



CHALMERS
UNIVERSITY OF TECHNOLOGY

Auralization of electric vehicles for the perceptual evaluation of acoustic vehicle alerting systems

Downloaded from: <https://research.chalmers.se>, 2024-09-20 03:29 UTC

Citation for the original published paper (version of record):

Müller, L., Kropp, W. (2024). Auralization of electric vehicles for the perceptual evaluation of acoustic vehicle alerting systems. *Acta Acustica*, 8(27). <http://dx.doi.org/10.1051/aacus/2024025>

N.B. When citing this work, cite the original published paper.



Auralization of electric vehicles for the perceptual evaluation of acoustic vehicle alerting systems

Leon Müller*  and Wolfgang Kropp 

Division of Applied Acoustics, Chalmers University of Technology, 41296 Gothenburg, Sweden

Received 2 February 2024, Accepted 11 June 2024

Abstract – Electric vehicles (EVs) typically emit little noise at low driving speeds, which increases the risk of accidents for vulnerable road users such as pedestrians. To reduce this risk, regulations demand that newly sold EVs have to be equipped with an acoustic vehicle alerting system (AVAS), which radiates artificial warning sounds. Developing AVAS sounds that provide a sufficient warning capability while limiting traffic noise annoyance requires laboratory listening experiments; such experiments need accurate auralization methods. Even though several auralization tools are already established in the research field, those frameworks require additional data to simulate EVs. This paper presents an electric vehicle auralization toolchain combined with an open-access database, including AVAS measurements, synthesis algorithms, and numerically calculated sound source directivities for three different electric passenger cars. The auralization method was validated numerically and in a listening experiment, comparing simulated EV passages to binaural in-situ recordings. The results of this perceptual validation indicate that stimuli generated with the presented method are perceived as slightly less plausible than in-situ recordings and that they result in a similar distribution of annoyance ratings but a higher perceived vehicle velocity compared to the reference recordings.

Keywords: Auralization, AVAS, Electric vehicles, Perceptual evaluation, Plausibility

1 Introduction

Electric vehicles (EVs) have become increasingly popular in the last few years, especially in urban environments [1]. EVs typically radiate less sound at low driving speeds than internal combustion engine vehicles [2, 3]. This reduced noise emission might, in general, be beneficial as it is known that high road traffic noise levels can cause serious health problems, such as chronic annoyance, sleep disturbance, and cardiovascular diseases [4]. However, the low sound emission of electric vehicles also increases the risk of accidents for pedestrians and vulnerable road users, such as the visually impaired, who can not localize approaching EVs using acoustic cues [5]. To reduce this risk, recently implemented regulations demand that all newly produced electric vehicles have to be equipped with an acoustic vehicle alerting system (AVAS) [6], i.e., a loudspeaker that radiates artificial warning sounds indicating the vehicle's location and driving speed. In the EU, these regulations specify minimum and maximum AVAS sound levels that the EV must comply with for driving speeds of up to approximately 20 km/h. Additionally, the AVAS signal should cover at least two third-octave bands, have content within or below the 1600 Hz third-octave band, and at least one tone should

shift in frequency proportional to vehicle speed by an average of at least 0.8% per 1 [7]. Regarding the sound type, the regulation states that “*the sound shall be similar to the sound of a vehicle of the same category equipped with an internal combustion engine*”, allowing manufacturers to design various vehicle-specific AVAS sounds.

A challenge in designing these AVAS sounds is to achieve a sufficient warning effect while simultaneously limiting the negative impact on the acoustic environment. Therefore, a fundamental requirement for developing efficient AVAS systems is to understand which sound characteristics are relevant to localize a vehicle acoustically and how those sounds affect bystanders. A common approach to studying such psychoacoustic effects is to conduct listening experiments in controlled acoustic environments. Such studies require an accurate reproduction of the acoustic scenes of interest, a process that is also referred to as auralization [8]. In the last decades, several different auralization models for internal combustion engine vehicles have been proposed. For example, passenger car and heavy-vehicle engine noise has been successfully synthesized using a sample-based synthesis approach, i.e., by deconstructing engine noise recordings into short grains that represent the sound emitted during one engine cycle and recombining those samples using sophisticated algorithms [9, 10]. Other researchers used a spectral modeling approach, i.e., adding

*Corresponding author: leon.mueller@chalmers.se

different sinusoids and filtering complex broadband source signals, to model both tire and propulsion noise [11]. Additionally, different auralization approaches specifically for tire/road noise exist, utilizing both statistical engineering methods and detailed numerical models [12, 13]. While some aspects of those established methods, such as the outdoor sound propagation modeling, could be directly applied to electric vehicle auralizations, the broad range of different AVAS signals makes it difficult to use one of those methods as a complete “one-fits-all” solution for AVAS signal synthesis.

The limited number of existing AVAS-related psychoacoustic studies often relied on auralization methods based on static recordings of EV passages obtained by, e.g., placing a microphone array on the side of a test track [14]. This approach generally results in realistic stimuli since it automatically includes complex acoustic properties, such as source directivity and sound propagation path. However, performing real-life recordings can be time- and resource-intensive and does not provide the flexibility of changing parameters such as vehicle speed or the AVAS signal after conducting the measurement. Other researchers combined recordings of microphones placed on a moving vehicle [15] and synthesized AVAS signals [16] with a general auralization model, resulting in more flexible simulations. Nevertheless, those auralizations do not necessarily include an accurate source directivity model, which can be expected to be perceptually relevant, especially for recently developed directional AVAS systems [17]. Even though several commercial and open-source auralization frameworks for rendering virtual acoustic scenes are already available [18–21], only few AVAS signals or EV directivity models are currently openly accessible to the scientific community, limiting the possibilities for researchers to conduct AVAS-related psychoacoustic studies. Additionally, most existing tire/road noise auralization methods are designed for velocities above 25 km/h, as combustion engine noise is typically considered dominating at lower speeds [22]. However, electric vehicles radiate significantly less motor noise than internal combustion engine vehicles; hence, even low-speed rolling noise can become audible and should, therefore, be included in the auralization.

This work offers resources and methods for researchers to conduct listening experiments related to Acoustic Vehicle Alerting Systems. It introduces an open-access database containing AVAS and tire-noise recordings, along with AVAS synthesis models for three electric vehicles. These models are paired with an auralization framework designed for headphone-based reproduction. Thereby, AVAS and tire/road noise are included, but the model omits propulsion and wind noise. While the presented methods can be used to model changes in velocity, acceleration-specific effects, such as increased tire/road friction forces, are not considered. The paper begins by outlining the measurement process and examining the characteristics of the recorded AVAS and tire/road noise signals in Section 2. Building on these measurements, Section 3 describes techniques for synthesizing AVAS and tire/road noise signals, including methodologies for modeling their radiation directivities

and sound propagation. In Section 4, these methods are applied to re-create measured vehicle passages, allowing for numerical validations through comparisons with the reference recordings. Finally, Section 5 presents an assessment of the perceptual quality of the auralization results through a laboratory listening experiment.

2 Measurements

In order to obtain reference data for the auralization model, pass-by measurements of three different electric vehicles were conducted: a Tesla Model Y 2021 (vehicle A, Fig. 1a), a Volkswagen ID.3 Pro Performance 2021 (vehicle B, Fig. 1b) and a Nissan Leaf 2018 (vehicle C, Fig. 1c). These vehicles can be classified as small to medium-sized electric passenger cars, which all utilize a single AVAS loudspeaker mounted in the front bumper. All vehicles were equipped with radial non-studded winter tires with an external rolling noise value of 72 dB according to EU regulation 2020/740 [23].

2.1 Methods

AVAS signals of each vehicle were measured by placing a microphone in front of the AVAS loudspeaker with approximately 10 cm distance as shown in Figure 1d. Additionally, a second microphone was mounted at 40 cm distance perpendicular to the tire as shown in Figure 1e to record isolated tire/road noise. The sound pressure at a stationary observer position was measured by placing a HEAD acoustics HMS V artificial head and a microphone at the roadside, as shown in Figure 1f. All microphones were omnidirectional, free-field equalized, and of the type GRAS 46AE with 90 mm foam windscreens. Vehicle velocity and position relative to the roadside observer were recorded via GPS using HEAD acoustics SQadriga III and SQobold data acquisition systems, allowing for the exact reproduction of the recorded scenarios and the comparison to their simulated counterparts. The measurements were conducted with a sampling frequency of 48 kHz and on a road with dense asphalt concrete surface under dry and windless conditions. Several passages with different constant and accelerating velocity profiles up to 30 km/h were recorded for each vehicle, driving both forward and backward. All measured data is openly accessible at [24].

2.2 Measurement results

The following section presents the measured AVAS signals for the three evaluated vehicles as well as the results of the tire/road noise recordings.

2.2.1 AVAS signals

Figure 2 shows the measurement results of the microphone placed in front of the AVAS loudspeaker for all three vehicles. For these plots, the measured time signals were downsampled to a sampling rate of 12 kHz and divided into overlapping blocks of 512 samples. Each block was assigned

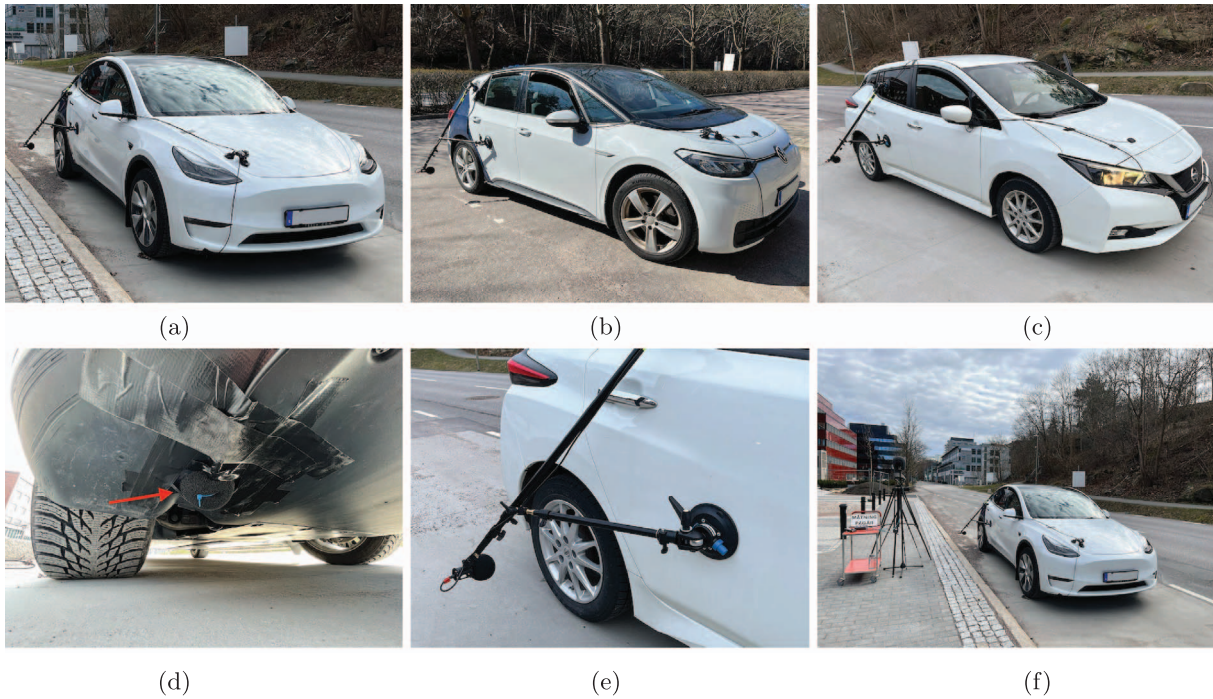


Figure 1. Vehicle A – Tesla Model Y 2021 (a), vehicle B – Volkswagen ID.3 Pro Performance 2021 (b), vehicle C – Nissan Leaf 2018 (c), microphone placed in front of vehicle A AVAS speaker (d), microphone placed in front of vehicle C tire (e), and measurement setup on roadside (f).

