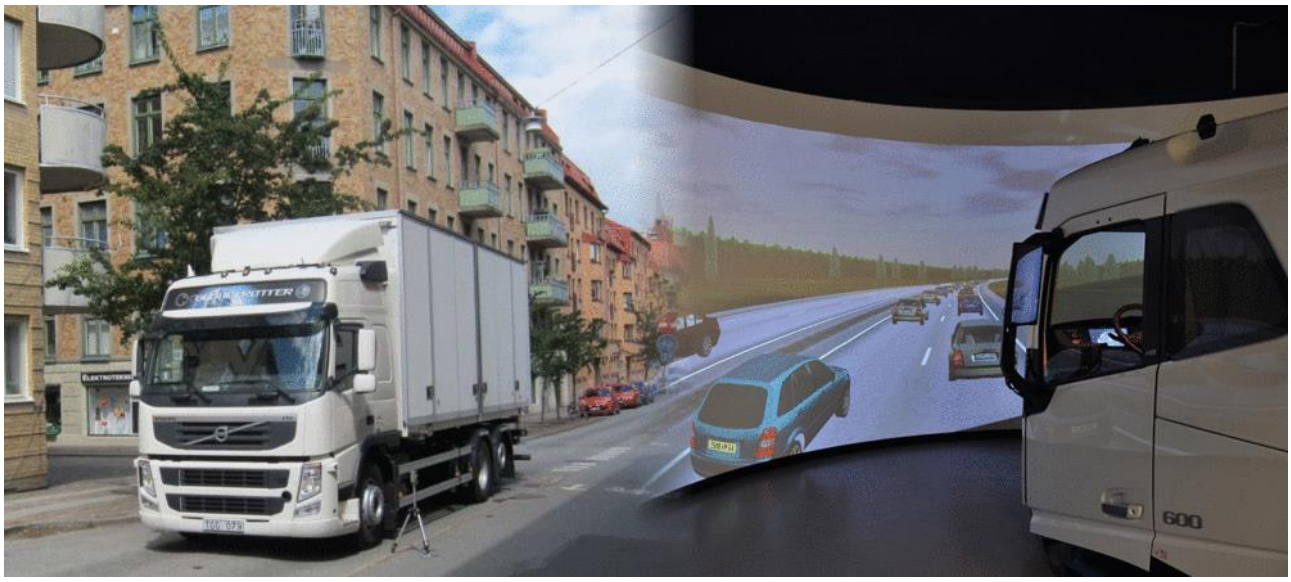
CHALMERS
UNIVERSITY OF TECHNOLOGY



Auralization Methodology for External and Internal Truck Sounds

Master's Thesis in the Master's programme in Sound and Vibration

WIKTOR ERIKSSON AND DANIEL NILSSON

Department of Civil and Environmental Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2016
Master's Thesis BOMX02-16-132

Auralization Methodology for External and Internal Truck Sounds

Master's Thesis BOMX02-16-132

Department of Civil and Environmental Engineering
Division of Applied Acoustics
Chalmers University of Technology
SE-41296 Göteborg
Sweden

Tel. +46-(0)31 772 1000

Reproservice / Department of Civil and Environmental Engineering
Göteborg, Sweden 2016

Auralization Methodology for External and Internal Truck Sounds
Master's Thesis in the Master's programme in Sound and Vibration
WIKTOR ERIKSSON and DANIEL NILSSON
Department of Civil and Environmental Engineering
Division of Applied Acoustics
Chalmers University of Technology

Abstract

Acoustical simulations are useful for work in community noise as well as sound quality. Applications within the respective fields are the auralization of urban vehicle pass bys and the synthesis of engine noise.

Endeavours to further develop more efficient transports have created an interest in external sound propagation within urban environments. A model has been constructed for this purpose, auralizing semi-anechoic recordings into a moving source on a generic city street. The amplification from the surroundings have been included as gain factors produced from a ray traced impulse response. Propagation phenomena such as air attenuation, Doppler effect, façade reduction and an indoor room impulse response have been integrated. Output are indoor and outdoor sound files as well as sound pressure level graphs over time and frequency spectra at the peak pressure time instance.

To improve the immersion in a driving simulator, the possibility of a sound quality improvement through the use of granular synthesis has been evaluated. In-cab recordings have been manually divided into sections of either acceleration or constant speed based on CAN bus data. An algorithm extracted grains from the sections with pitch periods dependant on engine rpm. Synthesis was performed by overlap-adding based on following a desired rpm profile.

Evaluation of the sound files resulting from both applications has been done with a listening test. The urban pass by results were also compared to measurements done in previous work at Landsvägsgatan in Göteborg. In the listening test, sounds were rated mainly by realism. A semantic differential test of realism, annoyance and loudness was used for the urban pass bys as well as a custom pairwise comparison. For the synthesised in-cab engine sounds, a direct pairwise comparison with the existing simulator sounds was done.

Keywords: Auralization, Source Modeling, Ray Tracing, Granular Synthesis, Pitch-Synchronous Overlap-Add (PSOLA), Digital Signal Processing

Contents

Acknowledgements

To our Chalmers supervisor Jens Forssén, the mastermind behind the auralization set up which we merely dusted off and bundled together into one application. For his guidance and for managing to be available despite his tight schedule and travels.

To our Volvo Group supervisor Krister Fredriksson for the help and encouragement throughout the project and for the occasional injection of realism into our optimistic time management.

To Patrik Höstmad at Chalmers University, for helping us understand the Landsvägsgatan measurements and for enduring our surprise consultation visits.

To Lars Hansson of LH akustik for, among other great insight and advice, the huge contribution with an "impulse train" algorithm which saved us (literally) weeks of staring at a Matlab "Busy" message.

To Alice Hoffmann for helping us navigate the jungle of listening test set up and literature on short notice.

To our classmates and the staff of Volvo Group who helped out with our listening test on such short notice, despite the lack of movie tickets.

1. Introduction

Transporting goods outside of normal work hours, i.e. night-time and mornings, is an example of a way to increase the traffic flow and reduce the congestion during day-time. In order to make that possible, the people living in the cities need to be open to the acoustic changes that follows and the vehicles need to be adjusted in order to be more acoustically acceptable (Fredriksson 2015). Auralization is a technique for creating and using virtual sounds to simulate sound events and environments. The use of the technique is widely spread, from city planning to virtual reality applications, meaning that it could be used as a tool when trying to get a better understanding of what type of vehicle set-up is necessary for this type of distribution and to help a sustainable urban development.

Driving simulators provide a good possibility to test experimental features for vehicles in a controlled environment. The best feedback is likely to come from individuals with every day experience of driving the vehicles in question. To make it easier to focus on the feature being tested, a degree of immersion is required. Engine sounds are a big factor for realism and has been raised as requiring improvement. Granular synthesis of sound is a well established method and is worth evaluating for this purpose.

1.1. Background

Volvo Group together with Chalmers, YKI and JABA had a project called "Quieter transport vehicles for more efficient distribution" that was concluded in 2015. The purpose of that project was to get a better understanding of how to create a quieter transport corridor for goods distribution through areas where people live (Fredriksson 2015). At the same time a project ran in parallel, made by Brett Seward, called "Auralization of a Heavy-Duty Truck with a Hybrid Engine using a Granular Approach". The project investigated the granular source model, i.e using small recorded signal segments and combining these to synthesize an engine sound for a heavy-duty truck with a hybrid engine (Seward 2014). Both projects generated information and insight that Volvo want to be able to use in their product development.

An area of interest was the use of auralization and to use this to evaluate truck sounds inside the vehicle, at the street and indoors in flats. The concluded project from 2015 had auralizations that was made for urban pass-by situations for heavy-duty road vehicles with combustion engines and electrical hybrid motors. The project used sounds that were synthesized using a synchronous granular synthesis approach, based on measurements of trucks in various driving conditions in a semi-anechoic laboratory, to create a pass-by. A model for the amplification within the street canyon at different façade positions was made, i.e how the sound coming from

the truck is amplified by the characteristics of the street it is passing through. Also made was a model for the façade reduction and room reverberation to get the corresponding indoor sounds, though the reverberation model showed problems with perceived realism. Listening tests were also made of the set-up but with limited validation (Forssén et al. 2013).

The methodology shows potential but currently with limited use and Volvo Group wants to continue the work with the objective to make it generally applicable. The granular synthesis itself also showed potential for real-time synthesis of truck sounds, something that could be used in a driving simulator.

1.2. Purpose

The purpose and goals of this thesis can be divided into two parts. Results from both parts are to be validated through listening tests and comparisons of real measurements

Part 1: Generalized script for indoors and façade sounds

Existing work, done by Jens Forssén and Brett Seward in previous projects mentioned in section 1.1, is to be made accessible for use with any new recordings of truck engine noise. Scripts from mentioned previous projects are to be optimized and generalized for this reason. The chief function of the project is to provide a tool for direct understanding of the effects resulting from truck drivetrain noise in a city street canyon environment.

Work will result in an application consisting of a several scripts, with proper documentation, usable for individuals without previous experience with Matlab. The application should accept measurement inputs and then output sounds and sound pressure levels in both time domain and frequency domain resulting from a simulated truck pass-by for a number of different cases. These cases should include apartment interiors or façades at specified heights, with different wall constructions. Outputs should be possible for simulations of different engine rotation speeds, truck travelling speeds and gears.

The theoretical basis for this thesis is drawn from the work done by researchers at Chalmers University of Technology together with staff at Volvo Group, explained in e.g. (Bergman et al. 2015), (Höstmad et al. 2015) and, in swedish, (Fredriksson 2015). All additions made will also be based on that work.

Part 2: Real time simulation inside a truck cabin

In part 2 a continued validation of granular synthesis, based on previous work made by Jens Forssén and Brett Seward, is to be made as an attempt to improve the currently implemented

sound simulation system in an existing truck driving simulator. Measurements made inside the truck's cabin are used as a base for the simulated sounds. The application should receive parameter inputs in form of torque/load, engine rotation speed, truck rolling speed and gear. Resulting output should be in form of simulated sounds corresponding to a real truck running with the given input conditions.

1.3. Delimitations

Part 1

The final compiled script will accept measurement data made with the same set up as the one described by e.g. Seward (2014). It will only consider drivetrain noise, why rolling noise is not included. It is also worth to note that the drivetrain noise is the main component for trucks at speeds below 50 km/h. Though rolling noise can, in some cases, already be the main component at constant speeds above 40 km/h (Sandberg 2001).

Shepard's method of inverse distance weighting will be used to calculate the contributions of the different microphone positions to the equivalent sound at the receiver position. This method showed the most promising result in the previous work made, so no other method is considered.

The model will use a ray traced generic street canyon with only limited possibilities of variation to its design. The street canyon will be in 2.5 D, meaning that height of the canyon will only be included as an extra propagation distance. No ground reflections are considered explicitly and only second order scattering is taken into account.

Only driving conditions of lower constant speed is considered, as no higher constant speed case measurements from Landsvägsgatan exist that are of good quality and have a match in the semi-anechoic chamber measurements. Acceleration is also not considered.

Part 2

Only rpm, gear and torque is considered when creating the grains. No modification, such as pitch shifting, to grains is made in order to match the driving conditions better. Because of part 1 having higher priority, only a limited validation is made for part 2.

Listening tests

The listening test is made with a limited amount of participants. Both parts are evaluated at the same instance, making test time a limiting factor.

2. Theory

2.1. Digital Signal Processing

In this section some digital signal processing concepts are introduced to give a basic overview before applying it to the work. How each concept is utilized will be explained in the implementation chapter further in the report.

2.1.1. Convolution

There exists a lot of different common mathematical operations; subtraction, addition, integration etc. In digital processing an equally common is convolution. Basically, convolution takes two signals and combines them to create a new signal. It can be used in linear systems to calculate the output signal with the help of the input signal and the impulse response of the system. Eq. 2.1 shows the mathematical expression for convolution. The product is a function depending on time (S. Smith 1999).

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau \tag{2.1}$$

where $x(t)$ is the input signal, $h(t)$ is the impulse response of the system and $y(t)$ is the output signal. For this work, the data of interest is taken from finite digital recordings, which means that the equation for convolution can be written in discrete form instead (Mulgrew et al. 2003). This can be seen in Eq. 2.2:

$$y[n] = \sum_{m=-\infty}^{\infty} x[m]h[n-m] \tag{2.2}$$

2.1.2. Correlation

Cross-correlation is useful for comparing two or more received signals in a measurement with several microphones, e.g. to determine and compensate for delay between different locations. It calculates the similarity between signals x and y with one of them delayed τ seconds, i.e. a function of τ (Rao Yarlagadda 2010). Eq. 2.3 shows the case where y is delayed.

$$R_{xy}(\tau) = \int_{\infty}^{\infty} x(t)y(t+\tau)dt \tag{2.3}$$

It gives the possibility of estimating the delay, τ, of signal y that makes it the most similar to signal x. The maximum attainable value is given from the *auto*correlation of x (the cross-correlation case where $y = x$) with zero delay, i.e. $\tau = 0$ (Rao Yarlagadda 2010). A real-valued signal correlation can also be calculated through convolution with the second signal time-reversed. Correlation however, is, unlike convolution, not commutative.

2.1.3. Digital Filtering

In DSP, digital filtering is an important and common function implemented on a system. The digital filter takes the discrete-time signal to perform mathematical operations in order to change parts of that signal. It can be used to separate signals from noise or from some sort of interference, for example background noise. It can also be used to restore signals that have been distorted, for example by poor measurement equipment.

There are various types of filters and the purpose, depending on the design, is to only let certain frequencies pass and attenuate or cut out the rest. Common examples are the lowpass, highpass and bandpass filters:

- **Lowpass filters:** Attenuates frequencies above the cut off frequency f_c.
- **Highpass filters:** Attenuates frequencies below the cut-off frequency f_c.
- **Bandpass filters:** Attenuates frequencies below the lower cut-off frequency f_{c-low} and above the higher cut-off frequency f_{c-high} (S. Smith 1999).

The frequency response of a filter, $H(\omega)$, can be calculated from the impulse response if that is known or for discrete-time systems, the Z-transform can be used. By taking taking the output transform $B(z)$ and dividing it by the input transform $A(z)$, the transfer function for a linear, time-invariant digital filter can be written in Z-domain according to Eq. 2.4

$$H(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} ..... + b_N z^{-N}}{1 - a_1 z^{-1} - a_2 z^{-2} ..... - a_N z^{-M}} \tag{2.4}$$

where the b_i coefficients characterises the feedforward part and a_i characterises the feedback part. This mathematical expression is the form of a recursive filter, i.e. a filter that re-uses at least one of its outputs as inputs. As long as $a_i \neq 0$ the filter will generate an infinite impulse response, also known as an IIR filter. If $a_u = 0$, the filter will generate a finite impulse response instead, also known as a FIR filter (Mulgrew et al. 2003).

With FIR filters it is possible to make them have zero phase, something that is not in general possible for IIR filters. This means that the filter lets all the frequencies pass through with zero phase shift. In order to achieve a zero phase filter, the frequency response needs to be real, even and fulfil the condition $H(\omega) > 0$ in the passband, meaning the frequencies that are passed through the filter. It is used in this work to preserve the characteristics of the original time waveform when filtered, without causing a delay (J. Smith 2007).

Windowing

A window function, is a mathematical function that will be equal to zero outside a selected interval. This means that when multiplied with another function, the product will also be zero-valued outside the interval, i.e everything will be zero except for where the functions overlap.

The reason for applying a window function on a signal is to focus on specific part instead of analysing the whole signal (Mulgrew et al. 2003). For this work, a Hann window is used in both the auralization and the granular synthesis model. The equation for this window function $w(n)$ can be seen in Eq. 2.5:

$$w(n) = \frac{1}{2}\left(1 - cos\left(\frac{2\pi n}{N-1}\right)\right) \tag{2.5}$$

where n is the current sample in the signal, N is the number of samples under consideration.

2.2. Source modelling and sound propagation

When sound propagates from a source to a receiver it's affected by multiple parameters. When modelling a source, some of these parameters change the sound more drastically, both in levels and in character, while others are more important for the perceived realism, i.e. in order to create a realistic sound. The source itself can be modelled in multiple ways, but in this work multiple sources are used, which means that a method such a set-up needs to be considered. In this section the theory behind the chosen propagation parameters considered and the chosen source modulation method is described.

2.2.1. Shepard's method

Several ways exist to solve interpolation problems for scattered data (Thacker et al. 2009). One way is to use Shepard's method of inverse distance weighting, which calculates a weighted average of the values at the different data points. The weighted average determines how important a specific data point is and is mostly based on the distance between the points. The classical form of this method can be expressed according to Eq. 2.6:

$$F(x,y) = \sum_{i=1}^{n} w_i f_i \tag{2.6}$$

where n is the number of points used for interpolation, f_i are the function values at the data points and w_i is the weight functions that is assigned to each data point. The weight function w can be expressed as:

$$w_i = \frac{h_i^{-p}}{\sum_{j=1}^{n} h_j^{-p}} \tag{2.7}$$

where h_i is the distance between the interpolation point and the different data points. The variable p is the weighting exponent, a positive real number with a standard value of 2. The distance h_i, in 3D, can be expressed according to Eq. 2.8:

$$h_i = \sqrt{(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2} \tag{2.8}$$

where (x,y,z) are the coordinates of the interpolation point and (x_i,y_i,z_i) are the coordinates of the data points. A problem with the classical weight function is that data points far away from the interpolation points are given too much influence (Thacker et al. 2009)(Franke and Nielson 1980). A modification to the classic model was developed by Franke and Nielson (1980), called the local modified Shepard model. The new modified weight function can be expressed according to Eq. 2.9:

$$w_i = \frac{\left[\frac{R-h_i}{Rh_i}\right]^2}{\sum_{j=1}^{n}\left[\frac{R-h_i}{Rh_i}\right]^2} \tag{2.9}$$

where R is the distance between the interpolation point and the most distant data point. This modified model eliminates the deficiencies mentioned above and has shown to give superior results compared to the classical method (Thacker et al. 2009)(Franke and Nielson 1980).

2.2.2. Spherical waves from point source

A sound source, if modelled as a point source, produces spherical pressure waves where the energy of these waves are spread out over the spherical surface area $4\pi r^2$, where r is the radius of the sphere (J. Smith 2010) . This means that the energy per unit area decreases proportional to $1/r^2$ when a wave is expanding. As the energy is also proportional to the amplitude squared, this leads to that the pressure amplitude of a spherical wave, using an inverse square law for energy, is proportional to $1/r$. Using two sets of coordinates $Q_1 = (x_1,y_1,z_1)$ and $Q_2 = (x_2,y_2,z_2)$, where Q_1 is the coordinates of the point source, the pressure amplitude p at a point Q_2 can be calculated according to Eq. 2.10:

$$p(Q_2) = \frac{p_1}{r_{12}} \tag{2.10}$$

where p_1 is the pressure amplitude at one radial unit from the point source and r_{12} is the distance between Q_1 and Q_2 and can be calculated according to Eq. 2.11:

$$r_{12} = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2} \tag{2.11}$$

2.2.3. Doppler effect

The Doppler effect is the apparent change in frequency of a sound wave caused by the relative motion between a source and an observer (Rennie and Law 2015). The frequency sent out by the source is changing because more sound waves are affecting the observer each second when source is travelling towards it. The opposite is happening when the source moves away from the observer, since fewer sound waves are affecting it each second. The perceived frequency can be calculated according to Eq. 2.12:

$$f_{perceived} = \left(\frac{c - u_0}{c - u_s}\right) f \tag{2.12}$$

where f is the true frequency sent out by the source, c is the speed of sound, u_0 is the speed of the observer and u_s is the speed of the source (Rennie and Law 2015). A higher frequency is perceived when the relative motion of the source and observer is such that they are moving towards each other and a lower frequency is perceived when they are moving away from each other.

2.2.4. Air attenuation

There are two mechanisms that contribute to the attenuation of sound in air; classical effects and relaxation effects. The classical effects include friction, heat conduction and molecular diffusion, where friction and heat conduction contributes the most to the absorption. The relaxation effects includes the additional losses caused by absorption of energy in the air molecules, which leads to vibration and rotation of the molecules and can cause them to re-radiate the sound at a later instance, causing interference (Bass et al. 1972).

According to standard, ISO 9613-1:2013, the atmospheric absorption, i.e. the attenuation coefficient α in decibels per metre, can be calculated with the help of Eq. 2.13

$$\begin{aligned}
\alpha &= 8.686 f^2 \left(X + \left(\frac{T}{T_0}\right)^{-5/2} \times (0.01275Y + 0.1068Z) \right) \\
X &= \left[1.84 \times 10^{-11} \left(\frac{p_a}{p_r}\right)^{-1} \left(\frac{T}{T_0}\right)^{1/2} \right] \\
Y &= \left[exp\left(\frac{2239.1}{T}\right) \right] \left[f_{rO} + \left(\frac{f^2}{f_{rO}}\right) \right]^{-1} \\
Z &= \left[exp\left(\frac{-3352.0}{T}\right) \right] \left[f_{rN} + \left(\frac{f^2}{f_{rN}}\right) \right]^{-1}
\end{aligned} \tag{2.13}$$

where $p_r = 101.325kPa$ is the reference atmospheric pressure, p_a is the actual atmospheric pressure, $T_0 = 293.15K$ is the reference temperature, f is the frequency and T is the actual atmospheric temperature. The variables f_{rO} and f_{rN} are relaxation frequencies of oxygen and nitrogen and can be calculated according to Eq. 2.14 and Eq. 2.15 respectively:

$$f_{rO} = \frac{p_a}{p_r} \left(24 + 4.04 \times 10^4 h \frac{0.02 + h}{0.391 + h} \right) \tag{2.14}$$

$$f_{rN} = \frac{p_a}{p_r} \left(\frac{p_a}{p_r}\right)^{-1/2} \times \left(9 + 280h \times exp\left(-4.170 \left[\left(\frac{T}{T_0}\right)^{-1/3} - 1 \right] \right) \right) \tag{2.15}$$

where h is the molar concentration of water vapour in percentage, i.e. the ratio between the partial pressure of water vapour and the atmospheric pressure, with usual values, at mean sea level, that ranges between 0.2% and 2%. It can be calculated according to Eq. 2.16:

$$h = h_r \left(\frac{p_{sat}}{p_r} \right) / \left(\frac{p_a}{p_r} \right) \tag{2.16}$$

where h_r is the relative humidity and p_{sat} is the saturation vapour pressure. The relationship p_{sat} / p_r can be calculated with the help of Eq. 2.17 and Eq. 2.18:

$$\frac{p_{sat}}{p_a} = 10^C \tag{2.17}$$

$$C = -6.8346 \left(\frac{T_{01}}{T} \right)^{1.261} + 4.6151 \tag{2.18}$$

where $T_{01} = 273.16K$ is the triple-point isotherm temperature. This work follows this standard when considering air attenuation.

2.2.5. Reduction index

Effects on sound travelling through objects such as walls or windows are most often summarized to a reduction index, R, a standardized calculation of wave pressure loss in decibel achieved at different frequencies (Kleiner 2012). The exact method of calculating a reduction index is shown in standards such as ISO 717. It is presented with averaged values for whole or 1/3-octave bands or, sometimes, as a single summed up value for the entire range of interest. The measured frequency interval is commonly between the 50 Hz and 4 kHz 1/3-octave bands.

The magnitude of transmission loss through a building element, such as a wall, depends on different material parameters in varying frequency intervals (Kleiner 2012). For the simplest case with a thin single wall panel, there are primarily four intervals and with several layered constructions there are more. The regions are, in order of occurrence starting from low frequency, the stiffness region; the resonance region; the mass law region; and the coincidence region. In e.g. the mass law interval, the main factor to the resulting transmission loss is the mass per unit area of the wall.

In an actual city environment, the radiated noise will travel into surrounding apartments mainly through walls, windows and ventilation (the latter often being the most transmissible). A *composite* reduction index is calculated to take all the elements into account.

2.3. Ray tracing theory

Simulation of a typical city street environment, or city canyon, is possible with use of ray tracing. With this method, portions of spherical waves are modelled as travelling from source to

receiver in the same manner as a beam of light (Long 2005). Complete models include an adequate number of sources and receivers to cover the parts of interest in the geometry. To simulate different travel paths between the source and receiver position, rays from the source are sent out at different angles.

In the case of a truck driving down a city street, a large number of evenly spaced positions along the road would account for the travel path. If the façades are the same in the entire canyon, only one location, per height above ground, is necessary to simulate a listening position by the road or in an apartment. The difference between floors however, necessitate more than one position along the height axis.

2.3.1. Reflections

A series of rays are sent out from the source at varying angles, resulting in varying amounts of reflections off of the environment on the way to the receiver (Long 2005). Reflections in ray tracing are modelled as specular, meaning that the reflected angle is the same as the angle of incidence. Surface effects such as absorption and transmission are bundled together in one coefficient that is factored onto the ray's wave amplitude. An example of the intensity (proportional to squared pressure) of a ray after being reflected and travelling the total distance r is seen in Eq. 2.19.

$$I_{spec} = \frac{WQ_{\phi}}{4\pi r^2}(1-\alpha_{abs})e^{-\alpha r} \tag{2.19}$$

Where W is the sound power of the source in watt, Q_{ϕ} is the source directivity at angle ϕ, α_{abs} is the wall absorption coefficient, α is the air absorption coefficient (see Eq. 2.13) and r the total propagated distance (Long 2005).

