CHALMERS
UNIVERSITY OF TECHNOLOGY



Auditory Size Perception in Acoustic Vehicle Alerting Systems (AVAS)

Virtual Pitch and Spatial Decorrelation for Heavy Vehicles under Electroacoustic and Environmental Constraints

Master's thesis in Sound and Vibration

Alve Ocklund

DEPARTMENT OF ARCHITECTURE AND CIVIL ENGINEERING

Chalmers University of Technology
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

Auditory Size Perception in Acoustic Vehicle Alerting Systems (AVAS)

Virtual Pitch and Spatial Decorrelation for Heavy Vehicles under Electroacoustic and Environmental Constraints

Alve Ocklund



CHALMERS
UNIVERSITY OF TECHNOLOGY

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

Auditory Size Perception in Acoustic Vehicle Alerting Systems (AVAS)
Virtual Pitch and Spatial Decorrelation for Heavy Vehicles under Electroacoustic and Environmental Constraints

© Alve Ocklund, 2026.

Supervisor Chalmers: Leon Müller, Division of Applied Acoustics.
Supervisor Volvo Trucks: Frédéric Wullens, Volvo Group Noise & Vibration Laboratory.
Examiner: Wolfgang Kropp, Division of Applied Acoustics

Master's Thesis 2026
Department of Architecture and Civil Engineering
Division of Applied Acoustics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Measurement setup to capture the in situ impulse reposes of the different AVAS speaker positions.

Typeset in LaTeX

Auditory Size Perception in Acoustic Vehicle Alerting Systems (AVAS):
Virtual Pitch and Spatial Decorrelation for Heavy Vehicles under Electroacoustic and Environmental Constraints

ALVE OCKLUND

Department of Architecture and Civil Engineering
Chalmers University of Technology

Abstract

The implementation of Acoustic Vehicle Alerting Systems (AVAS) has expanded as Electric Vehicles (EVs) become increasingly common on the roads. Due to their inherent lower acoustic radiation at low speeds, legislators have put regulatory demands, such as UN Regulation No. 138, on the acoustic radiation of these vehicles. These demands do not, however, differentiate between the type of vehicle, with the same requirements on passenger cars as for heavy electric trucks. This risks removing the auditory cues from the urban environment for differentiating size differences between vehicles, which may have a negative effect on the situational awareness of vulnerable road users and pedestrians, especially apparent for those visually impaired.

This thesis investigates auditory cues governing size perception to enhance the perceived size of heavy electric vehicles without increasing urban low-frequency noise pollution or exceeding the physical constraints of standard AVAS speakers. The results of two listening tests are presented: the first was conducted by 17 participants and showed that while spectral balance is the main cue for auditory size perception, spatial decorrelation and time structure also can influence the perceived auditory size of a signal. Further, participants showed a preference for signals containing tonal components over those composed of pure broadband noise.

The second test, conducted by 32 participants, applied the findings of the first, focusing on whether virtual pitch and spatial decorrelation could be used to increase the auditory size perception in an applied situation, convolving measured Impulse Responses (IRs) of a currently used AVAS speaker mounted at different positions on a representative truck. The results show that virtual pitch was highly effective at increasing the size perception of the signal, whilst spatial decorrelation was found not to be as effective due to the physical limitations of the implementation. Ultimately, utilizing the missing fundamental phenomenon to simulate virtual pitch allows for enhanced auditory size perception without the addition of low frequency components, effectively conveying vehicle mass without worsening urban noise pollution.

Keywords: Virtual Pitch, Spatial Decorrelation, Heavy Commercial Vehicles, AVAS, UN R138, Auditory Size Perception, Psychoacoustics.

Acknowledgements

I would like to express my sincere gratitude to my advisors, Leon Müller and Frédéric Wullens, for their guidance and support throughout the course of this thesis and whose work I am exited to continue to follow.

I also wish to extend my gratitude to the faculty at the Department of Applied Acoustics. In particular, I would like to thank my examiner, Wolfgang Kropp, whose inspiring teaching and charisma were primary motivations for my pursuit of this Master's program.

I extend my appreciation to Volvo Trucks and Chalmers University of Technology for enabling this research by providing access to their laboratories and measurement equipment.

A special thank you is owed to all the individuals who volunteered to participate in the listening tests. Your time and willingness to contribute were essential.

I am also deeply grateful to my family and friends who have throughout this process, some on a daily basis, endured being asked to evaluate the size of different (not seldom highly similar) sounds, and who have supported me through this process.

Finally, I would like to thank my peers in the Master's program who made the past two years a highly rewarding and enjoyable experience.

Alve Ocklund, Gothenburg, June 2026

List of Acronyms

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

AM	Amplitude Modulation
ASW	Apparent Source Width
AVAS	Acoustic Vehicle Alerting Systems
BPU	Binaural Playback Unit
BTL	Bradley-Terry-Luce
EV	Electric Vehicle
GPR	Glottal Pulse Rate
HEV	Hybrid Electric Vehicle
HRTF	Head Related Transfer Function
IACC	Inter-Aural Cross Correlation
ICEV	Internal Combustion Engine Vehicle
MF	Missing Fundamental
NHTSA	National Highway Traffic Safety Administration
PN	Pink Noise
UNECE	United Nations Economic Commission

Nomenclature

Below is the nomenclature of indices, sets, parameters, and variables that have been used throughout this thesis.

Indices

f	Frequency
t	Time
x, y	Evaluated stimuli (Bradley-Terry-Luce model)
ω	Angular frequency

Sets

$\hat{A}$	Set of available stimuli or alternatives (BTL model)

Parameters

α	Tilt factor for spectral balance
Δt	Interaural time delay
$\varepsilon(\omega)$	Frequency-dependent regularization factor
A	Cross-sectional area of the resonator's port
c	Speed of sound in the medium
f_{Nyq}	Nyquist frequency
K	Geometric scaling factor
l	Length of the resonator's port
T	Integration time limit for continuous IACF
t_1, t_2	Start and end times of the measurement integration window
V	Internal volume of the cavity

Variables and Functions

f_0	Resonance frequency
f_0'	Scaled resonance frequency
$H(f)$	Transfer function of the frequency domain tilt filter
$H(\omega)$	System transfer function
IACC	Inter-Aural Cross-Correlation
IACF(Δt)	Interaural Cross-Correlation Function
$P(x, \hat{A})$	Probability of choosing stimulus x from set $\hat{A}$
$p_l(t), p_r(t)$	Measured sound pressure signals at the left and right ear
$u(x)$	Scale value (or weight) of stimulus x on a ratio scale
$X(\omega)$	Input signal in the frequency domain
$\overline{X}(\omega)$	Complex conjugate of the input signal
$Y(\omega)$	Measured output signal in the frequency domain

Contents

List of Acronyms **ix**

Nomenclature **xi**

List of Figures **xv**

List of Tables **xvii**

1 Introduction **1**
1.1 Background . . . 1
1.2 Aim & Objective . . . 2
1.3 Research Questions . . . 2

2 Theory **3**
2.1 AVAS Regulations and UN R138 . . . 3
2.2 Physical Transducer Limitations . . . 4
2.3 The Psychoacoustics of Auditory Size Perception . . . 5
2.3.1 Pitch, Resonance Scale, and Spectral Balance . . . 5
2.3.2 The Lower Limit of Pitch and Virtual Pitch . . . 7
2.3.3 Temporal Dynamics: Amplitude Modulation . . . 8
2.3.4 Spatial Perception and Inter Aural Cross Correlation (IACC) . 9

3 Measurements **11**
3.1 Full system measurements . . . 12
3.1.1 Loudspeaker positions . . . 12
3.1.2 Receiver positions . . . 13
3.1.3 Directivity measurements . . . 15
3.2 Sweep Deconvolution . . . 17

4 Test I: Initial Listening Test **19**
4.1 Participants . . . 19
4.2 Equipment & Environment . . . 19
4.3 Procedure . . . 20
4.3.1 Bradley-Terry-Luce & Preference Tree Model . . . 20
4.4 Test Variables and Categorization . . . 21
4.4.1 Spatial Decorrelation . . . 22
4.4.2 Modulation . . . 22

4.4.3 Pitch & Modulation Shift . . . 23
4.4.4 Spectral Centroid and Time Structure . . . 24
4.4.5 Pitch Shift - AVAS . . . 26
4.5 Test I: Results . . . 28
4.5.1 Spectral Centroid and AVAS Preferences . . . 28
4.5.2 Temporal and Phase Characteristics . . . 29
4.5.3 Spatial and Tonal Preferences . . . 30

5 Test II: Spatial Configuration and Virtual Pitch Evaluation 31
5.1 Participants . . . 32
5.2 Stimuli Generation and Signal Processing . . . 32
5.2.1 Spatial Configurations & Proof of Concept . . . 36
5.3 Regulatory Calibration and Loudness Normalization . . . 40
5.4 Test II: Results . . . 42
5.4.1 Base Stimuli . . . 42
5.4.2 Influence of Amplitude Modulation . . . 43
5.4.3 Spatial Configurations . . . 44
5.4.4 AVAS Proof of Concept . . . 47

6 Discussion 49
6.1 Psychoacoustic Determinants of Perceived Source Size . . . 49
6.1.1 Frequency Content & Spectral Balance . . . 49
6.1.2 Tonal vs. Broadband Components . . . 50
6.1.3 Time Structure . . . 50
6.2 Synthesis & Efficacy of Virtual Pitch . . . 51
6.2.1 Missing Fundamental (MF) . . . 51
6.2.2 Stochastic vs. Deterministic MF . . . 54
6.3 Impact of Spatial Configuration and Directivity . . . 54
6.3.1 Interaural Cross Correlation (IACC) . . . 55
6.3.2 Speaker Position & Mounting . . . 56
6.4 Methodological Considerations and Limitations . . . 57
6.4.1 Spatial Configuration . . . 57
6.4.2 Initial Listening test IACC Category DSP Routing . . . 57
6.4.3 Ecological Validity of Headphone Auralization . . . 57

7 Conclusion 59

8 Appendix 1 61
8.1 Paired Comparison Matrices . . . 61
8.1.1 Test I . . . 61
8.1.2 Test II . . . 64

Bibliography 69

List of Figures

2.1 Interaction of GPR and resonance scale 6
2.2 Time structure of a 30 Hz tonal complex 7
2.3 Amplitude modulation characteristics of an ICE truck 8

3.1 Narrow-band frequency response of the heavy vehicle AVAS 12
3.2 Typical loudspeaker positions. 12
3.3 HRTF measurement setup 13
3.4 Magnitude spectra of the HRTFs at a 5 m distance 14
3.5 Coherence and SNR between input sweep and recorded output at a 5 m distance 14
3.6 Directivity measurement setup (isometric view) 15
3.7 Directivity measurement setup (top view) 16
3.8 Directivity patterns for the Center position AVAS 16
3.9 Directivity patterns for the Offset position AVAS 16

4.1 Isolated listening environment setup 19
4.2 Paired comparison interface (HEAD Acoustics SQala) 20
4.3 Time vs amplitude plots for Modulation stimuli 23
4.4 Spectrograms for Pitch & Modulation Shift stimuli 24
4.5 Spectra of the Spectral Centroid stimuli 26
4.6 1/3 octave band spectra during simulated acceleration 27
4.7 Spectrograms during simulated acceleration 27
4.8 Preference share (BTL): Spectral Centroid categories 28
4.9 Preference share (BTL): Pitch Shift AVAS 29
4.10 Preference share (BTL): Modulation and Pitch & Modulation Shift . 29
4.11 Preference share (BTL): Spectral Centroid - Time structure vs random phase 30
4.12 Preference share (PreTree): Spatial Decorrelation 30

5.1 Interface of ranked comparison utilized in Test II. 32
5.2 Synthesis of the 30 Hz harmonic complex (MF30) 33
5.3 Spectra of main base stimuli for Test II 34
5.4 Time structure of main base stimuli for Test II 34
5.5 Spectra of the phase-altered MF15 and zero-phase MF30 complexes . 35
5.6 Temporal waveforms of phase-manipulated missing fundamentals . . . 35
5.7 Frontal loudspeaker positions. 36
5.8 Spectrogram of unmodulated Pitch & Modulation Shift variants . . . 38

5.9 Spectrogram of modulated Pitch & Modulation Shift variants 38
5.10 Calibration process of unfiltered ICE truck recording 39
5.11 UN R138 calibration for reference pink noise 41
5.12 Preference share (BTL): Mono category 42
5.13 Average ranked comparison scores for Mono stimuli 43
5.14 Statistical significance of modulated vs unmodulated stimuli 43
5.15 Preference share (BTL): Pitch & Modulation Shift 44
5.16 Preference share (BTL): Stereo Correlated and Decorrelated 45
5.17 Average ranked comparison scores for spatial configurations 45
5.18 Preference share (BTL): Effect of spatial distribution across base stimuli 46
5.19 Preference share (BTL): Pedestrian Position 46
5.20 Preference share (BTL): Pedestrian Position (Secondary evaluation) . 47

6.1 Time structure comparison: MF30 vs phase-altered MF15 52
6.2 Spectral centroids vs time for raw base stimuli 53
6.3 Spectral centroids vs time for processed Mono stimuli 53
6.4 Calculated roughness for Mono category stimuli 54
6.5 Time structure comparison: PN vs MF stimuli 55
6.6 Frequency spectra and IACC of the Center speaker IR 56

List of Tables

2.1 Summary of overarching UN R138 SPL requirements 4
2.2 UN R138 third-octave band minimum thresholds 4

4.1 Overview of evaluation categories for Test I 21
4.2 Spatial Decorrelation stimuli descriptions and abbreviations 22
4.3 Modulation stimuli descriptions and abbreviations 23
4.4 Pitch & Modulation Shift stimuli descriptions and abbreviations . . . 24
4.5 Spectral Centroid stimuli descriptions and abbreviations 25
4.6 Pitch Shift AVAS stimuli descriptions and abbreviations 26

5.1 Overview of evaluation categories for Test II 33
5.2 Mono configurations for the center speaker position 36
5.3 Asymmetric stereo correlated configurations 37
5.4 Asymmetric stereo decorrelated configurations 37
5.5 Stimuli configurations evaluating pitch and AM shifts 37
5.6 Offset pedestrian position configurations 39
5.7 Proof-of-concept AVAS configurations 40

8.1 Result matrix: Spatial Decorrelation category 61
8.2 Result matrix: Modulation category 62
8.3 Result matrix: Pitch & Modulation Shift category 62
8.4 Result matrix: Pitch Shift AVAS category 62
8.5 Result matrix: Spectral Centroid category 63
8.6 Result matrix: Spectral Centroid - Random Phase category 63
8.7 Result matrix: Spectral Centroid category (Secondary evaluation) . . 63
8.8 Result matrix: Stereo Correlated category 64
8.9 Result matrix: Mono category 64
8.10 Result matrix: Stereo Decorrelated category 65
8.11 Result matrix: Spatial configuration of Pink Noise 65
8.12 Result matrix: Spatial configurations of Tone 65
8.13 Result matrix: Spatial configurations of 30 Hz Missing Fundamental . 66
8.14 Result matrix: Spatial configurations of Extracted MF 66
8.15 Result matrix: Pedestrian Position category 66
8.16 Result matrix: Pitch & Modulation Shift category (Test II) 67
8.17 Result matrix: AVAS Proof of Concept category 67

1
Introduction

1.1 Background

The global transition from Internal Combustion Engine Vehicles (ICEVs) to electric vehicles (EVs) in modern cities offers an opportunity to reduce urban noise pollution, which remains the second most significant environmental health risk after air pollution [1]. The lower noise pollution of EVs is particularly advantageous in low speed urban environments, an effect that is especially seen for heavy commercial vehicles. A study conducted by Glebe et al. investigated the acoustic impact of replacing diesel engine buses with electric variants and found a significant reduction in measured indoor traffic noise. The largest effect was noted in the low-frequency domain (defined as the range between 25 Hz and 200 Hz), yielding an average reduction of 10 dB for pass-by measurements [2]. This was also seen in the results published by Lan et al. where the A-weighted sound pressure levels measured for heavy ICE vehicles were significantly higher in the 1/3 octave bands between 31.5 Hz and 300 Hz compared to those of heavy EVs.

However, while EVs are considerably quieter at low driving speeds due to the absence of mechanical propulsion noise, this acoustic reduction introduces a safety hazard for pedestrians and other vulnerable road users who rely on auditory cues. A 2014 technical report published by the National Highway Traffic Safety Administration (NHTSA) showed that Hybrid Electric Vehicles (HEVs) were 1.66 times as likely to be involved in pedestrian incidents compared to conventional ICE vehicles during low-speed maneuvers [3]. At higher speeds, tire and aerodynamic noise dominate, rendering EVs and ICEVs acoustically comparable [1]. To mitigate this low-speed collision risk, international standards such as UN Regulation No. 138 mandate the implementation of Acoustic Vehicle Alerting Systems (AVAS). While necessary for pedestrian safety, the mandate to introduce artificial sound into the environment has sparked an ongoing debate. Critics express concern that poorly designed alerting sounds risk unnecessarily contributing to urban noise pollution [4] [5], effectively negating the low-speed noise reduction benefits of the EV transition [1].

Furthermore, current regulations dictate that AVAS signals can be arbitrary as long as they comply with minimum and maximum acoustic emission levels, and they also do not differentiate between vehicle categories. Because of this a critical dimension of acoustic information is lost in the transition to electric powertrains: the perception of vehicle size. Using identical AVAS signals for both passenger cars and

heavy commercial vehicles risks perceptual misjudgments by pedestrians, especially those who are visually impaired [6]. Therefore, determining how human hearing systems extract and interpret size information from complex acoustic stimuli is of importance. The design of a next-generation AVAS for heavy commercial vehicles must strike a balance between the need to convey the necessary impression of an approaching truck to guarantee safety, while minimizing the reintroduction of the low-frequency noise pollution that electrification inherently eliminates.

1.2 Aim & Objective

This thesis aims to investigate the possibility of increasing the size perception for heavy vehicle AVAS ensuring the signals convey an accurate sense of vehicle mass without violating regulatory limits or introducing unnecessary environmental noise pollution. The hypothesis is that this would help increase the situational awareness of the urban acoustic environment, especially for those who are visually impaired.

To fulfill this aim, the research is divided into the following objectives:

- **Literature research** - To identify the relevant parameters that influence the perceived source size, and investigate previous research.
- **Initial listening test** - To subjectively evaluate these parameters on a smaller scale to narrow down the focus.
- **Signal synthesis** - To design AVAS signals based on the initial findings and simulate an applied situation.
- **Final listening test** - To conduct a final subjective listening test with the synthesized signals in a simulated situation that accounts for the whole system.

1.3 Research Questions

To fulfill the overarching aims and objectives of this thesis, the investigation is structured to answer the following specific research questions:

RQ1: How can the perceived auditory size of an AVAS be increased without also increasing overall loudness or introducing energy at low frequencies?

RQ2: To what extent do associative acoustic attributes of heavy ICE vehicles, such as stationary and dynamic amplitude modulation, enhance the perception of size compared to purely deterministic signals?

RQ3: To what extent can principles of size perception derived from speech and musical instruments, mainly resonance scale and pitch, be translated to heavy vehicle warning sounds?

RQ4: Can these size enhancement strategies be implemented within the hardware constraints of current heavy commercial vehicles while maintaining compliance with UN R138?

2

Theory

This chapter handles the underlying theory and context relevant to auditory size perception in AVAS. It will first address regulatory demands put on an AVAS as well as physical limitations and then go on to explain what psychoacoustic concepts governs size perception based on previous research, as well as some more speculative parameters that might contribute to this perception.

2.1 AVAS Regulations and UN R138

The regulation governing the acoustic emissions of electrified vehicles in Europe is established by the United Nations Economic Commission (UNECE) under Regulation No. 138 [7]. To ensure pedestrian safety without contributing excessively to environmental noise pollution, the regulation dictates strict minimum and maximum acoustic thresholds for quiet road transport vehicles.

The regulatory demands placed on an AVAS consist of minimum overall A-weighted sound pressure levels, which scale depending on the vehicle's speed. Furthermore, the regulation requires a perceivable pitch shift of at least 0.8% per increased km/h between 5 km/h and 20 km/h to communicate acceleration. Table 2.1 illustrates a summary of the primary overall acoustic constraints.

Table 2.1: Summary of the overarching UN R138 requirements, detailing the minimum and maximum overall A-weighted sound pressure levels at varying operational speeds.

Requirement Type	Condition	Value / Limit	Notes
Operating Speed Range	Forward & Reverse	0–20 km/h	System must be active.
Sound Pressure (Min)	10 km/h (Forward)	≥ 50 dBA	Minimum overall level.
	20 km/h (Forward)	≥ 56 dBA	
	Reversing	≥ 47 dBA	
Sound Pressure (Max)	Forward Motion	≤ 75 dBA	Environmental noise cap.
Frequency Shift	5–20 km/h	≥ 0.8% / km/h	Must vary proportionally with speed.
Frequency Content	1/3 Octave Bands	Min. 2 bands	One band must be ≤ 1600 Hz.

In addition to overall broadband levels, UN R138 dictates spectral distribution requirements. The signal must possess sufficient energy in at least two one-third octave bands between 160 Hz and 5000 Hz. At least one of these bands must be situated below or within the 1600 Hz one. The specific minimum level requirements for these frequency bands are detailed in Table 2.2.

Table 2.2: Minimum A-weighted sound pressure levels required per one-third octave band according to UN R138. To achieve compliance, an AVAS signal must meet or exceed these thresholds in a minimum of two bands, of which at least one needs to be within or below 1600 Hz.

1/3 Octave Bands, Hz	160	200	250	315	400	500	630	800	1000	1250	1600	2000	2500	3150	4000	5000
10 km/h, dBA	45	44	43	44	45	45	46	46	46	46	44	42	39	36	34	31
20 km/h, dBA	50	49	48	49	50	50	51	51	51	51	49	47	44	41	39	36

2.2 Physical Transducer Limitations

Beyond regulatory constraints, the acoustic design of an AVAS is bound by the physical limitations of the hardware itself. Automotive applications are constrained by cost, weight, and available mounting space. Furthermore, because the AVAS module is mounted externally on the vehicle chassis, it must be highly durable to withstand harsh environmental conditions, moisture, and temperature fluctuations.

These physical and structural requirements result in transducers with a limited low-frequency response. As noted in a 2023 paper by Fortino et al. [8], the AVAS hardware needs to be taken in consideration when designing the warning signal. Their recommendation is to preform a sound radiation analysis on the full system to find frequency bands with favorable power-to-sound ratios and symmetrical radiation to reduce the energy consumption of the system.

2.3 The Psychoacoustics of Auditory Size Perception

The ability to discern the physical size of a source based on auditory cues is an innate biological trait [9], serving as a primary spatial and environmental sensor when a source is visually obscured. In the context of heavy commercial vehicles, understanding the mechanics that govern this perception may help ensuring the safety and situational awareness of vulnerable road users, particularly the visually impaired. However, a fundamental acoustic conflict exists: accurately conveying a large vehicle size requires prominence in the low-frequency region, which may negate the beneficial reduction of noise pollution at those frequencies inherit to heavy EVs. By making use of specific psychoacoustic phenomena, AVAS signals can be optimized to increase the perception of size without relying on these problematic frequencies. This section outlines the primary psychoacoustic mechanisms governing auditory size perception. It begins by examining the physical laws and spectral balances that link low-frequency prominence to size, followed by the temporal thresholds of pitch perception and the phenomenon of virtual pitch. Finally, it explores the influence of rhythmic amplitude modulation and spatial decorrelation on the perception of size.

2.3.1 Pitch, Resonance Scale, and Spectral Balance

The physical dimensions of a sound-producing object are inversely proportional to the frequencies of its radiated acoustic energy. This principle is governed by the general law of similarity, which states that if all linear dimensions of a resonator are scaled by a factor K, its natural vibrational modes decrease by a factor of $1/K$, provided the material properties remain constant [10]. As demonstrated by Fletcher and Rossing, this relationship is clearly illustrated by the resonance frequency of a Helmholtz resonator, defined as:

$$f_0 = \frac{c}{2\pi}\sqrt{\frac{A}{Vl}} \tag{2.1}$$

where c is the speed of sound in the medium, A is the cross-sectional area of the resonator's port (or neck), l is the length of the port, and V is the internal volume of the cavity.
Applying the geometric scaling factor K to all linear dimensions alters the cross-sectional area by K^2, the internal volume by K^3, and the port length by K. Substituting these scaled dimensions into the equation yields the new resonance frequency, f_0' [10]:

$$f_0' = \frac{c}{2\pi}\sqrt{\frac{K^2A}{(K^3V)(Kl)}} = \frac{c}{2\pi K}\sqrt{\frac{A}{Vl}} = \frac{1}{K}f_0 \tag{2.2}$$

Because of this, the human auditory system has evolved to utilize spectral balance, specifically, the distribution of high versus low-frequency energy, as a primary cue for estimating the physical size of an unseen source. A measure of this balance is the spectral centroid which indicate where the center of mass of the spectrum is located. This measure is also connected the pitch and "richness" of a sound, where a varying centroid is needed to enrich a signal, making it sound more lively [11].

