Auralization of a car pass-by using impulse responses computed with a wave-based method

Citation for published version (APA):
Georgiou, F., Hornikx, M., & Kohlrausch, A. (2019). Auralization of a car pass-by using impulse responses computed with a wave-based method. *Acta Acustica united with Acustica*, *105*(2), 381-391. https://doi.org/10.3813/AAA.919321

DOI:
10.3813/AAA.919321

Document status and date:
Published: 01/03/2019

Document Version:
Accepted manuscript including changes made at the peer-review stage

Auralization of a car pass-by using impulse responses computed with a wave-based method

Fotis Georgiou[1)], Maarten Hornikx[2)] & Armin Kohlrausch[3)]
[1,2)] Department of the Built Environment, Eindhoven University of Technology, Eindhoven, The Netherlands. f.georgiou@tue.nl
[3)] Human-Technology Interaction group, Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, The Netherlands.

Summary

Auralization of traffic noise can be of great value for city planners and in the communication to citizens. In inner city environments, auralization of traffic noise based on predictions might require a wave-based method for the low frequencies. In this paper, a method to auralize a car pass-by using binaural impulse responses (BIRs) computed with the wave-based pseudospectral time-domain method (PSTD) in two dimensions is presented. A dry synthesized car signal is convolved with the binaural impulse responses for different locations in the street, and cross-fade windows are used in order to create a smooth transition between the source positions. Auralizations were performed for the simplified scenarios where buildings are absent, and for an environment where a long flat wall is located behind the car, opposite from the receiver. A same-different listening test was carried out in order to investigate whether increasing the angular spacing between the discrete source positions affects the perception of the car pass-by auralizations. Signal detection theory (SDT) was used for the design and analysis of the listening test. Results showed that differences exist, although they are difficult to notice. On average, 52.3% of the subjects found it almost impossible to spot any difference between auralizations with larger angular spacing (up to 10°) compared to the reference auralization (2° angular spacing).

1 Introduction

Conventionally, the evaluation of outdoor sound environments is done via noise assessments, and the environment is characterized based on equivalent noise levels such as L_{den}. However, sources in the urban sound environments vary with time so they cannot be evaluated by equivalent sound levels only. In contrast to traffic noise, even-though they increase the overall sound levels, natural sounds such as bird song and sounds of water fountains actually lead to positive effects and improve the quality of the overall acoustic environment [1]. This highlights the importance of developing tools that will allow better understanding of urban sound environments and help towards their optimization. Such a tool is auralization [2]. Although the technique of auralization has been used in the context of architectural acoustics since the 1960s, it only entered the field of environmental acoustics in the last 15 years [3, 4, 5].

Modeling of outdoor sound propagation has been achieved with various methods. These methods can be split into three categories: geometrical acoustics methods, energy-based methods for a reverberant environment, and wave-based methods.

Popular geometrical acoustics methods are the ray-tracing [6] and the image source method (ISM) [7]. These methods could also be classified as energy-based methods when phase information between separate ray contributions is ignored. In these methods the behaviour of sound waves is modeled following the principle of light rays [7]. The main disadvantage is that those techniques cannot accurately simulate complex wave phenomena at low frequencies such as diffraction and modal effects [8, 9, 10, 11, 12]. For outdoor sound propagation and road traffic noise modeling, engineering methods based on the ISM have been developed like Harmonoise [10], CNOSSOS [13] and Nord2000 [9]. The greatest advantage of the engineering methods is that they are very computationally efficient.

An example of an energy based technique is the use of the diffusion equation into acoustics, which is based on a statistical method to compute a spatial distribution of acoustic energy over time as a diffusion process [14, 15]. The method assumes that the obstacles of the modeled space scatter sound uniformly and have the same mean-free path (the average distance that a sound ray travels before it hits an obstacle) [16]. These methods also fail to capture modal interferences, which can be present in urban configurations such as narrow street canyons at low frequencies [17]. However, these methods are computationally efficient and provide a convenient approach to model the reverberation tail of an impulse response [18].

Wave-based methods are techniques used to solve the governing partial differential equation(s) of linear acoustics with numerical approximations.

These methods are split into two main categories, the time-domain methods (operate in time-domain so time discretization is applied) and the frequency-domain methods (operate in the frequency-domain so frequency discretization is applied). Time-domain methods are in-line with the nature of sound propagation, which is a time-domain process. Time-domain methods can compute with a single run an impulse response over a broad frequency range. This property of the time-domain methods is one of the main advantages compared to the frequency-domain methods.

Wave-based methods are physically accurate and wave phenomena such as diffraction are inherently included in them [19]. Wave-based methods typically provide higher accuracy, especially at low frequencies, compared to the geometrical acoustics methods [19, 20].

Several methodologies have been developed for the auralization of urban environments with moving sources thus far.

Forssén *et al.* [21, 22] developed a method for auralization of car pass-bys with constant speed for an environment where only the ground is present and for an environment with a ground and noise barrier. The calculation of the propagation path relies on engineering methods. The Doppler effect is modeled by resampling the car signal. Finally, head-related transfer function (HRTF) processing is applied. The synthesized auralizations were compared against recordings, and approximately 50% of the subjects could not detect the differences [23].

Pieren *et al.* [4] auralized accelerating car pass-bys in real-time for an environment where only the ground is present. The sound propagation effects (propagation delay, Doppler effect, geometrical spreading, ground reflection and air absorption) are modeled using time-variant digital filters, based on geometrical acoustics. The synthesized source signals are processed with these filters. No subjective evaluation was conducted to evaluate this method.

Maillard and Jagla *et al.* [24, 25] developed a methodology for real-time auralization of urban environments including moving sources and receivers. The propagation paths which are defined by the source and receiver position are computed based on the Harmonoise model [10]. In [25] a listener moving along bike and pedestrian paths inside a city was auralized, and subjective tests were performed where subjects were asked to judge the perception of the auralization realism on a scale from 0-10. The average score was 6.9.

Viggen *et al.* [26] designed an outdoor auralization prototype using the Nord2000 [9] model to calculate sound propagation between source and receiver. They modeled a road with traffic as evenly spaced source points along the road, which represent the positions of the moving vehicle. The direct, reflected and diffracted sound paths are found separately including the information regarding their incoming angle and 1/3 octave-band levels. The propagation paths for the non-computed source points are estimated with interpolation. Next, they use an FFT-based overlap-add method where the transfer functions are applied corresponding to the position of the moving vehicle along its trajectory. The Doppler effect is also included and applied separately in the direct, reflected, and diffracted sounds by resampling the source signal. The model has not been evaluated.

Auralization for outdoor scenarios has also gained a lot of attention in the context of computer gaming using both wave-based and geometrical acoustics techniques [27, 28, 29]. However, for gaming applications it is sufficient to produce plausible approximations as long as it is computationally efficient and fits the computational budget [19]. Furthermore, these methods have not been validated via subjective tests.