2.3.2. Diffuse reflections

Environments consist to a large degree of uneven surfaces that send incident sound waves in several directions. When such a diffuse reflection occurs, the energy of the incident ray is assumed to be in part absorbed (and transmitted), specularly reflected and scattered (Long 2005). In a ray tracing model, the scattering effect is preferably included through a single number, frequency dependant, *scattering* coefficient (Cox and D'Antonio 2009). That is, a quantitative measure of how much of the total reflected energy is distributed non-specularly, applied in the same manner as the specular reflection coefficient. For a specular reflection off a diffusive area it becomes an additional factor $(1-\delta)$, as seen in Eq. 2.20.

$$I_{spec} = \frac{WQ_{\Phi}}{4\pi r^2}(1-\alpha_{abs})(1-\delta)e^{-\alpha r} \tag{2.20}$$

The scattering coefficient can be measured with two different methods, one random incidence and one free field (Cox and D'Antonio 2009). The method for determining the random incidence

coefficient is described in ISO 17497-1 and is simpler than the free field alternative. A table of random incidence coefficients for different irregular diffusive shapes can be found in work by Vorländer (2008). The effective range, where the scattering coefficient is high, depends on two dimensional properties of the diffusive element, average structural depth (h) and length (a). The elements are effective when the incident frequency fulfils the two following criteria (where c is the speed of sound in air):

$$f \approx c/2a, f \approx c/2h \tag{2.21}$$

That is, the elements scatter efficiently when half of the incident wavelength is of comparable dimension to the two mentioned structural dimensions. When the wavelength is much longer, the structural irregularities are too small and the wall can be seen as flat. In the opposite case, with a much smaller wavelength, each irregularity can be seen as a large flat object itself. In both of these cases, the incident wave is almost exclusively specularly reflected. Vorländer (2008) also describes how the coefficient varies with the ratio of average structural length to incident wavelength. As an example, below the ratio $a/\lambda = 0.125$ (where λ is wavelength in meters) the random incidence coefficient is generally smaller than 0.05.

In a scale model investigation done by Ismail and Oldham (2004) the scattering coefficient of building façades with typical surface irregularities was found to be in the $0.09 - 0.13$ range. The authors also concluded that there is a tendency for coefficient increase with increasing surface irregularity.

2.3.3. Lambert's emission law

The directionality of the scattering from the surface is described by the diffusion coefficient. In the ideal diffusion case it is Lambertian, a concept introduced in the coming paragraph. Another case is a surface with uniform distribution, which is the defined case for a diffusion coefficient of 1 (Vorländer 2008). The part of the energy that is scattered can be simulated with new rays by modelling the reflecting surface as a new source (Long 2005).

Viewing sound waves as rays leads to the use of optical laws when describing some of their propagation behaviour. Lambert's emission law, which explains the relation between diffusively radiated intensity and its outgoing angle, is such an example. Surfaces that follow this law are called Lambertian, meaning that their radiated intensity in any direction corresponds to the visible surface size from the perspective of the outgoing angle (W. J. Smith 2000).

A representation of Lambert's emission law is shown in Figure 2.1.

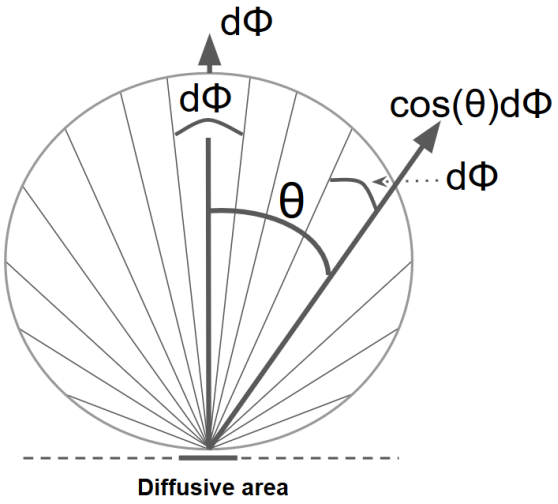


Figure 2.1.: Illustration of reflectance from a Lambertian diffusive wall area.

All directions in Figure 2.1 have the same angle portion $d\Phi$. The reflected intensity in each direction corresponds to the area of each portion of the circle. The cosine of the angle θ between the normal of the surface and the outgoing direction corresponds to the decrease in intensity compared to the perpendicular. This in turn relates to the size of the diffuser surface normal to the radiated angle, i.e. the apparent surface size from each outgoing direction.

2.4. Granular synthesis by pitch-synchronous overlap-add (PSOLA)

Sound synthesis can be done by extracting small, 1-100 ms long, segments out of recordings and then using the selections to reconstruct a new sound (Zölzer 2011). This type of method is referred to as *granulation*. Putting the segments in the order they were extracted makes it possible to construct a replica of the original signal. More importantly, using grains gives the possibility to modify sound, e.g. its duration (time stretching), by simple time-domain changes.

The pitch-synchronous overlap-add (PSOLA) framework utilizes the granulation process. PSOLA was originally developed for modification of speech without loss in perceived realism, but can be implemented with other types of sources, such as engines (Moulines and Charpentier 1990)(Jagla et al. 2012). There are different variants of the method and two of them are the time-domain and frequency-domain variants (TD-PSOLA and FD-PSOLA). The choice of method depends on if/how the grains are to be modified. Simplified, TD-PSOLA is easier to implement and less computationally demanding while FD-PSOLA allows for larger modifications, especially pitch shifting, with maintained sound quality (Moulines and Charpentier 1990).
The PSOLA process, as described by Moulines and Charpentier (1990), can be divided into three steps: analysis, modification (optional) and synthesis. Overall, PSOLA methods place the

bulk of the computation in the first two steps, which improves the synthesis performance and facilitates real-time implementations (Jagla et al. 2012). In the subsections below are general descriptions of the steps in the TD-PSOLA process. Grain modification however is not implemented in this thesis and is therefore left out.

2.4.1. Grain analysis

The analysis entails the creation of the grains, which starts with determining where (in time) the grain center points, called *pitch marks*, are to be placed (Zölzer 2011) (Moulines and Charpentier 1990). It is from these points the grains are extracted and the PSOLA method gives the possibility to use non-uniformly placed marks. That is, the placement is allowed to vary with the change of pitch.

For placing the markers a fundamental frequency called pitch period, that describes periodicity in the signal extracted from, needs to be established (Zölzer 2011). This is often one of the main issues of grain creation. In the case of an engine, the source signal is highly periodic as it consists of a number of cylinders firing off in a set order. When measuring an engine installed in a vehicle however, there are a lot of other noises that contribute (Jagla et al. 2012). There is often prominent low frequency noise that can ruin the "signal-to-noise" ratio (SNR) of the engine sound, making pitch marking more difficult.

Placement is commonly based on a "peak-search" approach, with an algorithm that finds global maxima of pressure/energy in the signal and places the first pitch mark there (Zölzer 2011). Thereafter, the algorithm searches for local maxima in each direction of the first mark, taking steps based on the pitch period. If the SNR is too low, there is a risk that the desired event, e.g. a cylinder firing, gets drowned out and that the pitch marker is placed erroneously.

Having set out pitch markers for each part of interest; a Hanning window is used to extract segments of length corresponding to e.g. two pitch periods, centred on the marker (Zölzer 2011). This length gives half a period on either side of the segment, two tails, that are used in the OLA-process when merging with other grains.

2.4.2. Grain synthesis

Once a database of grains has been built and possible modifications have been made, the synthesised signal can be constructed. The process consists of two steps: selection and addition. The addition process described here is called overlap-add, OLA for short.

Selection is the choice of whatever grain fits with the desired signal at any current time sample. This can become a complicated task and entail a large set of conditions to ensure a good match with previously selected grains.

To add two grains together seamlessly, a 128 sample long Hanning window is applied to each before the merge. The window is includes 64 samples from the core and 64 samples of the corresponding tail (left tail for the new core and the right tail for the previous core) as shown in Figure 2.2. The 128 sample intervals from both grains are then added together on the same 128 samples, creating a seamless continuation of the signal. Note that the sample amount used, 128, is arbitrary, but must not be too small or large to avoid the merge becoming audibly distinguishable.

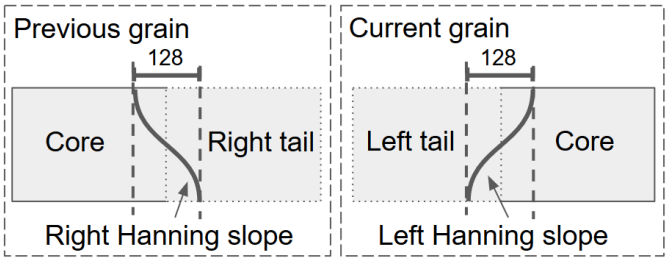


Figure 2.2.: Illustration of the overlap part of the OLA process.

3. Part 1 - External truck sound

The first part of this work is to simulate a truck pass-by in a street canyon by creating a computer model. The computer model for the truck pass-by is generalized to work with multiple truck models, different types of street canyon dimensions and various types of velocities. The outputs are plots of the sound pressure levels at chosen listening positions; both outdoors and indoors, at different floor levels and at both sides of the street. The sound pressure levels are calculated both in time-domain with the maximum value and the corresponding 1/3-octave band sound pressure levels at the maximum pressure time instance. The plots are both A-weighted and un-weighted. The possibility of choosing a fixed single point in the street canyon is also possible, for example right in front of the apartments using a recording of an idling truck. Also generated are corresponding sound files at the chosen listening positions.

The computer model, made in Matlab, consists of a core program, a set of function files and a user interface (GUI). An overview of how the program is set-up can be seen in the figure in Appendix A. The different function files contain subscripts not seen in the figure as they only exist to divide the function files for easier navigation. The GUI, seen in Figure 3.1 is created for easier user input.

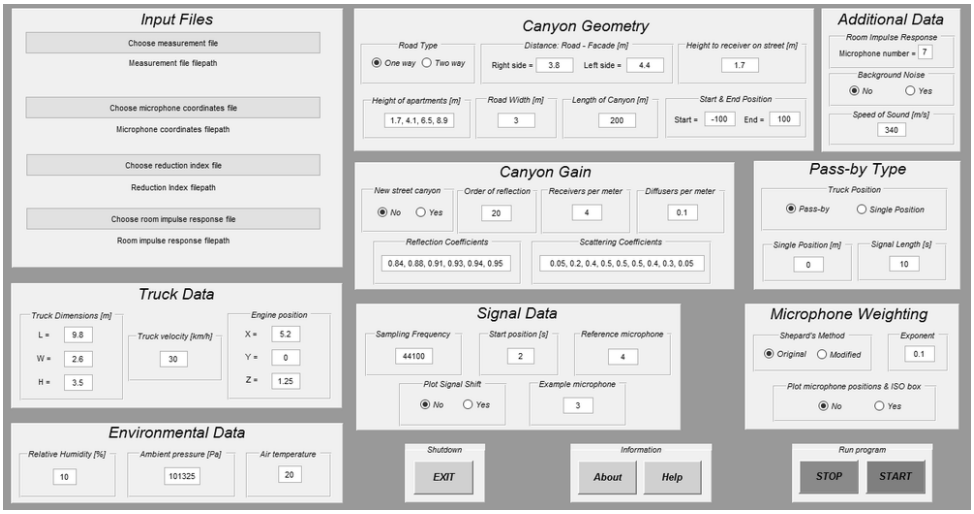


Figure 3.1.: MATLAB GUI

It contains all the data that is necessary to user to input in order for the model to work. Also included is a window to show the calculation status of the program and a user manual for easier use. The program is put together to stand-alone executables in order to make it user friendly and eliminate the necessity of having MATLAB. The sections below describe how the computer model is set up and what input data has been used in order to validate the model and find its limitations.

3.1. Street input data

For this work, the street canyon used is modelled after Landsvägsgatan. This was done in order to make the validation process of the computer model easy, as measurements from that street already existed. Four apartment floors are considered, using a height of 2.4 m for each floor and a listening position height of 1.7 m, i.e $[1.7, 4.1, 6.5, 8.9]\,m$. The height 2.4 m is a standard height in a Swedish apartment and the listening position is assumed to be the height of typical person. The canyon is modelled as $200\,m$ long with the apartments located in the middle of the canyon. The start and end position is located almost at the ends of the street canyon (80 m in both directions), but the option exist to change both positions. The canyon width is equal to $d_{canyon} = 11.2$ m with a $d_{road} = 3$ m road width (one way) and a distance of $d_{right} = 4.6$ m and $d_{left} = 3.8$ m to the right and left side façade respectively. Three receiver locations on both sides are considered; on the street, on the façade and inside the apartments. The street position is assumed to located at $d_{road}/2 + d_{right}/2$ for the right side and $d_{road}/2 + d_{left}/2$ for the left side and at a height of 1.7 m.

For the pass-by, the length of the generated signal is based on the start and end position of the truck, together with the velocity and sample frequency:

$$Signal\ length = \frac{x_{start} - x_{end}}{v/3.6} f_s \tag{3.1}$$

where x_{start} is the start position of the truck, x_{end} is the end position of the truck, v is the velocity in km/h and f_s is the sampling frequency. For the single position the signal length is chosen to $10s$, but any arbitrary value can be chosen.

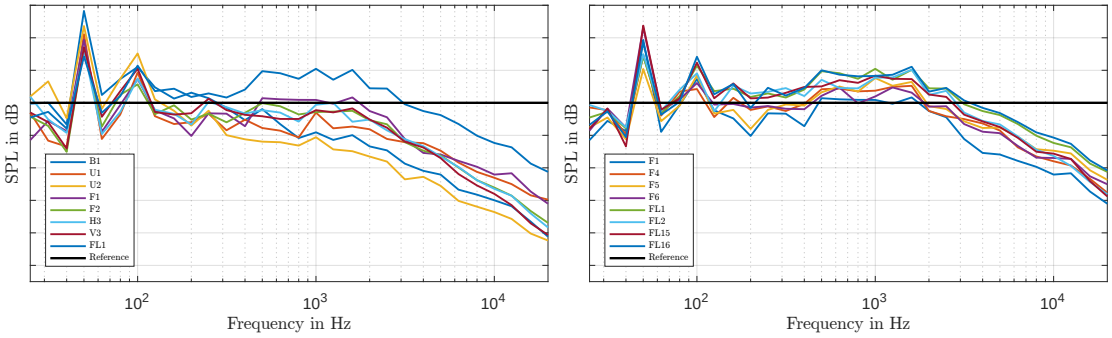
3.2. Input measurement

The measurement data was provided by the Volvo Group in Gothenburg, Sweden. The data, measured in Volvo Group's semi-anechoic laboratory, was previously used in the projects mentioned in Section 1. Though the data used in this section is taken from the truck "FM-499" with the dimensions (length × width × height): $9.8 \times 2.6 \times 3.5$, the model is generalized to work with any type of truck.

Microphones used to record the sound pressure data were of the type Brüel & Kjaer Type 4189, 1/2-inch free field microphones, with a sampling frequency of 44.1 kHz. The previous measurements included a total of 54 microphones; 22 were distributed, in a box shape, around the truck according to ISO 3744, 16 were placed at a drive-by distance of 7.5 m to the left and right side of the truck and 16 were placed on the floor around the truck. For the MATLAB model, only the 22 microphones in a box shape were considered in the beginning when simulating the sound source. A problem with this was the significant difference in frequency content between

the microphones and the lack of high frequency content for most of the microphones located in the top and back of the box. As each microphone is contributing to the source simulation (see Section 3.3 for the chosen method), the decision to add the floor microphones was made. This was done in order to lower the influence of the low frequency content microphones, i.e. increase the high frequency content that most of the floor microphones recorded which the top and back microphones did not, to get a higher better average.

The sound pressure level using a recording with constant speed of 30 km/h (1050 rpm and gear 9) for some of the microphones can be seen in Figure 3.2a. All microphones are distance compensated with the engine as the assumed point of origin. Each horizontal line is an increase of 10 dB and the black reference line exist in order to be able to compare with Figure 3.2b that plots a separate case with only front microphones being used instead of all microphones.



(a) All microphones used

(b) Selected microphones used

Figure 3.2.: Comparison of the sound pressure levels at different locations in the semi-anechoic chamber. Increments of 10 dB per horizontal line. Microphone names: B = Back, U = Upper, F = Front, H = Right, V = Left, FL = Floor.

As mentioned above and also seen in Figure 3.2a, the frequency content recorded differs a lot between the microphones with up to around 20 dB in frequencies above ~500 Hz. Since the sound pressure levels differ so much between the microphones, a separate case is also evaluated where only the low front microphones are included. Figure 3.2b shows the microphones used for this case. It can be seen that the frequency content is closer between the microphones.

The name of each microphone refers to a position in the measurement set-up and can be seen in Figure 3.3 below. It shows the set-up of the microphones of the measurement and the position of the engine as the assumed point of origin of the sound, at the coordinates $[x, y, z] = [5.2, 0, 1.25]$ in the ISO-box. All the microphone coordinates can be seen in Appendix 1.

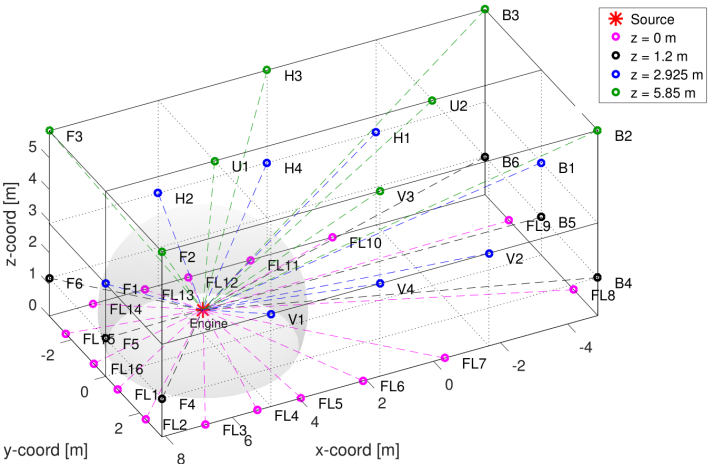


Figure 3.3.: Measurement set-up using ISO-Box microphones and floor microphones with engine as assumed source of origin. Microphone names: B = Back, U = Upper, F = Front, H = Right, V = Left, FL = Floor.

The microphones are mounted at four different heights: 0m, 1.2m, 2.925 and 5.85m and the six letters F,B,FL,H,V,U stands for Front, Back, Floor, Right, Left and Upper. The truck is located inside this box and where the half sphere is used for microphone weighting which is explained in Section 3.3.2 with the engine in the center of this half sphere. As mentioned, the MATLAB model is based on the previous measurements made according to ISO 3744 with the rectangular box and the addition of the floor microphones. This means that as long as new measurements on other trucks are made according to this measurement procedure the model will be able to calculate new outputs, independent of the number of microphones used. Figure 3.4 shows the set-up of the selected microphones:

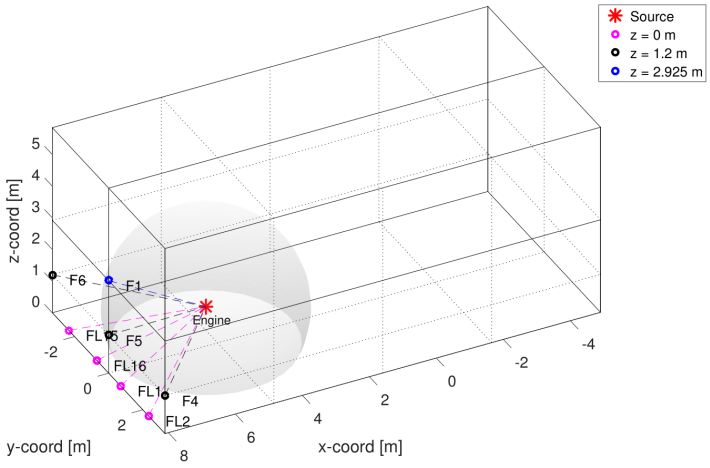


Figure 3.4.: Measurement setup - selected microphones. Microphone names: F = Front, FL = Floor

3.3. Source simulation

To simulate the truck as moving through the city canyon it is modelled as a single sound source, i.e. a point source, where each microphone contributes to the characteristics of the sound. The signals of the microphones need to be adjusted for time delay caused by extra travelling distance and distance between source and receiver. This is done by using cross-correlation and implementing Shepard's weighting.

3.3.1. Microphone Delay

Each microphone is placed at a certain distance from the truck which leads to a time delay of each signal. In order to compensate for this a reference microphone is chosen. Microphone *F5* located in front of truck is used as a initial choice because of the convenient location in front of the truck, but the option to change it exists as any microphone can be used. All the microphone signals are then shifted with regards to the reference microphone in a negative or positive direction. This is done by looking at a number of samples of the two signals and then using cross-correlation. The number of samples is proportionate to the extra distance the sound has to travel to get to the microphone located furthest away compared to the reference microphone An example of the effect of cross-correlation between two microphones can be seen in Figure 3.5.

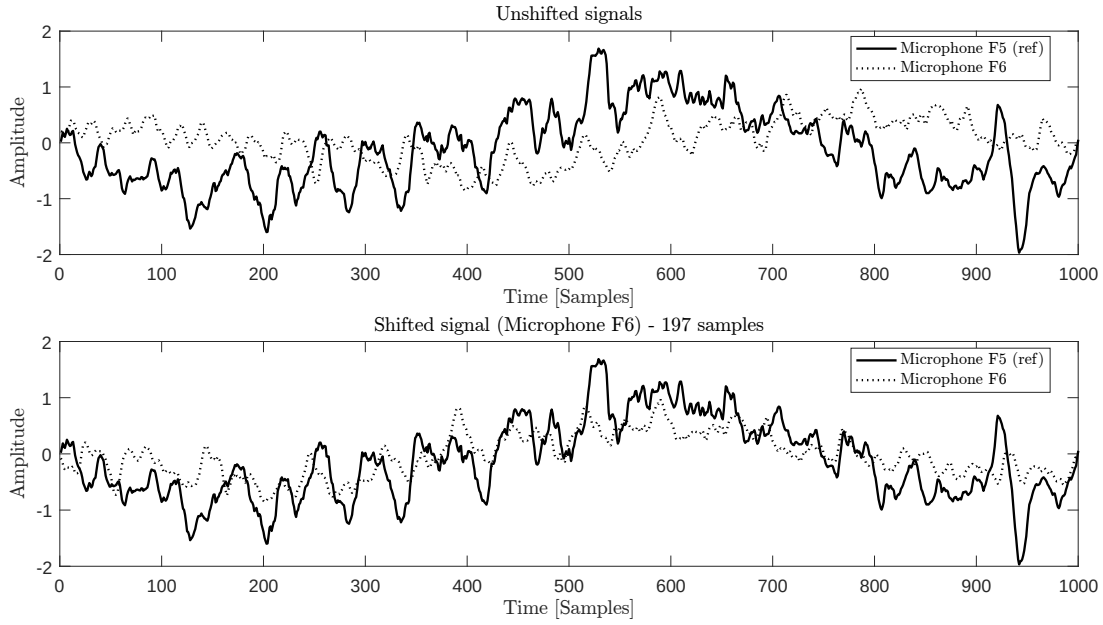


Figure 3.5.: Comparison of delay in samples between two microphone positions using cross-correlation

As can be seen in Figure 3.5, microphone F5 and F6 are quite close to each other since there is only a delay of 250 samples or 5.7 ms using the sampling frequency $f_s = 44.1$ kHz. As stated in Section 2.1.2, cross-correlation measures the similarity between two signals and since microphone F5 and F6 are quite close to each other the time delay is quite small compared to microphones located further away. This is due to both the distance and the fact that the character of the noise is different further away from the chosen reference position, for example because of contributions from the exhaust outlet or because of the trailer blocking parts of the sound.

3.3.2. Shepard's Weighting

The noise generated by the truck passing through the street canyon is modelled as a point source, where each microphone is contributing depending on the distance between the microphones and the receivers. Each microphone will be assigned a weighting value according to the original Shepard's method (see Section 2.2.1). The weighting values are changing depending on the position of the vehicle in the street canyon.

First, the microphones with their respective coordinates are projected onto a sphere with a radius r that is based on the volume of the chosen truck. The reason for doing this is to easily be able to calculate the distance between source and receiver without being forced to consider the real distance between them in the street canyon. The sphere projection is done by turning the Cartesian coordinates into spherical coordinates to get the azimuthal and the elevation angle, see Figure 3.6, and then turning the spherical coordinates back into Cartesian using the radius r to get the position of each microphone when projected on to the sphere.

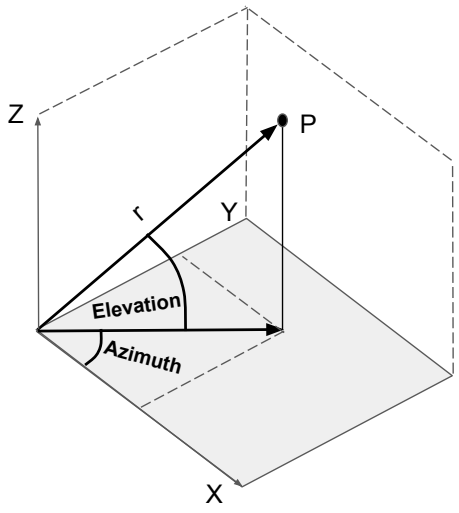


Figure 3.6.: Conversion of Cartesian to spherical coordinates

Figure 3.7 shows the microphones projected onto the sphere that could be seen in Figure 3.3 in Section 3.3 with the engine as the assumed point of origin of the generated sound and the radius r.