In human vocalization, pitch is determined by the fundamental frequency of the glottal pulse rate (GPR), while the resonance scale is determined by the physical length of the vocal tract, acting as a filter applied to the GPR [9]. Figure 2.1 shows a simplified representation of relationship of GPR and resonance scale.

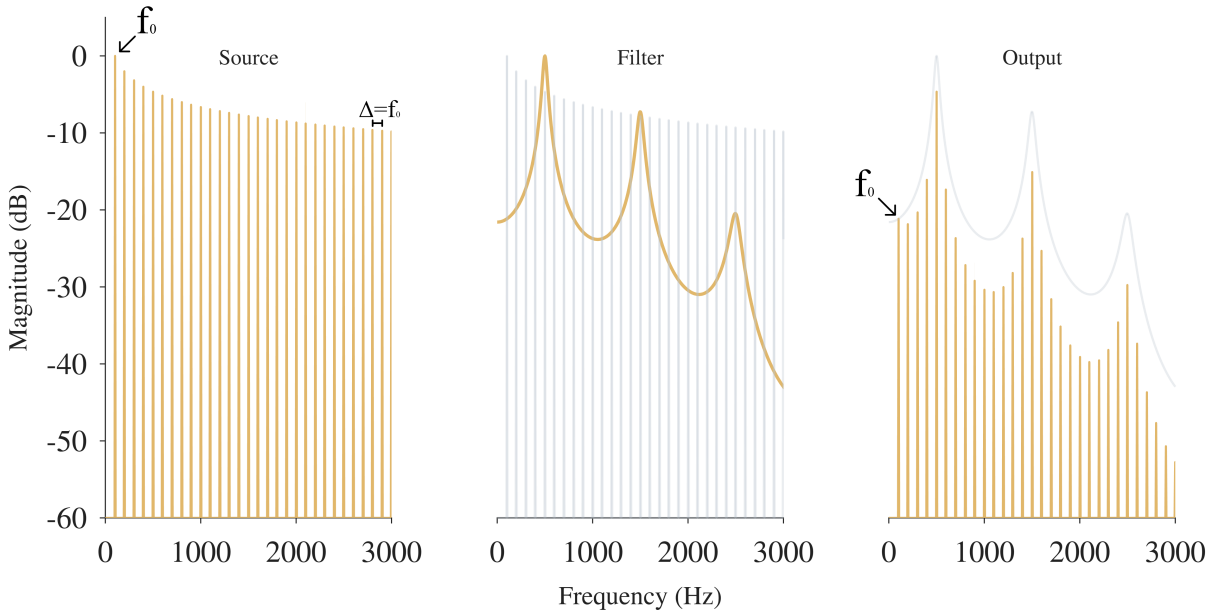


Figure 2.1: Simple representation of the interaction of GPR and resonance scale. Left most figure shows a simplified spectra of an hypothetical GPR harmonic complex, middle shows the filter of the vocal tract and right figure the combined output.

Pitch and resonance scale function as two of the main parameters the auditory system uses to determine the size of a source. This principle extends beyond human speech to both animal communication and musical instruments [12]. Vestergaard et al. demonstrated that the human auditory system extracts and processes these acoustic parameters to construct a representation of the sound source's physical size. Their research further found that this mechanism was active in newborn infants, indicating that to correlate low-frequency spectral dominance with physical size is an innate auditory function, evolved through selective pressure to aid in the acoustic detection of larger predators and potential suitors [9].

2.3.2 The Lower Limit of Pitch and Virtual Pitch

The pitch of a complex tone, defined as a combination of multiple simultaneous pure tones, has a perceived pitch determined by the frequency spacing of its spectral components [13]. If the time between two pulses is long enough the auditory system processes them as two separate events as opposed to one continuous one. This transition is not abrupt, instead it is gradual, and the perception can be described as going from fluctuation to roughness [13].

Krumbholz et al. tested pitch perception at different repetition rates and different numbers of harmonics, finding that for low frequencies (below 1000 Hz), the perception of pitch is limited by a temporal window rather than spectral resolution [14]. They concluded that the auditory system is fundamentally unable to process time intervals longer than approximately 33 ms, which corresponds to a lower limit of pitch at about 30 Hz. Another study by Pressnitzer et al. investigated the lower limit of melodic pitch, also identifying it to be approximately 30 Hz, provided there is energy in the stimulus below 800 Hz [15]. They further noted that while listeners still experience a pitch below the 30 Hz limit, the sensation is weak.

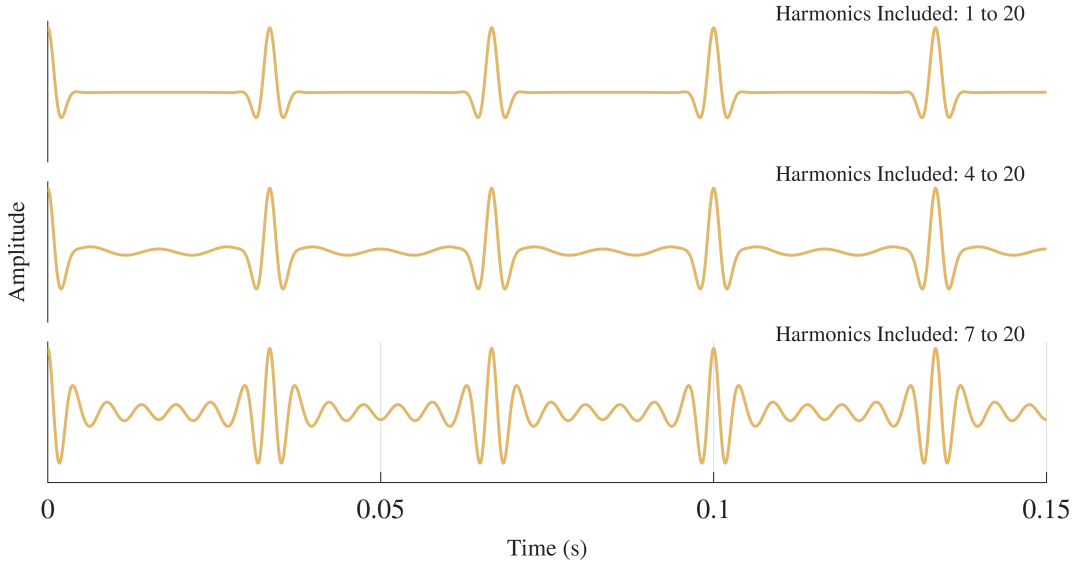


Figure 2.2: Time structure of a 30 Hz tonal complex with differing first harmonics.

Although the pitch of a complex tone is primarily determined by the fundamental frequency (f_0), a stimulus does not strictly need to contain physical energy at the fundamental frequency for the pitch to be perceived. The perceived pitch does not necessarily change if several of the lower harmonics are removed. This psychoacoustic phenomenon is known as residue pitch or virtual pitch [13]. This effect depends on the lowest frequency component of the stimulus, which varies with f_0. Fastl and Zwicker noted that for an f_0 of 50 Hz, the lowest harmonic should not be above 1000 Hz, meaning that components up to the 20th harmonic can still produce a virtual pitch [13].

Figure 2.2 illustrates three 30 Hz tonal complexes, each with a different number of harmonics. The top plot shows the signal containing the 30 Hz fundamental tone alongside the subsequent 20 harmonics. In the second plot, the first three harmonics are removed, and in the bottom plot, the first six harmonics are removed. As observed across the three plots, the temporal envelope maintains the same periodicity corresponding to the 30 Hz fundamental frequency, even when the lower harmonics are absent.

2.3.3 Temporal Dynamics: Amplitude Modulation

As established in previous sections, spectral content is highly important when judging the size of unseen objects. However, since merely lowering the spectral content of an AVAS will contribute to excessive noise pollution, and since the physical transducers currently used for AVAS playback are generally incapable of reproducing very low frequencies (as discussed in Section 2.2), this study also investigates manipulations of the signal's time structure to evaluate effects on size perception.

It is common for heavy ICE vehicles to produce a rhythmic amplitude modulation, an inherent characteristic of how their motors operate. Figure 2.3 displays the time structure of an ICE truck recording alongside the frequency spectra of the envelope of that same signal. The signal contains several modulation frequencies between 5 Hz and 40 Hz.

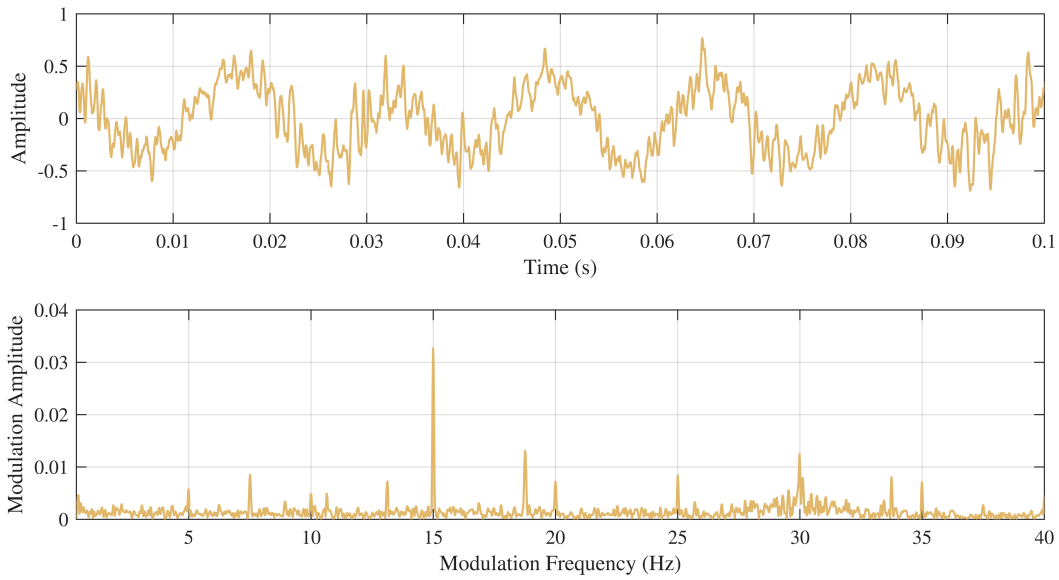


Figure 2.3: Amplitude modulation characteristics of an ICE truck at idle. (Top) Time-domain of the raw acoustic signal over a 0.1 s window. (Bottom) The corresponding analytic envelope calculated using a Hilbert filter with a length of 44.

2.3.4 Spatial Perception and Inter Aural Cross Correlation (IACC)

The horizontal width of a sound source is frequently described as the Apparent Source Width (ASW) [16]. Achieving a wide soundscape is considered desirable in concert halls, where an enveloping sound field typically serves as a measure of subjective hall quality [17].
A common measure of ASW is the Inter-Aural Cross-Correlation (IACC), which quantifies the correlation between the signals reaching the left and right ears. It is calculated using the cross-correlation function [18]:

$$\mathrm{IACC} = \max_{|\Delta t| \leq 1 \text{ ms}} |\mathrm{IACF}(\Delta t)| \tag{2.3}$$

where the Interaural Cross-Correlation Function (IACF) is defined as:

$$\mathrm{IACF}(\Delta t) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} p_l(t) p_r(t + \Delta t)\ dt \tag{2.4}$$

Here, p_l and p_r represent the measured sound pressure signals at the left and right ear, respectively, and Δt represents the interaural time delay. For a finite acoustic measurement, the normalized IACF is evaluated over a specific time window (t_1 to t_2). According to ISO 3382-1, this is calculated as:

$$\mathrm{IACF}_{t_1,t_2}(\Delta t) = \frac{\int_{t_1}^{t_2} p_l(t) \cdot p_r(t + \Delta t)\ dt}{\sqrt{\left(\int_{t_1}^{t_2} p_l^2(t)\ dt \int_{t_1}^{t_2} p_r^2(t)\ dt\right)}} \tag{2.5}$$

For the evaluation of ASW, the early IACC is typically calculated using an integration window from $t_1 = 0$ ms to $t_2 = 80$ ms.
Heavy ICE vehicles usually have large motors and chassis, which combine into a complex acoustic source. The size and complexity leads to a complex radiation which may result in a lower IACC compared to that of a relatively small AVAS speaker.
A study conducted by Potard et al. tested both the capacity of listeners to differentiate between various loudspeaker setups and their subjective perception of the resulting soundscapes, asking participants which configurations sounded more natural. They demonstrated that participants could easily discriminate between wide and narrow sources for broadband sounds (filtered white noise) in both the vertical and horizontal planes. Decorrelating the signals played through each source helped participants in assessing source dimensions. Furthermore, when participants rated a set of sounds (including that of a truck) based on perceived naturalness, a broad sound source was preferred 70% of the time over a narrow one [19]. This suggests that artificially increasing the ASW of an AVAS signal through signal decorrelation, might better align the spatial acoustic perception with the visual expectation of a large vehicle.

3

Measurements

This chapter outlines the measurements that were conducted to analyze the AVAS speaker used in the thesis. It also describes the acquisition of the Impulse Responses (IRs) that were used in the second listening test. It will first describe the speaker measurements conducted in a fully anechoic chamber with the speaker only, and then go on to explain the measurements done with the speaker mounted on a representative truck, conducted in a semi-anechoic chamber.

The frequency response function of the FIAMM AVAS loudspeaker used in this study was measured in an anechoic chamber to understand the physical limitations of the speaker. The measurement utilized a scientific free-field microphone placed 1.244 m in front of the loudspeaker with a HEAD acoustics SQuadriga III used for both the signal generation and the recording. A 15 second logarithmic sine sweep was used as the exciting signal.

The system evaluated in this thesis serves as a representative example of the hardware constraints discussed in Section 2.2. As illustrated in the narrow-band frequency response in Figure 3.1, the physical AVAS speaker displays a roll-off in radiated acoustic energy below approximately 300 Hz. The slope of this attenuation is approximately 18 dB per octave between 50 Hz and 300 Hz. This indicates that the system effectively act as a third-order high-pass filter with a cutoff frequency (f_c) close to 300 Hz.

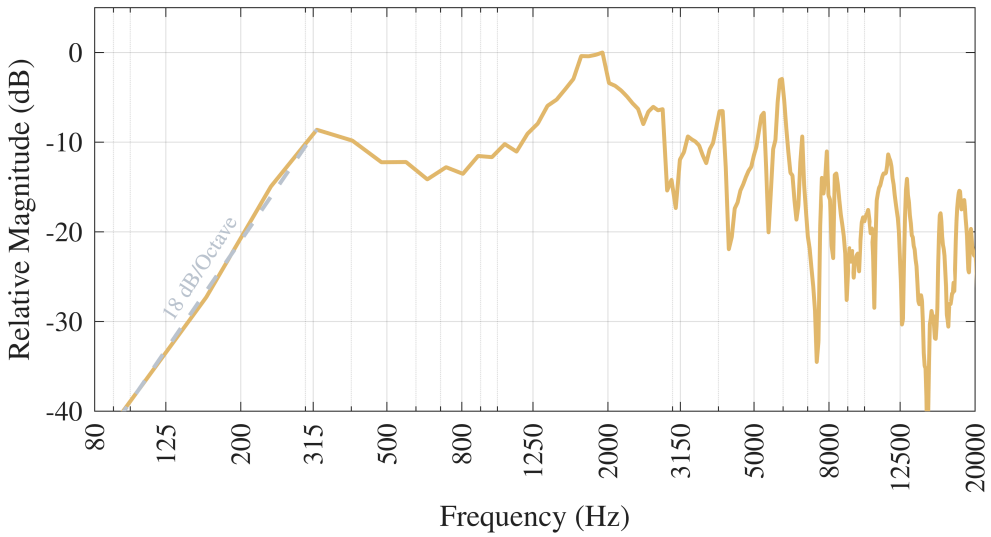


Figure 3.1: The normalized narrow-band frequency response of the heavy vehicle AVAS speaker used in this study. Measurements were conducted in an anechoic chamber using a free-field microphone, demonstrating ≈ 18 dB/octave drop-off in radiated SPL below 300 Hz.

3.1 Full system measurements

To be able to simulate a realistic playback for the planned finial listening test measurements were conducted in a semi-anechoic chamber with the currently used AVAS driver mounted at different position on a Volvo production truck. A sweep was then played and captured using an artificial head placed on the centerline 5 m horizontally facing the truck and with an ear height of 1.6 m. Additionally an array of 17 microphones placed in a half circle with a radius of 5 m was used to capture the directivity of each loudspeaker position.

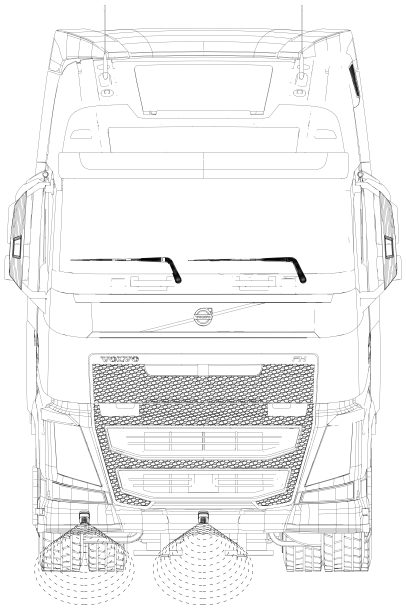


Figure 3.2: Typical loudspeaker positions.

3.1.1 Loudspeaker positions

Two loudspeaker positions where measured, these are listed below and seen in figure 3.2:

- **Center** - Loudspeaker mounted at the currently used location, on the center line of the truck behind the lower bumper.
- **Offset** - Loudspeaker mounted on a small stand 0.9 m lateral offset and ≈ 0.15 m vertically lower compared to the **Center** position.

3.1.2 Receiver positions

Two receiver positions were measured using an HEAD acoustics artificial head with ear height at 1.6 m relative to the ground. One of these were situated 5 m in front of the truck on the center line. The second position was measured at an approximate pedestrian position, 56.25°offset to the side, oriented 90° to the right. These are visualized in figure 3.3. The Head Related Transfer Function (HRTF) measurements were done for both loudspeaker positions. Additionally, a free-field microphone was used to capture the response at the same positions at two heights, 1.2 m and 1,7 m relative to ground level.

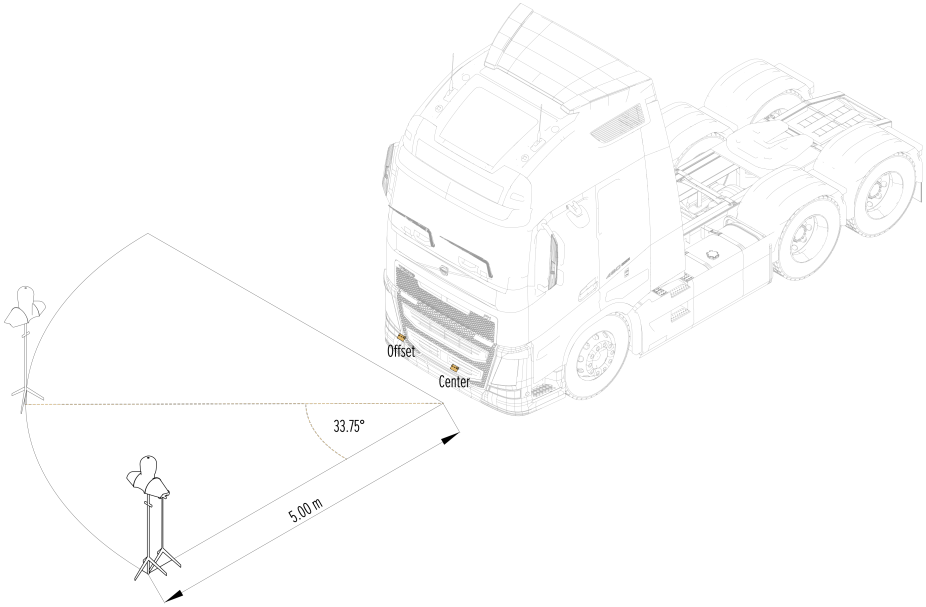


Figure 3.3: Isometric view of HRTF measurement setup with marked distance and angle to the pedestrian crosswalk position.

Figure 3.4 show the frequency response of the impulse response for the left and right channel gathered with the artificial head for both loudspeaker positions. The response of each position differ, especially in the range between about 200 Hz and 300 Hz, where the `Center` position show increased level at the low frequencies. The peak at approximately 300 Hz is also shifted between the two positions, where the `Center` position one is shifted to a slightly lower frequency ($\Delta f \approx 20$) Hz. This effect is also seen in the directivity measurement and most likely is due to mounting conditions shown in Figure **??**.

Figure 3.5 shows the coherence and SNR of the logarithmic input sweep and the recorded output at 5 m distance. It indicates that the differences in low frequency radiation seen between the different loudspeaker positions are a consequence of the physical setup and not necessarily due to the limited low frequency output of the speaker.

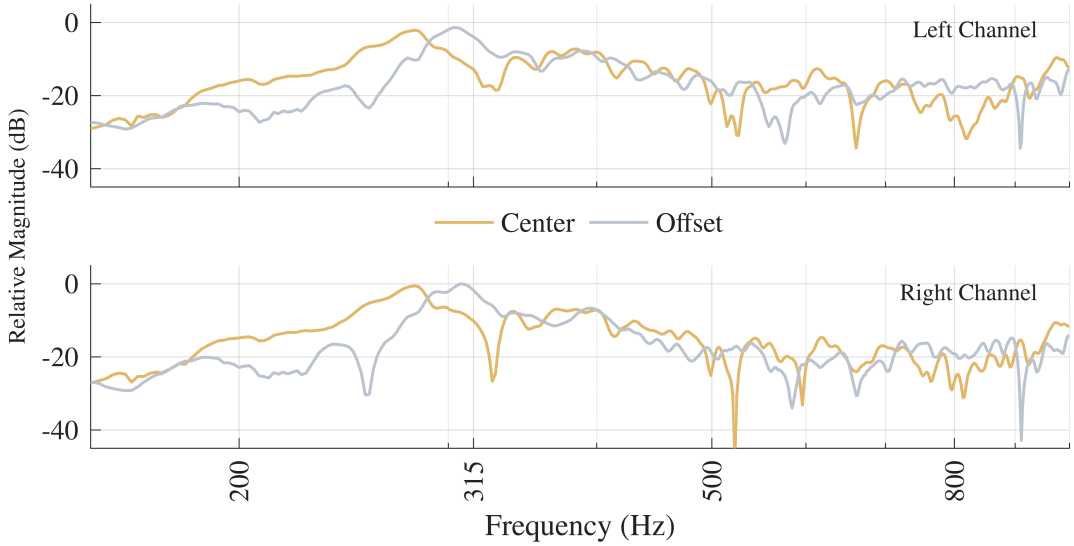


Figure 3.4: Magnitude spectra of the HRTFs captured at a 5 m distance for the two loudspeaker positions. Normalization is applied independently per ear, where the left and right channel responses are each normalized to their respective maximum amplitude (0 dB) across both positions.