Noise produced by moving vehicles, which are the predominant noise sources in urban sound environments, contains much energy at low frequencies (see Figure 1). This is an indication that accurate sound propagation modeling at low frequencies could be of high importance for the auralization of moving cars. Thus, developing a wave based method for auralization of urban environments could be necessary for authentic auralizations, which is the first goal of this paper. More specifically, a methodology for the auralization of a car pass-by from discrete source positions, which are represented virtually by binaural impulse responses (BIRs) computed with a wave-based method is developed. BIRs computed with wave-based methods have not been previously used for the auralization of a car pass-by.

Since auralizations are performed by cross-fading between discrete source positions, it is interesting to investigate the effect of the spacing between the discrete source positions in the perception of the auralized car pass-by noise. This is the second goal of the paper. The computations of the IRs and their convolution with the car signals is a time-consuming process. Because larger source spacing allows for fewer IRs computations, less computational storage requirements, and fewer convolutions for the creation of auralizations, it is important to investigate how large the spacing between the BIRs can be without affecting the perception of the car pass-by auralization.

In section 2, the computation and processing of BIRs for auralization is presented. In section 3, the auralization methodology is described. The subjective evaluation to investigate if the perception of the auralizations is affected by the increase of the spacing between the discrete source positions is presented in section 4. Afterwards, in section 5, the results of the subjective tests are presented and discussed. Finally,

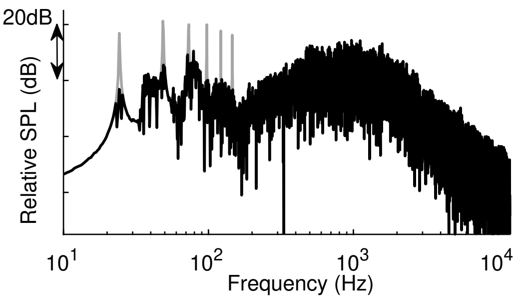

Figure 1: Spectrum of a dry car signal (70 km/h) with (grey) and without (black) tonal components used in the auralizations presented in this work.

in section 6, the conclusions and suggestions for future work are presented.

2 Computation and processing of BIRs for auralization

2.1 Overview

In this section, the methodology for computing BIRs with the pseudospectral time-domain (PSTD) [11] method is described. Those BIRs are used for auralization of car pass-bys, which is described in section 3. This section starts with an introduction to PSTD followed by the settings used in the PSTD simulation. Next, the method used to incorporate HRTFs in PSTD is described. In the last subsection, the post-processing applied on the computed BIRs is presented.

2.2 Pseudospectral time-domain method

Pseudospectral (PS) methods can be used to calculate the spatial derivative operator of time-dependent partial differential equations such as the linear acoustic equations [30]. The time derivatives are evaluated by another method and the combination of these two methods is referred to as PSTD. A popular PS methods is the Fourier PS. Fourier PS computes the spatial derivatives in the wave number domain using Fourier transforms [31]. Due to the spatial discretization of the domain and the use of discrete Fourier transforms, only two spatial points per wavelength are required to obtain highly accurate results. The use of discrete Fourier transforms causes an issue in PSTD regarding the modeling of boundary conditions, as the system then becomes spatially aperiodic. In the computations conducted for this paper the time derivative operator of the linear wave equations was calculated with a low storage optimized six-stage Runge-Kutta method developed by Bogey and Bailly [32]. Because urban environments are large, computational efficiency plays a very important role. Therefore, Fourier PSTD was chosen mainly because it is less computationally expensive compared to other wave-based methods. Moreover, it is a time-domain method, so it can directly compute IRs. More information regarding propagation modeling with PSTD can be found in [30, 11].

2.3 Computational settings

The auralizations were performed for the simplified cases where buildings are absent, and for an environment where a long flat wall is located behind the car. The layout of the 2D discretized domain in PSTD is shown in Figure 2. The reason that the simulation was implemented in 2D was to limit the computational costs in order to allow BIRs to be computed at higher frequencies. Due to its computational demands, a 3D simulation would have to be limited to low frequencies for such a large domain. Limitations of the used auralization method arise from the 2D simulation. These limitations are further discussed in subsection 3.4.

With PSTD, the spatial domain is discretized with an equidistant grid, and the spatial discretization was set to $\Delta x = \Delta y = c/(2f_{max})$ and the time step was set to $\Delta t = \Delta x/(2c)$, where $c = 343$ m/s is the speed of sound and f_{max}=9000 Hz is the maximum frequency that the PSTD grid can solve with two spatial points per wavelength. However, the source spectrum was truncated to 7.5kHz in order to avoid Gibbs phenomena (see [33] for more details on the design of the spatio-temporal source).

Thanks to the acoustic reciprocity principle, only one simulation is needed, in which the modeled source is placed at the physical receiver location and the receivers r_n are placed at actual car positions [34]. Because in this case there were two receivers (left and right ear), two simulations were executed in parallel. The source S was placed in the middle of the road segment and faced perpendicular to the line of receivers (see Figure 2). Thus, the auralization was implemented for the scenario where the listener is placed in the middle of the road segment and faced perpendicular to the direction of the moving source. There were only non-reflecting boundaries at the boundaries of the PSTD domain, which were modeled using a perfectly matched layer (PML) with a thickness of 50 grid points. The image receivers r'_n were the images of the receivers r_n mirrored on the wall as shown in Figure 2. These were used to simulate the 1^{st} order specular reflection from a wall. The distance between the source and the line of receivers r_n was 5 m as was the distance between the receivers and the wall. The BIRs were calculated at 0.25° intervals. The reason that the wall was not modeled as a boundary was to simplify the simulation, i.e. free field results as well as results with one wall are obtained by a single simulation.

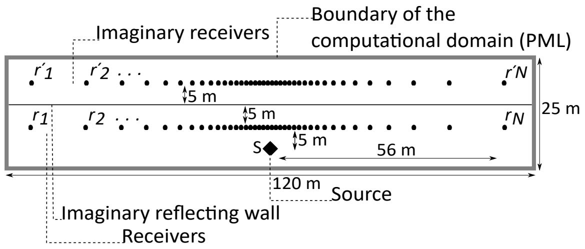

Figure 2: Schematic layout of the 2D PSTD domain where S is the source, r_n are the receivers used to simulate the direct sound and r'_n are the receivers that were used to synthesize the 1st order specular reflection.

2.4 HRTF processing and incorporation in PSTD

The HRTFs that were incorporated in the PSTD simulations are the TU Berlin KEMAR horizontal plane HRTFs measured at 3 m distance and in the angular range of 0° to 359° for every 1° [35]. The response of the loudspeaker used in the measurements was corrected for in the frequency range between 100 Hz-10 kHz [35].