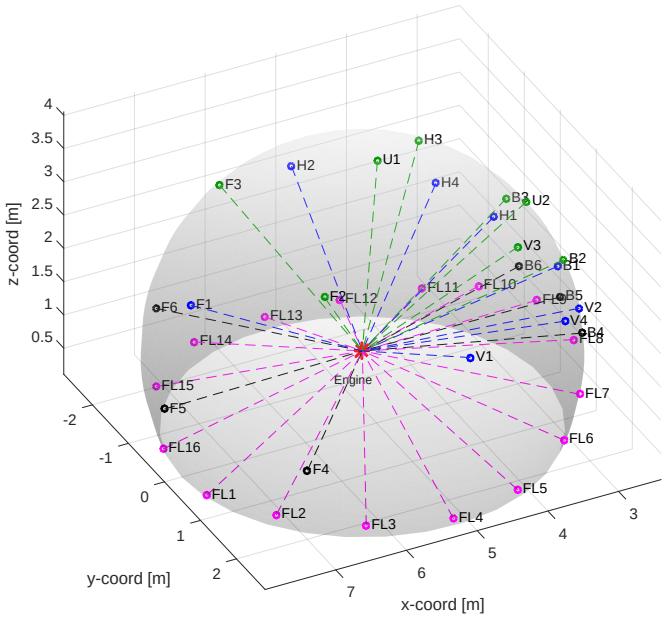


Figure 3.7.: Microphones projected onto the sphere with a radius based on the volume of the truck.

The receiver position is also converted into spherical coordinates and projected onto the sphere. This is done by first calculating the elevation angle between the source, i.e. engine, and the chosen apartment together with the time-varying azimuthal angle. The azimuthal angle (angle in x-y plane) between source and receiver is first generated with 4 points per degree between $0 - 180°$, i.e simulating the position of the vehicle going from $-\infty$ to ∞. The exact angle between the source and receiver is interpolated at a later instance. Using the calculated angles and the radius r, the spherical coordinates of the receiver are turned into Cartesian coordinates to be projected onto the sphere.

By using the position of each microphone on the sphere together with the position of the receiver, the angle between these two points can be calculated, which is done by taking the dot product. The angle is multiplied with the radius of the sphere to get the distance in between. This distance is used in Eq. 2.6 in Section 2.2.1 to get the different weighting values depending on microphone and receiver position. The weighting exponent p used in Eq. 2.7 was tested with different values between $0.1 - 2$. A higher value means that the microphones closer to the receiver has higher weighting values, i.e. influence the signal more, while a lower value means a more even contribution distribution between the microphones. Compared to a real case scenario, a higher value of p might be more accurate as the receiver is closer to certain parts of the vehicle

at a certain instance in time. But as could be seen in Figure 3.2a, the frequency content between the microphones differed quite much. This causes unnatural shifting of the sound when using a high value of p, i.e. when the most prominent microphone suddenly changes so does the character of the sound. The weighting exponent p was chosen to a value of 0.1 in order to prevent this from occurring. The modified Shepard's method was also tested but did not give better results for this model as it valued the closer microphones even more than the original method. Also the microphone furthest from the receiver wasn't included at all. Since the sound travels in all directions in the street canyon, all microphones should be included to some extent in order to simulate the truck more correctly.

The real position of the vehicle is calculated using the velocity and the length of the canyon together with the sampling frequency, as each sample in the signal represents a position on the road. The azimuth angle between the source and the receiver on the street or façade, on both sides of the street, is then calculated and used together with the pre-calculated angles to interpolate the real weighted values of each microphone depending on truck position. The signal of each microphone is multiplied with the weighted value and the distance between microphone and engine and then divided by $r_0 = 1$ (spherical spreading) to get the equivalent pressure at 1 meter distance from the engine. The signals are then summed up to create a single sound source, where each signal is contributing depending on the distance and weighting value assigned to it.

3.4. Sound propagation

When the source, i.e. the truck, is moving through the canyon, different parameters are affecting the sound propagation. Since it is a computer model not everything can be considered and some parameters need to be simplified in order for the model to work. For example, rolling noise is not considered in the model and in order to get indoor sound pressure levels, a measurement of a room impulse impulse response and estimated reduction indices from Landsvägsgatan are used. In the sections below, the implementation of how the parameters are considered is described. The influence of the street canyon, i.e. canyon gain, is not mentioned here, but is instead brought up in Section 3.7.

3.4.1. Distance compensation

As mentioned the truck is modelled as a point source that produces spherical waves, i.e. spherical spreading. According to the theory in Section 2.2.2 the pressure amplitude at the receiver position is proportional to $1/r$, where r is the distance between the engine and the receiver position located on the street or on the façade for the different apartment floors. For the façade position, since the receiver is modelled as mounted onto the façade, a doubling in pressure amplitude (+3dB) is applied in order to compensate for first order reflection.

3.4.2. Doppler effect

Doppler effect is mainly considered to increase the realism of the sounds and not so much for the sound pressure levels. The recorded signal is in time domain and not frequency domain. So in order to consider the Doppler effect, an initial time delay is first calculated by taking the distance from the start position to the receiver and dividing it by the speed of sound in air, c_0. The time vector t corresponding to each position $x(t)$ in the canyon is then re-sampled according to Eq. 3.2:

$$t_{resampled} = t - \frac{r(t)}{c_0} + t_{delay} \tag{3.2}$$

where $r(t)$ is the distance between source and receiver at the given time t and t_{delay} is the initial time delay. The new time vector $t_{resampled}$ is then used to interpolate new pressure values of signal at the position $x(t_{resampled})$. The new pressure values are the perceived values at the position $x(t)$.

3.4.3. Air attenuation

Applying air attenuation to the signal is done according to the standard ISO 913-1:2013 mentioned in theory, Section 2.2.4. Standard values are chosen: air temperature $T = 20°C$, relative humidity $h_r = 10\%$ and the ambient pressure of $1\,atm$, with the possibility of changing the values depending on the atmospheric conditions. The air attenuation coefficient in dB/m is calculated for each 1/3-octave band center frequency. Table 3.1 shows the air attenuation coefficient for the 1/3-octave band center frequencies used for calculation.

Table 3.1.: Air attenuation coefficients (dB/m) at different 1/3-octave band frequencies.

25 Hz	**31.5 Hz**	**40 Hz**	**50 Hz**	**63 Hz**	**80 Hz**	**100 Hz**	**125 Hz**	**160 Hz**	**200 Hz**
8.53e-05	1.29e-04	1.91e-04	2.69e-04	3.69e-4	4.91e-04	6.22e-04	7.71e-04	9.74e-04	1.22e-03
250 Hz	**315 Hz**	**400 Hz**	**500 Hz**	**630 Hz**	**800 Hz**	**1 kHz**	**1.25 kHz**	**1.6 kHz**	**2 kHz**
1.57e-03	2.11e-03	2.97e-03	4.23e-03	6.24e-03	9.48e-03	1.41e-02	2.09e-02	3.18e-02	4.55e-02
2.5 kHz	**3.15 kHz**	**4 kHz**	**5 kHz**	**6.3 kHz**	**8 kHz**	**10 kHz**	**12.5 kHz**	**16 kHz**	**20 kHz**
6.31e-02	8.5e-02	0.11	0.133	0.155	0.176	0.193	0.21	0.234	0.259

As seen in Table 3.1, the air attenuation has significant impact on the higher frequencies while it has almost no impact on the lower frequencies. For example, the reduction in the frequencies below 160 Hz is around 0.01 – 0.1 dB per 100 m distance while above 4 kHz it's around 10 – 26 dB per 100 m distance. Therefore considering air attenuation is of significant importance, especially when simulating a pass-by and positions far away from the receiver.

The signal is filtered into 1/3-octave bands, the air attenuation is applied as a scaling value and the new signal is calculated according to Eq. 3.3:

$$y(t) = \sum y_{thirds}(t) 10^{-\alpha r(t)/20} \tag{3.3}$$

where y_{thirds} is the signal in time for each 1/3-octave band frequency, α is the air attenuation coefficient and $r(t)$ is the distance, in time, between source and receiver.

3.4.4. Façade reduction & Room impulse response

To get the equivalent sound inside the apartments at the different floors, two parameters are considered; façade reduction and the impulse response of the room. The façade reduction indices are taken from previous work by *Simmons akustik & utveckling AB* (Forssén 2014), where five types of façade cases are implemented in the model. Case 1 and 2 use a steel frame wall (i.e. a lightweight wall) with two types of windows, where the choice of window has significant impact above 100 Hz. Case 3 and 4 use a concrete wall and with the two types of windows. Case 5 is an estimation of the reduction index for Landsvägsgatan and is used in the model to compare with the measurements performed there. The reason behind implementing the first four cases is to make the model more general by seeing the importance of different kinds of façade types. The reduction indices for each case can be seen in Figure 3.8:

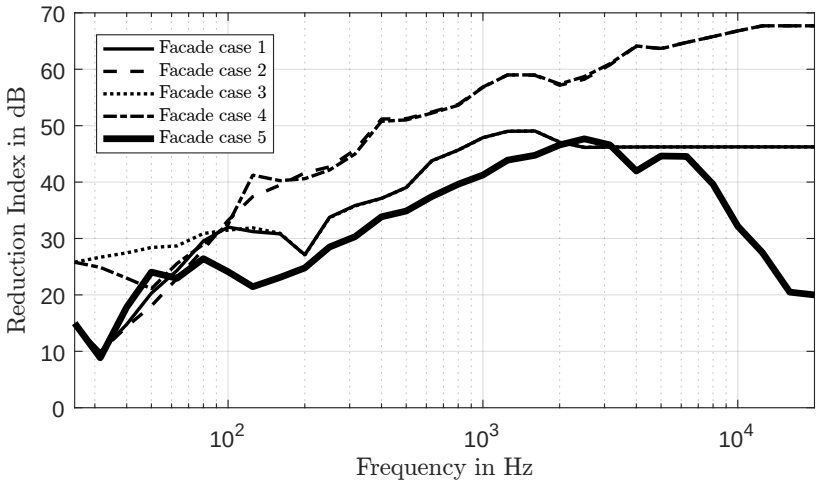


Figure 3.8.: Reduction indexes for five types of façade cases

The impulse response of a standard room at Landsvägsgatan is also taken in to account in order to add the character to the sound. The data is taken from a previous measurement consisting of seven microphones distributed at different positions inside the room. Only the microphone closest to the source of the measurement is used when applying the room impulse response, but any microphone position or measurement can be inserted into the model.

To apply the change of the signal caused by the properties of the room and the reduction indices, the signal is first convolved with the impulse response. The new signal, containing the characteristics of the room is then again converted into 1/3-octave bands to apply the reduction values of the five façade cases. The implementation of the façade reduction is done in the same way as air attenuation, meaning that it is applied as a scaling factor (see Section 3.4.3). An addition of $+3$ dB is added to the total reduction index in order to compensate for that the signal, before implementing the changes, is mounted on the façade.

3.4.5. Background noise

In order to get more accurate sound pressure levels and be able to compare with the measurements the background noise both outdoors and indoors are also considered. The data is also taken from the measurement previously made at Landsvägsgatan and the microphones used for that measurement can be seen in Figure 3.9.

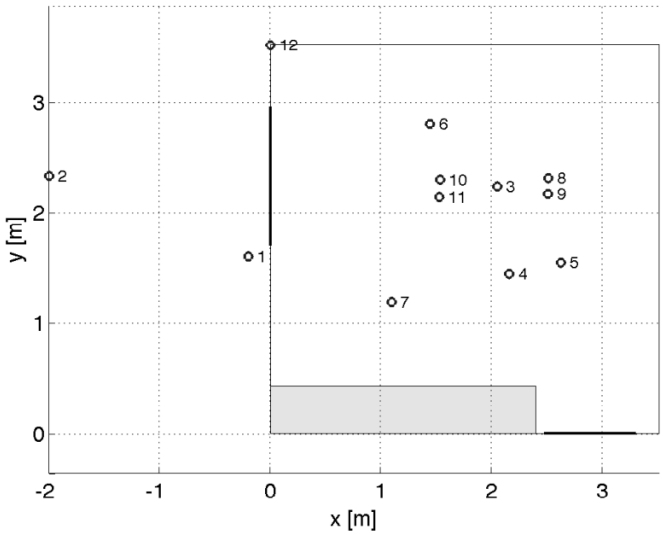


Figure 3.9.: Microphone setup of the measurement at Landsvägsgatan. Microphone 1 and 3 used for background noise recording.

For the outdoor background noise the microphone mounted on the façade, i.e. microphone 1, is chosen and for the background noise inside the apartment the microphone located in the middle of the room, i.e. microphone 3, is considered. Microphone 2 might be a better choice for the receiver located on the street, but since the comparison is focused on the façade positions, that microphone was chosen. A comparison between the two microphones showed a slight difference up to 2 - 3 dB in some frequencies caused by the microphone being mounted on the façade, but overall had matching levels. This means that the street position will be calculated with this error in mind.

The data is converted into 1/3-octave bands and an average for each band is taken. This is done in order to eliminate prominent peaks in the signal, i.e. people talking, birds, random impact noise or any other noise that stand out from the normal background noise. The implementation is made by converting the influence of the background noise to an amplitude correction factor. This factor depends on both the current signal and the background noise. The correction factor is in 1/3-octave bands but changes with time since it is also affected by the signal. Though the contribution of the background noise to the amplitude factor is always the same, as only averaged values are used, i.e. the background noise affects the signal equally at all time instances. Eq. 3.4 shows how the amplitude correction factor is calculated:

$$\begin{aligned} Lp_{tot} &= 10 \cdot log\left(\frac{p_1^2 + p_2^2}{p_0^2}\right) = 10 \cdot log\left(\frac{(k \cdot p_1)^2}{p_0^2}\right) \\ k &= \sqrt{1 + \frac{p_2^2}{p_1^2}} \end{aligned} \tag{3.4}$$

where p_2 is the pressure of the background noise, p_1 is the pressure of the signal and p_0 is the reference pressure. The average background noise levels used for the outdoor and indoor case can be seen in Figure 3.10.

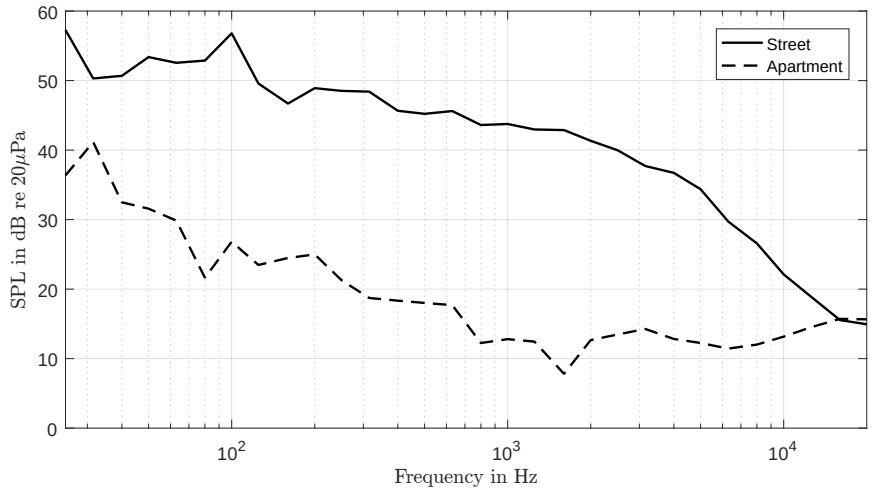


Figure 3.10.: Background noise sound pressure levels for street and apartment position

Since an average is used, no variation of the background is considered, meaning that while the total sound pressure levels will be closer to reality, the character of sound might still not be. Though the background noise is primarily considered in the model for better comparison to the sound pressure levels in the measurements from Landsvägsgatan.

3.5. Sound generation

After applying all the different parameters to the signals at each receiver position, the signals are converted into *.wav* files using the inbuilt MATLAB command *audiowrite* with a sample size of 24 bits. The pressure amplitudes of the signals are in some cases exceeding the maximum amplitude of 1 that is allowed for *.wav* files and therefore need to be scaled down in order to avoid clipping in the signal. A scaling factor is calculated which depends on the absolute maximum amplitude of the signal and then rounded upwards to nearest 10, i.e. 10, 20, 30 etc. If no scaling is necessary, the scaling factor is equal to 1. For the sound files used in this work, the maximum amplitudes are either below 1 or between 1 and 10, so the scaling factor assigned to each signal is either 1 or 10. The rounding upwards to nearest 10 is done in order to make it easier when listening to each sound file, as the user only have to scale each sound up with a factor of 10. Each signal is divided with the corresponding scaling factor before converted in to *.wav* sound files.

In order to avoid cut off in the ends of the signal a window function is used. The function, in this case a Hann window can be seen in Eq. 3.4 and is also mentioned in Section 2.1.3.

$$w = 0.5\left(1 - cos\left(2\pi\frac{n}{N}\right)\right) \tag{3.5}$$

where n is the current sample and N is the total number of samples the function should be applied on. In this case N corresponds to 400 ms and the Hann window is applied on both ends of the signal.

3.6. Sound pressure levels

For each chosen position, i.e. street, façade and indoors for all floors, the sound pressure level is calculated and plotted. The sound pressure levels are first plotted in time-domain, both un-weighted and A-weighted by applying a A-weighting filter. This is done by using a 'Fast' time weighting (0.125 s) according to standard IEC 61672-1. For each chosen position, a maximum value is also calculated. The corresponding 1/3-octave band sound pressure levels at that instant in time (maximum pressure instance) is also plotted; both un-weighted and A-weighted.

3.7. Ray tracing

To approximate the amplification on sound due to an urban environment, a generic effect is calculated from a ray traced impulse response. A basic city canyon geometry is a straight road with continuous multi-story buildings on both sides. The modelled canyon is based on the layout of the street *Landsvägsgatan* (shown in Figure 3.11) in Göteborg, Sweden. However, to get the generic effect, results from five street widths are averaged, reducing the impact of resonances due to wavelengths matching with geometry and/or diffuser positions. The width scaling factors used for the five iterations are shown in Table 3.2.

Table 3.2.: Chosen width scaling factors for the ray tracing model averaging.

Width factor	$\frac{1}{\sqrt{3}}$	$\frac{1}{\sqrt{2}}$	1	$\frac{e^1}{2}$	$\frac{\pi}{2}$

Genericism also means that smaller details, such as parked cars or vegetation, can be ignored. This limits the influential parameters in the model to the properties of the building façades.

Figure 3.11.: Street view of Landsvägsgatan in Göteborg, the basis for the ray tracing simulation. Photo courtesy of Google.

The model is two-and-a-half-dimensional (2.5 D) in the sense that it treats height only as an additional ray propagation distance and does not simulate additional rays due to different elevation angles. Building walls are therefore indirectly modelled as infinitely high. Likewise, reflections are only spread along the azimuth- and not the elevation angle. The road surface is also ignored since the measurements that are used as input into the simulation are performed in a semi-anechoic environment with hard ground. This carries the disadvantage of removing the possibility to include reflections interacting between the ground and walls.

An array of receivers are placed along the truck path to give the different impulse responses along the road. Only positions on one side of the street are calculated for. The difference between the two sides is assumed to be small, since the main difference is a short extra propagation distance. Geometrical symmetry is utilized to simulate only half the road length in the desired canyon. In the case of a 200 meter long canyon, only positions on half the road length (100 meters) are actually calculated. This carries the advantage of directly cutting the ray tracing calculation time in half. Nothing before or beyond the calculated 200 meter canyon is included, effectively making the ends free field boundaries.

Interpolation of the pressure amplitude in each third-octave band is used for positions in between the calculated ones. Getting the impulse response for the other half of the road length requires reversing the order of the calculated positions. A top-down sketch of the designed geometry is shown in Figure 3.12.

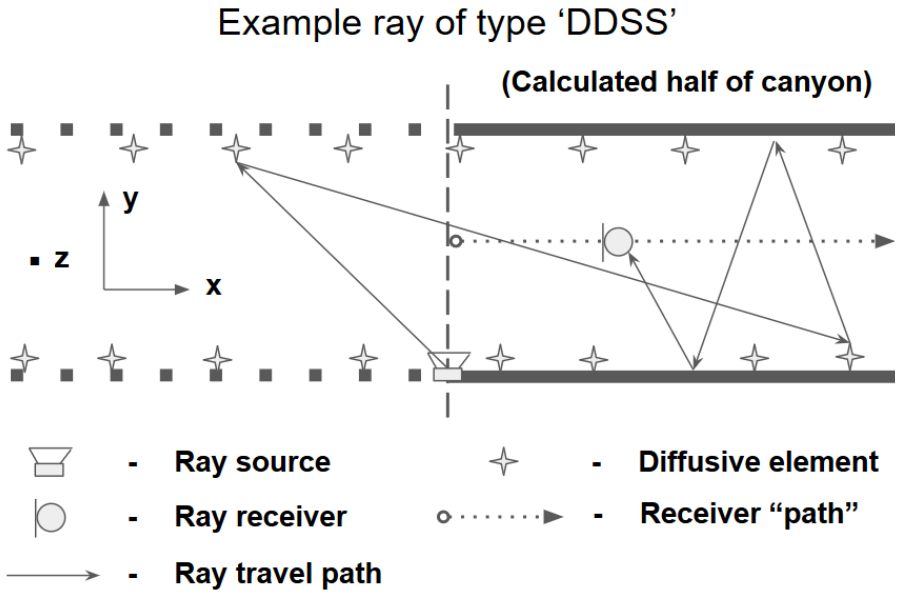


Figure 3.12.: Illustration of the ray tracing canyon geometry showing the travel path of an example ray of type 'DDSS'. The diffusers are semi-randomly placed along the entire road.

3.7.1. Diffusive area distribution

Diffusion of first and second orders, i.e. rays that reflect diffusively once and twice, are included in the model. Diffusive elements on the building façades are lumped together into fully covered areas as a ratio of length along the road. Rays that hit these areas can be diffusively reflected. Shown in Fig. 3.12 are centres of the diffusive wall areas on both walls along the canyon. Areas are placed also in the non-calculated half of the canyon in order to simulate rays, such as the one

shown in Figure 3.12, that propagate between the canyon halves. Once the calculated positions are flipped around the halfway section to give the impulse responses for the other canyon half, the distributed diffuser areas are flipped as well.

By estimation from viewing the street used as basis, shown in Figure 3.11, the ratio of diffusive to non-diffusive wall area is approximated to lie within 7 – 15%. In the model this means that for every 10 meters of wall along the road, 0.7 – 1.5 meters are said to be diffusive. Which value to choose within the range is determined by comparing achieved results with the performed measurements, see Section 6.1.3. The wall areas are given a 10 meter span along the road in which they are randomly placed. Diffusively reflected rays are calculated as reflecting off of the centres of the 0.7 – 1.5 meter wide diffusive wall areas. In the final model, after comparisons with measurements, this factor is set to 1.0 meter diffusive per 10 meter wall.

3.7.2. Absorption, reflection, scattering and diffusion coefficients

Coefficients for absorption by the walls due to specular and diffuse reflections are calculated as an average per third-octave band, based on the façade types on site. Roughly half of the walls are brick, while the other half are plaster on concrete. Furthermore, an approximated third of the wall area consists of windows and doors. Absorption coefficients of brick, plaster and window pane are therefore weighted together with a third of each. Table 3.3 shows the percentage values of the named materials in the six commonly specified octave bands of 125 – 4000 Hz (Long 2005).

Table 3.3.: Absorption coefficients (percentage of absorbed pressure) of the three different materials found at Landsvägsgatan, as well as the total after weighing a third of each.

Octave band (Hz)	125	250	500	1000	2000	4000
Brick	0.03	0.03	0.03	0.04	0.05	0.07
Plaster	0.12	0.09	0.07	0.05	0.05	0.04
Window pane	0.55	0.25	0.18	0.12	0.07	0.04
Weighted	**0.23**	**0.12**	**0.09**	**0.07**	**0.06**	**0.05**

Since the audible frequency spectrum between 20 Hz and 20 kHz is of interest, the coefficients need to cover this entire interval. It is assumed that the coefficient at the 125 Hz octave band remains constant down to the 25 Hz third octave band and likewise that the 4 kHz coefficient is constant up to 20 kHz. Narrowband coefficients are then interpolated from these third octave band values. The *reflection coefficients* for the model come from reducing the whole of the incident pressure with the calculated absorption coefficient, e.g. the operation $1 - \alpha_{weightedtotal}$ at each narrowband frequency.

For the diffusive reflections, the scattering coefficient is approximated with help of comparisons with diffusers. The model uses a generalised case to include a large variation of façade layouts. Average structural length, a, of the diffusive elements is assumed to vary between 10 to 20 cm, with a median of 15 cm. Average structural depth, h, is assumed to vary between 1 to 15 cm, with a median rounded up to 8 cm. This gives an average depth to length ratio (h/a) of 0.5. Depth and length are generalised estimations based on the size of elements at Landsvägsgatan, shown in Figure 3.11, which mainly consist of various window arrangements. Using the specified median element size values of $h = 0.08$ m and $a = 0.15$ m frequencies 2125 Hz and 1133 Hz will have half wavelengths that match the element depth and length respectively. Table 3.4 lists two examples from the random incidence scattering coefficient cases listed in Vorländer (2008) that, combined, have structural irregularities that can be compared to those found at Landsvägsgatan.

