CHALMERS

Auralization of a Heavy-Duty Truck with a Hybrid Engine using a Granular Approach

Master’s Thesis in the Master’s Program in Sound and Vibration

BRETT J. SEWARD

Department of Civil and Environmental Engineering
Division of Applied Acoustics
Vibroacoustics Group
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2014
Master’s Thesis 2014:149

Auralization of a Heavy-Duty Truck with a Hybrid Engine using a Granular Approach

Master's Thesis 2014:149

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 2014

Auralization of a Heavy-Duty Truck with a Hybrid Engine using a Granular Approach
Master's Thesis in the Master's program of Sound and Vibration
Brett Seward
Department of Civil and Environmental Engineering
Division of Applied Acoustics
Chalmers University of Technology

Abstract

This thesis details the development of a granular source model that was used to auralize various driving conditions for a hybrid heavy-duty truck, which is a vehicle that contains both an internal combustion engine and an electric motor. The granular approach is based on the storage and subsequent synthesis of short time pieces of recorded pressure signals of the hybrid truck. The method was applied using recordings provided by the Volvo Group for both constant vehicle speeds and accelerations. Granular synthesis was also used in this project to auralize sounds to accompany a video animation of a hybrid waste collection vehicle driving on a city street in Gothenburg, Sweden.

The sound quality of the synthesized sounds was evaluated using a same-different discrimination listening test for constant speeds and a listening test consisting of four semantic differential scales for accelerations. Listeners were unable to differentiate between recorded and synthesized sounds for eight of ten pairs of electric motor constant speed sounds for the hybrid truck. Synthesized acceleration sounds were evaluated using a repeated measures analysis of variance. The listeners found the perceived realism of the combustion engine's synthesized accelerations to be closer to a set of recorded reference sounds, as compared to the electric motor's synthesized accelerations which had a lower perceived realism relative to reference recordings.

Keywords: Source Modeling, Granular Synthesis, Overlap-and-Add (OLA), Pitch Synchronous Overlap-and-Add (PSOLA), Auralization, Hybrid Vehicle, Heavy-Duty Truck, Waste Collection Vehicle

Contents

Nomenclature

ANOVA	Analysis of variance
OLA	Overlap-and-add
FFI	*Fordonsstrategisk Forskning & Innovation* (Traffic Strategy, Research & Innovation)
PSOLA	Pitch synchronous overlap-and-add
RCV	Refuse collection vehicle

Acknowledgements

First, I would like to thank the U.S. Friends of Chalmers and the U.S.-Sweden Fulbright Commission. This once in a lifetime opportunity would not have been possible without the generous financial support of these two institutions. Additionally, thank you to VINNOVA, Renova AB, the Volvo Group and the city of Gothenburg for their funding and support of the SEND SMART project.

My advisor Jens Forssén deserves recognition for his guidance and promptness in answering my incessant midnight e-mails throughout the duration of my thesis. Thanks also to Patrik Andersson for his continued advice and feedback regarding my project work. I would like to thank Krister Fredriksson, the master of measurements, for his cooperation and ability to remain patient with me when I requested more data yet again.

Thanks to Jackie Forzelius for working collaboratively with me to create the Landsvägsgatan videos and allowing me to use her impressive animations. A huge thanks to Penny Bergman for her great assistance with facilitating my listening test and also to Astrid Pieringer and Alice Hoffmann for their work in the SEND SMART project. Thank you to Börje Wijk, Gunilla Skog and Wolfgang Kropp for their general willingness to lend a hand and their many contributions to the Division of Applied Acoustics.

Mendel Kleiner was the first person to help make Chalmers a reality for me, and I would like to thank him for his counsel and support. My classmates – now friends – have also helped to make this a much more fulfilling experience over the past two years. I will miss spending time with them and enjoying the idiosyncrasies that defined the character of our class.

Finally, my friends and family back in Indiana deserve my full gratitude for encouraging me to pursue this opportunity and extending support (and more importantly care packages) to me from afar, especially when my finishing date remained “just one more month” for about three months in a row.

1. Introduction

1.1. Project Background

One of the main objectives in the VINNOVA program "FFI Transporteffektivitet" (Traffic Strategy, Research & Innovation's collaborative project Transport Efficiency) is an investigation of quieter heavy-duty trucks that allow for more efficient nighttime operation. This includes the noise associated with a garbage truck that drives and collects waste in an urban setting. The conventional waste collection solution is restricted to use only during daytime hours, mainly due to the high noise levels emitted from the heavy-duty truck's internal combustion engine. These relatively high noise levels do not typically comply with urban nighttime sound emission regulations and therefore have the potential to negatively impact sleep quality and overall health if these vehicles are driven during the night.

To address the concern of high noise levels and to analyse the potential for earlier waste collection hours, a more silent concept was compared to a conventional one in the SEND SMART project. For the more silent concept, a hybrid refuse collection vehicle (RCV) with electrical bodywork was used in combination with noise abatement measures on waste bin handling. The aim of the project was to determine if the silent concept was sufficient for early morning operation. This work is significant because a change to earlier hours of operation has the potential to reduce congestion and improve traffic safety on busy urban streets during the day, while also decreasing the amount of disturbances inside of adjacent residential areas. In addition to noise reduction, the hybrid RCV provides improved fuel efficiency and lower $C0_2$ emissions as compared to the conventional RCV.

1.2. Purpose & Objectives

Audio synthesis is a valuable approach for modeling vehicle noise because it allows for complex source conditions, such as various driving speeds, to be modeled with a relatively small amount of computational power and resources as compared to traditional audio recordings. While audio recordings can initially be simple to gather and play back for a listener, this method does not scale well if a larger number of source conditions are required for analysis. The audio playback approach can also become cumbersome when trying to integrate large amounts of recorded data at the same time to form a complex source combination. Synthesis models tend to use significantly less memory than recordings, which permits a higher level of flexibility for the model-based approach [11]. When coupling all of this information with the fact that there is a general lack of auralization models for heavy-duty road vehicles, the synthesis approach becomes an especially attractive option to explore here in this work.

With this in mind, the purpose of this thesis is to investigate and develop the application of a granular source model for a heavy-duty truck with a hybrid engine at multiple positions surrounding the truck. The granular model effectively works by saving short pieces of laboratory recorded pressure signals and then combining these small signal segments together to synthesize an engine sound. This type of source model is useful because it allows for the modeling of different types of heavy-duty trucks operating for various driving conditions, including constant speeds and accelerations. The sounds auralized by this method can then be used to evaluate the noise levels and sound quality of the truck noise in both indoor and outdoor environments.

1.3. Thesis Overview

This report is separated into chapters that detail the various stages of the work. First, there is an overview of some fundamental digital signal processing principles in Chapter 2. The origins of the experimental data from the Volvo Group and the implementation of the granular detection and synthesis algorithms are explained in Chapter 3. Chapter 4 covers the results of two different listening tests that were conducted to evaluate the quality of the granularly modeled sounds for both constant speeds and accelerations. The process of auralizing sounds to accompany a video animation of both a hybrid and conventional garbage truck collecting waste on a city street is described in Chapter 5. Finally, Chapter 6 concludes the work by summarizing the key findings and describing some further approaches that could be extended for future work.

2. Theory

2.1. Digital Signal Processing Concepts

Before beginning with the methodology of the granular approach, it is first necessary to introduce some essential digital signal processing concepts. The direct application of these concepts will be covered in the following sections of this report.

2.1.1. Convolution & Cross-Correlation

Convolution is used to calculate the response of a linear system to an input signal. The linear time-invariant system is defined by its impulse response, and the convolution of this impulse response and the input signal is the output signal response. Once the impulse response of a system is known, it is possible to find the output of the system using convolution for any arbitrary input.

Since finite digital recordings are of interest in this work, the equation for the convolution of functions x and h can be given in discrete form as [8]:

$$x[n] * h[n] = \sum_{k=-\infty}^{\infty} h[k]x[n-k]$$

As seen in this equation, the output signal response is given by the area overlap between functions x and h as a function of the amount that x is translated.

Another useful mathematical operation in digital signal processing is cross-correlation. This operation is very similar to convolution and measures how similar two signals are to one another as a function of time delay. Because of this, cross-correlation is frequently used to find the time delay between two signals. In discrete form, the equation for the cross-correlation of functions x and h is given by [8]:

$$x[n] \star h[n] = \sum_{k=-\infty}^{\infty} h^*[k]x[n+k]$$

The cross-correlation of two real-valued signals would appear as a mirror image of the convolution of the same two signals. Intuitively, this is logical for real-valued signals since the equations for these two operations simply show a time reversal of one of the input signals. For complex-valued signals, taking the conjugate of h, which is written as h^*, helps to ensure that the imaginary components will be accounted for when aligning peaks in the cross-correlation calculation.

2.1.2. Digital Filtering

A digital filter is a tool used in digital signal processing to perform mathematical operations on a discrete-time signal. The overarching goal of filtering is generally to enhance (or diminish) some aspect of the signal. In the time domain, filters are usually designed to achieve a specific frequency response. Common applications of digital filters include low-pass, high-pass and band-pass filters, where only certain frequencies are passed and others are attenuated depending on the design of the filter. The passband, transition band, and stopband are shown in Figure 2.1 for a generic filter.

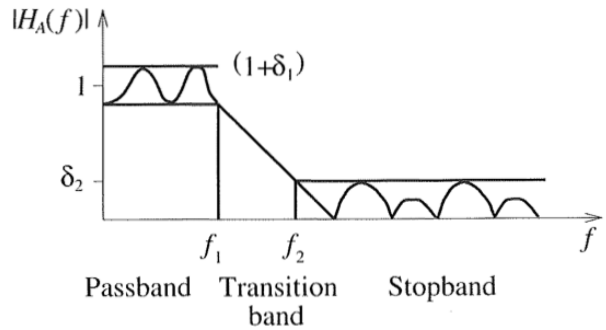

Figure 2.1.: Representation of the passband, transition band and stopband for a generic filter [8].

Using either the filter's impulse response or the Z-transform in discrete-time systems, the filter's frequency response $H(\omega)$ can be obtained. This transfer function, i.e. the ratio of the output transform over the input transform, can be expressed in the Z-domain for a linear, time-invariant digital filter as [8]:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{a_0 + a_1 z^{-1} + a_2 z^{-2} + a_N z^{-N}}{1 - b_1 z^{-1} - b_2 z^{-2} - b_M z^{-M}}$$

where a are feedforward coefficients and b are feedback filter coefficients. This is the general form of a recursive filter and produces an infinite impulse response (IIR) filter. If feedback is excluded and the denominator is set to unity, then this becomes a finite impulse response (FIR) filter. FIR filters can be made to have zero phase, while this is basically impossible for IIR filters. A filter is zero-phase when $H(\omega) > 0$ in the passband and when $H(\omega)$ is a real and even function of frequency [8].

In particular, zero-phase digital filters are of interest in this work because they help to preserve the characteristics of a filtered time waveform exactly where they occur in the unfiltered signal, while conventional filtering can introduce a small delay in the filtered time waveform. Zero-phase digital filtering is done by processing the input signal in both the forward and reverse directions [10].

Windowing

A window function is a mathematical function that is simply zero-valued outside a given interval. In general terms, a window function is useful because the product of the window function and another function is also zero-valued outside of the given interval, leaving only the part where the two functions overlap. For spectral analysis, the primary goal of windowing is to look at finer details of a given signal instead of considering the entire signal. The properties of the window function determine exactly how this is achieved; a wider window gives better frequency resolution but poorer time resolution, while the opposite is true for a narrower window function [8].

Though there are many different types of window functions, a window is typically chosen to be tapered, i.e. bell-shaped, for most digital signal processing applications due to its properties; more heavily tapered windows generally give better rejection in the stopband at the cost of a wider transition band between the passband and stopband (see Figure 2.1). One common type of tapered window is the Hanning window, and the coefficients of this function are given by [8]:

$$w(n) = \frac{1}{2}\left(1 - \cos\left(\frac{2\pi n}{N}\right)\right), \qquad 0 \leq n \leq N$$

where N is the width of the discrete-time, symmetrical window function in samples. The main advantage of the Hanning window is that it tends to reduce aliasing, though there is a small tradeoff of decreased resolution due to a wider main lobe [16]. Figure 2.2 shows a 128-point Hanning window in both the time and frequency domains, as created in Matlab.

As discussed in more detail later in the report, the Hanning window is used in the granular synthesis algorithm to taper and add multiple signals together to form one longer total signal.

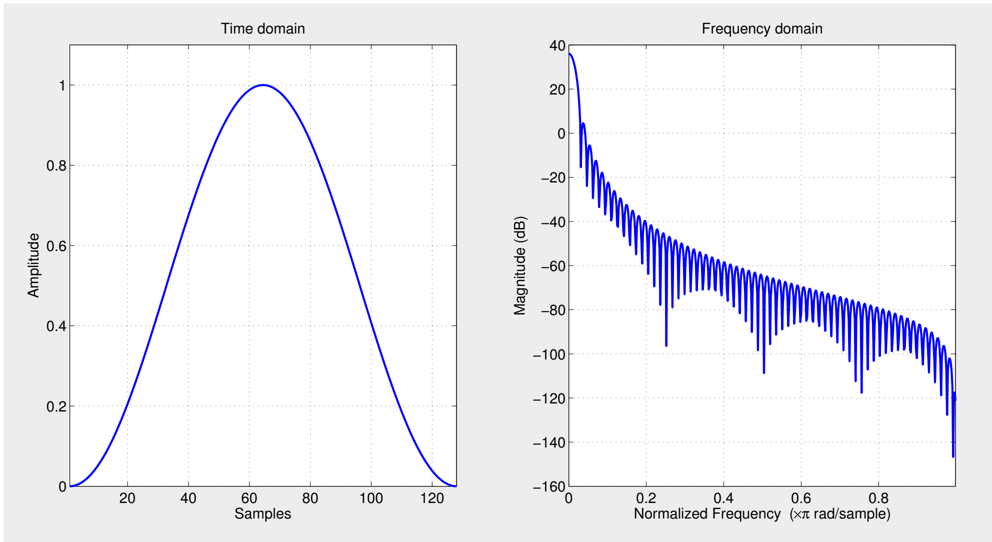

Figure 2.2.: Time and frequency domain representations of a 128-point Hanning window.

2.1.3. Granular Synthesis

Granular synthesis is a method in which a larger sound or group of sounds is broken into smaller grains and then reorganized to form a new sound. This method was originally used to synthesize music in the 1970s and 1980s by creating 1–50 ms long samples called grains. Through experimentation, it was found that different sonic and artistic effects, such as variable phase, level, and frequency, could be created by layering the grains together in different ways [12].

More recently in academic work, the granular approach has most commonly been applied to the synthesis of human speech. One specific way to apply granular synthesis is to use a method called pitch synchronous overlap-and-add (PSOLA). PSOLA is an overlap-and-add technique that is based on the decomposition of a signal into a series of successive waveforms each containing a pitch period of the signal , i.e. a grain. The overlap-and-add sum of these grains reconstructs the signal and allows for control of the pitch and length of the signal. One advantage of this type of synthesis is that in theory the source signal does not lose any details since it is operated on directly [14].

In terms of variables, PSOLA grain detection consists of deconstructing a signal $s(t)$ into a series of grains $s_i(t)$. The deconstruction is obtained by applying the window function $h(t)$ centered on time instants m_i [14]:

$$s_i(t) = h(t - m_i)s(t)$$

These m_i are also called pitch markers and are positioned pitch-synchronously, i.e. close to the local fundamental period of the signal [5]. This pitch marking condition is required to avoid deterioration of the grain during windowing. Once a database of grains has been established, PSOLA synthesis proceeds by overlap-and-add of the grains $s_i(t)$ repositioned on time instants $\tilde{m}_j$ to form a total synthesized signal $\tilde{s}(t)$ [14]:

$$\begin{cases} \tilde{s}_j(t) = s_i(t + m_i) \\ \tilde{s}(t) = \sum_j \tilde{s}_j(t - \tilde{m}_j) \end{cases}$$

Additional pitch shifting or time stretching modifications can be performed on the synthesized grains if desired. In speech processing, pitch is commonly changed by moving the grains further apart to decrease the pitch or closer together to increase the pitch. Similarly, grains can either be repeated or eliminated to increase or reduce the total length of the signal.

While PSOLA was originally designed for use with speech synthesis, it can be used for any application where the source signal content is both harmonic and suitable for deconstruction into grains via windowing. Since PSOLA relies upon periodicity, it is important to make note of any non-periodic or transient parts of a given signal to ensure that they are not saved as grains.

Keeping in mind the requirement of periodicity, a granular approach can also be used to model engine noise from a vehicle due to the physical nature of an engine. For the case of a heavy-duty truck, a pitch period or grain corresponds to one complete engine cycle, or one ignition for every engine cylinder (typically four or six cylinders total). The length of each grain can be approximated by correlating the engine speed to the frequency of the engine revolution [4]:

$$\begin{cases} F_{engine} = \frac{R}{60} \\ F_c = \frac{R}{120} \end{cases}$$

where R is the engine speed in RPM, F_{engine} is the frequency corresponding to the revolution of the engine and F_c is the fundamental frequency corresponding to a complete cycle of the engine, or two revolutions.

The basic application of this method to an engine is shown in Figure 2.3 for the synthesis of grains for a four-stroke internal combustion engine using a Hanning window and an overlap of 128 samples. Here the grains are shown with four sinusoidal cycles with additional small tails to the left and right of the core grain components. In this case, one long synthesized signal is successfully formed from four smaller variable length grains.

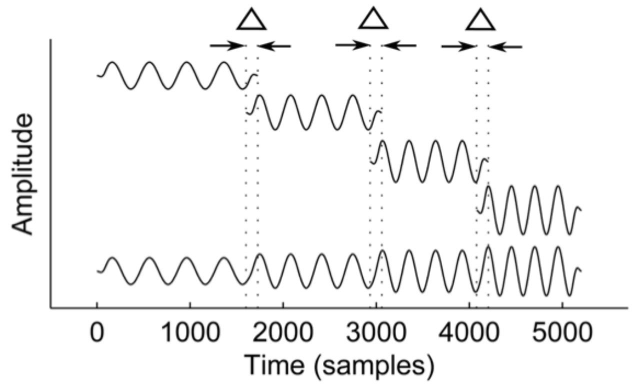

Figure 2.3.: Visual representation of OLA method for Δ = 128 sample overlap, individual grains (top) and total synthesized signal (bottom) [4].