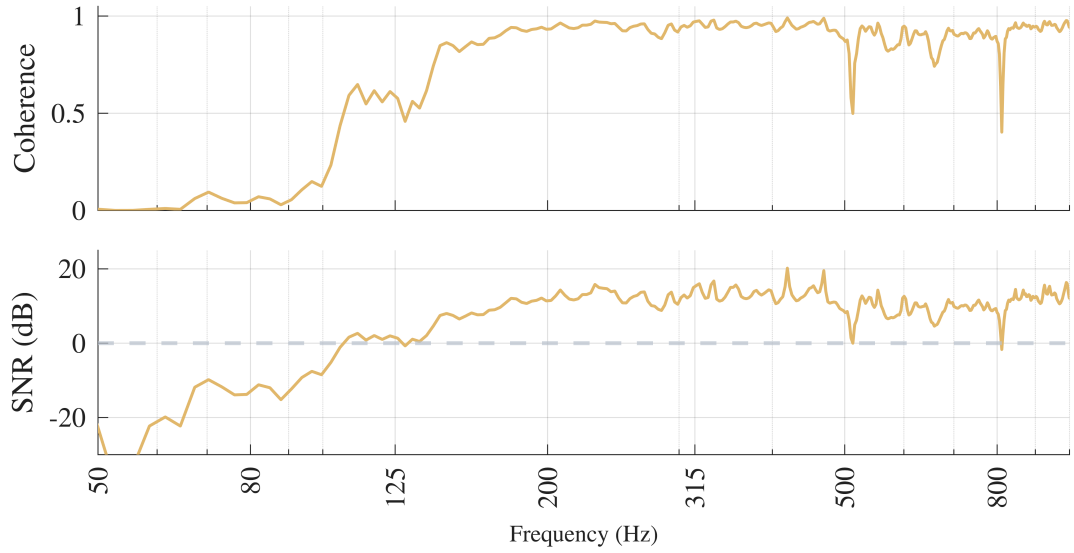


Figure 3.5: Coherence and SNR between the input sweep and recorded output at 5 m distance measured using a free-field microphone at 1.7 m height. Both measures are at an acceptable level from about 150 Hz.

3.1.3 Directivity measurements

Two directivity measurements where done for each loudspeaker position one at 1.2 m and one at 1.7 m relative to ground level. The measurement setup is shown in Figure 3.6 and 3.7 and the results are shown in Figures 3.8 and 3.9. As seen in these figures the directivity differs slightly between the plotted $\frac{1}{3}$-octave bands and speaker locations, but does not differ much between the two receiver heights.

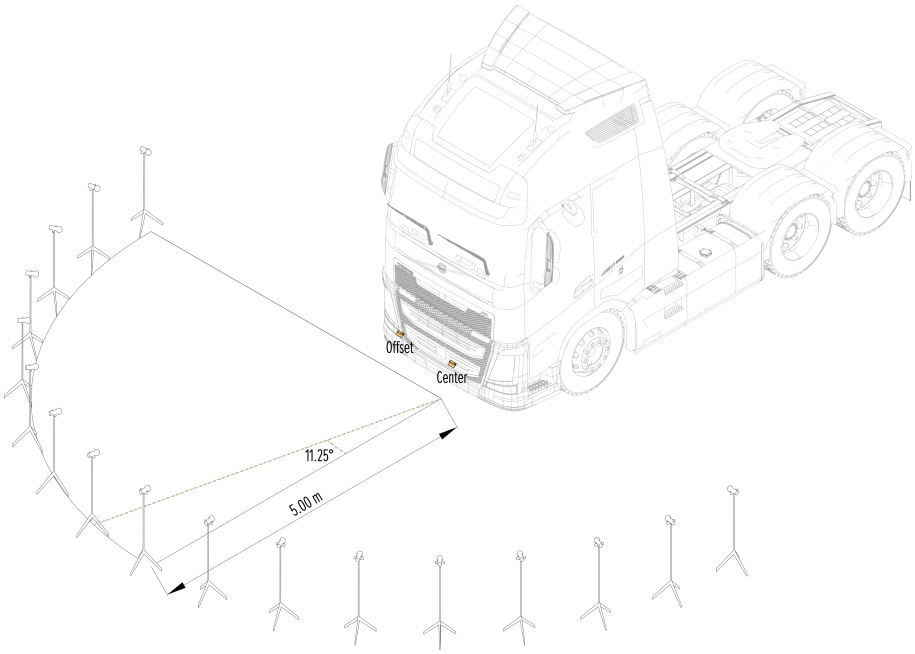


Figure 3.6: Isometric view of directivity measurement setup with marked distance to the center of the truck and angle between each microphones.

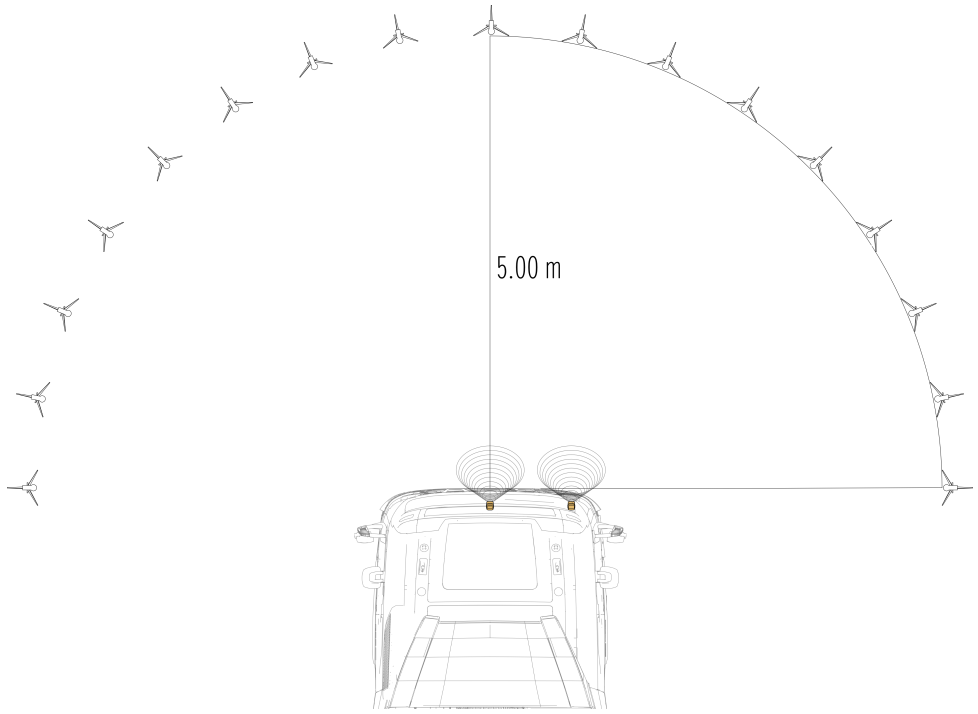


Figure 3.7: Top view of directivity setup with marked distances.

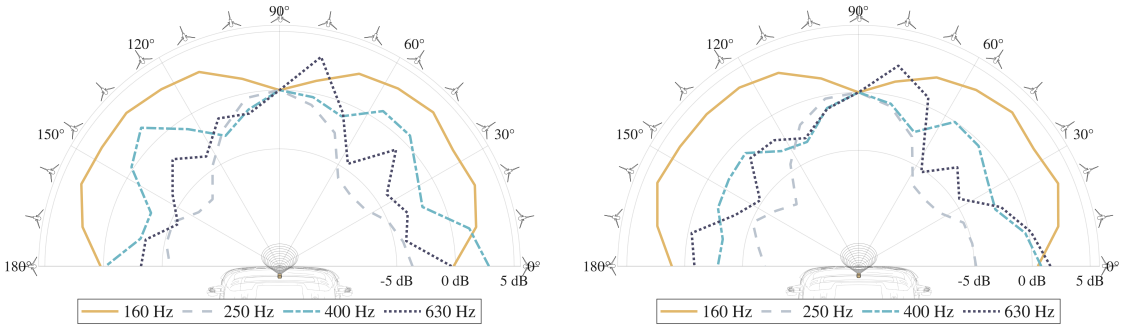


Figure 3.8: Directivity patterns for the Center position AVAS (1/3 octave bands). Left figure shows the directivity at 1.2 m height and right shows it at 1.7 m height. Amplitudes are normalized to the on axis microphone.

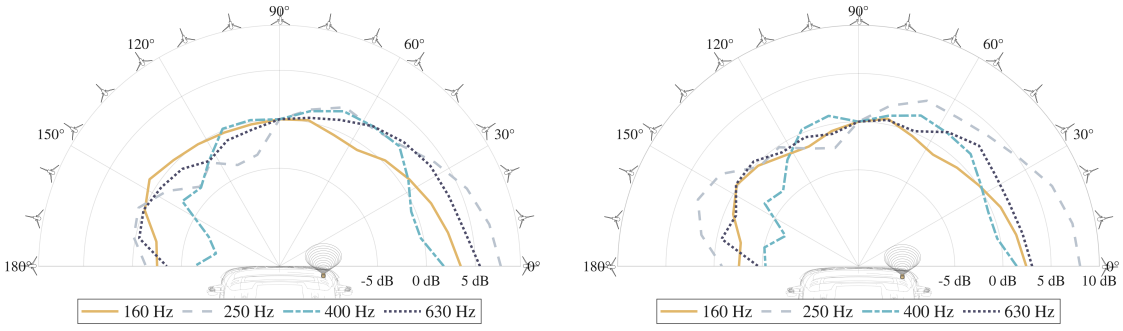


Figure 3.9: Directivity patterns for the Offset position AVAS (1/3 octave bands). Left figure shows the directivity at 1.2 m height and right shows the 1.7 m height. Amplitudes are normalized to the on axis microphone.

3.2 Sweep Deconvolution

As mentioned previously, logarithmic sweeps were used as exciting signals in the IR measurements. Since the goal is to find the transfer function, $H(\omega)$, and since both input and output were recorded one could rewrite the the simple frequency domain relation $Y(\omega) = H(\omega)X(\omega)$:

$$H(\omega) = \frac{Y(\omega)}{X(\omega)} \tag{3.1}$$

Where $Y(\omega)$ is the measured output and $X(\omega)$ is the input. Simply doing this can be unstable when the denominator (recorded input) contains low energy, which can cause background noise to be highly amplified.
One way of solving this is to use the method developed by Kirkeby [20]. By multiplying both the numerator and denominator by the complex conjugate of the input $\overline{X}(\omega)$ and introducing a regularization factor $\varepsilon(\omega)$ that has small values inside the sweep frequency range and large values outside of it this problem is solved:

$$H(\omega) = \frac{Y(\omega)\overline{X}(\omega)}{X(\omega)\overline{X}(\omega) + \varepsilon(\omega)} \tag{3.2}$$

Deconvolving the sweep also pushes nonlinearities produced by the speaker to negative time, which separates them from the linear behavior, making it easy to remove them.

4

Test I: Initial Listening Test

An initial listening test was conducted with 17 participants with the goal of finding parameters that effect the perceived auditory source size. The test was designed using paired comparison to analyze the effects of various parameters on size perception. The parameters were categorized into seven groups, and each sound within these groups was compared using a two-alternative forced-choice paired comparison. All sounds (except the ones in the binaural category) were convolved with binaural HRTFs simulating a source 0° in front of the listener.

4.1 Participants

The 17 participants age ranged between 24 and 60 years with an median age of 27. Seven of the participants had no prior experience conducting a listening test, 4 answered that they had seldom participated, 3 answered that they had participated several times and 3 answered that they had participated many times. To gauge familiarity with EV acoustics signals, participants were asked to rate their everyday awareness of EVs, 10 answered that they frequently observed EV sounds in their daily life, 4 answered occasionally and 3 answered that they rarely notice EV sounds signals. This indicates that all of the participants are fairly familiar with the existing AVAS on the roads.

4.2 Equipment & Environment

The test was conducted in a quiet, isolated environment where each participant was given a brief introduction before proceeding with the test individually. The digital signals were routed via USB to a HEAD acoustics Binaural Playback Unit (BPU) equipped with a 24-bit D/A converter. The audio was reproduced through calibrated, open dynamic HEAD acoustics HD IV.2 headphones. The frequency response of the playback system is specified to operate within -0.5 dB at 20 Hz and -1.5 dB at 20 kHz, ensuring accurate presentation of the stimuli.



Figure 4.1: Participant seated in the isolated listening environment.

4.3 Procedure

The paired comparison interface was designed and executed using HEAD acoustics ArtemiS SUITE (SQala). The test utilized forced choice paired comparisons, where participants were presented with two stimuli and asked to choose the one they perceived as being larger. Figure 4.2 shows the interface of the test. Each stimuli was presented in a random order.

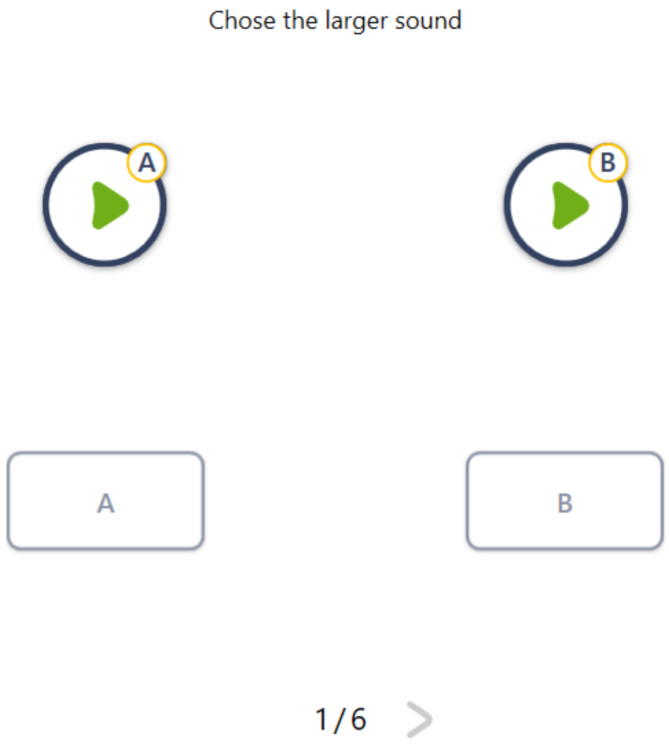


Figure 4.2: Paired comparison interface seen by the participant conducting the test - HEAD Acoustics SQala.

4.3.1 Bradley-Terry-Luce & Preference Tree Model

The results of the listening tests are presented using the Bradley-Terry-Luce (BTL) model, which aims to describe the probability of an outcome for paired comparisons, and was therefor used to process the results in this thesis. The authors that named the method proved that the probability of choosing a stimulus x from a set of alternatives A can be described with the equation [21]:

$$P(x, \hat{A}) = \frac{u(x)}{\sum_{y \in \hat{A}} u(y)} \tag{4.1}$$

where $u(x)$ represents the scale value (or weight) of the stimulus on a ratio scale. The BTL model assumes that each stimulus is unique, meaning the alternatives do not share attributes. However, if the evaluated stimuli are highly similar or share

attributes, this assumption fails.

To account for similarities between alternatives, different choice models have been developed. One of these is the Preference Tree (Pretree) model [21]. In a Pretree model, stimuli are organized in a way where multiple stimuli can be assigned a shared attribute in addition to their unique individual attributes, and these shared attributes are weighted individually.

One way of applying these theoretical models to collected data is via a numerical optimization algorithms, which can be easier to apply compared to analytical solutions. The results of the experiments only yield the observed choice frequencies of the participants. Because of this an optimization algorithm is used to work backward by iteratively adjusting the parameter weights (u) until it finds the specific combination that maximizes the likelihood of the observed experimental data occurring.

4.4 Test Variables and Categorization

To ensure that stimuli was being picked based on the parameter being manipulated all stimuli except for the stimuli in the **Spatial Decorrelation** category was convolved with 0° HRTFs taken from [22]. Further, to eliminate loudness as an influencing factor, each stimuli was normalized to a stationary loudness in accordance with ISO 532-1. The seven categories are briefly described in table 4.1

Table 4.1: Overview of what the first listening test categories tested.

Spatial Decorrelation	Time Structure Modifications	Spectral Modifications
Evaluated the effect of spatial configuration and IACC by comparing stimuli simulated as single- (mono) versus dual-speaker (stereo) sources.	Assessed the perceptual impact of amplitude modulation by applying varying modulation rates to a tonal signal.	Evaluated the perceptual impact of spectral centroid variations using three variations of an ICE truck recording, assessed under both their original temporal configurations and phase-randomized conditions.
	Examined the interaction between pitch shift (simulating vehicle acceleration) and dynamic rates of amplitude modulation (increasing with speed).	Tested three commercially implemented AVAS signals subjected to a uniform pitch shift against a band-passed recording of an accelerating ICE truck.
	Tested two of the original ICE truck recordings with their phase randomized counterparts to isolate the influence of temporal structure on size perception.	

4.4.1 Spatial Decorrelation

The spatial decorrelation category tested the effect of simulating one or several sources playing either pink noise (PN), tones, or both simultaneously. The sounds where convolved with HRTFs from [22]. The stimuli are described in table 4.2. The IACC was calculated to 0.95 for **ISD1** and 0.008 for the **ISD3**. The high decorrelation of **ISD3** was achieved by convolving two pink noise signals, each generated with different random number generation seeds, convolving both with the same HRTFs and then taking the left channel from one and the right channel from the other.

Table 4.2: Description of the Spatial Decorrelation stimuli and their corresponding abbreviations (Test I).

Stimuli Name	Test I: Spatial Decorrelation
ISD1	Mono point source in front of the listener playing pink noise (Reference PN).
ISD2	Mono point source in front of the listener playing a 150 Hz tone with three upper harmonics (300, 450, and 600 Hz) (Reference tone).
ISD3	De-correlated pink noise simulated as being played through two sources with an azimuth of 18.43° and −18.43° respectively.
ISD4	Reference tones simulated as being played through a point source with added synthetic early reflections.
ISD5	Combination of stimuli IA3 and IA4.

4.4.2 Modulation

To evaluate the influence of amplitude modulation on size perception a 250 Hz tone was synthesized at different rates of modulation. The rates were based on those found by analyzing the analytical envelope of several ICE truck recordings. The rates and stimuli names are displayed in table 4.3. The modulation depth of all signals were set to 0.8 based on subjective evaluation.

Table 4.3: Description of the Modulation stimuli and their corresponding abbreviations (Test I).

Stimuli Name	Test I: Modulation
IM1	Unmodulated reference tone.
IM2	0.6 Hz modulation.
IM3	8 Hz modulation.
IM4	16 Hz modulation.
IM5	30 Hz modulation.

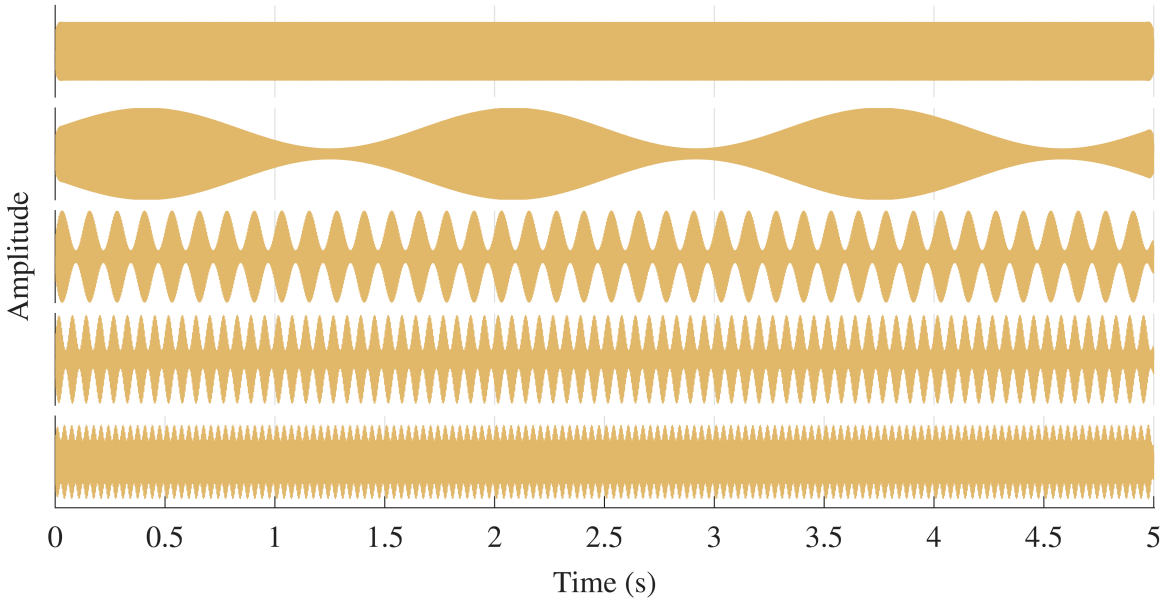


Figure 4.3: Time vs amplitude plots for each of the stimuli in the Modulation category. The rate of modulation increasing from top (unmodulated) to bottom (30 Hz modulation).

4.4.3 Pitch & Modulation Shift

This category tested the effect of adding dynamic modulation to the 0.8%/ km/h pitch shift demanded by regulation. The This effect mimics that of an ICE vehicles motor which naturally has these dynamic amplitude modulations due to the nature of the combustion engine, where it needs to increase the piston rates to deliver more power when accelerating. Each stimuli was composed of the same harmonic complex, with the fundamental frequency of 250 Hz and 3 upper harmonics, 500 Hz, 750 Hz and 1000 Hz and used the same AM depth (0.8) as the stationary AM category. Spectrograms of each stimulus are presented in Figure 4.4. These were computed using a 1024-point Hann window with 512 samples overlap and a 4096-point fast

Fourier transform. Spectral magnitudes are displayed in decibels relative to a unit reference dBFS.

Table 4.4: Description of the Pitch & Modulation Shift stimuli and their corresponding abbreviations (Test I).

Stimuli Name	Test I: Pitch & Modulation Shift
IPM1	Unmodulated reference pitch shift.
IPM2	0.1 Hz increased modulation per km/h.
IPM3	0.4 Hz increased modulation per km/h.
IPM4	0.8 Hz increased modulation per km/h.

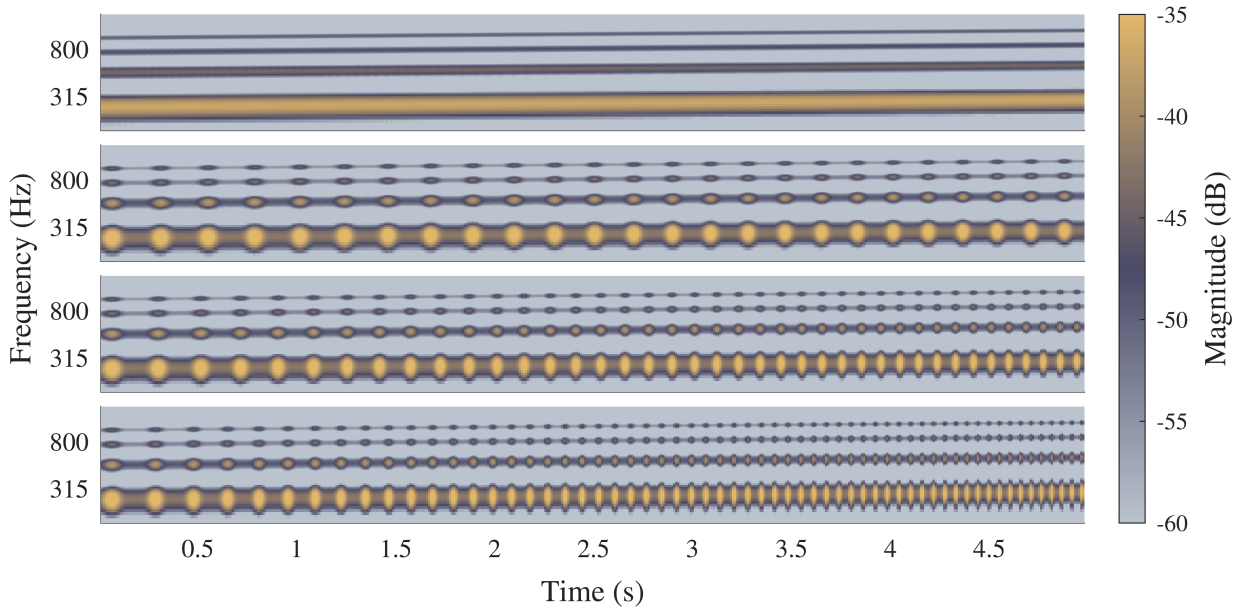


Figure 4.4: The spectrograms of each stimuli in the Pitch & Modulation Shift category. The top spectrogram represents the unmodulated IPM1 followed by IPM2, IMP3 and IMP4 in descending order. Magnitudes are represented in uncalibrated dBFS.

4.4.4 Spectral Centroid and Time Structure

This category was designed to investigate the influence of spectral balance versus temporal structure on perceived source size. The base stimulus was created from a recording of an idling ICE truck, high-pass filtered at $f_{\text{cutoff}} = 200$ Hz.

