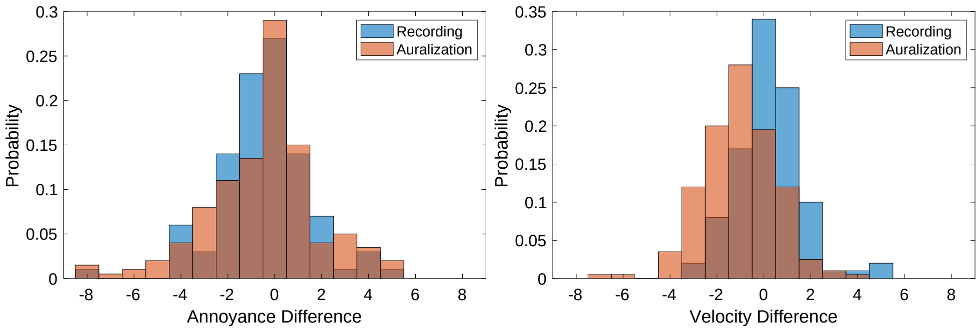


Figure 4.23: Distribution of differences in annoyance and vehicle velocity ratings between both experiment parts. The data combines the results for all evaluated vehicles.

allows assessing how consistently participants rate the repeated in situ recordings. Second, it enables evaluating how closely the auralizations reproduce the perceptual attributes of the real recordings.

To analyze these relations, the ratings obtained for the stimuli in the second part were compared with the ratings for the corresponding recordings in the first part. As discussed in Paper A, linear regression was used to illustrate overall trends, while the primary analysis relied on comparing the distributions of rating differences between both experiment parts, as described in the following subsection.

Results

The histograms in Figure 4.23 show the differences between the ratings obtained in the first and second experiment parts. The blue distributions represent repeated in situ recordings, while the orange distributions represent auralizations in the second part compared to their corresponding recordings (that is, stimuli based on the same velocity profile) in the first part. These distributions were compared using Wilcoxon signed-rank tests.

The results show that, even for repeated in situ recordings, participants were not perfectly consistent across the two experimental rounds. For the recordings, however, the difference distributions for both annoyance and perceived speed were approximately normal and centered on zero, indicating no systematic shift in ratings. A similar pattern was observed for the auralized annoyance ratings, whose distribution closely matched that of the recordings and did not differ significantly in the Wilcoxon test. This suggests that the auralizations reproduced perceived annoyance reasonably well.

For perceived vehicle speed, the pattern differed. The auralizations showed a clear bias, with the distribution centered below zero, indicating that auralized passages were generally rated as faster than their corresponding real recordings[10]. This shift was statistically significant according to the Wilcoxon signed-rank test. The direction of this bias is consistent with the numerical analyses in Paper A, which showed that the auralized pass-bys exhibit slightly sharper time-frequency transitions, likely due to limitations in the radiation directivity modeling and the absence of environmental reflections in the current implementation.

This outcome implies that the current implementation of the auralization framework is unsuitable for studies requiring accurate absolute estimates of perceived vehicle speed. However, it does not invalidate the method for relative comparisons. The auralizations were sufficiently consistent to detect differences between stimuli, for example, when evaluating whether different AVAS designs yield different perceived speeds. This observation directly motivated the follow-up work in Paper C, which focused specifically on pass-by speed perception.

4.2.3 Component-Level Perceptual Validation

In addition to system-level validation of the complete auralization chain, individual components of the framework can be validated separately. Such component-level validation may be motivated either by research questions that focus on a specific aspect of the auralization, or by methodological changes introduced after a system-wide plausibility assessment has been completed.

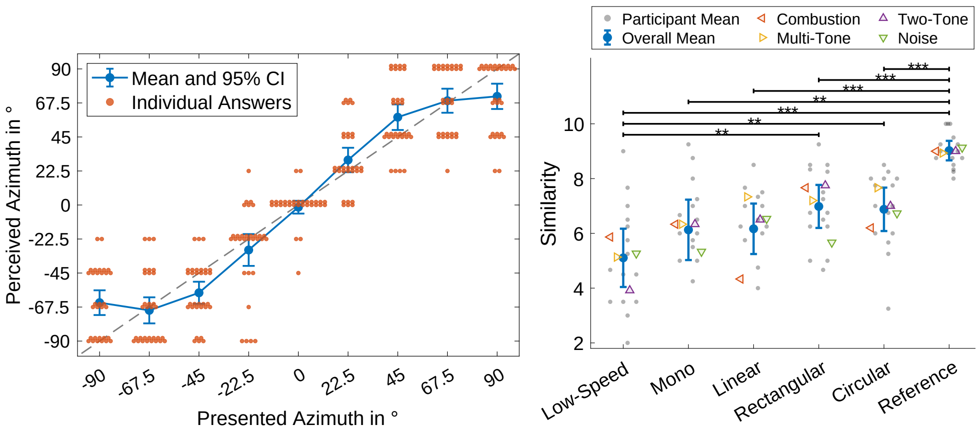
The latter motivation applies to the component-level validations performed in this thesis. The overall auralization chain was perceptually validated using a headphone-based reproduction system. In subsequent studies, additional reproduction methods were employed, most notably binaural crosstalk cancellation and wave field synthesis in the *Living Room Lab*. Rather than repeating a full system-level validation for each reproduction setup, these methods were evaluated individually at the component level. The following sections summarize the corresponding perceptual validation experiments and discuss their implications for the respective listening studies.

Binaural Crosstalk Cancellation

A component-level validation was conducted for the loudspeaker-based binaural reproduction with crosstalk cancellation employed in Paper C. For this reproduction method and the corresponding listening experiment's target outcome measure, perceived pass-by speed, the primary concern was that incomplete crosstalk suppression (cf. Section 4.1.5) could degrade horizontal localization cues and thereby unintentionally influence the task.

To assess this risk, a simple perceptual validation was conducted prior to the main experiment. Thirty-one participants were presented with nine binaural speech stimuli

[10] The differences were calculated by subtracting the ratings obtained in the second part from the ratings in the first part, hence a negative bias means the ratings in the second part were higher.



(a) **(b)**

Figure 4.24: Exemplary component-level perceptual validation results for binaural crosstalk cancellation in anechoic chamber (a), and for WFS reproduction in *Living Room Lab* (b).

distributed horizontally over azimuth angles ranging from $-90°$ to $90°$. Participants were instructed to report the perceived azimuth on a nine-point scale spanning from left ($90°$) to right ($-90°$). This procedure provided a coarse estimate of horizontal localization accuracy, as illustrated in Figure 4.24a.

The results showed that perceived azimuths were, on average, reasonably accurate for small angles but increasingly underestimated at larger azimuths. This behavior is consistent with known limitations of binaural crosstalk cancellation, where loudspeaker-based reproduction typically achieves lower interaural channel separation than headphone playback. For experiments that aim to quantify localization accuracy, such deviations would be unacceptable and would argue against the use of this reproduction method.

However, for the specific purpose of evaluating pass-by speed perception, the observed limitations were considered acceptable. The simulated vehicle trajectories in Paper C did not involve large azimuth angles, and the task did not require precise spatial judgments. The validation, therefore, supported the use of binaural crosstalk cancellation as a suitable compromise between spatial fidelity and experimental simplicity for this specific application.

Wave Field Synthesis Reproduction in the *Living Room Lab*

A more elaborate component-level validation was conducted for the wave field synthesis reproduction used in the *Living Room Lab*. The goal was not only to assess the perceptual plausibility of the existing linear loudspeaker array, but also to evaluate whether alternative array geometries could perceptually reduce artifacts associated with finite array length (cf. Section 4.1.5).

Defining an appropriate reference for this validation is nontrivial. While one could, in principle, record traffic pass-bys in real living rooms, finding an in-situ environment with a comparable facade, window construction, and controllable vehicle trajectories is practically infeasible. Instead, an indirect validation strategy was adopted. The previously conducted system-level listening experiment showed that the EV auralization toolbox produces plausible binaural renderings when reproduced via headphones. This allowed treating the auralized sound field at the window plane as a valid reference and focusing the validation on the reproduction method inside the *Living Room Lab*.

A numerical model of the complete *Living Room Lab* setup (see Section 4.1.5 and Paper B) was therefore used to simulate how different loudspeaker array geometries would reproduce the same exterior sound field. In addition to the existing linear array, hypothetical circular and rectangular array configurations were modeled and compared against an idealized reference consisting of an infinitely long, continuous loudspeaker array in free field, without reflections in the sending room. The numerical model accounted for room coupling, window transmission, and room reflections, and the resulting sound fields at a fixed listener position were rendered to binaural signals.

This approach effectively constitutes an auralization of different reproduction methods. Instead of actually setting up multiple arrays, their expected perceptual outcomes were evaluated via numerical simulation, followed by binaural rendering. The resulting stimuli represented pass-by scenarios reproduced with different array geometries as well as the ideal continuous reference.

As described in Paper B, a listening experiment was conducted with 15 participants, who rated the similarity between pairs of stimuli on an 11-point Likert scale ranging from “not at all similar” to “extremely similar”. The test set included the different array geometries, a hidden reference corresponding to the ideal continuous array, and several low-quality anchor stimuli. The results are shown in Figure 4.24b. While none of the finite WFS arrays reached the similarity ratings of the hidden reference, all were rated substantially higher than the low-quality anchors. Importantly, no statistically significant differences were observed between the tested array geometries.

Based on these results, the existing linear array was retained for subsequent experiments. In retrospect, participants reported considerable difficulty in identifying differences between the stimuli, suggesting that the chosen similarity-rating paradigm may have lacked sensitivity for this application. A forced-choice paradigm, such as *ABX* or *two-alternative forced choice (2AFC)*, might have been more suitable for detecting subtle perceptual differences between array configurations.

4.3 AVAS Perception and Safety

To demonstrate the applicability of the auralization methods in the present implementation, this thesis presents a series of controlled listening experiments that address different aspects of human responses to electric vehicle noise. These experiments were designed to evaluate behavioral, subjective, and physiological responses under well-defined and reproducible acoustic conditions. The following subsections focus on the studies related to AVAS perception and safety, outline the experimental frameworks, and briefly summarise the key observations related to the perception of pass-by speed (Section 4.3.1) and auditory localization (Section 4.3.2). Detailed results and interpretations are presented in the corresponding papers and discussed further in Chapter 6.

4.3.1 Pass-by Speed Perception

Background The investigation of pass-by speed perception was motivated by the system-wide perceptual validation described in Section 4.2.1, which indicated that auralized electric vehicle passages tended to be perceived as slightly faster than corresponding reference recordings. Inspection of the auralized signals suggested differences in time structure, that is, in how spectral content and level evolve during the pass-by. One plausible contributor to these differences is the radiation directivity used in the auralization, which shapes how sound energy is distributed before and after the vehicle passes the listener.

Since the influence of source radiation directivity on the perceived speed of moving sound sources has not been systematically investigated in previous work, a dedicated listening experiment was conducted and published in Paper C. The scenario involved a pedestrian standing at the roadside, judging the speed of a passing electric vehicle. The modular auralization toolchain described earlier was used to generate stimuli in which only selected components were varied. Specifically, different AVAS signal types, pass-by speeds, and radiation directivities, including omnidirectional and several directional patterns derived from numerical modeling and analytic shapes, were combined while all other elements of the rendering chain were kept identical. This allowed the perceptual influence of radiation directivity and AVAS type to be examined in isolation.

Procedure Because absolute speed judgments are known to be difficult and highly variable, a paired comparison paradigm was employed [178]. Participants were presented with pairs of pass-by stimuli and asked to indicate which one sounded faster, using a two-alternative forced-choice design that required a decision even when no clear difference was perceived. To limit the total number of comparisons, the stimuli were divided into several paired comparison groups, each targeting a specific research question, such as the effect of directivity within one AVAS type or differences between

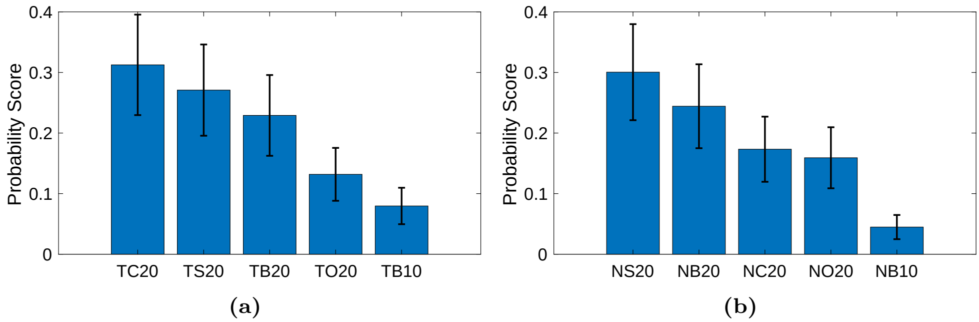


Figure 4.25: BTL probability scores and 95% confidence intervals for tonal AVAS (a) and narrowband noise AVAS (b) with different directivities (C = Cardioid, S = Star, B = BEM simulation, O = Omnidirectional), simulated at 10 km/h and 20 km/h.

tonal and narrowband noise signals. Each group included one lower-speed pass-by as an anchor stimulus. The experiment was conducted with 31 participants using a loudspeaker-based binaural reproduction in an anechoic chamber, as described in Section 4.1.5.

The paired comparison data were analyzed using a *Bradley–Terry–Luce (BTL)* model [179, 180], which estimates relative preference scores from binary-choice data. In this context, the scores represent the probability that one stimulus is judged as corresponding to a higher pass-by speed than another when presented as a pair, reflecting systematic tendencies in perceived speed ordering rather than absolute speed judgments.

Results Figure 4.25 shows exemplary BTL probability scores with 95% confidence intervals for tonal and narrowband noise AVAS signals under different radiation directivities. The results indicate that radiation directivity can influence perceived pass-by speed under certain conditions. In particular, some directional radiation patterns were more likely to be judged as faster than an omnidirectional reference when all other stimulus parameters were held constant.

However, this effect was not observed consistently across all tested conditions. Statistically significant differences occurred only for specific combinations of AVAS signal type and directivity, and the associated confidence intervals indicate relatively small effect sizes. The findings should therefore be interpreted as evidence for a modest and context-dependent influence of radiation directivity on speed perception rather than a general effect applicable to all AVAS designs.

Regarding AVAS signal type, no consistent overall difference in perceived speed between tonal and narrowband noise signals was found across participants. Instead, substantial inter-individual variability was observed, with some listeners systematically associating one signal type with higher perceived speed than the other. This suggests that pass-by speed perception in complex acoustic scenes depends on a combination of acoustic features, listener expectations, and prior experience rather than on a single signal characteristic.

Overall, the experiment demonstrates that the proposed implementation of the auralization framework is sufficiently sensitive to reveal subtle perceptual effects related to radiation directivity, while also highlighting the limitations of generalizing such effects beyond the specific experimental conditions. Detailed analyses are presented in Paper C, and broader implications are discussed in Chapter 6.