As compared to other modeling strategies, the main advantage of synthesizing grains from an existing database is that the bulk of the computational complexity is not performed in real time. Rather, the synthesis is limited solely to the selection of grains from a database and the addition of the overlapping samples. This relative simplicity makes the PSOLA granular approach suitable for real time applications. Prior research by Jagla et al. in this topic indicates that this method is significantly faster than any time or frequency domain additive synthesis methods and that over 1000 engine sounds can be simultaneously synthesized in real time on a standard personal computer using a single thread processor [4]. In practical terms, this method is appropriate for quickly synthesizing various driving conditions of the hybrid truck for many microphone positions.

3. Implementation

The implementation of a granular model for the Volvo FE Hybrid measurement data is explained in this section of the report. This approach for modeling the hybrid truck's sound sources is split into two distinct steps: first offline grain detection and then synthesis of the grains. The procedure detailed here is valid for both the combustion engine and electric motor propulsion sources, and any differences in the methodology for synthesizing constant speeds versus variable engine speeds, i.e. accelerations, are noted accordingly. An overview of the general workflow strategy for the detection and synthesis processes is seen below in Figure 3.1. It is important to note that real-time synthesis is computationally possible with the granular approach, though it was not of interest in this particular application.

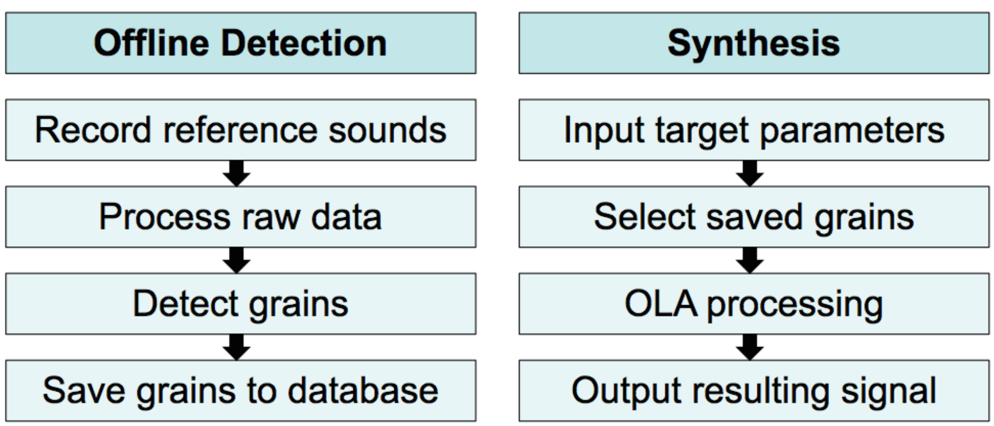

Figure 3.1.: Overview of granular approach workflow.

3.1. Granular Approach

3.1.1. Measurement Data

All measurement data was provided by the Volvo Group in Gothenburg, Sweden. These recordings were conducted for two heavy-duty trucks and were completed primarily in 2013 in the Volvo Group's semi-anechoic laboratory. In this project, the primary truck of interest is a Volvo FE Hybrid RCV. This is a heavy-duty truck with both a six-cylinder diesel combustion engine and an electric motor. The drivetrain is a parallel hybrid, which means that the two propulsion sources cooperate based on factors such as vehicle velocity, gear and the available charge in the electric motor's battery. During waste loading, the hybrid truck uses a separate electrical loading mechanism, and the hybrid drivetrain is in standby mode and therefore not active during this process.

For the measurements used in this work, the electric motor was operated for vehicle speeds up to 30 km/h, and the combustion engine was measured for higher vehicle speeds. The dimensions (length x width x height) of the hybrid truck are: 9.490 x 2.600 x 3.406 m. The hybrid truck is shown in the semi-anechoic laboratory in Figure 3.2.

Figure 3.2.: An overview of the recording setup for the Volvo FE Hybrid truck in a semi-anechoic laboratory [3].

Additionally, some measurements were taken of a conventional Volvo FE truck. The main difference between the conventional truck and the hybrid truck is that the conventional truck only uses the 6-cylinder diesel engine as its propulsion. The dimensions (length x width x height) of the conventional RCV are: 9.907 x 2.600 x 3.406 m.

The conventional and hybrid trucks were supplied by Renova AB and were chosen for analysis because they are commonly used to collect waste from dwellings in Gothenburg. The hybrid truck is primarily of interest with respect to the granular method since it contains both means of propulsion: a diesel engine and an electric motor.

The measurement data consists predominantly of two different driving cycles, the city cycle and the RCV cycle. The city cycle is a standard driving cycle that includes accelerations and constant speeds and is meant to represent typical distribution conditions of a heavy-duty truck. The RCV cycle was created specifically with waste collection in mind and was developed with the goal of gaining a better understanding of the sound levels of actual urban waste collection scenarios [3]. Both cycles were implemented in the semi-anechoic laboratory by driving the vehicles on top of a roller dynamometer. A list of completed measurement trials is available in Appendix A.

The sound pressure data was recorded with a sampling frequency of 44.1 kHz using Brüel & Kjær Type 4189, 1/2-inch free-field microphones. 22 of these microphones were distributed around the truck in a box-shaped array that fulfilled the requirements of ISO 3744 for sound power measurement. Additionally, 16 microphones were placed at a pass-by distance 7.5 m to the left and right of the truck, i.e. 8 microphones on each side, at a height of 1.2 m. A general schematic of the microphone positions is shown in Figure 3.3, and detailed information regarding the microphone coordinates can be seen in Appendix B.

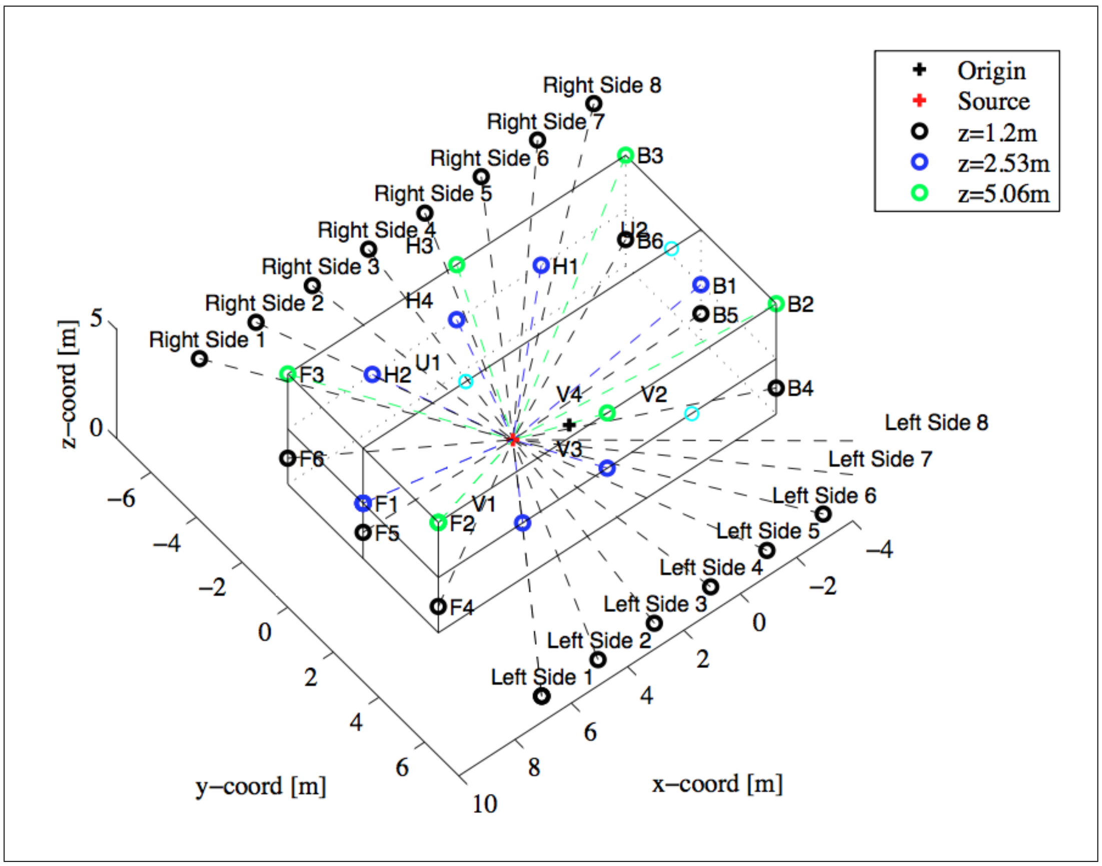

Figure 3.3.: Graphical layout of measurement microphone positions.

In addition to the sound pressure recordings, CAN-signals were recorded for each measurement. These CAN-signals were recorded with a sampling frequency of 100 Hz and included important vehicle information, such as engine speed, vehicle speed and torque on the rear wheels.

3.1.2. Raw Data Processing

Before operating on the signals to detect and synthesize grains, the raw data provided by the Volvo Group had to be processed in order to get all of the signals in a more convenient and consistent format. The recordings were first inspected both visually and aurally. This was done to identify any basic problems in the data, such as overloaded microphone channels or other recording errors. In most cases, the visual analysis included plotting of selected pressure signals, sound pressure and power levels, and spectrograms and a brief order tracking analysis for driving cycles that included accelerations, i.e. variable vehicle and engine speeds. Subjective listening is an important part of this process, as it can help to more easily identify problems that aren't apparent in the visual analysis of the data. Some unwanted noise sources included occasional engine cooling fan noise and very rarely noise in the recorded signals most likely due to clipping or electrical noise, poor connections, etc.

After the preliminary inspection of the signals, the long raw files were split into steps that correspond to each step of the particular driving cycle of interest. These files included the recorded data for every microphone position, along with vehicle and engine speeds from the CAN-signals. For some of the recordings, the time steps of the measurement data was inconsistent, so the data was resampled and saved in order to adhere to the same format as the rest of the data.

One example of exported motor speeds for multiple driving conditions is shown in Figure 3.4 for the hybrid truck. This plot includes both accelerations and constant speeds for this truck's electric motor. This data is critical for determining grain size in the impending granular detection and synthesis algorithms.

After the driving cycles were split into their appropriate steps, cross-correlation was used to determine the time delay between the microphone positions relative to a reference position. In reality, any microphone position can be used as a reference, but microphone F5 was chosen for this case due to its convenient location directly in front of the truck at a height of 1.2 m. An example of time shifted signals via cross-correlation is shown in Figure 3.5.

The microphone positions shown in Figure 3.5, positions F5 and F6, are quite close to each other and have a delay of only 209 samples, or 4.7 ms ($f_s = 44.1$kHz). As cross-correlation is a measure of how similar two signals are to one another, it is rather intuitive that microphone positions closer to the reference point tended to have less of a time delay as compared to microphones located to the sides and back of the trucks. This is due to both distance and the fact that the character of the noise was typically quite different farther away from the reference position. Depending on the position, it was

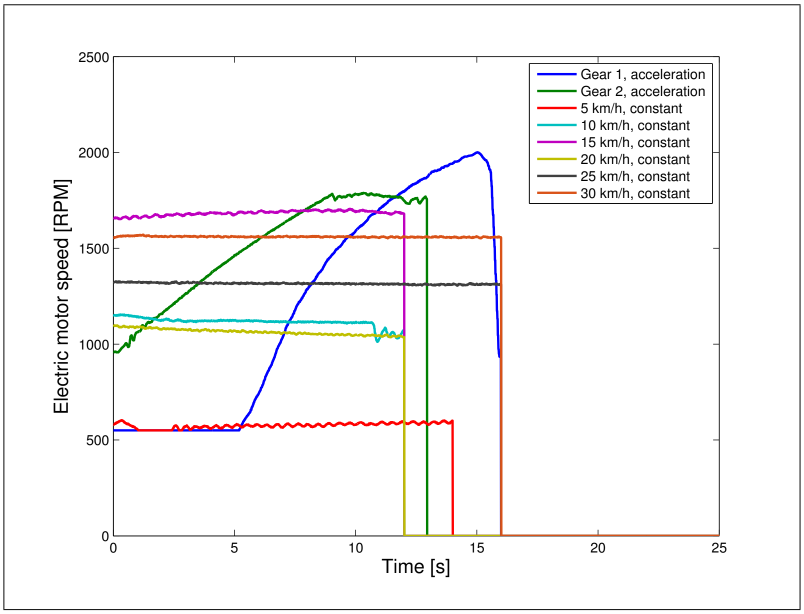

Figure 3.4.: Electric motor speeds separated into different steps for various driving conditions.

possible that the recording contained major contributions from the exhaust outlet or rolling noise from the rear tires, in addition to the noise from the engine. In any case, the maximum delay calculated via cross-correlation for any position was typically around 2000 samples, or about 45 ms. Using the speed of sound ($c_0 = 343$ m/s) to convert time to distance, this corresponds to approximately 15.4 m, which is a reasonable estimate of the distance between the front and rear microphone positions.

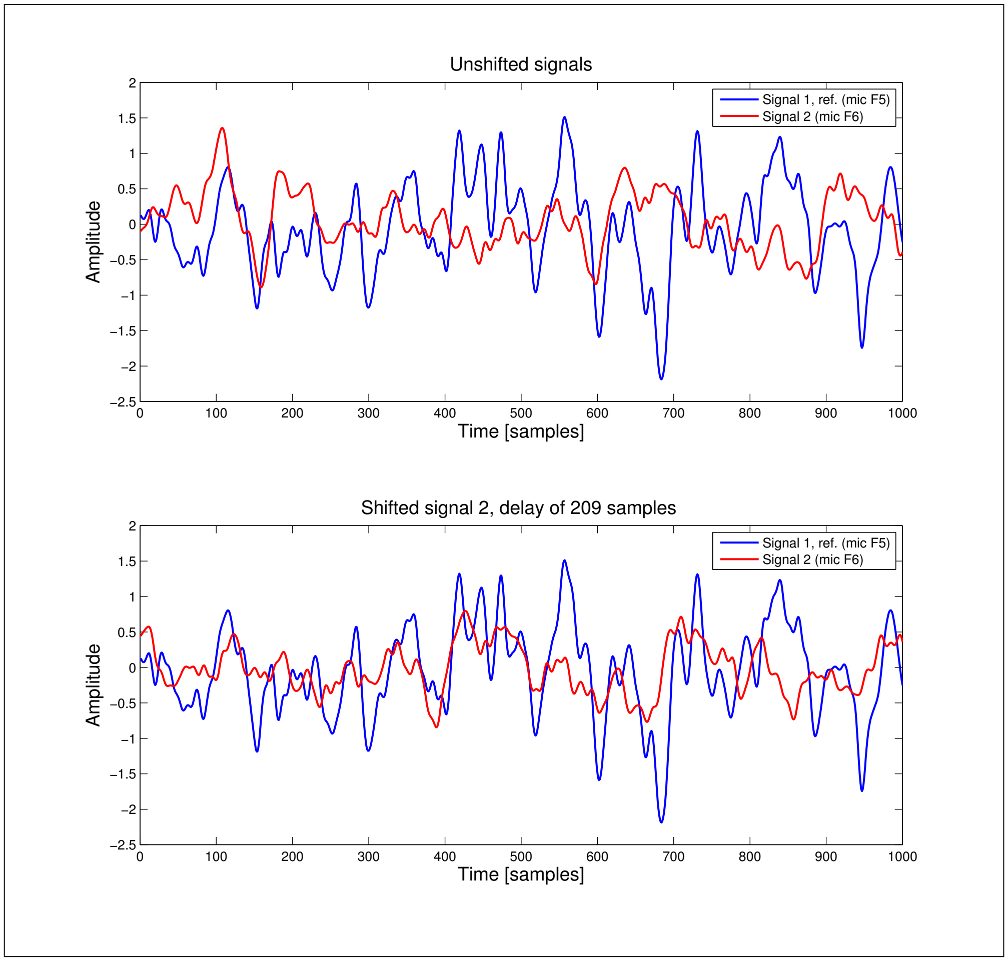

Figure 3.5.: Application of cross-correlation to find time delay between two microphone positions.

3.1.3. Grain Detection

The grain detection process begins with the user defining a range of driving conditions and microphone positions (microphone #2) to analyze. For each driving condition the engine speed and microphone distance compensation data, which includes the amount of delay between the signals relative to the reference signal (microphone F5), are loaded into the workspace in Matlab.

The reference signal and signal #2 are then imported for the relevant driving condition, and signal #2 is shifted relative to the reference signal based on the delay in samples that was calculated previously using cross-correlation. After accounting for the time delay between the signals, a high-pass filter is applied to both signals using a zero-phase filter with a cut-on frequency of 20 Hz.

Before detecting grains from the signals, it is first necessary to define some variables to set up the grain properties, such as the wavelet size, the search area and the amount of samples to skip to leave small margins for the left and right tails of the grain. The wavelet size of the grain is the number of samples that compose one ignition, or 1/6th of the total length of a typical combustion cycle. The search area is the number of samples, in either the positive or negative direction, to search for the grain and was defined as 55% of the size of the wavelet size in this case. The size of the skipped region at the beginning and end of the signal is typically determined to be slightly greater than six times the wavelet size (one complete cycle), but this value can be changed accordingly to fulfill the needs of different scenarios.

The actual detection algorithm can be started after defining these essential properties. The grain search and detection is contained in a "while" loop structure that continues detecting grains in small increments until the end of the signal is reached. Using the number of samples contained in the search area, a portion containing one combustion cycle, i.e. six wavelets along with some margin for the tails to account for variations, of the high-passed reference signal is extracted from the total signal in the time range of the current loop iteration. The start of this six-cycle segment signal, or the first relative maximum, is found using convolution with a "peakshape" function consisting of six ideal sinusoidal cycles whose period depends on the size of the current wavelet size. This relative maximum indicates where the grain begins both in time and amplitude, and the six peaks that follow can be taken to be the core of the grain. Some small simplifications have been made here in the explanation of the detection algorithm for the sake of brevity, but a more complete view of the grain detection algorithm is available for inspection in Appendix D.

For each iteration, the grain cores and left and right tails of the reference signal and signal #2 are saved in structure array databases for the corresponding vehicle type, propulsion type, driving condition and microphone position. The grain core refers to the central six cycles of the detected signal, and the left and right tails are the six cycles that immediately precede and follow the core, respectively. As there are many different measurement positions and driving conditions, accurate labeling of saved data is important to ensure that the correct databases are referenced during synthesis.

For constant speed driving conditions, 12–16 grains are typically saved in the databases, and the resulting MAT-files are around 1 MB in size for each microphone position. For accelerations, the size of the databases depends on the length of the reference recordings and can vary in size from a few MB up to 100 MB for each microphone depending on the number of grains in each database.