Table 3.4.: Table with two examples of random incidence scattering coefficients for two diffusive structures.

a/λ ratio	.125	.25	.5	1	2	4	8
If $a = 0.15m$, $c = 340m/s$	283 Hz	567 Hz	1.1 kHz	2.3 kHz	4.5 kHz	9.1 kHz	18.1 kHz
Trapezoidal grating($h/a \approx 0.5$)	.05	.05	.1	.9	.8	.9	.9
RPG "QRD" diffuser	.06	.15	.45	.95	.88	.91	-

The comparison is motivated by the fact that the ray tracing model treats the diffusive elements as if they were lumped together in 1 meter wide groups for every 10 meters of wall. If the diffusive elements would have been spread out evenly to match how they are actually spread on Landsvägsgatan, as seen in Fig. 3.11, this comparison would not work. If the diffusive elements were more realistically and evenly spread over the walls a scattering coefficient in the range 0.09 – 0.13, as found by Ismail and Oldham (2004), would have been more accurate.

In the two examples featured in the table, scattering is high from 1700 Hz and up (assuming $a = 0.2m$). Since it is assumed that the average structural factors can vary for different façades, the chosen scattering coefficient has a wide range of high effectiveness. However, the coefficient maximum is lowered by $\sim 45\%$ to take into account the difference between a designed diffusive element and those found on a typical building. Another reason for the lowered maxima is the assumed variation of the element lengths and depths in the span mentioned above.
Table 3.5 shows the chosen scattering coefficients at selected octave band center frequencies.

Table 3.5.: Chosen scattering coefficients for the ray tracing model.

Octave band, Hz	125	250	500	1000	2000	4000	8000	16000
Scattering coefficient	.155	.2	.4	.5	.5	.5	.4	.175

The quantity of scattered pressure is a portion out of the total reflected pressure given by the reflection coefficient. The scattered pressure is reduced further by the diffusion coefficient $cos(\theta)$ according to Lambert's emission law, where θ is the outgoing angle of the ray from the surface normal. This means that the diffusive areas are assumed to be ideally diffusive, a simplification that, while common in room acoustics modelling, is rarely fulfilled in reality according to Kuttruff (2009).

Each time a ray reflects it has a 10% chance to be reflecting off of a diffusive wall area. Therefore, 10% of the actual reflection coefficient α_{actual} is reduced by the scattering coefficient δ, as shown in Eq. 3.6.

$$\alpha_{actual} = 0.9\alpha_{weighted} + 0.1\Big[\alpha_{weighted}(1 - \delta_{scattering})\Big] \tag{3.6}$$

3.7.3. Calculation of rays for the geometry's impulse response

The amount of rays and their travel paths are predetermined by certain parameters. They are not stochastically calculated but divided into groups depending on whether they are only specularly reflected or diffusively reflected once or twice. Out of rays that are exclusively specularly reflected, only the ray with the shortest travel distance at each reflection order is counted. The set up can be compared with, if not seen as, that used with an image source method. For the diffusively reflected rays, each diffuser travel path combination possible is included as separate rays. Rays are simulated up until the 20th reflection order, which is a good limit for modelling sound pressure levels according to Kang (2007).

Factors important for describing each ray are travel distance, time of arrival as well as amount- and type of reflections. Knowing how many times rays are reflected gives how many times to reduce their amplitude with the reflection factor. One of the inputs to the simulation is the maximum allowed order of reflection, which then gives maximum reflection coefficient a ray can be reduced with (e.g. α^{20}_{actual} for a 20th order purely specular ray). The travel distances of rays that are only specularly reflected are calculated with the Pythagoras theorem, as shown in Figure 3.13.

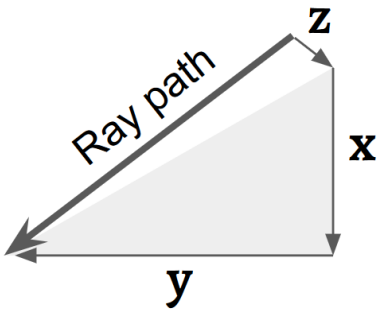


Figure 3.13.: Illustration of the way ray travel distances are calculated.

The specular rays travel the same distance along the road (x-coordinate) and height (z-coordinate) no matter which reflection order they are. The only change is in the y-coordinate which increases each time the ray passes over the width of the canyon.

For rays that are diffusively reflected, the model includes both 1st and 2nd order diffusion. The two orders are calculated in separate groups according to the following process:

First, the possible combinations of specular or diffusive reflections are calculated. Only rays with at least one, or two in the 2nd order case, diffusive reflection are included. All the orders of occurrence are saved as text strings with S and D signifying specular- and diffusive reflections respectively. As an example, a 4th reflection order ray with 2nd order diffusion could be "DDSS". This would mean that the first and second reflections are diffusive while the third and fourth are specular. Whether the letter position is even or odd determines which of the two canyon walls (left or right) the reflection happens at.

Second, there are a set amount of diffuser areas that the ray could hit for each diffusive reflection occurrence. Therefore the "DDSS" string gives a number of different rays, one of which is shown in the sketch in Figure 3.12. That ray could have travelled to any of the other 7 diffusers on the opposite wall and still fulfil the string combination. It is important to note that rays that have been diffusively reflected can be specularly reflected afterwards in the model. According to Kuttruff (2009) this is something that never happens. That is, energy that has been converted into "diffusive" after being scattered, never becomes specular again.

Since the outgoing angle from the diffusive element is important for Lambert's cosine law, the travel distances are divided into 2 or 3 intervals for 1st or 2nd order diffusive rays respectively. One distance before the diffuser and one after it (and one in between the two diffusers in the 2nd order case). Each distance interval is calculated as shown with Figure 3.13 and the outgoing angle from the diffuser is given from this as well. That is, $cos(\theta)$ (the Lambert cosine factor) is given by y_2/R_2 where y_2 is the y distance of the interval after the diffuser and R_2 is the total Pythagorean distance of the same interval. The amount of specular reflections before, between and after the diffusive reflections, are counted to be used when calculating the travel distance of each interval since each reflection adds to the y-distance.

At this point the amplitude factor of both diffusive and specular rays can be calculated. The direct ray, without reflections, has an amplitude factor of $1/R$ (R being the travel distance), i.e. scaled by spherical spreading. All other rays will have a larger R and will be scaled down further from the specular and/or diffusive reflections. The general equation for the amplitude scaling factor, in third-octave bands, of a second order diffusive ray such as the 'DDSS' example ray from Figure 3.12 is shown in Eq. 3.7.

$$Amplitude\,factor = \frac{\sqrt{(y_1/2R_1)}\,\sqrt{(y_2/2R_2)}}{R_1 + R_2 + R_3} \times F_{diff} \times F_{spec}^2 \tag{3.7}$$

Where $R_\#$ corresponds to the total length of each distance interval; $y_\#$ is the y-travel (width/across road) distance before the first or second diffusive reflection; and F_{diff} and F_{spec} are the third-octave diffusion- and specular reflection factors respectively. Since Lambert's emission law is defined for intensity, the cosine factors (written as $y_\#/R_\#$ in Eq. 3.7) are square rooted to give pressure. The factors are also halved to normalise the cosine over $\pi/2$ radians, corresponding to the half-circle that the diffuser can radiate on.

The last step of calculations is to add the contribution of all the rays from the three categories (specular, 1^{st} and 2^{nd} order diffusive) together into a single transfer function. In the actual script this is done together with the calculation of the ray amplitude factors described above. The process is in time domain and uses the amplitude factors and time of arrival of each ray. Each ray then becomes an impulse with delay according to time of arrival and amplitude from the amplitude factor.

Finally, the frequency domain transfer function that results is doubled to a double-sided spectrum and transformed to time domain. To make the signal double-sided, a duplicate of the single-sided part is added to the end of the signal. The duplicate has a flipped time axis and is the complex conjugate of the first part. The impulse responses are saved as a matrix in a ".mat" file for repeated use.

3.7.4. Application of the impulse responses as signal gain

The impulse response magnitude is added to the signal as third-octave band SPL factors. The process of calculating the gain factors is described in the following paragraph.

The end of each position's impulse response is treated with the right side of a Hanning window in order to avoid having any high frequency cut off influences on the gain. They are then transformed to frequency domain, converted to third-octave bands and calculated to SPL with reference pressure being the maximum peak for each IR. Thereafter, the SPL values from the fixed positions are interpolated to give an IR for each time sample of the truck signal. A copy of the interpolated values is time-inverted to form the SPL magnitude factors of the non-calculated half of the canyon. This copy is joined with the other to form the entire canyon signal. Finally, the gain factors in third-octave bands are applied to each sample of truck signal by adding the third-octave pressures of the two together.

4. Part 2 - Internal truck sound

In this part the implementation of using a granular approach for sound synthesis is described. A database of a single truck is created and different types of driving cases are set up, both acceleration and constant speed, to be used to simulate the sound inside a drivers cabin. The cases are created to compare and evaluate if the granular synthesis could be better than the current model that exists today for a truck simulator.

4.1. Grain database

The grain database is created using a single truck, the FM 490. The data comes from measurements previously made by Volvo Group in their semi-anechoic laboratory, performed in the same way as mentioned Section 3.2. The only difference is that instead of using the microphones located around the truck, the two microphones located inside the drivers cabin are used. The microphones are located on the left and right side, at the driver and passenger seat and since they are inside the cabin, they represent the actual sound that is heard better than the microphones outside.

The database is built on two types of measurement files; constant speed and whole city cycle cases. City cycle means that the truck is run in idling, various constant speeds and different types of accelerations, all at different gears, similar to that of driving in a city. The reason for using multiple measurement files is to get a big database. A big database means more accurate driving cases generated and higher rate of realism, since there is more data to choose from. Some of the recordings, or parts of them, contain noise from various types of sources such as warning signals, speech and so on, so they are removed. This is also why multiple measurement files are used, so that the database can contain at least the minimum number of grains needed.

4.1.1. Measurement data

The measurement data is loaded from the measurement files previously mentioned. Both microphone signal data and the CAN-bus signals, i.e. rpm, gear and torque are loaded. The microphone signals are recorded with a sampling frequency of 44.1 kHz while the CAN signals are recorded with a sampling frequency of 100 Hz. The CAN signals are therefore interpolated to have a matching amount of samples to the microphone signals, assuming the same value is constant over 441 samples.

4.1.2. Signal simulation

The signals used to create the grains are based on the data recorded from the two microphones. The truck simulator located at Volvo has two speakers, one behind the driver seat and one behind the passenger seat, same as the recorded signals. This means that the optimal solution would be to create a left and right signal. But for initial testing, the two recorded signals are combined in to one in a similar way as in Section 3.3. Cross-correlation is used to get the best match between the signals, by using microphone one as reference and shifting signal two a certain number of samples. A high-pass filter with a cut-off frequency of 20 Hz is applied to the signal in order to attenuate the frequencies below 20 Hz.

4.1.3. Creating grains

In order to create the different grains, the measurement data is first divided into sections based on rpm, gear and torque. An example of this can be seen with the black dashed vertical bars in the center graph of Figure 4.1

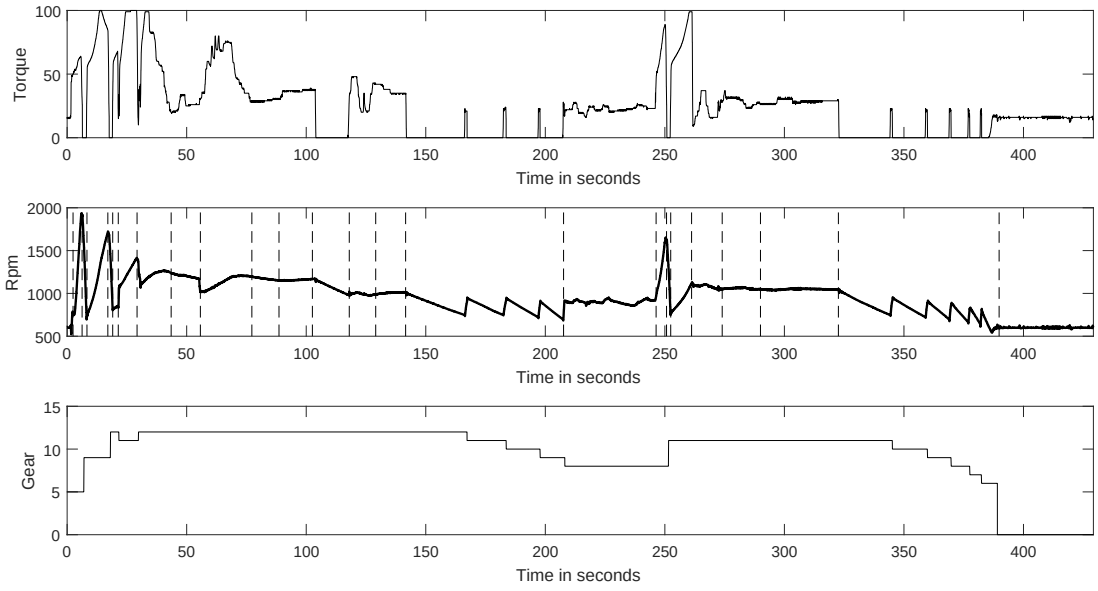


Figure 4.1.: Setting up grain database

Each section is defined as acceleration, constant speed or trash, i.e. no use for the grain database in this work. This is done in order to sort the constant grains and acceleration grains into separate groups. The sections labelled as "trash" are either when de-acceleration occurs or when the rpm is varying too much over a short period of time. Again, this is done because only acceleration and constant speed granular synthesis is of interest in this thesis. Though, for a real-time simulation model, de-acceleration should be included in order to fully cover a driving cycle.

The method of creating the grains follows the same method used by Seward (2014). In order to find the grains, some variables are set up: the wavelet size, the search area and the number of samples to skip so that the tails of the grains do not overlap. The wavelet size is the number of samples that equal to one ignition, i.e. 1/6th of a typical combustion cycle for a 6 cylinder ICE. The search area is the area to search for each grain, in both negative and positive time direction. The size of the area is defined as 55% of the wavelet size and the number of samples to skip is defined as 6.5 times the size of the wavelet size, i.e. slightly above one combustion cycle. A initial skip in samples of $0.1 f_s$ is applied for the first grain so that the left tail can be created.

The grains are detected by running through the whole signal, in small increments where each section, as mentioned before, is defined as either acceleration or constant speed. If the section is labelled as constant speed only 12 grains are created. The reason for using 12 is because Seward (2014) concluded that it would be enough to get good sound quality. If the section is labelled as acceleration then the whole section is run through. This means that size of the acceleration grain database is multiple times bigger than the constant speed database. So for each iteration, a portion is extracted from the samples contained in the search area. This portion is equal to one combustion cycle, i.e. six wavelets plus some samples to account for variations. A "peakshape" function is created to find the first relative maximum of this combustion cycle that consists of six sinusoidal wavelengths that is based on the current wavelet size. The function is convoluted with the signal to find the initial peak (relative maximum) of the grain. That and the five peaks that follow after are taken to be the core of the grain, i.e. six cylinder ignitions.

For each iteration, the core of the grain together with the, so called, left and right "tails" are saved in structure array database. The left and right tail has the length of six wavelets and are the six cycles that immediately precede and follows the core. Also saved together with the grain is the rpm that corresponds to the beginning and end of the core and the current gear for that grain.

4.2. Synthesis - assembling a signal

Overlap and add is used to place grains in a continuous signal. The difficulty lies in finding and choosing the correct grain for the desired scenario while merging the grains is a simple process. Computation speed is of lesser importance, since the purpose is "proof of concept" rather than on-line implementation. This means that one of the three perhaps main challenges, the other two being grain detection and grain selection, is ignored.

Input data for the synthesis consists of the driving case information, mainly engine rotation speed and gear. The length of the driving case directly determines the length of the synthesised signal. Engine torque is not utilized in the process, it is implicitly included due to the distinction between constant speed- and acceleration grains.

The synthesis is done by reading the desired gear, drive mode (acceleration or constant speed) and rpm at a particular time sample and then using either an acceleration function or a constant speed function to add one grain to the signal. After adding the grain, the driving case at the time sample by the new signal end is read and the process repeated. If the generated grain crosses over from an interval of constant speed drive mode to acceleration or vice versa, the signal is cut back to the point where the mode changes. That is, the script will always start an acceleration grain from the same time sample as expected from the desired driving case.

4.2.1. Constant speed

In the case of constant speed, the desired rpm curve will either remain flat or fluctuate around a certain value. Grains that are measured at the same gear, with appropriate ($<\sim$ 40%) torque and that contain the desired rpm value +/- 5 rpm are therefore included as candidates for the synthesis. If no grains are found with the initial tolerance, it is iteratively expanded up to +/- 20 rpm. This tolerance limit could be increased if there is a lack of grains in a certain rpm range.

With exception of the case when the grain to be added constitutes the start of the signal, conditions are set to avoid repetitive use of the same grain. This is to avoid the creation of signals that are perceived skipping or looping. The conditions do not allow the grain from the previous iteration ("AA" repetition) nor the grain from two iterations ago if the previous is the same as the one from three iterations ago ("ABAB" repetition). These conditions are ignored in cases where there are too few grains to choose from to fulfil them.

After the conditions have been passed, the grain with the rpm closest to the previously placed grain is chosen out of those that remain available. The chosen grain is then added to the existing signal with the OLA method, meaning that its core is put directly after the previous grain's core in the existing signal.

4.2.2. Acceleration

While the OLA process is the same for acceleration as for constant speed, the selection process is more stringent. Since the engine rpm increases continuously during acceleration, grains themselves will contain a, possibly steep, rpm gradient. To give the impression of a constantly accelerating engine, it is therefore important that the grains used have rpm gradients that match the desired rpm curve. The possible variations of gradient, start and end RPM, mean that a large amount of different grains is required to follow a given, arbitrary, curve.

A process to find the grain with the rpm characteristics closest to the desired case is implemented. The desired rpm at the time step where the new grain is to start is used to calculate an approximate length of a complete engine firing using Eq. 4.1.

$$n_{samples} \approx 120 * FS \,/\, rpm_{desiredstart} \tag{4.1}$$

The rpm values from the desired case for the entire duration are used to create an average rpm during the grain duration. Using the average, a new approximated grain length is calculated with Eq. 4.1. If the length exceeds the desired signal length, it is cut down to match that instead. This is used to find the approximate desired rpm at the end of the grain to be placed. Both the desired start and end rpm can now be used as filtering criteria in the database of acceleration grains. First, the difference between the available grains' start rpm and the start rpm of the desired case is calculated. The grain with the smallest difference and a few number (0 - 12) of grains with higher and lower start rpm than the closest are chosen. The difference between the selections' ending rpm and the desired end rpm is then calculated. This difference is added up with the previous to finally give the grain with the smallest total difference, which is then chosen for use. Unless the chosen grain is the first in the signal, the previously used grain is loaded and merged with the current using the OLA process described under constant speed.

4.3. Driving cases

As mentioned in the beginning, in order to evaluate the granular synthesis and compare it with the existing method used in the driving simulator, a couple of driving cases are set up. Both methods will be trying to simulate sounds similar to these cases. The driving are of two types: acceleration and constant speed. For the constant speed type, 5 cases are created: Idling, 20 km/h, 30 km/h, 50 km/h and 70 km/h. These can be seen in Figure 4.2:

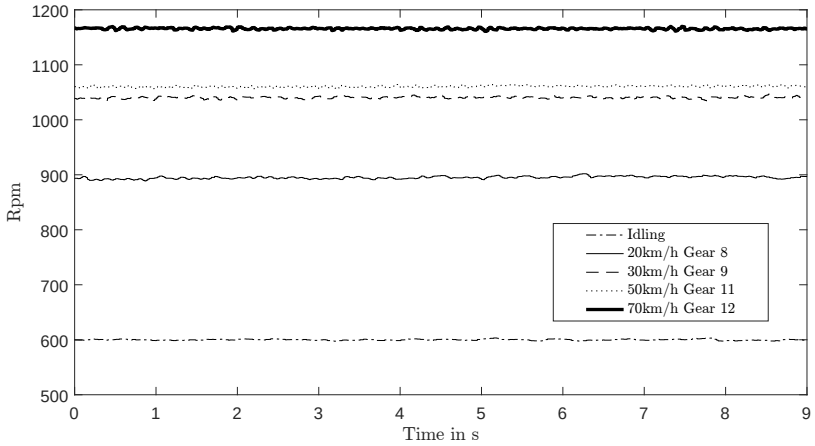


Figure 4.2.: Different driving cases - constant speed or idling.

For the acceleration type, three different gears are considered. The first, gear 9 can be seen in Figure 4.3:

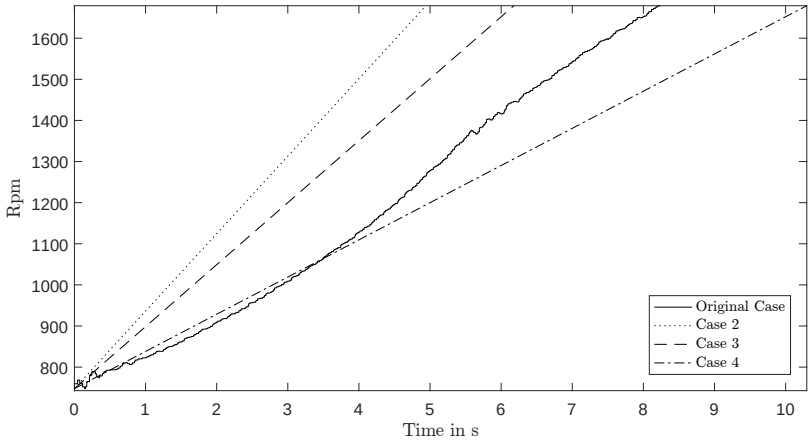


Figure 4.3.: Acceleration driving case - Gear 9.

The solid black line is the actual signal recorded and what the grains are created from. The three other lines are new cases that are created that the granular synthesis is supposed to be matching. Two of the cases accelerate faster (40% and 25%) than the original recording and one accelerates slower (25%). It can also be seen that the made up cases are linear which is not representative of real acceleration, but will have to suffice as an initial evaluation of the method. Figure 4.4 shows the second acceleration type, using gear 11. Three cases are created here as well, all of which are linear, where two are above the original recording (40% and 25% faster) and one is below (25% slower).

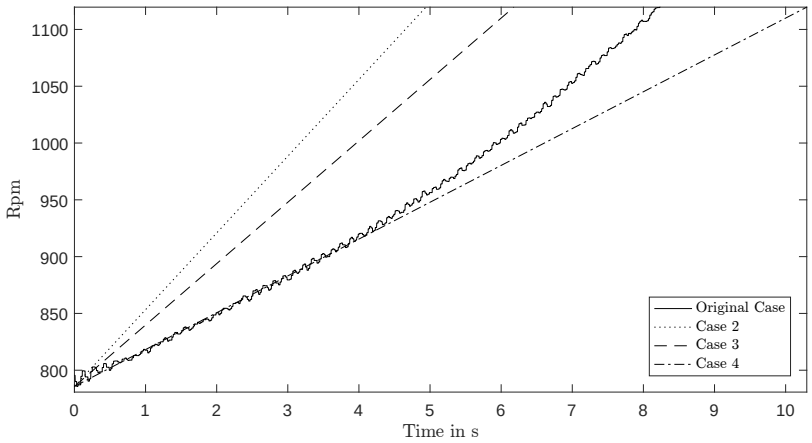


Figure 4.4.: Acceleration driving case - Gear 11.

For the last acceleration type, with gear 12 see Figure 4.5, only 2 cases are considered and both of them are located above the original recording (40% and 25% faster). The reason for this is because creating an acceleration case slower than the initial recording would be too long for a listening test. Also, the rpm is not varying too much in this case, which means that it would almost be perceived as a constant speed case.

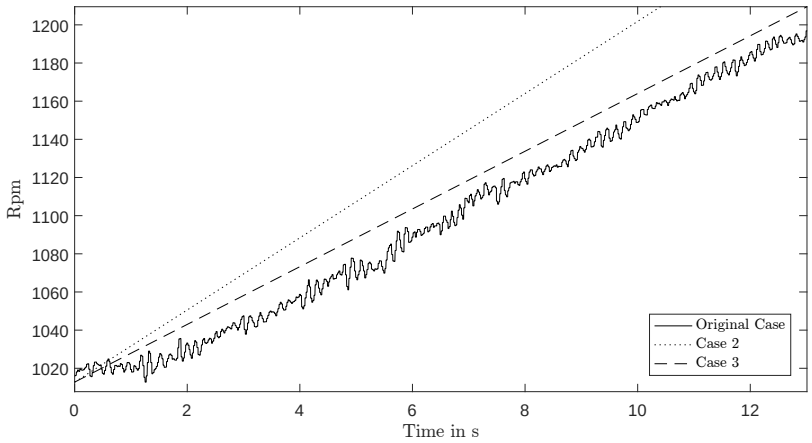


Figure 4.5.: Acceleration driving case - Gear 12.