HRTF databases have poor low frequency response (below 200 Hz) due to the poor low frequency performance of the measurement source and of the anechoic chamber [36]. Therefore, a correction needs to be applied to those frequencies. The fact that noise sources in the urban environment produce a lot of energy at low frequencies highlights the importance of this processing step for the auralization of urban environments.

The low frequency correction was applied using the method developed by Xie [36, 37]. Xie proposed a low frequency correction where the magnitude response of the HRTFs at frequencies below 100 Hz is set to a constant value based on the mean magnitude between 100-300 Hz and the phase is linearly interpolated. The correction was implemented with the code provided in [38].

Apart from the low frequency correction, diffuse-field equalization was also applied in order to remove the effects that are not incident-angle dependent like the ear canal resonance (KEMAR manikin has the microphones inside the ear canal) [39].

The HRTFs were incorporated in PSTD using the methodology developed in [33]. The HRTF source is modeled through spatial distributions in PSTD that relate to the spherical harmonics (SHs), and time-dependent functions are assigned to the spatial distributions in order to obtain the frequency content of the directivity. Since the computation is in 2D, circular harmonic functions (CH) were used instead of SH functions. CH functions are given by Fourier series and are the 2D analogue of SHs. CHs hold the same properties as SHs but for functions whose values vary only in one angular plane (e.g. 2D horizontal plane directivity). For more details regarding CHs, see Appendix 6.

2.5 BIR post-processing

The modeled source in PSTD did not excite the grid from time $t = 0$ s because the time function that was applied to the spatial distributions (see [33]) has some zero values at the beginning. In the simulation presented here, the source excited the grid after 30 ms, thus, the samples before this time limit were removed from the BIRs.

The BIR calculations include the characteristics of the source function in PSTD (Gaussian pressure distribution (Eq. 5)). To obtain a flat frequency response, a compensation filter was designed based on the response of the Gaussian pressure distribution using least-squares (LS) deconvolution with frequency-dependent regularization, which was implemented in time-domain [40, 41], and then was applied to the BIRs. The LS deconvolution minimizes the error between the target response p_0, which is a low-pass filtered delta function with cut-off frequency at 7.5 kHz (same as the maximum frequency of the source spectrum (see subsection 2.3)), and the transient response of the Gaussian pressure distribution that was recorded at 5 m distance p_1. A frequency weighting, which is referred to a regulaization filter b (time domain filter), is used to control the amount that the inverse filter will correct for across frequencies. b is a high-pass filter with cut-off at 7.5 kHz. This means that the regularization is the highest above 7.5 kHz, thus, the inverse filter will not try to correct the frequencies above 7.5 kHz. A regularization weighting factor β, which is a scalar parameter $\in [0, 1]$, is used to control the overall amount of the regularization. The inverse filter h_1 is calculated by the following expression:

$$h_1 = [\mathbf{P_1}^T\mathbf{P_1}+\beta\mathbf{B}^T\mathbf{B}]^{-1} \cdot \mathbf{P_1}^T p_0, \tag{1}$$

where $\mathbf{P_1}$ is the convolution matrix of p_1, $\beta = 0.4$ and $\mathbf{B}$ is the convolution matrix of the regularization filter b.

Another factor that needs to be considered is that the BIRs were computed in 2D. Hence, the pressure over distance reduces by a factor of $1/\sqrt{r}$ and not by $1/r$ as for 3D scenarios (where r is the distance from the source). In order to correct for this, the BIRs were multiplied by the following time-dependent correction factor $1/\sqrt{ct}$, where c is the speed of sound and t is the time vector of the BIRs samples.

Finally, a reflection factor of 0.9 was used for the wall. This was applied to the mirrored BIRs (calculations at r' in Figure 2) at all frequencies by multiplying their pressure values (in the time-domain) by this coefficient. A post-processed BIR is shown in Figure 3.

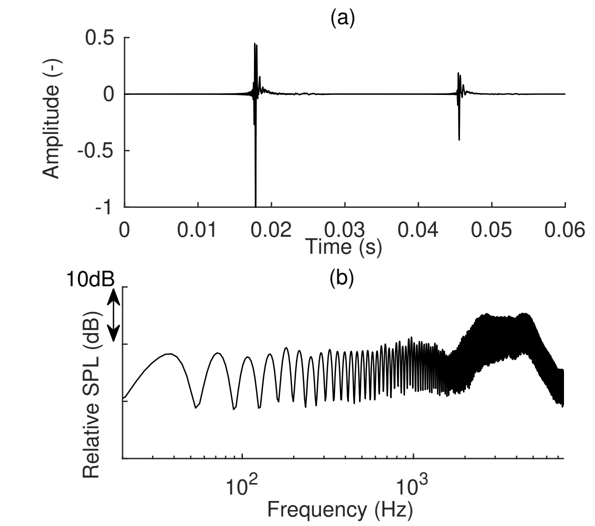

Figure 3: (a) Post-processed and normalized right ear BIR for the receiver at 35° and 6.1 m from the source including the corresponding mirrored image (see Figure 2) and (b) its frequency response.

3 Auralization methodology

3.1 Overview

In this section, the methodology used to auralize a car pass-by is presented. A synthesized car signal (see subsection 3.2) was split at increments based on the time it takes the car to travel between 3 discrete source locations. A sine window function was applied to these signals in order to achieve a smooth cross-fade between the signal from the source positions. Next, these signals were convolved with the corresponding BIRs (computed with the method presented in section 2) and shifted by the time that it takes the car to travel between two discrete neighbouring source positions. Finally, these signals were added together to create the final binaural auralization stimuli. The details of the auralization methodology is described in subsection 3.3. The limitations of the auralization methodology are outlined in the last subsection.

3.2 Dry car signal

The dry car signals used in this work were provided by Chalmers University of Technology and were synthesized with the Listen project simulator, in which project the same institute was partner[21]. Three different car signals were used for the auralizations: 1) a car moving at a speed of 50 km/h without its tonal components, 2) a car moving at a speed of 70 km/h with its tonal components, 3) a car moving with speed of 70 km/h without its tonal components. The dry signal of a car moving with 50 km/h with tonal components sounded very unnatural, so it was not used in the auralizations. The tonal components refer to the low frequency tonal components produced by the engine. The spectrum of the dry car signals for 70 km/h speed is shown in Figure 1.