The detection of three successive constant speed grains is shown in Figure 3.6. In this figure, the red line represents the high-passed filtered signal (microphone F5), the black line a low-frequency bandpassed version of the signal (for inspection only), and the blue and green data markers denote the three saved grain cores. The detection of the combustion engine is shown here because this type of signal is higher in amplitude than the electric motor and easier to observe the grain detection, but the algorithm works the same way for each propulsion source.

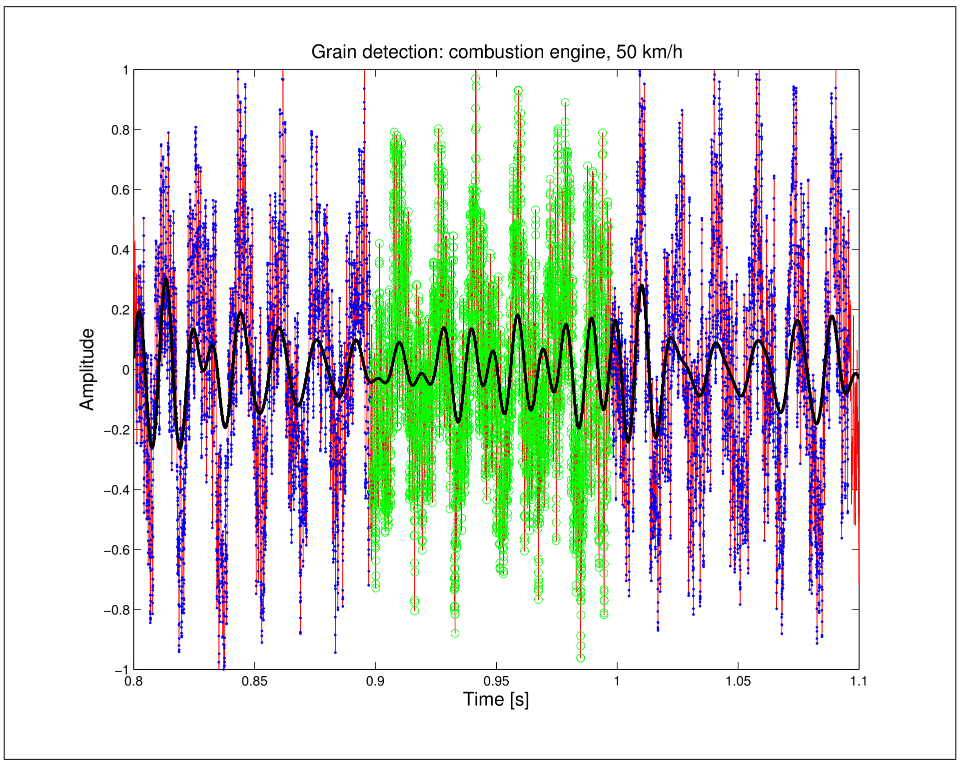

Figure 3.6.: Grain detection for combustion engine constant speed (1200 RPM), vehicle speed 50 km/h, microphone F5.

After saving the grains, the loop continues to the next iteration and the new wavelet size is calculated based on the updated engine speed. A correlation vector also updates the grain length to adapt to any compression or stretching of the signal. This process continues over and over until the end of the signal is reached or until a maximum number of grains is detected, as is the case for constant speeds. Since the algorithm updates the wavelet size for every iteration, it is possible to use the same code for either constant or variable engine speeds for both propulsion sources. Higher engine speeds contain less samples in each grain, while the opposite is true for lower engine speeds.

3.1.4. Grain Synthesis

The synthesis of the grains from their respective databases is broken down into either constant speed or acceleration synthesis functions. For both the combustion engine and the electric motor, there is a main script that allows the user to specify factors such as the driving condition, synthesis time, and microphone position for the synthesis. Based on these inputs, the appropriate function is called for the selected driving condition. The implementation of these algorithms is explained in more detail in the following sections.

Constant Speed Synthesis

For the synthesis of constant speed sounds, the grain databases for the reference microphone (F5) and the microphone of interest are first loaded into the workspace for the driving speed under study. Each database is typically either 12 or 16 grains long for constant speeds, depending on what was specified during the detection process. Regardless, either size of database works efficiently with the synthesis function and is capable of creating good auralizations, as detailed later in Chapter 4. The constant speed synthesis function has the potential to construct an infinite signal, and the user simply inputs the desired signal length in seconds into the main script.

Next, a pre-determined random order is calculated for use when adding the grains together. This is done to ensure that the same random order is used for each microphone position when multiple positions are analyzed at the same time. The only stipulation in the random order is that the grains should not be repeated in either an AA or ABAB structure. When these repetitions are included in the synthesized sounds, it is possible that the resulting auralizations sound unnatural, similar to a skipping CD with very quick, successive repetitions. Any repetitions in the grain order are corrected using simple conditional statements.

After being defined or loaded, all of the relevant variables are then passed to the synthesis function. These input variables include the two grain databases, the random grain order, the total number of grains, and the synthesis time. Using the random grain order, the grains are added to one another in succession until the length of the synthesized signal is equal to or exceeds the required synthesis time. The grains are added using the overlap-and-add method, with a 128-sample Hanning window to taper the tails of the grains. While the Hanning window is the only window function that was used in the synthesis, the code allows the user to select triangular and Hamming windows if desired. Other window functions could easily be defined as well.

In particular, the grains are joined together by multiplying the left and right tails of the grains with the corresponding left and right portions of the window functions. The total synthesized signal is formed into one long vector by concatenating the previous iteration of the synthesized signal with the overlapping part of the grains and the grain core of the current iteration. For additional clarification of this process, the MATLAB code for the constant speed synthesis is available for consultation in Appendix D. This procedure of adding tapered grain segments together is repeated over and over again for both the reference signal and signal #2 until the signals are long enough to fulfill the desired synthesis time.

The synthesized reference signal and signal for the second microphone are finally returned to the main script, and a WAV-file and a corresponding MAT-file are saved for each of the synthesized signals. An example of a synthesized signal is shown in Figure 3.7. Here a recorded signal is compared to a synthesized signal for the electric motor and a vehicle speed of 25 km/h (motor speed of approximately 1300 RPM). Spectrograms of each of these signals are shown in Figures 3.8 and 3.9 as well. The spectrograms are quite similar since the motor speed is stationary, though there is a large decrease in the amount very low frequency content in the synthesized signal due to the high-pass filter that was implemented at 20 Hz. While the resulting signals may appear slightly different in time in Figure 3.7 due to the nature of the granular synthesis, they sound extremely similar - if not indistinguishable - and are evaluated as part of a discrimination listening test later in Chapter 4.

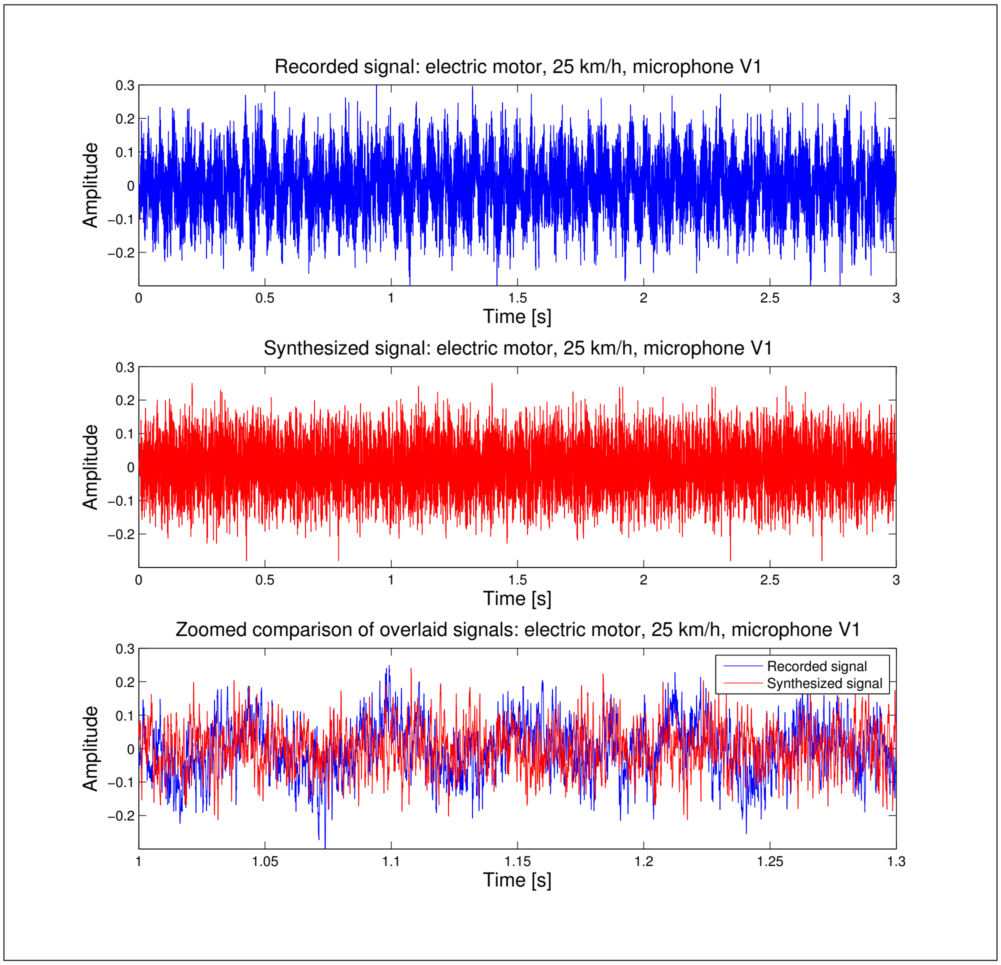

Figure 3.7.: Comparison of constant speed recording and synthesis for electric motor propulsion, vehicle speed 25 km/h, microphone V1.

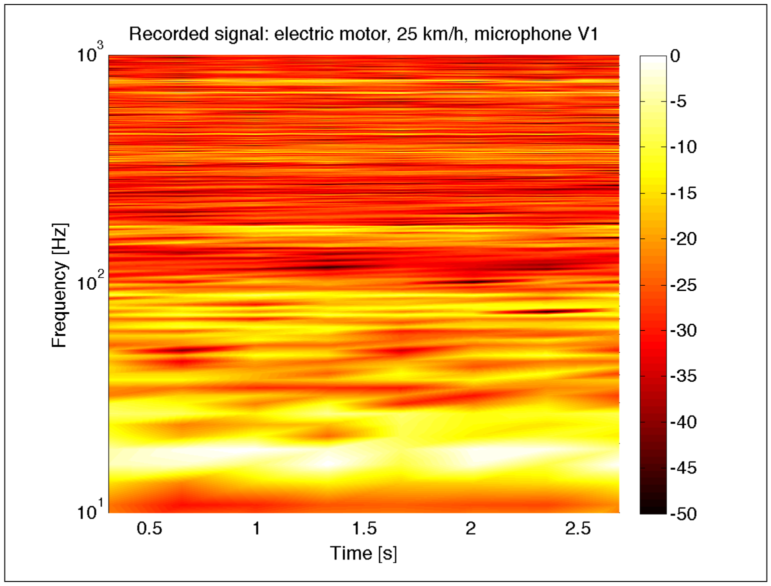

Figure 3.8.: Spectrogram of recorded signal for electric motor propulsion, vehicle speed 25 km/h, microphone V1.

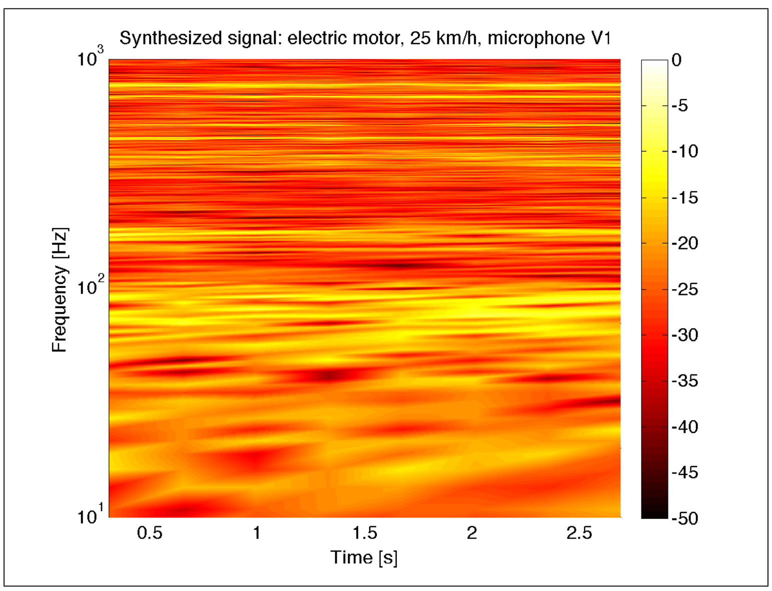

Figure 3.9.: Spectrogram of synthesized signal for electric motor propulsion, vehicle speed 25 km/h, microphone V1.

Acceleration Synthesis

The synthesis of acceleration sounds is quite similar to the constant speed synthesis procedure. Again, the grain databases are loaded for both the reference microphone and a second microphone of interest. These databases are of variables length depending on the engine speeds and length of the source recordings. For acceleration synthesis, the user defines a desired engine speed vector for the synthesis function to track. Special caution should be taken to ensure that the desired engine speed synthesis fits within the constraints of the corresponding grain database, i.e. an acceleration up to 2000 RPM should not be attempted to be synthesized if its database only contains detected grains up to 1500 RPM.

The primary input variables for the acceleration synthesis function are the desired engine speed vector as a function of time, the two grain databases, the synthesis time, and a vector that contains the lengths of every grain in the databases. The acceleration function adds the grains together in the same way that the constant speed function does (overlap-and-add with a Hanning window to taper the grain tails), but now the order of the synthesized grains is processed differently and is much more critical since a random order will not produce an adequate auralization for engine speeds that either increase or decrease with time.

For acceleration, the lengths of every grain in a given database are first sorted in ascending order, and the grain chosen for synthesis is the one that most closely matches the engine speed for a given instant in time. The target grain length, l_{grain} [samples], for each time iteration of the synthesis is determined by the relationship:

$$l_{grain} = \frac{120 * f_s}{RPM_{target}}$$

where f_s is the sampling frequency (44.1 kHz) and RPM_{target} is the target engine speed for synthesis. Once the requisite target grain length is calculated, the synthesized grain is chosen from the sorted list of grain lengths for the grain that most closely matches l_{grain} in size. Again, the algorithm is designed to avoid selecting potentially unnatural sounding grain repetitions in the form of either AA or ABAB. However, when a database is not sufficiently large enough to model the input synthesis conditions, then grains may be repeated due to a lack of grain selection choices.

This process is repeated for each of the two signals until the total signal lengths meet the desired synthesis time. Since this algorithm tracks the engine speed for every iteration, it is possible to use this approach for acceleration, deceleration, or even a combination of the two depending on the quality of the grains in the reference database. Sample MATLAB code is shown in Appendix D for the acceleration synthesis function.

The acceleration function returns the two synthesized signals, the targeted engine speed vector, and a matrix of the selected grain indices and their corresponding lengths in number of samples. As an example of this process, the measured and targeted motor speeds are shown in Figure 3.10 for an electric motor acceleration of 1065-1775 RPM. In this case, the synthesis time was equal to the length of the desired motor speed, so the signals are simply overlaid over one another. However, the synthesis algorithm will either stretch or compress the desired engine speed vector to match the synthesis time for signals that differ in length from the source data.

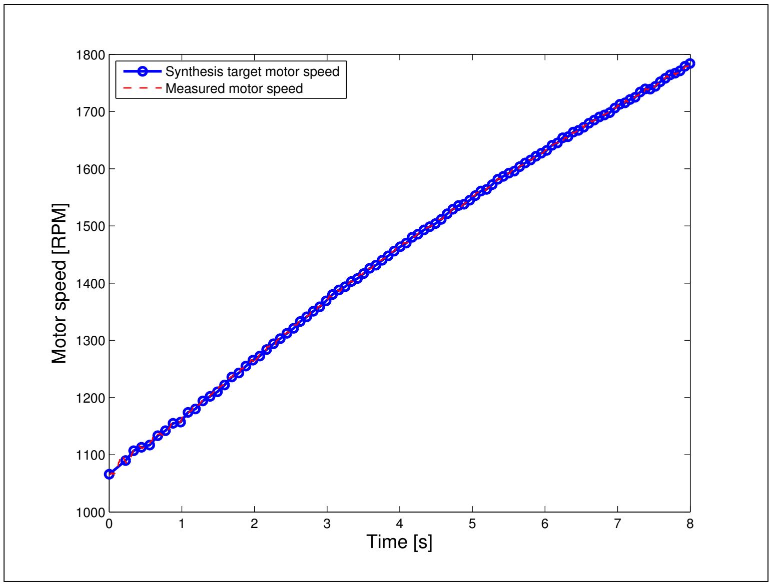

Figure 3.10.: Comparison of measured and targeted electric motor speeds for acceleration synthesis, motor speed 1065-1775 RPM, vehicle speed 20-33 km/h, microphone V1.

Similarly, a plot of the selected and ideal grain lengths and the relative error between the two are shown as functions of time in Figure 3.11 for the same synthesis. This plot helps to illustrate how the algorithm selects grains based on an ideal grain length and a sorted list of grain lengths in the database. Grain length varies inversely with the motor speed since increasing motor speeds correspond to decreasing grain lengths. While the algorithm attempts to avoid grain repetitions, there is not a sufficient number of grains in this particular database for some motor speeds, so there are some fluctuations and unintended repetitions for some time intervals in this plot. This of course causes the quality of the synthesized sound to suffer and produces obvious discontinuities in the synthesized sound. For this synthesis case, the relative error between the ideal and selected grain lengths is within $\pm 3\%$.