To test the effect of spectral balance, the spectral centroid of the base signal was shifted using a frequency domain tilt filter. The transfer function of the filter was

defined as:

$$H(f) = \left(\frac{f}{f_{\text{Nyq}}} \right)^{\alpha} \tag{4.2}$$

where f_{Nyq} is the Nyquist frequency and α is the tilt factor. Three spectral variants were created: a baseline with no shift ($\alpha = 0$, centroid at 2437 Hz), a low-frequency emphasis ($\alpha = -0.25$, centroid at 1853 Hz), and a high-frequency emphasis ($\alpha = 0.25$, centroid at 3264 Hz).

A second set of these three stimuli was generated by randomizing their phase spectrum. This process destroyed the original time structure while retaining an identical power spectrum.

The listening test evaluated these six stimuli across three sub-categories:

1. Comparisons between the three spectral centroids with intact time structures.
2. Comparisons between the three spectral centroids with randomized phases.
3. Comparisons between the intact and randomized-phase versions of the high- and low-centroid signals.

By comparing the stimuli within the categories the effect of the time structure could be compared to the effect of the spectral content. The stimuli and their corresponding abbreviations are summarized in Table 4.5, and their one-third octave band spectra (dBFS) are illustrated in Figure 4.5.

Table 4.5: Description of the Spectral Centroid stimuli and their corresponding abbreviations (Test I).

Stimuli Name	Test I: Spectral Centroid
ISC1	Centroid located at 2437 Hz.
ISC2	Centroid located at 1853 Hz.
ISC3	Centroid located at 3264 Hz.
ISC4	Centroid located at 2437 Hz - Random phase.
ISC5	Centroid located at 1853 Hz - Random phase.
ISC6	Centroid located at 3264 Hz - Random phase.

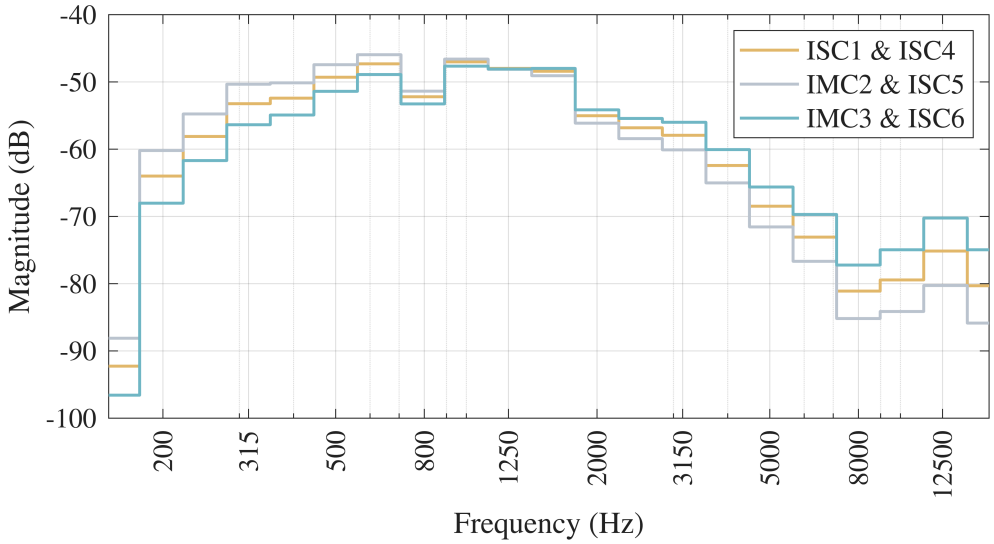


Figure 4.5: Shows the spectra of the Spectral Centroid stimuli. Note that each line represents two spectra.

4.4.5 Pitch Shift - AVAS

The final category evaluated three AVAS signals currently implemented commercially, one being from a passenger electric vehicle and two from heavy electric trucks, alongside a reference recording of an accelerating ICE truck. The stimuli simulated acceleration from 3 to 20 km/h. Each AVAS recordings was subjected to a dynamic pitch shift of 0.8%/km/h. Table 4.6 details the specific stimuli and their corresponding abbreviations.

One-third octave band plots of the four stimuli are seen in Figure 4.6 and their spectrograms are presented in Figure 4.7 which use the same settings as described in 4.4.3.

Table 4.6: Description of the Pitch Shift AVAS stimuli and their corresponding abbreviations (Test I).

Stimuli Name	Test I: Pitch Shift AVAS
IPSA1	AVAS of passenger car.
IPSA2	ICE truck acceleration.
IPSA3	AVAS of electric truck (B)
IPSA4	AVAS of electric truck (A)

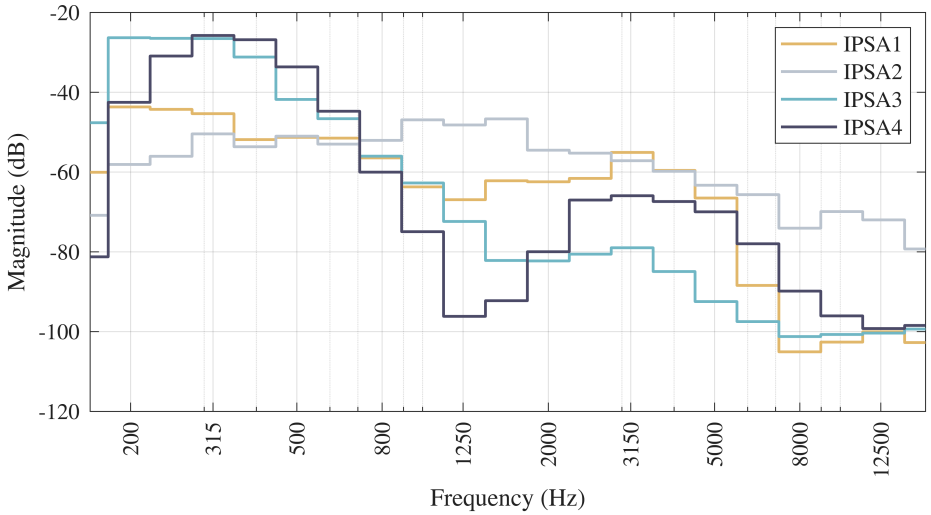


Figure 4.6: One-third octave band spectra of the four evaluated stimuli during a simulated acceleration from 3 to 20 km/h. Magnitudes are represented in uncalibrated dBFS.

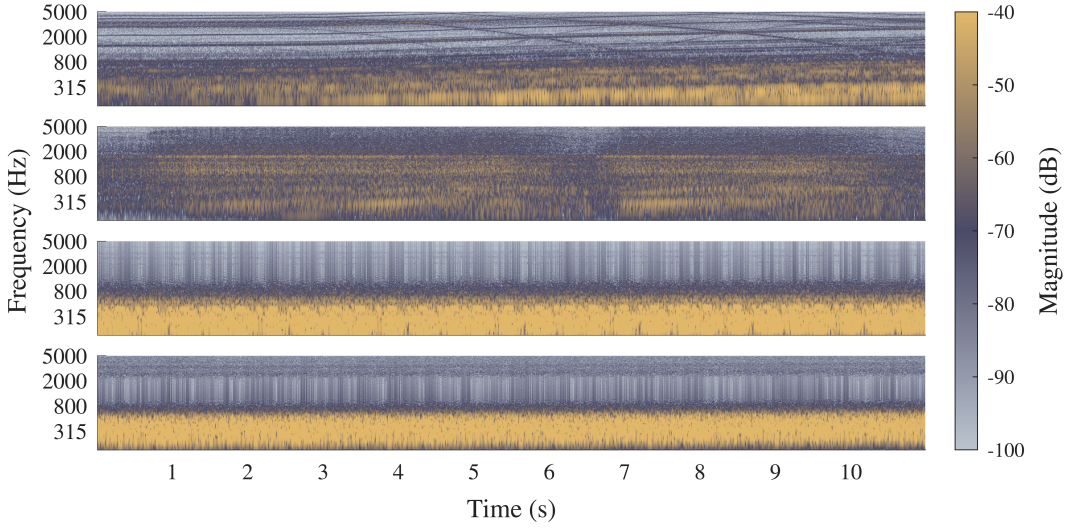


Figure 4.7: Spectrograms of the four evaluated stimuli during a simulated acceleration from 3 to 20 km/h. Magnitudes are represented in uncalibrated dBFS.

4.5 Test I: Results

The results of the initial listening test are presented in this section. Bradley-Terry-Luce (BTL) and the Pretree model was used, applying the MATLAB function published by Wickelmaier and Schmid [21]. The methods provide a share of preference model for each category. Goodness-of-fit was calculated for each model, where a model is rejected if $p < 0.1$, meaning that there is a less than 10% chance of the deviations between the model and the recorded results are due to random noise in the population. Because the model outputs a ratio scale value for each parameter, we can arbitrarily scale all stimuli so that the smallest prediction value equals 1. Consequently, the values of the other stimuli represent their relative preference compared to that baseline. For example, if three stimuli have the scaled values of 1, 2, and 4, the model predicts that in a direct paired comparison, the second stimulus is twice as likely to be chosen as the first. Similarly, the third stimulus is four times as likely to be chosen over the first, and twice as likely to be chosen over the second. The **Spatial Decorrelation** category used an additional categorization separating stimuli containing pink noise and stimuli not containing pink noise using the Pretree model. This was done since the BTL model did not yeld a model within the needed $p > 0.1$. The complete paired-comparison matrices for all categories in Test II, which show win rates and statistical significance, are provided in Appendix 8.

4.5.1 Spectral Centroid and AVAS Preferences

The results of the two categories comparing the idling truck signal with shifted spectral centroids display the same pattern, with a clear preference for the lower ones as indicated by the BTL model share of preference seen in Figure 4.8. All comparisons in these categories are within statistical significance as seen in the paired comparison matrices in Tables 8.6 and 8.7. The importance of spectral balance is further indicated by the results of the **Pitch Shift AVAS** category, where the BTL model results seen in Figure 4.9 indicate that AVAS (B) is preferred over the other 3 stimuli with statistical significance.

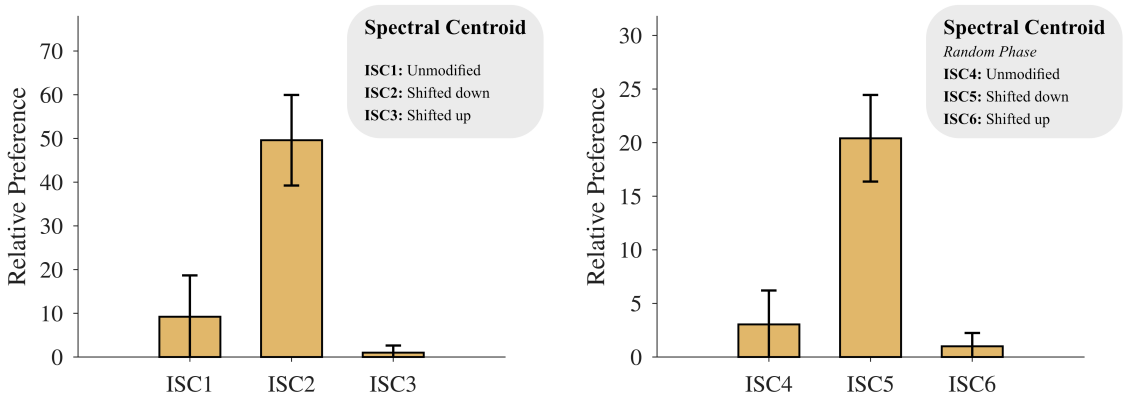


Figure 4.8: Share of preference for the initial listening test stimuli categories Spectral Centroid (left) and Spectral Centroid - Random Phase (right). Error bars represent 95% confidence intervals. Each category is scaled so the lowest ranked stimulus equals 1.

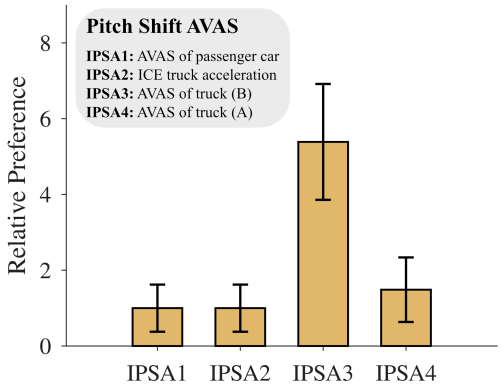


Figure 4.9: Share of preference for the initial listening test stimuli category Pitch Shift AVAS. Error bars represent 95% confidence intervals. Each category is scaled so the lowest ranked stimulus equals 1.

4.5.2 Temporal and Phase Characteristics

Amplitude modulation yielded different results depending on if the modulation was dynamic or not. The category comparing stationary rates of modulations show a very even result, only having a statistically significant result for one comparison (between the 8 Hz AM and unmodulated stimuli). The dynamic modulation category show a clear preference for the low rate of increasing modulation, both compared to the higher rates of modulation and compared to the unmodulated baseline. Both categories display a preference for low to medium rates of modulation, where both unmodulated and high rates of modulation decrease is less preferred. The effect is higher for the dynamic category. The BTL model estimations for the two categories are shown in Figure 4.10.

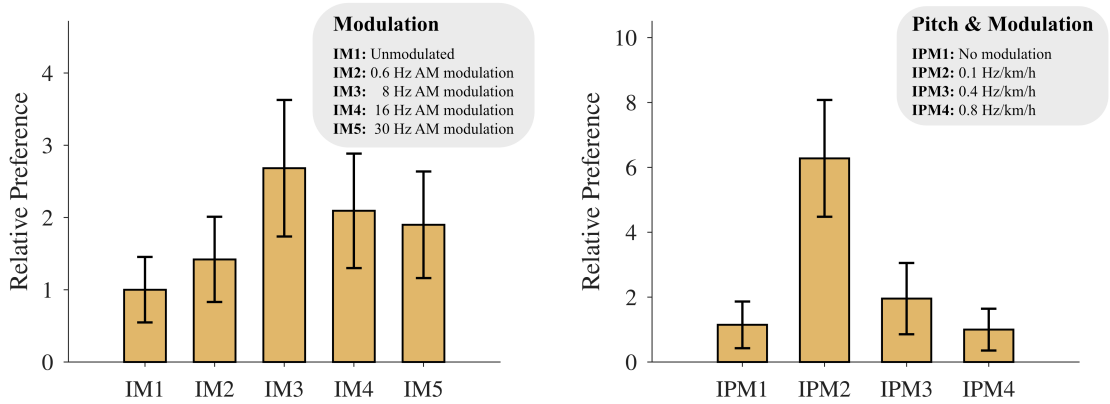


Figure 4.10: Share of preference for the initial listening test stimuli categories Modulation (left) and Pitch & Modulation Shift (right). Error bars represent the 95% confidence intervals. Each category is scaled so the lowest ranked stimulus equals 1.

The comparisons of the spectral centroids with intact time structure and their random phase counterparts show a preference for the signals with intact time structure, but this effect was more significant for the stimuli with higher frequency centroids. Figure 4.11 shows the results of the BTL model for this category.

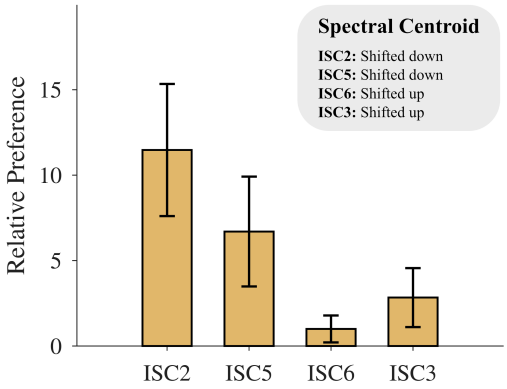


Figure 4.11: Share of preference for the initial listening test stimuli category Spectral Centroid - Time structure vs random phase. Error bars represent the 95% confidence intervals. Each category is scaled so the lowest ranked stimulus equals 1.

4.5.3 Spatial and Tonal Preferences

The category testing the effect of spatial decorrelation displayed a preference for stimuli containing tonal components over those composed of purely broad band noise. Decorrelated signals were also preferred over the correlated counterparts. Note that pretree model was applied here because the BTL model was rejected due to bad model fit. Figure 4.12 displays the pretree model prediction for this category.

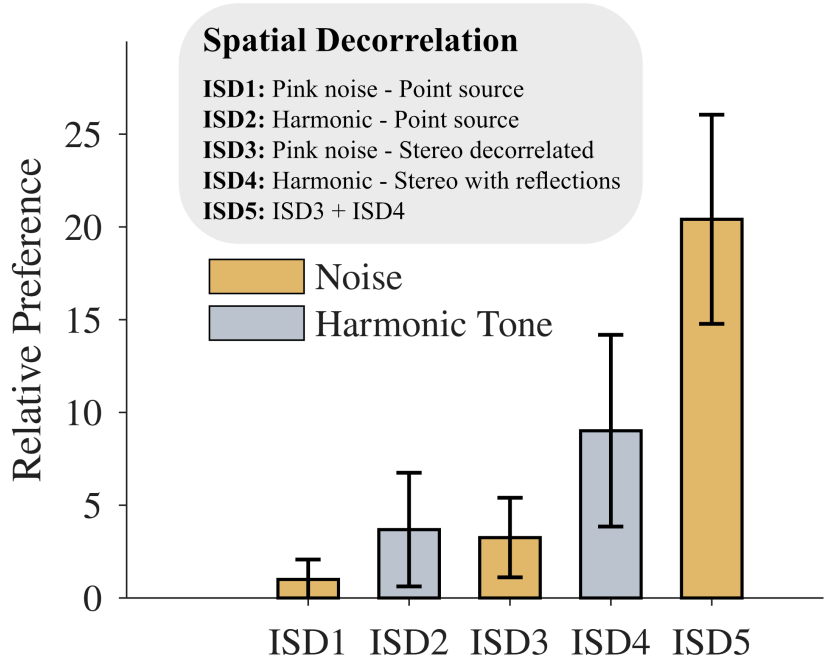


Figure 4.12: Share of preference for the initial listening test stimuli category Spatial Decorrelation. Error bars represent the 95% confidence intervals. Each category is scaled so the lowest ranked stimulus equals 1.

5

Test II: Spatial Configuration and Virtual Pitch Evaluation

A second paired-comparison listening test was conducted to evaluate the perceptual effects of spatial configuration and virtual pitch on the perceived size of heavy vehicles. The primary objective was to simulate a realistic scenario by convolving synthesized stimuli with measured HRTFs from the two AVAS speaker positions mounted on the truck. All signals were calibrated to comply with the minimum sound level requirements of UN Regulation No. 138. The experimental setup used the same equipment detailed in Section 4.2. The main hypotheses investigated were whether perceived source size could be increased by introducing an offset AVAS speaker playing either correlated or decorrelated versions of the stimuli and whether the missing fundamental phenomenon could successfully simulate acoustic mass by lowering the perceived pitch without requiring frequencies bellow 200 Hz to be present.

The test was designed in the same way as described in Section 4.3, mainly using the paired comparison method, but in addition to this two ranked comparisons was also included at the end of the test, where the participant was asked to rank a set of stimuli from small to large on a 0 to 10 scale. Figure 5.1 shows the interface for one of the ranked comparisons. The first of these included eight of the stimuli used in the paired comparison convolved with the single (Mono) center speaker position. The second compared spatial configurations of three of the base stimuli to get a scale of the perceived difference between these. The stimuli of these categories was presented in a random order.

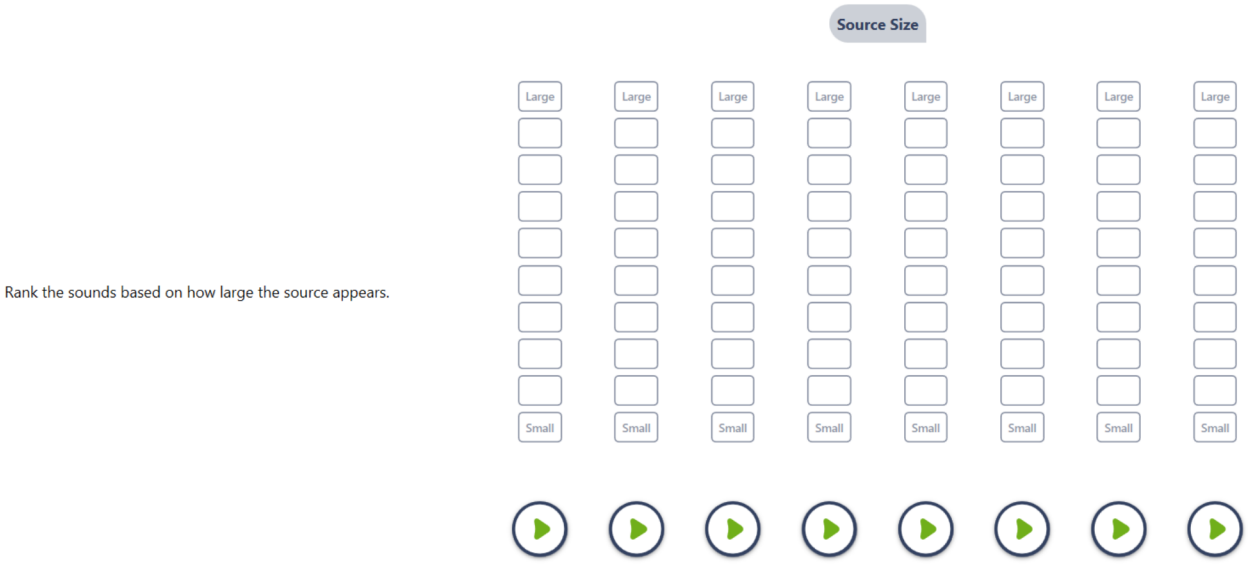


Figure 5.1: Interface of ranked comparison utilized in **Test II**.

5.1 Participants

A total of 32 participants conducted the listening test. Three participants were excluded from the final analysis due to self-reported hearing impairments or tinnitus, This gave a final sample size of $N = 29$. The median age of the participants was 32 and the range was 24 to 82 years of age. Ten participants had no prior experience with formal listening tests. All participants reported prior exposure to electric vehicle warning sounds in their everyday environment, and 12 participants had previously attended the initial listening test detailed in Section 4.

5.2 Stimuli Generation and Signal Processing

Based on findings from the initial listening test and existing literature, four primary signals were synthesized. All signals had similar spectral energy distributions between 200 Hz and 600 Hz. This lower limit was selected to ensure loudspeaker radiation efficiency and to mitigate low frequency environmental noise pollution, as discussed in Chapter 2. Table 5.1 shows an overview of the categories.
The baseline stimulus (**PN**) consisted of band passed pink noise with cut-off frequencies at 210 Hz and 600 Hz. The second signal (**Tone**) served as a regulatory compliant baseline, comprising two tonal components (210 Hz and 420 Hz) layered over the reference pink noise. The two tones were selected to comply with the UN R138 requirement of having a minimum of two one-third octave bands within the specified limits.
The third signal (**MF30**) was a complex tonal stimulus designed to exploit the missing fundamental phenomenon. This was achieved by synthesizing the 7th to the 20th harmonics of a 30 Hz fundamental frequency, resulting in 30 Hz-spaced

Table 5.1: Overview of the test categories of Test II.

Spatial Configurations	Time Structure Modifications	Positional Validity & Proof of Concept
The base signals were tested against each other with the three different spatial configurations.	Each base signal was compared to an 8 Hz AM variant of itself.	The two different missing fundamental methods were applied to the AVAS (A) as well as a combination of both of them. These were also compared to versions of an ICE recording run through the same DSP pipeline.
The three spatial configurations of each base signal was compared to each other.	Dynamic AM & Pitch shift was applied to the two **MF** as well as AVAS (A) and compared to variants without AM.	The different Spatial configurations of **MF30** was compared at a simulated crosswalk position.
Ranked copmparison of the different spatial configurations of the base signals (excluding **Tone**).	A phase altered missing fundamental variant, **MF15**, was compared to the **MF30** stilumus.	

tones spanning the 210 Hz to 600 Hz range. To mitigate edge tones and shift the spectral energy downwards, the complex tone was processed using a Gaussian filter (center frequency: 200 Hz, width: 100 Hz) before being layered onto the reference pink noise. The synthesis is visually represented in Figure 5.2. This generated a stimulus with a distinctly lowpitched, rough component.