5. Listening test

In order to evaluate the sounds generated from both parts of the thesis a listening test is set up. The test is made based on the *NORDTEST Method* for listening tests. It describes how a test should be set up; type of listening test, preparations, instructions, how it should be run etc. Some changes had to be made in order to fit the sounds being evaluated and the limited time frame. For part 1, i.e. the computer model, two types of tests are made; a semantic differential test and a comparison test between the measurements and the simulated sounds. Only outdoor sounds at floor four at the façade are used in the listening test as the indoor apartment sounds are affected by fan noise and voices. For part 2, i.e. the granular synthesis, a direct pair wise comparison test is set up, comparing the sounds created using granular synthesis with the sounds that exist for the driving simulator at Volvo Group today.

The three parts takes place at the same time and with a total of 16 participants, 3 female and 13 male. The average age of the participants is 35.7 years with a standard deviation of 12 years. One participant reports that he suffers from a slight hearing impairment on his right ear and the others report no hearing impairment.

The test set-up consists of a computer and a pair of Sennheiser HD650 headphones with the test made in MATLAB together with a GUI for easy user input. In Appendix E the user interface can be seen. The output from the headphones used for the listening is measured before the tests in order to know the sound pressure levels coming out. A setting is chosen based on the sounds being evaluated so that it won't be too loud for the listeners. Table 5.1 shows the recorded values for the pass-by sounds, both simulated with the computer model and from the measurements at Landsvägsgatan, at the ears of head dummy compared to output values from the computer, i.e. taken from the signal data. The values are fluctuating much so the table values should be seen as a rough estimate of the maximum values. As the pass-by signals are based on outdoor measurements, they have to be scaled down to avoid "clipping" in the signal caused by the *.wav* file format. The pressure amplitude is scaled down with a factor of 0.6, which is equal to a decrease of $\sim$ 4.4 dB. Together with volume set in the computer, the A-weighted levels perceived at the ears are around 2 - 3 dB lower compared to the actual levels. The un-weighted levels are between 5 and 10 dB off compared to the actual levels.

Table 5.1.: Measured output values at the ear compared to calculated values - Exterior sounds

		Exterior							
		20km/h Gear 7		20km/h Gear 8		30km/h Gear 9		Idle	
		Simulated	Measured	Simulated	Measured	Simulated	Measured	Simulated	Measured
Recorded	dB	77	79	77	79	77	81	73	75
	dB(A)	74	75	73	75	73	76	71	69
Real	dB	86.7	85.9	88.2	85.4	88.3	85.7	82	85
	dB(A)	76.8	77.1	74.6	77.3	76.6	77.9	69.4	70.4

Table 5.2 shows the recorded values for the interior truck sounds at the ears of a head dummy compared to output values from the computer, i.e. taken from the signal data. Only some of the constant speed cases are presented here. The acceleration cases are not measured as the levels increase over time. No modification to the part 2 sounds is necessary in order to be able to write to *.wav* files, which means that only the computer amplifier changes the levels. The difference between the actual levels and the measured ones can be seen here as well. Here the A-weighted levels are further off.

Table 5.2.: Measured output values at the ear compared to calculated values - Interior sounds

		Interior							
		20km/h		30km/h		50km/h		70km/h	
		Granular	Existing	Granular	Existing	Granular	Existing	Granular	Existing
Recorded	dB	75	79	77	81	80	81	80	81
	dB(A)	65	71	67	73	71	73	71	73
Real	dB	80.1	75.5	82.4	77.4	84	77.6	83.8	77.5
	dB(A)	57.7	64.8	60.3	67.3	61.5	67.4	61.9	67.7

For both sounds, i.e. exterior and interior, something is happening between the raw data and the sound recorded at the ears of the head dummy. Possible causes could be the frequency range of the headset, how the headset was mounted and the range of the microphones used for recording. This means that both the levels and the frequency content is different for the listeners compared to the real scenario. This should be considered and most likely affects the listening test result.
As mentioned before, the listening test consists of three parts:

- A semantic differential test for the computer model and the measurements at Landsvägsgatan. Evaluating parameters related to the sounds.
- A comparison test between the computer model and the measurements at Landsvägsgatan. Each sound is directly compared to its counterpart.
- A direct pair wise comparison test between the existing simulation sounds and the new made with granular synthesis. Each pair of sounds are only evaluated against each other.

The primary focus of all the test is realism. The reason for having two tests for part 1 is the thought of evaluating the sounds both separately and in pairwise comparison to each other. For part 2, the direct pairwise comparison is the best choice as primary reason is to compare the old model against the granular synthesis and to see if the granular synthesis is a better method. In order to make the test more valid, the play order varies randomly for each participant. This means that both the order of each sub-test and the order of the sounds is different. By doing this, the effect of having the test and sound in a specific order is minimized. The listener has the option of playing each sound infinite amount of times and to go back in each sub-test to change previous answers. The possibility to add comments for each part also exists if anything is unclear or an answer needs to be explained. First a short explanation is given orally and then in written form for each sub-test. The average test time is around 30 minutes.

6. Results and Discussion

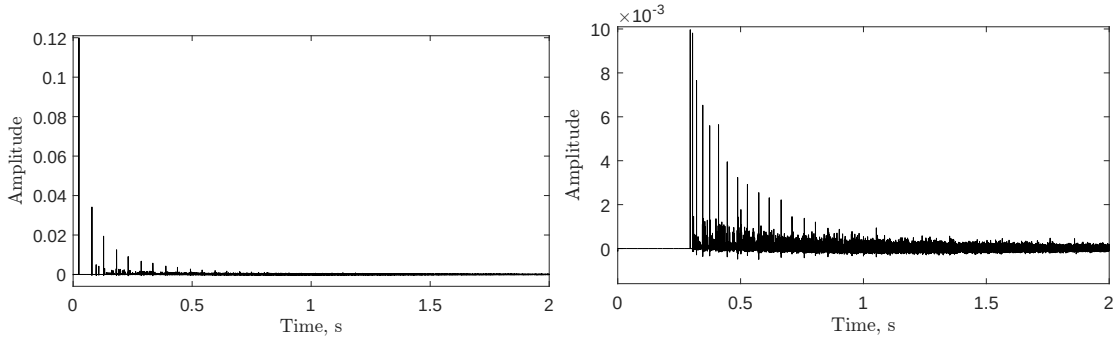
6.1. Ray Tracing: Impulse responses and parameter studies

Most of the presented simulation results show only two out of the large number of receiver positions along the road. The two chosen positions are by the source position (in the middle along the street canyon's length) and at the end of the canyon (100 meters away from the middle). Direct sound dominates at the first position while different reflections are big contributors at the second. Essentially, the further away from the middle the receiver gets, the smaller the propagation difference between reflections and the direct ray becomes. While most reflections are attenuated from absorption and scattering, the lower orders are still of amplitudes comparable to the direct ray.

The main purpose of this section is to show the reasoning behind the choice of certain parameters that had a less solid ground in theory, the scattering coefficients and diffusive area ratio. There are also comparisons between the outcome of running the model with and without certain features, such as averaging iterations with different canyon widths, to show their effect.

6.1.1. Impulse response

Figure 6.1 shows the modelled impulse response at the two mentioned positions. Since averaging is done with the gain factors that are calculated from the IR:s, the figure shows the response from the Landsvägsgatan dimensions.



(a) Receiver in front of source, both at mid canyon position along road.

(b) Receiver at end of road (100 meters "upstream" from source).

Figure 6.1.: Impulse responses of the modelled city canyon. Non-averaged, single width of Landsvägsgatan's dimension.

Between the direct sound and the first reflection in Figure 6.1a is a slightly wider delay than an actual, comparable, measurement would yield with geometry in question. The lack of ground reflections (due to ignoring ground all together) explains this. However, the energy loss is assumed to be compensated by the concrete floor reflections in the measurement carried out semi-anechoically that is used as input for the simulation.

The impulse responses in mid-canyon positions, where the propagation distances are short, are also affected by the removal of reflections of the truck due to the transparency assumption. For listener positions on lower floor levels, the truck could screen off or redirect some of the radiated sound. However, this is estimated to have a low impact for the overall sound pressure level since the direct field dominates at short propagation distance.

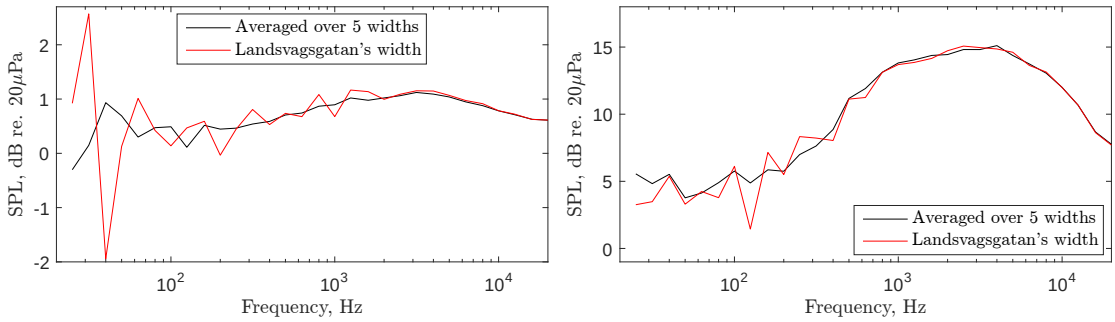
Figure 6.1b shows time instances with noticeably higher pressure than earlier instances; a side effect from the fact that the rays are predetermined and not propagated stochastically. Especially the diffuse reflections contribute since every diffusive element is included with at least one ray. All such rays, of the same order, that reflect off of diffusive areas distributed in between the source and receiver arrive at exactly the same time. These, together with the same order specularly reflected ray, creates the high peak that is just after and of almost the same amplitude as the direct ray. The same phenomenon explains the later peaks that are sometimes of higher amplitude than earlier ones.

6.1.2. Effect of averaging the gain factors

To achieve a more generic gain effect in the simulations, the gain factors from five different canyon width iterations, with scaling factors shown in Fig. 3.2, are averaged. Together the factors make out an average factor of 1.21, i.e. they scale up the width slightly compared to Landsvägsgatan's input width. This means that the average ray propagation distances along the width axis are 21% longer. Factors are chosen to not be multiples of each other, e.g. not 2 and 4, since that would not change the resonant wavelengths. A wavelength, λ, that is resonant at width scale 2 will correspond to 2λ and still be resonant at width scale 4. The desired effect is to have widths where different wavelengths are resonant at each iteration, in order to ultimately cancel out each iteration's "side" effect. Factors are also chosen to scale the input width realistically, hence the choice of factors within 0.6 from 1.

Important to remember is that the diffuser positions are semi-randomly changed between iterations, as described in Section 3.7.1, meaning that spectral differences aren't necessarily derived exclusively from width changes.

In Figure 6.2 the result of running five iterations and averaging the gain factors is compared with the case using only width factor 1.



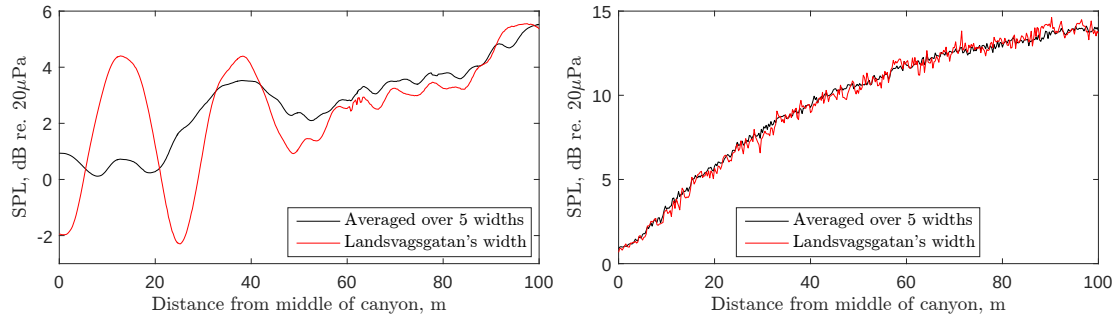
(a) Receiver in front of source, at same position along road.

(b) Receiver at end of road (100 meters "upstream" from source).

Figure 6.2.: Third-octave band spectra comparison between averaged and non-averaged gain factors. From one receiver position along road.

Averaging clearly removes most of the resonant behaviour both near and far from the canyon middle. The extra propagation distance from the higher average width does not show a noticeable gain reduction. This is possibly due to effects from changed diffuser positions or, more likely, due to the decibel scale not showing such minor differences.

Figure 6.3 shows a comparison between averaged gain factors and the "width factor 1" case, but at a single third-octave band over distance from the middle of the canyon out to the end (0 to 100 meters).



(a) Gain factors at the 40 Hz band.

(b) Gain factors at the 1 kHz band.

Figure 6.3.: Single third-octave comparison between averaged and non-averaged gain factors. Receiver positions from middle to upstream end ("0 - 100 m").

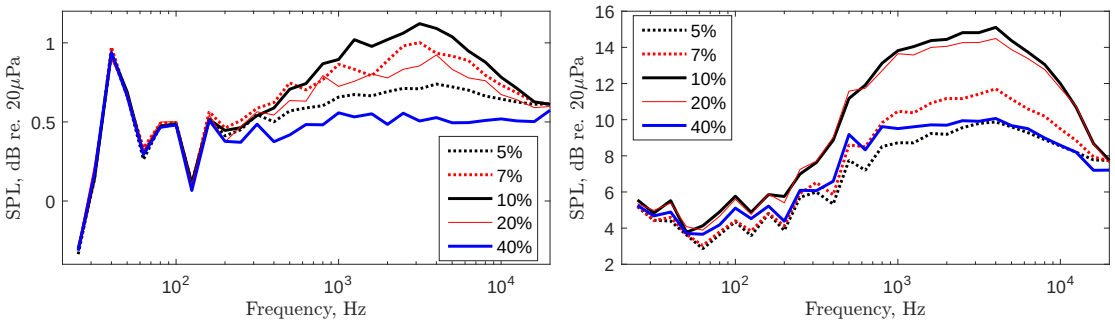
At the 40 Hz third octave band the gain shows a strong resonant behaviour that is not entirely cancelled out by the averaging. The averaged gain factors are also slightly higher in gain than the non-averaged case. At 1 kHz, the majority of the fluctuations are smoothed out. Clearly, averaging over several geometries is necessary for a simulation with a purpose of giving a generic city environment gain effect. The modelling only includes the physical causes and effects of a real amplification case to a certain degree and likely exaggerates some of the resonances in each iteration.

While the model is made to cover the entire audible frequency spectrum from 20 Hz to 20 kHz, data for material absorption in the entire spectrum is rarely stated in source material. This leads to the assumption of constant coefficients from the commonly provided 125 Hz and 4 kHz octave bands down to 20 Hz and up to 20 kHz respectively. The extrapolation simplifies the low/high frequency behaviours of the façades in the model and introduces an extra degree of uncertainty.

6.1.3. Influence from ratio between diffusive and flat/non-diffusive wall area

Measuring an exact ratio of diffusive element to flat wall area seems like an unnecessarily complicated approach for this model. The difficulty of evaluating the actual scattering coefficients, with the only literature found being a scale model investigation, is also complicated. These two parameters are therefore used as two "degrees of freedom" to tune the model to match better with the measurements from Landsvägsgatan.

In Figure 6.4 the influence of varying the "diffusive to flat wall area" ratio is shown for the two chosen positions in the modelled canyon.



(a) Position with source and receiver at same, mid canyon, location along road.

(b) Position with receiver at end of road, 100 meters upstream from source.

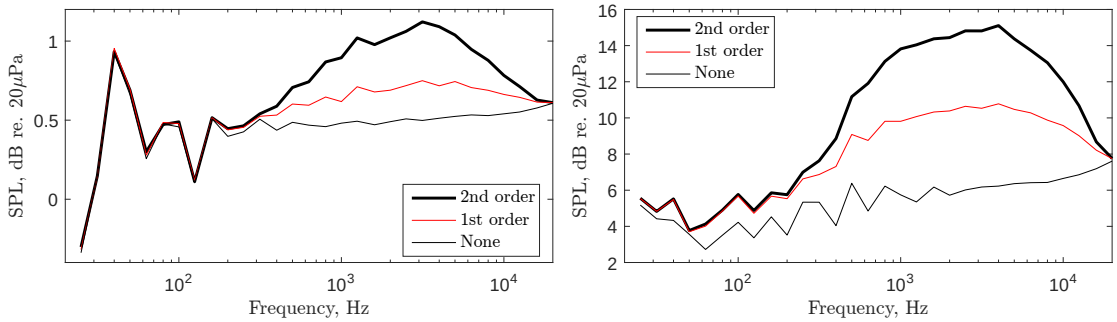
Figure 6.4.: Third octave band gain factors for different ratios of diffusive to non-diffusive wall area.

The comparison between the two positions reveals that there is a parabolic pattern in the gain resulting from increasing the ratio of diffusive wall area. There is a ratio, 10%, that gives a maximum over which the gain drops again. An explanation to this is found in the effect of increasing the amount of diffusive rays and in the way the model treats the reflection coefficient. More diffusive wall area results in more rays that contribute to the total pressure at the receiver. Only counting second order diffusive rays a ratio of 40% has 846280 rays while at a ratio of 5% the number is 80201. Simultaneously, an increasing ratio will reduce the reflection coefficient, as shown in Eq. 3.6 (Section 3.7.2), reducing the amplitude of all reflected rays. Using 1 kHz as an example, the coefficient is ~ 0.74 at 40% but ~ 0.91 at 5%, a difference of $\sim 21\%$. That means an extra 21% reduction in energy with each reflection order at 1 kHz, cancelling out the effect of increasing amount of rays, as seen when comparing the 40% and 5% cases. Since the

10% ratio gives the highest gain, it is chosen for the model. By maximising the effect of the diffusive area ratio, the scattering coefficient has maximal influence and becomes the last degree of freedom for the resulting gain factors.

6.1.4. Effect of allowed diffuse reflection order

To show the influence of diffusive reflections, Figure 6.5 shows the spectra at the two example positions for three cases: without, with first order, and with second order diffusive reflections.



(a) Position with source and receiver at same location along road.

(b) Position with receiver at end of road, 100 meters upstream from source.

Figure 6.5.: Third-octave band gain factors with 2nd order, 1st order and no scattering.

Scattered rays make out a large part of the total reflected energy in the gain factors. This makes it clear that an accurate scattering coefficient is essential in order for the model to have gain spectra that correspond well to reality. The coefficient implementation is dependant on the structure and dimension of the façade irregularities, which in turn are approximated. The chosen approach of aggregating the diffusive areas into 1 meter wide elements make this approximation even more important since it means an overall higher scattering coefficient. Since the model is so dependant on an approximation, reworking the implementation of scattering to eliminate the need for it would be well advised. This would be done by spreading the diffusive areas evenly along the walls and lowering the coefficient to $9 - 13\%$ as found in the scale model investigations by Ismail and Oldham (2004). Changing the area distribution would also require a change in the predetermination of the diffusive ray group's travel paths. Otherwise, there would be too many diffusively reflected rays in the model resulting in an exponentially longer computation time.

The lack of air attenuation on each individual ray is another factor that has a sizeable effect on the contribution of scattered energy. Especially at positions far from the receiver, where the reflections have a long propagation distance. The effect per meter of propagation, as shown in Table 3.1, could potentially reduce frequencies above 4 kHz with 10+ dB at the 100 meter position shown in Figure 6.5b. For positions close to the receiver, the effect is not as large; early reflections and the direct sound are most important for the SPL there.

6.2. Part 1 - Model validation

As mentioned in the section explaining how the auralization model is implemented, the computer model is modelled after Landsvägsgatan, in order to validate with already existing measurements from that street. Those measurements are made at the fourth floor with microphones located both on the outside and inside an apartment. The fourth floor is therefore used for the computer model as well. For the comparison the same microphones used in the background noise measurement are chosen (see Section 3.4.5) for the measurement files. The same truck is used for the Landsvägsgatan measurements as in the semi-anechoic chamber.

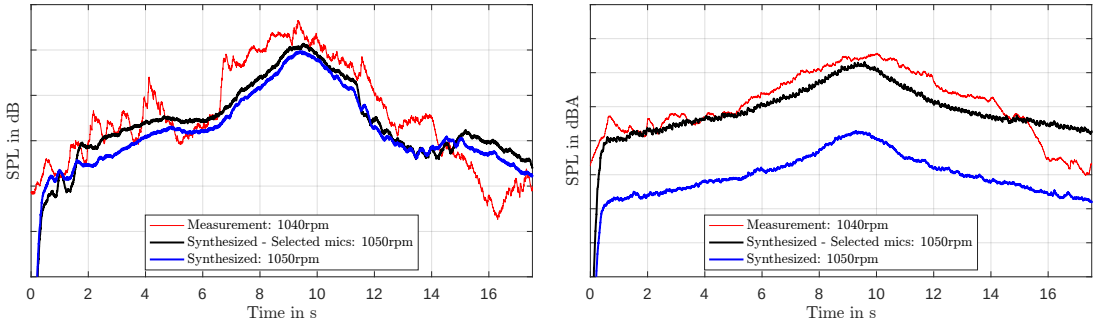
From the measurements one idling case and three pass-by cases are chosen to be used in the computer model: 20 km/h using gear 7, 20 km/h using gear 8 and 30 km/h using gear 9. The reason for choosing these four cases is due to each case at Landsvägsgatan being matchable with a case inside the semi-anechoic chamber; both gear and rpm match to a reasonable degree. In this section only one of the cases are presented, the 30 km/h case using gear 9. The rest of the cases can be see in Appendix C for sound pressure levels in time domain and Appendix D for the 1/3-octave band sound pressure levels.

Start position for the truck in the model is 80 meters to the left of the apartment and end position is 80 meters to the right. This is due to the fact that the recordings from the semi-anechoic laboratory could not generate a signal for a longer distance. The measurements at Landsvägsgatan are not made with this set-up, which is why the computer model has to be adjusted to be matched with the measurements. This means that the figures only show the length of the measurements.

6.2.1. SPL - Time domain

In Section 3.2 it is mentioned that two cases are considered; one using all microphones and one using a selected number of microphones located in the front of the truck. It can be seen from the two figures that the "Selected" case has a frequency content that is closer between the microphones. It can also be seen that levels are higher in frequencies above 200 Hz. The impact of this will be seen in the figures below.

Figure 6.6 shows both the un-weighted and the A-weighted sound pressure levels for a truck driving 30 km/h using gear 9. The red line is the measurement at Landsvägsgatan, driving with a slight rpm difference compared to the computer model. The blue line is the synthesized signal using all microphones and the black line shows the signal using only the selected front microphones. For the un-weighted case, see Figure 6.6a, there is hardly any difference between using all microphones and just the selected ones. When the A-weighting filter is applied, see Figure 6.6b, the 8 - 10 dB difference can clearly be seen.



(a) Un-weighted (b) A-weighted

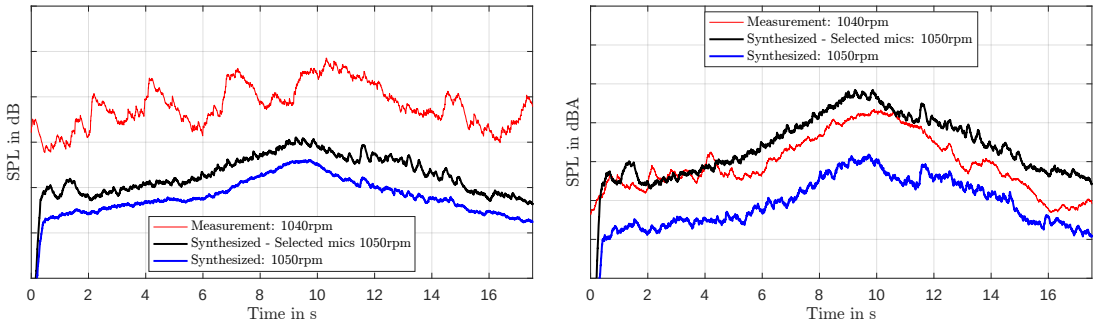
Figure 6.6.: Sound pressure levels at the façade for a truck driving 30 km/h, gear 9. Increments of 5 dB per horizontal line.

The computer model is quite close to the real measurement, with only a few dB off. For the computer model, the dip in area in the signal corresponding to 11.5 - 14.5 seconds seen in Figure 6.6a, is caused by distortion in the measurement signal and is not caused by the computer model. The dip in the beginning for the computer model, 0 - 0.4 seconds, is caused by the application of the Hann window mentioned in Section 3.5.

The primary difference between the model and the real measurement is the smoothness of the curves. The computer model uses an averaged canyon gain while the measurement is affected by the canyon characteristics typical for Landsvägsgatan. This means that while the model is close, it will not get the extreme cases that can happen in the real canyon. Also, the measurements contain noise from the truck that does not exist in the semi-anechoic room such as exhaust/ventilation noise. The background noise is not included in Figure 6.6 either and the influence of that can be seen later in Figure 6.11 in Section 6.2.2.

Another parameter that is not included in the model is tyre/rolling noise. As the measurements in the semi-anechoic room are made with the truck standing still, no rolling noise is recorded. This means no consideration to weight, tyre type or type of road surface is made. Implementing this would make the model more accurate but also more complex as more unknown parameters would be introduced. For trucks, rolling noise has a significant impact above driving speeds of 50 km/h (40 km/h if driving with constant speed), where it is dominant over propulsion noise. Theoretical values of the missing amount of pressure is -3 dB for 50 km/h and -5 dB for 80 km/h if it is not consideed. Even at lower speeds it still has an impact (Keulen and Duskov 2005). This means that it should be investigated further if the model is run for higher driving speeds.