Finally, a comparison of the recorded and synthesized signals for the same electric motor acceleration is displayed in Figure 3.12, and the corresponding spectrograms are

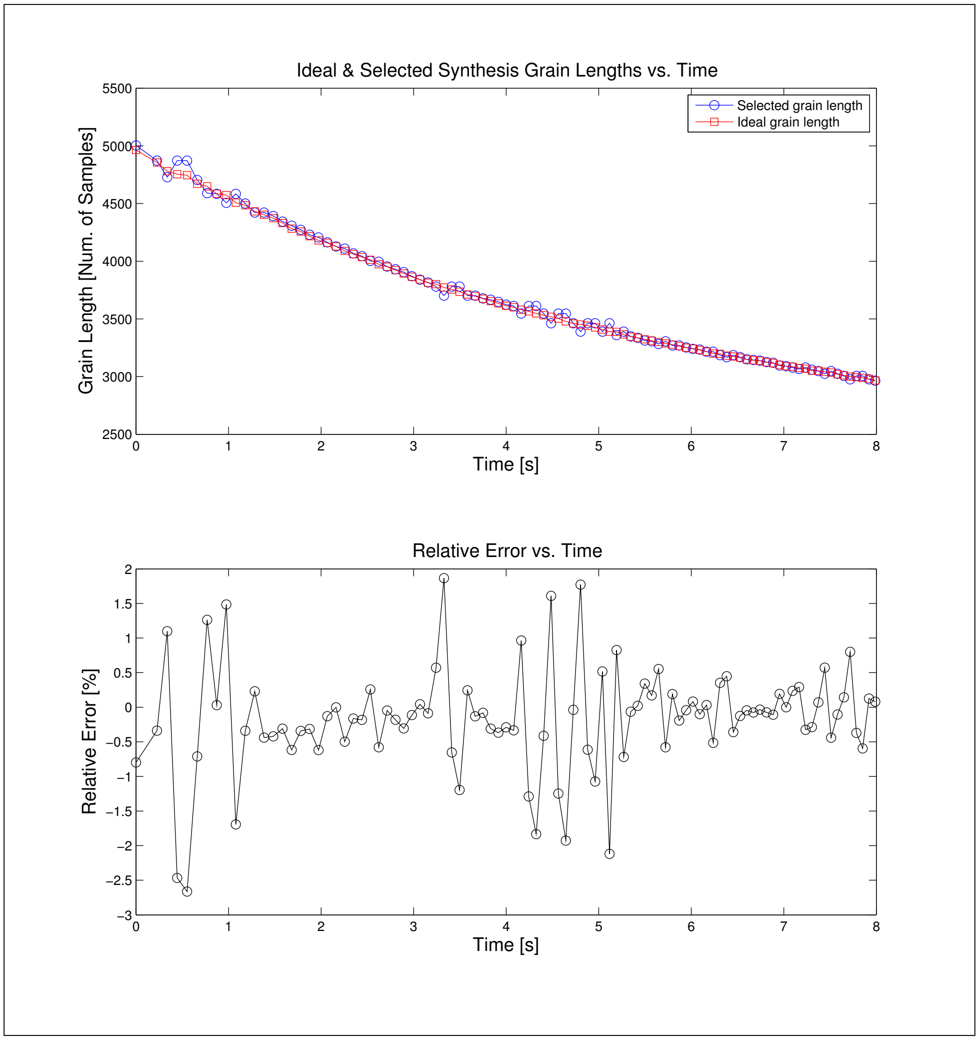

Figure 3.11.: Plots of grain length and relative error as functions of time for acceleration synthesis, motor speed 1065-1775 RPM, vehicle speed 20-33 km/h, microphone V1.

shown in Figures 3.13 and 3.14. The spectrograms show the level increases from about 300–500 Hz as the motor speed increases over the course of the acceleration, though higher order harmonics are slightly more apparent for the recording than for the synthesis. There are obvious amplitude differences in the signal content in Figure 3.12 for some time intervals, especially from 5-7 s, though it is difficult to qualitatively evaluate the synthesized sound from a visualized pressure signal alone. Largely because of this reason, both of these signals are tested in a semantic differential listening test in Chapter 4 along with other acceleration sounds.

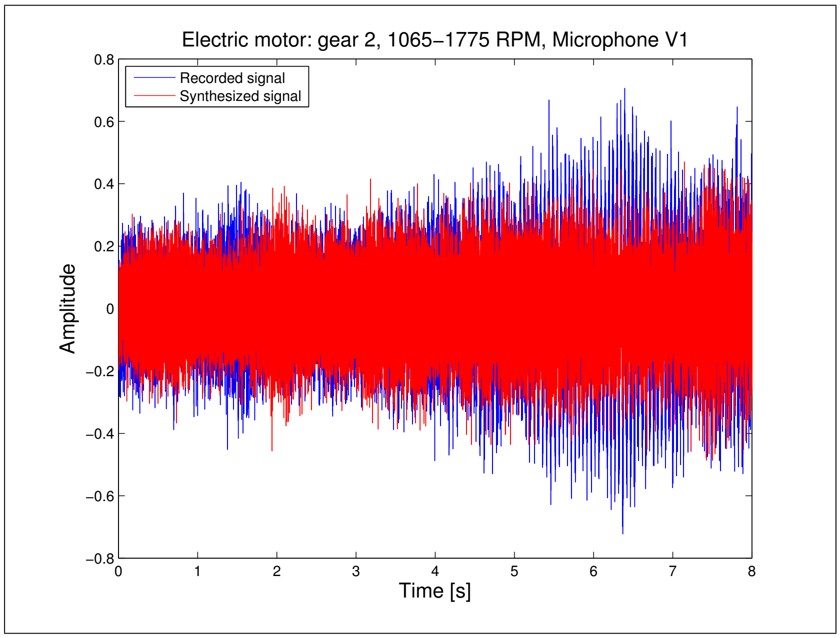

Figure 3.12.: Comparison of acceleration recording and synthesis for electric motor propulsion, motor speed 1065-1775 RPM, vehicle speed 20-33 km/h, microphone V1.

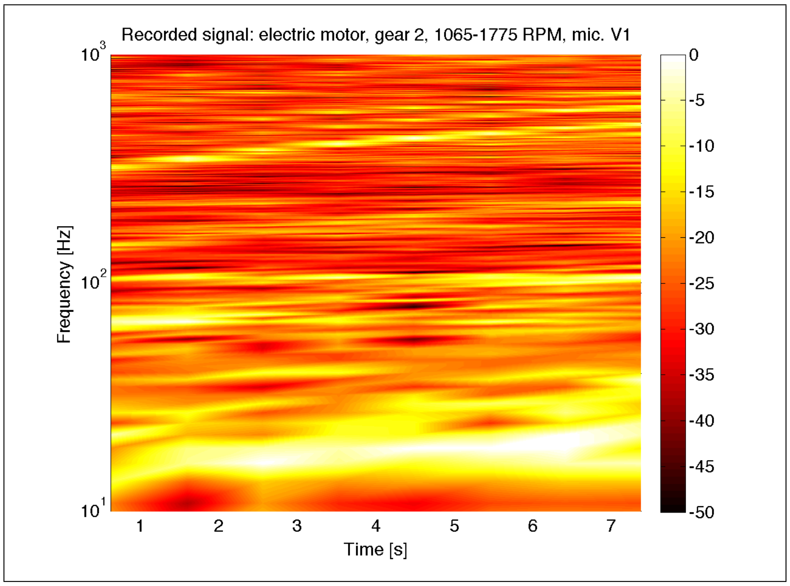

Figure 3.13.: Spectrogram of recorded acceleration signal for electric motor propulsion, motor speed 1065-1775 RPM, vehicle speed 20-33 km/h, mic V1.

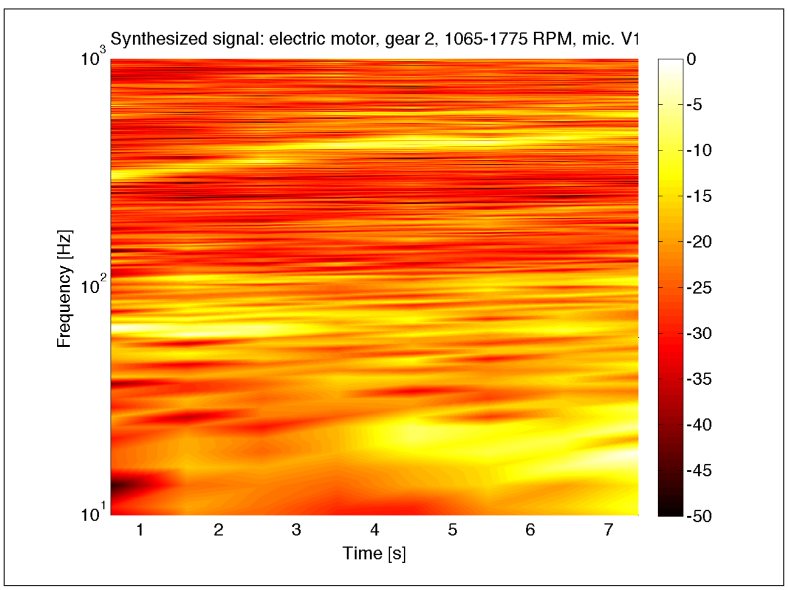

Figure 3.14.: Spectrogram of synthesized acceleration signal for electric motor propulsion, motor speed 1065-1775 RPM, vehicle speed 20-33 km/h, mic V1.

4. Evaluation of the Granular Approach

4.1. Listening Test Overview

The synthesized sounds were evaluated using different types of listening tests depending on the type of driving condition. Constant speeds for the hybrid truck's electric motor were evaluated in a same-different discrimination listening test, and accelerations for both the electric motor and the combustion engine of the hybrid truck were evaluated using a set of four semantic differential scales. A discrimination test was determined to not be useful for accelerations due to audible variations between the measured and synthesized signals.

Each of these listening tests took place concurrently and included a total of 16 listeners, of which ten were male and six were female. The mean age of the listeners was 27.4 years with a standard deviation of 3.3 years. One listener reported that he suffered from mild tinnitus, and the others reported no known hearing damage. The listening test took place in an acoustically treated listening room at the Chalmers Division of Applied Acoustics in Gothenburg. This test setup is shown in Figure 4.1.

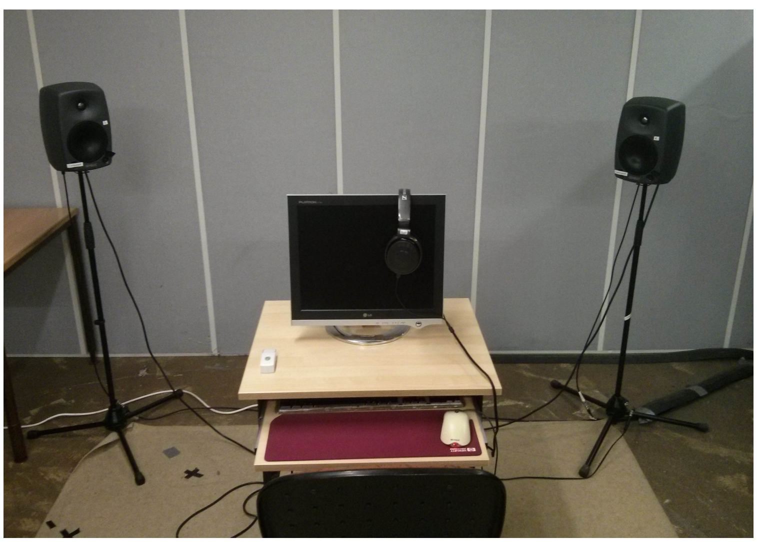

Figure 4.1.: The listening test setup in a acoustically treated room.

4.1.1. Constant Speed Sounds

While there are many ways of evaluating how similar an auralized sound is to a real sound, discrimination is the most strict way of evaluating this property [9]. This is due to the fact that any noticeable difference can be used to discriminate between the sounds. It also does not require that the listener knows which sound is auralized and which sound is real. Due to these strict test parameters, it is possible that auralized sounds that fail a same-different discrimination test are still perfectly useful as a source model depending on the desired application.

Aside from discrimination, other methods such detection, similarity, and psychoacoustic attributes can be used to evaluate auralized sounds. Detection requires the listener to detect whether or not a sound is real or auralized. For similarity, the listener rates how similar or dissimilar the auralized and real sounds are. Psychoacoustic attributes allow the listener to rate different qualities, such as perceived annoyance, loudness, or vehicle speed, using different types of rating scales.

Constant speeds had been previously evaluated in a discrimination test for prior work using the granular approach for the conventional heavy-duty truck with a diesel combustion engine. In this study, it was found that for nine out of ten pairs of sounds listeners could not differentiate between a recorded and synthesized sound at a 95% confidence interval [1]. In this test, speeds of 0 km/h (idle), 20 km/h, 50 km/h, 60 km/h and 70 km/h were evaluated for microphone positions in front of and to the left of the truck – microphones F5 and V1, respectively. The only sound that the test participants could discriminate between the synthesized sound and the recording sound at an above-chance performance level (significant for $p < .001$) was for the front microphone position at a speed of 20 km/h.

Since the granular model continued to produce very realistic sounds for the electric motor operating at constant speeds, it was again decided to test the hybrid truck's constant speeds using a same-different discrimination test. The truck speeds and corresponding electric motor speeds that were tested for this experiment can be seen in Table 4.1. As shown in this table, the same two microphones were again chosen for this analysis: one directly in front of the truck at a height of 1.2 m (F5) and one to the left side of the truck at a height of 2.53 m (V1). Exact microphone positions and coordinates are available in Appendix B.

This portion of the listening test took approximately 20–30 minutes to complete. Listeners were asked to determine whether a pair of stimuli were equal or not equal, and the stimuli were played as monaural signals through Sennheiser HD 650 dynamic headphones. The auralized and recorded sounds were each 3 seconds long, with tuning in and tuning out times of 10 ms. Each pair of sounds was combined into four different combinations: AA, AB, BB, and BA. These pairs were presented in a random order and the four combinations were tested twice for each participant.

Mic. position	Driving condition				
	10 km/h *(1125 RPM)*	15 km/h *(1675 RPM)*	20 km/h *(1050 RPM)*	25 km/h *(1300 RPM)*	30 km/h *(1575 RPM)*
F5	AB	AB	AB	AB	AB
V1	AB	AB	AB	AB	AB

Table 4.1.: Set of pairs from recordings (A) and synthesis (B) using 12 grains.

Binomial Analysis

The listeners' responses to the discrimination test were analyzed using a two-tailed, nondirectional binomial test in SPSS statistical software to determine if the participants could differentiate between the recordings and the synthesized sounds. By definition, a binomial test consists of identical, independent trials with one of two outcomes: a success or a failure [17]. For this experiment, a success would consist of either identifying stimuli AA or BB as the same or AB or BA as different, while a failure would be identifying AA or BB as different or AB or BA as the same. The results were analyzed using a hypothesis test with $\alpha = 0.05$, a null hypothesis of $H_0 : p = 0.5$, and an alternative hypothesis of $H_1 : p \neq 0.5$.

Finally, the p-values for each pair of recordings and synthesized sounds are listed in Figure 4.2. The null hypothesis was retained ($p > .05$) for eight out of ten pairs of stimuli, meaning that the recordings and synthesized sounds could not be differentiated by the listeners. The two pairs of sounds that rejected the null hypothesis ($p \leq .05$) were the front and left microphone positions for the highest speed of 30 km/h, meaning that the listeners could differentiate between the recorded and synthesized sounds at an above-chance performance level (significant for $p < .03$).

The majority of incorrect answers in the discrimination test were due to the different stimuli (either AB or BA) being perceived as the same. However, there was also a somewhat high proportion of false alarms. A false alarm occurs when two stimuli of the same origin - either AA or BB in this case - are perceived as different. For this experiment, the mean proportion of false alarms was 0.2, with a median proportion of 0.175. In the previous listening test of combustion engine constant speeds, the mean proportion of false alarms was less than 0.1, and it is believed that the relatively high number of false alarms observed here occurred because most of the participants were trained listeners who may have held pre-conceived expectations that the pairs of sounds would differ more than they did in reality. Past discrimination listening tests consisting of similar vehicle sounds have shown that trained listeners are more likely to correctly distinguish between real and synthesized sounds, and that they are more likely in general to indicate that sounds are different from one another [6]. It is therefore

Binomial Test

		Category	N	Observed Prop.	Prop.	p-value
10 km/h Mic F5	Group 1	1.00	75	.59	.50	.063
	Group 2	0.00	53	.41		
	Total		128	1.00		
10 km/h Mic V1	Group 1	1.00	65	.51	.50	.930
	Group 2	0.00	63	.49		
	Total		128	1.00		
15 km/h Mic F5	Group 1	1.00	71	.55	.50	.250
	Group 2	0.00	57	.45		
	Total		128	1.00		
15 km/h Mic V1	Group 1	1.00	73	.57	.50	.133
	Group 2	0.00	55	.43		
	Total		128	1.00		
20 km/h Mic F5	Group 1	1.00	75	.59	.50	.063
	Group 2	0.00	53	.41		
	Total		128	1.00		
20 km/h Mic V1	Group 1	1.00	68	.53	.50	.536
	Group 2	0.00	60	.47		
	Total		128	1.00		
25 km/h Mic F5	Group 1	1.00	74	.58	.50	.093
	Group 2	0.00	54	.42		
	Total		128	1.00		
25 km/h Mic V1	Group 1	1.00	71	.55	.50	.250
	Group 2	0.00	57	.45		
	Total		128	1.00		
30 km/h Mic F5	Group 1	1.00	77	.60	.50	.027
	Group 2	0.00	51	.40		
	Total		128	1.00		
30 km/h Mic V1	Group 1	1.00	77	.60	.50	.027
	Group 2	0.00	51	.40		
	Total		128	1.00		

Figure 4.2.: Binomial listening test results.

possible that a better balance of trained and untrained listeners would lead to a small reduction in the false alarm rate.

Subjective Judgements of Constant Speed Sounds

Through subjective listening, the constant speed synthesized sounds were deemed as very similar to the recordings with respect to spectral and temporal content. Temporal differences become slightly more apparent as the vehicle velocity increases due to the increased contributions of rolling noise components, which may help to explain the listeners' collective ability to differentiate the sounds at 30 km/h. Similarly, other audible effects may come from periodicities that are uneven in relation to the cycle of the motor.

Another potential source of error that must be considered is a small amount of human error in the driving of the truck during the measurement taking process. The truck speeds are assumed to be perfectly constant here, though in reality the truck speed was manually controlled by a driver and could vary up to ± 3 km/h. The three-second recordings that were extracted from measurement data were carefully selected with this fact in mind, but it is still possible that some small variations in speed exist in the sounds that were evaluated in the discrimination test. Variations in vehicle and motor speeds in turn affect both the granularly synthesized sounds and the real constant speed recordings.

4.1.2. Acceleration Sounds

The synthesized acceleration sounds were evaluated using a different type of listening test. This decision was made because there were obvious differences between the recordings and synthesized sounds, so a discrimination test was not a practical choice. In order to still be able to extract meaningful information from the synthesized accelerations, it was decided to use an attribute test to evaluate the quality of each sound [9]. This attribute test required the listener to rate various sounds on four semantic differential scales. The following four questions were asked on a nine-point scale:

- How realistic is the sound?
  - Range: Unrealistic (1) to Realistic (9)
- How annoying is the sound?
  - Range: Not annoying (1) to Very annoying (9)
- How pleasant is the sound?
  - Range: Unpleasant (1) to Pleasant (9)
- How activating is the sound?