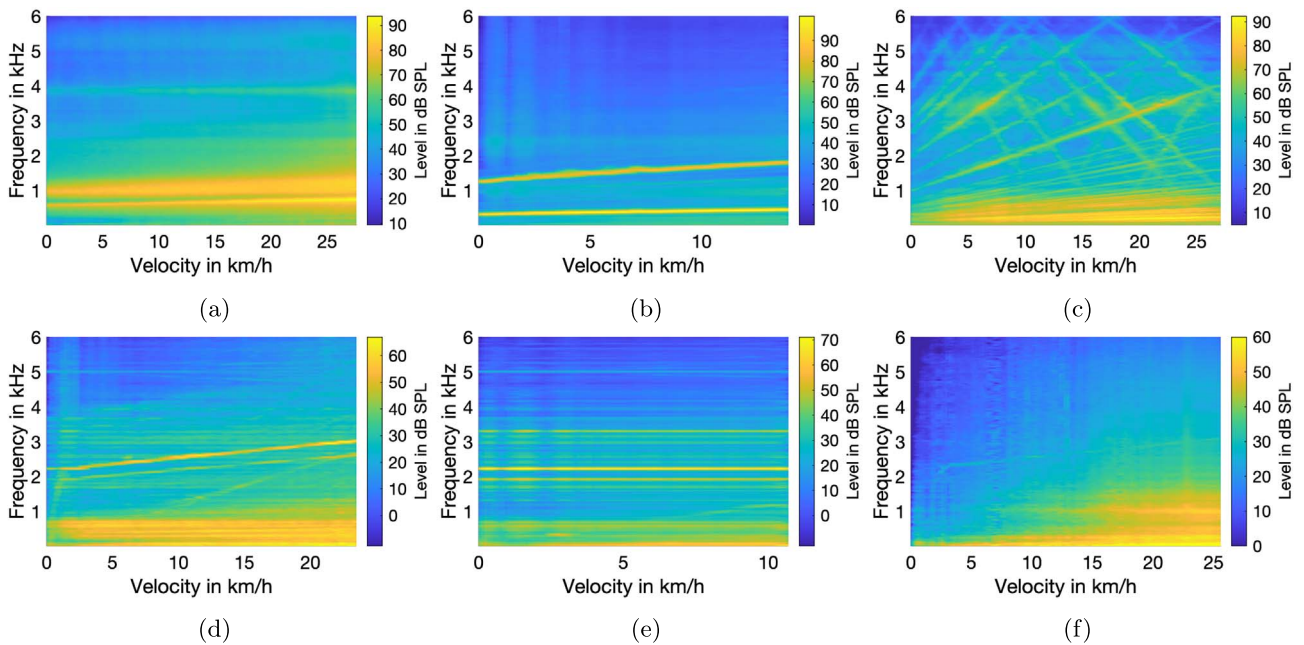


Figure 2. Measured velocity-dependent magnitude spectra $|H(f, v)|$ from microphones mounted on moving vehicle: vehicle A forward AVAS (a), vehicle A backward AVAS (b), vehicle B forward and backward AVAS (c), vehicle C forward AVAS (d), vehicle C backward AVAS (e) and vehicle C tire/road noise (f).

a velocity value based on the GPS recordings, transformed into a magnitude spectrum, and the spectra for blocks with similar velocity values were averaged. This results in

frequency over velocity plots, which visualize the characteristic velocity dependency of the AVAS signals. Thereby, the three measured electric vehicles differ quite substantially, as

described in the following. We recommend listening to the example sounds provided at [24] for a better understanding of the different signals.

Vehicle A As shown in Figure 2a, vehicle A radiates a band-pass filtered noise in combination with a lower, more narrow-band component when driving forward. The center frequency, as well as the bandwidth of both components, increases with vehicle velocity. When driving backward, vehicle A radiates two highly tonal components, which also increase in frequency with velocity (cf. Fig. 2b). Listening to the recorded signal reveals a clearly audible amplitude modulation, which is also visible in the form of sidebands when analyzing the corresponding frequency over time spectrogram (not shown here for reasons of conciseness).

Vehicle B Vehicle B radiates the same AVAS sound in both driving directions, consisting of a large number of different tonal components which partly seem to have a harmonic relation to each other as shown in Figure 2c. Most of those tones increase in frequency when accelerating, others decrease in frequency, and some are strongly amplitude-modulated. Of all measured AVAS signals, we perceive vehicle B as the most “chaotic”; it is also the only signal that contains strong tonal components above 3 kHz. Overall, the signal reminds us more of a science-fiction spaceship sound than the noise expected of a medium-sized passenger car. Whether or not this perceptual discrepancy has consequences for factors such as annoyance or localization accuracy compared to more conservative AVAS sounds is an example for a future study that could be performed using the methods presented in this paper.

Vehicle C The forward AVAS of vehicle C consists of both high-frequency tonal components and lower-frequency band-pass filtered noise components (cf. Fig. 2d). Compared to, for example, vehicle B, individual tonal components are perceptually not as pronounced and partly masked by background noise. When driving backward, vehicle C emits a recurring “pling” sound with a duration of 1 s per repetition that appears to be independent of the vehicle velocity as shown in Figure 2e.

Summarized, one can divide the evaluated AVAS signals into three categories: signals that mainly consist of a number of tones, signals that mainly consist of band-pass filtered noise components, and signals that consist of a repeating sound. Section 3.1 presents methods to synthesize those three different signal types for arbitrary vehicle velocities based on the presented reference measurements.

2.2.2 Tire/road noise

Figure 2f shows the measured tire/road noise spectrum for vehicle C. The results for vehicles A and B were found to be very similar and are, for reasons of conciseness, not presented here. It can be seen that the spectrogram contains a pronounced pressure maximum centered around 1 kHz above a velocity of approximately 15 km/h. This pressure maximum is characteristic of tire/road noise [25] and

increases in amplitude as well as slightly broadens in frequency range for higher vehicle speeds. Additionally, the spectrogram shows strong low-frequency components that increase in amplitude and upper-frequency limits for higher velocities. These components are assumed to correspond to wind-induced noise in the microphone [26], which are, due to the microphone positioning, expected to be more pronounced for the tire noise measurements than for the AVAS measurements where the car body partly shielded the microphone.

3 Auralization

The first step in auralizing a vehicle passage is to accurately synthesize all relevant source signals, which are, for the scope of this paper, limited to the AVAS signal as well as tire/road noise. As the previously described measurements showed, the three evaluated vehicles use substantially different AVAS signal types, requiring the implementation of different synthesis techniques as described in Section 3.1. Besides the source signal, the radiation characteristics of the different sources need to be taken into account, e.g., the sound radiation from an AVAS loudspeaker mounted to the front bumper of a car will most likely be strongest in the forward direction, and tire/road noise typically shows a strong, frequency dependent, amplification perpendicular to the contact surface due to the horn effect [27]. Section 3.2 presents a boundary element approach to numerically estimate the AVAS radiation directivity for all three evaluated vehicles and uses previously performed tire radiation measurements to construct a generic tire/road noise directivity pattern. All source directivities are encoded into spherical harmonic coefficients to simplify the subsequent processing. Once all source signals and radiation directivities are known, the sound propagation from all sources to a receiver position is modeled using spherical harmonic extrapolation while additionally taking into account vehicle movement, air attenuation, and binaural hearing as described in Section 3.3. The entire auralization model was implemented in Matlab R2023a; all relevant code is available under MIT license at <https://github.com/leonpaulmueller/evat>.

3.1 AVAS and tire/road noise synthesis

Based on the evaluation of the measured reference signals described in Section 2, we differentiate between three different synthesis methods for generating AVAS and tire/road noise signals: (i) subtractive synthesis, where broadband noise is filtered with vehicle velocity-dependent band-pass filters, (ii) additive synthesis, where sine wave oscillators are added up and modulated in amplitude and frequency depending on the vehicle velocity and (iii) sample-based synthesis, where a sample, i.e. a short sound recording, gets repeatedly played back and modified based on the vehicle velocity. The following sections describe methods to analyze the recorded reference signals and re-synthesize them using those three approaches. While this

paper focuses on recreating measured AVAS signals of existing vehicles, the synthesis methods are designed in a way that allows for creating arbitrary new AVAS sounds.

3.1.1 Subtractive synthesis

The subtractive synthesis model used in this paper assumes that the output signal, i.e., the AVAS or tire/road noise, can be described by convolving Gaussian white noise with a set of vehicle velocity-dependent impulse responses. The process of determining these impulse responses and subsequently generating a new signal can be divided into an analysis and a synthesis stage, as described below.

Analysis In order to determine a set of velocity-dependent filter functions describing the characteristics of a given AVAS or tire/road noise recording, the recorded pressure signal is divided into blocks of N samples, and each block is assigned a mean velocity value based on the recorded vehicle velocity signal. Those blocks are then transformed into magnitude spectra and averaged depending on the desired velocity resolution. Magnitude spectra for missing velocity values are linearly interpolated, and a smoothing function is applied to avoid discontinuities between individual velocity values, resulting in the final matrix of velocity-dependent magnitude spectra $|H(f, v)|$ (c.f. Fig. 2).

Synthesis To synthesize a signal for an arbitrary vehicle velocity, the new velocity signal is averaged over blocks of N samples. Since the magnitude spectra $|H(f, v)|$ do not contain any phase information, a minimum phase representation is computed using the real cepstrum [28], resulting in a set of velocity-dependent impulse responses $h(n, v)$. This allows block-wise computing of the output signal by convolving the impulse response corresponding to each block's velocity value with a block of Gaussian white noise. Additionally, the same noise block is convolved with the impulse response corresponding to the previous velocity value, and both results are cross-faded using a raised cosine function to avoid clicking artifacts between blocks.

Results The subtractive synthesis approach was used to re-synthesize the forward AVAS signals of vehicles A and C and all tire/road noise signals. Figure 3 shows an exemplary comparison between the measured and the re-synthesized forward AVAS signal of vehicle A using a window size of $N = 512$ samples at $f_s = 6$ kHz. Thereby, it can be seen that the synthesized signal resembles the measured signal very well up to 3 kHz, which was set as the upper-frequency limit for this synthesis. A perceptual validation of the resulting auralization is presented in Section 5, and both signals can be listened to via the supplementary online repository.

3.1.2 Additive synthesis

Additive synthesis describes the process of adding up multiple sinusoidal signals to create a complex sound [29, 30]. Thereby, the individual signal components are typically generated by individual oscillators, which, independent of

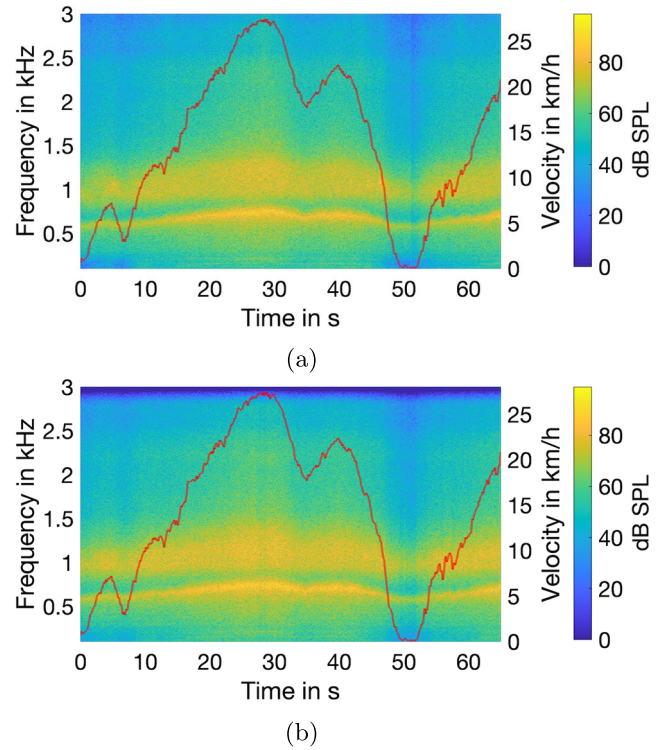


Figure 3. Vehicle A forward AVAS measurement (a) and synthesized forward AVAS signal (b) including measured vehicle velocity profile.

each other, may have a time-varying amplitude or frequency. This can either result in relatively simple sounds, such as the backward AVAS of vehicle A consisting of two tones that increase their pitch with velocity (c.f. Fig. 2b), or in more complex sounds such as the vehicle B AVAS, consisting of a large number of different tones that are partly in harmonic relation to each other (c.f. Fig. 2c). Figure 4 gives a high-level overview of the implemented additive analysis and synthesis model described in the following.