Figure 6.7 shows the un-weighted and A-weighted sound pressure levels inside an apartment located at floor four at Landsvägsgatan. The computer model uses the reduction index that was previously estimated and mentioned in Section 3.4.4 (case 5 in Figure 3.8). The big difference for the un-weighted case is caused by higher levels of the low frequency content in the measurement signal. This can be observed in the 1/3-octave band plots in Section 6.2.2. When the A-weighting filter is applied, these frequencies are attenuated and the measurement seems to match quite well with the computer model using the selected microphone case.



(a) Un-weighted. Increments of 10 dB per horizontal line.

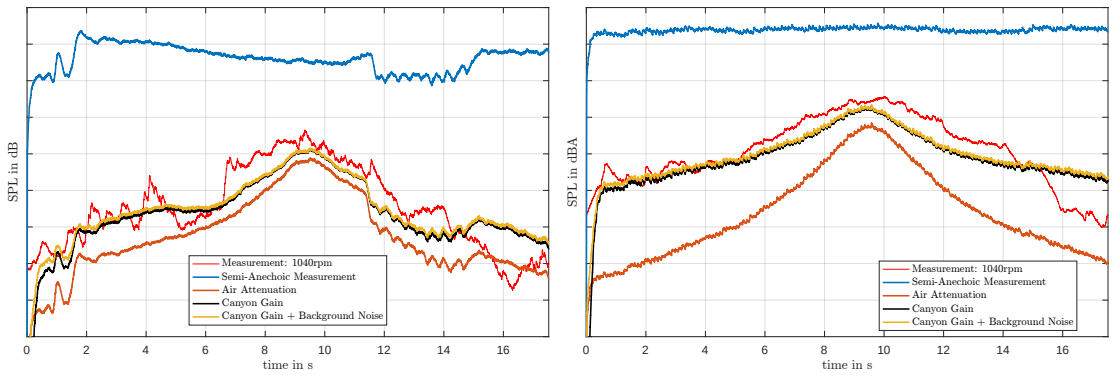
(b) A-weighted. Increments of 5 dB per horizontal line.

Figure 6.7.: Sound pressure levels indoors for a truck driving 30 km/h, gear 9.

Using all microphones shows a similar difference as in the façade case which leads to the conclusion that using only the selected microphones, located in the front, are recommended for use in the model. Using all the microphones might give a more realistic sound for listening purposes but will cause the model to generate sound pressure levels that are too low. Though sound pressure levels should be slightly lower and change characteristics when the truck is moving away from the receiver position, as the back of the truck is closest to the receiver during that time. This is something that the model is lacking and is caused by only using the front microphones. Figure 6.6 and Figure 6.7 are indicating just that, where the measurement signal is dropping both faster and lower when it moves away from the receiver.

The driving conditions are also worth mentioning. Even though the truck is driven with the same gear, a slight rpm difference exist. The vehicle speed in the measurements at Landsvägsgatan are also fluctuating a lot more compared to the speed kept in the semi-anechoic room. This is the case for all the driving conditions in Appendix C as well. This could be a contributing factor to the difference between the computer model and the measurements. As an example, in Figure 2 in Appendix C the primary peak has shifted when the truck is run with a slight increase in rpm.

It can also be of interest to see the impact of some of the parameters applied on the signal. Figure 6.8 shows the signal at the engine, named "Semi-Anechoic Measurement", after combining all the contribution of all the signals but before applying any other parameter. Also shown is the signal after both distance, Doppler effect and air attenuation is considered, named as "Air Attenuation" in the figure. "Canyon Gain" is the signal after all parameters are applied and also plotted is a version including background noise.



(a) Un-weighted

(b) A-weighted

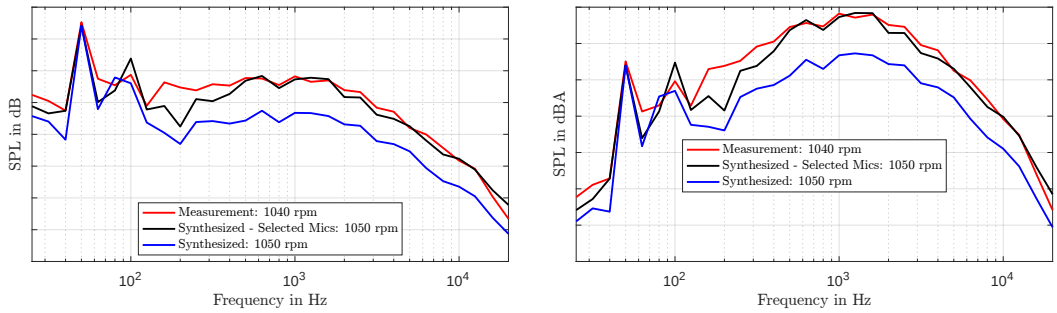
Figure 6.8.: Sound pressure levels at the facade, after application of different parameters, for a truck driving 30 km/h, gear 9. Increments of 5 dB per horizontal line.

It can be seen that there is some variation in the semi-anechoic measurements as well, fluctuating with around 5 dB at the more constant areas. The distortion at 0 - 2 s and 12 - 15 s is even more visible. Again, applying background noise to the signal has very little impact. The influence of the canyon gain is better seen in the A-weighted sound pressure levels.

6.2.2. SPL - Frequency domain

In this section the corresponding third octave band sound pressure levels are plotted at the maximum position, i.e. where the measurements at Landvägsgatan and the computer model show the highest SPL. Points before and after the maximum point are evaluated and show similar results and are therefore not included in the report to reduce the amount of plots.

Figure 6.9 shows the 1/3-octave band SPL:s at the façade. As mentioned before, the difference in SPL:s between the measurement and the computer model can primarily be explained by the high levels in the low frequency range. When the A-weighting filter is applied, these frequencies are attenuated which is why the levels seem more alike in Figure 6.6b.



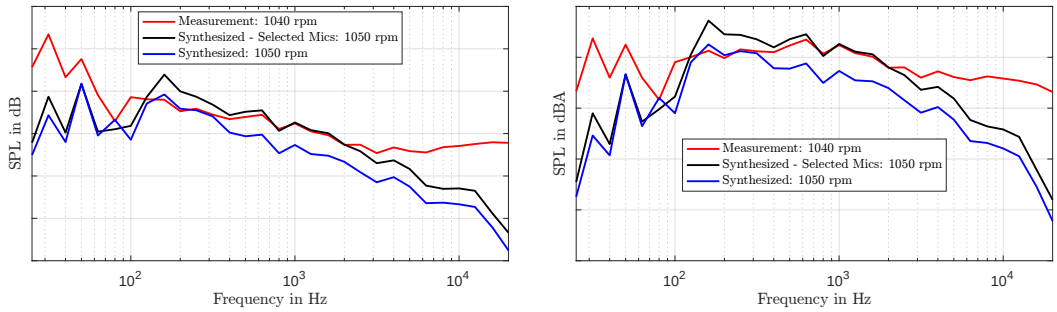
(a) Un-weighted. Increments of 5 dB per horizontal line.

(b) A-weighted. Increments of 10 dB per horizontal line.

Figure 6.9.: 1/3-octave band sound pressure levels at the façade. Truck driving 30 km/h, gear 9.

It can be seen that the big difference between using all microphones and only the selected ones is caused by the frequency content above 100 Hz. The peak at 50 Hz is unchanged between the two alternatives. When the signal is un-weighted, this peak is what dominating the SPL:s, as it is so much higher than anything else in the signal. But as mentioned, when the A-weighting filter is applied, this peak is attenuated and the mid frequency noise is contributing most to the sound pressure levels, which the all microphone case is lacking.

Figure 6.10 shows the sound pressure levels inside the apartment. One thing to notice is that while the levels for the computer model seem to be slightly below the measurements at 200 Hz - 1 kHz at the façade position, the levels are actually higher inside the apartment. This can also be seen in most of the cases in Appendix D. This implies that the estimated values of the reduction indices might be wrong.

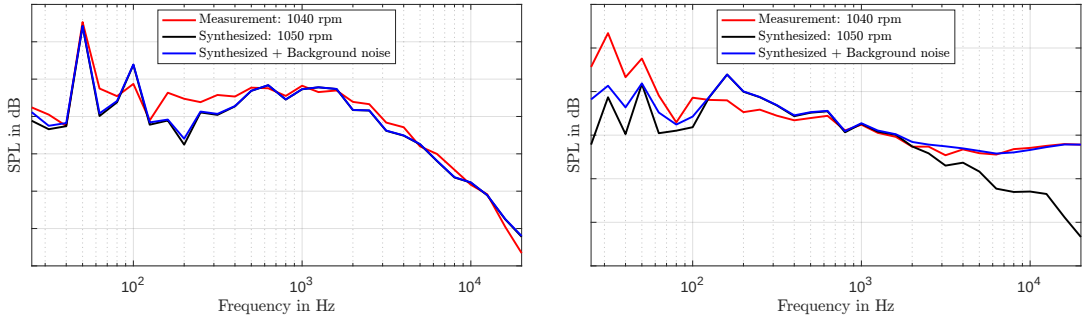


(a) Un-weighted. Increments of 20 dB per horizontal line.

(b) A-weighted. Increments of 10 dB per horizontal line.

Figure 6.10.: 1/3-octave band sound pressure levels indoors for a truck driving 30 km/h, gear 9.

As mentioned before, the background noise is affecting the measurements which is why it, also, is considered for the model. For the façade position, the noise from the truck is more prominent compared to outdoor background noise, which is why the background noise has almost no impact on the levels. But for the apartment, the influence of the background noise is bigger. This is clearly seen in frequency ranges 25 - 100 Hz and 2 k - 20 kHz.



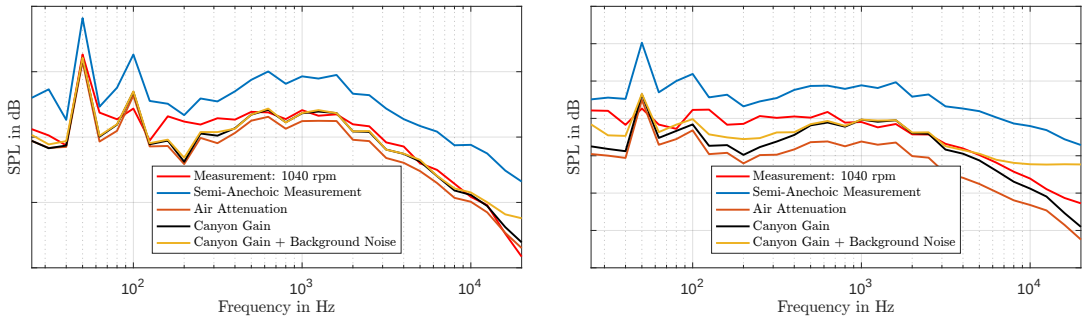
(a) At the façade. Increments of 10 dB per horizontal line.

(b) Indoors. Increments of 20 dB per horizontal line.

Figure 6.11.: Sound pressure levels including background noise. Truck driving 30 km/h, gear 9.

The difference at the very low frequencies, 25 Hz - 40 Hz, is also the explanation to why the un-weighted sound pressure levels in time domain, see Figure 6.10a differ so much. It seems that the computer model in general is lacking something for those frequencies. Possible explanations could be the missing rolling noise or the background noise at this instance in time, as the model only considers an averaged background noise. Another factor could be that the canyon gain factors are wrong in the low frequencies because of the assumptions in the coefficients below 125 Hz. The most likely reason is that the reduction index is estimated wrong, as the data is only calculated between 50 - 5000 Hz and the extended down to 25 Hz and up to 20 kHz.

Figure 6.12 shows the 1/3-octave band levels after applying some of the parameters considered in the model. The blue line is the sound at the engine, before applying any parameters. The orange line shows the sound after distance, Doppler effect and air attenuation is considered. The black and yellow line shows the sound after all parameters are considered with the yellow line also including background noise.



(a) Maximum position

(b) Around 61 meters left from the middle of the canyon

Figure 6.12.: Sound pressure levels at the façade, after application of different parameters, for a truck driving 30 km/h, gear 9. Increments of 20 dB per horizontal line.

It can be seen that the background noise has very low impact at the maximum position, in front of the apartment. However it is of significant importance when the truck is located further down in the street. The big difference between the measurement at Landsvägsgatan and the model around 130 Hz - 600 Hz can possibly be related to the measurements in the semi-anechoic chamber or because of rolling noise. Looking at Figure 13 in Appendix D the sound pressure levels are matching in this range. As this figure is showing the idling case, the truck is standing still and no rolling noise is influencing the signal. This shows indication towards that it is rolling noise contributing factor in this range.

6.2.3. Listening test

As mentioned before, a listening test is set up in order to evaluate the model further, but with focus on the sounds characteristics. The results are also used for further confirmation of the possible missing sources in the model.

The semantic differential test is evaluated with regards to three parameters. The main focus is realism since the goal of the computer model is to be as accurate as possible. However other parameters are still of interest for better understanding. The questions asked can be seen below.

- How realistic is the sound?
  - Unrealistic(1) to Realistic(9)
- How annoying is the sound?
  - Not annoying(1) to Very annoying(9)
- How loud is the sound?
  - Not loud(1) to Very loud(9)

A total of 8 sounds are evaluated; 4 from the measurements at Landsvägsgatan and 4 from the computer model. The driving conditions are the same as in Section 6.2: idling, 20 km/h gear 7, 20 km/h gear 8 and 30 km/h gear 9. Only the sounds at the outdoor position are evaluated. The length of the sounds are 5 seconds for the idling cases and between 10 - 12 seconds for the pass-by cases, depending on the driving speed. The length of the signals is based on the measurements at Landsvägsgatan, as they contained distortion and are fluctuating in driving speed, which means limited length. The measurements contain data of when the truck is passing by in front of the apartment, so the computer model sounds are cut in the same way. Since the measurement sounds are affected by background noise, the computer model also has background noise applied to each signal in order to make the conditions as close as possible.

Figure 6.13 shows the results from this sub-test. The table shows the averaged answers from the participants, where realism (the red bar) is of most importance. It can be seen that the measurements at Landsvägsgatan shows a slightly overall higher rate of realism.

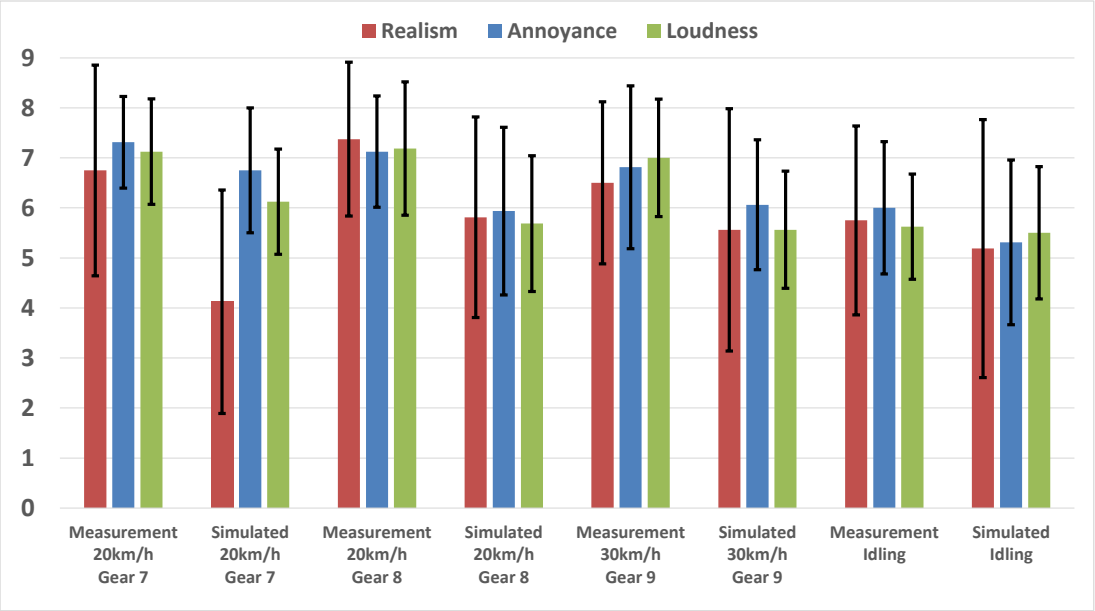


Figure 6.13.: Results of the semantic differential test. Answers averaged.

Focusing on the standard deviation, marked with the black bars, it can be seen that the listeners shows an overall better agreement when it comes to evaluating annoyance and loudness. When it comes to realism, the answers differ from listener to listener, especially for the simulated sounds. Realism is something very subjective and a contributing factor is obviously how closely the listener can related to truck sounds. Some of the listeners work closely with trucks, which affects the results. The measurements also contain specific sounds like background noise from the street and exhaust noise from the truck, which could make the listener biased towards the measurements. The biggest difference with regards to realism seems to be for the 20 km/h gear 7 driving case. This case also got the highest standard deviation.

For the idling case, the results are very similar between the measured and the simulated, which could be for multiple reasons. As mentioned in the sound pressure levels sections, one could be that rolling noise is not included when the truck is idling, i.e standing still, something the computer model does not consider. Also, since the truck is idling, the possibility of human error is less likely as the driver does not need to manually control the truck to keep a certain speed or rpm. Both reasons make the computer model more similar to the measurement as fewer parameters are affecting the sounds.

Also worth mentioning is the measurements done on the head dummy before the listening tests, see Table 5.1 in Section 5. They overall support the results with regards to loudness, as both Table 5.1 and Figure 6.13 indicates that the measurements at Landsvägsgatan are louder. It is possible that the listening test set-up is affecting the sounds, as the "real" levels are in most cases over 10 dB higher than the measured ones. This indicates that some frequencies are attenuated in the measurement set-up and could therefore affect the different parameters evaluated in the listening test.

The comparison test between the computer model and the measurements at Landsvägsgatan is evaluated with regards to realism. The listener has two sets of sounds, one from the computer model and one from the measurements, and the question posed is: "How do the two sounds compare with respect to realism?". The question is evaluated using a "slider" that the listener can drag from left to right. Dragging the slider all the way to the left side means that sound one is real while sound two is not and dragging it the right side means the opposite. If both sounds are similar with regards to realism the slider is to be kept in the middle. The listener can put the slider somewhere in between meaning that one sound is more real than the other, but both sounds are still real. Each set of sounds are evaluated twice, with the second time in reversed order in order to make the test more accurate and consider order specific answers.

Figure 6.14 shows the results from the "slider" comparison test. It shows the answers from each listener, where the crosses shows the original order where sound 1 is the measurement and sound 2 is the computer model. The diamonds shows the same but in reversed order.

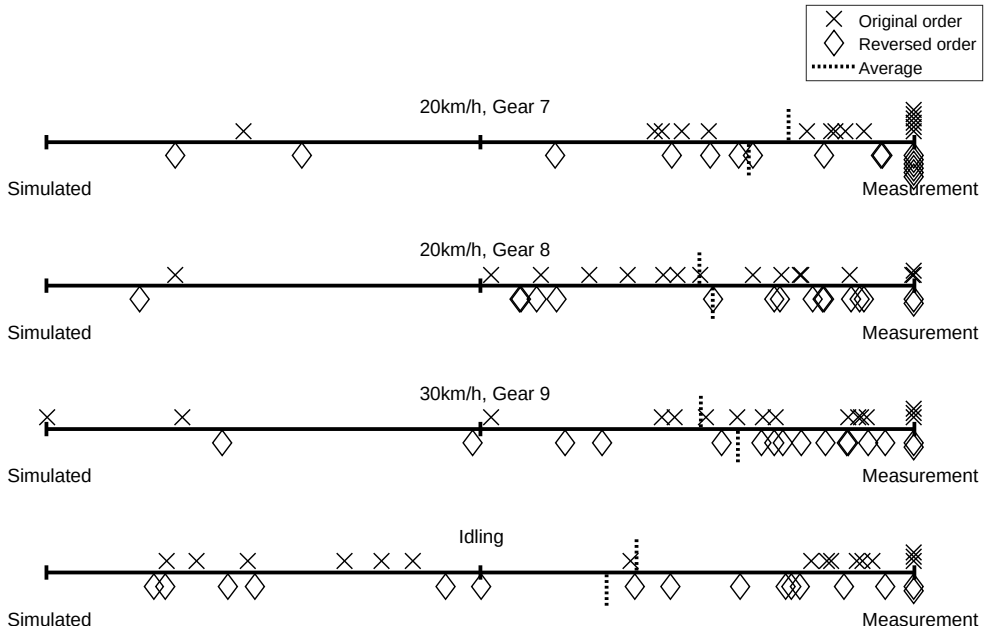


Figure 6.14.: Answer distribution of the comparison test.

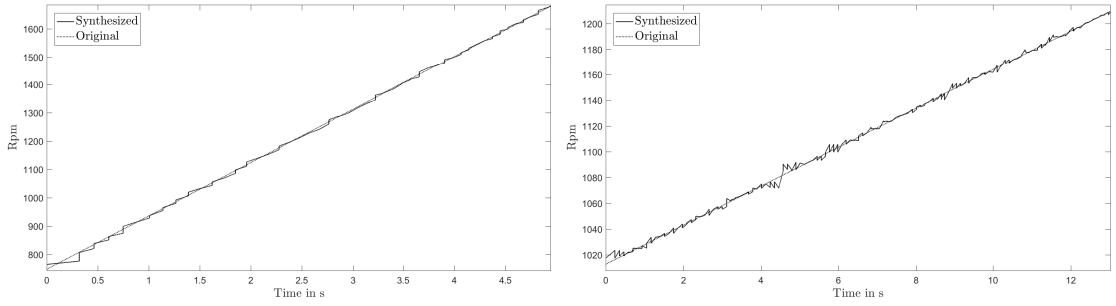
It can be seen here that idling case shows some support to the fact that they similar. Even though most listeners are biased towards the measurement, there are more people preferring the simulated sound in the idling case compared to the other driving cases. For the case with gear 7 and a driving speed of 20 km/h the simulated sound is perceived as not real at all by most listeners. This case also shows the lowest rate of realism in the semantic differential test. It could be that they misunderstood the question or that the simulated sound is missing too much sound characteristics to be perceived as a passing truck.

Based on the comments from listeners, a reason for choosing the measurement sounds is because it contains rolling noise and the ventilation/exhaust noise from the truck. Also the lack of low frequency noise is something the reduces the realism of the sounds. The results from the comparison test coincide with the results from the semantic test. The measured sounds are perceived slightly higher with regards to realism which is also supported in the comparison test. Comments about the simulated sounds containing too much engine noise can also be linked to the lower realism. This is most likely caused by using the "selected microphones" case, where only the the front microphones located next to the engine were considered.

6.3. Part 2 - Granular synthesis

The product of the acceleration synthesis and a general discussion over the implemented method are featured here. Results from constant speed synthesis are not shown, their rpm profiles are indistinguishable from the desired case and very similar to measurements in pressure. Also not shown are acceleration cases made to replicate the measurement the grain database was built on; that simply recreates the signal. Interesting to note is that, while real-time synthesis is not a goal with this thesis, the computation time with the implemented solution is short enough to eventually be used for producing sounds continuously.

Figure 6.15 compares the input desired rpm curve to the one achieved by acceleration synthesis. That is, the solid black curves show the rpm profile of the acceleration grains used and the black dotted curves show the input desired case. The vertical black lines shows the jumps in rpm between two consecutive grains, a larger/longer vertical line means a larger, more noticeable discrepancy.



(a) Acceleration on 9th gear (b) Acceleration on 12th gear

Figure 6.15.: Comparison of desired case ("original") and synthesized signal.

For Figure 6.15a the underlying grain database is created from a measurement of a slower, less steep acceleration. In this case the grains start slightly higher and end slightly lower in rpm than desired, but the error is no more than 10 rpm on either side. Focusing instead on Figure 6.15b, where the underlying database is based on a faster acceleration measurement, shows larger problems. The differences at the start and end of grains are overall comparable to the faster case, but extending the duration of the acceleration causes grains to repeat. An example of this can be seen as a sawtooth pattern 6 seconds into the signal, where a grain is repeated three times.

Together with other cases shown in Appendix F the results show the implemented method can synthesise accelerations that are faster than their underlying measurements without large error. However, if the gradient is increased too much the grains start/end will eventually be too far off. This could hypothetically be solved by only using portions of some grains to reduce their duration and allow a faster step up in rpm. In this context that would entail dividing up the grains into six separate pieces, one for each cylinder ignition (wavelet). Then it would be possible to use the first wavelet from one grain, the second from another and so on.

It is also important to note that the implemented method makes no modifications to the extracted grains. As an example, pitch shifting the grains as needed could possibly increase the rpm range in which each grain is useful. The drawback of pitch shifting is that it would speed up the entirety of the grain. That is, each cylinder ignition would be sped up as well, something that does not actually happen with increasing rpm. Instead, it is reducing the distance between each firing that is of interest, which suggests that dividing up the grains has the best potential.

A further improvement in the quality of the synthesised sounds could be gained from implementing stereo playback by making use of the fact that the cab recordings are made with two microphones.