– Range: Calm (1) to Highly activating (9)

Since an accurate source model was the goal of the synthesis, the primary semantic scale of interest was realism. However, it is still interesting to gain insight on how the other attributes were rated for different types of acceleration sounds, especially for the 8 second recordings and synthesized sounds.

The four questions listed above were asked for a set of twenty sounds: ten electric motor accelerations and ten combustion engine accelerations. The twenty sounds were played in a random order and then repeated again. The listener's responses for the first round of listening to the sounds were thrown out and only the second round of responses are included here in the results. This was done in order to allow the listener to have a chance to become familiar with the entire range of sounds and to avoid penalizing him or her for rating earlier sounds more strictly or leniently than later sounds.

The set of twenty acceleration sounds are shown in Table 4.2. As seen in the table, the synthesized sounds varied between 2–8 seconds in length, while all four of the recorded sounds were 8 seconds long. In an effort to produce a fuller, more spatially realistic sound, the signals from microphones F5, H2, U1, and V1 were normalized together into one file for each of the accelerations.

All stimuli were played over a calibrated pair of Genelec 8030A nearfield monitors, which are shown in Figure 4.1. Despite the fact that the signals were monaural, loudspeakers were chosen for this part of the listening test to give the listener a better sense of spatial accuracy since a sound such as a heavy-duty truck would not be encountered in real life through headphones.

The acceleration synthesis was severely limited by the available measurement data to detect and synthesize the grains. As shown in Table 4.2, the electric motor accelerations spanned a much wider range than the diesel engine accelerations. In an ideal scenario, the same engine speeds would have been evaluated for each of the two trucks, allowing for more direct comparisons between the combustion engine and the electric motor.

After the listening test was completed, a general overview of the semantic ratings was obtained by calculating the mean value of the listeners' response. These mean ratings for each semantic scale are shown for each tested sound in Figures 4.3 and 4.4 for the hybrid truck's combustion engine and electric motor, respectively.

In general, the perceived realism of the synthesized sounds for the combustion engine was rated closer to the realism of the recorded sounds, especially for the gear 4 accelerations (1150 - 1330 RPM). There is a clear difference between the synthesized and recorded sounds for the perceived realism of the electric motor accelerations, and the recorded sounds were rated as substantially more realistic. For the most part, there isn't much variance with respect to the other three attributes, and it is difficult to draw any noteworthy conclusions from inspection of mean ratings alone.

Propulsion	Driving condition			
	Gear 1 *(625-1910 RPM)*	Gear 2 *(1065-1775 RPM)*	Gear 4 *(1150 - 1330 RPM)*	Gear 6 *(1200 - 1400 RPM)*
Diesel engine (2 s)	-	-	S	S
Diesel engine (4 s)	-	-	S	S
Diesel engine (6 s)	-	-	S	S
Diesel engine (8 s)	-	-	R, S	R, S
Electric motor (2 s)	S	S	-	-
Electric motor (4 s)	S	S	-	-
Electric motor (6 s)	S	S	-	-
Electric motor (8 s)	R, S	R, S	-	-

Table 4.2.: List of sounds formed from recordings (R) and synthesis (S) for acceleration cases.

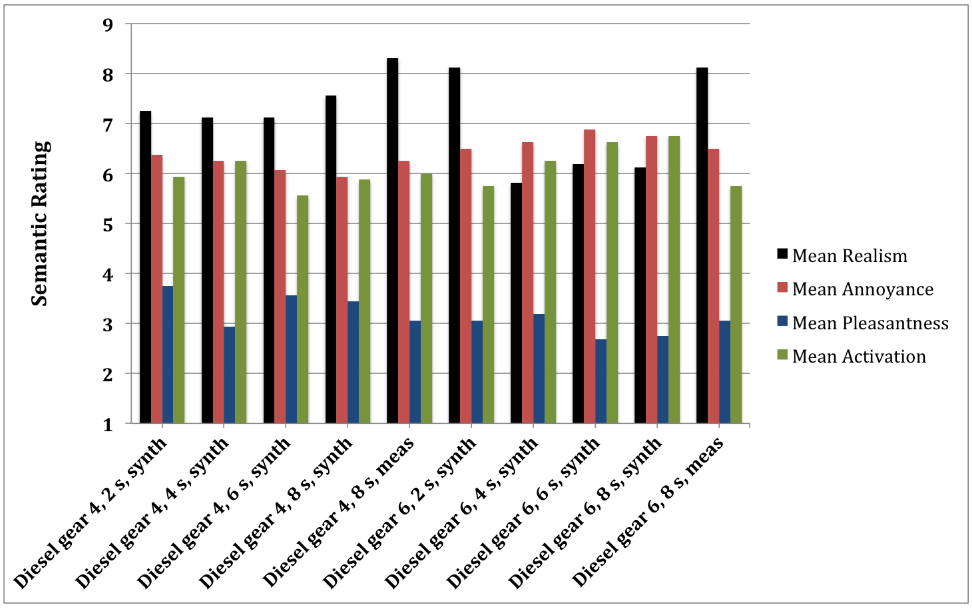

Figure 4.3.: Mean results of semantic differential scales for combustion engine accelerations: two gears, four synthesis times (synth) and two recordings (meas).

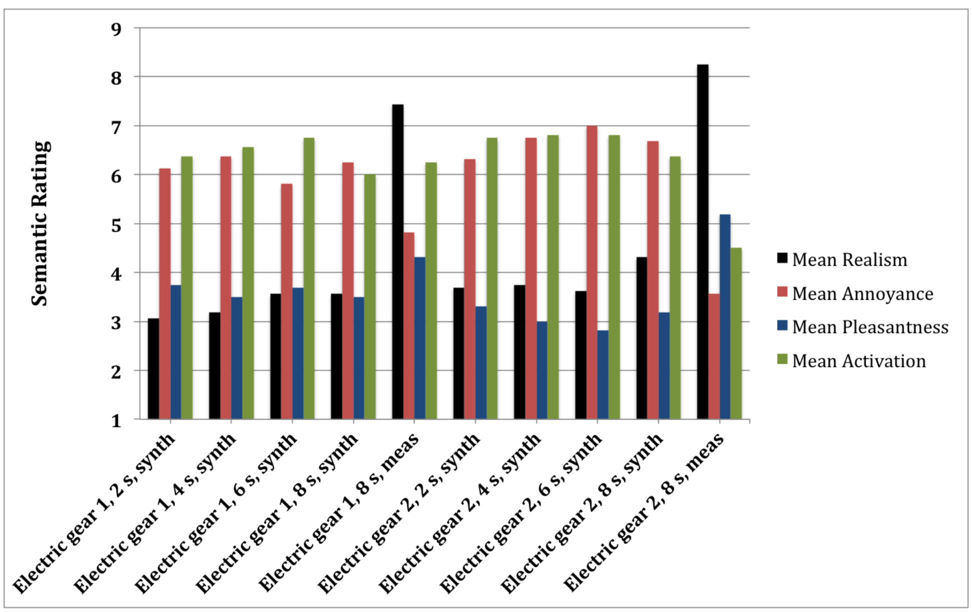

Figure 4.4.: Mean results of semantic differential scales for electric motor accelerations: two gears, four synthesis times (synth) and two recordings (meas).

Repeated Measures Analysis of Variance

A more in-depth look of the semantic differentials was completed using a repeated measures analysis of variance (ANOVA) using SPSS statistical software. A repeated measures ANOVA is used to compare multiple sample means when the means are from a single-factor within-subjects design. SPSS used the input results of the semantic differential responses to calculate F-statistics, which are used to determine statistical significance. These F-values are ultimately calculated as functions of variance referred to as the mean squared within-subjects (MS_{effect}) and the mean squared error (MS_{error}) [7]:

$$F = \frac{MS_{effect}}{MS_{error}}$$

For each trial, the F-statistic is reported in SPSS along with a probability (p), effect size ($\eta^2_{partial}$) and the degrees of freedom of the effect and error. When the F-value is sufficiently large, a small p-value is obtained, indicating that the set of means differ significantly from each other. In other words, for $p < .05$ the null hypothesis is rejected and the alternative hypothesis is accepted.

The purpose of $\eta^2_{partial}$ is to provide a standardized measure of significant effects. Effect size varies directly with $\eta^2_{partial}$, so a larger effect is given by larger values of $\eta^2_{partial}$. This value is calculated as a function of the sum of squares of a given effect (SS_{effect}) and the sum of squares of the error associated with that effect (SS_{error}) [7]:

$$\eta^2_{partial} = \frac{SS_{effect}}{SS_{effect} + SS_{error}}$$

These repeated measures ANOVA were completed in different stages in order to adequately analyze different effects. First, the eight synthesized sounds for each truck were separately analyzed for each of the two gears. The truck gear was isolated due to the fact that the range of engine speeds is quite different for each of the trucks, as discussed previously. It is also important to note that the repeated measures ANOVA was repeated four times for each trial in order to separately account for the four semantic differentials.

The full ANOVA tables of this analysis are shown in Appendix C, and the significant effects are listed below. In these lists, “gear” refers to the two gears tested for each of the hybrid truck’s types of propulsion and “time” to the four different lengths of synthesized signals.

For accelerations of the hybrid truck’s combustion engine for two gears and four times, there were statistically significant effects of:

- gear on realism, $F(1, 15) = 11.765$, $p = .004$, $\eta^2_{partial} = .440$

- gear on annoyance, $F(1,15) = 9.698$, $p = .007$, $\eta^2_{partial} = .393$
- gear on pleasantness, $F(1,15) = 5.714$, $p = .030$, $\eta^2_{partial} = .276$
- gear on activation, $F(1,15) = 7.061$, $p = .018$, $\eta^2_{partial} = .320$
- time on pleasantness, $F(3,45) = 3.172$, $p = .033$, $\eta^2_{partial} = .175$

Clearly, the gear for the combustion engine is a factor for each of the four attributes. The main difference between the two gears is that the truck is traveling at a higher speed in gear 6 as compared to gear 4, which causes the gear 6 group of synthesized sounds to be higher in level and a little more aggressive in nature. This likely causes the differences in perception since collectively the gear 4 sounds were rated on average as more realistic, more pleasant, less annoying and less activating than the gear 6 sounds. The effect size of time on pleasantness is not particularly large, but it is seen that pleasantness generally decreases slightly with increasing synthesis time.

For accelerations of the hybrid truck's electric motor for two gears and four times, there were statistically significant effects of:

- gear on realism, $F(1,15) = 8.136$, $p = .012$, $\eta^2_{partial} = .352$
- gear on annoyance, $F(1,15) = 6.160$, $p = .025$, $\eta^2_{partial} = .291$
- gear on pleasantness, $F(1,15) = 12.638$, $p = .003$, $\eta^2_{partial} = .457$

The gear is again a significant effect for the electric motor. In this case, the main difference between the two gears is that gear 1 spans a larger range of motor speeds as compared to gear 2. This likely causes the differences in perception, though there are no characteristics of the two groups of sound that immediately stand out when listening subjectively. The largest effect is for the gear on pleasantness, and Figure 4.4 shows that the mean pleasantness of the synthesized sounds for gear 1 are all rated as slightly higher than the synthesized sounds for gear 2.

Next, the effects of the hybrid truck's source of propulsion and the length of the synthesized sounds were analyzed. In this case, "propulsion" refers to the two drive systems of the hybrid truck: the diesel combustion engine and the electric motor. There were only two significant effects – both for realism – for this ANOVA, though the full results are located in Appendix C.

For accelerations of the hybrid truck's combustion engine and electric motor for four lengths of times, there were statistically significant effects of:

- propulsion on realism, $F(1,15) = 58.059$, $p = .000$, $\eta^2_{partial} = .795$
- time on realism, $F(3,45) = 3.064$, $p = .037$, $\eta^2_{partial} = .170$

As expected, the propulsion source has a rather large effect on realism and is further supported by the difference between the combustion engine and electric motor mean realism ratings in Figures 4.3 and 4.4. Time has a much smaller effect on realism, and realism is seen to generally increase slightly as the synthesis time increases.

Finally, the variance of the real and synthesized eight second sounds was analyzed for the two propulsion types. These are referred to as the "sound type" in the ANOVA. Again, the complete results are available at the end of Appendix C, and this analysis shows that there are both main effects and an interaction effect between propulsion and sound type that are significant for all of the tested semantic attributes except activation. An interaction effect is when the impact of one factor depends on the level of the other factor [7].

For accelerations of the hybrid truck's combustion engine and electric motor for two sound types, there were statistically significant effects of:

- propulsion on realism, $F(1,15) = 24.153$, $p = .000$, $\eta^2_{partial} = .617$
- sound type on realism, $F(1,15) = 48.867$, $p = .000$, $\eta^2_{partial} = .765$
- propulsion*sound type on realism, $F(1,15) = 32.904$, $p = .000$, $\eta^2_{partial} = .687$
- propulsion on annoyance, $F(1,15) = 6.555$, $p = .022$, $\eta^2_{partial} = .304$
- sound type on annoyance, $F(1,15) = 25.058$, $p = .000$, $\eta^2_{partial} = .626$
- propulsion*sound type on annoyance, $F(1,15) = 15.943$, $p = .001$, $\eta^2_{partial} = .515$
- propulsion on pleasantness, $F(1,15) = 8.606$, $p = .010$, $\eta^2_{partial} = .365$
- sound type on pleasantness, $F(1,15) = 7.507$, $p = .015$, $\eta^2_{partial} = .334$
- propulsion*sound type on pleasantness, $F(1,15) = 11.812$, $p = .004$, $\eta^2_{partial} = .441$

The results in this final ANOVA mesh well with the previous findings. Propulsion, sound type, and the interaction between the two are all significant effects for three of the four attributes. These three effect sizes are all rather large for realism, which makes sense logically because the combustion engine sounds were rated as more realistic than the electric motor sounds, and the measured sounds were all rated as more realistic than the corresponding synthesized sounds.

Overall, these repeated measures ANOVA help to demonstrate how perception of synthesized sounds depends on many different factors. In this case, the attribute "activation" only had a significant effect (gear) for the first ANOVA on the combustion engine, but the other three attributes that were tested contained more significant effects throughout the analyses. Also, only one type of statistically significant interaction (propulsion*sound type) was observed, though this interaction was significant for three different attributes. As mentioned previously, realism was the primary focus of this study, and it was found to have main effects for all of the analyzed trials. Based on $\eta^2_{partial}$ values, the propulsion type, sound type, and propulsion*sound type were the largest calculated effects for realism.

Subjective Judgements of Acceleration Sounds

There are a number of reasons that could explain why the combustion engine accelerations were perceived as more realistic than the electric motor accelerations. For one, the combustion engine sounds could have been perceived as more realistic because the tested engine speed ranges were much smaller than the electric motor's, so the diesel engine's acceleration was rather slow and gradual while the electric motor's speed increased much more rapidly in the test sounds. Also, the combustion engine noise is in general much rougher and contains more low frequency content than the electric motor, which could make discontinuities in the synthesized grains less obvious for the combustion engine due to low frequency masking [18]. It is plausible that each propulsion source had a similar number of discontinuities in the synthesized sounds, but that the discontinuities for the electric motor tended to be more obvious and jarring and removed the listener's sense of realism more easily due to the spectral content of this type of source.

Another factor to consider is the spatial representation of the tested sounds. The recordings were taken for a stationary truck set up on top of a roller dynamometer. In turn, the synthesized sounds were created to simply match this set up and multiple microphone positions were mixed together to create a fuller, more accurate sound for both the synthesized sounds and the recordings. However, a stationary accelerating vehicle is not a common phenomenon in real life, so while the tested accelerations were representative of the capabilities of the granular synthesis model, perhaps the spatial accuracy of the sounds could have been improved by implementing a stereo pass-by model where the sound shifted from the listener's right to his or her left. Hypothetically, this could have helped to improve the perceived position of the vehicle as it accelerated past the listening position, and thus potentially enhanced the listener's sense of realism to some extent.

5. Application of Granular Model for Animation

5.1. Animation Overview

As a part of the SEND SMART project regarding urban waste collection, a video animation including auralized sounds was requested for two different types of heavy-duty trucks driving on Landsvägsgatan in Gothenburg: 1) a conventional waste collection truck operating with a diesel combustion engine, and 2) a hybrid waste collection truck operating with an electric motor. The overarching purpose of these videos was to visually and aurally demonstrate the effects of a quieter waste collection concept that had been studied in this project. All video animations were created by Jacqueline Forzelius of Visual Arena Research in Gothenburg.

5.2. Truck Noise Synthesis

The most substantial and important part of the auralization process was to apply the granular synthesis model to this animation case. First, the truck's position as a function of time and a stationary listening position were provided by the animator, as shown in Figure 5.1.

A coordinate system was developed by placing the x-axis along the center of the street and setting the relative origin of the x-axis ($x = 0$ m) at each of the truck's front axle. Each truck is 9 m long, and the distance from front axle to the front edge of the truck is approximately 1.4 m. For the conventional truck a source position of $(0.45, 0, 1)$ m was used in the model, and a source position of $(-1, 0, 1)$ m was used for the hybrid truck. These approximations were based on the locations of the diesel engine and the electric drivetrain system. The x-coordinates of all measurement microphone positions were also normalized with respect to the front axle. As prescribed by the animator, the listening position was at a location of $(29, 5.5, 1.8)$ m - that is to say 29 m up the street from the truck's starting position at a distance of 5.5 m from the center of the street at a height of 1.8 m.

