Politecnico di Torino

Department of Control and Computer Engineering

Master of Science in Mechatronic Engineering

Master's Degree Thesis

Engine sound enhancement and customization based on granular synthesis

Academic supervisor: Luciano Lavagno

Company supervisor: Daniele Marenco

Candidate: Silvia Falchi

October 2020

Abstract

Everyday car makers face important challenges in the automotive design and development. One of these is represented by the absence of acoustic feedback from the electric vehicles, leading to reduced safety for road users and worse driving experience, as engine sound is essential for taking decisions when traveling without the need for looking at the rev and speed indicators. Another core topic in the automotive field is the growing request for customization of the motor sound, aiming at making it more powerful and attractive for car users.

This thesis work, developed in collaboration with the Electronics division of Marelli Europe S.p.A. located in Venaria Reale (TO), proposes a solution to emphasize the motor sound but also to create it in the case of noise-free vehicles, using a single recording of a thermal engine that identifies the vehicle acoustic fingerprint.

After an overview on acoustics, the state of the art about systems which address car sound generation and exaltation is presented. Later, a model stating the relationship between perceived engine sound, car status parameters and noise sources is built up. Then, the practical activities performed to acquire a sound recording of an internal combustion engine along with the corresponding vehicle state variables (mainly engine speed, vehicle speed, engaged gear and acceleration) from the CAN bus are exposed.

The designed procedure involves first the analysis of the audio recording and the extraction of its harmonic content from the stochastic noise sources that

surround it. The resulting signal represents the input for the developed algorithm, granular synthesis based, which addresses the generation of a new sound depending on the operating conditions of the engine, identified by new vehicle state variables profiles, acquired in real-time from the on-board network.

Sound customization is achieved by modulating independently the power content of each harmonic contribution of the audio signal, producing original sound textures.

The presented solution provides quite pleasant and realistic engine sounds that can be played in the cabin and outside the vehicle.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Urban mobility continuously evolves, providing car makers with new challenges. Nowadays the automotive field has to deal with important and topical issues, some of which are air pollutant emissions, safety on the road, vehicle performance and users requests.

One of its rising business is currently represented by electric vehicles, whose diffusion is experiencing a significant growth, especially due to the reduced impact they have on the environment with respect to the ones equipped with a thermal engine.
Such diffusion has to face with law obligations, imposing these vehicles to be equipped with an acoustic system to make them audible from road users, as the electric motor noise is almost inappreciable. This aims at reducing the probability of accidents, especially ones at low speed, generally involving pedestrians and riders.

Another core topic on which automotive business focuses on is the vehicle customization: it's common to see car users modifying various aspects of their

own vehicles, so to make them unique and closer to their own expectations. Such changes can be either related to the chassis or to the car sound. Generally, modifications in the vehicle acoustic system are performed so to enhance the timbre of the motor and to achieve a more powerful sound.
Furthermore, since nowadays cars are more widely used with respect to other transport methods, improving the driving experience is an important point on which car makers focuses on when designing a car. For this purpose, usually *Active Sound Design*[1] is employed. In particular, Active Sound Design adopts a technique referred to as *Engine Sound Enhancement*, which can be suitable for both the electric vehicles needs and the will of customization. Details about it are provided in the following.

1.1 Overview on electric vehicles

Electric vehicles feature a portable energy storage and an Electric Motor (EM) which provides the traction effort.
Different electric vehicles can be distinguished, according to their degree of hybridization[2], as shown in Figures 1.1a - 1.1d (available at [3]):