Synthesis We assume that the desired signal consists of the sum of U different simple harmonic oscillators where the u -th oscillator at time sample n has an amplitude of $A_u(n)$ and a frequency $f_u(n)$ that both change depending on the vehicle velocity $v(n)$. Furthermore, the amplitude of each oscillator is modulated by an additional time-varying harmonic oscillation with amplitude $\tilde{A}_u(n)$ and frequency $\tilde{f}_u(n)$. This means that an individual oscillator is characterized by the four parameters $A_u(n)$, $f_u(n)$, $\tilde{A}_u(n)$ and $\tilde{f}_u(n)$, which all depend on the vehicle velocity. For each oscillator, the relation between those parameters and the velocity is described by four sets of polynomial coefficients $C_{u,q}^f$, $C_{u,q}^A$, $C_{u,q}^{\tilde{f}}$ and $C_{u,q}^{\tilde{A}}$ which can either be manually set to design an arbitrary new AVAS signal or can be obtained from analyzing an existing AVAS recording as described in the next paragraph. This means that, for example, the frequency of the u -th oscillator at time sample n can be calculated from the polynomial coefficients $C_{u,q}^f$ with degree Q and the vehicle velocity $v(n)$ as

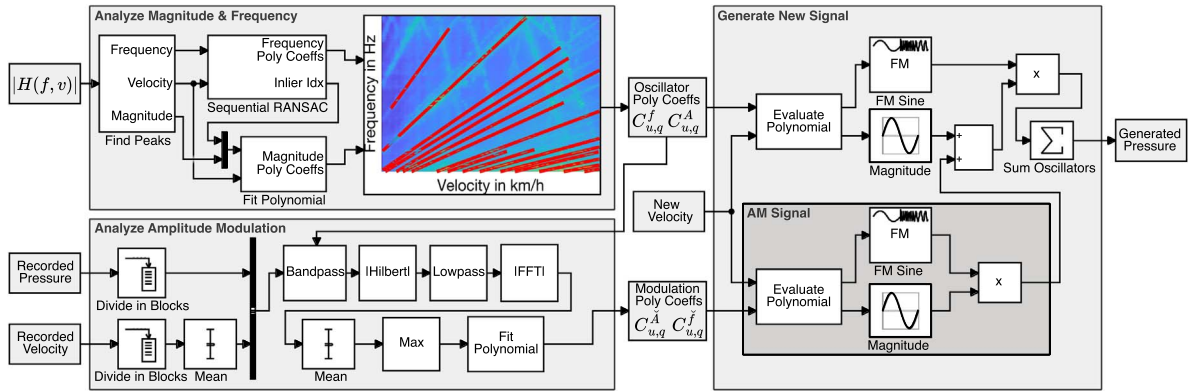


Figure 4. Additive analysis and synthesis model. The red lines in the spectrogram visualize RANSAC results for vehicle B, i.e., amplitude and frequency of individual oscillators. The velocity-dependent magnitude spectrum $|H(f, v)|$ is obtained from the analysis stage of the subtractive synthesis method described in Section 3.1.1.

$$f_u(n) = \sum_{q=0}^Q C_{u,q}^f v(n)^q. \quad (1)$$

Since the modulation signal for each oscillator is expected to change in amplitude and frequency over time and the phase argument of such a time-varying signal is proportional to the integral of the instantaneous frequency [31, Sec. 5.6], the modulation signal $\check{s}_u(n)$ can be constructed as

$$\check{s}_u(n) = \check{A}_u(n) \cos\left(\frac{2\pi}{f_s} \sum_{i=0}^n \check{f}_u(i)\right). \quad (2)$$

The output signal $s(n)$ then corresponds to the sum of all U oscillators, which each consist of a frequency-modulated sinusoidal multiplied with the corresponding amplitude modulation signal:

$$s(n) = \sum_{u=1}^U (A_u(n) + \check{s}_u(n)) \cos\left(\frac{2\pi}{f_s} \sum_{i=0}^n f_u(i)\right). \quad (3)$$

Analysis To determine the coefficient sets $C_{u,q}^f$, $C_{u,q}^A$, $C_{u,q}^{\check{f}}$ and $C_{u,q}^{\check{A}}$ from a recorded AVAS signal, the same velocity-dependent magnitude spectra $|H(f, v)|$ used in Section 3.1.1 are analyzed to find peaks with a user-defined prominence, threshold and inter-peak distance for each velocity value. Those peaks' amplitude and frequency values are then processed using a sequential random sample consensus (RANSAC) approach [32, 33] as visualized in Figure 5. Thereby, N_R random subsets, each containing N_s peak values, are used to fit N_R different polynomials with degree Q in the frequency-velocity plane. The polynomial that covers the most data points within a certain user-defined distance, i.e., corresponds to the model with the most inliers, is selected as frequency over velocity description for the first oscillator, $C_{1,q}^f$. Those inlier points are then removed from the peak data set and used to fit the corresponding magnitude-velocity polynomial coefficients, $C_{1,q}^A$. The remaining data points are re-iterated to find descriptors for all U oscil-

lators. This process results in two sets of polynomials, $C_{u,q}^f$ and $C_{u,q}^A$, describing the velocity-dependent frequency and amplitude behavior of all oscillators. The red lines in the spectrogram shown in Figure 4 visualize those polynomials for the Vehicle B AVAS signal, where each line represents an individual oscillator. As can be seen, the estimated polynomials cover most but not all of the tonal components included in the analyzed AVAS signal.

Additionally, the strength and frequency of the amplitude modulation for each oscillator are analyzed by block-wise band-pass filtering the recorded AVAS signal according to the previously determined frequency for each oscillator. Taking an FFT of the Hilbert envelope of those band-pass filtered signals then allows determining the velocity-dependent amplitude modulation frequency and strength which are then described by the two polynomial sets $C_{u,q}^{\check{f}}$ and $C_{u,q}^{\check{A}}$.

Results Figure 6 shows a time-frequency representation of the measured and additively re-synthesized AVAS signal for vehicle B using $U = 27$ oscillators, a window size of $N = 512$ samples at $f_s = 6$ kHz, frequency-velocity polynomial coefficients $C_{u,q}^f$ with degree $Q_f = 1$ and magnitude-velocity polynomial coefficients $C_{u,q}^A$ with degree $Q_A = 6$. It can be seen that the generated signal successfully reproduced a large number of tonal components contained in the original recording, but not all of them. This lower number of oscillator voices is not a limitation of the synthesis method itself but is caused by the sequential RANSAC approach failing to reliably identify more than 27 different tones in the velocity-dependent magnitude spectrum $|H(f, v)|$ for this specific AVAS recording. This could be improved by either fine-tuning the peak detection and RANSAC settings or by manually adding more harmonics to the coefficient set $C_{u,q}^f$.

Unlike the recorded signal, the synthesized tones show a fast frequency modulation caused by small fluctuations in the measured vehicle velocity signal. This undesired modulation can be avoided by smoothing the recorded velocity signal as it was done for the auralized signals evaluated in

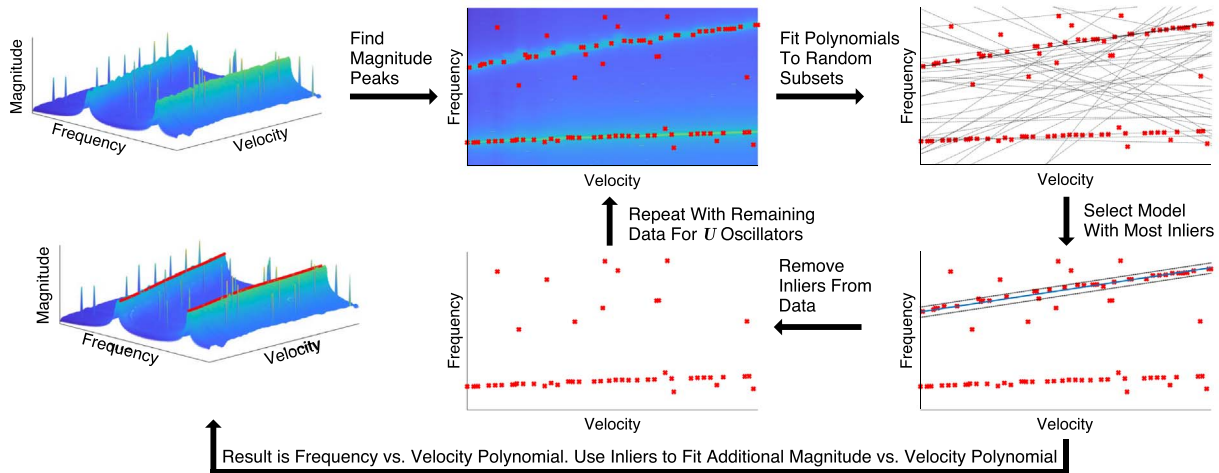


Figure 5. Sequential RANSAC tone detection procedure.

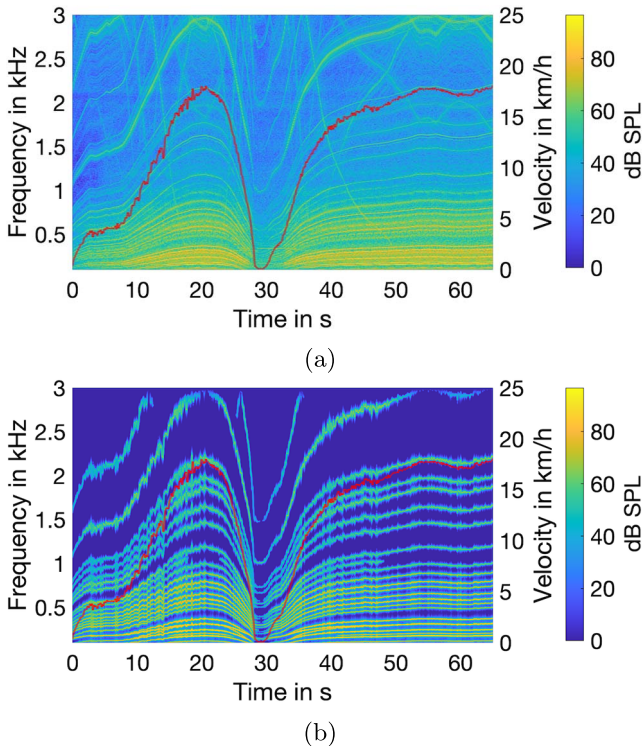


Figure 6. Vehicle B AVAS measurement (a) and additively synthesized AVAS signal (b) including the measured vehicle velocity profile.

Sections 4 and 5. When using generated instead of recorded velocity profiles, this problem does not occur. Therefore, it was decided not to include this velocity smoothing step in the additive synthesis algorithm but instead assume that the input velocity signal has already been pre-processed.

Additionally, the recorded and the generated signals differ in background noise level as the synthesized signal solely consists of pure tones while the recorded signal contains a broadband background noise of around 30 dB. However,

in this specific case, the background noise most likely corresponds to wind-induced noise in the microphone and cross-talk from tire/road noise; hence, it is assumed not to be part of the AVAS signal and should not be included in the source signal synthesis. However, other AVAS signals might contain strong tonal and broadband noise components, requiring a combination of additive and subtractive synthesis.

3.1.3 Sample-based synthesis

In contrast to subtractive and additive synthesis, sample-based synthesis uses a pre-recorded sound, a so-called sample, instead of simple sinusoids or noise as a synthesis foundation. Different variations of sample-based synthesis have been previously used to, for example, auralize combustion engine noise [9, 10], using sophisticated algorithms like granular synthesis or pitch-synchronous overlap-add to modify sample properties such as pitch or time scale. However, for this paper, only the backward AVAS of vehicle C, consisting of a simple “plinging” sound played back repeatedly with a constant pitch, is of interest for a sample-based synthesis approach. Therefore, the implemented method was limited to only modulating sound pressure level and repetition rate depending on the vehicle velocity.

Synthesis The implemented sample-based synthesis model constructs an output signal based on repetitions of a pre-recorded sound sample. Thereby, both the equivalent continuous sound pressure level and the repetition rate are assumed to be velocity dependent and are described by a set of polynomial coefficients, similar to the polynomials used for the additive synthesis approach in Section 3.1.2. To synthesize an AVAS signal for an arbitrary new velocity, the instantaneous repetition rate and sound pressure level are calculated from those polynomials according to equation (1) and then used to distribute the pre-recorded samples in an output signal vector as well as scale them to achieve the desired RMS-values.

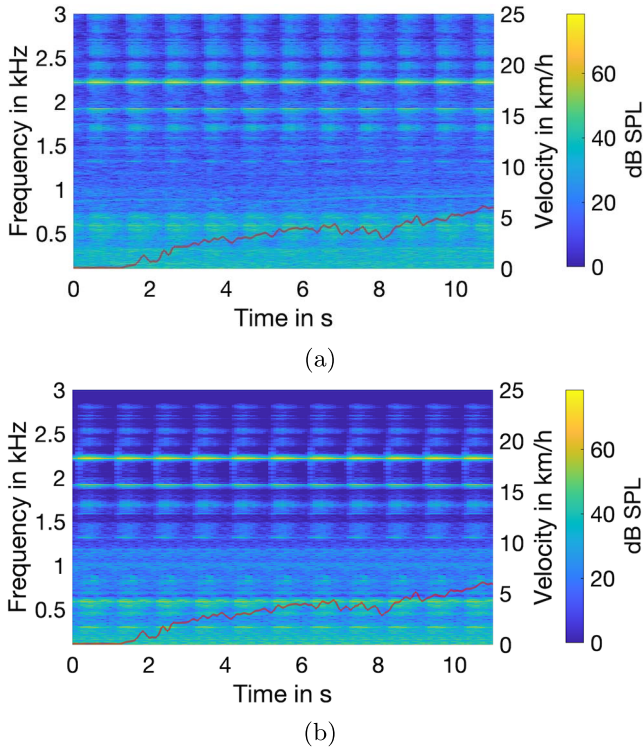


Figure 7. Vehicle B AVAS measurement (a) and additively synthesized AVAS signal (b) including the measured vehicle velocity profile.

Analysis Analyzing a recorded AVAS signal for sample-based synthesis requires manually selecting one period of the desired sound sample in the recording. The entire recording is then analyzed by calculating the cross-correlation between the selected reference signal excerpt and a time-shifted copy of the entire recording to find repetitions of the selected sound. The RMS value of each repetition and the spacing between repetitions is assigned to the corresponding recorded vehicle velocity value, which allows for the fitting of polynomials describing the velocity dependency of both parameters. Signal repetitions correlating strongly with the selected reference sound are then averaged in the time domain to obtain a clean signal sample and suppress potential background noise.