Once the truck's position was known for each time step, it became possible to calculate the velocity as well. This is shown in Figure 5.2. As shown in both the velocity and position plots, the truck starts off stationary (idling), then accelerates forward in

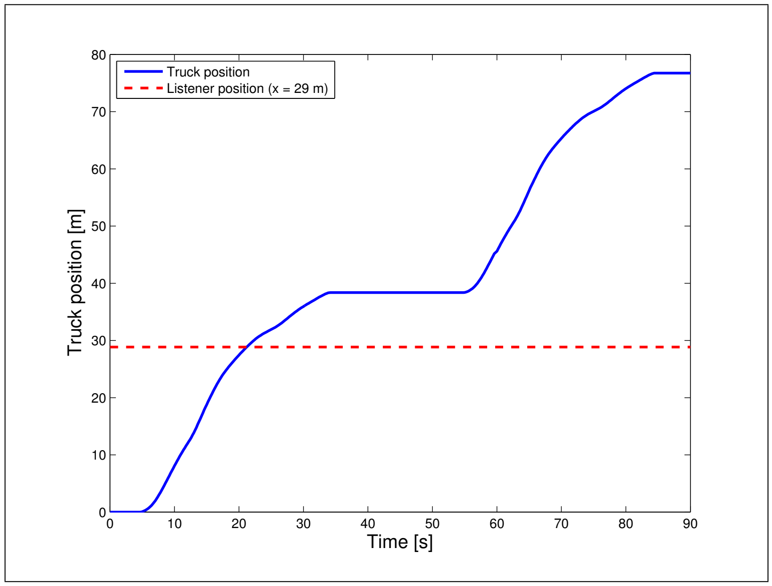

Figure 5.1.: Position of truck on Landsvägsgatan as a function of time.

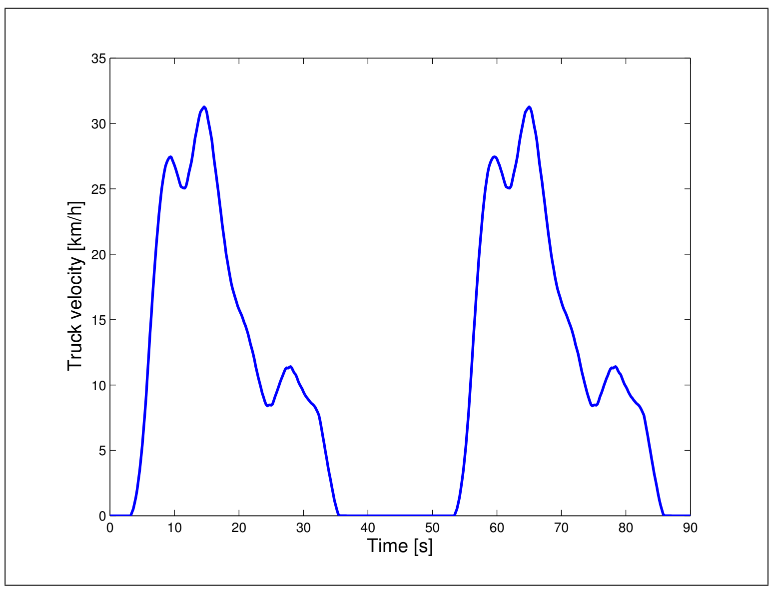

Figure 5.2.: Velocity of truck on Landsvägsgatan as a function of time.

first gear and shifts into second gear before reaching a top speed of around 30 km/h. The truck then decelerates quite slowly before shifting down into first gear and finally stops just to the left of the listener and begins idling again. During this idling, waste is collected from a bin in an apartment building across the street. This waste bin is rolled out to the truck and dumped into the reservoir. The truck then takes off again and repeats the same driving cycle and the video ends with the truck about 45 m down the street to the left of the listening position. It should be noted that in the actual video that the middle idling and waste collection step is actually longer than shown in these plots, but this is rather trivial to account for since the truck is stationary during this process and only shifts the impending movement farther forward in time.

After determining the needed velocity function to complete the driving, it was then possible to work backwards and determine the necessary engine speeds required to achieve these vehicle velocities. For the hybrid truck operating with its electric motor, grains were detected from an appropriate distribution cycle consisting of acceleration and then deceleration with the vehicle velocity increasing up to 30 km/h and then decreasing back to 0 km/h. The electric motor speed in this case varied from 0 – 1900 RPM.

The conventional truck grains were also detected from a representative distribution cycle. The engine speed ranged from roughly 600 (idle) – 2200 RPM. The target engine/motor speeds that were used for the granular synthesis for each respective truck are seen in Figure 5.3.

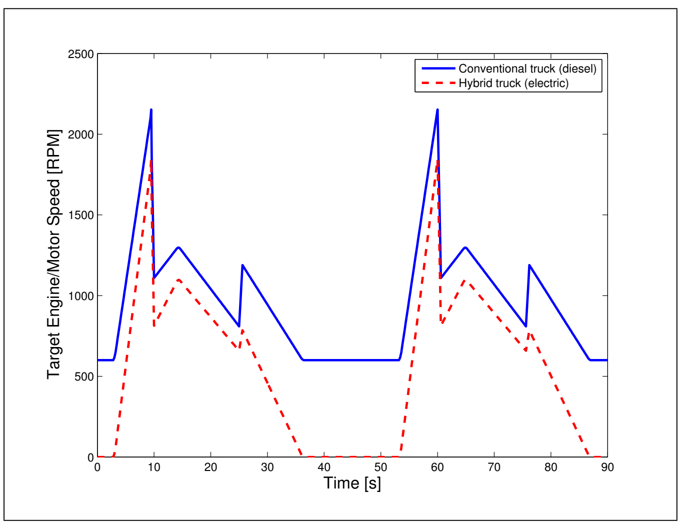

Figure 5.3.: Target engine/motor speeds of synthesized sounds as a function of time for each truck.

Figure 5.4 shows a still image of the conventional truck accelerating toward the listening position, while Figure 5.5 shows the hybrid truck accelerating down the street away from the listening position. For each truck, a total driving cycle for the video was created using the various segments of idling, acceleration, and deceleration by joining them together synchronously using a Hanning window with 128 samples of overlap. Though there were some approximations made in correlating the truck velocities to particular engine speeds, the final synthesized sounds were satisfactory for highlighting the differences in source character between the two different types of trucks.

Figure 5.4.: Conventional (diesel engine) truck accelerating toward listener on Landsvägsgatan [2].

Figure 5.5.: Hybrid (electric motor) truck accelerating away from listener on Landsvägsgatan [2].

5.2.1. Propagation Effect & Microphone Weighting

The synthesized sound obtained in the previous section does not include any distance normalization, so this must be added since the position of the truck changes drastically with respect to the listener. This distance dependence can be calculated by first finding the distance vector, R_{engine}, between the source and receiver for every time step:

$$R_{engine} = \sqrt{(x_{rec} - x_{engine})^2 + (y_{rec} - y_{engine})^2 + (z_{rec} - z_{engine})^2}$$

The proper signal with respect to the free-field is found by simply dividing the previously synthesized time signal with the R_{engine} position vector, or:

$$signal_{re.free-field} = signal_{synth} / R_{engine}$$

This process of normalizing the synthesized signal amplitudes with respect to distance was repeated for all microphone positions on the left side of the truck due to the fact that the listening position is situated to the left of each truck.

Still, it is not enough to simply add all of the normalized signals together. It is necessary to consider how the position of the truck affects the character of the sound that actually arrives at the listener. That is to say that from the listener's perspective the truck noise should be coming primarily from the front of the truck when the truck is to the listener's right, from the side of truck when the truck is parallel to the listener, and primarily from the rear of the truck once the truck has finally passed by the listener.

This type of effect can be accounted for by weighting the amplitude of each microphone signal with respect to the distance to the listener. In simple terms, microphones that are closest to the listener should have a higher amplitude, while microphone positions that are farther from the listener should have a more negligible impact on the observed sound.

To accomplish this effect, Shepard's method for inverse distance weighting allows for the variable microphone positions to be continually updated as the trucks pass by the listener along the side of the street. The appropriate equations for this method are described by [15]:

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

where n is the number of scatter points (microphones) in the data set, f_i are the function values of each scatter point (pressure as a function of time), and w_i are the weight functions assigned to each microphone. These weight functions are found using:

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

where p is the weighting exponent and is typically assigned a value between 1 and 2. h_i is the distance from each microphone to the interpolation point, or the listening position, which is located at (x, y, z) m. h_i is simply found by calculating the distance between these points, or:

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

where (x_i, y_i, z_i) m are the coordinates of each microphone. This distance equation can them be plugged into the original weight function, which is then simplified to:

$$w_i = \frac{\frac{R - h_i}{R h_i}^p}{\sum_{j=1}^{n} \frac{R - h_i}{R h_i}^p}$$

where R is the distance from the listening position to the most distant microphone.

For this case of the trucks driving on Landsvägsgatan, there are eight microphones that are included in the final auralized sounds. These microphones are labeled F2, F4, B2, B4, V1, V2, V3 and V4 and were chosen for inclusion because they span the entire length and height of the left side of the truck. Their exact coordinates relative to a common origin in the middle of the truck are found in Appendix B.

The weighting exponent for Shepard's method in this case was chosen as $p = 1$. Since the truck's position on the street is known as a function of time (as shown in Figure 5.1), it is easy to calculate the corresponding microphone distances to the listening position. These calculated weight functions are seen in Figure 5.6 for all eight microphones used in the auralization.

The peak in the weight functions corresponds to when each individual microphone passes the listening position. Intuitively the weight functions make sense because the microphones in front of the truck have the highest weighting amplitudes when the truck is to the right of the listener. After the truck passes the listener, the rear microphones dominate the majority of the sound.

5.2.2. Doppler Effect

In addition to the source-receiver distance compensation, the Doppler effect was taken into account since the source is moving relative to the listening position, which results in a slight shift in the observed frequency, f, and the emitted frequency, f_0, whenever the trucks are in motion. This relationship is true whenever the velocities of the source and receiver are lower than the velocity of the waves in the medium and is written as [13]:

$$f = \left(\frac{c + v_r}{c + v_s} \right) f_0$$

where c is the speed of sound in air, v_r is the velocity of the receiver, and v_s is the velocity of the source and is positive if the source is moving away from the receiver and negative if the source is moving toward the receiver. It is important to note that there is a slight source amplitude change on the order of $\sim \frac{1}{1-M}$, where $M = \frac{v_s}{c}$, but this amplitude term is neglected here since the vehicle speeds are quite low. Since the receiver is not moving in this case on Landsvägsgatan and the change in source amplitude is neglected, the Doppler equation can be simplified to:

$$f = \left(\frac{c}{c + v_s} \right) f_0$$

In practical terms, the trucks reach a maximum speed of approximately 30 km/h, or 8.33 m/s, in the animations. This means that the observed frequency is up to $\pm 3\%$ relative to the emitted frequency depending on whether or not the trucks are approaching or driving away from the listener. Using a maximum source velocity of 8.33 m/s, various frequencies for f and f_0 can be calculated and are shown in Table 5.1.

These results indicate that the Doppler effect plays a small - but not insignificant - effect on the frequency content of the auralized sound that is observed at the listening

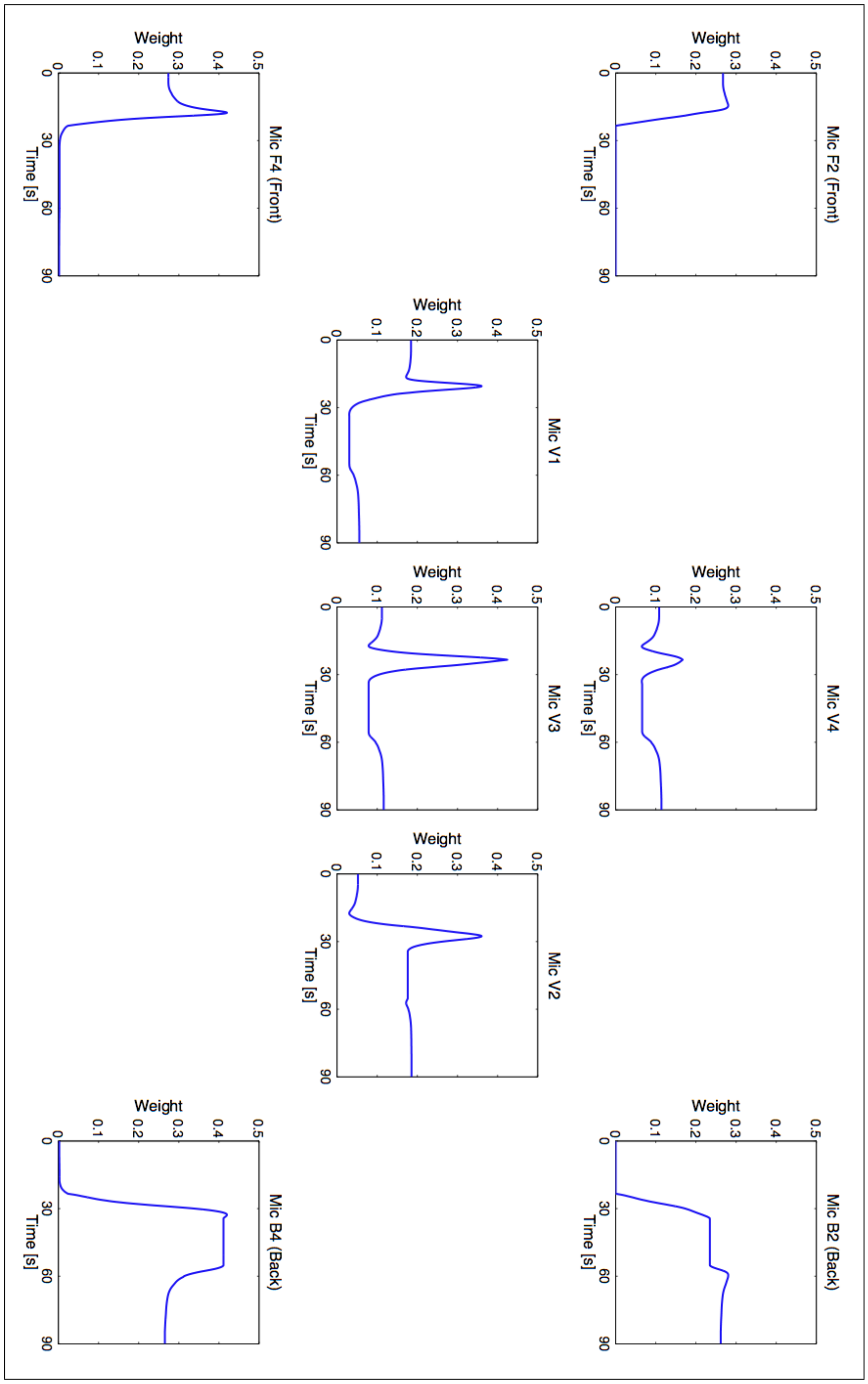

Figure 5.6.: Microphone amplitude weights as functions of time.

Emitted frequency, f_0	100 Hz	1,000 Hz	10,000 Hz
Observed frequency, f (approaching listener)	103 Hz	1025 Hz	10,251 Hz
Observed frequency, f (driving away from listener)	98 Hz	976 Hz	9,761 Hz

Table 5.1.: Comparison of emitted and observed frequencies using the Doppler effect.

position. This effect would obviously increase in driving conditions that call for greater vehicle speeds.

5.2.3. Additional Auralized Sounds

Aside from the primary synthesized engine sounds, there were a few other sounds that needed to either be created or modified in some way in order to finish the sounds that accompany the animations.

Rolling Waste Bin

The first additional sound was for a rolling waste bin. There are two variations of this sound: one treated rolling bin with softer wheels and one rolling untreated bin that had previously been in use in the city of Gothenburg. A still image from the animation of the untreated bin being rolled out towards the conventional truck is shown in Figure 5.7.

For each of these waste bins, recordings were provided by the Volvo Group of the bins being rolled over a rough net which was used to simulate the effect of a cobblestone street in a laboratory setting. Due to the short duration of the recordings, these sounds were looped using Audacity audio editing software in order to provide an adequately long sound.

The relative levels of the rolling bins were treated using the inverse square law and a free-field assumption. This equation is given by:

$$\Delta L_p = 10log_{10}\left(\frac{r_2^2}{r_1^2}\right) = 20log_{10}\left(\frac{r_2}{r_1}\right)$$

where r_2 is the distance from the rolling waste bin to the listening position and r_1 is the distance from the listening position to the middle of the street where the truck stops to load the waste. The listening position was assumed to have roughly the same height and lateral position as the waste bin, so the distance varies only in the y-direction. The listener is 5.5 m from the middle of the street, and the waste bin travels a total of 13.5 m, which is comprised of 7 m through a tunnel within the apartment building and another

Figure 5.7.: Rolling waste bin towards conventional truck on Landsvägsgatan [2].

6.5 m from the building facade to the middle of the street. This gives a maximum relative loss due to distance of:

$$\Delta L_p = 20log_{10}\left(\frac{13.5}{5.5}\right) = 7.8\ dB.$$

Linear interpolation was used to approximate the losses due to distance for intermediate bin positions between the starting and end positions. Additionally, a small amount of reverberation was applied to the signal in Audacity when the bins were in the tunnel in order to roughly approximate the some of the reflections that occur there due to the compact geometry. This entire procedure was essentially repeated in reverse after the bin was emptied into the truck and rolled from the middle of the street back into the tunnel and toward the building's interior courtyard.

Waste Dumping

The final sounds to complete the videos consisted of waste dumping from the bin into each of the two garbage trucks. For the hybrid truck, a quieter solution was demonstrated by dumping a load of soft, cut-up tires. The loading of the conventional truck was louder due to the fact that its load contained hard plastic pipes that created a substantial impact sound when they contacted the interior of the truck.

Both of the waste dumping sounds were obtained from prior measurements taken at the Volvo Group. Aside from a small amount of compression for the louder load of plastic piping, the only editing that was done to these sounds was cutting the sound down to the proper length of time for each video. A still image of the rear-loading waste collection process for the hybrid truck is seen in Figure 5.8.

Figure 5.8.: Dumping waste into hybrid truck on Landsvägsgatan [2].

6. Conclusions

6.1. Summary of Results

Granular synthesis was used to auralize constant speed and accelerations propulsion noise of a heavy-duty truck with a hybrid engine. The constant speed auralizations were especially effective; a group of listening test participants were unable to differentiate between recorded and synthesized sounds for eight out of ten pairs of sounds of the hybrid truck's electric motor. The two pairs of sounds that were detected as different (significant for $p < .03$) were for the two microphone positions of the highest tested vehicle speed of 30 km/h. These results are comparable to the results of a previous listening test that evaluated a similar heavy-duty truck's combustion engine using granular synthesis, where listeners were not able to differentiate between nine of ten constant speed pairs of sounds.

The acceleration auralizations were evaluated in a listening test using semantic differentials for both the electric motor and combustion engine. The listeners found the realism of the synthesized combustion engine accelerations to be closer to reference recordings as compared to the realism ratings of the electric motor accelerations relative to its reference recordings. Annoyance, pleasantness and activation are three other attributes that were also evaluated with semantic differentials and analyzed using a repeated measures ANOVA. For each of these attributes, various factors such as the vehicle's propulsion source and driving gear were found to be significant effects on the listeners' perception of the synthesized sounds.