6.3.1. Listening test

The listening test for part 2 is the primary source to evaluate the granular synthesis method. All the sounds evaluated are created from the driving cases mentioned in Section 4.3. A semantic differential test could also have been beneficial in order to evaluate each sound separately, but is not included as the direct pairwise comparison is deemed more useful for initial evaluation.
The direct pairwise comparison test is only evaluated with regards to realism. The goal is to see if the granular syntheses can be a replacement to the current method used in the driving simulator. The listener has two sets of sounds, one from the granular synthesis and one from the current method, and is to decide which sound is more real. Only the option of choosing one sound over the other is given. A total of 13 sounds from each method are evaluated: 1 idle case, 4 constant speed cases and 8 acceleration cases, also mentioned in Section 4.3. The length of each signal vary depending on the driving case, where the idle case and the constant speed cases are cut to

5 seconds long and the acceleration cases vary between ~5 - 13 seconds. In the same way as in the comparison test, the sets of sounds are played twice, the second time in reversed order.

In Figure 6.16 the results for the driving simulator sounds are shown. The results show how many listeners prefer the new sounds made by the granular synthesis over the old sounds. The blue and green bars are the answers from when the sounds are put in the original and reversed order. The red bar shows the minimum of how many listeners change their answers when the sounds are played in the reversed order.

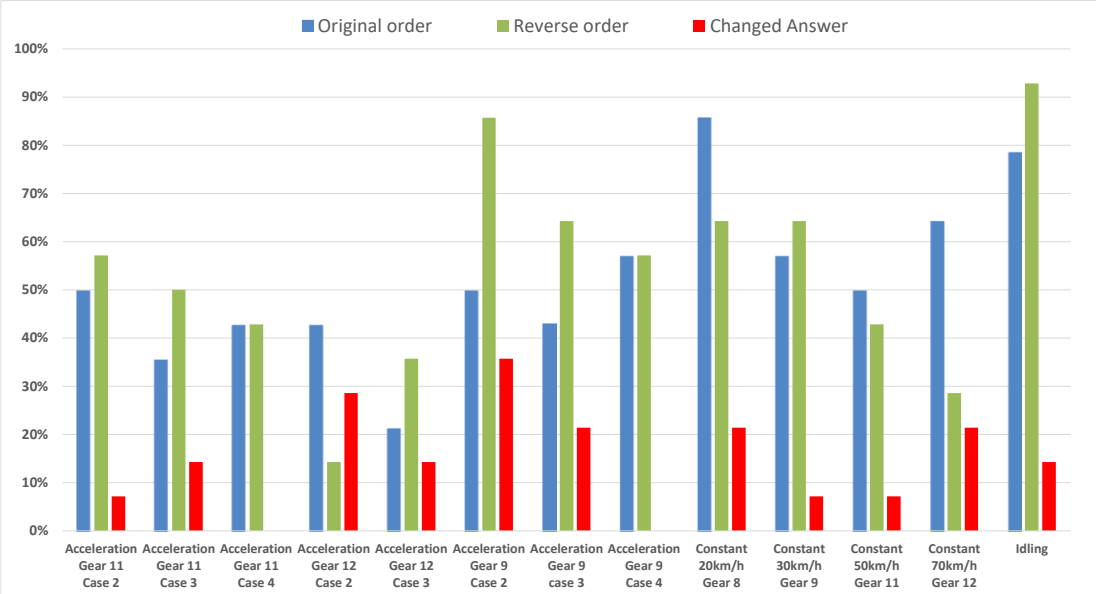


Figure 6.16.: Results from the direct pairwise comparison test. The number of people preferring the new sets of sounds over the old method.

The slower constant speed cases created by granular synthesis are overall preferred by most listeners. That the constant speed cases show good results is already known from work by Seward (2014), where the listeners, in 9 out of 10 cases, could not tell the difference between the synthesized and real measurements. The same can also be said for the idling case. At higher speeds the answers seem to be around 50/50. Presenting the results in this way only shows the minimum of listeners that changed their answers. An answer close to 50% could indicate that the sounds are of similar rate of realism, which is why this way of analysing the results is very limited.

For the acceleration cases the listeners prefer the sounds created with the current method overall. It can also be seen that acceleration cases that are accelerating faster, i.e. gear 11 case 2 and gear 9 case 2, show better results compared to the slower acceleration cases, for example gear 12 case 3. These cases change rpm faster over time compared to the slow acceleration with gear 12,

which means that the grains do not have to be repeated and no abrupt jumps between rpm have to be made. This can also be seen in Figure 6.15 in Section 6.3 where the slower case indicates precisely that. If the grain database was bigger, i.e. containing multiple acceleration cases, both slower and faster, there would be more to choose from and a better match would be possible. This would most likely help to eliminate the abrupt jumps, reduce grain repeats and make the signal smoother. Another possible solution already mentioned above is to use portions of the grains (wavelets) to eliminate the abrupt jumps in rpm.

A factor that could contribute to the difference and also make the comparison less valid is the sound source. The granular synthesis sounds are made from measurements made on the "FM-490" truck, inside the drivers cabin, while the existing sounds are made from a previous measurement of unknown origin (most likely outside the truck). This means that while the main focus for the listening test is to evaluate the model/method for creating sounds for a truck simulator, it is highly likely that sound origin also influenced the results.

7. Conclusions

The auralization model shows good results for the driving speeds considered. According to sources, rolling noise is most prominent in speeds 50 km/h or above, which will most likely make the model less valid at those speeds. It is also a likely factor for the difference that exists in the lower speeds. In order to evaluate the model further and see the effects of rolling noise, higher velocities should be tested. This will decide the speed limitations of the model. New measurements at Landsvägsgatan and/or other streets could be used to evaluate the model further as well. The new measurements should minimize the human error in trying to keep the truck at a certain speed and avoiding fluctuations. Background noise is a significant factor to consider when looking at indoor sounds to verify the model. Also the measurements indoors indicates wrong estimation of the lower frequencies' reduction indices used in the model. A second estimation with a non-furnished room, to achieve something closer to a diffuse field, would be useful in further investigations.

Microphone weighting is an important part of the model which affects sound pressure levels but also the sound files. This is why the Shepard's method with the weighting exponent is used in a way so that the difference between the microphones is minimized. A sound power measurement should be performed for the microphone set-up in order to get a better understanding of the difference in frequency content. A possible change, to increase the realism of the sound, would be to use a time varying exponent p in the Shepard's equation, where p assumes lower values closer to the center of the canyon and higher values further away. In order to do so, the mid/high frequency content needs to be increased in the microphones located on the sides and back of the truck.

Because of its importance to the environmental gain factors, reworking the scattering to be more realistic and evenly distributed is advisable. At the same time, it has to be remembered that the scattering parameters were partly used to adjust the gain factors to match the Landsvägsgatan measurement results. By changing how scattering is handled, the overall gain from the model might be lowered. As long as rolling noise is not accounted for, some compensation will be required.

Low frequency noise is something that is lacking in the model and is also mentioned in the comments from the listening test. This reduces the realism of the sounds. The listening test is to some degree biased towards the measurements as they contain ventilation/exhaust noise from the truck. This made the listening test less useful for evaluation of the model in this thesis. If new measurements were to be made, this type of noise should be minimized if listening tests are to be used for further evaluation.

The granular synthesis shows promising results for constant speed but less so for acceleration. A bigger grain database is needed in order to have more options and cover most possible cases and better fit the expected rpm values. Fewer repeats of grains and fewer sudden increase/decrease in rpm will most likely result in a higher rate of perceived realism.

Two potential ways of improving the quality of the acceleration synthesis are the modification of grains (pitch shifting) and increasing the grains' flexibility by separating each one into their respective six wavelets. In the latter case, the issue with the duration of the acceleration grain overshooting the desired rpm profile could be counteracted. This would also mean that the selection algorithm wouldn't have to approximate the ending rpm as far away in time, leading to a smaller error in the selection process.

Bibliography

Bass, H.E., H.J. Bauer, and L.B. Evans. 1972. "Atmospheric absorption of Sound: Analytical Expressions." *The Journal of the Acoustical Society of America* (February 15).

Bergman, P., A. Pieringer, J. Forssén, and P. Höstmad. 2015. "Perceptual validation of auralized heavy-duty vehicles." *EuroNoise.*

Cox, T. J., and P. D'Antonio. 2009. *Acoustic Absorbers and Diffusers.* Theory, design and application. 2nd edition. Taylor & Francis. ISBN: 0-203-89305-0.

Forssén, J. 2014. *Suggested facade cases for study of sound insulation considering wall, window and air intake.* Department of Civil and Environmental Engineering, Chalmers University of Technology, April.

Forssén, J., P. Andersson, P. Bergman, K. Fredriksson, and P. Zimmerman. 2013. "Auralization of truck engine sound – preliminary results using a granular approach."

Franke, R., and G Nielson. 1980. "Smooth interpolation of large sets of scattered data." *International journal for numerical methods in engineering* 15:1691–1704.

Fredriksson, K. 2015. *Slutrapport för projekt 2011-0805 Tystare transportfordon för effektivare distribution.* FFI.

Höstmad, P., P. Bergman, and J. Forssén. 2015. *Towards a low noise truck specification* S15-04. Chalmers University of Technology.

Ismail, M.R., and D.J. Oldham. 2004. "A scale model investigation of sound reflection from building façades." *Applied Acoustics,* no. 66.

Jagla, J., J. Maillard, and N. Martin. 2012. "Sample-based engine noise synthesis using an enhanced pitch-synchronous overlap-and-add method." *Acoustical Society of America* 132 (5).

Kang, J. 2007. *Urban Sound Environment.* Taylor & Francis. ISBN: 978–0–203–00478–4.

Keulen, IR.W. van, and Dr.ir.M Duskov. 2005. *Inventory study of basic knowledge on tyre/road noise.* October.

Kleiner, M. 2012. *Acoustics and audio technology.* 3rd. J. Ross Publishing. ISBN: 978-1-60427-052-5.

Kuttruff, H. 2009. *Room Acoustics, Fifth Edition.* Taylor / Francis, June 2. ISBN: 978-0-203-87637-4.

Long, M. 2005. *Architectural Acoustics.* Elsevier Academic Press, December 23. ISBN: 978-0-12-455551-8.

Moulines, E., and F. Charpentier. 1990. "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones." *Speech communication* (9). doi:10.1016/0167-6393(90)90021-Z.

Mulgrew, B., P. Grant, and J. Thompson. 2003. *Digital Signal Processing: Concepts and Applications - Second Edition.* Palgrave Macmillan.

Rao Yarlagadda, R. K. 2010. *Analog and Digital Signals and Systems.* Springer US, August 5. ISBN: 978-1-4419-0034-0. doi:10.1007/978-1-4419-0034-0.

Rennie, R., and J. Law. 2015. *A Dictionary of Physics (7 ed.)*

Sandberg, U. 2001. "Tyre/road noise - Myths and realities." In *Inter-noise.* VTI särtryck 345, August 27.

Seward, B. 2014. "Auralization of a Heavy-Duty Truck with a Hybrid Engine using a Granular Approach." Master thesis, Chalmers University of Technology.

Smith, J.O. 2007. "Introduction to Digital Filters with Audio Applications." Online Book. Accessed February 17, 2016. http://ccrma.stanford.edu/~jos/filters/.

———. 2010. "Physical Audio Signal Processing." Online Book. Accessed April 28, 2016. http://ccrma.stanford.edu/~jos/pasp/.

Smith, S.W. 1999. *The Scientist and Engineer's Guide to Digital Signal Processing - Second Edition.* California Technical Publishing.

Smith, W. J. 2000. *Modern Optical Engineering.* 3rd ed. McGraw-Hill. ISBN: 0-07-136360-2.

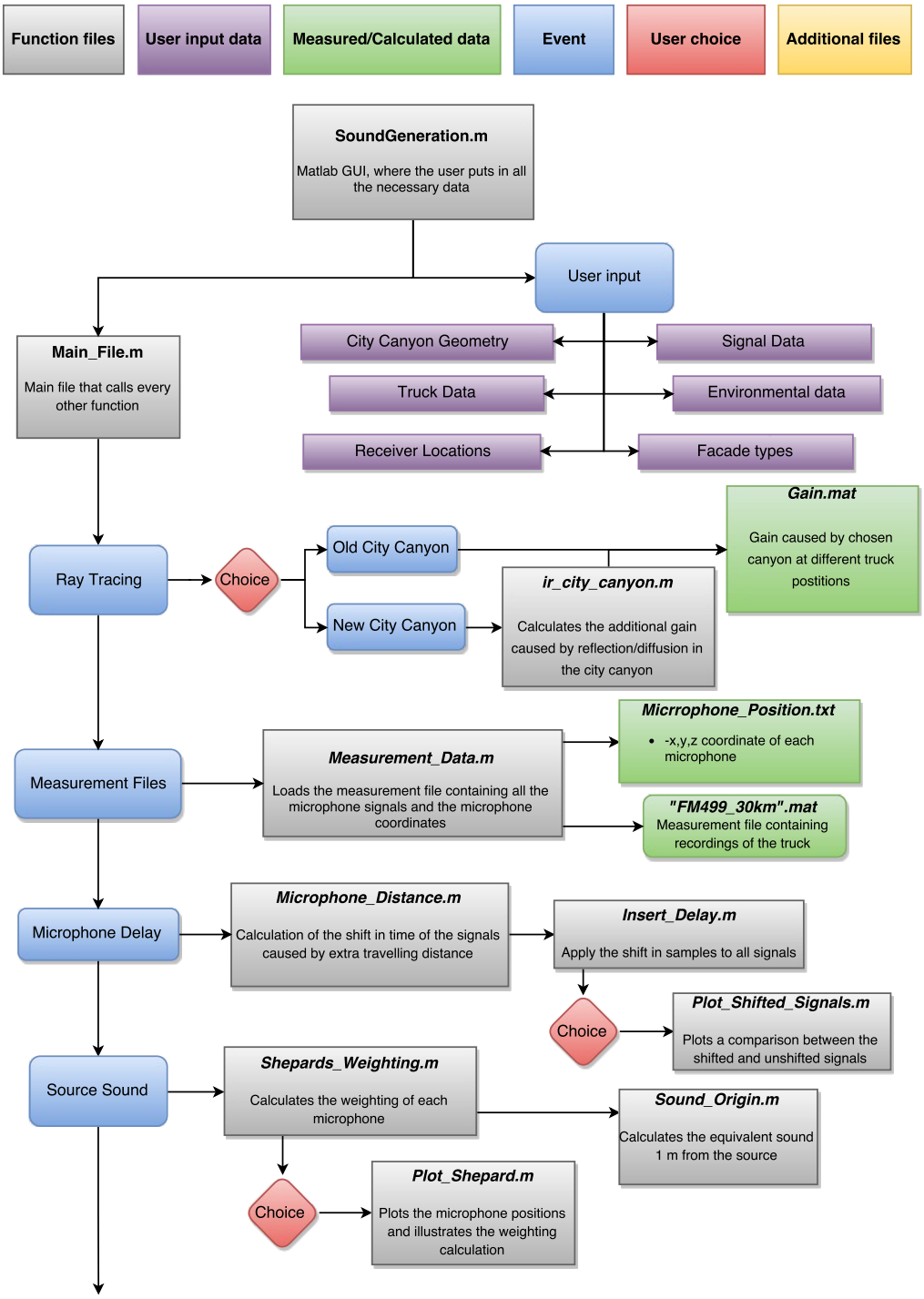
Thacker, W.I., J. Zhang, L.T. Watson, J.B. Birch, M.A. Iyer, and M.W. Berry. 2009. *Algorithm XXX: SHEPPACK: Modified Shepard Algorithm for Interpolation of Scattered Multivariate Data.* Technical report. Department of Computer Science, Virginia Polytechnic Institute & State University.

Vorländer, M. 2008. *Auralization.* ISBN: 978-3-540-48829-3.

Zölzer, U. 2011. *DAFX: Digital Audio Effects.* Second. John Wiley & Sons, Ltd. ISBN: 978-0-470-66599-2.

Appendices

A. Flow Chart



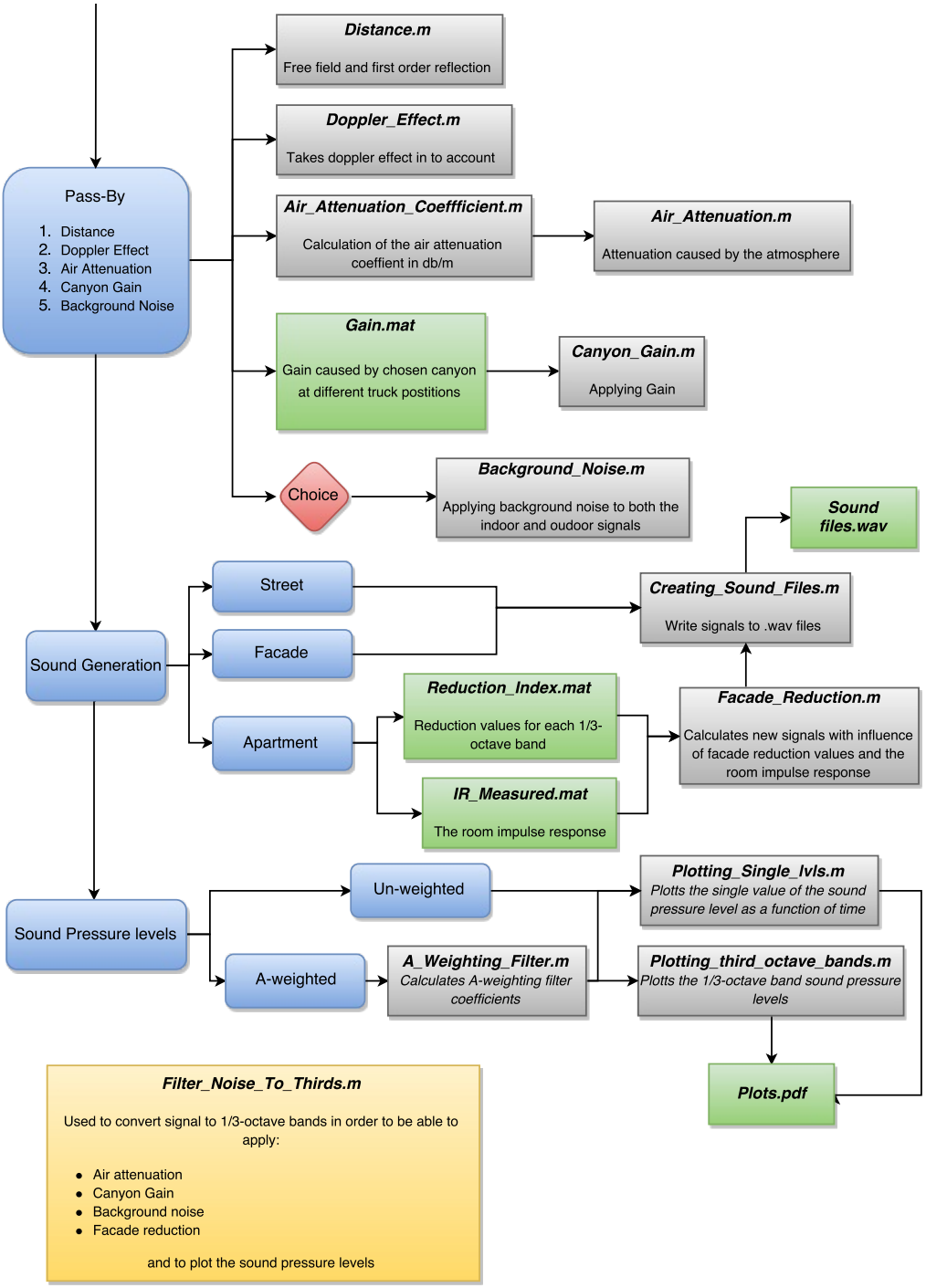


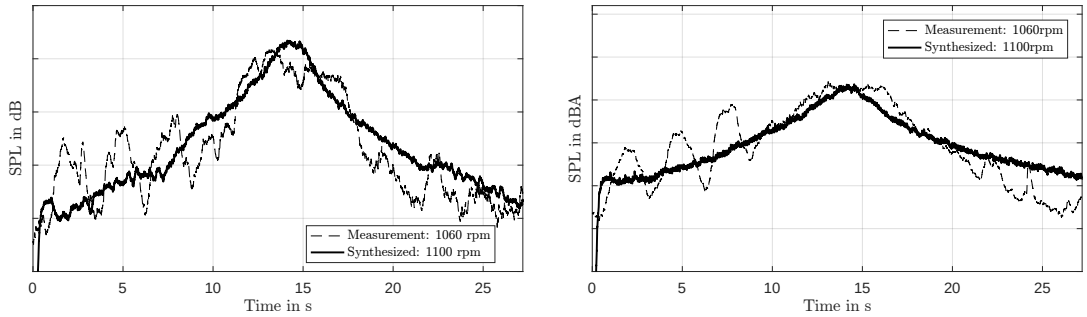
Figure 1.: Flow chart of the Matlab computer model

B. Microphone Coordinates

Table 1.: Microphone Coordinates

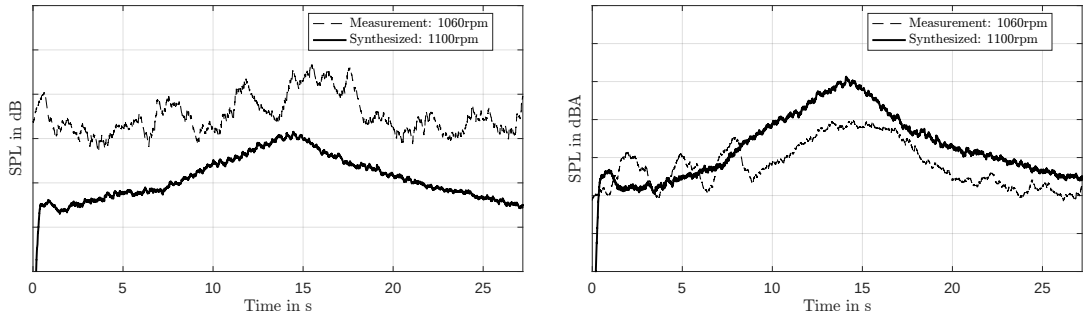
Microphone	x (m)	y (m)	z (m)
B1	-4.9	0	2.925
B2	-4.9	3.3	5.85
B3	-4.9	-3.3	5.85
B4	-4.9	3.3	1.2
B5	-4.9	0	1.2
B6	-4.9	-3.3	1.2
F1	8.1	0	2.925
F2	8.1	3.3	5.85
F3	8.1	-3.3	5.85
F4	8.1	3.3	1.2
F5	8.1	0	1.2
F6	8.1	-3.3	1.2
H1	-1.65	-3.3	2.925
H2	4.85	-3.3	2.925
H3	1.6	-3.3	5.85
H4	1.6	-3.3	2.925
U1	4.85	0	5.85
U2	-1.65	0	5.85
V1	4.85	3.3	2.925
V2	-1.65	3.3	2.925
V3	1.6	3.3	5.85
V4	1.6	3.3	2.925
FL1	8.1	0.696	0
FL2	8.1	2.339	0
FL3	6.805	3.3	0
FL4	5.256	3.3	0
FL5	3.944	3.3	0
FL6	2.1	3.3	0
FL7	-0.339	3.3	0
FL8	-4.9	1.89	0
FL9	-4.9	-1.89	0
FL10	-0.339	-3.3	0
FL11	2.1	-3.3	0
FL12	3.944	-3.3	0
FL13	5.256	-3.3	0
FL14	6.805	-3.3	0
FL15	8.1	-2.339	0
FL16	8.1	-0.696	0

C. Single SPL: different velocities



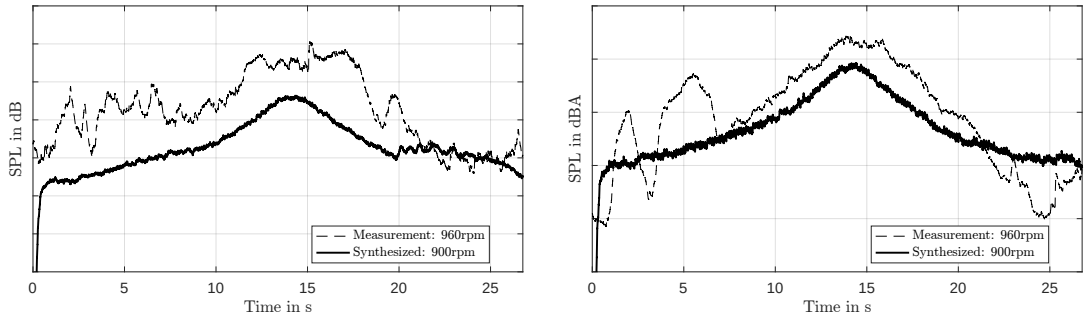
(a) Un-weighted (b) A-weighted

Figure 2.: Sound pressure levels at the façade for a truck driving 20km/h, gear 7. Increments of 5 dB per horizontal line.