3.3 Cross-fading between discrete source positions

The auralization methodology aims to achieve a continuous switch between the discrete source positions such that the car is continuously moving along the line (see Figure 2). The dry car signal was split into blocks of lengths equal to the time it takes the car to travel from location r_{n-1} to r_{n+1}. The signal blocks were chosen randomly so that they were not correlated. Afterwards, the signals were multiplied by a non-symmetrical cross-fade sine window in order to achieve a continuous and smooth switch between signals from the discrete source positions (see Figure 4 (a)). The windows were designed such that their center is at the location of r_n and their length is equal to the time it takes the car to travel from r_{n-1} to r_{n+1}. The window function is shown in the equation below:

$$w(n) = \begin{cases} \sin\left(\frac{\pi}{2} \cdot \frac{n}{N_1}\right) & \text{for } 0 \leq n \leq N_1 \\ \sin\left(\frac{\pi}{2} + \frac{\pi}{2} \cdot \frac{n-N_1}{N_2}\right) & \text{for } N_1 + 1 \leq n \leq N_1 + N_2, \end{cases} \quad (2)$$

where N_1 is the time (in samples) it takes for the car to travel from r_{n-1} to r_n and N_2 is the time it takes the car to travel from r_n to r_{n+1}. Next, the windowed dry car signals were convolved with their corresponding BIRs (r_n) and shifted by the time it takes the car to travel between r_{n-1} and r_n as shown in Figure 4 (b). Finally, the signals were summed in order to create the final auralization signals (see Figure 5).

The car signal blocks were handled as being uncorrelated. Therefore, the sine window was applied in order to achieve an equal power cross-fade and to avoid audible fluctuations of the signal loudness [42].

As mentioned in subsection 2.3, the BIRs were calculated every 0.25°. However, the auralizations with such small spacing produced audible artefacts (repetitive amplitude fluctuations of low frequency noise bursts). This artefact is produced due to the fast switch between the discrete positions and due to the window function, which assumes that the signals are uncorrelated. However, the short signal segments contain some correlated components, mainly at low frequencies (i.e. tonal components shown in Figure 1), which the window did not cross-fade well. For auralizations with the dry car signal moving at 70 km/h with tonal components, this artefact was still present even for a source spacing of 1°. The shortest angular increment that did not produce any audible artefact was 2°. For dry car signals without tonal components, the shortest increment that did not produce any artefact was 1°.

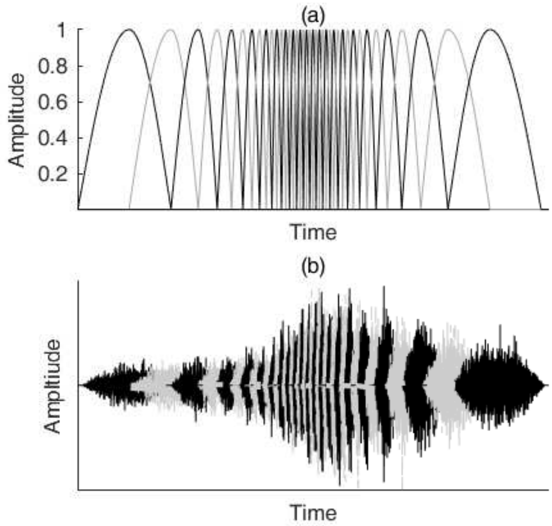

Figure 4: (a) Window functions according to equation (2) that were multiplied by the dry car signal blocks. (b) Overlapping windowed car signals blocks convolved with their corresponding BIR.

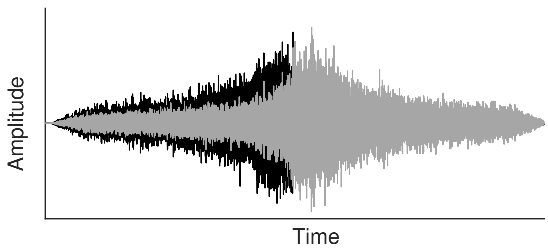

Figure 5: Final auralization signals for left (black) and right (grey) ears. These signals were created by adding the signal blocks shown in Figure 4 (b) (which has to be done for the left and right signal separately).

3.4 Limitations of the auralization method

The main limitations of the presented auralization method are: 1) The BIRs were computed in a 2D domain and not in 3D; 2) The Doppler effect was not simulated; 3) Car directivity was not incorporated.

A limitation that arises from the 2D domain, is the absence of the ground reflection. The scenario presented in this paper is a close approximation of a source on a dense asphalt ground surface.

In the literature, simplified methods have been developed to model the Doppler effect using variable delay lines and interpolation of the signal (asynchronous resampling process) [4, 43, 44]. These methods have potential applicability here. The Doppler shift can be applied to the dry car signal blocks, but it should be ensured that the same Doppler shift is applied to both signals in the overlapping region (see Figure 4).

As mentioned in the introduction of this paper, the purpose of the listening experiments (described in detail in section 4) is to investigate the audibility of switching between the discrete source positions when the distance between the discrete source positions is increased. The Doppler effect is a parameter that will affect the perception of the auralized signals, but the main focus of the listening tests presented in this paper is the perception of the switch between the discrete source positions. Thus, it was decided to exclude the Doppler effect from the listening tests stimuli and focus entirely on the switch between the discrete source positions. The impact of the Doppler effect on the perception of the switch between the discrete source positions when the source increment is increased is left for future work.

4 Subjective test set up

4.1 Overview

The listening test, which was conducted to investigate whether increasing the angular spacing between the discrete source positions affects the perception of the auralizations, is described here. The section starts with an explanation regarding the choice of the reference increment that was used in the listening tests followed by a presentation of the test signals. Next, the test methodology is outlined. Finally, the information regarding the test subjects is presented.

4.2 Reference increment

The minimum audible angle, which is both dependent on the type of stimulus and the direction, is the smallest angular separation that humans can detect between two sound sources. It is 1° for the horizontal frontal plane [45]. However, since the 1° increment produced artefacts, especially in the auralizations with the car signal with tonal components (see subsection 3.3), it was not used as the reference increment.

As mentioned in subsection 3.3, the 2° increment was the smallest angular increment that produced a smooth and continuous cross-fade for cars moving at speeds of 50 km/h and 70 km/h (including the tonal components). Therefore, the angular increment of 2° was chosen as a reference for all test conditions.

It should be noted that 2° is also very close to the lowest reported minimum audible movement angle (MAMA) [46, 47, 48], defined as the minimum angular increment that a source needs to move in order for it to be discriminated from a static source or a source moving in the opposite direction [47]. Reported angles of MAMA vary between 1.5°-21° and are dependent on the velocity and the bandwidth of the moving source [46]. MAMA decreases with increasing frequency bandwidth and increases with velocity. Also, the review conducted by Carlile and Leung [46] based on the previous work of [49, 50, 51, 52, 48] suggests that MAMA is 2 to 3 times larger than the minimum audible angle of a static source in the horizontal plane when measured under the same conditions. This supports the conclusion that the choice of 2° as the reference increment is also a motivated choice from the MAMA perspective.

4.3 Test signals

The reference auralization was compared against different increments for the test scenarios with buildings absent and with a long building block with flat wall behind the car, as well as for different speeds and car signals, as presented in Table 1.

Table 1: List of the 11 different test conditions and their notation.