Results Figure 7 shows a comparison between the recorded and the re-synthesized sample-based backward AVAS signal of Vehicle C. Both signals appear to be very similar both in time and frequency structure, apart from the fact that the synthesized signal shows a lower overall background noise level. However, similarly to the additive synthesis results, we assume that the background noise in the recording is rather an artifact than part of the AVAS signal, which means that the lower noise level in the generated signal is beneficial for auralization purposes.

3.1.4 Source signal synthesis conclusion

Based on the characteristics of the measured AVAS and tire/road noise signals, three different source signal

synthesis methods have been implemented. All three methods have been shown to work well for the signals generated by the evaluated vehicles. However, additional fine-tuning might be required when adapting the methods to synthesize other types of signals. Thereby, the choice of the method should, of course, be determined by the character of the signal of interest.

3.2 Directivity

For the auralization approach presented in this paper, we assume that the sound radiated by an electric vehicle is a superposition of two types of sources: tire/road noise and the AVAS warning sound. While the previous section described the synthesis of the corresponding source signals, one also has to consider sound radiation properties, which are expected to be both frequency and space-dependent. Whether or not this spatial directivity is perceptually relevant could be a subject of a future study; to allow such research, this work aimed to reproduce the source directivities as accurately as possible. Therefore, the AVAS radiation patterns for all three vehicles were numerically estimated using the boundary element method (BEM) as described in Section 3.2.1, and the tire/road noise directivity was modeled based on previous measurements (c.f. Section 3.2.2). Both AVAS and tire directivities were encoded using spherical harmonic expansion as described in Section 3.2.3 to allow more straightforward spatial processing and propagation modeling.

3.2.1 AVAS directivity

Methods In order to obtain a numerical estimate of the AVAS radiation directivity, all three vehicles were modeled using the boundary element method (BEM) in Comsol Multiphysics 6.1. The vehicle geometries were based on simplified, commercially available 3D models of the individual cars where the AVAS loudspeaker was substituted by a single disk with a 5 cm radius embedded in the vehicle chassis. This disk was then excited with a velocity proportional to $\frac{1}{j\omega}$, which means that the radiated sound pressure would be constant over all frequencies if the source would be a monopole in free field. An infinite, sound-hard ground was included in the simulation by introducing a symmetry boundary condition. A simple, porous absorber impedance model was assigned to the vehicle floor to avoid numerical problems caused by resonances between the vehicle and the ground. The resulting sound pressure was then evaluated at 5810 points of a 131st order Lebedev sphere [34] with 3 m radius surrounding the vehicle as visualized in Figure 8. Since the complex sound pressure on a surface enclosing all sources is known, this pressure can be extrapolated to any position outside of the sphere [35], in this case by using spherical harmonic expansion as described in Section 3.2.3. Thereby, the fact that the introduced symmetry boundary condition leads to a mirrored pressure field below the ground plane automatically results in correct ground reflections when extrapolating the pressure from the evaluation sphere. This correct extrapolation would

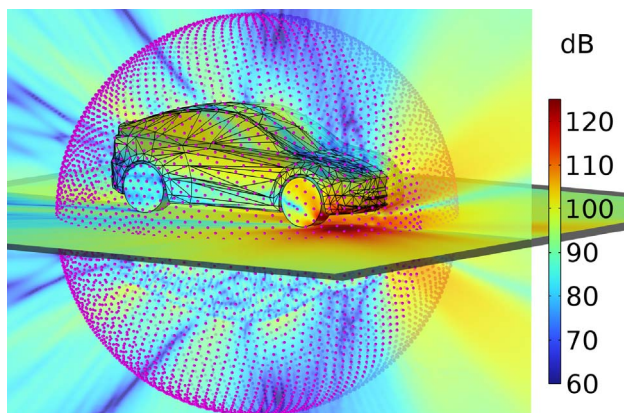


Figure 8. Simplified 3D model of vehicle A with BEM results for radiated sound pressure at $f = 2$ kHz and evaluation points on Lebedev grid. The mirrored pressure below the ground plane is a consequence of the symmetry boundary condition used to model an infinite sound-hard ground.

not be the case when setting the pressure below ground to zero or only evaluating the upper half-sphere; one can also interpret this approach as introducing an additional image source below the ground.

The model was solved up to 3 kHz in 30 Hz steps, which, when transformed to time domain, results in impulse responses describing the propagation from the AVAS loudspeaker to receiver positions on the evaluation grid with a sampling rate of 6 kHz and a duration of 33.3 ms. This upper-frequency limit was set due to high computational demands and effectively limits the maximum possible frequency for the AVAS signal auralization to 3 kHz. Since none of the measured AVAS signals except for vehicle B showed significant contents above 3 kHz and higher frequency BEM calculations would have been infeasible with the available computational resources, it was decided to accept this limitation for the purposes of this study. When higher frequency radiation patterns are needed, the BEM model could be solved using more computational power or be extended by a less computationally demanding approach for higher frequencies, such as ray tracing.

Results Figure 9 shows polar representations of the calculated AVAS radiation directivity for vehicle A. These results show that the radiation in the horizontal plane is focused towards 330° and that the sound radiation in the frontal plane is also skewed towards this direction for all evaluated frequency bands. This directivity appears reasonable as the AVAS loudspeaker of this vehicle is mounted on the right side of the front bumper (c.f. Fig. 1d). While the details of the radiation patterns obtained from the BEM calculations might not be perfectly accurate due to, e.g., deviations in the vehicle geometry and unknown material properties, they strongly indicate that this specific vehicle radiates the AVAS signal mostly to the front right relative to its driving direction and less to the back and to the left side. This observation was also confirmed by in-situ directivity measurements of the vehicle A AVAS system, which we, in order to limit the extent of this paper, do not describe

in further detail. The radiation directivities calculated for vehicles B and C were calculated using the same methods and are attached in Appendix A (Fig. A1). How exactly these patterns differ and whether or not they are beneficial for the intended AVAS warning purposes would be a relevant topic for a follow-up study. For the scope of this paper, we conclude that the AVAS radiation in the relevant frequency range is not omnidirectional and should, as long as it is not shown to be perceptually irrelevant, be included in the auralization.

3.2.2 Tire directivity

The implemented tire directivity is based on static laboratory measurements performed in [27]. Thereby, a microphone was flush mounted to the ground below the rolling surface of a commercial tire (type 155SR13) with a distance of 10 mm to the tire/road contact point. The tire itself was installed on a middle-class car, and a loudspeaker was placed with 7.5 m distance to the center of the tire and 1.2 m above the ground at a horizontal angle of 0° , 15° , 30° , 45° , 60° , 75° , 90° relative to the tire normal axis. Transfer functions between the loudspeaker and the microphone were measured outdoors on a dense asphalt surface using the maximum length sequence technique. Assuming reciprocity, these transfer functions describe the radiation from the tire/road contact point to the environment, which were then normalized by the transfer function measured for 0° and mirrored to construct a full sphere 360° radiation pattern. Half of this radiation directivity was manually attenuated by up to 12 dB to compensate for the car body blocking parts of the radiated sound, which resulted in a polar pattern as shown in Figure 10. For the tires on the opposite side of the car, the pattern was rotated by 180° . While the resulting directivity is certainly not as accurate as, for example, a detailed radiation simulation of the exact tires used for the reference pass-by measurements, the main focus of this work is the accurate auralization of the specific AVAS systems more than the exact reproduction of the tire/road noise. We, therefore, assume that the obtained generic tire/road noise radiation pattern is sufficient for this purpose.

3.2.3 Spherical harmonic expansion

The previously described radiation directivities, i.e., the transfer functions from the sound source to a spherical grid of evaluation points, can be used to directly calculate the sound propagation from a moving electric vehicle to a receiver position. However, doing so is not very convenient, especially when considering that other researchers might want to embed the presented radiation directivities in their own auralization tools. One complication with directly using the obtained transfer functions for the auralization is that the AVAS and tire directivities have different spatial resolutions, and the exact coordinates of the evaluation points need to be known for further processing. Additionally, calculating the directivity for a polar angle that does not lie on the evaluation grid requires some form of interpolation, and

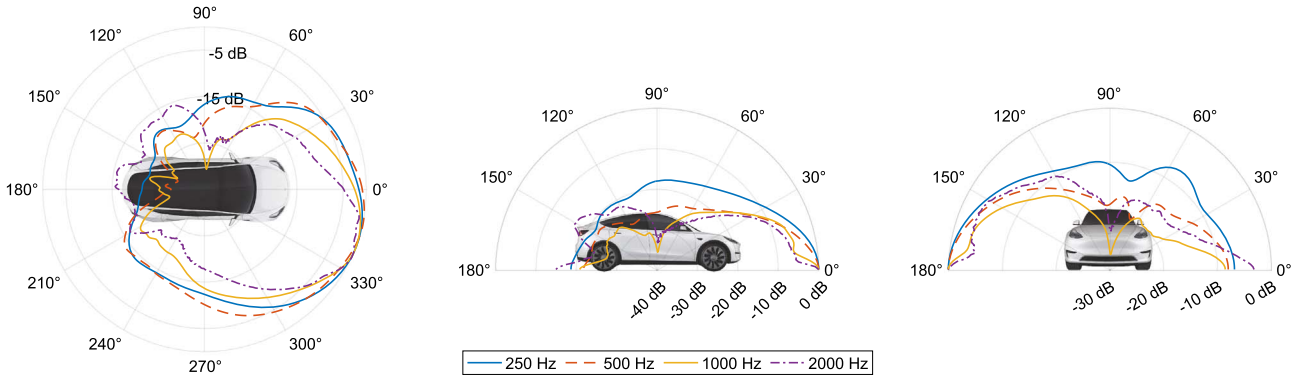


Figure 9. Vehicle A AVAS radiation directivity results of BEM calculation for the horizontal, median, and frontal plane. Normalized to the maximum for each frequency band.

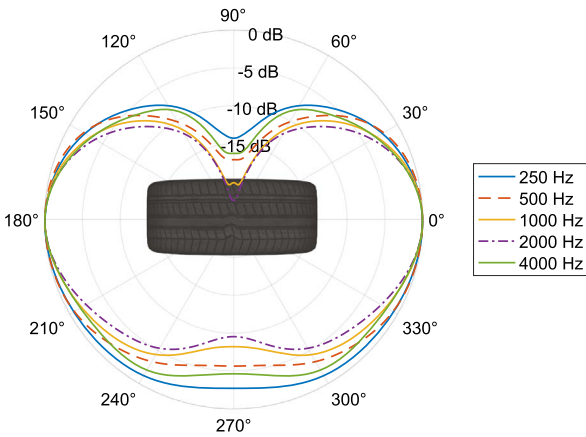


Figure 10. Tire Directivity in octave bands, normalized to the maximum for each band and attenuated towards the direction of the vehicle body.

for real-time auralizations, the spatial resolution of the AVAS directivities might need to be decreased.

All these problems are simplified by encoding the radiation patterns using spherical harmonic (SH) expansion [36–39]. While already established for virtual acoustics applications, this concept of handling complex radiation patterns has also gained entry to other fields such as railway acoustics [40] or road traffic noise auralizations [41]. Thereby, the fundamental approach is that a sound field defined on the surface of a sphere with radius r_0 is decomposed into a sum of orthogonal spherical harmonic basis functions, resulting in a set of spherical harmonic coefficients. This set of SH coefficients then allows the extrapolation of the pressure at an arbitrary angle and distance as well as the reduction of the spatial resolution by truncating the order of the spherical harmonics. The following paragraphs describe the implemented spherical harmonic encoding and extrapolation methods and evaluate how well the AVAS radiation patterns are reproduced using these techniques.

Encoding If $p(r_0, \phi, \theta, \omega)$ represents the pressure on an observation sphere with radius r_0 , which, in our case,

corresponds to the BEM calculations and tire radiation measurement results, this pressure can be expanded into a set of spherical harmonic expansion coefficients $W_l^m(r_0, \omega)$ as [35, Ch. 6.3.3]

$$p(r_0, \phi, \theta, \omega) = \sum_{l=0}^{\infty} \sum_{m=-l}^l W_l^m(r_0, \omega) Y_l^m(\phi, \theta). \quad (4)$$

Thereby, $Y_l^m(\phi, \theta)$ represents the SH basis functions with order l and degree m for the azimuth angle ϕ and the colatitude angle θ . When limiting the maximum SH order to $l = L$ and considering spatially discrete pressure observation points, equation (4) can be written in vector-matrix form as [39]

$$\mathbf{p} = \mathbf{Y} \mathbf{W}. \quad (5)$$

If the number of discrete observation points in \mathbf{p} is greater than or equal to $\eta_g \cdot (L + 1)^2$, equation (5) can be solved in a least-squares sense to obtain the SH coefficient matrix \mathbf{W} [42]. Thereby, η_g represents the degree of overdeterminacy of the pressure sampling scheme, which corresponds to $\eta_g = 1.3$ for the Lebedev grid used in this study [43]. The maximum SH order L determines the spatial resolution of the encoded directivities; an advantage of the spherical harmonic encoding is that the spatial resolution can be reduced without any additional interpolation or down-sampling by simply truncating $W_l^m(r_0, \omega)$ to a lower SH order. Based on the number of pressure observation points obtained from the BEM calculations and the tire measurements, the AVAS directivities were encoded with SH order $L = 64$ and the tire directivity with $L = 16$ which, based on studies such as [44, 45] and the perceptual validation performed in Section 5, is expected to be sufficient for auralization purposes.