4.3.2 Auditory Localization

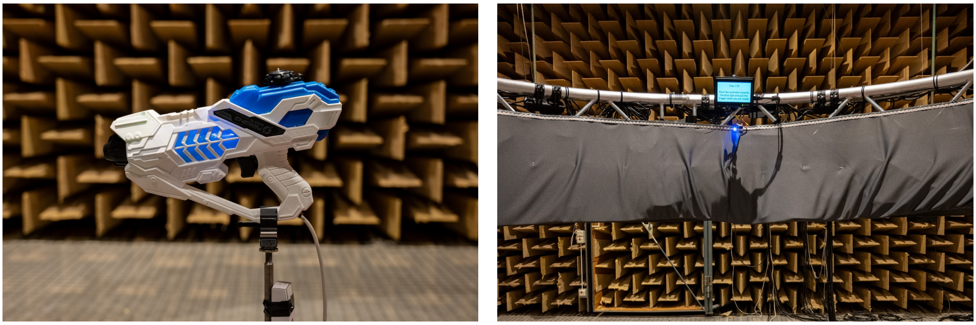
Background One of the central research questions of this thesis concerns the localizability of different AVAS designs, particularly in scenarios involving multiple vehicles simultaneously present. To investigate this question, a dedicated laboratory experiment was developed that combines a high-resolution reproduction method with a custom-built response interface. This experiment exemplifies the close coupling between auralization techniques and evaluation methods, as the reproduction setup and response paradigm were explicitly designed to address small differences in localization performance.

Procedure The experiment was conducted in an anechoic chamber using a concealed circular loudspeaker array consisting of 24 loudspeakers (Figure 4.18). Participants were positioned at the center of the array and exposed to simulated static vehicles modeled as point sources, with each vehicle assigned to a single loudspeaker. Depending on the trial, one, two, or three vehicles were presented simultaneously. Each vehicle reproduced either one of three AVAS signals or a combustion engine reference sound.

To measure localization performance with high angular resolution, a custom pointing device was developed by modifying a toy waterblaster to house two microcontrollers equipped with motion sensors and a trigger switch (Figure 4.26a). The device's horizontal pointing angle was tracked in real time and mapped to a circular LED strip mounted on the inside of the loudspeaker array. The currently indicated azimuth was visualized as a blue LED (Figure 4.26b), providing participants with precise visual feedback of their pointing direction[11]. A small display located at the 0° position provided task instructions and timing information. Similar experiment setups, although often with lower spatial resolution, have been used in the context of spatial hearing research [181, 182].

Each trial began with a visual countdown of 3 s, followed by a stimulus presentation of 10 s. During this interval, participants were instructed to indicate the perceived

[11]The strip consisted of 695 LEDs, resulting in $360°/695$ LEDs $\approx 0.5°$ angular resolution for the visual feedback.



(a) (b)

Figure 4.26: Pointing device with inertial sensors used for localization responses (a) and participant point of view with blue LED light indicating current pointing direction (b).

direction of each audible vehicle by aiming the device and pulling the trigger as quickly as possible. In trials with multiple vehicles, participants were asked to localize all perceived sources in arbitrary order. The number of vehicles per trial was not disclosed in advance. After 10 s, the trial ended, and participants were instructed to return to the 0° position and pull the trigger to initiate the next trial.

In total, three AVAS designs and one combustion engine sound were evaluated in single-, dual-, and triple-vehicle configurations. This resulted in 72 trials per participant, all presented in randomized order. The experiment was completed by 55 participants, of whom three were excluded from the following analysis as their overall localization error, averaged over all trials, was more than 1.5 interquartile ranges above the upper quartile. Further details on the experimental procedure and stimulus design are provided in Paper D.

Results Localization performance was evaluated using two primary outcome measures, absolute localization error and response time. Absolute error was defined as the angular deviation between the reported direction and the true loudspeaker position. Response time was measured from stimulus onset to trigger activation. Repeated-measures analyses of variance were used to assess the effects of sound type and number of vehicles on both measures, with Bonferroni-corrected post hoc tests applied for pairwise comparisons. Figure 4.27 illustrates exemplary localization error results for single- and dual-vehicle conditions.

In addition, the percentage of failed localizations was analyzed. A localization was considered failed if a participant did not respond within the 10 s time limit or if the reported direction deviated by more than 90° from the true source position. Because these data violated normality assumptions, failed localization rates were evaluated using nonparametric Friedman tests.

The results show a clear dependence of localization performance on sound type

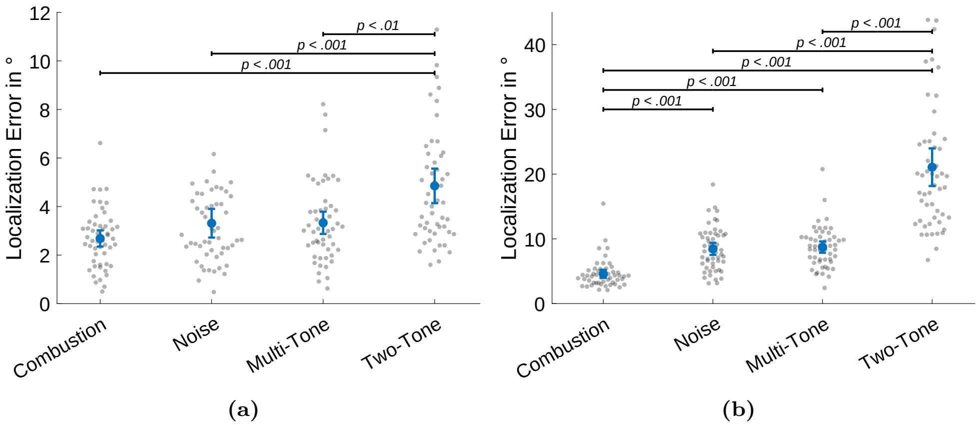


Figure 4.27: Exemplary localization error results for single-vehicle (a) and dual-vehicle (b) conditions as a function of sound type.

and scene complexity. Combustion engine noise consistently yielded the smallest localization errors and fastest response times across all conditions. Among the AVAS designs, the two-tone AVAS exhibited the poorest localization performance, while no significant differences were observed between the multi-tone AVAS and the noise AVAS in single-vehicle conditions.

Crucially, the differences between sound types increased substantially in multi-vehicle scenarios when multiple vehicles emitted the same sound. In these conditions, all three AVAS signals were localized significantly less accurately than the combustion engine reference. The two-tone AVAS performed worst, with mean localization errors exceeding those of the combustion engine sound by up to 20° and response times up to 2 s longer. The proportion of failed localizations also increased markedly in this condition, with most participants localizing fewer than 60 % of the presented vehicles when three two-tone AVAS signals were played simultaneously.

The fact that mean localization errors smaller than 3° were achieved for some stimuli demonstrates that both the reproduction method and the response interface provided angular resolution beyond typical human localization limits. Lower-resolution approaches, such as headphone-based reproduction with generic HRTFs or pointing methods without continuous visual feedback, would likely have introduced additional response variability and masked the observed differences between AVAS designs.

4.4 Noise Effects in Living Environments

The second major research question addressed in this thesis concerns the effects of road traffic noise on humans in living environments. To study these effects under controlled yet realistic conditions, the *Living Room Lab* with its wave field synthesis-based reproduction and physical facade transmission (see Section 4.1.5) was employed. Two studies conducted in this environment, Paper E and Paper F, investigated human responses to low-level road traffic noise. Paper E focuses on electric vehicles and AVAS, while Paper F investigates low-frequency traffic noise with different temporal structures. Both studies combined subjective, cognitive, and physiological outcome measures to capture complementary dimensions of human response. The following subsections summarize the rationale for the selected measures and outline the main experimental approaches, with detailed analyses reported in the corresponding papers.

4.4.1 Attention

Attention is a central cognitive function and belongs to the class of cognitive performance measures described in Section 3.4.1. Together with working memory, it is among the most frequently studied cognitive domains in noise research because it is sensitive to distraction and task interference caused by environmental sound. Previous studies have shown that noise can impair attentional performance, particularly under conditions of high sound levels, irrelevant speech, or highly variable and unpredictable noise [20, 183]. However, much of this literature relies on relatively elevated exposure levels and simplified noise stimuli, such as broadband noise or stationary sounds, rather than acoustically realistic traffic scenarios.

This gap motivated focusing on attention as the primary measure of cognitive performance in the present work. In residential settings, traffic noise is typically encountered at relatively low indoor levels, yet it remains a frequent source of reported disturbance. In addition, AVAS signals introduce sound components that are intentionally designed to attract attention in outdoor contexts, raising the question of whether such signals may have negative cognitive effects when transmitted indoors, even at low absolute levels.

Attention is not a unitary construct but can be divided into several partially distinct subprocesses. Sustained attention refers to the ability to maintain focus over extended periods of time. Selective attention involves filtering relevant information from competing stimuli, while executive control supports conflict resolution and inhibition of irrelevant responses [184]. Different noise characteristics may affect these subprocesses in different ways, as suggested by findings that continuous white noise can selectively alter performance over time in sustained attention tasks, whereas irrelevant speech or environmental sounds modulate selective and executive aspects of attention control [185–187]. In the following, attention is used as an umbrella term encompassing sustained attention, selective attention, and executive control processes relevant for task performance under noise exposure. To address these different components, two complementary attentional paradigms were employed across the studies, as described below.

Conners' Continuous Performance Test

Background Sustained attention was examined in Paper F using a *Continuous Performance Test (CPT)*. CPTs are commonly used to assess sustained attention and inhibitory control by requiring frequent responses to target stimuli while suppressing responses to rare non-targets. The implemented paradigm resembles the Conners' CPT, which is widely used in clinical and experimental settings [188, 189]. To the best of the author's knowledge, this type of CPT has not previously been applied to investigate the effects of complex traffic noise on cognitive performance.

Procedure During the task, participants were presented with a sequence of single letters on a computer screen and instructed to press a response button for all letters except the letter "X", which appeared in 10% of trials. Each CPT round consisted of 360 letters organized into six cycles, with variable interstimulus intervals of 1 s, 2 s, or 4 s and a fixed stimulus duration of 250 ms, resulting in a total task duration of approximately 14 minutes. The high target-to-distractor ratio required frequent responding and occasional response inhibition, making the task sensitive to lapses of attention and impulsive responding. Performance was quantified primarily by the number of commission errors, that is, responses to the non-target letter "X". Response times were analyzed secondarily.

Results Figure 4.28 shows the mean number of commission errors and mean response time for the three sound conditions evaluated in Paper F, namely close traffic noise, far traffic noise, and silence. Both traffic noise conditions were normalized to the same indoor equivalent sound pressure level of $L_{\mathrm{A,eq}} = 40$ dB, while the silence condition corresponded to a background level of approximately $L_{\mathrm{A,eq}} = 10$ dB. Forty-two participants completed the experiment in a fully randomized within-subject design, performing one CPT for each sound condition, with breaks in between.

A repeated-measures analysis of variance revealed a significant effect of sound condition on the number of commission errors ($F(2, 82) = 3.68$, $p = .029$, partial $\eta^2 = .082$). While post hoc comparisons did not show a significant difference between the close and far traffic conditions, the far traffic condition yielded a significantly higher number of commission errors than silence after Bonferroni correction. A planned contrast analysis further showed that, when both traffic noise conditions were combined, they differed significantly from silence. These results indicate that, even at moderate indoor levels, road traffic noise can exert a small but measurable effect on sustained attention compared with an acoustically very quiet baseline. At the same time, they demonstrate that the *Living Room Lab* environment and the applied auralization approach are sufficiently sensitive to detect subtle cognitive effects under controlled indoor exposure conditions.

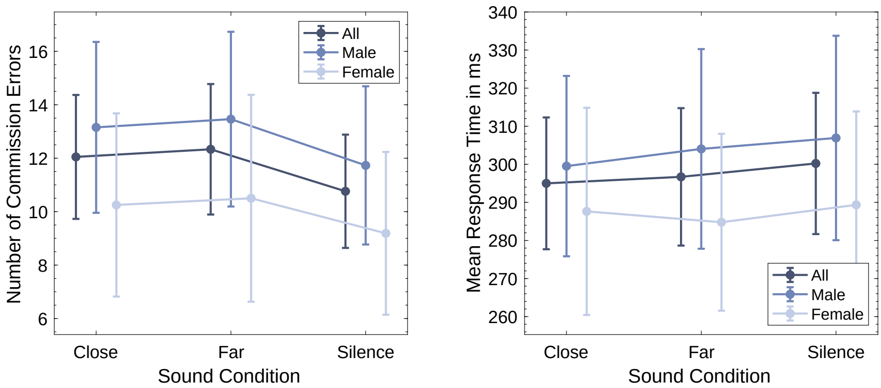


Figure 4.28: Number of commission errors (a) and mean response time (b) for the CPT attention test performed in Paper F.

Combined Flanker and Spatial Stroop Test

Background The CPT results in Paper F showed that sustained attention can be affected by traffic noise, but only weakly and at moderate indoor levels. For the electric vehicle experiment in Paper E, two additional considerations motivated the use of a different attentional paradigm. First, the expected indoor exposure levels were substantially lower ($L_{\mathrm{A,eq}} \leq 21.5$ dB), suggesting that any performance effects would likely be even smaller. Second, the CPT requires long and monotonous task blocks, which would have resulted in unreasonably long experiment sessions when combined with four sound conditions. The aim was therefore to employ a more time-efficient task that probes additional attentional control mechanisms and may be more sensitive to subtle acoustic influences.

Procedure Paper E employed a combined *Eriksen Flanker* and *spatial Stroop* task. Whereas the CPT primarily targets sustained attention and response inhibition, the Flanker and Stroop paradigms probe executive attention under conditions of stimulus conflict. The Eriksen Flanker task assesses selective attention and interference control by requiring participants to respond to a target stimulus while suppressing conflicting information from surrounding distractors. In the version used here, the visual target was a central arrow pointing left or right, flanked by four additional symbols: arrows pointing in the same direction as the target (*congruent*), arrows pointing in the opposite direction (*incongruent*), or non-directional dashes (*neutral*). The spatial Stroop component introduced a conflict between stimulus identity and spatial location by presenting the entire stimulus array either centrally, shifted left, or shifted right on the screen. By combining these manipulations, the task targets multiple components of executive control, including conflict monitoring, selective attention, and spatial processing. Figure 4.29a shows the resulting stimulus set.

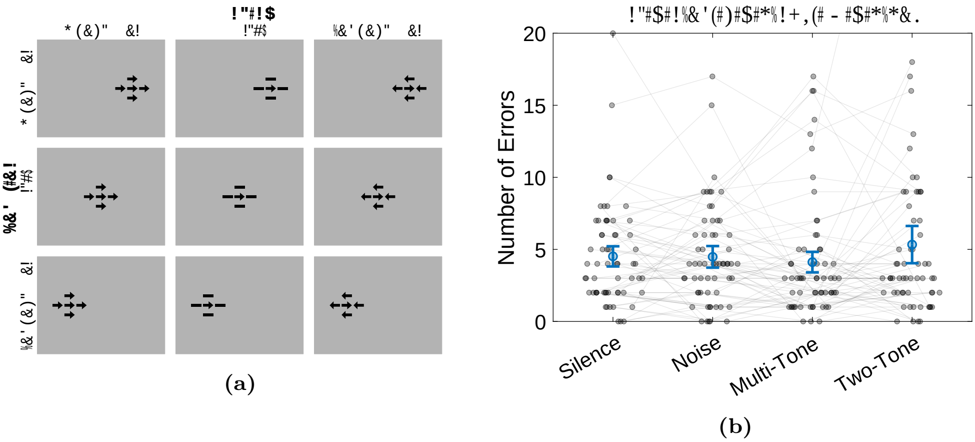


Figure 4.29: Combined flanker and spatial Stroop stimulus configuration (a) and number of errors, including arithmetic mean, within-subject confidence intervals, and Friedman test statistics (b).