Auralizations were also created to accompany video animations of conventional and hybrid refuse collection vehicles operating on a city street in Gothenburg. The propulsion noise sources were modeled using synthesis from grain databases that were detected from reference recordings provided by the Volvo Group. The auralizations also took the propagation effect of distance and the Doppler effect on frequency into account as the vehicle's position on the street changed relative to a stationary listening position. Shepard's inverse distance weighting method was implemented to mix eight microphone channels together with variable amplitudes based on the truck's position as a function of time, thereby modeling the varying directionality of the source. In addition to the granularly synthesized propulsion noise, auralizations were also created for waste dumping and rolling waste bins. Overall, the auralizations were well received by project partners and stakeholders, and this work is an especially good example of the possible applications and capabilities of a granular model.

6.2. Suggestions for Further Work

There are some notable improvements that could be made to the current granular model that could, at least hypothetically, increase the overall quality of the grain databases and synthesized sounds.

6.2.1. Source Recordings

New source recordings of the hybrid truck's driving cycles would be quite useful for producing higher quality grain databases and auralizations. The original measurements were not made with the granular synthesis application in mind, so there is a somewhat limited amount of data that is available for analysis, especially for the electric motor.

A broader range of accelerations, including both fast and slow accelerations, would be valuable in particular. A fast acceleration through a gear, typically around 3–5 s in length, tends to sound more violent and aggressive and has a higher torque compared to slower accelerations, which can take over 10 s to complete. If data could be gathered in a systematic manner to span the entire range of possible accelerations, then it is plausible that the model could produce a higher quality, more realistic synthesized sound. Additionally, the granular model would be applicable to a wider range of driving conditions if more reference recordings were incorporated into the grain databases.

6.2.2. Grain Detection

As far as the grain detection algorithm is concerned, it would likely be highly advantageous to incorporate torque tracking into the acceleration detection process, in addition to engine speed tracking. As discussed above, in some cases the variable length accelerations have vastly different noise characteristics and torque. Even currently with a limited amount of acceleration measurement data, torque is not accounted for in any way in the algorithms. One way to accomplish this would be to sort grains into different torque bins, i.e. low, medium, high torque scenarios, so that all grains in a given database have a similar noise character, in addition to the appropriate grain size which is based on engine speed.

The acceleration algorithm could also be refined to automatically detect "bad" grains. A "bad" grain is classified as a detected grain that has some form of unwanted noise in the recorded signal and therefore produces a lower quality sound when it is chosen for use by the synthesis algorithm. With additional work and research into digital signal processing methods, it should be possible to detect a "bad" grain when the engine signal-to-unwanted noise ratio is especially poor, and thus not save it in the corresponding database. Currently, every grain within a user-specified time interval is automatically detected and saved to the database, so subjective listening is in effect the only means of grain quality control when determining which time intervals to detect. As discussed in the analysis of the synthesized sounds, it only takes one bad grain to create a prominent discontinuity in a synthesized sound, and noticeable discontinuities can cause a listener's sense of perceived realism to decrease dramatically, as was observed in the case of the evaluated electric motor acceleration auralizations.

6.2.3. Elector Motor Synthesis

Since the character of the electric motor noise is fundamentally different from the combustion engine noise, it may be beneficial to entirely change the way that the grains are synthesized for the electric motor. The combustion engine noise spectrum contains more low frequencies with many higher order harmonics, and its character is much rougher and complex than the electric motor's in general. The electric motor noise spectrum tends to be more broadband and typically centers around frequencies between 500–1000 Hz.

For electric motor accelerations, it could be possible to use the granular approach to synthesize the low frequency noise and then use bandlimited noise that tracks the speed of the motor to auralize the middle and high frequencies. Ideally, this would help to reduce the discontinuities observed at higher frequencies and create an overall "smoother" auralized sound. However, to accomplish this in reality it would first be necessary to gain a better level of understanding of the relationship between electric motor speed and radiated sound power from higher order frequencies. In previous work, the relationship between these two quantities was not found to be linear. Instead, the frequency range of the radiated sound power likely depends on specific properties of the electric motor, such as magnetic field strength and torque.

References

[1] Forssén, J., P. Andersson, P. Bergman, K. Fredriksson, and P. Zimmerman. *Auralization of truck engine sound - preliminary results using a granular approach*. Division of Applied Acoustics, Chalmers University of Technology and Volvo Group. Göteborg, Sweden, 2012.

[2] Forzelius, J. *Landsvägsgatan Refuse Collection Vehicle Animations*. Visual Arena Research. Göteborg, Sweden, September, 2014.

[3] Fredriksson, K, M. Wolkesson, and K. Bodlund. *Engineering Report: 633916 - Indoor sound power level measurements of hybrid and conventional refuse collection vehicles*. Volvo Group. Göteborg, Sweden, November, 2013.

[4] Jagla, J., J. Maillard, and N. Martin. *Sample-based engine noise synthesis using an enhance pitch-synchronous overlap-and-add method*. The Journal of the Acoustical Society of America. 132(5), November, 2012.

[5] Kortekaas, R. *Physiological and psychoacoustical correlates of perceiving natural and modified speech*. TU, Eindhoven, Holland, 1997.

[6] Maillard, J., and J. Jagla. *Real Time Auralization of Non-Stationary Traffic Noise: Quantitative and Perceptual Validation in an Urban Street*. Centre Scientifique et Technique du Batiment, Grenoble, France. Presented at AIA-DAGA in Merano, Italy, March, 2013.

[7] Montgomery, D., and G. Runger. *Applied Statistics and Probability for Engineers, Sixth Edition*. Hoboken, NJ: John Wiley & Sons, 2014.

[8] Mulgrew, B., P. Grant, and J. Thompson. *Digital Signal Processing: Concepts and Applications, Second Edition*. New York, NY: Palgrave Macmillan, 2003.

[9] Nilsson, M., J. Forssén, P. Lunden, A. Peplow, and B. Hellström. *LISTEN Auralization of Urban Soundscapes*. Stockholm University, Chalmers University of Technology, Sonic Studio, KTH Royal Institute of Technology, University College of Arts, Crafts and Design, September, 2011.

[10] Oppenheim, A., R. Schafer, and J. Buck. *Discrete-Time Signal Processing*. Upper Saddle River, NJ: Prentice Hall, 1999.

[11] Pendharkar, C. *Auralization of road vehicles using spectral modeling synthesis*, Master's Thesis. Division of Applied Acoustics, Chalmers University of Technology, Göteborg, Sweden, 2012.

[12] Roads, C. *Introduction to granular synthesis*. Computer Music Journal, 1990.

[13] Rosen, J., and L. Gothard. *Encyclopedia of Physical Science: the Doppler Effect*. Infobase Publishing, 2009.

[14] Schnell, N., G. Peeters, S. Lemouton, P. Manoury, and X. Rodet. *Synthesizing a choir in real-time using Pitch Synchronous Overlap Add (PSOLA)*. IRCAM - Centre Georges-Pompidou, Paris, France, 2002.

[15] Shepard, D. *A two-dimensional interpolation function for irregularly spaced data*. Harvard College, Cambridge, Massachusetts, 1968.

[16] Weisstein, E. *Hanning Function*. MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/HanningFunction.html. Accessed October, 2014.

[17] Wuensch, Karl. *Testing Hypotheses with the Binomial Probability Distribution*. East Carolina University, 2012.

[18] Zwicker, E., and H. Fastl. *Psychoacoustics: Facts and Model, Second Edition*. Munich, Germany: Springer-Verlag, 1990.

A. Description of Measurements

Case	Date	Truck	Propulsion	Vehicle mass	Cycle Type
Ba	2013-09-20	Hybrid	Engine	20000 kg	City cycle
Bb	2013-09-20	Hybrid	Engine	20000 kg	City cycle
Bd	2013-09-20	Hybrid	Engine	20000 kg	City cycle
Ca	2013-09-20	Hybrid	Engine	15000 kg	City cycle
Cb	2013-09-20	Hybrid	Engine	15000 kg	City cycle
Cc	2013-09-20	Hybrid	Engine	15000 kg	City cycle
Ea	2013-09-23	Hybrid	Electric	20000 kg	RCV cycle
Eb	2013-09-23	Hybrid	Electric	20000 kg	RCV cycle
Ec	2013-09-23	Hybrid	Electric	20000 kg	RCV cycle
Fb	2013-09-23	Hybrid	Electric	20000 kg	Constant speed 30 km/h
Fc	2013-09-23	Hybrid	Electric	20000 kg	Constant speed 25 km/h
Ja	2013-09-24	Hybrid	Engine	15000 kg	Constant speeds
Jb	2013-09-24	Hybrid	Engine	15000 kg	RCV cycle
Jc	2013-09-24	Hybrid	Engine	15000 kg	RCV cycle
Jd	2013-09-24	Hybrid	Engine	15000 kg	Drive-by accel.
Ka	2013-09-24	Hybrid	Engine	20000 kg	RCV cycle
Kb	2013-09-24	Hybrid	Engine	20000 kg	RCV cycle
Kc	2013-09-24	Hybrid	Engine	20000 kg	Drive-by accel.
Kd	2013-09-24	Hybrid	Engine	20000 kg	Constant speeds
La	2013-09-24	Hybrid	Electric	20000 kg	RCV cycle
Lb	2013-09-24	Hybrid	Electric	20000 kg	RCV cycle
Lc	2013-09-24	Hybrid	Electric	20000 kg	Constant speed 30 km/h
Ld	2013-09-24	Hybrid	Electric	20000 kg	Constant speed 25 km/h
Ma	2013-09-24	Hybrid	Electric	15000 kg	RCV cycle
Mb	2013-09-24	Hybrid	Electric	15000 kg	RCV cycle
Mc	2013-09-24	Hybrid	Electric	15000 kg	Constant speed 30 km/h
Md	2013-09-24	Hybrid	Electric	15000 kg	Constant speed 25 km/h

Table A.1.: List of recordings and measurement details for the Volvo FE Hybrid truck.

Case	Date	Truck	Propulsion	Vehicle mass	Cycle Type
Na	2013-10-01	Conventional	Engine	15000 kg	City cycle
Nb	2013-10-01	Conventional	Engine	15000 kg	City cycle
Nc	2013-10-01	Conventional	Engine	15000 kg	City cycle
Oa	2013-10-01	Conventional	Engine	20000 kg	City cycle
Ob	2013-10-01	Conventional	Engine	20000 kg	City cycle
Oc	2013-10-01	Conventional	Engine	20000 kg	City cycle
Pa	2013-10-02	Conventional	Engine	20000 kg	RCV cycle
Pb	2013-10-02	Conventional	Engine	20000 kg	RCV cycle
Qa	2013-10-03	Conventional	Engine	20000 kg	Constant speeds
Ra	2013-10-03	Conventional	Engine	20000 kg	Drive-by accel.
Rb	2013-10-03	Conventional	Engine	20000 kg	Drive-by accel.
Sa	2013-10-03	Conventional	Engine	15000 kg	RCV cycle
Sb	2013-10-03	Conventional	Engine	15000 kg	RCV cycle
Ta	2013-10-03	Conventional	Engine	15000 kg	Drive-by accel.
Ua	2013-10-03	Conventional	Engine	15000 kg	Constant speeds

Table A.2.: List of recordings and measurement details for the conventional Volvo FE truck.

B. Measurement Microphone Positions

Index	Label	x[m]	y[m]	z[m]
1	B1	-4.68	0.00	2.53
2	B2	-4.68	3.30	5.06
3	B3	-4.68	-3.30	5.06
4	B4	-4.68	3.30	1.20
5	B5	-4.68	0.00	1.20
6	B6	-4.68	-3.30	1.20
7	F1	7.32	0.00	2.53
8	F2	7.32	3.30	5.06
9	F3	7.32	-3.30	5.06
10	F4	7.32	3.30	1.20
11	F5	7.32	0.00	1.20
12	F6	7.32	-3.30	1.20
13	H1	-1.68	-3.30	2.53
14	H2	4.32	-3.30	2.53
15	H3	1.32	-3.30	5.06
16	H4	1.32	-3.30	2.53
17	U1	3.67	0.00	5.06
18	U2	-3.63	0.00	5.06
19	V1	4.32	3.30	2.53
20	V2	-1.68	3.30	2.53
21	V3	1.32	3.30	2.53
22	V4	1.32	3.30	5.06
23	Left Side 1	7.05	7.5	1.20
24	Left Side 2	5.05	7.5	1.20
25	Left Side 3	3.05	7.5	1.20
26	Left Side 4	1.05	7.5	1.20
27	Left Side 5	-0.95	7.5	1.20
28	Left Side 6	-2.95	7.5	1.20
29	Left Side 7	-4.95	7.5	1.20
30	Left Side 8	-6.95	7.5	1.20
31	Right Side 1	7.05	-7.5	1.20
32	Right Side 2	5.05	-7.5	1.20
33	Right Side 3	3.05	-7.5	1.20
34	Right Side 4	1.05	-7.5	1.20
35	Right Side 5	-0.95	-7.5	1.20
36	Right Side 6	-2.95	-7.5	1.20
37	Right Side 7	-4.95	-7.5	1.20
38	Right Side 8	-6.95	-7.5	1.20

Table B.1.: Measurement microphone labels and coordinates.

C. Listening Test Questions & ANOVA Tables

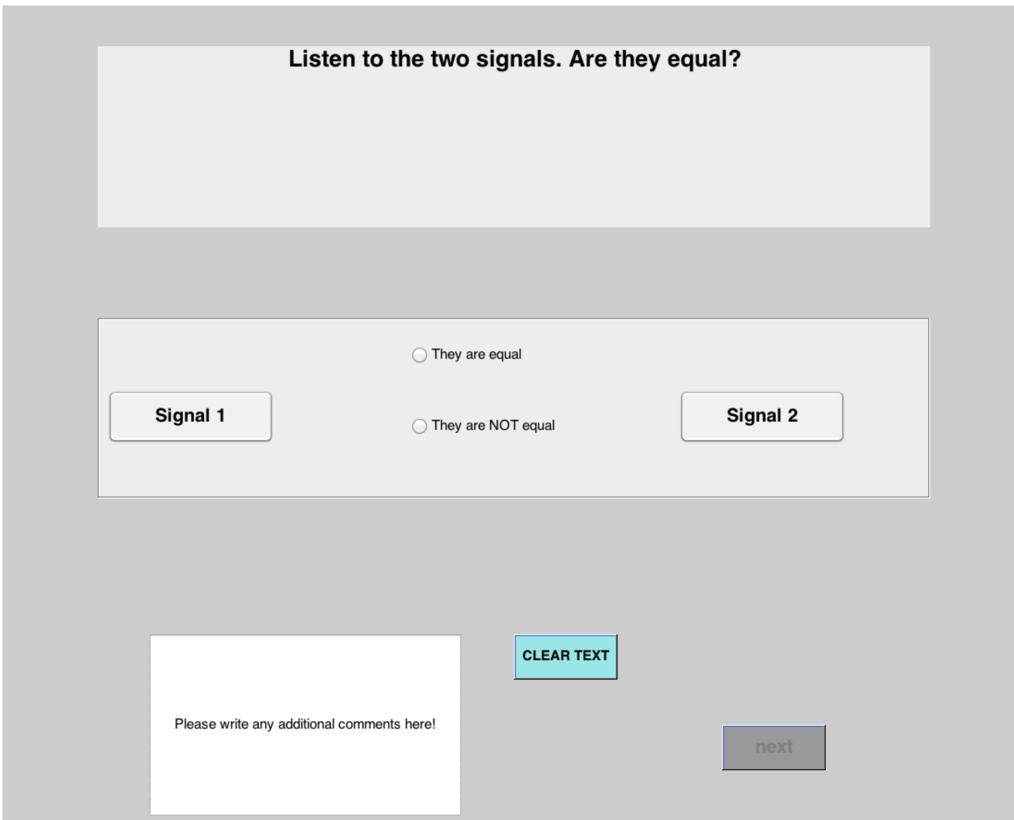

Figure C.1.: Graphical user interface for same-different discrimination listening test.

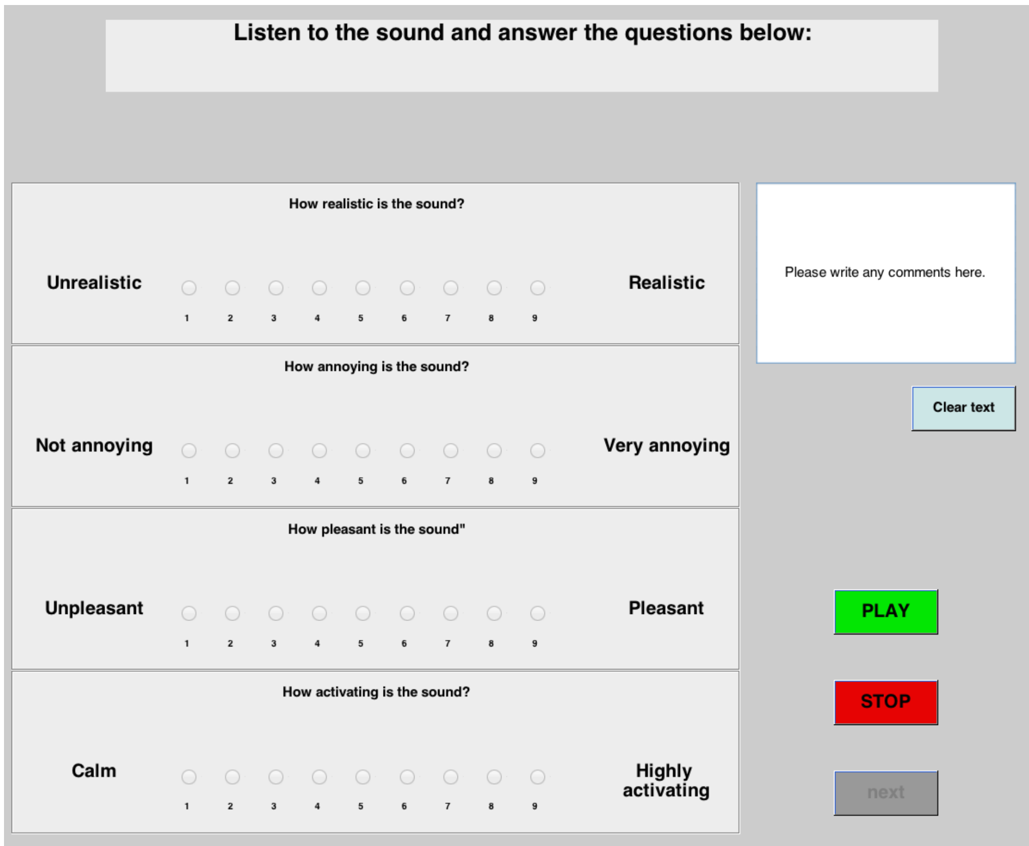

Figure C.2.: Graphical user interface for semantic differentials listening test.