Extrapolation Assuming that the reference pressure was observed on a sphere with radius r_0 , the pressure $p(r, \phi', \theta', \omega)$ at any position with $r \geq r_0$ can be extrapolated by multiplying the SH coefficients, $W_l^m(r_0, \omega)$, and the SH basis functions for those new positions, $Y_l^m(\phi', \theta')$, and scaling the result with the l -th order spherical Hankel function of the first kind, $h_l(kr)$, as [35, Eq. 6.94]

$$p(r, \phi', \theta', \omega) = \sum_{l=0}^L \frac{h_l(kr)}{h_l(kr_0)} \sum_{m=-l}^l W_l^m(r_0, \omega) Y_l^m(\phi', \theta'). \quad (6)$$

For the AVAS directivities, the symmetry condition embedded in the BEM model results in a mirrored pressure field below the ground (c.f. Section 3.2.1), which, when included in the SH expansion, leads to the extrapolation method correctly reproducing all ground reflections. For the tire/road noise, we assume that the sound source corresponds to the tire contact point on the ground; hence, there are no first-order ground reflections. Second-order ground reflections, such as tire/road noise scattered from the tire or vehicle chassis to the ground, are not correctly extrapolated using the measured tire directivities.

Validation To evaluate the accuracy of the SH encoding and extrapolation method, the modal assurance criterion (MAC) [46] between the pressure extrapolated from the SH coefficient set with $r_0 = 3$ m to a 2861-point spherical validation grid with $r = 6$ m and the BEM results for the same validation positions was calculated as

$$\text{MAC}(\omega) = \frac{|\mathbf{p}_{\text{SH}}^{\text{H}}(\omega)\mathbf{p}_{\text{BEM}}(\omega)|^2}{(\mathbf{p}_{\text{BEM}}^{\text{H}}(\omega)\mathbf{p}_{\text{BEM}}(\omega))(\mathbf{p}_{\text{SH}}^{\text{H}}(\omega)\mathbf{p}_{\text{SH}}(\omega))}. \quad (7)$$

Thereby, \mathbf{p}_{SH} and \mathbf{p}_{BEM} correspond to the complex pressure values obtained from the SH extrapolation and the BEM simulation; the superscript H marks the Hermitian transpose. Similar to a correlation coefficient, the MAC describes the degree of linearity between the extrapolated SH pressure and the pressure at the validation grid obtained directly from BEM. A MAC value of 1 indicates a perfectly linear spatial dependency between both pressure sets, which, compared to spatially independent measures such as the RMS error, gives more meaningful insights into the actual similarity of the radiation patterns. This comparison allows estimating how well the SH extrapolation method can reproduce pressure at positions other than the ones used to calculate the SH directivities.

Figure 11 shows the MAC results as a function of SH order and frequency. It can be seen that higher frequencies, in general, require higher SH orders to reproduce the AVAS radiation pattern correctly. At the maximum evaluated SH order of $L = 64$, both pressure sets have a perfectly linear dependency up to a frequency of 1600 Hz. This result indicates that the SH extrapolation correctly reproduces the direct pressure radiated by the AVAS speaker as well as captures the ground reflections and scattering on the vehicle body included in the BEM simulations. Around 2500 Hz, the MAC for $L = 64$ drops to zero, indicating a significant error between the SH extrapolated data and the BEM results. This comparably large error for low SH orders is a consequence of the acoustic center of the sound source not being placed in the center of the evaluation sphere [47], which results in a more complicated pressure pattern on the evaluation sphere and hence requires a higher SH order than when the evaluation sphere would be aligned with the acoustic center of the sound source. This

influence of the evaluation sphere position was confirmed by calculating an additional set of directivities with the Lebedev pressure evaluation grid centered on the AVAS loudspeaker instead of the center of the vehicle. This shift resulted in a significantly lower required SH order of between $L = 10$ and $L = 24$, depending on the exact vehicle type. While the resulting auralizations do not differ perceptually, those lower-order speaker-centered directives are a more efficient way to describe the sound radiation and result in lower computational demand. However, since scattering from the vehicle has to be considered part of the sound source, the entire car must be enclosed by the evaluation sphere. Centering the sphere on the AVAS speaker mounted in the front bumper hence requires a larger radius r_0 , resulting in an increase of the minimum auralization distance from 3 m to 5 m. This makes the speaker-centered directivities less suitable for the auralization of close-distance vehicle passages compared to the vehicle-centered directivities used in this paper.

While the upper-frequency limit of the vehicle-centered directivities with $L = 64$ is sufficiently high to cover the most relevant parts of the measured AVAS signals, the MAC validation indicates that not all AVAS components are reproduced correctly without increasing the SH order even further. Fortunately, several perceptual studies have shown that high-frequency deviations in sound source directivity are often not audible and that much lower SH orders may be sufficient for auralization purposes [44, 45]. Additionally, the modal assurance criterion evaluates the similarity of both magnitude and phase, whereas, in practice, it is often considered acceptable only to reproduce the correct magnitude of sound source directivities, significantly reducing the required SH order [39]. Since the overall goal of this work was to create perceptually accurate rather than numerically perfect simulations, it was concluded that an AVAS SH order of $L \leq 64$ for a vehicle-centered directivity is sufficient for this purpose. This assumption was further investigated in a listening experiment presented in Section 5.

3.3 Propagation

The movement of the outdoor sound source was implemented using the concept of moving Green's functions [48]. This means that the desired trajectory of the vehicle is spatially discretized according to the desired sampling frequency, i.e., each audio sample is assigned to a corresponding source position. Transfer functions describing the propagation from each of those discrete source positions to the receiver position are then calculated by extrapolating the SH directivities using equation (6) applying an inverse Fourier transform results in a set of Green's functions $g_j(n)$ for all discrete source positions j . The number of source positions N is equal to the number of output samples and source signal samples; the length of each Green's function depends on the frequency resolution of the SH coefficients. Combining these Green's functions with a source signal $s(n)$ obtained from the synthesis methods described in Section 3.1 allows for calculating the resulting

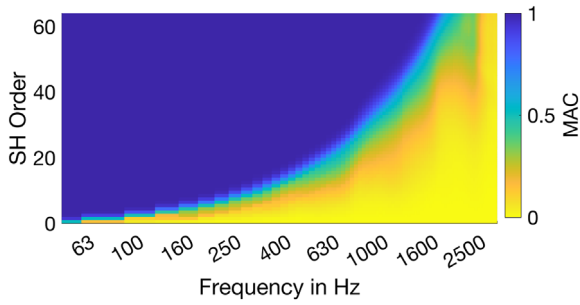


Figure 11. Modal assurance criterion between SH directivities extrapolated to validation grid and corresponding BEM results as a function of SH order and frequency.

pressure at the receiver position $p_{\text{rec}}(n)$ by convolving each Green’s function with the corresponding sample of the source signal as

$$p_{\text{rec}}(n) = \sum_{i=0}^{N-1} s(n-i) \cdot g_{n-i}(i). \quad (8)$$

To account for atmospheric absorption, the resulting pressure is filtered according to ISO 9613-1:1993 [49] by attenuating individual third-octave bands depending on the instantaneous distance between the source and the receiver. For a headphone-based reproduction, the pressure signal is then block-wise convolved with head-related transfer functions (HRTFs) corresponding to the instantaneous angle between the source and the receiver, resulting in a binaural output signal. Alternatively, the HRTF for each sample n can be convolved with the corresponding Green’s function to directly obtain a binaural signal from equation (8). Since the auralization model assumes that the overall vehicle radiation is composed of five different sound sources, i.e., four independent tires and the AVAS loudspeaker, which all have a different spatial orientation relative to the receiver position, the previously described process is performed separately for all five sound sources and the resulting binaural pressure signals are added up to obtain a summation of all sound sources. Finally, binaural ambient noise recorded at the exact location used for the measurements in Section 2 is added to the output signal to create a more lifelike scenario instead of simulating a vehicle passing by in a perfectly silent environment.

4 Evaluation

In order to numerically evaluate the quality of the auralization model, several passages of all vehicles were simulated using the velocity profiles of the corresponding recordings. Since those recorded velocity signals contain small fluctuations that can result in unwanted modulations during the synthesis step, as shown in Section 3.1.2, they were smoothed by replacing them with a 5th-order polynomial fitted to the velocity recordings. Using the same velocity profile and vehicle type means that, ideally, the auralization should result in a signal identical to the

recorded signal. Figure 12 shows an exemplary comparison between a recorded and auralized passage of vehicle B, including the measured and smoothed vehicle velocity profile for a roadside observer position. Comparing the spectrograms of the recorded and generated signals (c.f. Fig. 12a, Fig. 12b) shows that the generated signal misses some of the weaker tonal components above 630 Hz, resulting in temporary differences of up to 6 dB in the 2 kHz and 4 kHz octave bands. However, this difference in the number of tones is not as pronounced in the final auralization as when comparing the isolated AVAS source signals (see Fig. 6). This could be explained by directivity and propagation modeling attenuating those weak tones to a level below the background noise floor.

When listening to both the recording and the auralization results, a noticeable difference is that the tire/road noise in the recording contains transient crackling sounds, most likely caused by small stones on the road surface. While the subtractive tire/road noise synthesis method alone is not able to reproduce such sounds, the auralization could be improved by combining the subtractive synthesis method for broadband tire/road noise components with a sample-based synthesis approach for those crackling sounds.

Additionally, the tonal components in the generated signal decrease drastically as soon as the vehicle passes the observer at ca. 15 s, while, in the recorded signal, tonal components fade out more slowly. This difference is also evident in the octave band comparison over time shown in Figure 12d. A possible explanation for these deviations might be that the numerically estimated AVAS directivity is inaccurate, i.e., the AVAS radiation is more omnidirectional than the BEM results indicate. This might be caused by the simplified vehicle geometry and the lack of surface roughness and diffuse reflections in the BEM model. Another factor that could lead to such an overly pronounced magnitude change during the vehicle passage is that the simulated signals, except the ground reflections included in the directivities, assume free field sound propagation. The recordings, however, were made in proximity to buildings, resulting in additional reflections and, hence, a more diffuse sound field at the receiver position. Even if the AVAS radiation is highly directional, additional diffuse reflections would automatically decrease the influence of the source directivity during the passage. Implementing an image source model by mirroring the calculated directivities would allow the inclusion of those reflections in the simulation, potentially resulting in more accurate auralizations at the cost of higher computational demand. Finally, omitting the time structure and comparing both signals in third-octave bands, as shown in Figure 12c, reveals that the auralization relatively accurately reproduces the overall time-averaged sound pressure levels, which might be of interest for research in the context of traffic noise regulations. To summarize, the numerical comparison between the recorded and synthesized passage of vehicle B revealed differences in the time/frequency structure of the signals, which could originate from an incorrect directivity model or a lack of environmental reflections. Similar differences

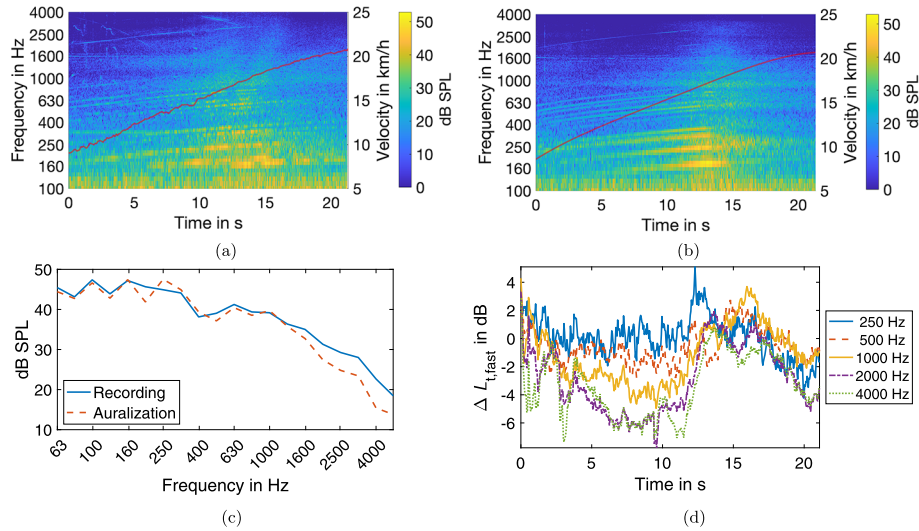


Figure 12. Recorded (a) and auralized (b) passage of vehicle B including the measured (a) and smoothed (b) vehicle velocity, third-octave band levels of vehicle B recording and auralization (c) and difference in octave-band fast-weighted levels between vehicle B recording and auralization (d).

are also visible when comparing auralizations and reference recordings for the other evaluated vehicle as shown in [Appendix A \(Fig. A2\)](#).