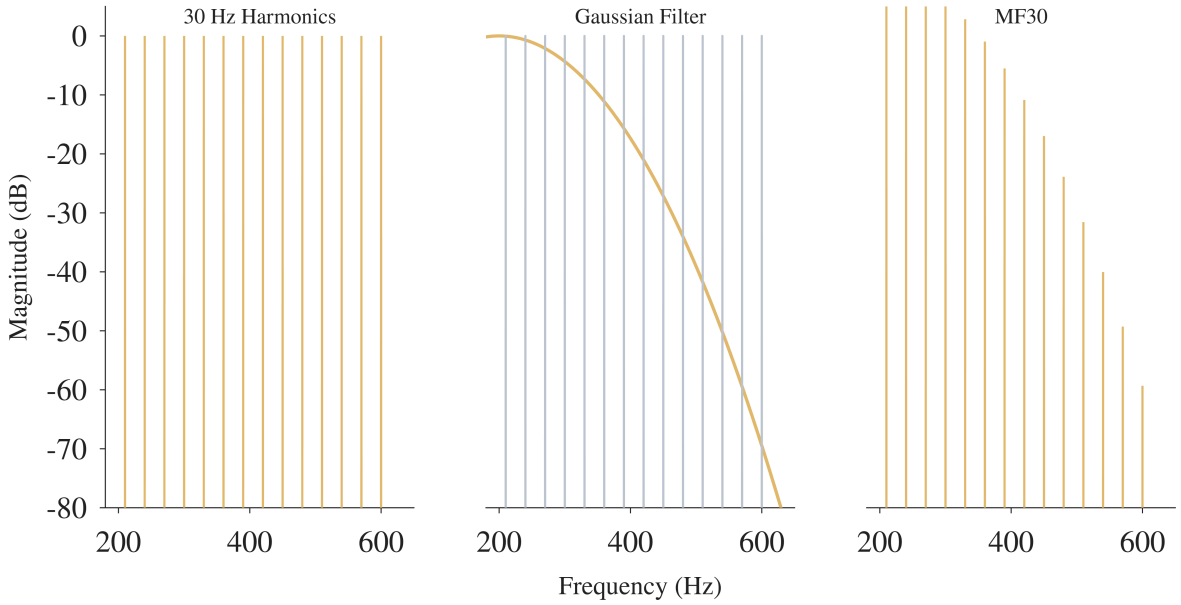


Figure 5.2: Visual representation of the synthesis of the 30 Hz harmonic complex utilizing the missing fundamental phenomenon (**MF30**) stimulus.

The final signal (**Ext**) was a stochastic missing fundamental variant created by extracting 5 Hz-wide frequency bands from pink noise, centered at the harmonics used in the **MF30** signal. Based on subjective pre-evaluations, an exponential filter was applied to this signal rather than a Gaussian filter. The reference pink noise was excluded from this stimuli to prevent auditory masking of the extracted signal. Figure 5.3 show the spectra of each base stimuli and Figure 5.4 shows their time structure, before calibration, convolution and normalization.

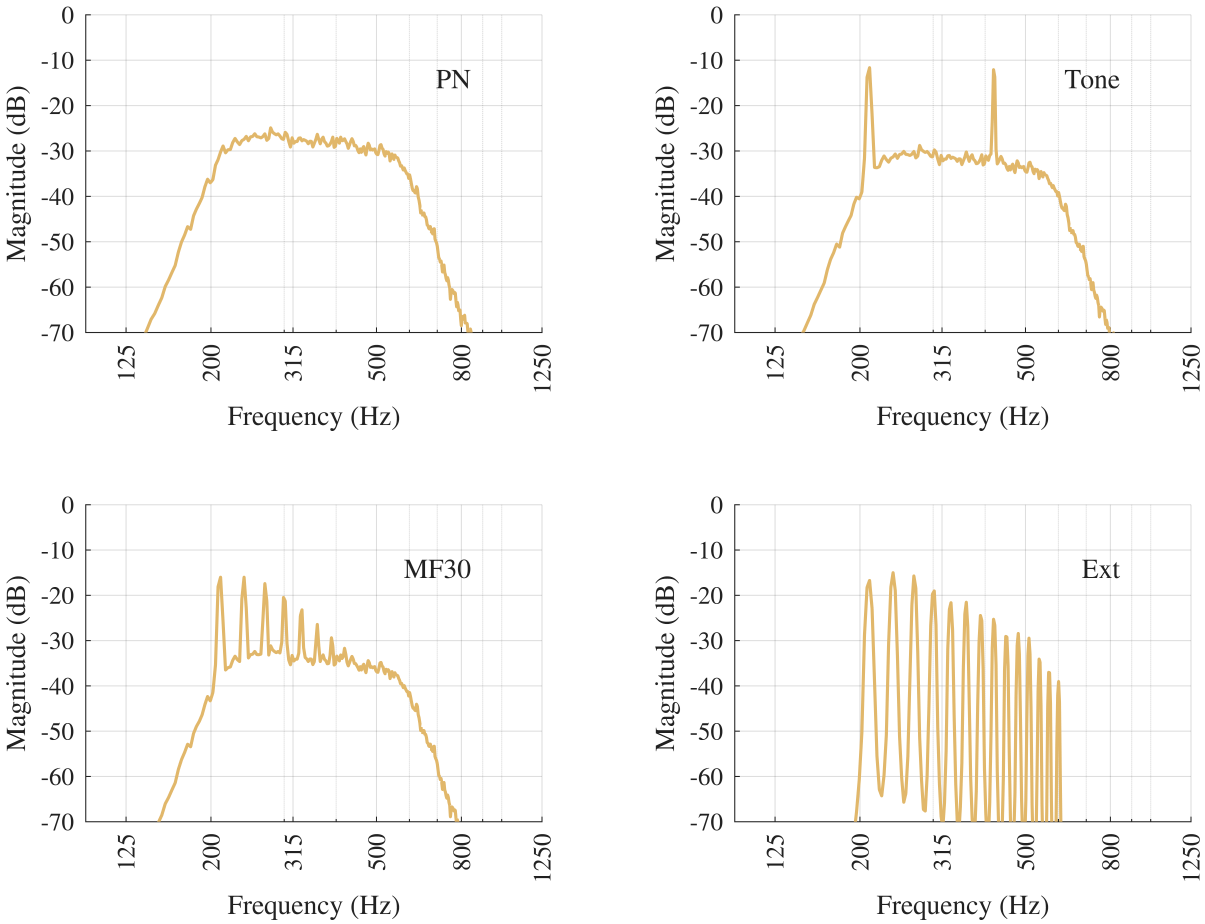


Figure 5.3: Spectra of main base stimuli for **Test II**, before calibration, convolution and normalization. Abbreviations of each of the stimuli is seen in the top right of the plot.

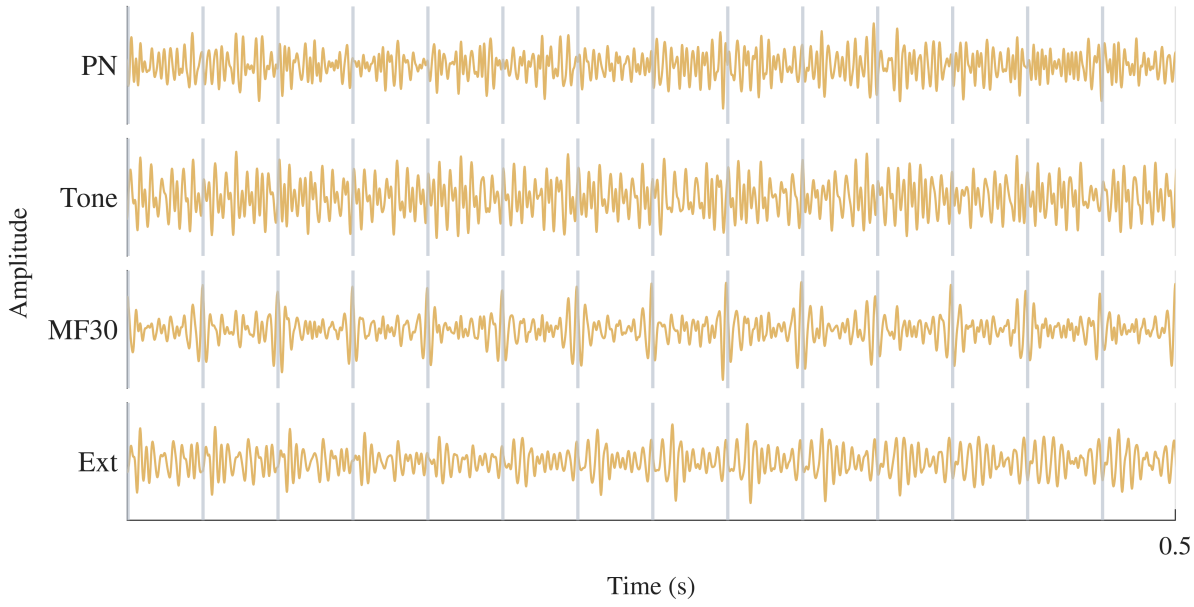


Figure 5.4: Time structure of main base stimuli for **Test II**, before calibration, convolution and normalization. Gray vertical lines are spaced with 33 ms separation (limit of pitch perception).

A second version of the missing fundamental complex was created and included in the listening test denoted as **MF15**. This variant was composed of the 14th to 40th harmonics of a 15 Hz fundamental and utilized the same Gaussian amplitude envelope as the **MF30** signal, but every odd harmonic was phase shifted by $\pi/2$ radians, which reduced the repetition rate from 66.7 ms to the needed 33 ms limit of pitch perception[14]. Figure 5.5 shows the spectra of the raw phase altered stimuli. The temporal structure shift is shown in Figure 5.6 which shows the unaltered and phase altered variants of the harmonic complex.

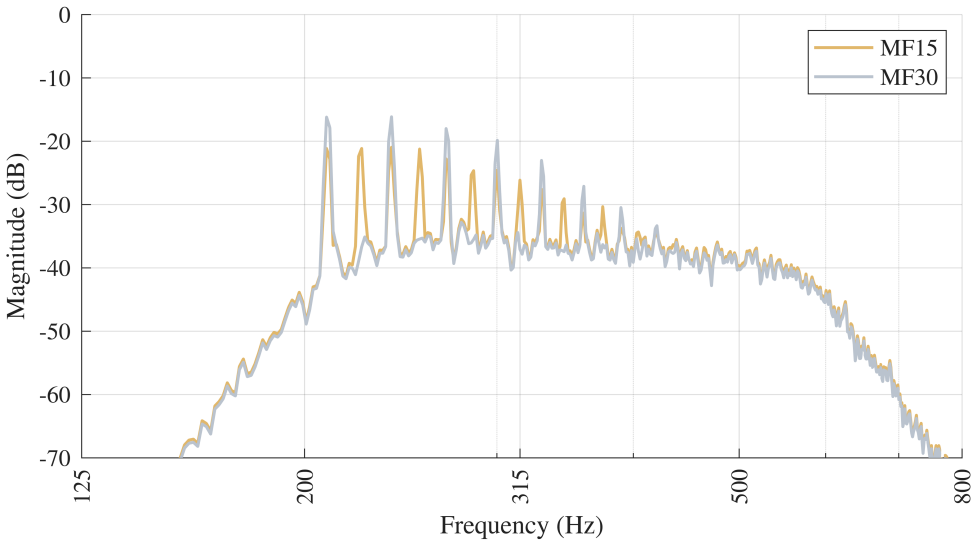


Figure 5.5: Spectra of the raw phase altered 15 Hz missing fundamental complex and the zero phase 30 Hz harmonic complex.

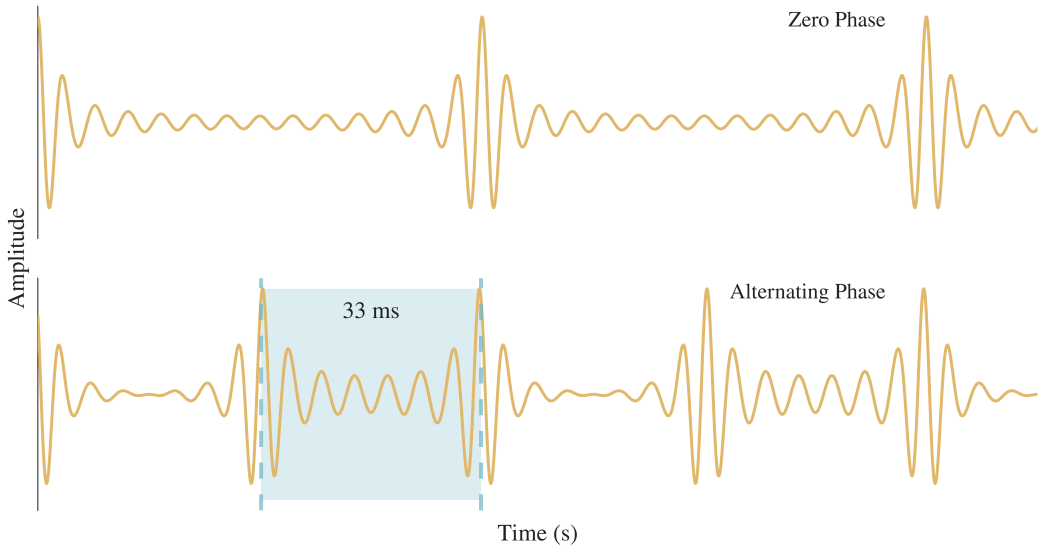


Figure 5.6: Temporal waveforms illustrating the phase manipulation of the 15 Hz missing fundamental complex. (Top) The zero-phase variant with a envelope period of 66.7 ms. (Bottom) The alternating-phase variant, creating a secondary peak that halves the envelope period to $\approx$ 33.3 ms, aligning with the temporal lower limit of pitch perception.

5.2.1 Spatial Configurations & Proof of Concept

To test the effect of decreased IACC in a realistic scenario, the four base signals were routed through different simulated speaker configurations. Mono configurations used the primary center speaker, while stereo configurations used both the center and an offset speaker to decrease IACC. In the stereo configurations the added speaker played either correlated or decorrelated pink noise. The decorrelation was achieved by generating the two pink noises using different seeded random number generation. Tables 5.2, 5.3 and 5.4 briefly overviews the spatial categories and their stimuli.



Figure 5.7: Frontal loudspeaker positions.

Table 5.2: Mono speaker stimuli configurations evaluated, simulating playback through the center speaker position.

Stimuli Name	Test II: Mono
M1	Band-passed pink noise (210–600 Hz).
M2	M1 + tonal components (210 Hz and 420 Hz).
M3	M1 + synthetic 30 Hz missing fundamental (MF) complex.
M4	Stochastic MF (5 Hz noise bands extracted at M3 harmonic center frequencies).

The new stimuli was also evaluated dynamically applying a pitch and modulation shift to asses the effect of applying a pitch shift as well as dynamic modulation on size perception. The dynamic amplitude modulation applied used the same parameters as the IPM2 stimulus in the initial test (**Test I**, Table 4.4). An pedestrian crosswalk position was also tested to asses weather the spatial effect was influenced by listening position. Tables 5.5 and 5.6 briefly describes the stimuli and their abbreviations. Figures 5.10 and 5.9 show spectrograms of the modulated and modulated pitch shifting stimuli.

Table 5.3: Stereo correlated configurations evaluated in an asymmetric two speaker setup, where the additional second offset speaker simulated outputting correlated PN.

Stimuli Name	Test II: Stereo Correlated
S1	Dual-speaker correlated pink noise (210–600 Hz).
S2	Primary-anchored tonal components (210 and 420 Hz) + S1.
S3	Primary-anchored 30 Hz missing fundamental (MF) complex + S1.
S4	Primary-anchored stochastic MF (5 Hz noise bands) + correlated pink noise.

Table 5.4: Stereo decorrelated configurations, where the additional second offset speaker simulated outputting decorrelated PN.

Stimuli Name	Test II: Stereo Decorrelated
D1	Dual-speaker decorrelated pink noise (210–600 Hz).
D2	Primary anchored tonal components (210 and 420 Hz) + D1.
D3	Primary-anchored 30 Hz missing fundamental (MF) complex + D1.
D4	Primary-anchored stochastic MF (5 Hz noise bands) + decorrelated pink noise.

Table 5.5: Stimuli configurations evaluating pitch and amplitude modulation shifts, simulating vehicle acceleration from 0 to 20 km/h.

Stimuli Name	Test II: Pitch & Modulation shift
PM1	Pitch-shifting baseline AVAS (A).
PM2	PM1 + dynamic AM.
PM3	Pitch-shifting 30 Hz MF complex (M3).
PM4	PM3 + dynamic AM.
PM5	Pitch-shifting stochastic MF (M4).
PM6	PM5 + dynamic AM.

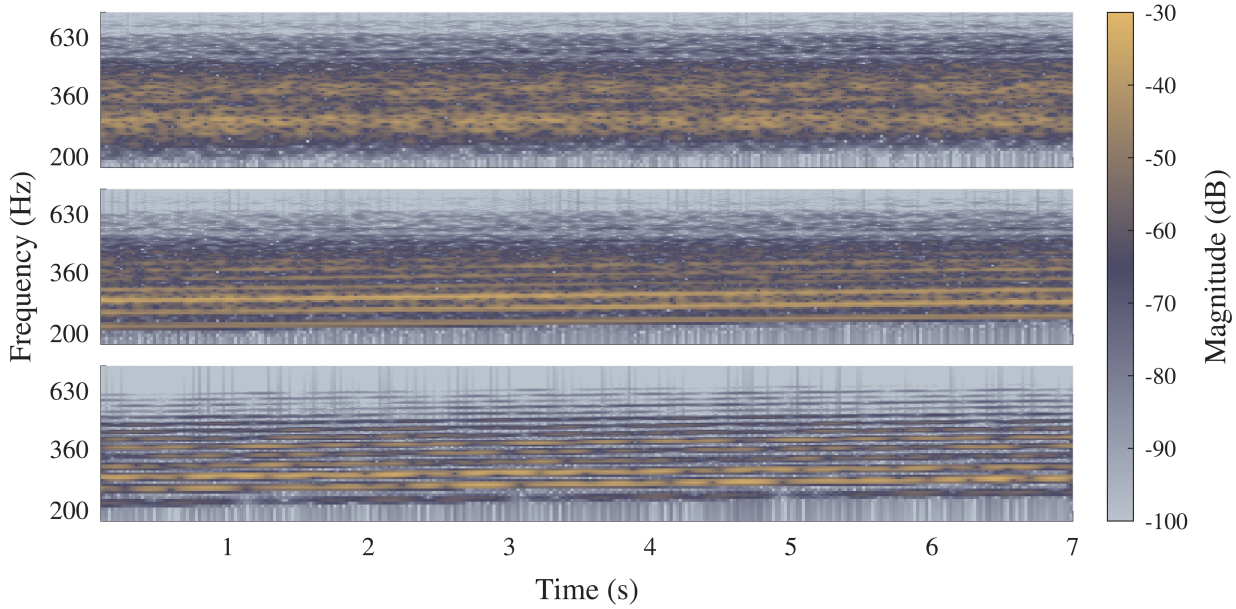


Figure 5.8: Spectrogram of the unmodulated variants (**PM1**, **PM3** and **PM5**) from the Pitch & Modulation shift category. The STFT was computed using an 8192-point Hamming window with a 87.5% overlap to be able to see the tonal components in the stimuli.

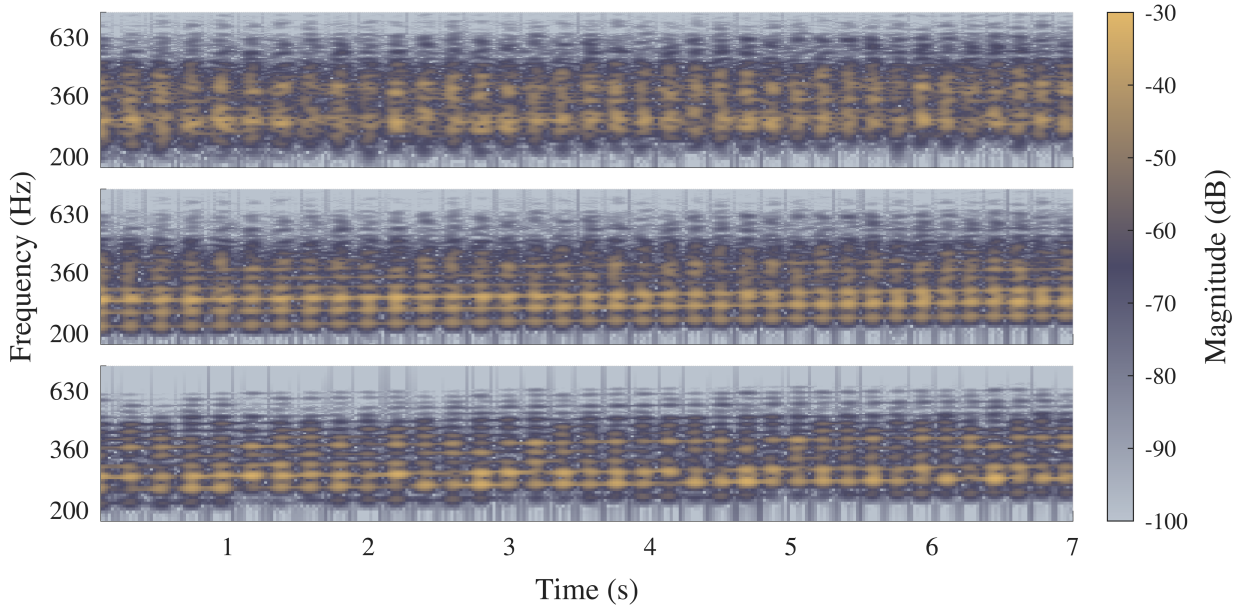


Figure 5.9: Spectrogram of the modulated variants (**PM2**, **PM4** and **PM6** from the Pitch & Modulation shift category. The STFT was computed using an 8192-point Hamming window with a 87.5% overlap to be able to see the tonal components in the stimuli.

The final paired comparison category was a proof of concept, where the missing fundamental mechanics where applied to the excising AVAS (A) and compared to each other as well as the unaltered AVAS signal. Additionally, two high-passed (with a cut-off frequency 200 Hz) variants of an ICE truck recording was added to this category to asses how these new signals would fare against having the AVAS speaker play these. Both ICE truck recordings were run through the same calibration/normalization as the other signals, detailed in Section 5.3 below. This meant that result of the only high-passed ICE recording had a high spectral centroid compared to the other stimuli. The second ICE truck recording was band-passed between 200 Hz and 600 Hz to be a more fair comparison to the other signals. Both signals were high-passed with a cut-off frequency of 200 Hz for the reasons discussed in Chapter 2.

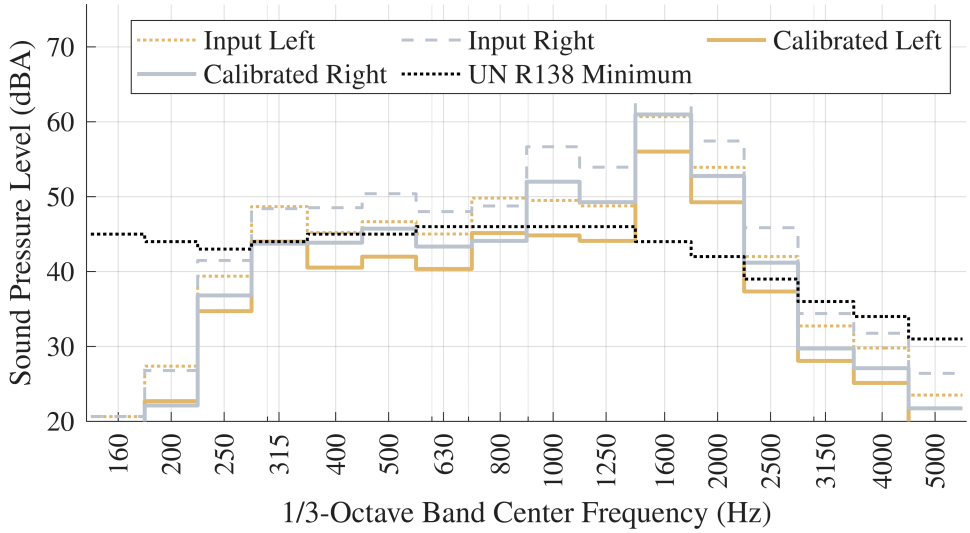


Figure 5.10: Calibration process of the unfiltered ICE truck recording, because of the R138 requirement of at least one band being bellow 1600 Hz the spectra of the resulting stimuli is skewed to higher frequencies compared to other signals in the comparison.