During the experiment, participants were instructed to indicate the direction of the central arrow as fast as possible using left or right response keys while ignoring both the flankers and the spatial offset. Each sound condition block consisted of 180 trials and lasted approximately five minutes, making the task substantially shorter and less fatiguing than the CPT while still providing a sufficient number of trials for statistical analysis.

Results Figure 4.29b shows individual participant data, arithmetic means, and within-subject confidence intervals for the total number of errors in the attention task, performed by 60 participants. Both the descriptive statistics and the Friedman test, which is also reported in Figure 4.29b, indicate that there was no significant effect of sound condition on error rates. Although the mean number of errors and the corresponding confidence intervals show a tendency toward slightly higher values for the two-tone AVAS condition, this trend did not reach statistical significance. Mean response time and response time variability likewise showed no significant effects of sound condition. Since no significant main effects were observed, no post hoc tests were conducted.

In summary, these results suggest that low-level electric vehicle road traffic noise does not significantly affect attention as assessed using a combined Eriksen Flanker and spatial Stroop task. At least for short-duration indoor exposures at realistic closed-window levels, executive control mechanisms appear relatively robust to traffic noise. The absence of measurable performance effects also helps contextualize subsequent subjective and physiological findings, indicating that changes in annoyance or perceived workload are not necessarily accompanied by detectable decrements in executive attention.

4.4.2 Perceived Workload and Annoyance

Background In addition to objective performance measures, both Paper E and Paper F assessed subjective responses using a modified *NASA Task Load Index (NASA-TLX)* questionnaire (see Section 3.2). The motivation for including subjective workload measures was twofold. First, perceived workload provides complementary information to cognitive performance metrics, particularly when performance differences are small or absent. Second, in the context of electric vehicle noise and indoor traffic noise, subjective responses such as mental demand, frustration, and annoyance are directly relevant outcome variables, even when objective task performance remains stable. Paper E additionally included explicit noise annoyance ratings to capture the affective response to different AVAS signals.

Procedure In Paper E, participants completed the workload questionnaire after each attention test round, that is, once per sound condition, resulting in a repeated-measures design. The standard NASA-TLX item *physical demand* was replaced by a noise annoyance question phrased as "How annoyed were you by the noise you heard during the experiment?", with response options ranging from "not at all" to "very". The remaining items assessed *mental demand*, *temporal demand*, *performance*, *effort*, and *frustration*.

In Paper F, the original NASA-TLX questionnaire was used without modification. In that study, where the attention task with subsequent breaks required much more time than in Paper E, the questionnaire was administered only once after the first attention test round, resulting in a between-subject design. This difference in experimental design reflects the distinct aims of the two studies and limits the direct comparability of their absolute workload ratings.

Results The perceived workload and annoyance data in Paper E were analyzed using separate Friedman tests for each questionnaire item, followed by Bonferroni-corrected Wilcoxon signed-rank tests for post hoc comparisons. Figure 4.30 shows individual participant ratings from 60 participants, together with arithmetic means, within-subject confidence intervals, and the corresponding test statistics.

Overall, the results indicate that sound condition significantly influenced both annoyance and perceived workload. Participants reported significantly different annoyance ratings across the evaluated sound conditions, with the two-tone AVAS consistently rated as most annoying. Importantly, this signal was also the only condition that produced a significant increase in perceived mental demand compared with silence, as well as a near-significant increase in perceived frustration. No other sound condition led to significant changes in subjective workload relative to silence.

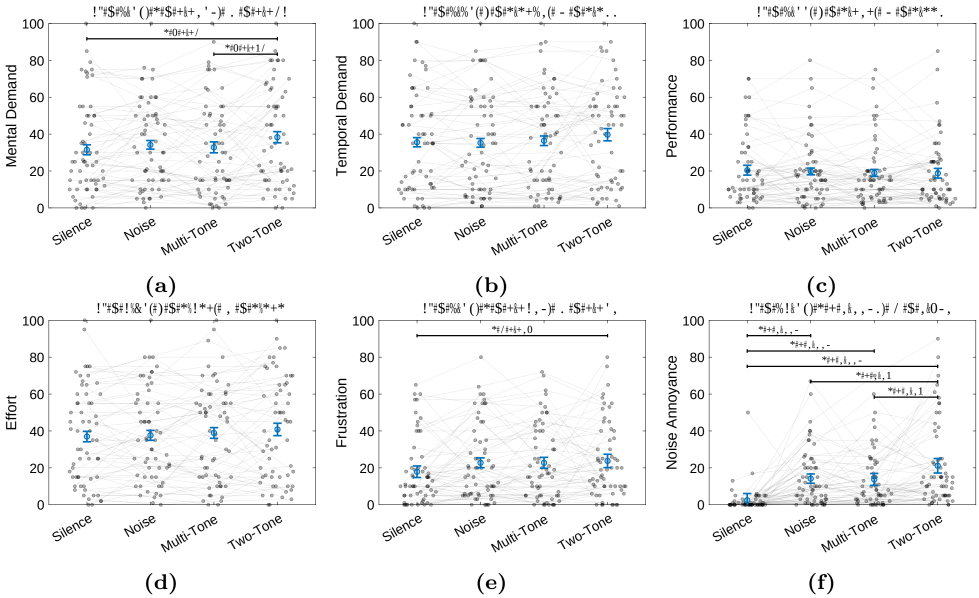


Figure 4.30: Paper E perceived workload and annoyance results showing individual subject data (grey dots) and arithmetic mean with within-subject 95%-confidence intervals (blue errorbars) for mental demand (a), temporal demand (b), performance (c), effort (d), frustration (e), and noise annoyance (f). The subplot titles report Friedman test results, and the horizontal bars represent significant ($p < 0.05$) and near-significant ($p < 0.075$) Bonferroni-corrected Wilcoxon signed-rank post hoc tests.

These findings show that, even in the absence of measurable performance decrements in the attention tasks, different AVAS designs can substantially affect how demanding and unpleasant a task is perceived to be. This dissociation between objective performance and subjective workload underscores the importance of including self-report measures when evaluating human responses to low-level traffic noise.

Paper F, which evaluated low-frequency traffic noise with different temporal structures, likewise reported significant effects of sound condition on perceived mental demand. However, since the present chapter focuses on the application of the framework to electric vehicle noise, the reader is referred to Paper F for a detailed presentation and discussion of those results.

4.4.3 Electrodermal Activity

Background The only physiological measure investigated in this thesis is *Electrodermal Activity (EDA)*, introduced in Section 3.3.2. EDA reflects changes in skin conductance caused by sweat gland activity under sympathetic nervous system control and is commonly used as an indicator of physiological arousal, stress, and cognitive workload [140, 190]. It consists of a slowly varying tonic component, the *Skin Conductance Level (SCL)*, which reflects general arousal, and faster phasic components, *Skin Conductance Responses (SCRs)*, which are associated with transient increases in sympathetic activation. Changes in either component can therefore provide an objective complement to subjective workload and annoyance ratings in response to environmental noise.

Procedure In Paper E, EDA was recorded continuously throughout the experiment, i.e., while participants performed the attention task under different sound conditions, using Ag/AgCl electrodes placed on the palmar side of the middle and ring fingers of the left hand. Signals were acquired at a sampling rate of 6 kHz and analyzed only during the approximately five-minute attention task blocks for each sound condition. To separate tonic and phasic activity, the EDA signals were filtered using second-order Butterworth low-pass and high-pass filters with a cutoff frequency of 0.05 Hz.

For each participant and sound condition, the mean SCL was computed from the tonic component. From the phasic component, SCRs were identified using a peak-detection algorithm, and two metrics were derived: the number of SCRs per minute and the summed SCR amplitude per minute. Because all metrics were averaged across entire task blocks and the experiment employed a within-subjects design, minor movement-related artifacts were unlikely to bias comparisons between sound conditions systematically.

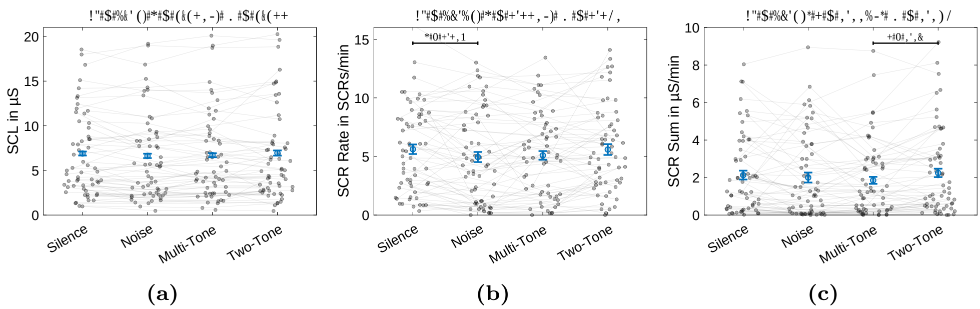


Figure 4.31: Paper E electrodermal activity results showing individual subject data (grey dots) and arithmetic mean with within-subject 95%-confidence intervals (blue errorbars) for skin conductance level (a), SCR rate (b), and sum of SCR amplitudes per minute (c) for the four evaluated sound conditions. The subplot titles report Friedman test results for the main effect of sound condition. The horizontal bars represent significant ($p < 0.05$) and near-significant ($p < 0.075$) Bonferroni corrected Wilcoxon signed-rank post hoc tests.

Results Figure 4.31 shows individual participant data, arithmetic means, and within-subject confidence intervals for SCL, SCR rate, and summed SCR amplitude. For all three measures, Friedman tests indicated statistically significant main effects of sound condition, with small effect sizes as reported in the figure titles.

Post hoc Wilcoxon signed-rank tests did not reveal any significant pairwise differences for SCL or SCR rate after Bonferroni correction, suggesting that the main effects reflect subtle distributed differences rather than strong contrasts between specific sound conditions. For the summed SCR amplitude, however, a significant difference was observed between the multi-tone and two-tone AVAS conditions, with the two-tone signal eliciting higher overall SCR amplitudes.

Overall, these findings indicate that even low-level road traffic noise can affect physiological arousal, as reflected in electrodermal activity. Although the observed effects are small and should be interpreted cautiously, the consistently elevated SCR amplitudes for the two-tone AVAS suggest that this signal may evoke stronger sympathetic activation than other AVAS designs. Together with the subjective workload and annoyance results, the EDA findings support the conclusion that low-level EV noise can influence subjective and physiological responses even when objective task performance remains unaffected.

CHAPTER 5

Summary of Included Papers

This chapter presents concise summaries of the included papers, written in the context of the overall aims of this thesis. The papers collectively build on a shared methodological foundation, while differing in how they implement and apply it to specific research questions. The summaries are not copies of the abstracts but highlight the elements that matter for the later discussion, such as study aims, design choices, key results, and limitations relevant here. The goal is to give the reader enough context to follow the synthesis without consulting the full articles. Chapter 6 then integrates these results and discusses their implications for methods, regulations, and future work.

5.1 Paper A

Auralization of Electric Vehicles for the Perceptual Evaluation of Acoustic Vehicle Alerting Systems
Leon Müller, Wolfgang Kropp
Published in *Acta Acustica*, 2024, 8, 27.

This paper introduces a foundational auralization approach for simulating outdoor passages of electric vehicles with a focus on Acoustic Vehicle Alerting Systems. Based on in situ recordings of three passenger EVs, it integrates AVAS and tire-road synthesis methods with boundary element-derived source directivities and a propagation model using moving Green's functions with spherical harmonics expansion. All data and algorithms were made publicly available in an open-access database.

Validation included numerical comparisons to reference recordings and a two-part listening experiment with 20 participants. The perceptual results showed that auralizations achieved relatively high plausibility but were judged less realistic than in situ recordings, likely due to the absence of tire-road transient noise. Annoyance ratings were consistent with the references, while vehicle speed was systematically overestimated, pointing to a possible role of radiation directivity in speed perception and motivating Paper C. Overall, the proposed auralization approach provides a validated and openly accessible foundation for AVAS-related listening experiments and subsequent studies in this thesis.

5.2 Paper B

Loudspeaker Array-Based Auralization of Electric Vehicle Noise in Living Environments
Leon Müller, Jens Ahrens, Wolfgang Kropp
Published in *Proceedings of Forum Acusticum*, 11th Convention of the European Acoustics Association, Forum Acusticum 2025, Málaga, Spain.

Paper B extends the electric vehicle auralization approach introduced in Paper A to indoor environments by reproducing vehicle pass-bys in the *Chalmers Living Room Lab*. Using a 24-loudspeaker array and wave field synthesis, outdoor vehicle sound fields are projected onto a real window separating two coupled rooms, allowing authentic transmission into a furnished living space. This setup enables perceptual experiments on how electric vehicle noise, including AVAS signals, is experienced indoors. The study investigates various loudspeaker array geometries to minimize time-structure distortions resulting from a finite array aperture. A numerical model of the whole source-window-receiver path was developed, and binaural auralizations were generated for perceptual testing. While simulations suggested that a circular array yields the closest match to a continuous free-field reference, listening tests with 15 participants revealed no significant perceptual differences between array types at realistic indoor levels. Together, these results demonstrate that the approach can be adapted to indoor contexts, providing a validated basis for controlled studies on the perception and potential health impacts of electric vehicle noise in residential environments.

5.3 Paper C

On the Influence of AVAS Directivity on Electric Vehicle Speed Perception
Leon Müller, Wolfgang Kropp

Published in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, INTER-NOISE 2024, Nantes, France.

This paper builds on the auralization approach developed in Paper A, applying it for the first time beyond plausibility validation to investigate perceptual effects of AVAS radiation directivity. Motivated by the findings of Paper A, where perceptual validation suggested that mismatches in modeled radiation directivity might influence speed perception, a paired-comparison listening experiment with 31 participants was conducted to test whether directivity shapes the perception of electric vehicle pass-by speed. Tonal and narrowband AVAS signals were combined with different radiation patterns and reproduced via loudspeaker-based binaural crosstalk cancellation. Results show that directivity can significantly affect speed perception, with omnidirectional patterns generally perceived as slower than more directional ones. While no overall difference between tonal and narrowband AVAS was observed, some listeners showed strong and consistent preferences, indicating that signal type may influence perception in subjective ways. This study provides the first systematic investigation of how AVAS radiation directivity affects vehicle speed perception, demonstrating the value of the auralization approach for generating novel perceptual insights and contributing to a better understanding of how spatial emission patterns of moving sources shape human perception.

5.4 Paper D

Auditory Localization of Multiple Stationary Electric Vehicles
Leon Müller, Jens Forssén, Wolfgang Kropp
Published in *Journal of the Acoustical Society of America*, 157, 2025.