5 Perceptual validation

While the previously presented numerical validations already revealed that the auralization results are no perfect reproductions of the reference recordings, a perceptual validation is necessary to determine whether or not those differences are relevant to the method’s intended purpose of performing AVAS-related listening experiments. Thereby, one has to first specify the needs of this application to be able to decide on which “quality level” is required. For example, does the auralization necessarily have to be numerically identical to in-situ measurements? Or is it sufficient if the auralization is “authentic”, i.e., perceptually indistinguishable in direct comparison to a real sound acting as external reference [50, 51]? Or might the quality of the auralization already be acceptable if it is perceived as “plausible”, meaning the simulation corresponds to a listener’s expectation of the corresponding real event [52] based on an internal reference that is built up by everyday life experiences [53]. In the context of virtual acoustic environments, one could also say that authenticity means that all perceivable “quality features” [54] of an acoustic environment are copied, while plausibility means that only the features required for a specific purpose are simulated. Following this definition, one could argue that it is enough for any application to strive for “plausibility” as there is no point in reproducing unnecessary features. This, however, implies that one needs to know exactly which features are required for a specific application. In the context of psychoacoustic experiments, that is not always possible, as the sole purpose of such studies might lie in estimating which features of a complex auditory scene are perceptually

relevant. This could be seen as an argument to always strive for authenticity in the context of auralizations for listening experiments.

From a more pragmatic point of view, we concluded that, while there are some areas of virtual acoustics where authenticity might be achievable, an authentic auralization of complex acoustic scenarios such as electric vehicle passages is very ambitious and comes at the cost of high effort and low flexibility. Listening to the auralization results presented in [Section 4](#), it becomes clear that some small audible differences would stand out in a direct A/B comparison, even when the results are perceived as very similar to real-life recordings. To design and fine-tune an AVAS system for a specific existing vehicle, the overall auralization should be as authentic or, when it comes to estimating compliance with regulations, even as numerically correct as possible. Nevertheless, for our goal of investigating the human response to AVAS signals, we argue that it is secondary whether the auralization sounds exactly like an existing electric vehicle as long as it is perceived as plausible and we are aware of and have complete control over all signal properties. Even when the overall perception of a stimulus is “only” plausible, there could nevertheless be some individual features that are perceived as authentic. For example, in this work, we prioritize the AVAS signal auralization over the tire/road noise, which could mean that the isolated AVAS signal is perceived as authentic while the overall combination of AVAS and tire/road noise is not.

Based on these considerations, we evaluated plausibility in terms of “sounds like it could be an electric vehicle passage” by performing a laboratory listening experiment with 20 participants. Additionally, an indirect parametric comparison to the reference recordings was performed by asking the subjects to rate perceptual attributes such as annoyance and vehicle speed for both auralizations and recordings. The following sections describe the experiment setup and discuss the implications of the obtained results.

5.1 Procedure and stimuli

The listening experiment was divided into two parts: In the first part, the participants were presented with ten binaural in-situ recordings, two for each vehicle and driving direction. The subjects were informed that they were listening to real recordings and were asked to rate perceived vehicle speed and annoyance for each passage. When recruiting a group of non-experts, one has to assume that not all subjects have sufficient experience to rate plausibility without any training phase, especially since electric vehicle sounds are not yet well established in our everyday lives. The purpose of this first experiment part was, therefore, to familiarize the participants with the sound of electric vehicle passages and, by this, build up an internal reference while, at the same time, obtaining a “ground truth” for the subjective vehicle speed and annoyance ratings.

In the second part of the experiment, participants were again presented with five out of the ten in-situ recordings from the first experiment part, one for each vehicle and driving direction. Additionally, the subjects listened to 20 generated passages synthesized using a smoothed version of the vehicle velocity profiles belonging to the corresponding reference recordings. This means that, ideally, the generated signals in the second experiment part should be indistinguishable from the reference recordings presented in the first experiment part. Of those 20 generated stimuli, ten passages were rendered with spherical harmonic order $L = 64$, five passages were rendered with $L = 16$, and five passages were low-quality renderings included to act as anchors. This group of low-quality renderings consisted of (i) amplitude-panned white noise, (ii) amplitude-panned white noise combined with a binaural ambience recording, (iii) an auralization without ambience sounds, and (iv) auralizations without spatialization for two different vehicles. The stimuli without spatialization were generated by using a static HRTF for 0° azimuth angle instead of HRTFs varying according to the source position. This means they include radiation directivity and distance attenuation but no binaural movement cues. For each stimulus, the participants were asked to rate perceived annoyance, vehicle speed, and plausibility compared to their internal reference based on the recordings presented in the first experiment.

5.2 Participants and implementation

The experiment was performed by 20 participants (12 male, 7 female, 1 preferred not to specify) recruited from Chalmers students and faculty members. The participants were aged between 22 and 37 years, with a median age of 27 years. All participants had self-reported normal hearing and an educational background in acoustics and gave their written consent for participation as well as collection and processing of their personal data. All stimuli were presented via calibrated headphones (Sennheiser HD 650), and the experiment was conducted using a HEAD acoustics SQala jury testing system. The order of stimuli within both experiment parts was randomized for each participant. All auralizations were rendered using HRTFs measured for a HEAD

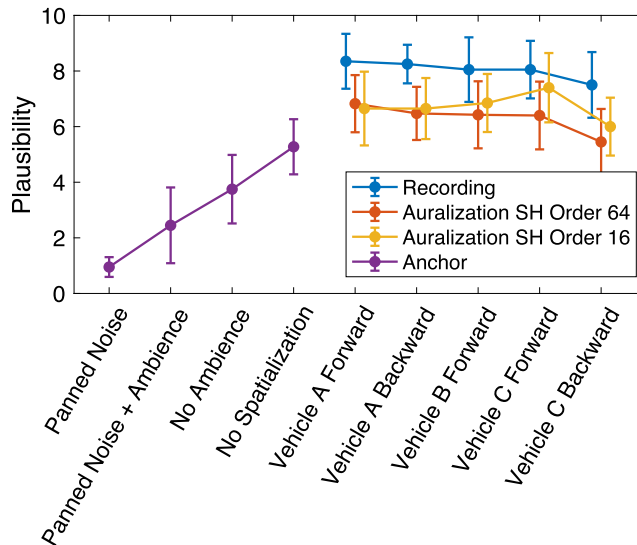


Figure 13. Arithmetic mean of plausibility ratings with 95% confidence intervals. Vehicle B was only evaluated driving forward since its AVAS signal does not change with driving direction.

acoustics HMS II.3 artificial head [55], which has the exact dimensions and ear shape as the HMS V artificial head used for the reference measurements.

5.3 Results

5.3.1 Plausibility

In the second part of the experiment, participants rated the plausibility of in-situ recordings, auralizations with spherical harmonic order $L = 64$ and $L = 16$ as well as low-quality anchor signals on a unipolar numerical 11-point interval scale ranging from the value 0 (“not at all plausible”) to the value 10 (“extremely plausible”). Figure 13 shows the arithmetic mean and 95% confidence intervals of the obtained results. Independent of vehicle type, it can be seen that the in-situ recordings consistently scored the highest plausibility rating and that the anchor signals are rated as least plausible, with the amplitude-panned noise achieving the lowest plausibility score. Comparing the different vehicle types, the backward passages of vehicle C were rated as least plausible for both recordings and auralizations. This could be due to the fact that the vehicle C backward AVAS consists of a constantly repeating “plinging” sound that, in itself, could be perceived as artificial. Furthermore, there seems to be no consistent difference pattern between both evaluated spherical harmonic orders.

To further investigate the difference between in-situ recordings and auralizations, the plausibility ratings were averaged over vehicle type as shown in Figure 14. It can be seen that the recordings achieved a mean plausibility of around 8.0 out of 10 while the auralization scored a lower average plausibility rating of 6.3 for SH order $L = 64$ and 6.7 for $L = 16$. The overall difference between those three stimuli groups was determined as statistically significant according to a repeated measures analysis of variance with

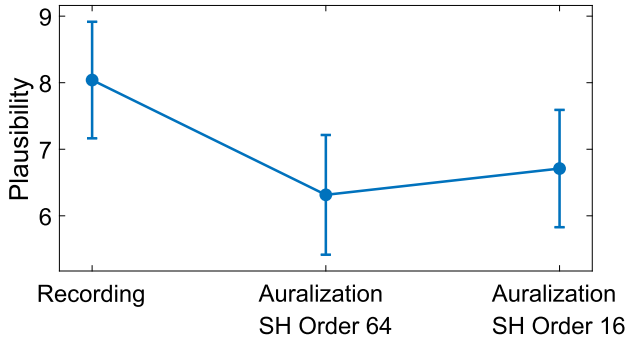


Figure 14. Arithmetic mean of plausibility ratings with 95% confidence intervals, averaged over all vehicle types.

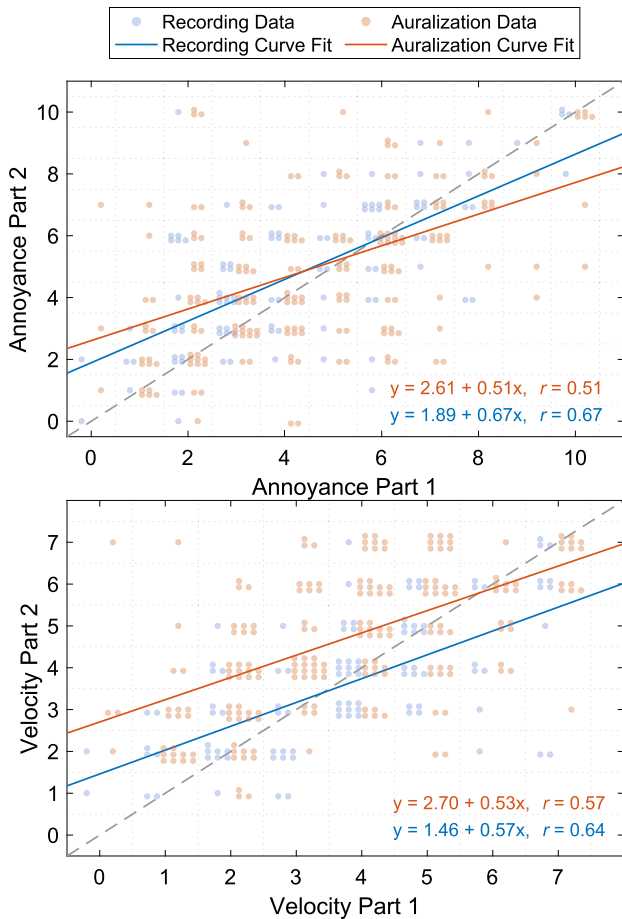


Figure 15. Annoyance and vehicle velocity ratings for both experiment parts and linear regression with correlation coefficient r . Observe that the used interval scale only allows for integer answers; data points with identical values were slightly offset to better visualize the distribution. The data combines the results for all evaluated vehicles; only auralizations with SH order $L = 64$ are included.

Greenhouse-Geisser correction ($F(1.502, 28.540) = 15.134$, $p < .001$, $\text{partial } \eta^2 = .443$). A Bonferroni-adjusted post-hoc analysis revealed a significant difference between in-situ recordings and auralizations with SH order $L = 64$ (MD = 1.725, 95%-CI[0.719, 2.731], $p < .001$) as well as between the recordings and auralizations with $L = 16$ (MD =

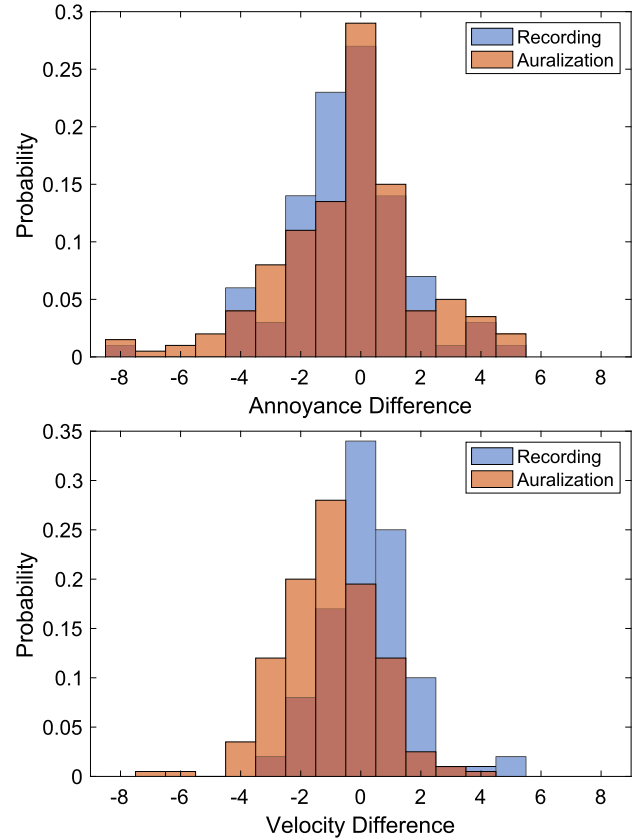


Figure 16. Distribution of differences in annoyance and vehicle velocity ratings between both experiment parts. The data combines the results for all evaluated vehicles; only auralizations with SH order $L = 64$ are included.