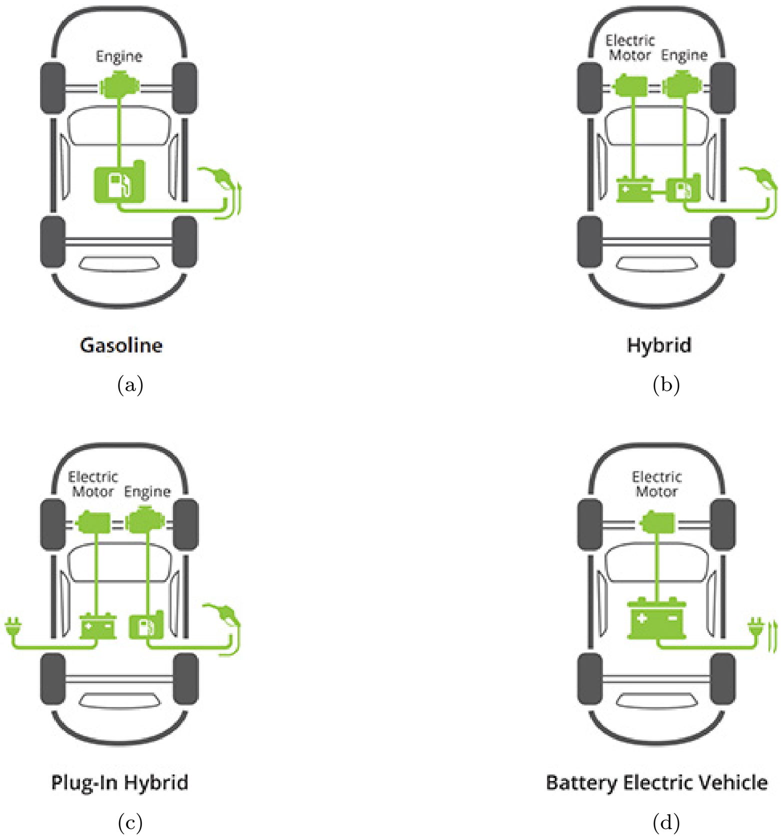


Figure 1.1: Vehicles classification

- *Conventional vehicles*

  They have an Internal Combustion Engine (ICE), which generally uses petrol or diesel as fuel.

- *Hybrid Electric Vehicles (HEVs)*

  They have both an ICE and an EM, working in mutual exclusion: as long as the HEV works at low speed and loads, it is powered by the electric motor, then, the ICE takes over as they increase. The work of the two motors is governed by an electronic control unit, which aim is to optimize the driving conditions and, thus, the consumption[2].

  The battery recharge exploits either the ICE or the so called *Regenera-*

tive Braking (RB): basically, the braking action is exploited to recharge the battery, exploiting the heat dissipated during the maneuver.

According to [4] and [5], they can be further divided in:

  - *Micro Hybrid Vehicles*

    They exploit a Start&Stop mechanism to keep at bay consumption at idle and feature a limited regenerative braking capability.

  - *Mild Hybrid Vehicles*

    They have an additional feature with respect to the latter ones, that consists in using the electric motor to support the internal combustion engine in heavier operating conditions, like increased load.

  - *Full Hybrid Vehicles*

    Their advantage with respect to the formerly presented vehicles is represented by the limited propulsion capability provided by the electric motor under restrained load and low speed conditions, as battery life is quite short.

- *Plug-In Hybrid Electric Vehicles (PHEVs)*

  They present both an ICE and an EM.

  Batteries, having a longer life than the previous case, are recharged either through the RB or by means of suitable supply station[2].

- *All Electric Vehicles (EVs)*

  Also referred to as *Battery Electric Vehicles (BEVs)*, they are only equipped with an electric motor, supplied by high capacity batteries. The charge is performed by means of external supply stations[2].

Advantages in using electric vehicles[6] can be summarized as follow:

- reduction of pollutant emissions and fuel consumption;
- low running and maintenance costs;
- same or better efficiency than gasoline or Diesel engine vehicles;
- no need for fossil fuel availability.

A significant drawback in the e-mobility field is that, in general, electric vehicles produce a weaker noise than conventional ones, especially at low speeds when it's almost inappreciable. In particular:

- traffic rules requires that silent vehicles, as electric ones, shall be equipped with an acoustic system so that they can be easily heard by road users, as internal combustion engine vehicles. The absence of this acoustic signal results in reduced safety, since electric motor is totally noise-free, thus while driving it's not perceivable by other drivers, riders and pedestrians;
- the lack of motor noise results in a worse driving experience: when considering conventional vehicles, as they produce a typical sound depending mainly on velocity, gear, gas pedal pressure and engine speed, the driver can perceive the vehicle status without looking at the speed and rev indicators, thus it is essential for driving decisions, like gear shifting[7]. This

is not valid when considering electric vehicles, because the driver doesn't receive an acoustic feedback from the motor: the only sounds reaching the cabin are basically represented by the rolling of tyres on the road, the wind on the windshield and the environmental noise sources.