While the experiment in Paper C served as an exploratory, academically oriented investigation with limited direct implications for traffic safety, Paper D presents the first large-scale, safety-relevant application of the electric vehicle auralization approach, investigating the auditory localization of electric vehicles, including scenarios with multiple simultaneously present EVs emitting similar AVAS sounds. Using a concealed 24-loudspeaker array in an anechoic chamber, static single- and multi-vehicle scenarios were simulated with combustion engine noise, a two-tone AVAS, a multi-tone AVAS, and a noise-based AVAS. Results from 52 participants show that the two-tone AVAS is localized significantly worse than broadband alternatives and combustion noise, and that presenting multiple vehicles with similar AVAS sounds further degrades performance, drastically increasing localization errors and the number of failed responses. These findings indicate that sounds performing well in detection tasks are not necessarily well-suited for localization, especially in complex traffic scenarios. This study not only provides the first systematic evidence on the localizability of electric vehicles but also demonstrates that the auralization approach developed in Paper A can produce robust and practically relevant results in listening experiments that go beyond plausibility validation.

5.5 Paper E

Effects of Low-Level Electric Vehicle Noise on Attention, Electrodermal Activity, Workload, and Annoyance
Leon Müller, Jens Forssén, Wolfgang Kropp
Published in *Journal of the Acoustical Society of America*, 159, 2026.

Building on the auralization approach developed in Paper A and its indoor extension introduced in Paper B, Paper E investigates the effects of low-level indoor electric vehicle noise on human cognitive, physiological, and subjective responses. Using a wave field synthesis-based auralization in the *Chalmers Living Room Lab*, 60 participants completed attention tasks while electrodermal activity was continuously recorded, and perceived workload and noise annoyance were assessed. The same three commonly implemented AVAS types used in Paper D were presented at realistic indoor levels corresponding to outdoor vehicle passages on a nearby street ($L_{\mathrm{A,eq}} \leq 21.5$ dB). Results show that the two-tone AVAS, which showed the poorest localization in Paper D, also elicits the highest physiological arousal, perceived mental demand, and annoyance, even though attention performance remains unaffected. These findings demonstrate that low-level, highly tonal AVAS signals can impose measurable perceptual and physiological strain on non-involved listeners in living environments. The study not only provides the first empirical evidence on human responses to indoor EV noise at very low levels but also confirms that the auralization methods developed in this thesis can be successfully applied to health- and comfort-related research, thereby extending their utility beyond perception and traffic safety studies.

5.6 Paper F

Traffic Noise at Moderate Levels Affects Cognitive Performance: Do Distance-Induced Temporal Changes Matter?
Leon Müller, Jens Forssén, Wolfgang Kropp
Published in *International Journal of Environmental Research and Public Health*, 20(5), 2023

Even though Paper F predates the other studies and does not focus on electric vehicle noise, it is included in this thesis because its methodology, auralization of traffic noise in an indoor living environment, provided the foundation for the implementation presented in Paper B and the experiment performed in Paper E. The study investigates the effects of low-frequency road traffic noise with different time structures caused by varying road-to-facade distances, while maintaining the same indoor equivalent sound level ($L_{\mathrm{A,eq}} \approx 40$ dB). Forty-two participants performed a continuous performance test to assess sustained attention and inhibitory control and completed a NASA-TLX questionnaire to measure perceived task load while being exposed to different traffic stimuli in the *Chalmers Living Room Lab*. Results show that low-frequency road traffic noise at moderate levels can affect both cognitive performance and perceived workload, with significant differences between noise and silence, but no statistically significant differences between close and distant traffic conditions. In the context of this thesis, the study provided both methodological groundwork and new insights into human responses to moderate levels of indoor traffic noise.

CHAPTER 6

Discussion

This chapter discusses the findings of the preceding studies, starting with methodological insights gained from applying the proposed framework to controlled investigations of traffic noise. Rather than reiterating individual results, the discussion integrates evidence across experiments to identify consistent patterns, trade-offs, and limitations across perceptual, cognitive, and physiological outcomes. The chapter concludes with broader implications for future research and experimental design.

6.1 Methodological Insights

This thesis demonstrates that, under suitable conditions, listening experiments on electric vehicle noise can be conducted in virtual acoustic environments with results that are informative for design and policy. The following points discuss the main methodological insights of this thesis. Here, *auralization chain* refers to the concrete electric-vehicle-specific implementation used across the papers, while *framework* denotes the more general methodological principles introduced earlier.

6.1.1 Modular and Task-Driven Auralization

Investing in a universal yet task-adaptable auralization chain proved worthwhile. Source models for AVAS were developed once and then reused across experiments, with modules that could be swapped or tuned for the task at hand. This enabled, for example, using a shared tire-noise component in Paper E, generating traffic sequences with randomized velocity profiles where AVAS, tire noise, and trajectories remain consistent, applying different radiation directivities in Paper C, and creating quasi-constant signals with slightly different velocities for Paper D. Off-the-shelf recordings or a one-size-fits-all auralization tool would not have provided this flexibility. The trade-off is higher engineering effort and the need for validation.

For efficient listening experiments in virtual acoustic environments, the evaluation method should inform the auralization, and reproduction constraints should inform evaluation choices. Here, the *outcome measure* is the primary variable the study aims to estimate, for example, localization error or perceived pass-by speed. In Paper D, horizontal localization was the primary outcome measure. Headphones with generic HRTFs risk degraded localization accuracy, so a loudspeaker approach with single hidden sources was used. This choice prioritized stable interaural cues over exact wavefront reproduction and yielded mean angular errors below 3°, a performance rarely achieved with generic HRTFs or low-order spatial arrays. By contrast, Paper C examined speed judgments under free-field assumptions where small directivity changes, once spectrally equalized, do not strongly alter interaural cues. Immersion and simplicity were therefore prioritized through crosstalk-cancellation binaural playback with generic HRTFs, which provided a headphone-free, immersive experience at the cost of less accurate interaural cues.

The general conclusion is to select auralization and reproduction methods that best serve the primary outcome measure, rather than maximizing physical accuracy in dimensions that will not affect that outcome. Unfortunately, predicting which acoustic aspects influence a given human response is not always straightforward. For instance, one might assume that the spatial characteristics of indoor noise, i.e., the fact that one perceives some vehicle movement even after propagation through a facade, does not alter human responses. However, there is currently no direct evidence that confirms this assumption. Hence, a fully comprehensive auralization that encompasses all plausible aspects would be ideal in principle, but is rarely feasible in practice. Researchers should therefore make motivated choices about what to include, state omissions explicitly, and avoid overclaiming transferability. For example, a study that examines noise effects using white noise from a single loudspeaker can yield valid insights for that stimulus and setup, yet its findings do not automatically generalize to traffic or workplace noise. When possible, small sensitivity checks that toggle candidate features, such as added spatial structure, can indicate whether the main conclusions depend on those aspects and inform future study designs.

6.1.2 System-Level Plausibility with Targeted Component Checks

A system-level perceptual validation, here via binaural renderings compared with reference recordings, showed that the synthesized electric-vehicle passages in Paper A were not indistinguishable from reality but reached a comparable level of plausibility. Perceptually validating the auralization chain in a free-field pass-by scenario, then later allowed for assuming sufficient plausibility in static scenarios (Paper D) and only required additional validation of the different reproduction techniques used in Paper C and Paper E. System-level checks are valuable because individual steps can compensate for each other, even when some internal modules are imperfect. This has a practical upside. When the final signal at the listener position sounds plausible, effort spent on details that become inaudible after propagation may not add value. For example, a source model with small, unwanted high-frequency components may contribute negligibly at the ear after air absorption, facade or window transmission, and HRTF filtering. In this thesis, the synthesized multi-tone AVAS signals sounded noticeably different from their recorded references when heard in isolation. Once combined with tire noise, propagation, and source motion, these differences became subtle and did not undermine the scene's plausibility.

The converse risk is that scene-level plausibility does not guarantee that each module is valid in other contexts. A mis-specified distance attenuation can be masked in one scene by compensating gain settings, yet it will bias any task that relies on distance cues. Likewise, a directivity model that is spectrally accurate on-axis but inaccurate off-axis may pass a static plausibility check yet distort outcomes as the source moves. Component-level validation is therefore necessary for some modules that are intended to generalize across studies or that are expected to contribute substantially to human response. In this thesis, perceptual component-level validations were conducted for the reproduction methods used in Paper C and Paper E. However, such validation does not always need to be perceptual. Some aspects, such as distance attenuation or facade transmission, can be numerically compared with analytical solutions. Nevertheless, a system-wide perceptual validation remains essential, because even if all modules match numerical expectations, the resulting auralization can still be judged implausible.

6.1.3 Study Design, Analysis, and Open Data

During the planning of the studies in this thesis, care was taken to align the study design and analysis with what participants could reliably report. For subjective measures, this meant favoring response formats that people can judge with confidence and statistical models that match those formats. The main studies, therefore, used fully balanced repeated measures designs with standard parametric or nonparametric analyses for continuous outcomes (Paper D–F). When it seemed unlikely that participants could deliver a stable point estimate, comparative judgments were preferred. For example, in Paper C, it was unlikely that participants, without reference, would be able to judge perceived speed on a continuous scale consistently enough to detect differences in radiation directivity. Hence, the experiment relied on paired comparisons of perceived speed. While this required familiarization with a different set of statistical evaluation methods, it ultimately might have paid off by yielding more stable ranking results. This, however, is only an assumption that would need to be confirmed by running a similar experiment using continuous response scales. Nevertheless, planning the response scale in advance reduced ambiguity during piloting and improved statistical power when effects were expected to be small.

A second consideration throughout this thesis was transparent exposure control and open data. Re-recording the exact stimuli at the listener position with a calibrated artificial head, as in Paper C to Paper F, provided traceable evidence of what participants actually heard and reduced reliance on long calibration narratives. Making these recordings, together with raw data and analysis routines, openly available in an online repository helps reviewers and readers retrace the analysis and check assumptions. When working with novel sounds that many readers have not encountered, being able to listen to the stimuli makes the studies more comprehensible than relying solely on descriptions or spectrograms. This open data approach was particularly valuable in experiments at relatively low levels, such as those in Papers E and F, where small mismatches between intended and delivered signals could otherwise raise doubts about validity. In addition, publishing complete datasets, including factors not analyzed in depth in the papers, for example, potential gender effects in Paper E, enables others to examine these aspects and to combine the data with their own results. Finally, sharing the auralization toolbox, reference measurements, and stimuli enables other researchers to replicate or extend the experiments without having to repeat substantial portions of the development.

6.2 EV Noise Findings

The findings related to electric vehicle noise can be grouped along the two thematic strands introduced earlier in this thesis: AVAS perception and safety, and noise effects in living environments. The following sections synthesize the findings for each strand and present a combined interpretation of the results in terms of regulatory implications.

6.2.1 AVAS Perception and Safety

The results of Paper C and Paper D suggest that the human response to electric vehicle noise in outdoor traffic scenarios depends strongly on the AVAS implementation. In terms of localizability, the two-tone AVAS, which is based on a regulation-compliant implementation in an existing vehicle, performed significantly worse than all other evaluated signals, particularly in situations with multiple vehicles exhibiting similar AVAS. From a safety-oriented perspective, this reduced localizability may be critical, as accurate spatial attribution is a key component of situational awareness, i.e., the ability to perceive relevant elements in the environment and anticipate how they will evolve in order to guide timely and appropriate action in traffic situations involving vulnerable road users.

Although the two-tone AVAS clearly stood out as the least suitable candidate in terms of localizability, the results do not identify a single "best" AVAS from a safety perspective, as the noise and multi-tone AVAS performed equally well in the localization experiment. While neither matched the localizability of the combustion-engine reference, both outperformed the two-tone AVAS, with localization errors small enough that practical relevance is likely limited for typical single-vehicle situations. In this context, it should be noted that this thesis does not include a pure detection experiment in which participants indicate when they first hear an approaching vehicle. Based on related literature, a tonal signal would likely perform well in such a detection task under broadband background noise. Whether such a detection advantage would be sufficient to compensate for reduced localizability and increased noise annoyance remains an open question.

Furthermore, the results of Paper C showed that extreme differences in AVAS radiation directivity can influence pass-by speed perception and that, in a direct comparison between multi-tone and noise AVAS, neither is universally perceived as faster-sounding across the examined participant pool. Instead, the data suggest that the majority of subjects can be divided into two groups that consistently choose either the noise AVAS or the multi-tone AVAS to sound faster, whereas only a few participants alternated between the two AVAS types. This underlines how personal experience and expectations can influence even abstract perceptual aspects, such as perceived pass-by speed, thereby underscoring the need to account for human perception in both AVAS design and regulatory evaluation.

Finally, informal interviews suggested that the noise AVAS more closely resembles sounds that listeners already associate with vehicles. This is relevant for situational awareness, which requires not only detection and localization but also correct identification and interpretation of a sound as an approaching car. While short-term training can improve localization of unfamiliar signals, it remains unclear whether lifelong exposure to combustion-engine sounds confers an advantage that transfers to EV contexts. Moreover, even with accurate detection and localization, delayed recognition of an unfamiliar AVAS could hinder timely responses. Direct comparisons of identification and action responses for familiar combustion sounds and novel AVAS designs are largely absent from the literature and should be addressed in future work.

6.2.2 Noise Effects in Living Environments

From the perspective of resident indoor noise exposure, the same AVAS designs exhibit different patterns of relevance. In Paper E, no significant differences in cognitive performance were observed between the noise and multi-tone AVAS at the low indoor exposure levels studied. Apart from annoyance, where both signals received slightly higher ratings than silence, neither AVAS type led to measurable increases in mental demand, frustration, or electrodermal activity relative to the silent reference.

In contrast, the two-tone AVAS consistently elicited higher ratings of annoyance, mental demand, and frustration, as well as a small but systematic increase in electrodermal activity. These results indicate that certain AVAS designs can have detrimental effects on perceived workload and physiological arousal in living environments, even when objective task performance remains unaffected. The outcome is that the two-tone AVAS, which already performed worst in localizability, also exhibits the most negative indoor noise effects, raising the question of whether implementing such highly tonal designs is favorable when both outdoor safety and indoor noise effects are considered.

Beyond perceptual responses, propagation-related aspects are also relevant for indoor exposure. In Paper E, all three AVAS types were equalized to the same facade level of $L_{\mathrm{A,eq}} = 60$ dB with a maximum level of $L_{\mathrm{A,max}} = 66$ dB. After transmission through the window, the indoor level at the participant position was slightly higher for the two-tone AVAS than for the multi-tone and noise AVAS. This highlights that facade transmission can depend on signal type, not only on overall level. For a two-tone signal, it is plausible that a substantial fraction of the energy coincides with local dips in the facade reduction index, leading to elevated indoor levels. For broadband signals with the same outdoor level, energy is distributed across the frequency spectrum, reducing the impact of spectral coincidences. While the effect observed here was subtle for the specific window construction, other facade types may produce larger differences. A targeted comparison using measured and simulated reduction curves for representative facades would therefore be a valuable extension, particularly for informing AVAS level regulations that are currently defined only at the vehicle exterior.

6.2.3 Regulatory Implications

The objective outcomes of this thesis suggest that, under the controlled conditions and AVAS designs studied here, AVAS signals consisting of only two pure tones are disadvantageous for both localization accuracy and localization time. Such signals were also found to increase annoyance, perceived mental demand, and physiological response. A straightforward regulatory implication would therefore be to discourage strongly tonal designs and to require broader spectral content. For example, regulations could mandate that the signal distributes energy across a larger number of nonadjacent one-third-octave bands, or specify bounds on spectral balance and modulation characteristics.