Table 5.6: Pedestrian position configurations testing the effect at a different position to see if the effect is position dependent.

Stimuli Name	Test II: Pedestrian Position
PP1	Band-passed pink noise (210–600 Hz).
PP2	PP1 + synthetic 30 Hz missing fundamental complex.
PP3	PP2 + Added offset speaker playing correlated pink noise.
PP4	PP2 + Added offset speaker playing de-correlated pink noise.

Table 5.7: Proof of concept AVAS configurations, testing the virtual pitch methods when layered directly over an existing AVAS.

Stimuli Name	Test II: AVAS Proof of Concept
PC1	Baseline AVAS (A).
PC2	PC1 + synthetic 30 Hz missing fundamental (MF) complex.
PC3	Stochastic MF (5 Hz noise bands extracted from PC1).
PC4	ICE truck recording.
PC5	Band-passed ICE truck recording (210–600 Hz).
PC6	PC3 + synthetic 30 Hz MF complex.

5.3 Regulatory Calibration and Loudness Normalization

To ensure the synthesized stimuli complied with the band-level requirements of UN Regulation No. 138 (Table 2.2), a deterministic calibration algorithm was applied. Before convolution, a 5 ms amplitude fade was applied to the dry signals to mitigate any impulse transients.

The calibration process calculated the precise linear gain required for each stimulus to meet the UN R138 10 km/h thresholds. The dry signals were convolved with the impulse responses of the edge directivity microphones located at 0° and 180°. Because the directivity measurements were captured at a radius of 5 m, free-field acoustic propagation was assumed and an attenuation of approximately −8 dB was applied to map the regulatory 2 m limits to the 5 m measurement distance. The calibration process is visualized in Figure 5.11, which displays the signal convolved with the directivity positions before and after calibration. It is also important to note that the stimuli excludes other sources such as tire noise, which in reality would contribute to the radiation and therefore may reduce the needed level output of the AVAS speaker.

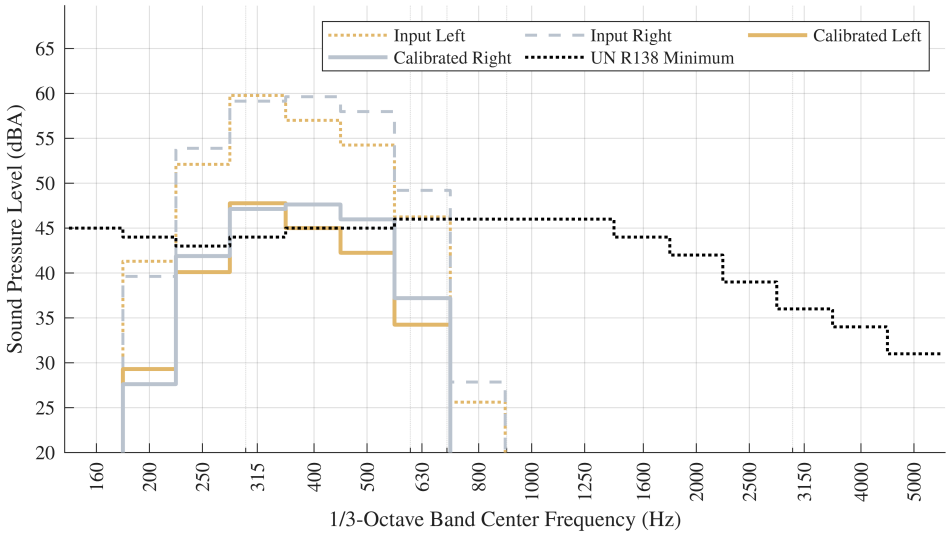


Figure 5.11: UN R138 calibration (simulating 10 km/h) for the reference pink noise stimulus. The overall level was attenuated to precisely fulfill the minimum two-band requirement (indicated by the dashed line). Levels are extrapolated to the regulatory 2 m distance from the empirical 5 m measurements, assuming free-field propagation.

The signals were then A-weighted, and the continuous sound pressure level (SPL) was calculated for one-third-octave bands. Here an algorithm was used to find the precise gain needed for each band to meet the threshold, these were then evaluated to ensure that a minimum of two bands, at least one of which was positioned below 1600 Hz, was within the requirement. The algorithm independently evaluated the left and right channels and used the worst-case scenario as a linear scalar to the unconvolved dry signal. The scaled signals were then convolved with the HRTF impulse responses for binaural spatialization. To remove loudness as a factor, stationary free-field loudness was calculated for all binaural signals. In this study, loudness was modeled according to ISO 532-1. The maximum calculated loudness across all stimuli was established as the universal normalization target.

An iterative minimization algorithm was used to normalize each stimuli to this loudness. The objective function was defined as the absolute error between the target loudness and the calculated loudness in Sones: $f(g) = |L(x \cdot g) - L_{\text{target}}|$ where L is the ISO 532-1 loudness calculation, x is the binaural acoustic signal, and g is the scalar linear gain. The algorithm iteratively solved for the optimal gain g that minimized this error, ensuring that all stimuli fulfilled regulatory minimums.

5.4 Test II: Results

The results of the final listening test are presented in this section. As with the initial test, subjective preferences were evaluated using the BTL choice model to generate share-of-preference scores. The same goodness-of-fit threshold of $p < 0.1$ was used. In addition to paired comparisons, the average scores from the ranked category comparisons are presented. The complete paired-comparison matrices for all categories in Test II, which show win rates and statistical significance, are provided in Appendix 8.

5.4.1 Base Stimuli

The result of the category comparing the four base signals in the **Mono** spatial configuration show a clear preference for both the stimuli utilizing the missing fundamental phenomenon, over both the other stimuli, within statistical significance. Between the two preferred stimuli there is a slight preference for the stochastic **Ext** version, but not within statistical significance. This trend is also seen in the ranked comparison result seen in Figure 5.13 where these stimuli has higher average scores compared to all other stimuli in the comparison. Notably their average scores are very similar. Furthermore, the comparison between the zero-phase **MF30** stimulus and its phase-altered counterpart (**MF15**) yielded no statistically significant preference, indicating that both stimuli were perceived as highly similar in size.

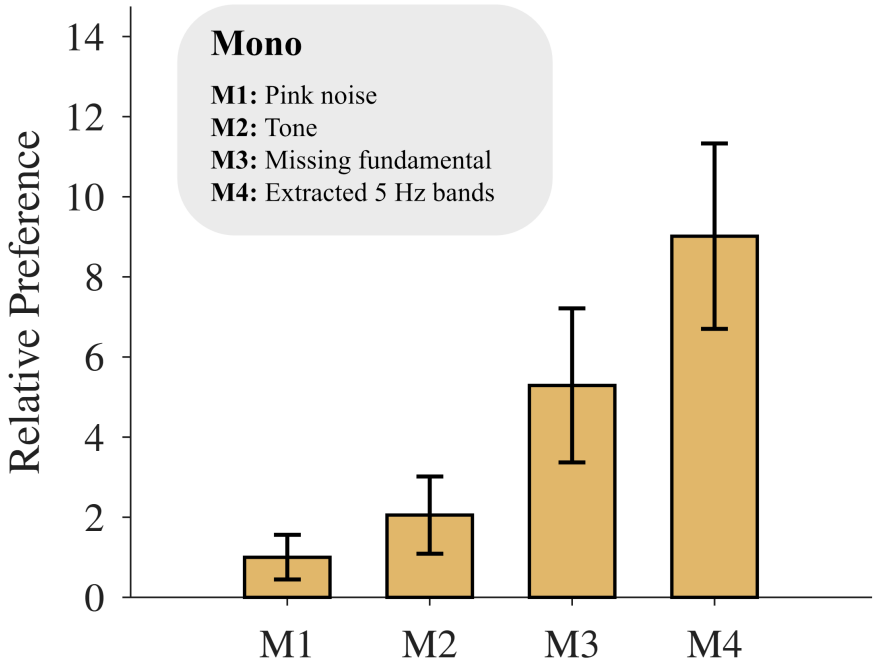


Figure 5.12: Share of preference (BTL scores) for the final listening test category Mono. Error bars represent the 95% confidence intervals. Each category is scaled so the lowest ranked stimulus equals 1.

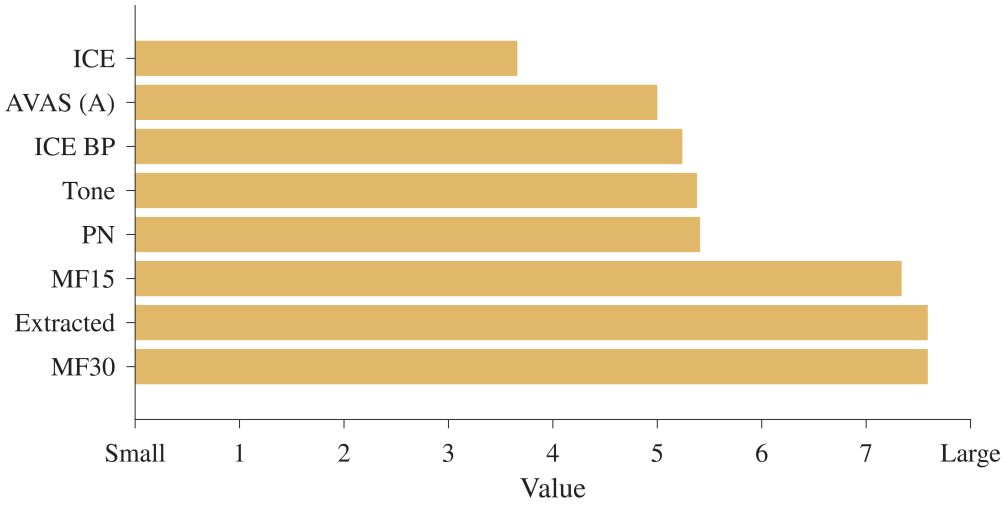


Figure 5.13: Average ranked comparison score reported form the participants for the eight different stimuli. All stimuli was simulated as being played by the center (Mono) speaker position.

5.4.2 Influence of Amplitude Modulation

The results of the comparisons of modulated and unmodulated base stimuli is seen in Figure 5.14. No preference for either version of the pink noise stimuli (**PN**) is seen, but the modulated version of the **Tone** stimulus was preferred over the unmodulated variant. Both the stimuli utilizing the missing fundamental phenomenon had the reverse result, where the unmodulated variant was preferred.

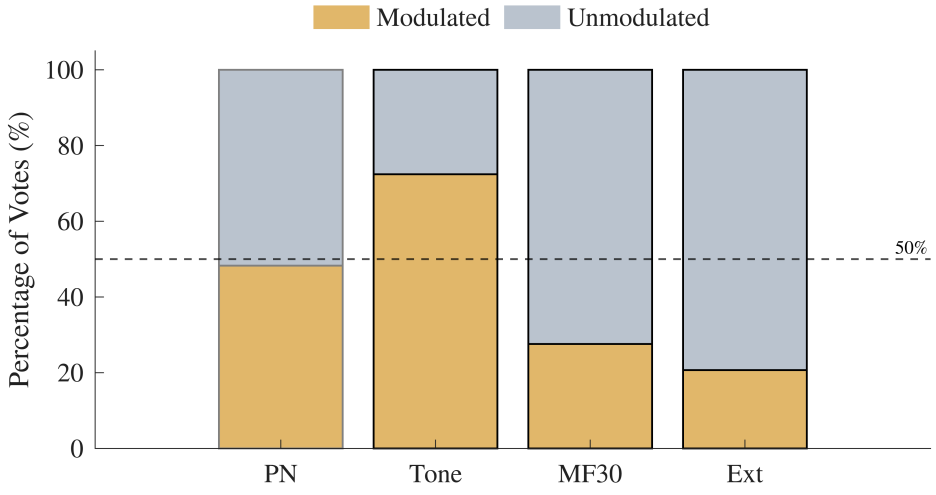


Figure 5.14: Results of the modulated vs unmodulated variants of the base stimuli. A black outline represents if the result was statistically significant.

The result of the dynamic modulation category sen in Figure 5.15 showed an overall preference for the stimuli with the added dynamic modulation, although this effect was only significant for the unmodulated and modulated AVAS (A) signal.

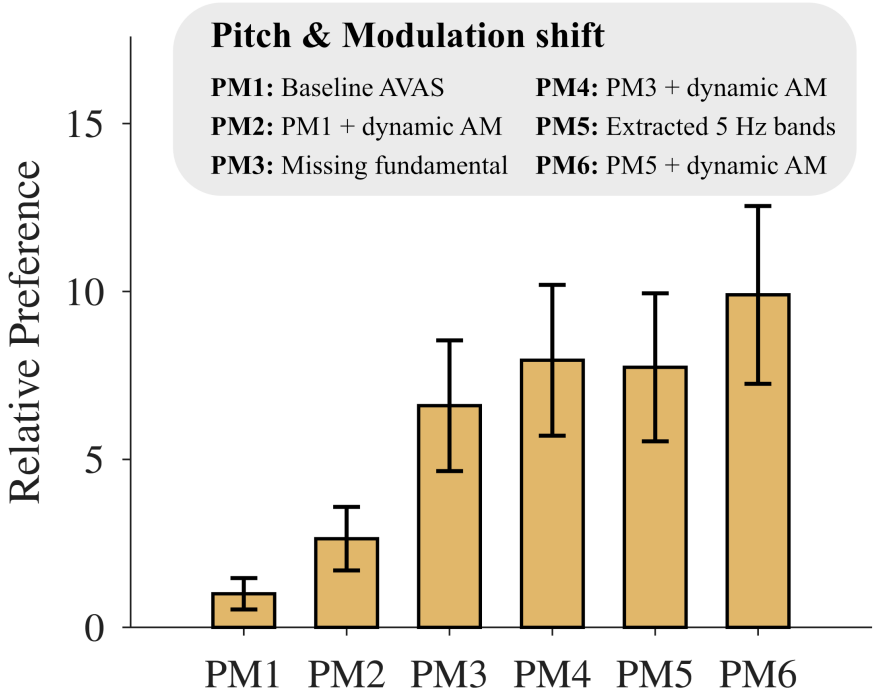


Figure 5.15: Share of preference (BTL) for the Pitch & Modulation Shift category, evaluating the dynamic acoustic response of the virtual pitch variants. Error bars represent the 95% confidence intervals. Each category is scaled so the lowest ranked stimulus equals 1.

5.4.3 Spatial Configurations

The same trend as in the Mono category is seen for the two other spatial configurations comparing the four base signals, where again the stimuli utilizing the missing fundamental phenomenon is highly preferred over the other base stimuli. However, there is a slight shift in preference between these within the categories, where the correlated category has a more even preference prediction for the two stimuli compared to both the Decorrelated and the Mono categories, none of the spatial configurations does however indicate a statistically significant result in the comparison between the two missing fundamental stimuli. The BTL model predictions for the **Stereo Correlated** and **Stereo Decorrelated** categories are shown in Figure 5.16.

The results of the ranked comparison seen in Figure 5.17 also show a similar result, having the missing fundamental phenomenon stimuli being rated with a higher average than the band passed pink noise signals. The differences between these higher ranked stimuli are very small, and this is also true for the different spatial configurations, only resulting in a slightly higher average scoring than the other variants.

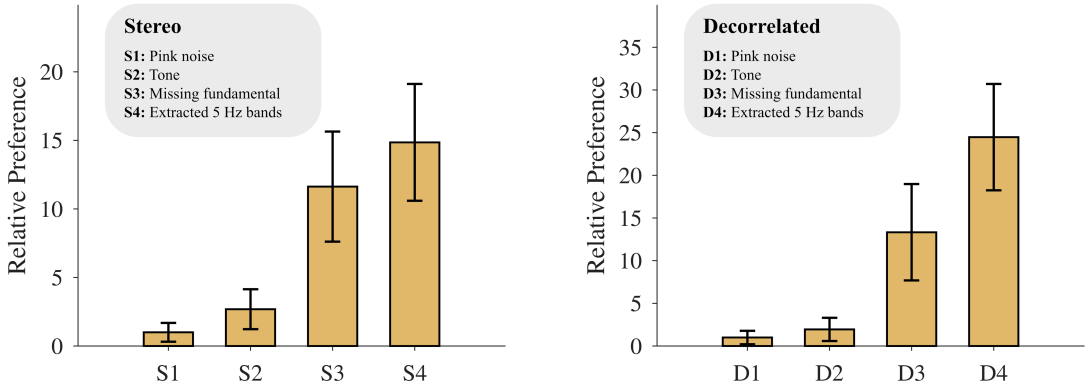


Figure 5.16: Share of preference (BTL) for the final listening test categories Stereo Correlated (left) and Stereo Decorrelated (right). Error bars represent the 95% confidence intervals. Each category is scaled so the lowest ranked stimulus equals 1.

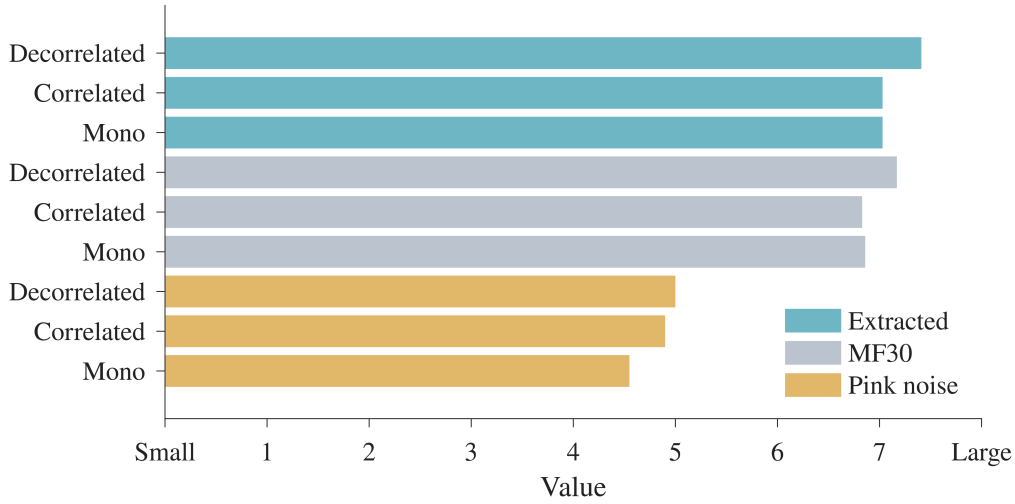


Figure 5.17: Average ranked comparison score reported form the participants for the spatial configuration variants of **Ext**, **MF30** and **PN**.

The comparisons between the same base signal with different spatial configurations has a spread result with only 4 statistically significant comparisons split between the 3 categories. The BTL model predictions are shown in Figure 5.18. No statistically significant result is recorded for the **PN** or the **Tone** category. The **MF30** category show a preference for the Mono and Stereo Decorrelated variants over the Stereo Correlated variants, whilst the **Ext** show a preference for the Mono over both the other variants. The category evaluating the pedestrian position (Figure 5.19) show a similar result, with a slight preference for the decorrelated variant over the two other spatial configurations.

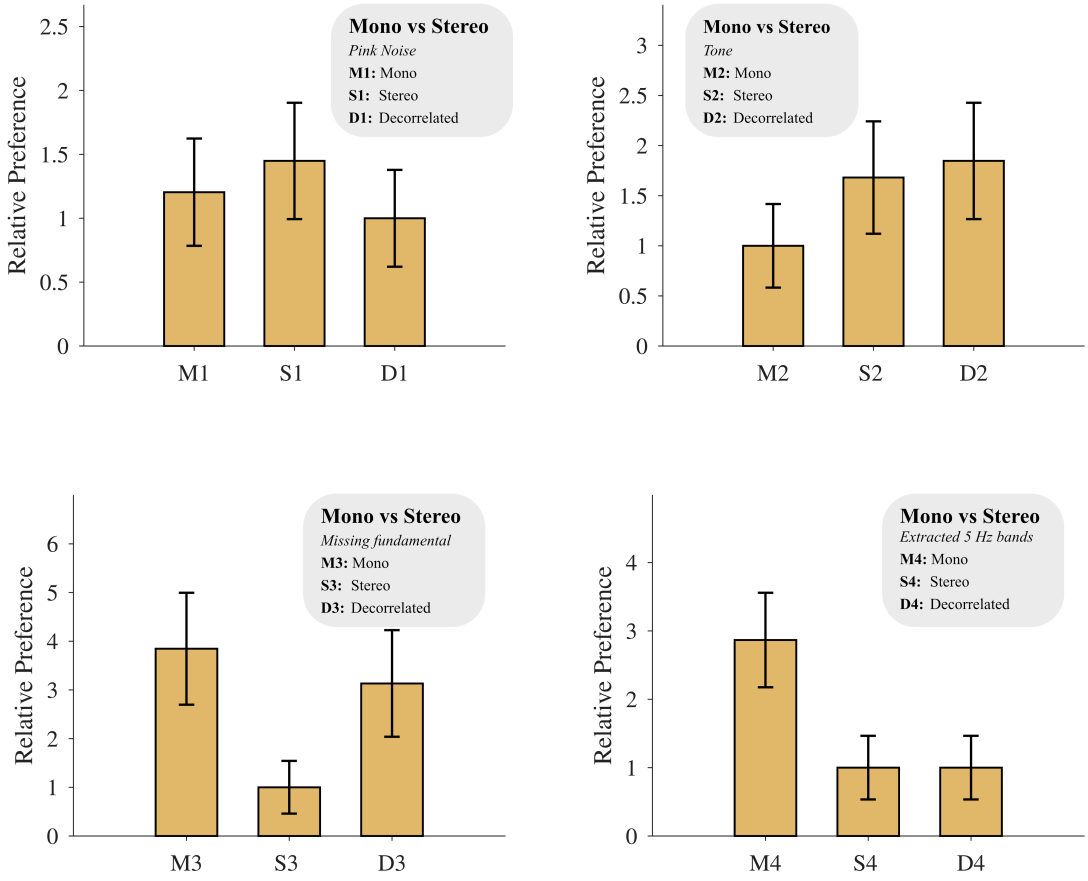


Figure 5.18: Share of preference (BTL) evaluating the effect of spatial distribution across the four base stimuli: **PN** (top left), **Tone** (top right), **MF30** (bottom left), and **Ext** (bottom right). Error bars represent the 95% confidence intervals. Each category is scaled so the lowest ranked stimulus equals 1.

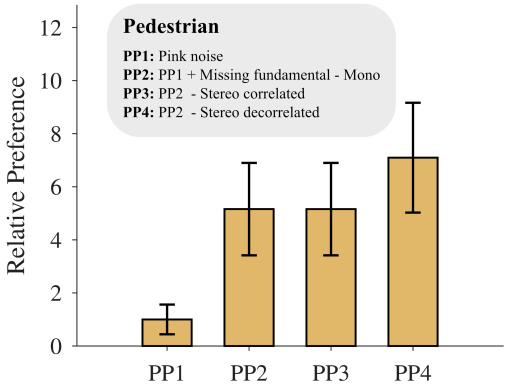


Figure 5.19: Share of preference (BTL) for the Pedestrian Position category. Error bars represent the 95% confidence intervals. Each category is scaled so the lowest ranked stimulus equals 1.

5.4.4 AVAS Proof of Concept

Finally, the two virtual pitch methods were applied into an existing regulatory AVAS signal (Proof of Concept). As seen in Figure 5.20, the modified AVAS signals using the missing fundamental methods were significantly preferred over the unmodified baseline AVAS. The existing AVAS preformed very similar to the band passed ICE truck recording, while the raw ICE truck recording was the least preferred.