1.2 Systems on market

The challenges concerning the engine sound have been addressed by means of *Active Sound Design (ASD)*[1]: it's a technique that aims at creating a desired engine sound or enhancing an existing timbre. Basically, it makes use of two methods, referred to as *Engine Sound Enhancement (ESE)* and *Active Noise Cancellation (ANC)*. ESE aims at creating and exalting the car engine sound. It is often associated with ANC, whose goal is to improve the quality of the sound perceived in the cabin, canceling the tyre friction noise, the aerodynamic one on windshield and car windows and allowing the motor sound and audio from loudspeakers to emerge and to be clearly audible by passengers.
The purposes of sound design are:

- to ensure a better driving experience, giving customers the opportunity to have the engine sound they prefer on their own vehicle, that is a customized one or a completely different sound, featuring the timbre of another car; furthermore, drivers prefers to hear the motor sound in a clear and powerful way in the cabin;
- safety, equipping silent vehicles with an acoustic fingerprint, as they cannot be easily perceived by road users.

On the market several versions of the engine sound generation and exaltation system are present. Some examples are listed below:

- Acoustic Vehicle Alerting System (AVAS).

  As mentioned in [8], this system produces warning sounds to notify pedestrians the presence of the vehicle, only when the car is driven by the electric motor and solely when reversing or driving at low speed, lower than 20 km/h: as speed increases, the tyre noise becomes more appreciable and so the vehicle presence can be perceived by means of the wheels rolling on the pavement. The unique purpose of AVAS is safety, without regard for sound quality and customization.

- EngineVOX engine sound generator.

  This is a car sound synthesizer that, based mainly on the engine speed, generates the engine sound[9] almost in real-time. The sound is provided through the car speakers in the cabin and outside.

- Bose Active Sound Management.

  It's based on engine harmonic cancellation and enhancement[10]:

  - the harmonic cancellation algorithm aims at reducing noise by synthesizing suitable compensating tones: in this way the engine sound emerges from the disturbances, generated from a variety of sources, that will be explained in Chapter 3;
  - the harmonic enhancement aims at emphasizing some car sound

tones, to achieve a desired sound effect.

- Active Sound Design on BMW M5.

  BMW proposes a system, that, depending on the engine speed and load information, play inside the vehicle the car sound, recorded in real-time[11].

- Engine sound enhancement on Jaguar.

  Differently from previous systems, in which engine sound enhancement is performed through electronic techniques, here mechanical principles are exploited: the exhaust system valves opening is controlled by the car parameters, like engine speed and throttle valve position: in this way the flow of gasses is modulated and thus the sound changes accordingly[12].

1.3 Thesis goal

The aim of this work is to propose an algorithm which serves basically two purposes: the first one consists in improving the safety on road, providing noise-free vehicles with an sound feedback, while the second one is to allow the customization of the roar of the engine, by emphasizing more certain harmonics.

In particular, this thesis wants to provide an approach to sound generation for silent vehicles, capable to synthesize the sounds for all the possible vehicle operating conditions in terms of vehicle parameters (mainly engine speed, acceleration, vehicle velocity and engaged gear) and sound customization for both ICE and electric vehicles.

The development passes towards different steps. At the beginning an overview

on acoustics, signal processing and audio synthesis techniques is provided along with the analysis of the state of the art relative to the engine sound enhancement. Then, by means of recordings of the engine motor sound found on the Internet (available at https://www.zapsplat.com/), that identify the acoustic fingerprint of various vehicles, a model stating the dependency between engine sound, vehicle parameters and noise sources is built up.
The latter is used as starting point for the realization of the proposed algorithm, based on granular synthesis, which permits to create the electric motor sound given a new vehicle state variables profile to be followed. Furthermore, the discussion deals with the sound customization and the action that have been initiated to improve the proposed algorithm, and, especially, to obtain an integrated sound system, which works by recording the engine audio along with the vehicle parameters, obtained from the CAN messages. Additionally, some basics on CAN bus communication are provided to clarify the presented topics.
In the conclusive part, possible future progresses in this followed procedure are addressed.

This project has been developed in collaboration with the Electronics division of Marelli Europe S.p.A, located in Venaria Reale (TO).

Chapter 2

Overview on sound theory

In order to better understand the following of this thesis work, a short overview on acoustics principles and techniques of sound processing and synthesis is provided.

2.1 Basics of sound theory

The sound is a longitudinal pressure wave, generated by the oscillation of the particles composing the medium in which it travels, like air and water[13].
Its representation is the audio signal, that is the pressure profile in the medium over the time.
Sound is defined by three attributes[14]:

- *Pitch*

  It is determined by the fundamental frequency of the sound, thus, as the it increases, the pitch grows too.