However, such prescriptive rules would likely increase certification complexity and may face opposition from industry stakeholders.[1] Moreover, the present work did not

[1] The records documenting the development of the corresponding U.S. regulation, FMVSS No. 141, show that several commenters urged the National Highway Traffic Safety Administration (NHTSA) to adopt more flexible guidance and opposed tighter prescriptive constraints [191].

include a pure detection task, which is the paradigm used most often in regulatory contexts. Previous studies have shown that strongly tonal signals can improve detectability under certain masking conditions. Therefore, any regulatory shift should weigh detectability, localizability, and community impact together. Such a weighting, however, requires further research, as it is currently unclear whether detectability without reliable localization translates into higher real-world risk for exposed road users.

A more flexible path would be to prioritize performance-based criteria over detailed signal prescriptions. Regulators could, for instance, set limits on tonal prominence and require minimum detection or localization performance in standardized tests across representative background environments, rather than prescribing exact acoustic properties. Such tests could also include a standardized annoyance evaluation, and regulations could, where appropriate, specify that AVAS designs should not exceed a certain annoyance threshold. This aligns requirements with real-world outcomes while preserving manufacturer design freedom.

Notably, during development of FMVSS No. 141, the 2013 Notice of Proposed Rulemaking explored an objective and repeatable jury-based recognizability option, including fixed jury composition, pass-fail criteria, and power analyses that implied panel sizes between roughly 28 and 140 listeners depending on the recognition targets [191]. The present findings support such an approach, because signals with similar overall levels and spectral composition can still evoke markedly different human responses. In the final record, NHTSA treated jury testing as an analyzed alternative but did not adopt it, citing practicality and repeatability concerns raised in comments, as well as the administrative burden of human-subject certification.[2]

A pragmatic compromise would be to certify AVAS modules rather than each vehicle model individually. Manufacturers could submit a small number of standardized AVAS designs to a jury-based or hybrid performance test once, then deploy those modules across different vehicle models without repeating human-subject certification. Such an approach would, however, still require verification that the vehicle-specific loudspeaker position yields sound pressure levels consistent with the jury-validated AVAS. From an acoustics perspective, a limited portfolio of consistent AVAS signatures may also aid public learning and recognition over time. This latter point is a motivated hypothesis rather than a confirmed finding and should be tested.

Finally, Paper D shows that localization, and likely also detection, degrade when multiple vehicles with identical AVAS are present. This could be read as an argument against module unification. An alternative is to allow controlled stochastic variability within a certified design, for example, small, bounded randomizations of spectral balance or modulation that prevent two cars from sounding exactly the same while preserving the certified performance envelope. Current measurement-only certifica-

[2] The NHTSA estimated costs of approximately $20,000 to run a 50-person laboratory jury per vehicle, using headphone playback of binaural pass-by recordings [192].

tions tend to favor strict reproducibility within make and model,[3] which complicates such variability. However, a laboratory jury or hybrid certification could accommodate families of closely related variants by requiring each variant to meet the same performance targets. In summary, while this thesis does not prescribe a single regulatory path, it shows that, despite full compliance with current requirements, some of the evaluated AVAS signals leave clear room for improvement.

6.3 Broader Implications and Transferability

6.3.1 Applications beyond EVs

Although developed for EV alerting sounds, the methodological choices and validation steps in this thesis are applicable to other moving sources, such as e-scooters, delivery robots, and emergency vehicles. In these applications, the same balance between detectability, localizability, and noise disturbance applies, even if priorities differ. A modular source-path-receiver auralization chain, with system-level plausibility checks and targeted component validation, as used here, supports rapid testing while keeping results interpretable for design and policy purposes.

The framework also extends to indoor exposure problems, for example, railway or aircraft noise. For such sources, the auralization chain would need to be adapted, because isolated source signals at different velocities are not easily recorded. Suitable options include parametric source models, synthesis from pass-by recordings, or hybrid models that combine measurements and simulation. Auralizations that accurately model building transmission, whether via numerical modeling or by incorporating a real facade in the laboratory, enable assessments that go beyond single-number metrics. Such assessments can help bridge the gap between observational exposure studies, which often rely on facade-level noise maps, and laboratory experiments that examine effects without accounting for outdoor or facade propagation.

6.3.2 Standardization and Open Science

The notion of performance-based standardization rather than purely signal-based prescriptions, as discussed in Section 6.2.3, is not limited to AVAS. In adjacent domains such as medical devices, industrial alarms, and public warning signals, most standards are specification-driven. They prescribe signal structures and acoustic margins, for example, minimum levels above background, stability, and environmental robustness, rather than mandating human performance outcomes like minimum detection rates or maximum localization error in standardized tasks. In this context,

[3] FMVSS No. 141 states "Any two vehicles of the same make, model, model year, body type, and trim level . . . shall be designed to have the same pedestrian alert sound when operating under the same test conditions and at the same speed." [17]

the ideas here are best viewed as complementary. Where appropriate, outcome-oriented tests for detection, localization, and annoyance could augment current requirements without replacing the established specification framework.

A practical suggestion would be a laboratory protocol that tests detection and localization across a small set of representative backgrounds, with published stimuli, clear pass-fail thresholds, and transparent reporting. Where sound radiates into public spaces, and alerting is not the only priority, annoyance can be measured alongside performance, enabling designers to balance audibility and community impact. The same open science practices used in this thesis, for example, sharing stimuli, analysis code, and listener position control recordings, would make such tests more comparable across laboratories.

6.4 Limitations

This thesis combines methodological development with an application to electric vehicle noise, and its limitations reflect both aspects. Although many constraints were already discussed alongside specific models and experiments in Chapter 4, this section summarizes the main limitations at a higher level to clarify the scope and interpretation of the findings.

6.4.1 Simplified Tire-Road Noise

One limitation of this thesis is that tire-road noise was modeled less accurately than the AVAS. While the tire-road auralization was judged perceptually plausible and produced realistic spectral levels, its radiation directivity is overly simple, and factors such as drive torque, pavement types, or gravel on the road were not included. In Paper E, the same tire-road synthesis model was used for all passages. This has the advantage that tire-road noise could be treated as a controlled background, so differences between conditions are not confounded by changes in tire noise. At the same time, it is not very realistic because in real-world traffic, tire-road noise can vary strongly across tire models, road surfaces, and meteorological conditions.

This limitation implies that conclusions about relative differences between AVAS designs remain valid, since they were evaluated against the same tire-road baseline, but absolute statements about overall EV noise should be interpreted with care. In particular, differences in tire-road noise characteristics could affect outcomes when tire noise is a significant contributor to the total level. This applies especially to Paper E, where some passages were rendered at velocities high enough for tire-road noise to become more prominent.

6.4.2 Experimental Context and Exposure Conditions

The experiments in this thesis were conducted under controlled laboratory conditions that differ in several respects from everyday listening environments. While some studies were conducted in the *Living Room Lab*, others used anechoic conditions, and background noise was deliberately controlled rather than representative of typical indoor or outdoor contexts. In Paper E, additional background noise was introduced to achieve realistic indoor exposure levels, but overall variability remained lower than in real-world situations.

Furthermore, exposure durations were relatively short and confined to a single experimental session, so the results primarily reflect acute responses rather than long-term adaptation, fatigue, or annoyance accumulation. In addition, the acoustic scenes were limited to a small set of trajectories, speeds, and vehicle configurations, often with few sources and without complex intersections, occlusions, or dense traffic. These constraints were necessary to isolate specific factors under study, but they limit the direct transfer of the findings to more dynamic and heterogeneous real-world environments.

These constraints are typical of laboratory work and were necessary to isolate specific factors. However, they limit generalizability. In everyday settings, EV alert sounds will interact with a wider range of background noise, including competing traffic, voices, and intermittent sources such as construction or household noise. Listeners will also be engaged in other tasks and may not attend as carefully as participants in a focused experiment. As a result, the absolute effect sizes reported here, particularly at low indoor levels, should be interpreted with caution when extrapolated to busier environments or longer exposure durations. The findings are most directly applicable to controlled scenarios with similar signal-to-noise ratios and short-term exposure, and they should be complemented by studies in more complex acoustic and behavioral contexts.

6.4.3 Range of AVAS Designs

A further limitation concerns the variety of AVAS designs covered in this thesis. The EV studies focused on three main classes: a broadband noise AVAS, a two-tone AVAS, a multi-tone AVAS, and a combustion reference. These categories still represent many currently implemented concepts, but more recent systems may be more complex. For example, real vehicle implementations can include multiple spectral layers, richer mappings of speed and acceleration, more elaborate amplitude modulation, and branding-related sound signatures. Within each of the three categories considered here, there is, in practice, a wide design space defined by carrier frequencies, bandwidth, modulation depth and rate, onset and offset behavior, and level-speed curves.

The findings in this thesis, therefore, speak most directly to the specific vehicle models synthesized and evaluated. They support relative statements such as "two-tone designs of this type can perform poorly on localization and annoyance, even when compliant", but they should not be assumed to generalize to all possible two-tone, multi-tone, or noise-like designs. Alternative choices of frequency range, modulation pattern, or dynamic mapping could yield different outcomes. Future work should sample more densely within each AVAS family and include additional classes, for example, signals with mild impulsive components, to establish how general the observed patterns are.

6.4.4 Limited Sample Diversity and Preregistration

While the sample sizes used in this thesis were generally large enough to achieve sufficient statistical power, the composition of the participant pool was relatively homogeneous. Across all six experiments, a total of 220 participants were included in the analyses. Of these, 115 identified as male, 101 as female, 2 as non-binary, 1 as other, and 1 preferred not to specify. Participants were between 20 and 64 years old, with study-wise median ages ranging from 25.5 to 27 years. This homogeneity is partly a consequence of recruiting at a university. The majority of participants were students or staff at Chalmers, and many had an educational background in acoustics or related fields. As a result, the range of age, socioeconomic background, and familiarity with technical listening tasks is limited. These factors limit the generalizability of the findings to other groups, such as older adults, people with hearing impairments, or listeners without prior critical listening experience. On the other hand, an approximately balanced gender distribution was achieved across the studies, which is not always the case in acoustic human response research.

One final limitation is that the studies in this thesis were not preregistered. Preregistration means publicly time-stamping hypotheses, primary outcomes, exclusion rules, and analysis plans before the actual experiment is conducted. The purpose is to increase transparency and credibility by fixing primary analyses in advance, reducing hindsight bias and analytical flexibility, and making clear which results are confirmatory and which are exploratory [193]. It is standard in clinical trials and is increasingly used in psychology and human factors, but it has not yet been established in acoustics research. Adopting preregistration in future experiments would sit naturally alongside the open-data and transparent reporting practices used in the present work, and help ensure that analyses follow a pre-specified plan while still allowing clearly labeled exploratory work.

6.5 Future Research Suggestions

6.5.1 Listening Experiments in Virtual Environments

This thesis shows that listening experiments in virtual acoustic environments can yield meaningful results. External validity beyond plausibility checks was not in scope, however. For some paradigms, the translation to real settings is likely. For example, there is little reason to expect that horizontal localization with single hidden loudspeakers in anechoic conditions would not transfer to AVAS loudspeakers mounted on a car chassis in a parking lot. Even then, details such as the pointing method, vertical offsets, and early reflections may shift absolute performance. Future work should therefore include targeted validations that quantify when laboratory findings generalize and when they do not.

A first priority is to compare reproduction methods for horizontal localization tasks directly. In a related study on robot sounds that used headphone playback with generic HRTFs, minimum localization errors were larger for the same AVAS signals as used in this thesis, which could reflect the reproduction method or differences in the task design [194]. A factorial comparison would be informative, for example, reproduction method (loudspeaker array with single hidden sources versus headphones with generic HRTFs), pointing interface (high-precision tracker versus off-the-shelf VR controller), and dynamic versus static rendering. Such a study would show whether future work can rely on simpler headphone-based protocols.

There is evidence that each of these factors can affect localization, yet current results do not quantify these effects specifically for vehicle localization. To judge adequacy, a study should predefine a maximum allowed performance drop attributable to the reproduction and measurement method. Setting that threshold requires knowing which errors matter in practice. As discussed in Section 6.5.2, the localization error magnitude that impacts situational awareness in traffic scenarios is not yet established. If errors of 2° alter behavior, then the experimental setup must be capable of reproducing and measuring with even higher spatial accuracy. If conditions below 15° do not affect behavior, then a setup that adds approximately 5° of error may still be sufficient to detect the differences that matter.

A second question is whether the wave field synthesis setup used in Papers E and F is actually necessary for studying noise effects in living environments. So far, the plausibility of this setup has been validated only against continuous ideal references, not against reduced playback configurations. It would be useful to compare spatially accurate array playback with simplified options, for example, two loudspeakers or a single loudspeaker with matched level and time structure. Thereby, one should consider both scene plausibility as judged by listeners and whether subjective, cognitive, and physiological measures change significantly. If not, simplified setups could be adopted more widely.

A third aspect is the role of signal-to-noise ratio in the laboratory. In the *Liv-*

ing Room Lab, background levels are below 17 dBA, so even very quiet stimuli are clearly audible. At more typical apartment backgrounds, a 40 dBA stimulus would be less salient, and very quiet conditions may feel unnatural or even uncomfortable, as reported by some participants in Paper F. Many studies treat silence as the ideal reference condition, yet it is unclear whether silence yields the best behavioral, subjective, or physiological baselines. Laboratory isolation also varies. Higher background levels in less isolated rooms can reduce audibility and complicate comparisons across studies, even when nominal stimulus levels are matched. Future experiments should therefore vary the baseline environment, for example, silence, realistic indoor backgrounds, and shaped masking noise, while measuring subjective, behavioral, and physiological responses. This would clarify when absolute silence is appropriate and when a more ecological baseline improves interpretability.

6.5.2 Electric-Vehicle Alerting Signals

This thesis examined localizability and annoyance, as well as cognitive and physiological responses, across three AVAS designs. Detectability, which has been the primary consideration in most regulatory work, was not tested directly. Future studies should therefore evaluate detectability, localizability, and annoyance within a single paradigm, across single- and multi-vehicle scenes, and across representative background noise levels.

A practical design would extend Paper D by using approaching vehicles rather than static sources. On each trial, participants would first indicate detection, then provide a localization response, followed by annoyance ratings. After each block, the perceived workload could be collected. This combined protocol would quantify trade-offs under identical acoustic conditions. As a simpler alternative, a detection-only study could be conducted using the same stimuli and scene geometry. Those results could then be combined with the localization and response data from Papers D and E to form an integrated picture.

Clear conclusions will depend on whether such detection outcomes align with the localization and annoyance patterns reported here. If a follow-up study found that the two-tone AVAS, which performed worst on localization and subjective measures, yielded the best detection results, the regulatory implications would be difficult to interpret. To address this, future work should include applied paradigms that speak directly to safety relevance. Examples include identifying thresholds for minimum detection distance and acceptable localization error for pedestrians, cyclists, and people with visual impairments, using controlled six-degree-of-freedom virtual environments and realistic traffic scenarios. Complementary observational work could also help, for example, by linking incident records to AVAS types and by conducting post-incident interviews to determine whether vehicles were detected and localized.