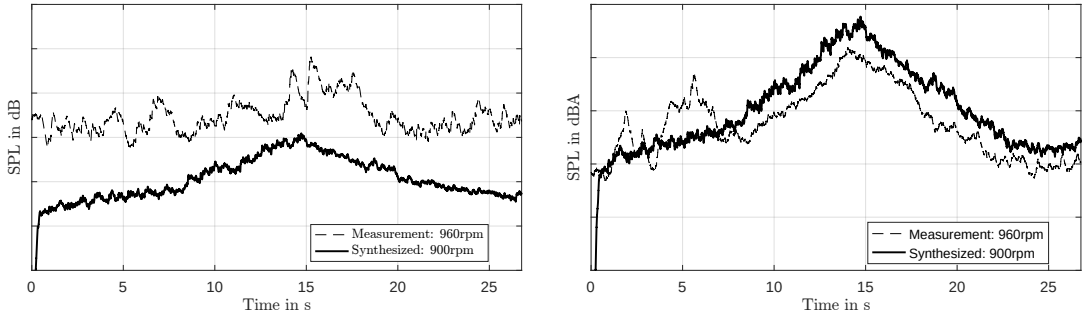
(a) Un-weighted. Increments of 10 dB per horizontal line. (b) A-weighted. Increments of 5 dB per horizontal line.

Figure 3.: Sound pressure levels indoors for a truck driving 20km/h, gear 7.



(a) Un-weighted (b) A-weighted

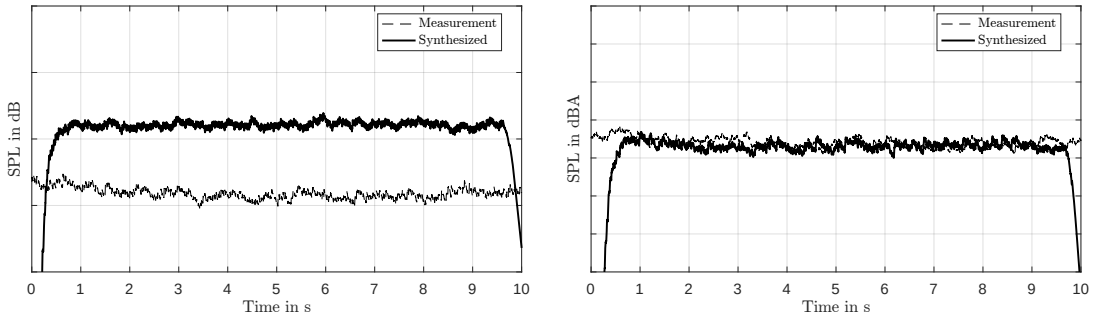
Figure 4.: Sound pressure levels at the façade for a truck driving 20km/h, gear 8. Increments of 5 dB per horizontal line.



(a) Un-weighted. Increments of 10 dB per horizontal line.

(b) A-weighted. Increments of 5 dB per horizontal line.

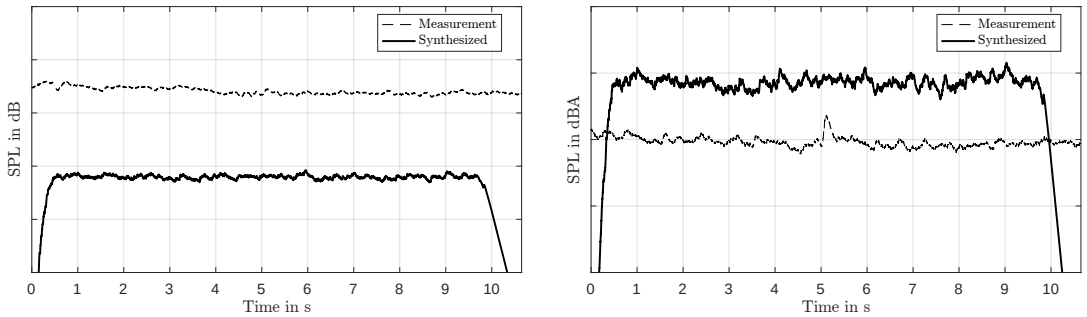
Figure 5.: Sound pressure levels indoors for a truck driving 20km/h, gear 8.



(a) Un-weighted. Increments of 5 dB per horizontal line.

(b) A-weighted. Increments of 2 dB per horizontal line.

Figure 6.: Sound pressure levels at the façade for a truck in Idle mode.

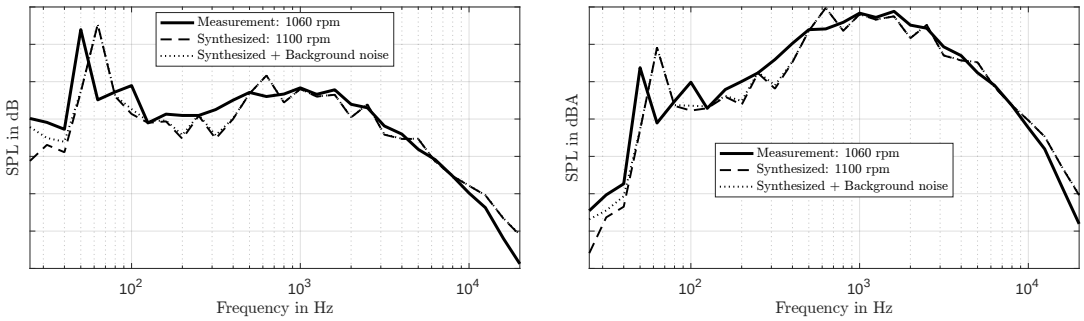


(a) Un-weighted. Increments of 10 dB per horizontal line.

(b) A-weighted. Increments of 5 dB per horizontal line.

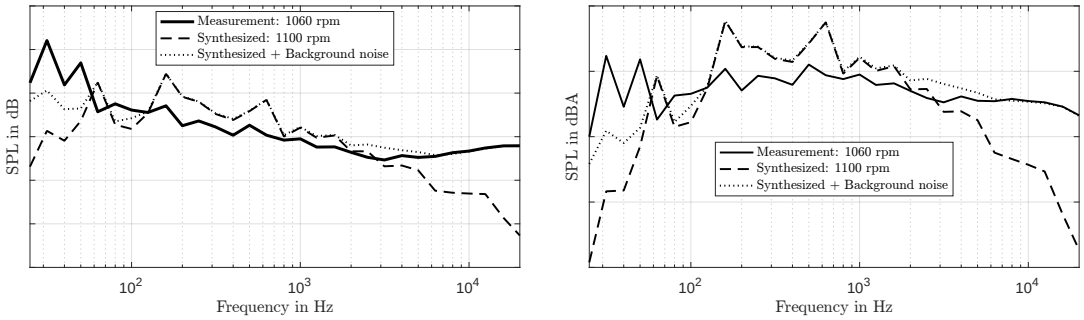
Figure 7.: Sound pressure levels indoors for a truck in Idle mode.

D. 1/3-octave band SPL: different velocities



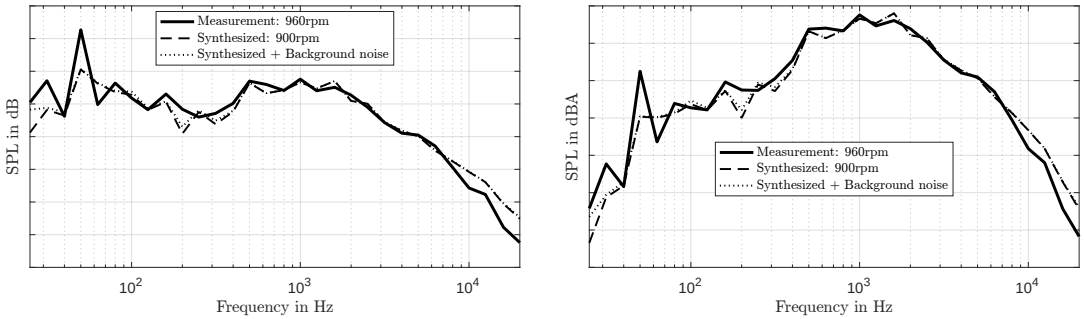
(a) Un-weighted. (b) A-weighted.

Figure 8.: 1/3-octave band sound pressure levels at the façade for a truck driving 20km/h, gear 7. Increments of 5 dB per horizontal line.



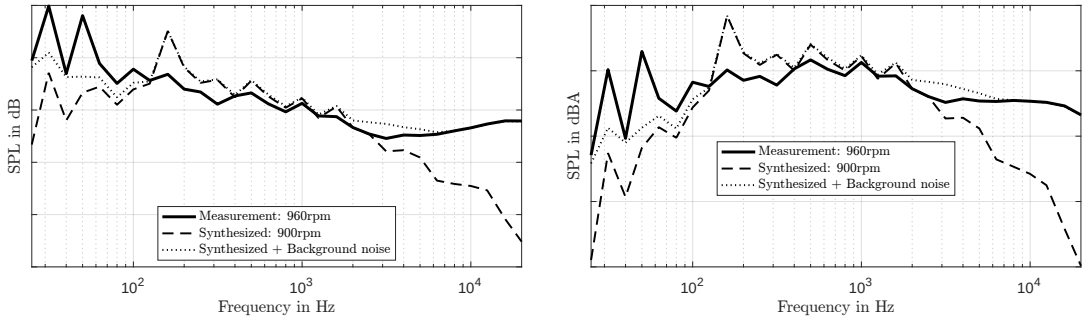
(a) Un-weighted (b) A-weighted

Figure 9.: 1/3-octave band sound pressure levels indoors for a truck driving 20km/h, gear 7. Increments of 20 dB per horizontal line.



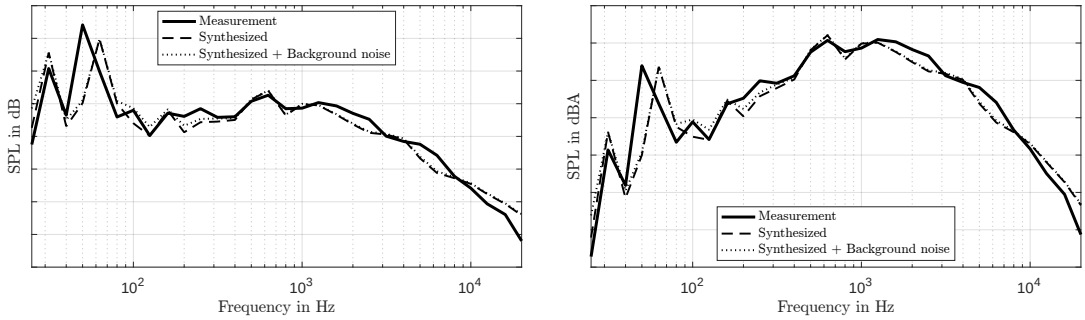
(a) Un-weighted (b) A-weighted

Figure 10.: 1/3-octave band sound pressure levels at the façade for a truck driving 20km/h, gear 8. Increments of 10 dB per horizontal line.



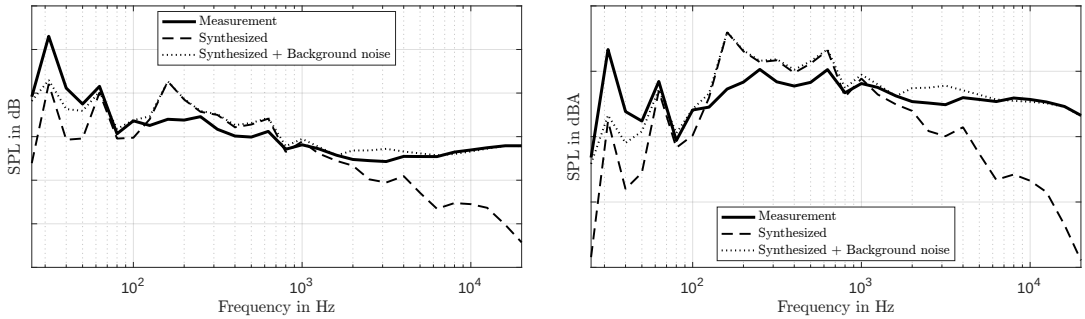
(a) Un-weighted (b) A-weighted

Figure 11.: 1/3-octave band sound pressure levels indoors for a truck driving 20km/h, gear 8. Increments of 20 dB per horizontal line.



(a) Un-weighted (b) A-weighted

Figure 12.: 1/3-octave band sound pressure levels at the façade for an idling truck. Increments of 10 dB per horizontal line.



(a) Un-weighted (b) A-weighted

Figure 13.: 1/3-octave band sound pressure levels indoors for an idling truck. Increments of 20 dB per horizontal line.

E. Listening test GUI

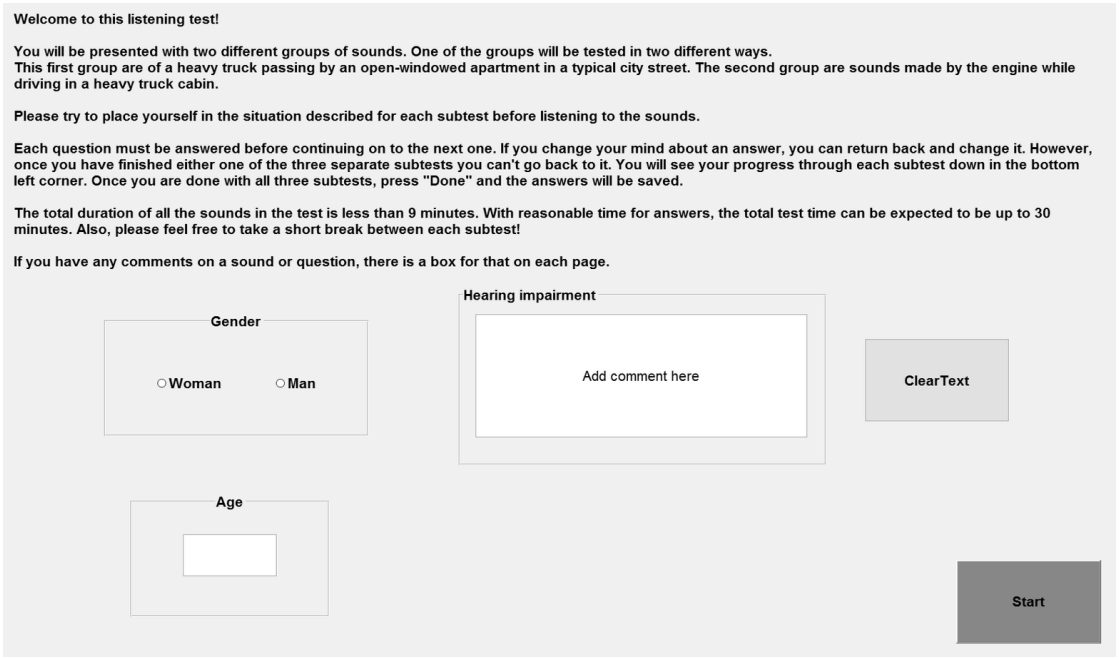


Figure 14.: Listening test GUI

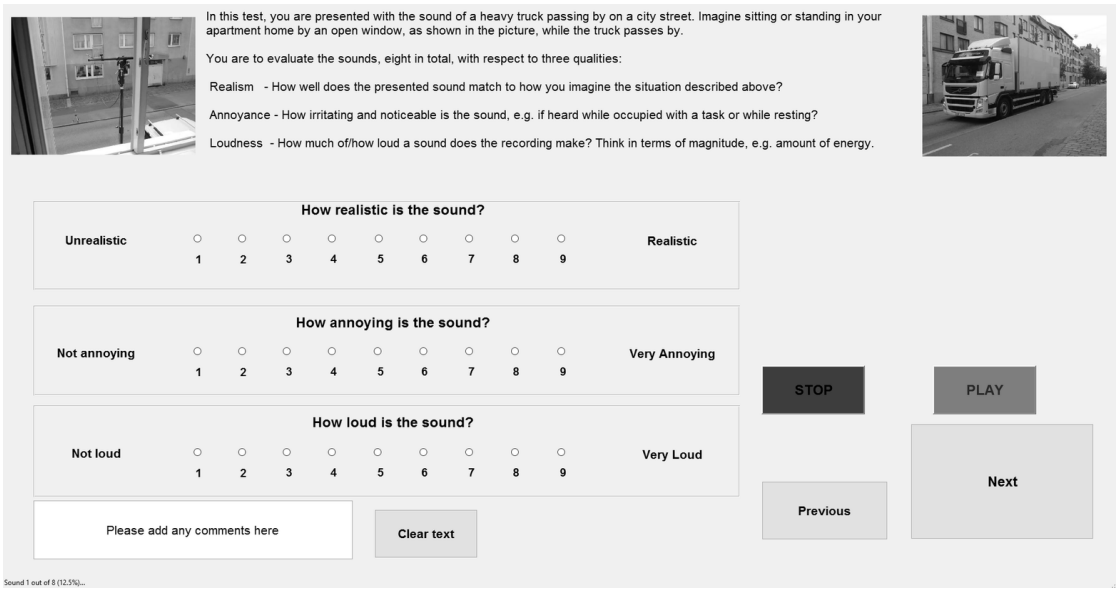


Figure 15.: Semantic test GUI

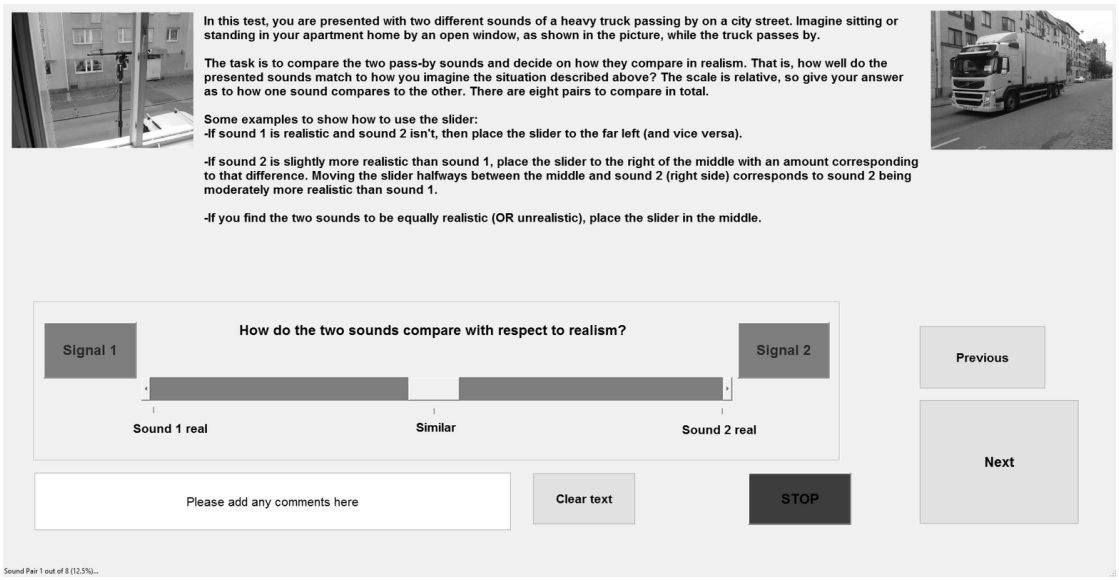


Figure 16.: Slider comparison GUI

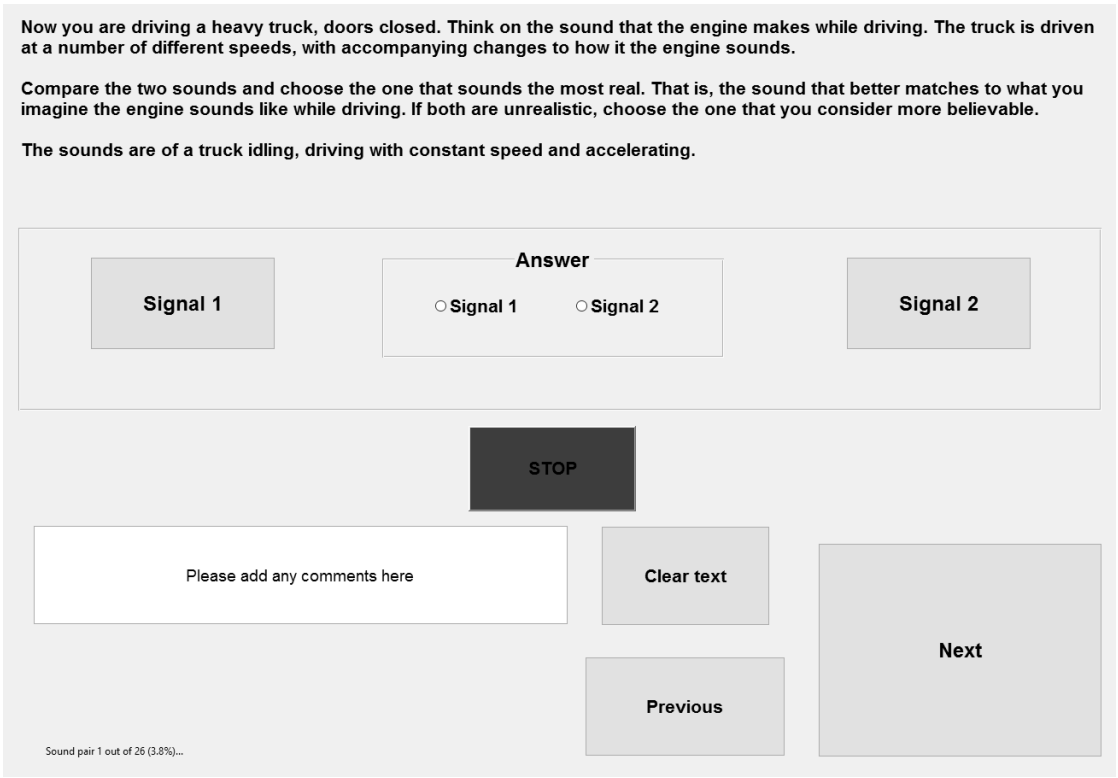
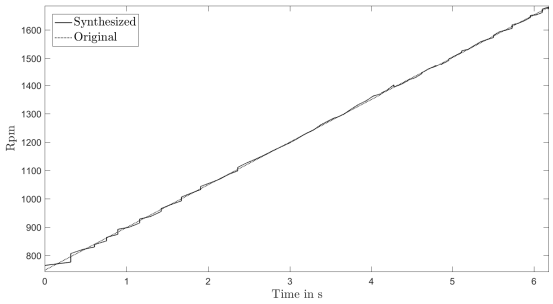
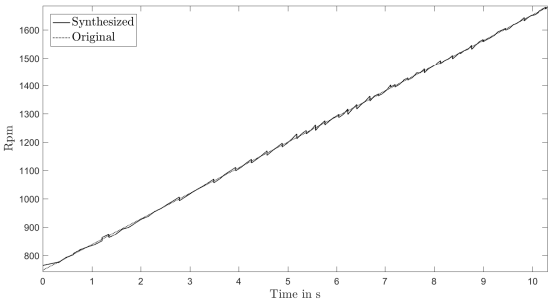


Figure 17.: Pair-wise comparison GUI

F. RPM curve comparisons - synthesised acceleration cases

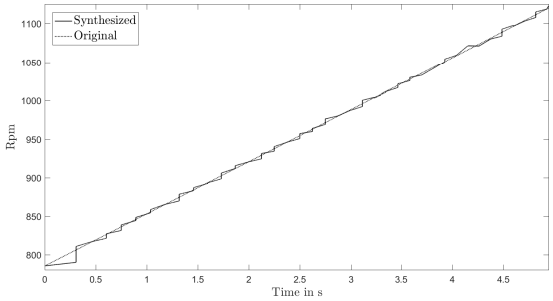


(a) Acceleration on 9th gear. Faster than underlying measurement.

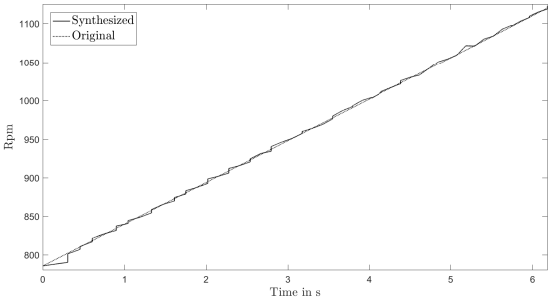


(b) Acceleration on 9th gear. Slower than underlying measurement.

Figure 18.: Comparison of desired case ("original") and synthesized signal.

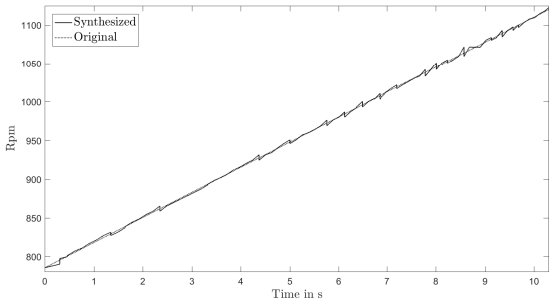


(a) Acceleration on 11th gear. Faster than underlying measurement.

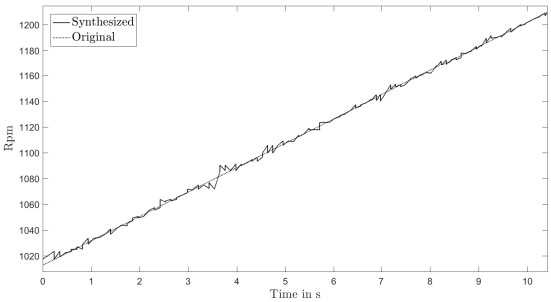


(b) Acceleration on 11th gear. Faster than underlying measurement.

Figure 19.: Comparison of desired case ("original") and synthesized signal.



(a) Acceleration on 11th gear. Faster than underlying measurement.



(b) Acceleration on 12th gear. Slower than underlying measurement.

Figure 20.: Comparison of desired case ("original") and synthesized signal.