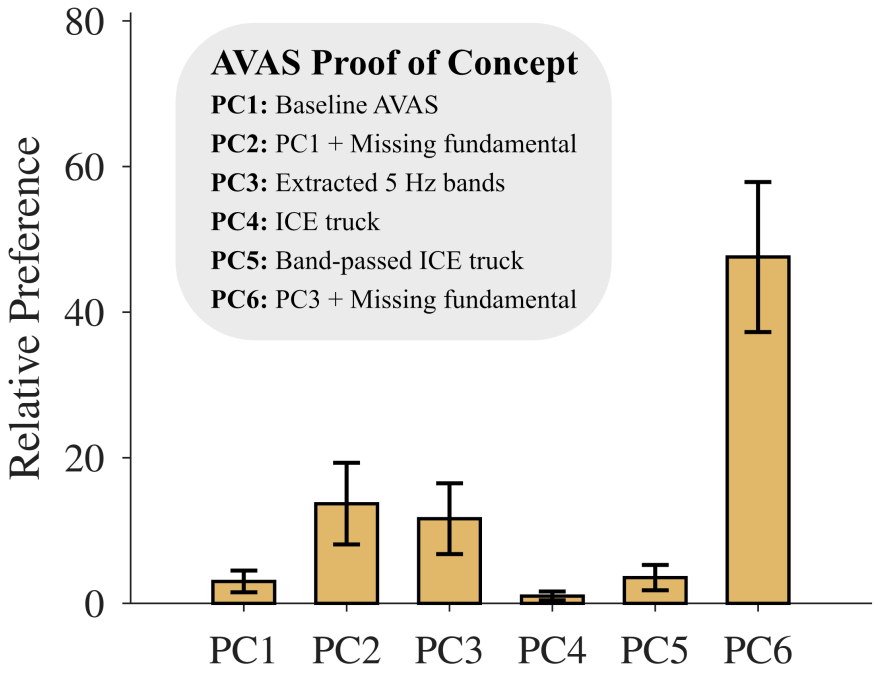


Figure 5.20: Share of preference (BTL) for the Pedestrian Position category. Error bars represent the 95% confidence intervals. Each category is scaled so the lowest ranked stimulus equals 1.

6
Discussion

The primary goal of this thesis was to identify and synthesize acoustic parameters that enhance the perceived size of AVAS signals for heavy electric vehicles while adhering to the regulatory demands, minimizing low frequency noise pollution and considering hardware limitations. This Chapter interprets the results from the two listening tests, first handling signal properties such as spectral balance, pitch and time structure. It then discusses the efficiency of utilizing the missing fundamental phenomenon (virtual pitch) to lower the pitch whilst keeping the signal within the chosen frequency band limitation (200-600 Hz). In Section 6.3 the results of the spatial configurations using the stereo configurations are analyzed and the physical limitations of using several sources to decrease the IACC are addressed. Finally some methodological limitations and potential future study considerations are outlined.

6.1 Psychoacoustic Determinants of Perceived Source Size

The auditory perception of size is closely connected to spectral balance and pitch, where the prominence of low frequency content is the dominating acoustic cue. This section evaluates the signal properties isolated during the listening tests, specifically analyzing how shifts in spectral balance, the presence of tonal structures, and temporal modulation influence the perception of auditory size.

6.1.1 Frequency Content & Spectral Balance

The primary cues for size perception seem to be spectral balance and pitch, which aligns with literature and with fundamental acoustic principles, as larger physical bodies radiate longer wavelengths and thus a lower fundamental frequency, resulting in a lower pitch as discussed in Chapter 2. The results of the spectral centroid category show a clear correlation between shifting the spectra towards low frequencies and increasing the perceived size, where all comparisons were statistically significant, with participants consistently selecting the stimuli possessing the lower centroid. This trend was observed throughout the various test categories, and it is one of the reasons as to why the stimuli in the second test all had very similar spectra. In the category comparing currently implemented AVAS signals this same trend is seen, where AVAS (B), which had the spectra most skewed towards low frequencies was heavily preferred over the other stimuli as seen in Table 8.4. This is also in line with the findings of Kullukcu et al. that indicated that energy in the 100 Hz to 250 Hz

range produced a truck like sensation [23], which may be associated with a larger auditory size. While this result is not necessarily surprising it introduces a conflict of interest; low frequencies are highly important for increasing size perception, but introducing high energy at low frequencies negates one the main benefits that electric heavy vehicles have over conventional ICE heavy vehicles, namely their reduced noise pollution in urban environments, which was found to be most prominent at low frequencies by both Glebe et al. [2] and Ziqin et al. [24].

6.1.2 Tonal vs. Broadband Components

The data indicate a clear preference for stimuli containing tonal components over purely stochastic broadband noise. This preference is seen in the initial Spatial Decorrelation evaluations and was maintained as a trend during the mono-speaker comparisons in the final listening test.
One hypothesis as to why this is the case is that the noise is more diffuse and harder to attribute to a single source while tones are more easily placed as coming from a single source. Further, the pitch of low-passed noise corresponds to its cut-off frequency [13] which means that many of the noise stimuli used in this study would have had a higher pitch than the tonal components they were compared to.

6.1.3 Time Structure

The results of the temporal modifications, namely the categories involving AM and phase randomization, display a complex relationship. On the one hand, adding a stationary rate of AM yielded almost no statistically significant effect, indicating that stationary AM does not contribute heavily to size perception. On the other hand, when a dynamic amplitude modulation was added to the regulatory demanded pitch shift there was a clear preference for the slowest AM increase per km/h compared to both the other dynamically modulated signals and the modulated reference. Additionally, the highest increasing modulation per km/h was least preferred, although this result did not achieve statistical significance.

Furthermore, comparisons between stimuli with identical spectral centroids but differing temporal structures (preserved versus random phase) compared in the initial test, indicated that the time structures role is frequency dependent. Stimuli retaining the intact time structure of an idling ICE truck were consistently preferred over their random-phase counterparts. Notably, this preference was most significant for stimuli with a higher spectral centroid. It could be speculated that the familiar temporal envelope of a combustion engine allows listeners to associate the sound with a heavy vehicle, compensating for the physical lack of low frequency cues allowing listeners to associate the sound with a heavy vehicle. These findings provide an answer to **RQ2**, revealing that while stationary amplitude modulation is ineffective, dynamic modulation and preserved ICE-like temporal envelopes can enhance size perception compared to purely deterministic signals.

In the second listening test all base stimuli were compared against their AM coun-

terparts which indicate that the modulation can disrupt the virtual pitch and thus have a negative effect on size perception. Whilst the AM increased the perceived size of the **Tone** stimuli it had a significant inverse effect on both the stimuli composed of harmonic complexes. It has been noted by Fastl and Zwicker that a slow rate of amplitude modulation can reduce pitch strength which might explain this effect [13].

6.2 Synthesis & Efficacy of Virtual Pitch

Low frequencies pose a problem when designing the AVAS signal, partly because of the potential physical limitation of the speaker. Another reason is that low frequency content can contribute excessively to noise pollution because of the large wave lengths more easily propagating through building facades and into the indoor environment [1]. Because of these factors a lower limit of 200 Hz was put on all the stimuli in the second listening test. This limitation of lower frequency content is in direct conflict with increasing the perceived size of the signal, since as previously discussed, one, if not the most important cue for the auditory system for judging size is spectral balance and prominence at low frequencies. But another closely connected cue is pitch, which has been proved to be one of the deciding factors for size perception of speakers (people), instruments [12] as well as animals [9]. In response to **RQ3**, these results confirm that psychoacoustic principles of size perception translate effectively to heavy vehicle warning sounds. Even if pitch is closely connected to spectral content by being ruled by the fundamental frequency of the harmonic complex, this has some lee way, in that a virtual pitch can be synthesized by making use of the missing fundamental phenomenon. This phenomenon is a product of how the auditory system processes pitch and is limited by its temporal processing window. This limitation of lower frequency content is in direct conflict with increasing the perceived size of the signal. Addressing **RQ4**, the successful application of the missing fundamental phenomenon above this 200 Hz threshold proves that virtual pitch can indeed be implemented within current hardware limitations while maintaining UN R138 compliance.

6.2.1 Missing Fundamental (MF)

As discussed in Chapter 2 Section 2.3.2 an harmonic complex can have the same pitch even if the early harmonics are removed. This enables for a low pitch without the inclusion of spectral content below our defined lower limit of 200 Hz. One of the main goals of the second test was to test if the virtual low pitch could increase size perception in a meaningful way and the results indicate that this was successful. Throughout the spatial configurations the **MF30** stimulus utilizing lowest perceivable pitch defined as 30 Hz by Krumbholz et al. [14] which included the 14th to the 20th harmonic of that fundamental, was consistently preferred over that of both the pink noise stimuli and the **Tone** stimuli. This directly addresses **RQ1** by demonstrating that the missing fundamental phenomenon can effectively increase perceived auditory size without introducing physical low-frequency energy or increasing overall loudness.

The comparison between the phase altered **MF15** and the zero phase **MF30** showed no preference for either stimuli. Since both stimuli exhibit equally spaced peaks, as seen in Figure 6.1, it could be speculated that they also produce the same virtual pitch, which could then explain the result.

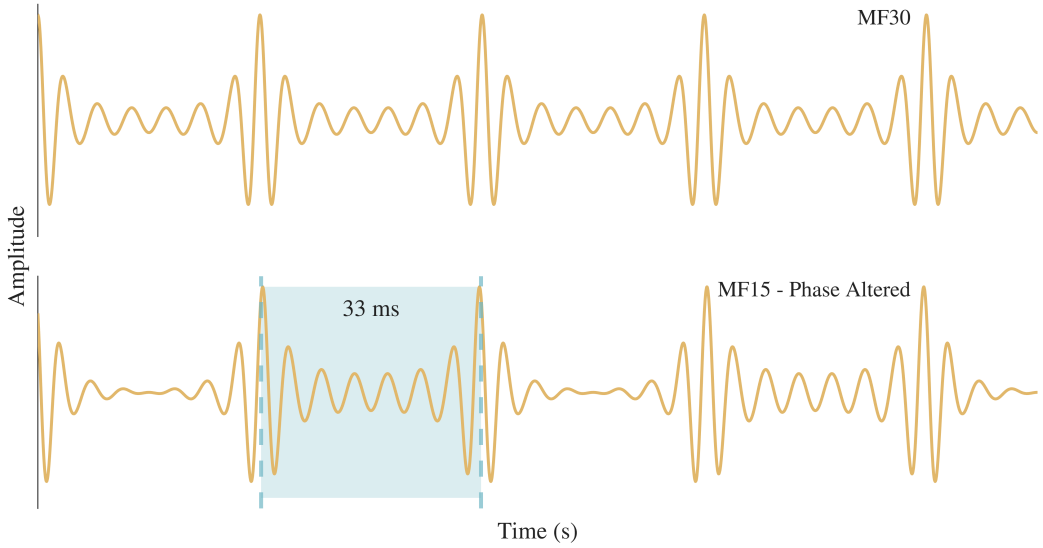


Figure 6.1: Time structure comparison between the zero phase missing fundamental tonal complex, MF30, and the phase altered MF15. Both signals exhibit highly similar envelopes with peaks located just within the 33 ms temporal processing window.

It is worth noting that adding the missing fundamental harmonic complex shifts the median spectral centroid down compared to the other stimuli due to the Gaussian filter gradually decreasing the upper harmonics which is seen in Figure 6.2. This effect may not be negligible, and further research is needed to conclude to what extent the preference is simply due to this shift. It is however noteworthy that even if the comparison did not yield a statistically significant result, there seem to be a slight preference for the **Tone** stimulus over the **PN** stimulus despite the raw signal having a higher spectral centroid. This indicate that the spectral centroid, while a good indicator, may not be enough to predict what signal is going to be perceived as being larger. Further, it highlights the importance of the hardware and it's effect on the raw signal. As seen when comparing Figure 6.2 to Figure 6.3 the spectral centroid of the **Tone** stimulus has been shifted down after the after the raw signal has been convolved with the HRTFs.

Furthermore, the missing fundamental complex relies on a fundamental frequency of 30 Hz, which is located at the lower threshold of pitch perception [14]. This sits in the smooth transition between fluctuation strength and roughness. Because of this, the interaction of the tightly spaced harmonics introduces roughness to the signal [13]. As demonstrated by Kullukcu et al., increased roughness is correlated with "truck-likeness" [23]. This time varying roughness is illustrated in Figure 6.4, which shows a higher median roughness for the virtual pitch stimuli. Therefore, it may be that the roughness component increase the perceived size of the vehicle due to the associations with large ICE vehicles.

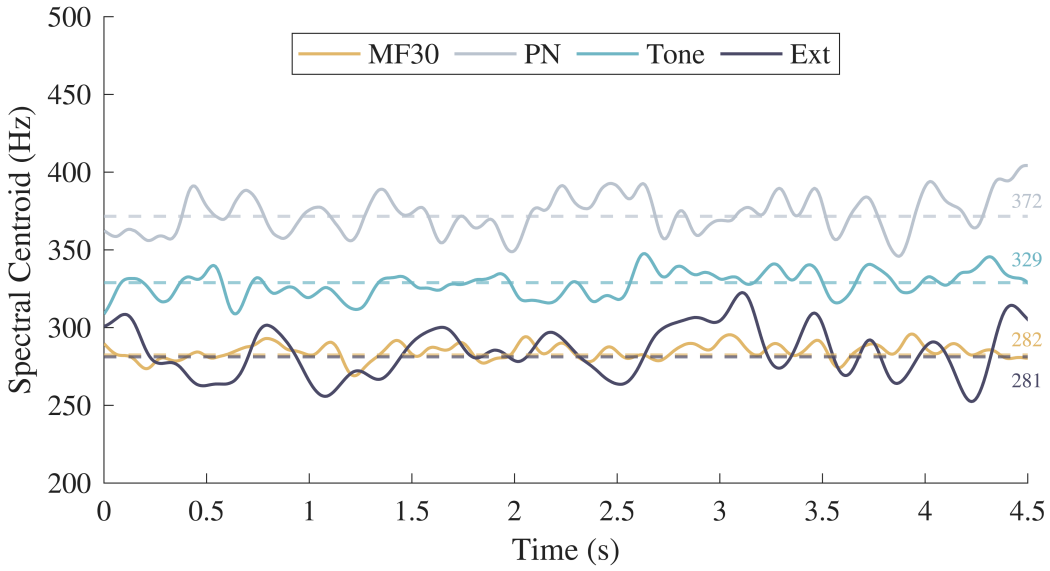


Figure 6.2: Spectral centroids vs time for the 4 raw base stimuli. The median value of each stimuli is represented with a dashed line in the same color of the stimuli and the value is written in the same color at the end of those lines. A smoothing function (20-sample gaussian filter) has been applied to improve readability.

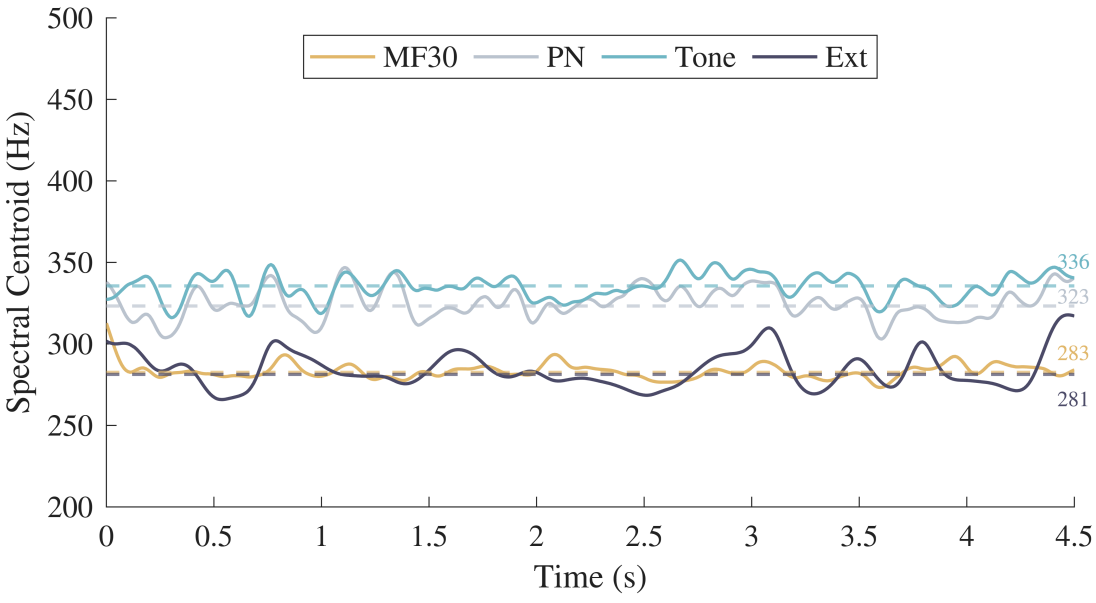


Figure 6.3: Spectral centroids vs time for the 4 base stimuli used in the **Mono** category, so after calibration, convolution and normalization. The median value of each stimuli is represented with a dashed line in the same color of the stimuli and the value is written in the same color at the end of those lines. A smoothing function (20-sample gaussian filter) has been applied to improve readability.

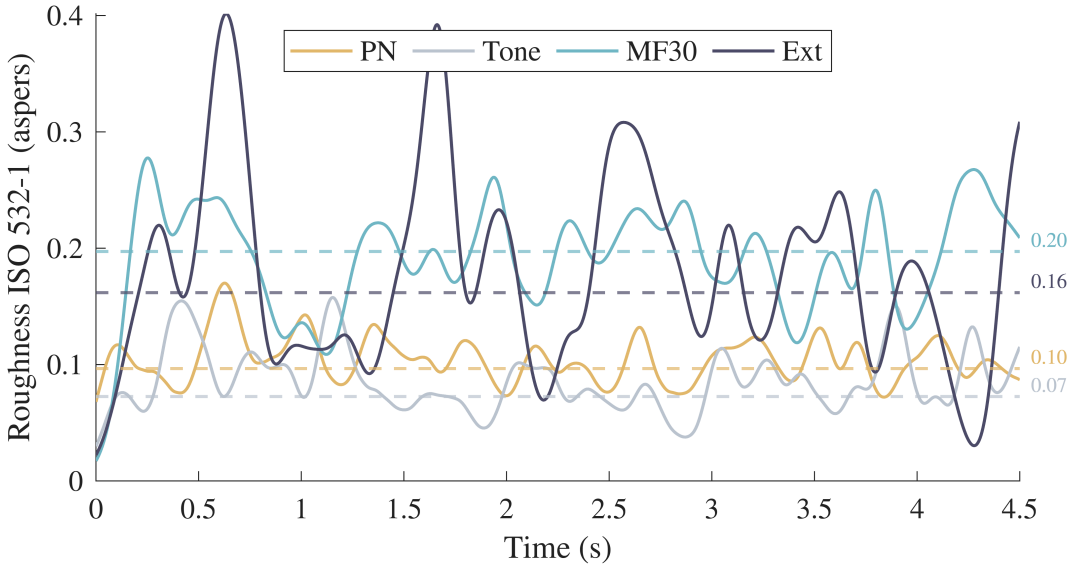


Figure 6.4: Roughness calculated for the stimuli used in the **Mono** category. A smoothing function (512-sample gaussian filter) has been applied to improve readability. The median roughness of each stimulus is represented with a dashed line in the same color of the stimuli and the value is written in the same color at the end of those lines.

6.2.2 Stochastic vs. Deterministic MF

The **Ext** stimuli also utilized the same phenomena but with an added stochastic nature. By extracting the 5 Hz wide bands at the harmonic indices of the 30 Hz fundamental the random phase of the pink noise is incorporated into the stimuli, resulting in a pitch and amplitude fluctuating signal that shifts stochastically. The nature of the signal is seen in Figure 6.5 where the time structure of the band-passed pink noise is compared to the normal and stochastic missing fundamental complexes. The results indicate that the stochastic missing fundamental complex is slightly preferred over the one created with pure synthetic sines, but the preference is not substantial and as seen in the ranked comparisons (Figures 5.13 and 5.17) both stimuli achieved an average score very similar to each other.

6.3 Impact of Spatial Configuration and Directivity

While the spectra and time structure of the synthesized signals establish the baseline acoustic signature of the vehicle, the physical radiation and spatial distribution of these sounds introduce an additional layer of considerations. One of the hypothesis of this study was that increasing the perceived width of the signal by decreasing the IACC at the listener position, would enhance the perceived size of the vehicle. This section evaluates the efficacy and limitations of this spatial approach, discussing the physical restrictions caused by the frequency range of the evaluated signals. Further it addresses the effect of mounting position on the radiation.

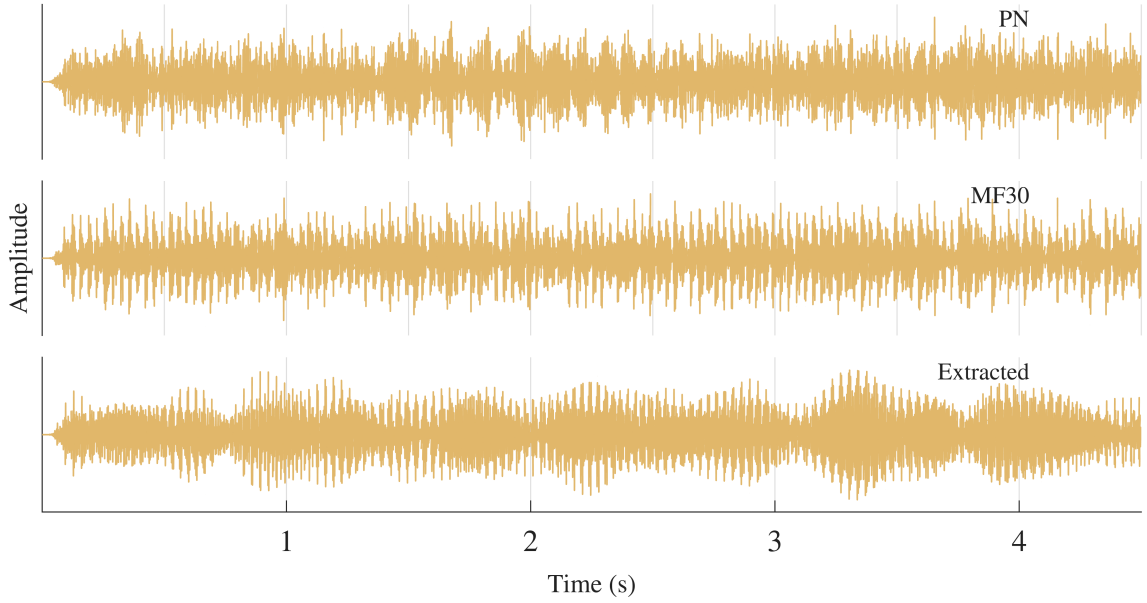


Figure 6.5: Time structure comparison between the band passed pink noise (**PN**) and the two types of missing fundamental stimuli used in the second test (**MF30** and **Ext**).

6.3.1 Interaural Cross Correlation (IACC)

In the initial listening test there seemed to be a preference for decorrelated stereo signals compared to mono signals, where, for instance, 16 of 17 participants perceived decorrelated pink noise with a calculated IACC of 0.008 as larger than its mono counterpart (IACC of 0.9454). This was in line with the hypothesis that increasing the ASW by decreasing the IACC would increase the perceived size. The next question to answer was whether adding an additional speaker on to a heavy truck that played a signal decorrelated from the main signal would have a the same effect on the perceived size of the vehicle.

To test this, an additional speaker (Offset) was simulated, radiating a band-passed pink noise signal generated with a different random number seed than the primary speaker (Center). Cross-correlating the raw digital signals confirmed they were uncorrelated (IACC ≈ 0). Furthermore, evaluating the broadband HRTFs for these positions yielded IACCs of 0.85 for the Center position, 0.74 for the Offset position, and 0.71 for the combined superimposed impulse responses. This validates that the physical geometry of the multi-speaker setup inherently provides a degree of spatial decorrelation.

However, analysis of the final stimuli revealed that this decorrelation was attenuated through the DSP pipeline. The IACC of the resulting binaural channels measured IACC values between 0.9911 for the mono pink noise and 0.9695 for the decorrelated stereo pink noise. This increase in correlation can be attributed to the frequency range of the signals. The stimuli were all designed to contain a similar spectra, ranging between 210 Hz and 600 Hz. In this low-frequency region, acoustic wavelengths range from approximately 0.57 m to 1.6 m, which exceeds the diameter of

the human head. Consequently, the sound waves easily diffract around the head increasing the IACC. This effect is displayed in Figure 6.6 where the spectra of the center positioned speakers IR is plotted. The IACC of the full frequency range is calculated to 0.85 whilst for the shorter range between 200 Hz and 600 Hz this is instead calculated to 0.98.