1.330, 95%-CI[0.381, 2.279], $p = .005$) but not between auralizations with $L = 16$ and $L = 64$ (MD = -0.395 , 95%-CI[-0.961 , 0.171], $p = .248$). This indicates that the auralizations are not perceived as plausible as the in-situ recordings and that there is no significant difference in plausibility between spherical harmonic orders $L = 16$ and $L = 64$ for this specific application. The fact that even the in-situ recordings were not rated as perfectly plausible shows that, despite the training phase in the first experiment part, not all participants had a sufficiently strong internal reference to identify the authentic signals reliably. Even though not as good as the recordings, the auralization plausibility ratings are still relatively high on the scale and significantly better than for the low-quality anchor signals. Looking at the difference between the individual anchor signals, it becomes clear that both the added binaural ambient noise and the spatialization are relevant features for the overall perceived plausibility since the signals rendered without those attributes are rated as less plausible than the complete auralizations.

5.3.2 Annoyance and vehicle velocity

In both experiment parts, the participant rated perceived annoyance and vehicle velocity for all stimuli. Thereby, annoyance was measured on a unipolar numerical

11-point scale labeled from 0 (“not at all annoying”) to 10 (“extremely annoying”) as recommended by ISO/TS 15666 [56], the perceived vehicle velocity was rated on an 8-point unipolar scale labeled as <5 km/h, 5–10 km/h, 10–15 km/h, 15–20 km/h, 20–25 km/h, 25–30 km/h, 30–35 km/h and 35–40 km/h. Since absolute velocity values are not of interest for the following evaluation, those eight velocity categories were translated to integer values ranging from 0 (<5 km/h) to 7 (35–40 km/h).

Using the ratings obtained for the in-situ recordings in the first experiment part as a reference allows for evaluating how well the auralizations presented in the second experiment part reproduce features relevant to the perception of annoyance and vehicle velocity. Since half of the in-situ recordings from the first experiment part were also repeated in the second part, we can additionally determine how consistent those subjective ratings are throughout the experiment. Figure 15 compares the annoyance and vehicle velocity ratings obtained for auralizations and recordings in the second experiment part to the results from the first part. Thereby, only ratings for stimuli pairs that exactly match each other were compared for each participant, i.e., values for the exact same recordings in experiment parts one and two, as well as ratings for auralizations in part two, matched with ratings for the recordings in part one that they aimed to reproduce. If the ratings in experiment part two would perfectly match the results of the first experiment part, all data points in Figure 15 would lie on the identity line. However, that is not the case, neither for the recordings nor for the auralizations.

In order to better understand the difference between the results of both experiment parts, a simple linear regression was performed. The resulting regression lines presented in Figure 15 indicate that, for both repeated in-situ recordings and auralizations, participants tend to less extreme ratings in the second experiment part than in the first part, i.e., all regression lines have a similar slope smaller than one. This tendency could be statistically explained by a regression to the mean effect [57], i.e., assuming the annoyance and velocity ratings of each subject to be random variables with a certain distribution around a mean value, it is statistically more likely that subjects who gave an extreme rating in the first experiment part tend to ratings closer to this mean value in the second experiment part. Alternatively, this trend could be interpreted as a repetition priming effect [58], meaning that, after hearing all stimuli of the first experiment part, the participants might have adjusted their internal reference, resulting in more conservative ratings in the following part. Similar response patterns were described as “simple order effect” in other laboratory noise annoyance studies such as [59]. Independent of the cause, this observation means that it is not sufficient for the perceptual validation of the auralization only to compare the difference between both experiment parts since even a “perfect” auralization that exactly reproduces the in-situ recordings would show this inconsistency in the subjective ratings.

While the correlation between the data for both experiment parts is slightly higher for the repeated in-situ recordings than for the auralizations, both data sets are only

moderately correlated ($r < 0.7$). This means that the relation of the data obtained for both experiment parts is not perfectly linear. Therefore, a simple linear regression might not be the most suitable tool for evaluating whether the difference between ratings for auralizations and in-situ recordings is statistically significant. Instead, the distribution of differences between both experiment parts was compared by subtracting the ratings obtained in the second experiment part from the ratings for the first part as shown in Figure 16. These distributions were then compared using a Wilcoxon signed-rank test, which showed that the distribution of differences in annoyance ratings compared to the reference in-situ recordings in the first experiment part, averaged over vehicles for each participant, does not significantly differ between recordings and auralizations ($Z = -.081, p = .936$). The difference in vehicle velocity rating was found to be statistically significant ($Z = 3.825, p < .001$) which, in combination with the shape of the distributions shown in Figure 16 and the offset between the recording and auralization velocity regression lines in Figure 15, leads to the conclusion that the auralization results in higher subjective vehicle velocity ratings than obtained for the corresponding in-situ recordings.

Based on the numerical comparison between auralizations and in-situ recordings discussed in Section 4, we assume that these differences in perceived vehicle velocity are related to the fact that the auralizations tend to show a more drastic change in time/frequency structure when the vehicle passes the observer which could be a consequence of inaccurate radiation directivities and missing environmental reflections. However, more research is needed to determine which features of an auralized vehicle passage influence the perceived velocity. While the perceptual validation results indicate that the implemented auralization method is unsuitable for experiments where the authenticity of the overall perceived vehicle speed is essential, it still allows for comparisons between stimuli, e.g., whether two different AVAS signals result in different speed perceptions.

6 Conclusion

This paper presented an auralization approach for electric vehicle passages based on in-situ measurements of three electric passenger cars. Different AVAS and tire/road synthesis methods were combined with radiation directivity and propagation models to generate pass-by auralizations suitable for AVAS-related psychoacoustic experiments. The numerical validation of the auralization results shows that, while the reproduction of the AVAS source signals is accurate, some discrepancies in the propagation modeling may be caused by inaccurate radiation directivities or missing environmental reflections. The auralization method achieved relatively high plausibility ratings in a perceptual evaluation, even though the generated stimuli were perceived as less plausible than in-situ recordings. While perceived annoyance ratings for the auralization results are consistent with the ratings for in-situ recordings, there is

a statistically significant difference in velocity ratings between measurements and auralizations, which requires further investigation. Overall, we conclude that, while there are possibilities for improvement, the presented methods constitute a suitable foundation for AVAS-related listening experiments.

Acknowledgments

This research was funded by FORMAS – a Swedish research council for sustainable development – under grant agreement FR-2020-01931. The HEAD-Genuit-Foundation provided the hardware used for the listening experiment under grant P-22101-W.

Conflicts of interest

The authors declare no conflict of interest.

Data availability statement

The data and audio examples associated with this article are available on Zenodo under the reference <https://doi.org/10.5281/zenodo.10610490>. All relevant code is published on GitHub under <https://github.com/leonpaulmueller/evat>.

Informed consent statement

Informed consent was obtained from all subjects involved in the listening experiment.

References

1. IEA: Global EV outlook 2023. International Energy Agency, Paris, Tech. Rep. 2023.
2. L.M. Iversen, G. Marbjerg, H. Bendtsen: Noise from electric vehicles – ‘State-of-the-art’ literature survey. INTER-NOISE and NOISE-CON Congress and Conference Proceedings 247 (2013) 267–271.
3. M.-A. Pallas, M. Bérengier, R. Chatagnon, M. Czuka, M. Conter, M. Muirhead: Towards a model for electric vehicle noise emission in the European prediction method CNOS-SOS-EU. Applied Acoustics 113 (2016) 89–101.
4. EAA: Environmental noise in Europe, European Environment Agency, Luxembourg, 2020.
5. M. Wessels, S. Kröling, D. Oberfeld: Audiovisual time-to-collision estimation for accelerating vehicles: the acoustic signature of electric vehicles impairs pedestrians’ judgments. Transportation Research Part F: Traffic Psychology and Behaviour 91 (2022) 191–212.
6. UNECE: Regulation No 138 of the Economic Commission for Europe of the United Nations (UNECE) – Uniform provisions concerning the approval of Quiet Road Transport Vehicles with regard to their reduced audibility [2017/71]. Economic Commission for Europe of the United Nations, Tech. Rep. 2017.
7. EU: Commission Delegated Regulation (EU) 2017/1576. European Union, Tech. Rep. 2017.
8. M. Kleiner, B.-I. Dalenbäck, P. Svensson: Auralization – an overview. Journal of the Audio Engineering Society 41, 11 (1993) 861–875.
9. J. Jagla, J. Maillard, N. Martin: Sample-based engine noise synthesis using an enhanced pitch-synchronous overlap-and-add method. The Journal of the Acoustical Society of America 132, 5 (2012) 3098–3108.
10. J. Forssén, P. Andersson, P. Bergman, K. Fredriksson, P. Zimmermann: Auralisation of truck engine sound – preliminary results using a granular approach, AIA-DAGA, Merano, 2013. [Online]. Available: <https://research.chalmers.se/publication/188498>.
11. R. Pieren, T. Büttler, K. Heutsch: Auralization of accelerating passenger cars using spectral modeling synthesis. Applied Sciences 6, 1 (2015) 5.
12. J. Forssén, A. Hoffmann, W. Kropp: Auralization model for the perceptual evaluation of tyre-road noise. Applied Acoustics 132 (2018) 232–240.
13. W. Kropp, C. Hoever, J. Theyssen: Auralisation of tyre/road noise. In: Fortschritte der Akustik – DAGA, 2024.
14. M.J. Roan, L. Neurauder, M. Song, M. Miller: Probability of detection of electric vehicles with and without added warning sounds. The Journal of the Acoustical Society of America 149, 1 (2021) 599–611.
15. M. Wessels, C. Zähme, D. Oberfeld: Auditory information improves time-to-collision estimation for accelerating vehicles. Current Psychology 42, 27 (2023) 23195–23205.
16. L. Steinbach, M.E. Altinsoy: Influence of an artificially produced stationary sound of electrically powered vehicles on the safety of visually impaired pedestrians. Applied Acoustics 165 (2020) 107290.
17. N. Kournoutos, J. Cheer: Investigation of a directional warning sound system for electric vehicles based on structural vibration. The Journal of the Acoustical Society of America 148, 2 (2020) 588–598.
18. B.U. Seeber, S. Kerber, E.R. Hafter: A system to simulate and reproduce audio-visual environments for spatial hearing research. Hearing Research 260, 1–2 (2010) 1–10.
19. G. Grimm, J. Lubradzka, V. Hohmann: A toolbox for rendering virtual acoustic environments in the context of audiology. Acta Acustica United with Acustica 105, 3 (2019) 566–578.
20. IHTA: Virtual acoustics – a real-time auralization framework for scientific research. Institute for Hearing Technology and Acoustics – RWTH Aachen University, accessed on 2023-11-07. [Online]. Available: <http://www.virtualacoustics.org/>.
21. D. Gonzalez-Toledo, L. Molina-Tanco, M. Cuevas-Rodríguez, P. Majdak, A. Reyes-Lecuona: The binaural rendering toolbox. A virtual laboratory for reproducible research in psychoacoustics. In: Forum Acusticum 2023 – 10th Convention of the European Acoustics Association, 2023.
22. K. Heutsch, E. Bühlmann, J. Oertli: Options for reducing noise from roads and railway lines. Transportation Research Part A: Policy and Practice 94 (2016) 308–322.
23. EU: Regulation (EU) 2020/740 of the European parliament and of the council on the labelling of tyres with respect to fuel efficiency and other parameters. Official Journal of the European Union, 2020.
24. L. Müller, W. Kropp: Dataset: “Auralization of electric vehicles for the perceptual evaluation of acoustic vehicle alerting systems”. Zenodo, 2024. [Online]. Available: <https://doi.org/10.5281/zenodo.10610490>.
25. U. Sandberg, J.A. Ejsmont: Tyre/road noise reference book. Informex Ejsmont & Sandberg handelsbolag, 2002.
26. M. Strasberg: Dimensional analysis of windscreen noise. The Journal of the Acoustical Society of America 83, 2 (1988) 544–548.