Speed (km/h)	Tones	Environment	Angles tested°	Notation
50	No	Buildings absent	4,6,8,10	50NA4,6,8,10
50	No	Building	4	50NB4
70	No	Buildings absent	4,6,8	70NA4,6,8
70	Yes	Buildings absent	4,6,8	70TA4,6

These test signals were selected based on pilot tests with 3 subjects. These preliminary tests aimed to indicate the difficulty of detecting the difference between the auralizations with the reference increment and the auralizations with larger increments in order to aid in the selection of the test signals that would be used in the test. For the environment with a long building with flat wall behind the car (modeled with the mirrored image calculated at r'_n in Figure 2), only one increment of 4° was tested, because results from the preliminary test showed that when the spacing between the sources was increased it was easier to perceive the difference due to the comb filter effect caused by the reflection (see Figure 3). Also, the 70NA and 70TA were not tested for 10° increment in order to shorten the duration of the test.

In Figure 6 the spectrum of the 70NA8 vs the 2° reference is plotted. The spectrum of the two test signals is very similar. Differences are noticeable at low frequencies below 50 Hz.

Since the signals will be compared against each other in a same/different task (see subsection 4.4), short signals will make the task easier for the subjects (low short-term memory load). Therefore, the test signals from Table 1 were truncated for a moving distance from -28.8 m left to 28.8 m right from the subject, even-though the length of the simulation grid was 120 m (Figure 2). This truncation led to test signals of 2.9 s for the speed of 70 km/h and 3.9 s for 50 km/h. The sound files were stored in WAV format at 44.1 kHz and 32 bits.

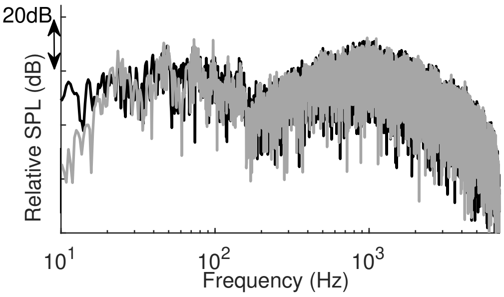

Figure 6: Spectrum of auralization signals 70TA2 (reference) (black line) and 70TA8 (grey line). The spectrum was computed by taking the FFT of the total duration (2.9 s) of the test signals. See Table 1 for notations. SPL stands for sound pressure level.

4.4 Test methodology and test interface

A same/different test [53] was conducted in which the subjects were presented with two stimuli and had to indicate whether the stimuli were the same or different. The reason that a same/different test was chosen is that the task is subject-friendly because the responses "same" and "different" are familiar and easily understandable [54]. The design and the analysis of the test was based on signal detection theory (SDT). The basics of SDT that were used in this research can be found in section 4.4 of Georgiou [55] and further details can be found in [56, 57]. Each of the 11 test conditions presented in Table 1 consisted of 50 trials; 25 of those trials were reference signal versus reference signal and the remaining 25 reference signal versus a signal computed from a larger angular increment. Each test condition was tested individually; i.e. subjects had to complete 50 trials of that condition in one run. For every subject the test conditions order and the trials of each condition were randomized.

The test interface was programmed in Cycling '74 Max 7 [58]. The stimuli were played once with a 200 ms break between them. The subjects were only allowed to listen to each trial once. Subjects were given both oral and written instructions. Subjects were encouraged to take as many breaks as they wanted. They were also given the opportunity to stop the test and continue on another day if they felt tired. After completing the 50 trails of each test condition, they were asked to write a short feedback on how difficult they found the test and how they identified the differences between the stimuli. This step was not mandatory.

The test was implemented inside a quiet room at TU Eindhoven. The stimuli were presented via Sennheiser HD 800 headphones, which were connected to a laptop via the E-MU 0204 USB audio interface. The maximum sound level that the subjects were exposed to was approximately 84 dB(A) (instantaneous sound level), which occurred when the car was directly in front of the receiver (max peak in Figure 5). Subjects were asked if the level was comfortable for them, and they were allowed to decrease or increase it. However, their chosen level had to remain the same over the whole duration of the test. On average, subjects completed the 50 trials of each test condition within 7-8 minutes, and the total test duration was around 1 hour 40 minutes including breaks.

4.5 Test subjects

Fourteen subjects participated in the listening test. One subject quit the test after 20 minutes because (s)he claimed that (s)he was unable to perceive any differences between the stimuli. Therefore, the analysis is based on the 13 subjects who completed the tests. Four subjects completed the test in two different days because they found the task demanding and required a long break. They were strictly advised not to discuss the experiments with other subjects. From the subjects 10 were male and 3 were female. The average age of the participants was 34.2 years. Ten of the subjects had participated previously in a perceptual test. Eleven of the subjects have a profession related to sound and acoustics. None of the subjects reported any hearing problems. Ten of the subjects were fine with the playback level (maximum instantaneous SPL of 84 dB(A), one of them reduced it to 79 dB(A), one to 81 dB(A) and one increased it to 86 dB(A).

5 Results and discussion

A fundamental measure of difference in the perception of the two stimuli is the d' value, which is also referred to as sensitivity. The sensitivity d' is a measure of the subjects' ability to identify whether the stimuli are the same or different; the greater the d', the easier it is for the subjects to recognize the differences [56]. Here, the limit to consider two auralizations to be different was $d' \geq 1$, which is a limit commonly used in the literature [59]. The equation to compute d' for the same/different test conducted here is shown below and is taken from chapter 9 of Macmillan & Creelman [56]:

$$\begin{aligned} p_c &= \Phi\left(\frac{z(H) - z(F)}{2}\right) \\ d' &= 2z\left(\frac{1}{2}\left\{1 + \sqrt{2p_c - 1}\right\}\right) \end{aligned} \tag{3}$$

where p_c is the unbiased proportion correct, H and F are the hit and false-alarm rates, Φ is the cumulative normal distribution function and z is the inverse cumulative normal distribution. The hit in the same/different test is defined when the two stimuli are different and the response is "Different". The false-alarm is defined when the stimuli are the same and the response is "Different".

The d' for every test condition, shown in Table 1, was computed separately for each subject. Figure 7 demonstrates with Box-and-Whisker plots the d' scored in every test condition. In Figure 8 the d' scores for every subject are plotted (each colored marker represents a different test condition).