Property	**Effect**	$F(df_{effect}, df_{error})$	p	$\eta^2_{partial}$
Realism	Gear	$F(1,15) = 11.765$	.004	.440
Realism	Time	$F(3,45) = .596$	.621	.038
Realism	Gear*Time	$F(3,45) = .574$	.635	.037
Annoyance	Gear	$F(1,15) = 9.698$	.007	.393
Annoyance	Time	$F(3,45) = .375$	.772	.024
Annoyance	Gear*Time	$F(3,45) = .348$	.791	.023
Pleasantness	Gear	$F(1,15) = 5.714$	.030	.276
Pleasantness	Time	$F(3,45) = 3.172$	.033	.175
Pleasantness	Gear*Time	$F(3,45) = .545$	.654	.035
Activation	Gear	$F(1,15) = 7.061$	.018	.320
Activation	Time	$F(3,45) = .486$	.693	.031
Activation	Gear*Time	$F(3,45) = 1.055$	.378	.066

Table C.1.: ANOVA table for two gears and length of synthesized sounds (2, 4, 6 and 8 s), combustion engine.

Property	**Effect**	$F(df_{effect}, df_{error})$	p	$\eta^2_{partial}$
Realism	Gear	$F(1,15) = 8.136$	.012	.352
Realism	Time	$F(3,45) = 2.360$	.084	.136
Realism	Gear*Time	$F(3,45) = .504$	.681	.033
Annoyance	Gear	$F(1,15) = 6.160$	.025	.291
Annoyance	Time	$F(3,45) = .556$	.647	.036
Annoyance	Gear*Time	$F(3,45) = 1.550$	.215	.094
Pleasantness	Gear	$F(1,15) = 12.638$	.003	.457
Pleasantness	Time	$F(3,45) = .482$	.697	.031
Pleasantness	Gear*Time	$F(3,45) = .557$	.646	.036
Activation	Gear	$F(1,15) = 3.046$	.101	.169
Activation	Time	$F(3,45) = 2.352$	.085	.136
Activation	Gear*Time	$F(3,45) = .176$	.912	.012

Table C.2.: ANOVA table for two gears and length of synthesized sounds (2, 4, 6 and 8 s), electric motor.

Property	**Effect**	$F(df_{effect}, df_{error})$	p	$\eta^2_{partial}$
Realism	Propulsion	$F(1,15) = 58.059$	.000	.795
Realism	Time	$F(3,45) = 3.064$	.037	.170
Realism	Propulsion*Time	$F(3,45) = .260$	.854	.017
Annoyance	Propulsion	$F(1,15) = .000$	.983	.000
Annoyance	Time	$F(3,45) = .392$	.759	.025
Annoyance	Propulsion*Time	$F(3,45) = .502$	.683	.032
Pleasantness	Propulsion	$F(1,15) = .359$	.558	.023
Pleasantness	Time	$F(3,45) = 2.427$	.078	.139
Pleasantness	Propulsion*Time	$F(3,45) = .553$	.649	.036
Activation	Propulsion	$F(1,15) = .711$	.412	.045
Activation	Time	$F(3,45) = 1.279$	.293	.079
Activation	Propulsion*Time	$F(3,45) = 1.043$	.383	.065

Table C.3.: ANOVA table for propulsion type (combustion engine and electric motor) and length of synthesized sounds (2, 4, 6 and 8 s).

Property	**Effect**	$F(df_{effect}, df_{error})$	p	$\eta^2_{partial}$
Realism	Propulsion	$F(1,15) = 24.153$	.000	.617
Realism	SoundType	$F(1,15) = 48.867$	.000	.765
Realism	Propulsion*SoundType	$F(1,15) = 32.904$	.000	.687
Annoyance	Propulsion	$F(1,15) = 6.555$	.022	.304
Annoyance	SoundType	$F(1,15) = 25.058$	.000	.626
Annoyance	Propulsion*SoundType	$F(1,15) = 15.943$	.001	.515
Pleasantness	Propulsion	$F(1,15) = 8.606$	.010	.365
Pleasantness	SoundType	$F(1,15) = 7.507$	.015	.334
Pleasantness	Propulsion*SoundType	$F(1,15) = 11.812$	.004	.441
Activation	Propulsion	$F(1,15) = .328$	.575	.021
Activation	SoundType	$F(1,15) = 4.095$	.061	.214
Activation	Propulsion*SoundType	$F(1,15) = 2.258$	.393	.049

Table C.4.: ANOVA table for propulsion type (combustion engine and electric motor) and type of sound (recorded and synthesized).

D. Sample MATLAB Code

All MATLAB code developed in cooperation with Jens Forssén, Chalmers University of Technology.

Grain Detection (extract)

```
gear_idx = 1;      % driving condition
mic2_idx = 1;      % microphone #2 index
micref_idx = 11;   % reference microphone (mic F5)

load engine_data
load mic_distance_data

load Data/Gear1/CaseXX_gear1_mic_F5.mat
s1_ref = load_data;
load(strcat('Data/Gear1/CaseXX_gear1_mic_',char(lbm(mic2_idx)),'.mat'))
s2 = load_data;

T = length(s1_ref)/FS-(mic_distance_data.N_shift(mic2_idx, gear_idx)+1)*dt;
t = 0:dt:T-dt;
N = length(t);

% shift signal 2 based on previous cross-correlation calculation
idx_s1 = 1:N;
if mic_distance_data.shift_direction(mic2_idx, gear_idx) == 0
s1_ref = s1_ref(idx_s1);
s2 = s2((idx_s1)+mic_distance_data.N_shift(mic2_idx, gear_idx));

% high-pass filter signals
f_high = 20;
Wn = f_high/FS*2;
b = fir1(1024, Wn, 'high');
s1_ref_filt_high = filtfilt(b,1,s1_ref);
s2_filt_high = filtfilt(b,1,s2);

%% set up grain

% over 1 cylinders (1/6th of the total length)
N_wavelet = round(engine_N_wavelet(round(load_start_n/FS*FS_can)));
```

```
% N_searcharea is +/- samples to search for grain
N_searcharea = ceil(0.55*N_wavelet);
% (6.5*N_wavelet) to have margin for the tails
N_skipatstart = round(6.5*N_wavelet);
N_skipatend = round(N_skipatstart);

n_tmp = round(N_skipatstart);
idx_tmp_start = 1;

% grain detection
while (n_tmp + 2*N_searcharea + 1 + N_skipatend) < N

    % wavelet over 1 cylinder (1/6th of the total length)
    N_wavelet = round(engine_N_wavelet(round((load_start_n+ ...
        idx_tmp_start)/FS*FS_can)));

    Ni = ceil(0.05*N_wavelet); % search within +/- 5 % of period
    N_search_samples = 2*N_searcharea + 1;
    N_search_period = 2*Ni + 1;

    s_6_tmp = s1_ref_filt_high((n_tmp+1-N_searcharea):round((n_tmp+...
        wavelet_n*N_wavelet+N_searcharea)));

    correlation_vector=zeros(N_search_period,2); % value, idx

    % account for grain compression or stretching
    for ni = -Ni:Ni
        nii = ni + Ni + 1; % index starting at 1
        N_tmp = N_wavelet + ni;

        peakshape_1_tmp = peakshape_jf3(N_tmp)*0.1;
        peakshape_6_tmp = [peakshape_1_tmp; peakshape_1_tmp; ...
            peakshape_1_tmp; peakshape_1_tmp; ...
            peakshape_1_tmp; peakshape_1_tmp];
        hann_wind_big=hann(6*N_tmp,'periodic'); % window entire grain

        % convolve high-passed signal with peak shape
        conv_6_tmp = conv(s_6_tmp, flipud(peakshape_6_tmp.* ...
            hann_wind_big),'valid')/(6*N_tmp);
        % find first maximum
        idx_first_max_tmp = find(diff(sign(diff(conv_6_tmp))) == -2, 1) + 1;
        max_6_val_tmp = conv_6_tmp(idx_first_max_tmp);

        correlation_vector(nii, 1) = max_6_val_tmp;
        correlation_vector(nii, 2) = idx_first_max_tmp;
    end

    [val_tmp, idx_correlation_vector_max] = max(correlation_vector(:, 1));
```

```
    nii = idx_correlation_vector_max;
    ni = nii - Ni - 1;
    N_wavelet_new = N_wavelet + ni;

    idx_new_max = correlation_vector(nii,2);
    idx_tmp_start = n_tmp + idx_new_max;
    idx_tmp_end = idx_tmp_start + 6*N_wavelet_new - 1;

    s_grain_tmp = s1_ref_filt_high(idx_tmp_start:idx_tmp_end);
    S(n_grain).core = s_grain_tmp;  % core = 6 ignitions
    s_grain_tmp_B1 = s2_filt_high(idx_tmp_start:idx_tmp_end);
    S_2(n_grain).core = s_grain_tmp_B1;

    N_tail = 6*N_wavelet_new;

    % define left and right tails outside of grain core
    S(n_grain).left = s1_ref_filt_high(idx_tmp_start-N_tail:idx_tmp_start-1);
    S(n_grain).right = s1_ref_filt_high(idx_tmp_end+1:idx_tmp_end+N_tail);
    S_2(n_grain).left = s2_filt_high(idx_tmp_start-N_tail:idx_tmp_start-1);
    S_2(n_grain).right = s2_filt_high(idx_tmp_end+1:idx_tmp_end+N_tail);

    N_wavelet = N_wavelet_new;
    n_tmp = n_tmp + 6*N_wavelet_new;
    n_grain = n_grain + 1

end

save('database_name.mat', 'S', 'S_2')
```

Grain Synthesis: Constant Speed (extract)

```
function [s1_ref_synth_tot, s2_synth_tot, tsynth] = ...
    constant_speed_synth(S, S_2, Tsynth, dt, n_rand)

Nsynth = ceil(Tsynth/dt);
N_taper = 128; % number of samples overlap

n = 1;
n_grain = n_rand(n); % choose a random grain

s1_ref_synth = [S(n_grain).left; S(n_grain).core];  % reference microphone
s2_synth = [S_2(n_grain).left; S_2(n_grain).core]; % microphone #2
Ntmp = length(s1_ref_synth);
Ntmp2 = length(s2_synth);

while Ntmp < Nsynth || Ntmp2 < Nsynth
```

```
    n_grain_old = n_grain;
    n = n + 1;
    n_grain = n_rand(n);

    s_tmp_old = S(n_grain_old).core;
    s_tmp = S(n_grain).core;

    s_tmp_old_2 = S_2(n_grain_old).core;
    s_tmp_2 = S_2(n_grain).core;

    [w_tmp_left, w_tmp_right] = window_half_jf(N_taper, N_taper, 'hann');

    % join grain segments together
    s_tmp_right = S(n_grain_old).right;
    s_tmp_right2 = [s_tmp_old(end-N_taper/2+1:end); ...
        s_tmp_right(1:N_taper/2)].*w_tmp_right;
    s_tmp_left = S(n_grain).left;
    s_tmp_left2 = [s_tmp_left(end-N_taper/2+1:end); ...
        s_tmp(1:N_taper/2)].*w_tmp_left;
    s_tmp_overlap = s_tmp_right2 + s_tmp_left2;

    s_tmp_right_2 = S_2(n_grain_old).right;
    s_tmp_right2_2 = [s_tmp_old_2(end-N_taper/2+1:end); ...
        s_tmp_right_2(1:N_taper/2)].*w_tmp_right;
    s_tmp_left_2 = S_2(n_grain).left;
    s_tmp_left2_2 = [s_tmp_left_2(end-N_taper/2+1:end); ...
        s_tmp_2(1:N_taper/2)].*w_tmp_left;
    s_tmp_overlap_2 = s_tmp_right2_2 + s_tmp_left2_2;

    s1_ref_synth = [s1_ref_synth(1:end-N_taper/2); s_tmp_overlap; ...
        s_tmp(1+N_taper/2:end)];
    s2_synth = [s2_synth(1:end-N_taper/2); s_tmp_overlap_2; ...
        s_tmp_2(1+N_taper/2:end)];
    Ntmp = length(s1_ref_synth);
    Ntmp2 = length(s2_synth);

end

s1_ref_synth_tot = s1_ref_synth(1:Nsynth);
s2_synth_tot = s2_synth(1:Nsynth);
tsynth = (0:Nsynth-1)*dt;

return
```

Grain Synthesis: Acceleration (extract)

```
function [s1_ref_synth_tot, s2_synth_tot, tsynth, rpm_tracking, n_grain_mat] = ...
    accel_synth(desired_rpm, gear_idx, S, S_2, grainlengths, Tsynth, dt)

FS_can = 100; % CAN-signal sampling frequency
FS = 44100;    % recording sampling frequency

Nsynth = ceil(FS*Tsynth);
N_taper = 128; % number of samples overlap

RPM = desired_rpm;  % synthesis target RPM vector

[sortlist, idx_grainlength] = sort(grainlengths');

target_rpm_tmp = RPM(1);
grain_length_target_tmp = 120*FS/target_rpm_tmp;

if grain_length_target_tmp > max(sortlist)
    grain_idx_tmp = max(find(grain_length_target_tmp > sortlist));
    n_grain = idx_grainlength(grain_idx_tmp);
else
    grain_idx_tmp = min(find(grain_length_target_tmp < sortlist));
    n_grain = idx_grainlength(grain_idx_tmp);
end

sorted = sortlist(grain_idx_tmp);

s1_ref_synth = [S(n_grain).left; S(n_grain).core];
s2_synth = [S_2(n_grain).left; S_2(n_grain).core];
Ntmp = length(s1_ref_synth);
Ntmp2 = length(s2_synth);

rpm_tracking(1, 1) = 1/FS;
rpm_tracking(2, 1) = target_rpm_tmp;
n_grain_mat(1,1) = n_grain;
n_grain_mat(2,1) = sorted;
n_grain_mat(3,1) = 0;
l_count = 1;
n_grain_old2 = 0;

while Ntmp < Nsynth || Ntmp2 < Nsynth

    Ntmp_old = Ntmp;
    target_rpm_old = target_rpm_tmp;

    if l_count >= 2
        n_grain_old2 = n_grain_old;
    end

    n_grain_old = n_grain;
```

```
sorted_old = sorted;

target_rpm_tmp = RPM(round((1+Ntmp)*FS_can/FS));
grain_length_target_tmp = 120*FS/target_rpm_tmp;

if grain_length_target_tmp > max(sortlist)
    grain_idx_tmp = max(find(grain_length_target_tmp > sortlist));
    n_grain = idx_grainlength(grain_idx_tmp);
    sorted = sortlist(grain_idx_tmp);
else
    grain_idx_tmp = min(find(grain_length_target_tmp < sortlist));
    n_grain = idx_grainlength(grain_idx_tmp);
    sorted = sortlist(grain_idx_tmp);
end

    % check for AA repetition
    if n_grain_old == n_grain && l_count > 1

        if grain_idx_tmp == 0 || grain_idx_tmp == 1
           grain_idx_tmp = grain_idx_tmp + 1;
        elseif grain_idx_tmp == length(sortlist)
            grain_idx_tmp = grain_idx_tmp - 1;
        else
            if target_rpm_tmp >= target_rpm_old
                grain_idx_tmp = grain_idx_tmp - 1;
            elseif target_rpm_tmp < target_rpm_old
                grain_idx_tmp = grain_idx_tmp + 1;
            end
        end

        n_grain = idx_grainlength(grain_idx_tmp);
        sorted = sortlist(grain_idx_tmp);
    end

    % check for ABAB repetition
    if n_grain == n_grain_old2
        if grain_idx_tmp == 0 || grain_idx_tmp == 1
           grain_idx_tmp = grain_idx_tmp + 1;
        elseif grain_idx_tmp == length(sortlist)
            grain_idx_tmp = grain_idx_tmp - 1;
        else
            grain_idx_tmp = grain_idx_tmp + sign(randn(1));
        end
        n_grain = idx_grainlength(grain_idx_tmp);
    end

s_tmp_old = S(n_grain_old).core;
s_tmp = S(n_grain).core;
```

```
    s_tmp_old_2 = S_2(n_grain_old).core;
    s_tmp_2 = S_2(n_grain).core;

    [w_tmp_left, w_tmp_right] = window_half_jf(N_taper, N_taper, 'hann');

    % join grain segments together
    s_tmp_right = S(n_grain_old).right;
    s_tmp_right2 = [s_tmp_old(end-N_taper/2+1:end); ...
        s_tmp_right(1:N_taper/2)].*w_tmp_right;
    s_tmp_left = S(n_grain).left;
    s_tmp_left2 = [s_tmp_left(end-N_taper/2+1:end); ...
        s_tmp(1:N_taper/2)].*w_tmp_left;
    s_tmp_overlap = s_tmp_right2 + s_tmp_left2;

    s_tmp_right_2 = S_2(n_grain_old).right;
    s_tmp_right2_2 = [s_tmp_old_2(end-N_taper/2+1:end); ...
        s_tmp_right_2(1:N_taper/2)].*w_tmp_right;
    s_tmp_left_2 = S_2(n_grain).left;
    s_tmp_left2_2 = [s_tmp_left_2(end-N_taper/2+1:end); ...
        s_tmp_2(1:N_taper/2)].*w_tmp_left;
    s_tmp_overlap_2 = s_tmp_right2_2 + s_tmp_left2_2;

    s1_ref_synth = [s1_ref_synth(1:end-N_taper/2); s_tmp_overlap; ...
        s_tmp(1+N_taper/2:end)];
    s2_synth = [s2_synth(1:end-N_taper/2); s_tmp_overlap_2; ...
        s_tmp_2(1+N_taper/2:end)];

    Ntmp = length(s1_ref_synth);
    Ntmp2 = length(s2_synth);
    l_count = l_count + 1;

    % save target engine speed and selected grain for each iteration
    rpm_tracking(1, l_count) = (Ntmp_old+1)/FS;
    rpm_tracking(2, l_count) = target_rpm_tmp;
    n_grain_mat(1, l_count) = n_grain;
    n_grain_mat(2, l_count) = sorted;
    n_grain_mat(3, l_count) = n_grain_old2;

end

s1_ref_synth_tot = s1_ref_synth(1:Nsynth);
s2_synth_tot = s2_synth(1:Nsynth);
tsynth = (0:Nsynth-1)*dt;

return
```