Beyond the noise, two-tone, and multi-tone AVAS classes, an additional signal type with mild impulsive components should be explored. A design that resembles light gravel in the tire tread would be broadband yet punctuated by low-intensity transients. Given evidence that click or impulse-train signals can improve localization in related psychoacoustic contexts, such features may aid localization while remaining familiar to the sound that we currently associate with approaching vehicles.

The influence of such familiarity should be investigated systematically. Listeners respond differently to familiar voices than to unfamiliar voices [195], but it is not well understood how familiarity with environmental sounds shapes detection, localization, appraisal, and action. Emotional and contextual factors are likely moderators, for example, whether a sound is associated with one's own car or a neighbor's, attitudes toward electric mobility, or nostalgic associations with combustion engines. Targeted experiments that manipulate familiarity and measure its influence on performance and appraisal would clarify these effects.

Finally, inclusive design suggests extending performance tests to include a range of listeners, including older adults, individuals with mild hearing loss, and individuals with visual impairments. Reporting performance by subgroup would help identify cases in which an alert is easily detected by one group but not by another, which is important for safety and equity.

CHAPTER 7

Conclusions and Outlook

This chapter presents the main conclusions of the thesis, summarizes its contributions, and outlines a focused outlook. A detailed integration of results, methodological reflections, and limitations is provided in Chapter 6.

7.1 Key Conclusions

This thesis set out to develop and demonstrate a methodological framework for studying human responses to sound using virtual acoustic environments. Electric vehicle noise, and in particular acoustic vehicle alerting systems, served as a concrete application through which the framework was implemented, tested, and refined.

First, this thesis shows that a modular, task-driven approach that combines auralization and human-response evaluation can enable controlled listening experiments with sufficient plausibility and experimental control. When reproduction methods, validation strategies, and response measures are selected based on the primary outcome of interest, virtual acoustic environments can yield robust and interpretable results across subjective, behavioral, and physiological outcomes.

Second, applying this framework to electric vehicle noise indicates that strongly tonal two-tone AVAS designs can perform poorly across several response domains despite regulatory compliance. In this thesis, two-tone signals were localized less accurately and more slowly than other designs and elicited higher annoyance, perceived mental demand, and physiological arousal (Section 6.2). These results suggest that compliance with current level and bandwidth requirements does not necessarily imply perceptual or functional suitability.

Third, noise-based and multi-tone AVAS designs performed similarly across most evaluated conditions. Both were localized more accurately than the two-tone AVAS and, under the very low-level indoor exposure conditions studied, did not produce measurable increases in cognitive performance decrements, physiological arousal, or subjective workload relative to silence, apart from slightly elevated annoyance ratings. The results, therefore, point to multiple viable design directions rather than a single optimal AVAS.

7.2 Contributions

Methodological contributions This thesis introduces a reusable, modular auralization framework that links physically informed sound synthesis and propagation with outcome-driven experimental design. It demonstrates a validation workflow that combines system-level perceptual plausibility checks with targeted component-level evaluations for specific modules critical to the outcome measure. The work also documents practical procedures for calibration and exposure control and supports transparency by openly sharing stimuli, data, and analysis code.

Empirical contributions on EV alerting sounds The thesis provides new empirical evidence on how AVAS design choices relate to localizability, perceived speed, annoyance, perceived workload, and electrodermal activity. It extends laboratory research on AVAS beyond single-source scenarios by testing multi-vehicle localization and by connecting outdoor safety-oriented effects with indoor exposure responses under controlled, facade-transmitted conditions.

Applied contributions toward regulation and design The findings highlight that AVAS signals with similar overall levels and compliance status can still evoke markedly different human responses. This supports considering performance-oriented evaluation criteria that reflect detection, localization, and community impact, alongside conventional acoustic specifications. The thesis also outlines a practical direction for integrating human-response evidence into design workflows through early human-in-the-loop testing.

7.3 Implications

The key findings of this thesis have the following implications:

For researchers Human-response studies benefit from outcome-driven auralization design. Modeling details, reproduction methods, and validation strategies should be aligned with the intended response measure and documented transparently. Modular frameworks enable controlled manipulation of individual factors while maintaining plausibility, and open sharing of stimuli and analysis routines strengthens reproducibility and comparability across laboratories.

For regulators The findings support adopting performance-based criteria that jointly consider detectability, localizability, and annoyance. Limiting tonal prominence and requiring minimum localization performance in standardized tests would better align requirements with real-world human responses than compliance checks that focus primarily on the exterior level.

For industry The results suggest avoiding narrowly tonal AVAS designs and instead exploring broadband or multi-tone signal families. Incorporating human-in-the-loop evaluations early in development can help identify unfavorable trade-offs between safety-relevant performance and community impact before implementation.

7.4 Outlook

Future research should prioritize combined paradigms that evaluate detection, localization, and annoyance for approaching vehicles across representative background environments. Such studies would clarify trade-offs under identical acoustic conditions and help calibrate performance targets that are meaningful for certification and safety assessment.

Further work should also explore AVAS designs that occupy the space between tonal and noise-like signals, for example, mildly impulsive or modulated broadband designs, which may improve localization while maintaining acceptable annoyance and a familiar sound character.

Finally, the coupling between AVAS spectral design and facade transmission deserves more systematic attention. Integrating representative building transmission characteristics with outdoor performance evaluation could help ensure that exterior requirements translate more reliably to indoor exposure and impact.

References

[1] European Environment Agency (EEA), "Environmental noise in Europe 2025," Publications Office of the European Union, Luxembourg, Tech. Rep., 2025.

[2] World Health Organization (WHO), *Burden of disease from environmental noise: quantification of healthy life years lost in Europe*, 2011.

[3] United Nations (UN), *Transforming our world: the 2030 Agenda for Sustainable Development*, Resolution adopted by the General Assembly, A/RES/70/1, 2015.

[4] International Energy Agency (IEA), *Global EV Outlook 2025*, 2025.

[5] F. Anfosso-Lédée, M. Paviotti, and S. Kephalopoulos, "Common noise assessment methods in Europe (CNOSSOS-EU) – To be used by the EU Member States for strategic noise mapping following adoption as specified in the Environmental Noise Directive 2002/49/EC," Joint Research Centre: Institute for Health and Consumer Protection, Tech. Rep., 2012.

[6] World Health Organization (WHO), *Environmental Noise Guidelines for the European Region*, 2018.

[7] L. Garay-Vega, A. Hastings, J. K. Pollard, M. Zuschlag, and M. Stearns, "Quieter Cars and the Safety of Blind Pedestrians: Phase 1," National Highway Traffic Safety Administration, Washington, DC, Tech. Rep. DOT HS 811 304, 2010.

[8] M.-A. Pallas, M. Bérengier, R. Chatagnon, M. Czuka, M. Conter, and M. Muirhead, "Towards a model for electric vehicle noise emission in the European prediction method CNOSSOS-EU," *Applied Acoustics*, vol. 113, pp. 89–101, 2016.

[9] M. R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors: The Journal of Human Factors and Ergonomics Society*, vol. 37, no. 1, pp. 32–64, 1995.

[10] European Blind Union (EBU), *European Blind Union response to the European Commission public consultation on the AVAS Pause Function Prohibition*, https://www.euroblind.org/sites/default/files/2019-01-10_ebu_response-to-avas-consultation.pdf (accessed 2025-09-12), 2019.

[11] R. Hanna, "Incidence of Pedestrian and Bicyclist Crashes by Hybrid Electric Passenger Vehicles," National Highway Traffic Safety Administration, Washington, DC, Tech. Rep. DOT HS 811 204, 2009.

[12] J. Wu, R. Austin, and C.-L. Chen, "Incidence rates of pedestrian and bicyclist crashes by hybrid electric passenger vehicles: an update.," National Highway Traffic Safety Administration, Washington, DC, Tech. Rep. DOT HS 811 526, 2011.

[13] P. A. Morgan, L. Morris, M. Muirhead, L. K. Walter, and J. Martin, "Assessing the perceived safety risk from quiet electric and hybrid vehicles to vision-impaired pedestrians," Transport Research Laboratory for UK Department for Transport, Tech. Rep. PPR525, 2011.

[14] X. Hou, M. Su, C. Liu, Y. Li, and Q. Ma, "Examination of the Factors Influencing the Electric Vehicle Accident Size in Norway (2020–2021)," *World Electric Vehicle Journal*, vol. 15, no. 1, p. 3, 2023.

[15] P. J. Edwards, S. Moore, and C. Higgins, "Pedestrian safety on the road to net zero: cross-sectional study of collisions with electric and hybrid-electric cars in Great Britain," *Journal of Epidemiology & Community Health*, vol. 78, no. 8, pp. 487–492, 2024.

[16] United Nations Economic Commission for Europe (UNECE), *Regulation No 138 of the Economic Commission for Europe of the United Nations (UNECE) — Uniform provisions concerning the approval of Quiet Road Transport Vehicles with regard to their reduced audibility [2017/71]*, 2017.

[17] National Highway Traffic Safety Administration (NHTSA), *FMVSS No. 141 - Minimum Sound Requirements for Hybrid and Electric Vehicles*, National Highway Traffic Safety Administration - Department of Transportation, 2016.

[18] D. A. Laird, "The Effects of Noise: A summary of experimental literature," *The Journal of the Acoustical Society of America*, vol. 1, no. 2A, pp. 256–262, 1930.

[19] D. A. Laird, "The influence of noise on production and fatigue, as related to pitch, sensation level, and steadiness of the noise," *Journal of Applied Psychology*, vol. 17, no. 3, pp. 320–330, 1933.

[20] S. J. Schlittmeier and J. E. Marsh, "Review of research on the effects of noise on cognitive performance 2017-2021," in *The 13th ICBEN Congress on Noise as a Public Health Problem*, Stockholm, 2021, pp. 1–48.

[21] J. Dratva *et al.*, "Transportation Noise and Blood Pressure in a Population-Based Sample of Adults," *Environmental Health Perspectives*, vol. 120, no. 1, pp. 50–55, 2012.

[22] W. Babisch, B. Beule, M. Schust, N. Kersten, and H. Ising, "Traffic Noise and Risk of Myocardial Infarction," *Epidemiology*, vol. 16, no. 1, pp. 33–40, 2005.

[23] A. Muzet, "Environmental noise, sleep and health," *Sleep Medicine Reviews*, vol. 11, no. 2, pp. 135–142, 2007.

[24] L. C. Erickson and R. S. Newman, "Influences of Background Noise on Infants and Children," *Current Directions in Psychological Science*, vol. 26, no. 5, pp. 451–457, 2017.

[25] M. Basner *et al.*, "ICBEN review of research on the biological effects of noise 2011-2014," *Noise & Health*, vol. 17, no. 75, pp. 57–82, 2015.

[26] C. L. Themann and E. A. Masterson, "Occupational noise exposure: A review of its effects, epidemiology, and impact with recommendations for reducing its burden," *The Journal of the Acoustical Society of America*, vol. 146, no. 5, pp. 3879–3905, 2019.

[27] K. P. Waye and E. v. Kempen, "Non-auditory effects of noise: an overview of the state of the science of the 2017-2020 period," in *The 13th ICBEN Congress on Noise as a Public Health Problem*, Stockholm, 2021.

[28] T. L. Lash, T. J. VanderWeele, S. Haneause, and K. Rothman, *Modern Epidemiology*. Philadelphia, US: Wolters Kluwer Health, 2021, ISBN: 9781975166281.

[29] W. Shadish, T. Cook, and D. Campbell, *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston, Massachusetts: Houghton Mifflin, 2002, ISBN: 9780395615560.

[30] N. Zacharov, *Sensory Evaluation of Sound*. Boca Rato: CRC Press, 2018, ISBN: 9781498751360.

[31] M. Vorländer, *Auralization, Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Heidelberg: Springer Berlin, 2008, ISBN: 9783540488293.

[32] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization - An Overview," *Journal of the Audio Engineering Society*, vol. 41, no. 11, pp. 861–875, 1993.

[33] G. Kramer, *Auditory Display: Sonification, Audification, and Auditory Interfaces*. Boca Raton: CRC Press, 1994, ISBN: 0201626047.

[34] M. Grieves and J. Vickers, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," in *Transdisciplinary perspectives on complex systems: New findings and approaches*, Springer, 2017, pp. 85–113, ISBN: 9783319387567.

[35] B. Ketzler, V. Naserentin, F. Latino, C. Zangelidis, L. Thuvander, and A. Logg, "Digital Twins for Cities: A State of the Art Review," *Built Environment*, vol. 46, no. 4, pp. 547–573, 2020.

[36] J. Ahrens and L. Müller, "Perceptually Transparent Binaural Auralization of Simulated Sound Fields," *Journal of the Audio Engineering Society*, vol. 74, no. 3, pp. 118–129, 2026.

[37] R. Pieren, T. Bütler, and K. Heutschi, "Auralization of Accelerating Passenger Cars Using Spectral Modeling Synthesis," *Applied Sciences*, vol. 6, no. 1, p. 5, 2015.

[38] C. Dreier and M. Vorländer, "Drone auralization model with statistical synthesis of amplitude and frequency modulations," *Acta Acustica*, vol. 8, p. 35, 2024.

[39] R. Pieren, K. Heutschi, J. M. Wunderli, M. Snellen, and D. G. Simons, "Auralization of railway noise: Emission synthesis of rolling and impact noise," *Applied Acoustics*, vol. 127, pp. 34–45, 2017.

[40] J. Forssén, A. Hoffmann, and W. Kropp, "Auralization model for the perceptual evaluation of tyre–road noise," *Applied Acoustics*, vol. 132, pp. 232–240, 2018.

[41] A. R. Aumann, S. A. Rizzi, and S. L. Heath, "Receiver-based auralization of broadband aircraft flyover noise using the NASA Auralization Framework," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Chicago, USA, 2018, pp. 411–420.

[42] C. G. Dreier, "Virtual acoustics: synthesis of traffic sound sources," Ph.D. dissertation, Rheinisch-Westfälische Technische Hochschule Aachen, 2025.

[43] C. H. Kasess, C. Reichl, H. Waubke, and P. Majdak, "Perception Rating of the Acoustic Emissions of Heat Pumps," in *Forum Acusticum*, Lyon, France, 2020, pp. 2453–2458.

[44] W. Kropp, C. Hoever, and J. Theyssen, "Auralisation of Tyre/Road Noise," in *Fortschritte der Akustik – DAGA*, Hannover, Germany, 2024, pp. 455–458.

[45] S. A. Rizzi and A. K. Sahai, "Auralization of air vehicle noise for community noise assessment," *CEAS Aeronautical Journal*, vol. 10, no. 1, pp. 313–334, 2019.

[46] J. W. Beauchamp, "Additive Synthesis of Harmonic Musical Tones," *Journal of the Audio Engineering Society*, vol. 14, no. 4, pp. 332–342, 1966.

[47] J. Forssén, P. Andersson, P. Bergman, K. Fredriksson, and P. Zimmermann, "Auralisation of truck engine sound – preliminary results using a granular approach," in *Proceedings of the International Conference on Acoustics including the 40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA)*, Merano, Italy, 2013.

[48] J. Jagla, J. Maillard, and N. Martin, "Sample-based engine noise synthesis using an enhanced pitch-synchronous overlap-and-add method," *The Journal of the Acoustical Society of America*, vol. 132, no. 5, pp. 3098–3108, 2012.