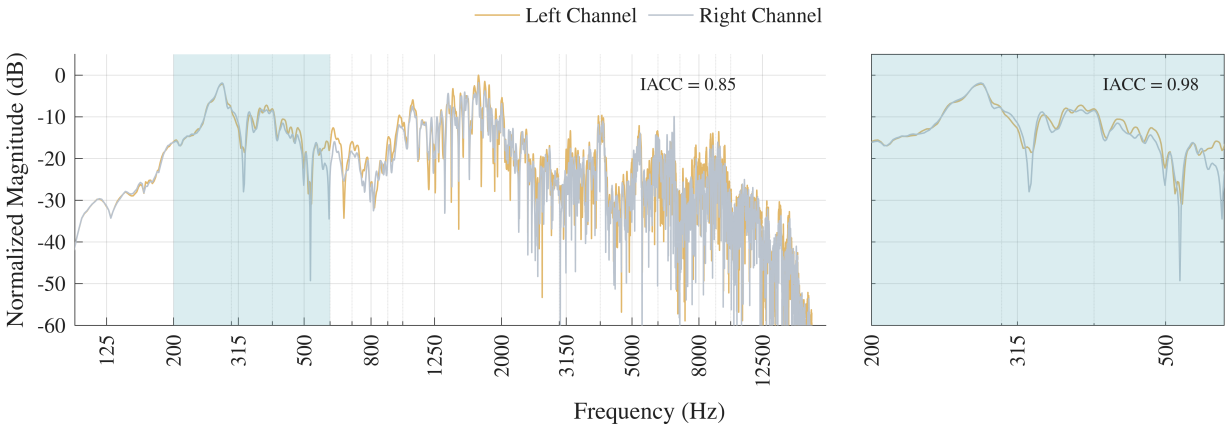


Figure 6.6: Shows the frequency spectra of the Center speaker location IR normalized to the global maximum. The calculated IACC is displayed for the full range as well as the 200-600 Hz range.

This together with the narrow angle ($\approx 10°$) created by a 0.9 m speaker separation at a 5 m listening distance, caused left and right ears to receive highly coherent signals. The high IACC does not indicate a flaw in the signal generation instead it reflects the physical reality that achieving a wide ASW using front mounted speakers with low frequency AVAS signals at typical pedestrian distances is inherently constrained. It is however worth noting that the auditory system is sensitive to small interaural differences when the reference sound is highly correlated as found by Gabriel and Colburn [25]. The just noticeable difference (JND) of IACC for the the 400 Hz bandwidth would according to their findings be between 0.01-0.02, which means that the difference between the correlation of the mono and the stereo decorrelated stimuli would be just noticeable ($\Delta\text{IACC} = 0.0216$). This might explain the mixed results for the spatial distribution categories, since while noticeable, the differences in IACC between the stimuli are small.

6.3.2 Speaker Position & Mounting

The difference in frequency response between the two speaker positions shown in Figure 3.4 where the Center speaker exhibited a 5 dB to 10 dB increase in the 185 Hz to 290 Hz frequency band compared to the Offset speaker can possibly be attributed to the differing mounting conditions and the complex geometry of the vehicle's underbody. The Center speaker is mounted in front of the engine block, which possibly acts as a reflective acoustic boundary that increases the forward radiation of that speaker location compared to the Offset one. The difference in spectra in this critical range likely contributed to the mixed results of the spatial configuration categories where because of this the median spectral centroid was shifted upwards

by between 5-8 Hz for the mono compared to stereo stimuli. Even if the effect is not extensive it still highlights the importance of mounting and placement of the speaker. The directivity seen in Figures 3.8 and 3.9 further highlights that the radiation pattern is frequency dependent, and that, as noted by Fortino et al. [8], this needs to be considered when designing the warning signal.

6.4 Methodological Considerations and Limitations

While the methods used in this study found several parameters that correlate to size perception and succeeded in increasing the size perception of an existing heavy vehicle AVAS signal certain limitations will be acknowledged in this section.

6.4.1 Spatial Configuration

The decorrelation of the two simulated speakers for the stereo configurations was complicated by two physical constraints. The introduction of the decorrelated pink noise would have had an effect on the balance of tonal signal and noise signal since the stimuli were all calibrated and loudness normalized. This might have had an effect on the virtual pitch perception where the shifted balance might have slightly masked the tonal components of the signal compared to the Mono case. This would have been especially apparent for the missing fundamental complex extracted from pink noise, which as mentioned did not contain any broad band pink noise in it's mono variant, but where the stereo variants had the added pink noise from the second speaker.

6.4.2 Initial Listening test IACC Category DSP Routing

An error in the DSP routing occurred during the synthesis of the stimuli used in the initial listening test. For the categories evaluating the effect of synthetic reflections, the reflections were applied to the audio signals after the convolution with the HRTFs. This means the reflections did not receive the correct spatial filtering inherent to the simulated environment. While this routing error compromises the absolute spatial accuracy of those specific stimuli, subjective comparisons between the flawed signals and corrected variants revealed that the difference between these was small. Because of this, conclusions drawn from the reflection categories in the initial test have been treated with caution throughout this discussion.

6.4.3 Ecological Validity of Headphone Auralization

Finally, while the use of processed HRTFs and calibrated playback systems does provide a controlled environment headphone auralizations do lack ecological validity. A quiet isolated room is vastly different from the environment vehicles would be located in, where the situational awareness is based on multi-sensory perception involving not only acoustic cues but also physical ground vibrations and visual context. In a typical urban setting, background noise masking would likely affect the perception of the synthesized signals. While the use of missing fundamental

complexes proved effective in a controlled environment, further testing which test the signals in the environment they are implemented in, would be needed to ensure these findings are true under these conditions.

7
Conclusion

This thesis investigated the influence of parameters on perceived auditory size and the feasibility of utilizing these in an applied situation to enhance the perception of heavy electric vehicles. It showed that spectral balance and pitch, which have been found to be the main cues of auditory size perception in human voices, animal calls, and musical instruments, remain effective for stochastic noise and tonal complexes. Further, it was found that reduced IACC has a positive effect on size perception and that associative modifications such as amplitude modulation do affect this perception, although less prominently and depending on whether it was stationary or dynamically applied. Finally, virtual pitch synthesis was found to be an efficient way to increase perceived source size without adding energy at frequencies below 200 Hz, adhering to hardware limitations, reducing low-frequency noise pollution, and maintaining compliance with UN R138 regulations. Spatial decorrelation proved difficult to implement given the physical restraints of the speaker placements relative to the listening position, which made the perceptual differences between the different spatially distributed stimuli small.

8

Appendix 1

8.1 Paired Comparison Matrices

The following section presents the paired-comparison matrices for each tested category. The cell values indicate the absolute number of times the stimulus in a given row was chosen over the stimulus in the corresponding column. To highlight the performance of the stimuli, statistical significance ($p < 0.05$) is denoted by a bordered box in the lower-left triangle of each matrix. The second-to-last column (Sum) displays the row totals, representing the total number of wins. Finally, the rightmost column (%) shows the win rate, calculated by taking the sum and dividing it with the number of comparisons the stimuli was included in.

8.1.1 Test I

Table 8.1: Result matrix for the Spatial Decorrelation category. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test I: Spatial Decorrelation	ISD	ISD1	ISD2	ISD3	ISD4	ISD5	Sum	%
Mono point source in front of the listener playing pink noise.	ISD1		4	1	2	1	8	11.8
Mono point source in front of the listener playing a 150 Hz tone with three upper harmonics (300, 450, and 600 Hz).	ISD2	13		11	0	1	25	36.8
De-correlated pink noise simulated as being played through two sources with an azimuth of 18.43° and −18.43° respectively.	ISD3	16	6		4	3	29	42.7
Reference tones simulated as being played through a point source with added synthetic early reflections.	ISD4	15	17	13		5	50	73.5
Combination of stimuli ISD3 and ISD4.	ISD5	16	16	14	12		58	85.3

Table 8.2: Result matrix for the Modulation category. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test I: Modulation	IM	IM1	IM2	IM3	IM4	IM5	Sum	%
Unmodulated reference tone.	IM1		8	3	7	5	23	33.8
0.6 Hz modulation.	IM2	9		7	7	7	30	44.1
8 Hz modulation.	IM3	14	10		9	10	43	63.2
16 Hz modulation.	IM4	10	10	8		10	38	55.9
30 Hz modulation.	IM5	12	10	7	7		36	52.9

Table 8.3: Result matrix for the Pitch & Modulation Shift category. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test I: Pitch & Modulation Shift	IPM	IPM1	IPM2	IPM3	IPM4	Sum	%
Unmodulated reference pitch shift.	IPM1		5	3	10	18	35.3
0.1 Hz increased modulation per km/h.	IPM2	14		14	14	42	82.4
0.4 Hz increased modulation per km/h.	IPM3	12	3		11	26	51.0
0.8 Hz increased modulation per km/h.	IPM4	7	6	3		16	31.4

Table 8.4: Result matrix for the Pitch Shift AVAS category. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test I: Pitch Shift AVAS	IPSA	IPSA1	IPSA2	IPSA3	IPSA4	Sum	%
AVAS of passenger car.	IPSA1		8	3	7	18	35.3
ICE truck acceleration.	IPSA2	9		3	6	18	35.3
AVAS of electric truck (B)	IPSA3	14	14		14	42	82.4
AVAS of electric truck (A)	IPSA4	10	11	3		24	47.1

Table 8.5: Result matrix for the Spectral Centroid category. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test I: Spectral Centroid	ISC	ISC1	ISC2	ISC3	Sum	%
Centroid located at 2437 Hz.	ISC1		1	14	15	44.1
Centroid located at 1853 Hz.	ISC2	16		15	31	91.2
Centroid located at 3264 Hz.	ISC3	3	2		5	14.7

Table 8.6: Result matrix for the Spectral Centroid - Random Phase category. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test I: Spectral Centroid - Random Phase	ISC	ISC4	ISC5	ISC6	Sum	%
Centroid located at 2437 Hz - Random phase.	ISC4		2	16	18	53.0
Centroid located at 1853 Hz - Random phase.	ISC5	15		16	31	91.2
Centroid located at 3264 Hz - Random phase.	ISC6	1	1		2	5.9

Table 8.7: Result matrix for the Spectral Centroid category. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test I: Spectral Centroid	ISC	ISC2	ISC5	ISC6	ISC3	Sum	%
Centroid located at 1853 Hz.	ISC2		10	14	16	40	78.4
Centroid located at 1853 Hz - Random phase.	ISC5	7		15	11	33	64.7
Centroid located at 3264 Hz - Random phase.	ISC6	3	2		3	8	15.7
Centroid located at 3264 Hz.	ISC3	1	6	14		21	41.2

8.1.2 Test II

Table 8.8: Result matrix for the Stereo Correlated category. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test II: Mono	M	M1	M2	M3	M4	Sum	%
Band-passed pink noise (210–600 Hz).	M1		9	5	3	17	17.7
M1 + tonal components (210 Hz and 420 Hz).	M2	20		7	6	33	34.8
M1 + synthetic 30 Hz missing fundamental (MF) complex.	M3	24	22		10	56	58.3
Stochastic MF (5 Hz noise bands extracted at M3 harmonic center frequencies).	M4	26	23	19		68	70.8

Table 8.9: Result matrix for the Mono category. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test II: Stereo	S	S1	S2	S3	S4	Sum	%
Dual-speaker correlated pink noise (210–600 Hz).	S1		10	1	1	12	12.5
Primary-anchored tonal components (210 and 420 Hz) + S1.	S2	19		5	7	31	32.3
Primary-anchored 30 Hz missing fundamental (MF) complex + S1.	S3	28	24		11	63	65.6
Dual-speaker correlated stochastic MF (5 Hz noise bands).	S4	28	22	18		68	70.8

Table 8.10: Result matrix for the Stereo Decorrelated category. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test II: Stereo Decorrelated	D	D1	D2	D3	D4	Sum	%
Dual-speaker decorrelated pink noise (210–600 Hz).	D1		12	0	0	12	13.5
Primary-anchored tonal components (210 and 420 Hz) + D1.	D2	17		5	3	25	26.0
Primary-anchored 30 Hz missing fundamental (MF) complex + D1.	D3	29	24		10	63	65.1
Dual-speaker decorrelated stochastic MF (5 Hz noise bands).	D4	29	26	19		74	76.6

Table 8.11: Result matrix for the spatial configuration of Pink Noise. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test II: Mono vs Stereo - Pink Noise	MSP	M1	S1	D1	Sum	%
Band-passed pink noise (210–600 Hz).	M1		12	17	29	50.0
Dual-speaker correlated pink noise (210–600 Hz).	S1	17		16	33	56.9
Dual-speaker decorrelated pink noise (210–600 Hz).	D1	12	13		25	43.1

Table 8.12: Result matrix for the spatial configurations of Tone. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test II: Mono vs Stereo - Tone	MST	M2	S2	D2	Sum	%
Band-passed pink noise + tonal components.	M2		12	9	21	36.2
Dual-speaker correlated pink noise + Primary-anchored tonal components.	S2	17		15	32	55.2
Dual-speaker decorrelated pink noise + Primary-anchored tonal components.	D2	20	14		34	58.6

Table 8.13: Result matrix for the spatial configurations of 30 Hz Missing Fundamental. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test II: Mono vs Stereo - MF	MSMF	M3	S3	D3	Sum	%
Band-passed pink noise + synthetic 30 Hz missing fundamental complex.	M3		23	16	39	67.2
Dual-speaker correlated pink noise + Primary-anchored synthetic 30 Hz missing fundamental complex.	S3	6		7	13	22.4
Dual-speaker decorrelated pink noise + Primary-anchored synthetic 30 Hz missing fundamental complex.	D3	13	22		35	60.3

Table 8.14: Result matrix for the spatial configurations of Extracted MF. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test II: Mono vs Stereo - Extracted	MSE	M4	S4	D4	Sum	%
Stochastic MF (5 Hz noise bands extracted at the 30 Hz harmonic center frequencies).	M4		21	22	43	74.1
Dual-speaker correlated stochastic MF (5 Hz noise bands).	S4	8		14	22	37.9
Dual-speaker decorrelated stochastic MF (5 Hz noise bands).	D4	7	15		22	37.9

Table 8.15: Result matrix for the Pedestrian Position. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test II: Pedestrian Position	PP	PP1	PP2	PP3	PP4	Sum	%
Band-passed pink noise (210–600 Hz).	PP1		5	4	4	13	14.9
PP1 + synthetic 30 Hz missing fundamental complex.	PP2	24		10	17	51	58.6
PP2 + Added offset speaker playing correlated pink noise.	PP3	25	19		15	59	67.8
PP2 + Added offset speaker playing de-correlated pink noise.	PP4	25	12	14		51	58.6

Table 8.16: Result matrix for the Pitch & Modulation Shift category. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test II: Pitch & Modulation Shift	PM	PM1	PM2	PM3	PM4	PM5	PM6	Sum	%
Pitch-shifting baseline AVAS (A).	PM1		7	5	1	4	4	21	14.5
PM1 + dynamic AM.	PM2	22		10	5	6	7	50	34.5
Pitch-shifting 30 Hz MF complex (M3).	PM3	24	19		13	14	14	84	57.9
PM3 + dynamic AM.	PM4	28	24	16		11	12	91	62.8
Pitch-shifting stochastic MF (M4).	PM5	25	23	15	18		9	90	62.1
PM5 + dynamic AM.	PM6	25	22	15	17	20		99	68.3

Table 8.17: Result matrix for the AVAS Proof of Concept. Cell values indicate the absolute number of times the row stimulus was chosen over the column stimulus. A bordered box in the lower-left triangle indicates statistical significance ($p < 0.05$). The second-to-last column (Sum) displays the total wins, and the final column (%) shows the overall dominance.

Test II: AVAS Proof of Concept	PC	PC1	PC2	PC3	PC4	PC5	PC6	Sum	%
Baseline AVAS (A).	PC1		6	3	23	15	1	48	33.1
PC1 + synthetic 30 Hz missing fundamental (MF) complex.	PC2	23		16	27	23	7	96	66.2
Stochastic MF (5 Hz noise bands extracted from PC1).	PC3	26	13		25	22	5	91	62.8
ICE truck recording.	PC4	6	2	4		6	0	18	12.8
Band-passed ICE truck recording (210–600 Hz).	PC5	14	6	7	23		3	53	36.6
PC3 + synthetic 30 Hz MF complex.	PC6	28	22	24	29	26		129	88.6

Bibliography

[1] Jorge P. Arenas. "On the impact of electric vehicle transition on urban noise pollution". en. In: *Current Opinion in Environmental Science & Health* 45 (June 2025), p. 100623. ISSN: 24685844. DOI: 10.1016/j.coesh.2025.100623. URL: https://linkinghub.elsevier.com/retrieve/pii/S2468584425000327 (visited on 04/19/2026).

[2] Dag Glebe, Juan Parra, and Kerstin Persson Waye. "Replacing diesel buses with electric buses reduced residential low frequency noise". en. In: *Transportation Research Part D: Transport and Environment* 137 (Dec. 2024), p. 104516. ISSN: 13619209. DOI: 10.1016/j.trd.2024.104516. URL: https://linkinghub.elsevier.com/retrieve/pii/S1361920924004735 (visited on 04/20/2026).

[3] "Incidence Rates of Pedestrian and Bicyclist Crashes by Hybrid Electric Passenger Vehicles: An Update1". en. In: *Annals of Emergency Medicine* 64.2 (Aug. 2014), pp. 195–196. ISSN: 01960644. DOI: 10.1016/j.annemergmed.2014.05.017. URL: https://linkinghub.elsevier.com/retrieve/pii/S0196064414004223 (visited on 05/07/2026).

[4] Jim Motavalli. *Anti-Noise Activists Oppose Sounds for Electric Cars*. en. Section: Autos. URL: https://archive.nytimes.com/wheels.blogs.nytimes.com/2010/06/21/anti-noise-activists-oppose-sounds-for-electric-cars/ (visited on 04/24/2026).

[5] Leon Müller, Jens Forssén, and Wolfgang Kropp. "Effects of low-level electric vehicle noise on attention, electrodermal activity, workload, and annoyance". en. In: *The Journal of the Acoustical Society of America* 159.1 (Jan. 2026), pp. 285–299. ISSN: 1520-8524. DOI: 10.1121/10.0042190. URL: https://pubs.aip.org/jasa/article/159/1/285/3377322/Effects-of-low-level-electric-vehicle-noise-on (visited on 06/09/2026).

[6] "Acoustic Vehicle Alerting Systems (AVAS) for Electric Trucks: Initial Thoughts for the Requirements". In: *DAGA 2024 (Annual Conference of the German Acoustical Society)*.

[7] *UN Regulation No. 138 - Uniform provisions concerning the approval of Quiet Road Transport Vehicles*. Tech. rep. United Nations Economic Commission for Europe (UNECE).

[8] A. Fortino and J. Kotzian. "Design guidelines for improved AVAS application using sound radiation analysis". en. In: *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*. Turin, Italy: European Acoustics Association, Jan. 2024, pp. 1063–1066. ISBN: 978-88-88942-

67-4. DOI: 10.61782/fa.2023.0628. URL: https://dael.euracoustics.org/confs/landing_pages/fa2023/000628.html (visited on 03/08/2026).

[9] Martin D. Vestergaard et al. "Auditory size-deviant detection in adults and newborn infants". en. In: *Biological Psychology* 82.2 (Oct. 2009), pp. 169–175. ISSN: 03010511. DOI: 10.1016/j.biopsycho.2009.07.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0301051109001367 (visited on 04/19/2026).

[10] Neville H. Fletcher and Thomas D. Rossing. *The Physics of Musical Instruments*. en. New York, NY: Springer New York, 1991. ISBN: 978-0-387-94151-6 978-1-4612-2980-3. DOI: 10.1007/978-1-4612-2980-3. URL: http://link.springer.com/10.1007/978-1-4612-2980-3 (visited on 04/26/2026).

[11] Udo Zölzer. "DAFX: Digital Audio Effects". en. In: ().

[12] Ralph Van Dinther and Roy D. Patterson. "Perception of acoustic scale and size in musical instrument sounds". en. In: *The Journal of the Acoustical Society of America* 120.4 (Oct. 2006), pp. 2158–2176. ISSN: 0001-4966, 1520-8524. DOI: 10.1121/1.2338295. URL: https://pubs.aip.org/jasa/article/120/4/2158/982645/Perception-of-acoustic-scale-and-size-in-musical (visited on 01/20/2026).

[13] Hugo Fastl and Eberhard Zwicker. *Psychoacoustics*. en. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. ISBN: 978-3-642-51765-5 978-3-540-68888-4. DOI: 10.1007/978-3-540-68888-4. URL: http://link.springer.com/10.1007/978-3-540-68888-4 (visited on 01/21/2026).

[14] Katrin Krumbholz, Roy D. Patterson, and Daniel Pressnitzer. "The lower limit of pitch as determined by rate discrimination". en. In: *The Journal of the Acoustical Society of America* 108.3 (Sept. 2000), pp. 1170–1180. ISSN: 0001-4966, 1520-8524. DOI: 10.1121/1.1287843. URL: https://pubs.aip.org/jasa/article/108/3/1170/551134/The-lower-limit-of-pitch-as-determined-by-rate (visited on 04/06/2026).

[15] Daniel Pressnitzer, Roy D. Patterson, and Katrin Krumbholz. "The lower limit of melodic pitch". en. In: *The Journal of the Acoustical Society of America* 109.5 (May 2001), pp. 2074–2084. ISSN: 0001-4966, 1520-8524. DOI: 10.1121/1.1359797. URL: https://pubs.aip.org/jasa/article/109/5/2074/549172/The-lower-limit-of-melodic-pitch (visited on 04/06/2026).

[16] Johannes Kasbach et al. "Assessing and modeling apparent source width perception". en. In: (2016).

[17] Tim Ziemer. *Psychoacoustic Music Sound Field Synthesis: Creating Spaciousness for Composition, Performance, Acoustics and Perception*. en. Vol. 7. Current Research in Systematic Musicology. Cham: Springer International Publishing, 2020. ISBN: 978-3-030-23032-6 978-3-030-23033-3. DOI: 10.1007/978-3-030-23033-3. URL: http://link.springer.com/10.1007/978-3-030-23033-3 (visited on 04/28/2026).

[18] Gary S. Kendall. "The Decorrelation of Audio Signals and Its Impact on Spatial Imagery". en. In: *Computer Music Journal* 19.4 (1995), p. 71. ISSN: 01489267. DOI: 10.2307/3680992. URL: https://www.jstor.org/stable/3680992?origin=crossref (visited on 04/28/2026).

[19] Guillaume Potard and Ian Burnett. “A study on sound source apparent shape and wideness”. en. In: (2003).

[20] Angelo Farina. “Advancements in impulse response measurements by sine sweeps”. en. In: (2007).

[21] Florian Wickelmaier and Christian Schmid. “A Matlab function to estimate choice model parameters from paired-comparison data”. en. In: *Behavior Research Methods, Instruments, & Computers* 36.1 (Feb. 2004), pp. 29–40. ISSN: 0743-3808, 1532-5970. DOI: 10.3758/BF03195547. URL: http://link.springer.com/10.3758/BF03195547 (visited on 04/29/2026).

[22] B Bernschütz. “A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU100”. en. In: (2013).

[23] Berkay Kullukcu, Robert Rosenkranz, and M Ercan Altinsoy. “Acoustic Vehicle Alerting Systems (AVAS) for Electric Trucks: Initial Thoughts for the Requirements”. en. In: (2024).

[24] Ziqin Lan et al. “Study of the Traffic Noise Source Emission Model and the Frequency Spectrum Analysis of Electric Vehicles on Urban Roads in China”. en. In: *Acta Acustica united with Acustica* 104.6 (Nov. 2018), pp. 989–998. ISSN: 1610-1928. DOI: 10.3813/AAA.919265. URL: https://dael.euracoustics.org/landing_pages/aaua/52599.html (visited on 05/20/2026).

[25] Kaigham J. Gabriel and H. Steven Colburn. “Interaural correlation discrimination: I. Bandwidth and level dependence”. en. In: *The Journal of the Acoustical Society of America* 69.5 (May 1981), pp. 1394–1401. ISSN: 0001-4966, 1520-8524. DOI: 10.1121/1.385821. URL: https://pubs.aip.org/jasa/article/69/5/1394/774073/Interaural-correlation-discrimination-I-Bandwidth (visited on 05/21/2026).

DEPARTMENT OF ARCHITECTURE AND CIVIL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se



CHALMERS
UNIVERSITY OF TECHNOLOGY