27. W. Kropp, F.-X. Bécot, S. Barrelet: On the sound radiation from tyres. *Acustica United with Acta Acustica* 86 (2000) 769–779.
28. A. Oppenheim, R. Schaffer: *Discrete-time signal processing*, 3rd edn., Prentice-Hall Signal Processing Series, Pearson, New Jersey, 2014.
29. Beauchamp, W. James: Additive synthesis of harmonic musical tones. *Journal of the Audio Engineering Society* 14, 4 (1966) 332–342. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=1129>.
30. J.O. Smith, *Spectral audio signal processing*. W3K Publishing, 2011. [Online]. Available: <http://ccrma.stanford.edu/~jos/sasp/>.
31. Couch: *Digital and analog communication systems*, 8th edn., Prentice-Hall International Editions, Pearson, New Jersey, 2013.
32. M.A. Fischler, R.C. Bolles: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24, 6 (1981) 381–395.
33. P. Torr, A. Zisserman: MLESAC: a new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding* 78, 1 (2000) 138–156.
34. V.I. Lebedev, D.N. Laikov: A quadrature formula for the sphere of the 131st algebraic order of accuracy. *Doklady Mathematics* 59 (1999) 477–481.
35. E.G. Williams: *Fourier Acoustics*, Academic Press, London, 1999.
36. G. Weinreich, E.B. Arnold: Method for measuring acoustic radiation fields. *The Journal of the Acoustical Society of America* 68, 2 (1980) 404–411.
37. F. Zotter: *Analysis and synthesis of sound-radiation with spherical arrays*. Ph.D. dissertation, University of Music and Dramatic Arts, Graz, 2009.
38. M. Pollow: *Directivity patterns for room acoustical measurements and simulations*. Ph.D. dissertation, RWTH Aachen, 2014.
39. J. Ahrens, S. Bilbao: Computation of spherical harmonic representations of source directivity based on the finite-distance signature. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29 (2020) 83–92.
40. J. Theyssen, T. Deppisch, A. Pieringer, W. Kropp: On the efficient simulation of pass-by noise signals from railway wheels. *Journal of Sound and Vibration* 564 (2023) 117889.
41. M. Alkmim, G. Vandernoot, J. Cuenca, K. Janssens, W. Desmet, L.D. Ryck: Real-time sound synthesis of pass-by noise: comparison of spherical harmonics and time-varying filters. *Acta Acustica* 7 (2023) 37.
42. D.N. Zotkin, R. Duraiswami, N.A. Gumerov: Regularized HRTF fitting using spherical harmonics. In: *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2009.
43. B. Bernschütz: *Microphone arrays and sound field decomposition for dynamic binaural recording*. Ph.D. dissertation, Technical University of Berlin, 2016.
44. M. Frank, M. Brandner: Perceptual evaluation of spatial resolution in directivity patterns. In: *Fortschritte der Akustik – DAGA*, 2019.
45. T. Lübeck, C. Pörschmann: Investigation of the minimum required spatial resolution of moving sound sources. In: *Fortschritte der Akustik – DAGA*, 2023.
46. R.J. Allemang: The modal assurance criterion—twenty years of use and abuse. *Sound and Vibration* 37 (2003) 14–23.
47. I.B. Hagai, M. Pollow, M. Vorländer, B. Rafaely: Acoustic centering of sources measured by surrounding spherical microphone arrays. *The Journal of the Acoustical Society of America* 130, 4 (2011) 2003–2015.
48. J. Ahrens, S. Spors: Reproduction of moving virtual sound sources with special attention to the Doppler effect. In: *124th Convention of the AES*, 2008. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=14493>.
49. ISO: ISO 9613-1:1993 Acoustics – attenuation of sound during propagation outdoors – Part 1: Calculation of the absorption of sound by the atmosphere. International Organization for Standardization, Tech. Rep., 1993.
50. J. Blauert: *Spatial hearing: the psychophysics of human sound localization*, The MIT Press, Cambridge, Massachusetts, 1996.
51. F. Brinkmann, A. Lindau, S. Weinzierl: On the authenticity of individual dynamic binaural synthesis. *The Journal of the Acoustical Society of America* 142, 4 (2017) 1784–1795.
52. A. Lindau, S. Weinzierl: Assessing the plausibility of virtual acoustic environments. *Forum Acusticum 2011* (2011) 1187–1192.
53. C. Kuhn-Rahloff: *Realitätstreue, Natürlichkeit, Plausibilität, Perzeptive Beurteilungen in der Elektroakustik*. Springer Berlin, Heidelberg, 2012.
54. R.S. Pellegrini: Quality assessment of auditory virtual environments. In: *Proceedings of the 2001 International Conference on Auditory Display*, Espoo, Finland, 2001.
55. C. Pörschmann, J.M. Arend, R. Gillioz: Spherical headgear HRIR compilation of the Neumann KU100 and the head acoustics HMS II.3 [Data set]. In: *EAA spatial audio signal processing symposium*, Paris, France, 2020.
56. ISO: ISO/TS 15666 acoustics – assessment of noise annoyance by means of social and socio-acoustic surveys. International Organization for Standardization, Tech. Rep., 2021.
57. A.G. Barnett, J.C.v.d. Pols, A.J. Dobson: Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology* 34, 1 (2005) 215–220.
58. D.L. Schacter, R.L. Buckner: Priming and the brain. *Neuron* 20, 2 (1998) 185–195.
59. B. Schäffer, R. Pieren, U.W. Hayek, N. Biver, A. Grêt-Regamey: Influence of visibility of wind farms on noise annoyance – a laboratory experiment with audio-visual simulations. *Landscape and Urban Planning* 186 (2019) 67–78.

Appendix A

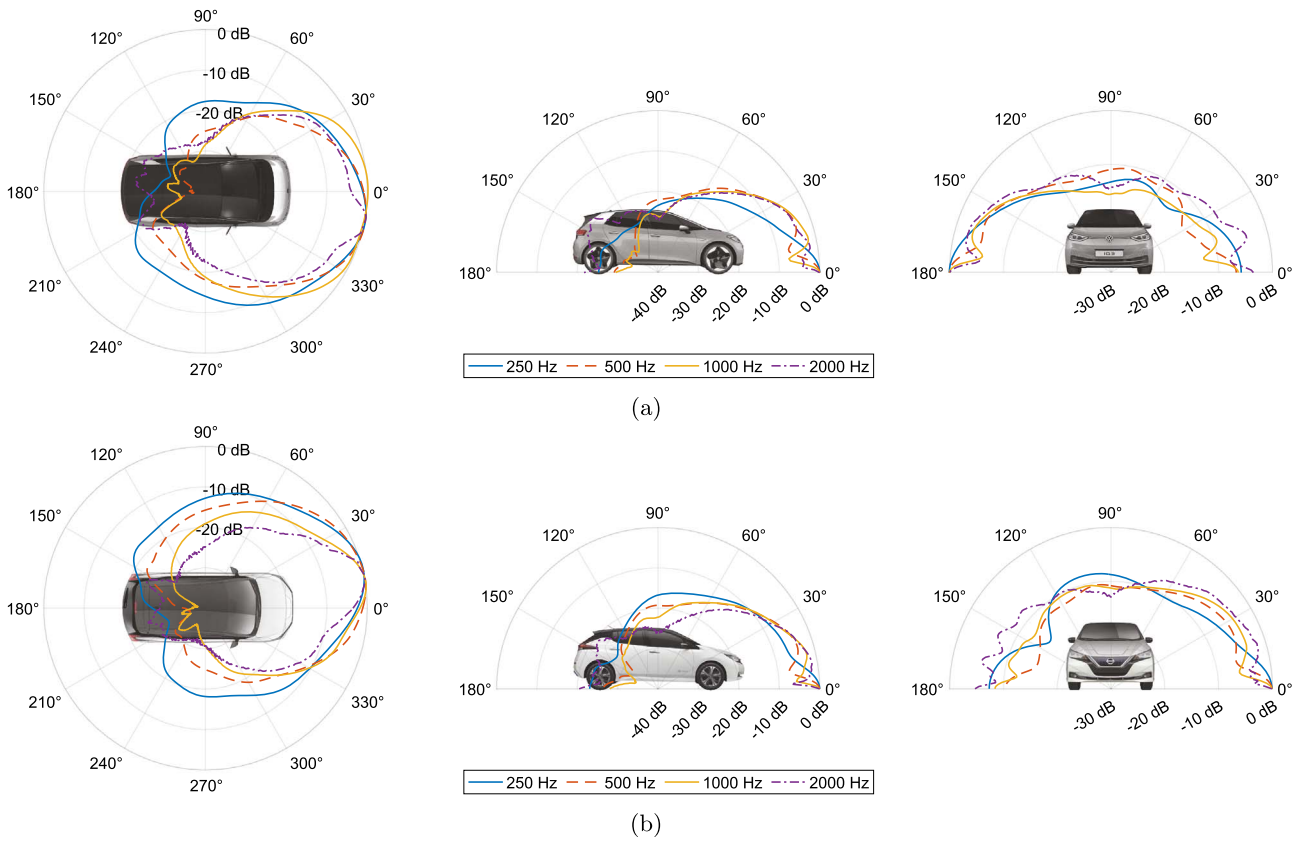


Figure A1. Vehicle B (a) and vehicle C (b) AVAS radiation directivity results of BEM calculation for the horizontal, median, and frontal plane. Normalized to the maximum for each frequency band. For vehicle A, see [Figure 9](#).

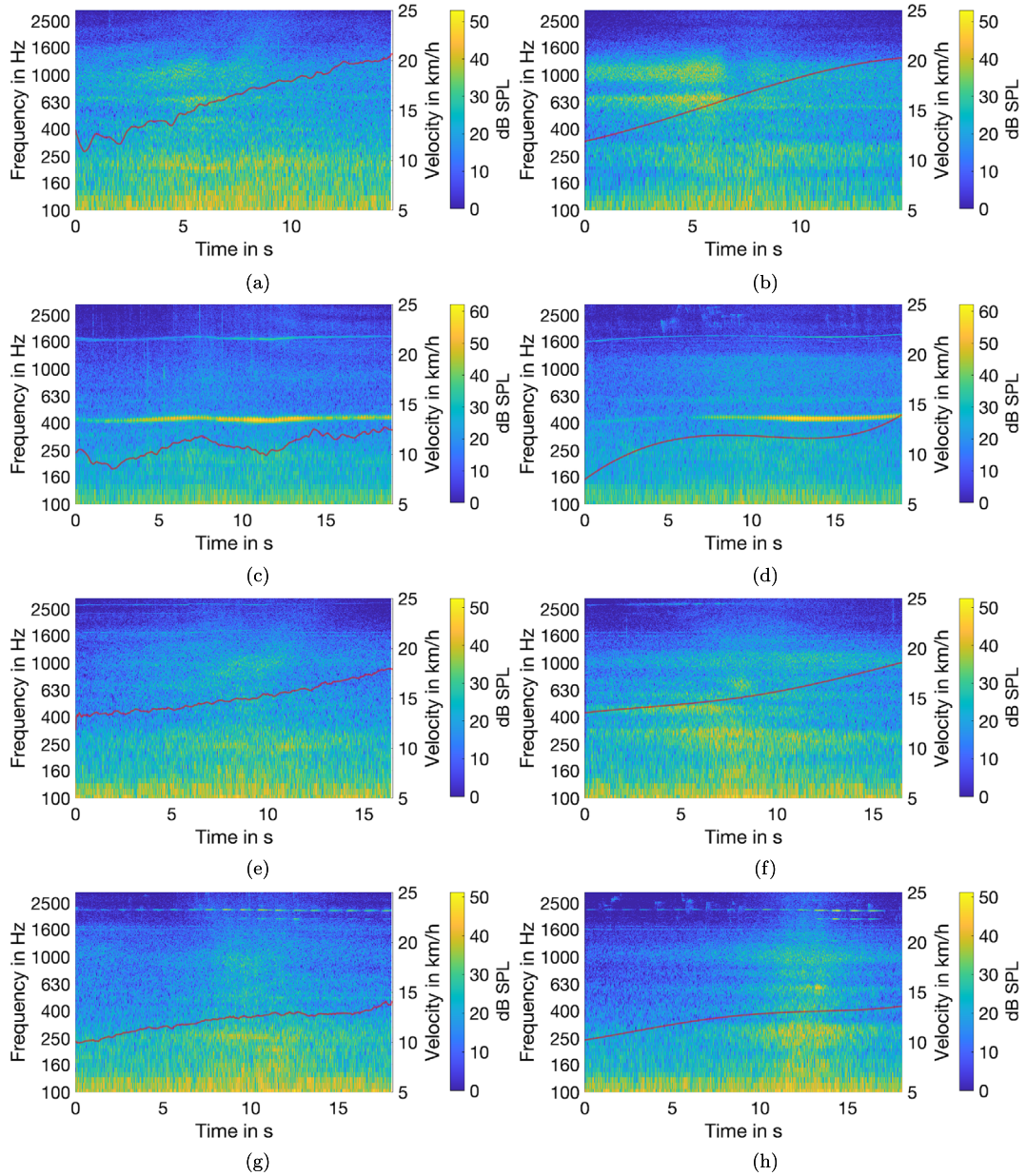


Figure A2. Vehicle A forward recording (a) and auralization (b), vehicle A backward recording (c) and auralization (d), vehicle C forward recording (g) and auralization (h) and vehicle C backward recording (i) and auralization (j). The red lines visualize the recorded and smoothed velocity profiles. For vehicle B, see [Figure 12](#).

Cite this article as: Müller L. & Kropp W. 2024. Auralization of electric vehicles for the perceptual evaluation of acoustic vehicle alerting systems. Acta Acustica, 8, 27.