[49] H. Jonasson, U. Sandberg, G. v. Blokland, J. Ejsmont, G. Watts, and M. Luminari, "Harmonoise project. Final report work package 1.1: Source modelling of road vehicles," SP Sveriges Provnings- och Forskningsinstitut, Tech. Rep., 2004.

[50] H. G. Jonasson and S. Storeheier, "Nord 2000. New Nordic prediction method for road traffic noise," SP Swedish National Testing and Research Institute, Tech. Rep., 2001.

[51] R. A. McCarthy, Y. Zhang, S. A. Verburg, W. F. Jenkins, and P. Gerstoft, "Machine Learning in Acoustics: A Review and Open-source Repository," *npj Acoustics*, vol. 1, no. 1, p. 18, 2025.

[52] R. Liu, Y. Yin, Y. Peng, and X. Zheng, "Predicting Vehicle-Engine-Radiated Noise Based on Bench Test and Machine Learning," *Machines*, vol. 13, no. 8, p. 724, 2025.

[53] X. Yin, M. Fallah-Shorshani, R. McConnell, S. Fruin, and M. Franklin, "Predicting Fine Spatial Scale Traffic Noise Using Mobile Measurements and Machine Learning," *Environmental Science & Technology*, vol. 54, no. 20, pp. 12 860–12 869, 2020.

[54] S. Wirler, S. J. Schlecht, and V. Pulkki, "Machine Learning Based Auralization of Rigid Sphere Scattering," *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, vol. 00, pp. 1–8, 2021.

[55] D. Xie and M. Hornikx, "Real-time Auralization based on a Convolutional Neural Network Trained by Low-Frequency Wave-Based Calculations," in *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum*, Turin, Italy, 2023, pp. 3189–3195.

[56] Z. Kong, W. Ping, J. Huang, K. Zhao, and B. Catanzaro, "DiffWave: A Versatile Diffusion Model for Audio Synthesis," in *International Conference on Learning Representations (ICLR)*, Online, 2021.

[57] A. v. d. Oord *et al.*, "WaveNet: A Generative Model for Raw Audio," in *9th ISCA Workshop on Speech Synthesis Workshop (SSW 9)*, Sunnyvale, CA, USA, 2016, p. 125.

[58] L. Wyse, P. Kamath, and C. Gupta, "Sound Model Factory: An Integrated System Architecture for Generative Audio Modelling," in *Artificial Intelligence in Music, Sound, Art and Design. EvoMUSART*, Madrid, Spain, 2022, pp. 308–322.

[59] W. Kropp, F.-X. Bécot, and S. Barrelet, "On the Sound Radiation from Tyres," *Acustica united with acta acustica*, vol. 86, pp. 769–779, 2000.

[60] J. A. Hargreaves, L. R. Rendell, and Y. W. Lam, "A framework for auralization of boundary element method simulations including source and receiver directivity," *The Journal of the Acoustical Society of America*, vol. 145, no. 4, pp. 2625–2637, 2019.

[61] P. Luizard *et al.*, "A 256-microphone measurement system to estimate 3D directivity patterns," in *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*, Turin, Italy, 2023, pp. 4151–4156.

[62] C. Dreier and M. Vorländer, "Speed-dependent directivity patterns of road-traffic vehicles," *The Journal of the Acoustical Society of America*, vol. 157, no. 4, pp. 2735–2749, 2025.

[63] S. Bilbao, J. Ahrens, and B. Hamilton, "Incorporating source directivity in wave-based virtual acoustics: Time-domain models and fitting to measured dataa)," *The Journal of the Acoustical Society of America*, vol. 146, no. 4, pp. 2692–2703, 2019.

[64] U. Sloma and A. Neidhardt, "Investigations on the Impact of Listener Movement on the Perception of Source Directivity in Virtual Acoustic Environments," in *Fortschritte der Akustik*, München, Germany, 2018, pp. 1433–1436.

[65] T. Robotham, O. S. Rummukainen, and E. A. P. Habets, "Towards the Perception of Sound Source Directivity Inside Six-Degrees-of-Freedom Virtual Reality," in *5th International Conference on Spatial Audio*, Ilmenau, Germany, 2019.

[66] L. M. Wang and M. C. Vigeant, "Evaluations of output from room acoustic computer modeling and auralization due to different sound source directionalities," *Applied Acoustics*, vol. 69, no. 12, pp. 1281–1293, 2008.

[67] N. R. Shabtai, G. Behler, M. Vorländer, and S. Weinzierl, "Generation and analysis of an acoustic radiation pattern database for forty-one musical instruments," *The Journal of the Acoustical Society of America*, vol. 141, no. 2, pp. 1246–1256, 2017.

[68] E. G. Williams, *Fourier Acoustics.* London: Academic Press, 1999, ISBN: 9780127539607.

[69] J. Ahrens and S. Bilbao, "Computation of Spherical Harmonic Representations of Source Directivity Based on the Finite-Distance Signature," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 83–92, 2020.

[70] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*. Princeton, NJ: Princeton University Press, 1968, ISBN: 0691084254.

[71] H. Kuttruff, *Room Acoustics*, English, 6th ed. Boca Raton: CRC Press, 2016, ISBN: 9781315372150.

[72] L. Savioja and U. P. Svensson, "Overview of geometrical room acoustic modeling techniques," *The Journal of the Acoustical Society of America*, vol. 138, no. 2, pp. 708–730, 2015.

[73] J. Forssén, L. Müller, E. Hedlund, and W. Kropp, "Hybrid prediction tool implemented for acoustic design studies of open plan office spaces," *Applied Acoustics*, vol. 248, p. 111 284, 2026.

[74] T. E. Vigran, *Building Acoustics*. New York, NY: Taylor & Francis, 2008, ISBN: 9780367865214.

[75] ISO 12354, *3:2017 – Building acoustics – Estimation of acoustic performance of buildings from the performance of elements – Part 3: Airborne sound insulation against outdoor sound*, (International Organization for Standardization, Geneva, Switzerland), 2017.

[76] ISO 9613, *1:1993 Acoustics – Attenuation of sound during propagation outdoors — Part 1: Calculation of the absorption of sound by the atmosphere*, (International Organization for Standardization, Geneva, Switzerland), 1993.

[77] F. Rietdijk, K. Heutschi, and J. Forssén, "Modelling sound propagation in the presence of atmospheric turbulence for the auralization of aircraft noise," *The Journal of the Acoustical Society of America*, vol. 136, no. 4, pp. 2286–2286, 2014.

[78] ISO 9613, *2: Acoustics – Attenuation of sound during propagation outdoors – Part 2: General method of calculation*, 2006.

[79] A. D. Pierce, *Acoustics, An Introduction to Its Physical Principles and Applications*. Cham, Switzerland: Springer, 2019, ISBN: 9783030112134.

[80] J. Ahrens and S. Spors, "Wave field synthesis of moving virtual sound sources with complex radiation properties," *The Journal of the Acoustical Society of America*, vol. 130, no. 5, pp. 2807–2816, 2011.

[81] J. Ahrens and S. Spors, "Reproduction of Moving Virtual Sound Sources with Special Attention to the Doppler Effect," in *124th Convention of the AES*, Amsterdam, The Netherlands, 2008, pp. 1–12.

[82] J. Theyssen, T. Deppisch, A. Pieringer, and W. Kropp, "On the efficient simulation of pass-by noise signals from railway wheels," *Journal of Sound and Vibration*, vol. 564, p. 117 889, 2023.

[83] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial Decomposition Method for Room Impulse Responses," *Journal of the Audio Engineering Society*, vol. 61, pp. 17–28, 2013.

[84] F. Zotter and M. Frank, *Ambisonics, A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Cham, Switzerland: Springer Nature, 2019, ISBN: 9783030172060.

[85] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, 1997.

[86] E. Verheijen, "Sound Reproduction by Wave Field Synthesis," Ph.D. dissertation, Delft University of Technology, 1998.

[87] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, Massachusetts: The MIT Press, 1996, ISBN: 9780262268684.

[88] H. Møller and C. S. Pedersen, "Hearing at low and infrasonic frequencies," *Noise & health*, vol. 6, no. 23, pp. 37–57, 2004.

[89] F. Zotter, H. Pomberger, and M. Noisternig, "Energy-Preserving Ambisonic Decoding," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 37–47, 2012.

[90] H. Møller, "Reproduction of Artificial-Head Recordings through Loudspeakers," *J. Audio Eng. Soc*, vol. 37, no. 1/2, pp. 30–33, 1989.

[91] J. Ahrens, *Analytic Methods of Sound Field Synthesis*. Heidelberg: Springer Berlin, 2012, ISBN: 9783642257421.

[92] H. Hamdad, C. Pézerat, B. Gauvreau, C. Locqueteau, and Y. Denoual, "Sensitivity analysis and propagation of uncertainty for the simulation of vehicle pass-by noise," *Applied Acoustics*, vol. 149, pp. 85–98, 2019.

[93] H. Fastl and E. Zwicker, *Psychoacoustics, Facts and Models*, 3rd ed. Heidelberg: Springer Berlin, 2007, ISBN: 9783642517655.

[94] P. Majdak, C. Hollomey, and R. Baumgartner, "AMT 1.x: A toolbox for reproducible research in auditory modeling," *Acta Acust.*, vol. 6, p. 19, 2022.

[95] F. Brinkmann, A. Lindau, and S. Weinzierl, "On the authenticity of individual dynamic binaural synthesis," *The Journal of the Acoustical Society of America*, vol. 142, no. 4, pp. 1784–1795, 2017.

[96] A. Lindau and S. Weinzierl, "Assessing the plausibility of virtual acoustic environments," in *Forum Acusticum 2011*, Aalborg, Denmark, 2011, pp. 1187–1192, ISBN: 978-84-694-1520-7.

[97] C. Kuhn-Rahloff, *Realitätstreue, Natürlichkeit, Plausibilität, Perzeptive Beurteilungen in der Elektroakustik*. Heidelberg: Springer Berlin, 2012, ISBN: 9783642220715.

[98] R. S. Pellegrini, "Quality Assessment of Auditory Virtual Environments," in *Proceedings of the 2001 International Conference on Auditory Display*, Espoo, Finland, 2001, pp. 161–168.

[99] W. A. Munson and M. B. Gardner, "Loudness Patterns—a New Approach," *The Journal of the Acoustical Society of America*, vol. 22, no. 2, pp. 177–190, 1950.

[100] J. Boley and M. Lester, "Statistical Analysis of ABX Results Using Signal Detection Theory," in *Proceedings of the 127th Convention of the Audio Engineering Society*, New York, NY, USA, 2009, p. 7826.

[101] International Telecommunication Union (ITU), *Recommendation ITU-R BS.1534-3: Method for the subjective assessment of intermediate quality level of audio systems*, 2015.

[102] CATT, *CATT Acoustic*, `https://www.catt.se` (accessed 2025-10-03).

[103] Odeon A/S, *ODEON*, `www.odeon.dk` (accessed 2025-10-03).

[104] Treble Technologies, *TREBLE*, `www.treble.tech` (accessed 2025-10-03).

[105] Institute for Hearing Technology and Acoustics (IHTA), RWTH Aachen University, *RAVEN*, `https://www.virtualacoustics.org/RAVEN/` (accessed 2025-10-03).

[106] Institute for Hearing Technology and Acoustics (IHTA), RWTH Aachen University, *Virtual Acoustics – A real-time auralization framework for scientific research*, `http://www.virtualacoustics.org/` (accessed 2025-09-04).

[107] M. Noisternig, B. F. G. Katz, S. Siltanen, and L. Savioja, "Framework for Real-Time Auralization in Architectural Acoustics," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 1000–1015, 2008.

[108] G. Grimm, J. Luberadzka, and V. Hohmann, "A Toolbox for Rendering Virtual Acoustic Environments in the Context of Audiology," *Acta Acustica united with Acustica*, vol. 105, no. 3, pp. 566–578, 2019.

[109] K. d. Haas, M. Schutte, and S. Ewert, "Real-time Virtual Environment and Room Acoustics Simulator," in *Proceedings of the 11th Convention of the European Acoustics Association Forum Acusticum / EuroNoise 2025*, Málaga, Spain, 2025, pp. 6533–6540.

[110] A. R. Aumann, B. C. Tuttle, W. L. Chapin, and S. A. Rizzi, "The NASA Auralization Framework and plugin architecture," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, San Francisco, USA, 2015, pp. 1932–1943.

[111] M. Gochfeld, "Chronologic History of Occupational Medicine," *Journal of Occupational and Environmental Medicine*, vol. 47, no. 2, pp. 96–114, 2005.

[112] M. E. Paladino, M. Belingheri, R. Mazzagatti, and M. A. Riva, "Noise-induced hearing loss in the pre-industrial era: early contributions in De Morbis Artificum by Bernardino Ramazzini (1633–1714)," *The Journal of Laryngology & Otology*, vol. 138, no. 1, pp. 7–9, 2024.

[113] F. E. Thurston, "The worker's ear: A history of noise-induced hearing loss," *American Journal of Industrial Medicine*, vol. 56, no. 3, pp. 367–377, 2013.

[114] H. Lane and B. Tranel, "The Lombard Sign and the Role of Hearing in Speech," *Journal of Speech and Hearing Research*, vol. 14, no. 4, pp. 677–709, 1971.

[115] E. C. Cherry, "Some Experiments on the Recognition of Speech, with One and with Two Ears," *The Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, 1953.

[116] D. Broadbent, *Perception and Communication.* London: Pergamon Press, 1958, ISBN: 978-1-4832-0079-8.

[117] International Organization for Standardization, "ISO 15666: Acoustics – Assessment of noise annoyance by means of social and socio-acoustic surveys," International Organization for Standardization, Tech. Rep., 2021.

[118] ISO 12913, *3: Acoustics – Soundscape – Part 3: Data analysis*, (International Organization for Standardization, Geneva, Switzerland), 2025.

[119] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Advances in Psychology*, vol. 52, pp. 139–183, 1988.

[120] S. G. Hart, "Nasa-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, 2006.

[121] F. R. H. Zijlstra, "Efficiency in work behaviour: A design approach for modern tools," Ph.D. dissertation, Delft University Press, 1985.

[122] P. Susini, S. McAdams, S. Winsberg, I. Perry, S. Vieillard, and X. Rodet, "Characterizing the sound quality of air-conditioning noise," *Applied Acoustics*, vol. 65, no. 8, pp. 763–790, 2004.

[123] M. Basner and S. McGuire, "WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Effects on Sleep," *International Journal of Environmental Research and Public Health*, vol. 15, no. 3, p. 519, 2018.

[124] M. Basner *et al.*, "Auditory and non-auditory effects of noise on health," *The Lancet*, vol. 383, no. 9925, pp. 1325–1332, 2014.

[125] B. M. Savers, H. A. Beagley, and W. R. Henshall, "The Mechanism of Auditory Evoked EEG Responses," *Nature*, vol. 247, no. 5441, pp. 481–483, 1974.

[126] R. Näätänen and I. Winkler, "The concept of auditory stimulus representation in cognitive neuroscience.," *Psychological Bulletin*, vol. 125, no. 6, pp. 826–859, 1999.

[127] J. Ke, J. Du, and X. Luo, "The effect of noise content and level on cognitive performance measured by electroencephalography (EEG)," *Automation in Construction*, vol. 130, p. 103 836, 2021.