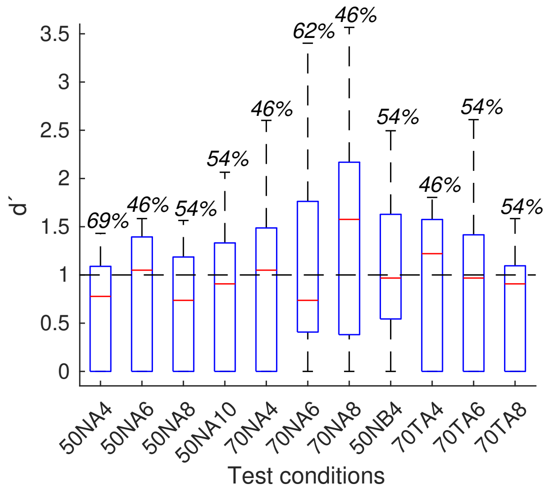

Figure 7: (Colour online) Box-and-Whisker plots of the d' scored for every test condition (x-axis). The plot shows the lower and upper quartile values, and the median value (red line). The whiskers represent the remainder of the data. The numbers on top of the plots show the percentage of subjects who scored $d' < 1$. See Table 1 for notations.

Results show that differences were identifiable, but they are difficult to notice. On average, from all test conditions, 52.3% (numbers above each whisker line in Figure 7) of the subjects scored $d' < 1$, meaning that they found it very difficult to identify any differences between the reference auralizations and the auralizations with larger increments. Subject 8 scored $d' < 1$ in two test conditions only and subjects 6, 7 and 13 scored $d' < 1$ in four conditions. The rest of the subjects scored $d' < 1$ on average in 6.8 test conditions. Generally, the subjects did not score higher values of d' in the test conditions with larger angular source spacing, as seen in Figure 8. For example, subject 12 scored $d' > 1$ for the conditions 70TA4 and 70TA6 showing a good discriminability, but scored $d' = 0$ in the 70TA8 condition. Multiple examples like the one described previously can be observed in Figure 8.

Paired t-tests were also conducted in order to investigate if there were significant differences between the results of the test conditions of different angular increment i.e. between the test condition shown on each row of Table 1 e.g. 70TA4 vs 70TA6, 70TA4 vs 70TA8, 70TA6 vs 70TA8. The p-values were larger than 0.05 (varied between 0.35-0.82), meaning that

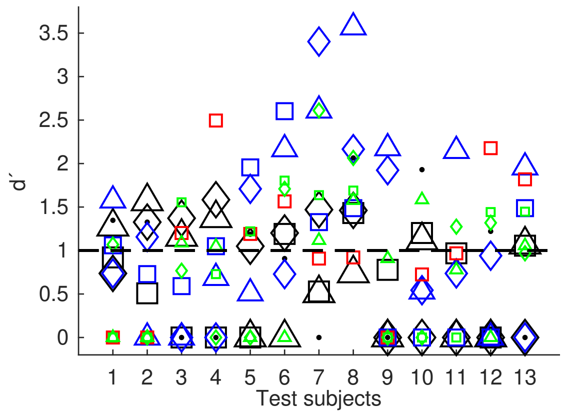

Figure 8: (Colour online) Plot of the d' scored by each subject for the 11 different test conditions. The colors represent the first 3 columns of Table 1 (speed, car signal, building block behind the car or not). Black: 50NA, Blue: 70NA, Red: 50NB, Green: 70TA. The markers represent the angular increments. Square: 4°, Diamond: 6°, Triangle: 8°, Dot: 10°.

there are no significant differences between them. This is also visible in Figure 7 from the amount of overlap between the box-plots. Before performing the t-tests, Kolmogorov-Smirnov tests were conducted to test for normality. The p-values varied between 0.32-0.99, so the assumptions to conduct the t-test were met. T-tests were also conducted between the test conditions with the same angular increment. In all conditions $p > 0.05$, which means that there were no significant differences between the different test conditions of the same angular increment. The p-values below 0.35 were at 50NA8 vs 70NA8 ($p = 0.066$) and 70NA8 vs 70TA8 ($p = 0.075$).

The results highlight three main aspects:

1. Speed and tonal components do not affect the results.

2. The specular reflection did not play a role in the perception of the switch between the BIRs for the increment of 4°. This result was unexpected because the subjects who participated in the preliminary tests could detect the differences in this scenario much more easily.

3. The fact that there is no significant difference in the sensitivity measure between the test conditions with the lowest angular increment and the highest suggest that the increment might be further increased without affecting the subjects' perception.

It should be noted that the subjective tests were performed for short duration around the time of passage (car passing-by from -28.8 m left to 28.8 m right from the subject). Therefore, the finding from the listening tests should not be extrapolated for distances further than the ones tested, as well as for different car speeds.

In the feedback that all subjects gave after each test condition and at the end of the test, 10 out of 13 subjects mentioned that it was very difficult to perceive any differences, and as such in many cases they simply guessed. All of the subjects mentioned that the task was demanding and that they had to be very focused to identify any possible differences. This finding could also be an indication that subjects may have lost interest during the test soon after the task revealed itself to be so demanding. However, the results did not indicate any rise or drop of the subjects' performance with time. Subjects also wrote other remarks on the feedback forms, but it was not possible to extract any consistent information because the responses were very random. Also, the performance of the subjects who completed the test in two days did not vary. Finally, the results are valid for the playback levels used in this listening test. It might be possible that at higher playback levels the spectral differences at low frequencies between the reference auralization and the auralizations with larger source angular spacing (see Figure 6) could be more perceivable, thereby making the task easier.

6 Conclusions

In this paper a methodology to auralize a car pass-by using BIRs computed with the PSTD method was presented, in order to be able to accurately model low frequency urban sound propagation. This is the first time that BIRs computed with a wave-based method were used for the auralization of car pass-by. The auralizations were performed for the simplified scenarios where buildings are absent, and for an environment where a long flat wall is located behind the car. The auralization methodology can be split in the following steps:

- BIRs were calculated with PSTD and stored based on a 0.25° angular increment between the source position and receiver positions. Thanks to the acoustic reciprocity principle, only one simulation is needed, in which the modeled source is placed at the physical receiver location and the receivers are placed at actual car positions.

- A dry car signal was split at increments based on the time it takes the car to travel between 3 discrete source locations. A sine window function was applied to these signals in order to achieve a smooth cross-fade between the source positions.

- The signals of step 3 were convolved with the corresponding BIRs and shifted by the time that it takes the car to travel between two discrete neighbouring source positions. Finally, these signals were added together to create the final auralization stimuli.

This auralization methodology has not been implemented previously. The main limitations of the auralization method in its current implementation are: 1) The BIRs were computed with a 2D PSTD simulation and not 3D; 2) The Doppler effect was not modeled; 3) The car source directivity was not incorporated.

A subjective evaluation was carried out in order to investigate whether the subjects can perceive any differences when the angular spacing between the discrete source positions is increased. No previous research has conducted such an investigation.

The main outcomes from the subjective analysis were:

- Differences exist between the reference auralization and the auralization with larger increments, but they are difficult to notice and are not recognized by all subjects.
- On average, 52.3% of the subjects found it almost impossible to identify any differences between the reference auralization and the auralization with larger increment, even for the 8° and 10° angular spacing.
- The responses of the subjects did not significantly change when the increment was increased.