[128] S. Luck, *An Introduction to the Event-Related Potential Technique, second edition.* Cambridge, Massachusetts: MIT Press, 2014, ISBN: 9780262525855.

[129] A. Roye, T. Jacobsen, and E. Schröger, "Personal significance is encoded automatically by the human brain: an event-related potential study with ringtones," *European Journal of Neuroscience*, vol. 26, no. 3, pp. 784–790, 2007.

[130] L. Han, Y. Liu, D. Zhang, Y. Jin, and Y. Luo, "Low-Arousal Speech Noise Improves Performance in N-Back Task: An ERP Study," *PLoS ONE*, vol. 8, no. 10, e76261, 2013.

[131] L. Müller, W. Kropp, G. Zachos, and J. Forssén, "Investigating Low Frequency Sound from Traffic in a Living Room Lab," in *Fortschritte der Akustik*, Vienna, Austria, 2021, pp. 160–163.

[132] E. Wascher *et al.*, "Frontal theta activity reflects distinct aspects of mental fatigue," *Biological Psychology*, vol. 96, pp. 57–65, 2014.

[133] B. Ross and K. Tremblay, "Stimulus experience modifies auditory neuromagnetic responses in young and older listeners," *Hearing Research*, vol. 248, no. 1-2, pp. 48–59, 2009.

[134] C. S. Herrmann, U. Oertel, Y. Wang, B. Maess, and A. D. Friederici, "Noise affects auditory and linguistic processing differently: an MEG study," *NeuroReport*, vol. 11, no. 2, pp. 227–230, 2000.

[135] R. Luke *et al.*, "Analysis methods for measuring passive auditory fNIRS responses generated by a block-design paradigm," *Neurophotonics*, vol. 8, no. 2, pp. 025 008–025 008, 2021.

[136] J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, *Handbook of Psychophysiology*, 4th ed. Cambridge: Cambridge University Press, 2016, ISBN: 9781107415782.

[137] W. Babisch, "Cardiovascular effects of noise.," eng, *Noise & health*, vol. 13, pp. 201–4, 2011.

[138] H. Ising and C. Braun, "Acute and chronic endocrine effects of noise: Review of the research conducted at the Institute for Water, Soil and Air Hygiene.," eng, *Noise & health*, vol. 2, pp. 7–24, 2000.

[139] S. Laborde, E. Mosley, and J. F. Thayer, "Heart Rate Variability and Cardiac Vagal Tone in Psychophysiological Research – Recommendations for Experiment Planning, Data Analysis, and Data Reporting," *Frontiers in Psychology*, vol. 08, p. 213, 2017.

[140] W. Boucsein, *Electrodermal Activity*. New York, NY: Springer, 2012, ISBN: 9781461411253.

[141] S. A. Stansfeld and M. P. Matheson, "Noise pollution: non-auditory effects on health," *British Medical Bulletin*, vol. 68, no. 1, pp. 243–257, 2003.

[142] H. Ising and B. Kruppa, "Health effects caused by noise: evidence in the literature from the past 25 years.," eng, *Noise & health*, vol. 6, pp. 5–13, 2004.

[143] C. D. Wickens, S. E. Gordon, and Y. Liu, *An Introduction to Human Factors Engineering*, 2nd ed. Upper Saddle River, NJ: Pearson, 2013, ISBN: 9781292022314.

[144] E. Strauss, E. M. S. Sherman, and O. Spreen, *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*, 3rd ed. New York, NY: Oxford University Press, 2006, ISBN: 0195159578.

[145] D. Kahneman, *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall, 1973, ISBN: 0130505188.

[146] P. Salamé and A. Baddeley, "Disruption of short-term memory by unattended speech: Implications for the structure of working memory," *Journal of Verbal Learning and Verbal Behavior*, vol. 21, no. 2, pp. 150–164, 1982.

[147] R. M. Yerkes and J. D. Dodson, "The relation of strength of stimulus to rapidity of habit-formation," *Journal of Comparative Neurology and Psychology*, vol. 18, no. 5, pp. 459–482, 1908.

[148] S. Hygge, "Classroom experiments on the effects of different noise sources and sound levels on long-term recall and recognition in children," *Applied Cognitive Psychology*, vol. 17, no. 8, pp. 895–914, 2003.

[149] S. J. Schlittmeier, J. Hellbrück, R. Thaden, and M. Vorländer, "The impact of background speech varying in intelligibility: Effects on cognitive performance and perceived disturbance," *Ergonomics*, vol. 51, no. 5, pp. 719–736, 2008.

[150] J. E. Marsh, J. Gädtke, and S. J. Schlittmeier, "A Review of the Effect of Noise on Cognitive Performance 2021-2023," in *14th ICBEN Congress on Noise as a Public Health Problem*, Belgrade, Serbia, 2023, pp. 1–64.

[151] G. A. Gescheider, *Psychophysics: The Fundamentals*. New York: Psychology Press, 1997, ISBN: 9780203774458.

[152] W. A. Yost, A. N. Popper, and R. R. Fay, *Auditory Perception of Sound Sources*. New York, NY: Springer, 2008, ISBN: 9780387713045.

[153] S. S. Stevens and E. B. Newman, "The Localization of Actual Sources of Sound," *The American Journal of Psychology*, vol. 48, no. 2, p. 297, 1936.

[154] R. Baumgartner, P. Majdak, and B. Laback, "Modeling sound-source localization in sagittal planes for human listeners," *The Journal of the Acoustical Society of America*, vol. 136, no. 2, pp. 791–802, 2014.

[155] V. Best, F. J. Gallun, S. Carlile, and B. G. Shinn-Cunningham, "Binaural interference and auditory grouping," *The Journal of the Acoustical Society of America*, vol. 121, no. 2, pp. 1070–1076, 2007.

[156] C. Lorenzi, S. Gatehouse, and C. Lever, "Sound localization in noise in normal-hearing listeners," *The Journal of the Acoustical Society of America*, vol. 105, no. 3, pp. 1810–1820, 1999.

[157] R. A. Magill and D. I. Anderson, *Motor Learning and Control: Concepts and Applications*. New York, NY: McGraw Hill, 2021, ISBN: 9781260240702.

[158] A. Smith, "Noise, performance efficiency and safety," *International Archives of Occupational and Environmental Health*, vol. 62, no. 1, pp. 1–5, Jan. 1990.

[159] F. Gravetter and L. Forzano, *Research Methods for the Behavioral Sciences*, 4th ed. Belmont, CA: Cengage Learning, 2011, ISBN: 9781111342258.

[160] B. C. Moore, *An Introduction to the Psychology of Hearing*, 4th ed. London, UK: Academic Press, 1997, ISBN: 0125056273.

[161] J. L. Myers, A. D. Well, and R. F. L. Jr, *Research Design and Statistical Analysis: Third Edition*. New York: Routledge, 2010, ISBN: 9780203726631.

[162] G. Cumming, "The New Statistics," *Psychological Science*, vol. 25, no. 1, pp. 7–29, 2013.

[163] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. New York: Routledge, 1988, ISBN: 9780203771587.

[164] W. J. Conover, *Practical Nonparametric Statistics*, 3rd ed. New York: Wiley, 1999, ISBN: 978-0-471-16068-7.

[165] P. Misra and P. Enge, *Global Positioning System: Signals, Measurements, and Performance*, 2nd ed. Lincoln, Massachusetts, US: Ganga-Jamuna Press, 2006, ISBN: 0970954417.

[166] ISO 11819, *2:2017 – Acoustics – Measurement of the influence of road surfaces on traffic noise – Part 2: The close-proximity method*, (International Organization for Standardization, Geneva, Switzerland), 2017.

[167] A. Oppenheim and R. Schafer, *Discrete-time Signal Processing*, 3rd. Harlow, England: Pearson, 2014, ISBN: 9780132067096.

[168] V. I. Lebedev and D. N. Laikov, "A Quadrature Formula for the Sphere of the 131st Algebraic Order of Accuracy," *Doklady Mathematics*, vol. 59, pp. 477–481, 1999.

[169] M. Frank and M. Brandner, "Perceptual Evaluation of Spatial Resolution in Directivity Patterns," in *Fortschritte der Akustik – DAGA*, Rostock, 2019, pp. 74–77.

[170] T. Lübeck and C. Pörschmann, "Investigation of the minimum required spatial resolution of moving sound sources," in *Fortschritte der Akustik – DAGA*, Hamburg, Germany, 2023, pp. 1108–1111.

[171] I. B. Hagai, M. Pollow, M. Vorländer, and B. Rafaely, "Acoustic centering of sources measured by surrounding spherical microphone arrays," *The Journal of the Acoustical Society of America*, vol. 130, no. 4, pp. 2003–2015, 2011.

[172] F. Rietdijk, J. Forssén, and K. Heutschi, "Generating Sequences of Acoustic Scintillations," *Acta Acustica united with Acustica*, vol. 103, no. 2, pp. 331–338, 2017.

[173] J. Forssén, "Scintillating and decorrelating signals for different propagation paths in a random medium," *Applied Acoustics*, vol. 221, p. 110 038, 2024.

[174] V. E. Ostashev and D. K. Wilson, *Acoustics in Moving Inhomogeneous Media*, 2nd ed. London: CRC Press, 2015, ISBN: 9780429176494.

[175] B. Masiero, J. Fels, and M. Vorländer, "Review of the crosstalk cancellation filter technique," in *Proc. International Conference on Spatial Audio*, Detmold, 2011.

[176] M. A. Akeroyd *et al.*, "The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics," *The Journal of the Acoustical Society of America*, vol. 121, no. 2, pp. 1056–1069, 2007.

[177] P. Majdak, B. Masiero, and J. Fels, "Sound localization in individualized and non-individualized crosstalk cancellation systems," *The Journal of the Acoustical Society of America*, vol. 133, no. 4, pp. 2055–2068, 2013.

[178] H. A. David, *The Method of Paired Comparisons*. London: C. Griffin, 1988, ISBN: 9780852642900.

[179] R. A. Bradley and M. E. Terry, "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324–345, 1952.

[180] F. Wickelmaier and C. Schmid, "A Matlab function to estimate choice model parameters from paired-comparison data," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 1, pp. 29–40, 2004.

[181] H. Mali and N. K. Venkateswaran, "Auditory spatial assessment in real and virtual environments: a scoping review of current evidence and future directions," *The Egyptian Journal of Otolaryngology*, vol. 42, no. 1, p. 74, 2026.

[182] M. Ramírez, J. M. Arend, P. v. Gablenz, H. R. Liesefeld, and C. Pörschmann, "Toward Sound Localization Testing in Virtual Reality to Aid in the Screening of Auditory Processing Disorders," *Trends in Hearing*, vol. 28, pp. 1–17, 2024.

[183] S. J. Schlittmeier, A. Feil, A. Liebl, and J. Hellbrück, "The impact of road traffic noise on cognitive performance in attention-based tasks depends on noise level even within moderate-level ranges," *Noise & Health*, vol. 17, no. 76, pp. 148–157, 2015.

[184] A. V. Fisher, "Selective sustained attention: a developmental foundation for cognition," *Current Opinion in Psychology*, vol. 29, pp. 248–253, 2019.

[185] B. D. Auerbach and H. J. Gritton, "Hearing in Complex Environments: Auditory Gain Control, Attention, and Hearing Loss," *Frontiers in Neuroscience*, vol. 16, p. 799 787, 2022.

[186] J. Egeland, O. Lund, I. Kowalik-Gran, A. K. Aarlien, and G. B. W. Söderlund, "Effects of auditory white noise stimulation on sustained attention and response time variability," *Frontiers in Psychology*, vol. 14, p. 1 301 771, 2023.

[187] K. Trimmel, J. Schätzer, and M. Trimmel, "Acoustic Noise Alters Selective Attention Processes as Indicated by Direct Current (DC) Brain Potential Changes," *International Journal of Environmental Research and Public Health*, vol. 11, no. 10, pp. 9938–9953, 2014.

[188] C. K. Conners, *Conners' Continuous Performance Test II (CPT II) Computer Program for Windows, Technical Guide and Software Manual*, 2000.

[189] C. Riccio, C. Reynolds, and P. Lowe, *Clinical Applications of Continuous Performance Tests: Measuring Attention and Impulsive Responding in Children and Adults*. New York: Wiley, 2001, ISBN: 9780471216797.

[190] H. D. Critchley, "Review: Electrodermal Responses: What Happens in the Brain," *The Neuroscientist*, vol. 8, no. 2, pp. 132–142, 2002.

[191] National Highway Traffic Safety Administration (NHTSA), "Notice of proposed rulemaking - Minimum Sound Requirements for Hybrid and Electric Vehicles," Department of Transportation - National Highway Traffic Safety Administration, Tech. Rep. NHTSA-2011-0148, 2013.

[192] National Highway Traffic Safety Administration (NHTSA), "Minimum Sound Requirements for Hybrid and Electric Vehicles FMVSS No. 141: Final Regulatory Impact Analysis," National Highway Traffic Safety Administration, Tech. Rep., 2016.

[193] B. A. Nosek, C. R. Ebersole, A. C. DeHaven, and D. T. Mellor, "The preregistration revolution," *Proceedings of the National Academy of Sciences*, vol. 115, no. 11, pp. 2600–2606, 2018.

[194] M. Wessels, L. Müller, A. L. Maier, and J. Kraus, "Auditory localization and subjective assessment of autonomous cleaning robot sounds: A VR experiment on speed, operating mode and alerting signals," *Under review in Acta Acustica*, 2026.

[195] P. H.-s. Yahav, A. Sharaabi, and E. Z. Golumbic, "The effect of voice familiarity on attention to speech in a cocktail party scenario," *Cerebral Cortex*, vol. 34, no. 1, 2023.

Appended Papers

PAPER A

Auralization of Electric Vehicles for the Perceptual Evaluation of Acoustic Vehicle Alerting Systems

Leon Müller, Wolfgang Kropp

Published in *Acta Acustica*, 2024, 8, 27.

PAPER B

Loudspeaker Array-Based Auralization of Electric Vehicle Noise in Living Environments

Leon Müller, Jens Ahrens, Wolfgang Kropp

Published in *Proceedings of Forum Acusticum*, 11th Convention of the European Acoustics Association, Forum Acusticum 2025, Málaga, Spain.

PAPER C

On the Influence of AVAS Directivity on Electric Vehicle Speed Perception

Leon Müller, Wolfgang Kropp

Published in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, INTER-NOISE 2024, Nantes, France.

This paper is accessible at research.chalmers.se/en/publication/548371

PAPER D

Auditory Localization of Multiple Stationary Electric Vehicles

Leon Müller, Jens Forssén, Wolfgang Kropp

Published in *Journal of the Acoustical Society of America*, 157, 2025.

PAPER E

Effects of Low-Level Electric Vehicle Noise on Attention, Electrodermal Activity, Workload, and Annoyance

Leon Müller, Jens Forssén, Wolfgang Kropp

Published in *Journal of the Acoustical Society of America*, 159, 2026.

PAPER F

Traffic Noise at Moderate Levels Affects Cognitive Performance: Do Distance-Induced Temporal Changes Matter?

Leon Müller, Jens Forssén, Wolfgang Kropp

Published in *International Journal of Environmental Research and Public Health*, 20(5), 2023