For future work, the auralization methodology presented here should be applied in 3D realistic street environments.

Appendix

Circular harmonics (CHs) are real functions. Any 2D directivity function $d(\phi)$ can be expressed as a weighted sum of series of CH functions as in the following equation:

$$\begin{aligned} d(\phi) &= \sum_{n=0}^{\infty} L_n G_n(\phi), \\ G_n &= \frac{1}{\sqrt{2\pi}} \qquad \text{for n=0,} \\ G_n &= \frac{1}{\sqrt{\pi}}\big(cos(n\phi), sin(n\phi)\big) \qquad \text{for n>0,} \end{aligned} \tag{4}$$

where ϕ is the angle ($0\leq\phi\leq2\pi$), G_n are the CH functions of degree n, and L_n are the CH weights. In numerical calculations the summation of Eq. 4 for the CH functions is truncated at degree l (see [60]). The decomposition of $d(\phi)$ into CH weights is implemented with the same methodology as used for SH weights in [33].

For sound propagation calculations with PSTD in this work, the design of the initial spatial distributions (ISDs) related to CHs is implemented by the combination of higher order spatial derivatives of the initial Gaussian pressure distribution (Eq. 5) in the x and y directions [61] (similarly to the modeling of the ISDs related to SHs presented in [33]).

$$p(r) = e^{-\frac{|\mathbf{r}-\mathbf{r}_s|^2}{w^2}} \qquad \text{at} \quad t=0, \tag{5}$$

where $|\mathbf{r}-\mathbf{r}_s| = \sqrt{(x-x_s)^2+(y-y_s)^2}$, (x_s, y_s) are the coordinates of the distribution center and w determines the width of the Gaussian signal (see [33]). ISDs related to CHs are presented in Table 2.

Table 2: Design of initial spatial distributions related to CH functions of different degree.

Degree n:	Initial distribution:
$n=0$	$e^{-\frac{\lvert\mathbf{r}-\mathbf{r}_s\rvert^2}{w^2}} G_0$
$n=1$	$-\frac{2\lvert\mathbf{r}-\mathbf{r}_s\rvert}{w^2} e^{-\frac{\lvert\mathbf{r}-\mathbf{r}_s\rvert^2}{w^2}} G_1$
$n=2$	$\frac{4\lvert\mathbf{r}-\mathbf{r}_s\rvert^2}{w^4} e^{-\frac{\lvert\mathbf{r}-\mathbf{r}_s\rvert^2}{w^2}} G_2$
$\vdots$	$\vdots$

References

[1] J.Y. Jeon, P.J. Lee, J. You, J. Kang: Perceptual assessment of quality of urban soundscapes with combined noise sources and water sounds. J. Acoust. Soc. Am. **127**-3 (2010) 1357-1366.

[2] M. Kleiner, B.-I. Dalenbäck, P. Svensson: Auralization-An overview. J. Audio Eng. Soc. **41**-11 (1993) 861-875.

[3] A. Krokstad, S. Strom, S. Sørsdal: Calculating the acoustical room response by the use of a ray tracing technique. J. Sound Vib. **8**-1 (1968) 118-125.

[4] R. Pieren, T. Bütler, K. Heutschi: Auralization of accelerating passenger cars using spectral modeling synthesis. Appl. Sci. **5**-6 (2016) 1-27.

[5] M. Vorländer: Auralization - Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality. First, Springer, 2008.

[6] A. Pompei, M.A. Sumbatyan, N.F. Todorov: Computer models in room acoustics: The ray tracing method and the auralization algorithms. Acoust. Phys. **55**-6 (2009) 821-831.

[7] J.B. Allen: Image method for efficiently simulating small-room acoustics. J. Acoust. Soc. Am. **65**-4 (1979) 943-950.

[8] D.O. Elorza: Room acoustics modeling using the ray-tracing method: implementation and evaluation. University of Turku, 2005.

[9] H. Jonasson: Acoustic source modelling of Nordic road vehicles, SP Rapp. **12** (2006).

[10] H. Jonasson, U. Sandberg, G. van Blokland, J. Ejsmont, G. Watts, M. Luminari: Source modelling of road vehicles, Deliv. 9. Harmon. Proj. (2004) 1-52.

[11] M. Hornikx, R. Waxler, J. Forssén: The extended Fourier pseudospectral time-domain method for atmospheric sound propagation. J. Acoust. Soc. Am. **128**-4 (2010) 1632-1646.

[12] M. Hornikx: Ten questions concerning computational urban acoustics. Build. Environ. **106** (2016) 409-421.

[13] S. Kephalopoulos, M. Paviotti, F. Anfosso-Lédée: Common noise assessment methods in europe (CNOSSOS-EU). Eur. Comm. Dir. Jt. Res. Cent. (2012).

[14] P. Luizard, J.-D. Polack, B.F.G. Katz: Sound energy decay in coupled spaces using a parametric analytical solution of a diffusion equation. J. Acoust. Soc. Am. **135**-5 (2014) 2765-2776.

[15] V. Valeau, J. Picaut, M. Hodgson: On the use of a diffusion equation for room-acoustic prediction. J. Acoust. Soc. Am. **119**-3 (2006) 1504-1513.

[16] N. Xiang, Y. Jing, A.C. Bockman: Investigation of acoustically coupled enclosures using a diffusion-equation model.. J. Acoust. Soc. Am. **126**-3 (2009) 1187-1198.

[17] S.M. Pasareanu, M.C. Remillieux, R.A. Burdisso: Energy-based method for near-real time modeling of sound field in complex urban environments. J. Acoust. Soc. Am. **132**-6 (2012) 3647-3658.

[18] P. Luizard, J.-D. Polack, B.F.G. Katz: Auralization of coupled spaces aased on a diffusion equation model. Proc. Sound Music Comput. Conf. (2013) 616-621.

[19] V. Välimäki, J.D. Parker, L. Savioja, J.O. Smith, J.S. Abel: Fifty years of artificial reverberation. IEEE Trans. Audio. Speech. Lang. Processing **20**-5 (2012) 1421-1448.

[20] K. Kowalczyk: Boundary and medium modelling using compact finite difference schemes in simulations of room acoustics for audio and architectural design applications. Queenś University Belfast, 2008.

[21] J. Forssén, T. Kaczmarek, P. Lundén, M.E. Nilsson, J. Alvarsson: Auralization of traffic noise within the LISTEN project - Preliminary results for passenger car pass-by. Proc. Euronoise, Edinburgh, UK, 2009.

[22] A. Peplow, J. Forssén, P. Lundén, M.E. Nilsson: Exterior auralization of traffic noise within the LISTEN project. Proc. Forum Acust., Aalborg, Denmark, 2011: 665-669.

[23] M.E. Nilsson, M. Rådsten-Ekman, J. Alvarsson, P. Lundén, J. Forssén: Perceptual validation of auralized road traffic noise, Proc. Inter-Noise, Osaka, Japan, 2011

[24] J. Maillard, J. Jagla: Auralization of non-stationary traffic noise using sample based synthesis - Comparison with pass-by recordings. Proc. Internoise, Saint-Martin-d'Hères, 2012.

[25] J. Maillard, A. Kacem: Auralization applied to the evaluation of pedestrian and bike paths in urban environments. Proc. Internoise, Hamburg, Germany, 2016: 3473-3482.

[26] E.M. Viggen, S. Audun, J. Venerød, H. Olsen: Development of an outdoor auralisation prototype with 3D sound reproduction. Proc. 18th Int. Conf. Digit. Audio Eff., Trondheim, Norway, 2015.

[27] R. Mehra, N. Raghuvanshi, L. Antani, A. Chandak, S. Curtis, D. Manocha: Wave-based sound propagation in large open scenes using an equivalent source formulation. ACM Trans. Graph. **32**-2 (2013) 19:1-19:13.

[28] N. Raghuvanshi, J. Snyder, R. Mehra, M. Lin, N. Govindaraju: Precomputed wave simulation for real-time sound propagation of dynamic sources in complex scenes. ACM Trans. Graph. **29**-4 (2010) 68:1-68:11.

[29] H. Yeh, R. Mehra, Z. Ren, L. Antani: Wave-ray coupling for interactive sound propagation in large complex scenes. ACM Trans. Graph. **32**-6 (2013) 165:1-165:11.

[30] M. Hornikx: Numerical modelling of sound propagation to closed urban courtyards. Chalmers University of Technology, 2009.

[31] J.S. Hesthaven: S. Gottlieb, D. Gottlieb, Spectral methods for time-dependent problems. Cambridge University Press, 2007.

[32] C. Bogey, C. Bailly: A family of low dispersive and low dissipative explicit schemes for flow and noise computations. J. Comput. Phys. **194**-1 (2004) 194-214.

[33] F. Georgiou, M. Hornikx: Incorporating directivity in the Fourier pseudospectral time-domain method using spherical harmonics. J. Acoust. Soc. Am. **140**-2 (2016) 855-865.

[34] A.D. Pierce: Acoustics, an Introduction to its physical principles and applications. The Acoust. Soc. Am. , 1989.

[35] H. Wierstorf, M. Geier, A. Raake, S. Spors: A Free database of head-related impulse response measurements in the horizontal plane with multiple distances. Proc. Audio Eng. Soc. 130th Conv., 2011.

[36] B. Xie: Head-Related Transfer Function and Virtual Auditory Display. Second ed., J. Ross Publishing, 2013.

[37] B. Xie: On the low frequency characteristics of head-related transfer functions. Chinese J. Acoust. **28**-2 (2009) 1-13.

[38] B. Xie: www.github.com/spatialaudio/lf-corrected-kemar-hrtfs. (2016).

[39] W.G. Gardner: 3-D audio using loudspeakers, Springer US, 1998.

[40] Z. Schärer, A. Lindau: Evaluation of equalization methods for binaural signals. Proc. Audio Eng. Soc. 126th Conv., Munich, Germany, 2009.

[41] S. Norcross, M. Bouchard, G. Soulodre: Inverse filtering design using a minimal-phase target function from regularization. Proc. Audio Eng. Soc. 121st Conv., San Francisco, USA, 2006.

[42] A.B. Greenblatt, J.S. Abel, D.P. Berners: A hybrid reverberation crossfading technique. Proc. ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., Texas, USA, 2010: 429-432.

[43] J. Smith, S. Serafin, J. Abel, D. Berners: Doppler simulation and the Leslie. Proceeding Work. Digit. Audio Eff., Hamburg, Germany, 2002: 188-191.

[44] H. Strauss: Implementing Doppler shifts for virtual auditory environments. Proc. Audio Eng. Soc. 104th Conv., Amsterdam, The Netherlands, 1998.

[45] J. Blauert, Spatial hearing: The psychophysics of human sound localization. Revised ed., MIT Press, 1997.

[46] S. Carlile, J. Leung: The perception of auditory motion. Trends Hear. **20** (2016) 1-19.

[47] D.W. Chandler, D.W. Grantham: Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity. J. Acoust. Soc. Am. **91**-3 (1992) 1624-1636.

[48] D.R. Perrott, K. Saberi: Minimum audible angle thresholds for sources varying in both elevation and azimuth. J. Acoust. Soc. Am. **87**-4 (1990) 1728-1731.

[49] W.O. Brimijoin, M.A. Akeroyd: The moving minimum audible angle is smaller during self motion than during source motion. Front. Neurosci. 8 (2014) 1-8.

[50] D.W. Grantham, B.W.Y. Hornsby, E.A. Erpenbeck: Auditory spatial resolution in horizontal, vertical, and diagonal planes. J. Acoust. Soc. Am. **114**-2 (2003) 1009-1022.

[51] D.W. Grantham: Detection and discrimination of simulated motion of auditory targets in the horizontal plane. J. Acoust. Soc. Am. **79**-6 (1986) 1939-1949.

[52] J.D. Harris, R.L. Sergeant: Monaural-binaural minimum audible angles for a moving sound source. J. Speech, Lang. Hear. Res. **14**-3 (1971) 618-629.

[53] H.S. Lee, D. van Hout, M. Hautus, M. OḾahony: Can the same-different test use a β-criterion as well as a τ-criterion?. Food Qual. Prefer. **18**-4 (2007) 605-613.

[54] J. Liang, Y. Zhou, Z. Liu: Examining the standard model of signal detection theory in motion discrimination. J. Vis. **16**-7 (2016) 1-12.

[55] F. Georgiou: Modeling for auralization of urban environments: Incorporation of directivity in sound propagation and analysis of a framework for auralizing a car pass-by. Eindhoven University of Technology, 2018.

[56] N.A. Macmillan, C.D. Creelman: Detection theory: A userś guide. Second ed., Lawrence Erlbaum Associates, Publishers, 2005.

[57] H. Stanislaw, N. Todorov: Calculation of signal detection theory measures. Behav. Res. Methods, Instruments Comput. **31**-1 (1999) 137-149.

[58] Max7: https://cycling74.com/

[59] S. Le, T. Worch: Analyzing sensory data with R. Chapman and Hall/CRC, 2014.

[60] R. Duraiswami, D.N. Zotkin, N. Gumerov: Interpolation and range extrapolation of HRTFs. Proc. IEEE Int. Conf. Acoust. Speech Speech Process., Montreal, Canada, 2004: 45-48.

[61] S. Sakamoto: Directional sound source modeling by using spherical harmonic functions for finite-difference time-domain analysis. Proc. ICA 2013, Montreal, Canada, 2013.