- *Timbre*

  It describes the acoustic fingerprint of a sound. It depends on the energy content of each sound overtone.

- *Loudness*

  It refers to the perception of the sound level. Loudness depends on the amplitude of the vibration: considering a fixed frequency value, the wider the sound wave, the more intense and stronger the sound.

Sounds have a recognizable shape and organization, definable by a mathematical expression: this is what makes them distinguishable from noise, which features, on the contrary, an undefined and random structure. Differences between noise and sound tones structure can be appreciated in Figure 2.1.

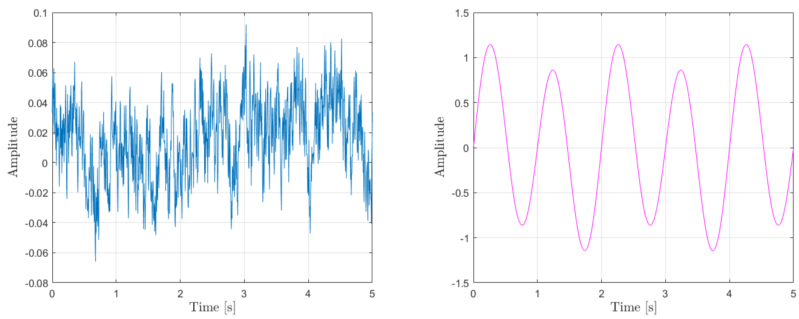


Figure 2.1: Noise (left plot) and sound (right plot) comparison

According to [15], a simple sound signal y, referred to as tone, can be represented as a periodic sine wave, whose general expression is given by Equation 2.1.

$$y(t) = A \cdot sin\left(\frac{\omega \cdot t}{f_s} + \phi_0\right) \tag{2.1}$$

It depends on five parameters, reported in Figure 2.2:

- amplitude A, which denotes the waveform pressure at a certain point in time;
- phase $\omega \cdot t$ is the angular frequency of the sine wave, function of the frequency f, being $\omega = 2 \cdot \pi \cdot f$.
- f_s is the sampling frequency, in Hertz;
- phase offset ϕ_0, which is the angular shift of the sine wave. As the phase, it is measured in radiant or degrees.

The frequency f, measured in Hertz, represents the number of oscillations per unit of time. Its reciprocal is the period T (see Equation 2.2), expressed in seconds, which measures the amount of time between two identical points of the wave:

$$T = \frac{1}{f} \tag{2.2}$$

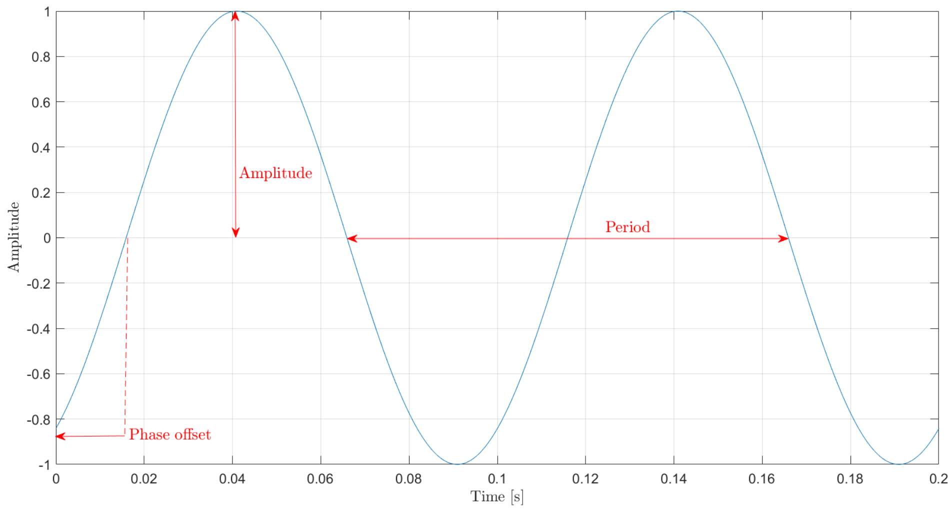


Figure 2.2: Sound wave and its parameters

Generally, acoustics deals with complex sounds, made by overlapping pure tones: this means that their shape is no longer sinusoidal, but it is affected by the way in which tones are combined.

Some examples of how tones can be merged are reported in Figure 2.3: in the first column two tones are presented along with their sum, while in the others their superposition (purple signal) in shown in the case of two different delays of the second sine wave (blue signal).

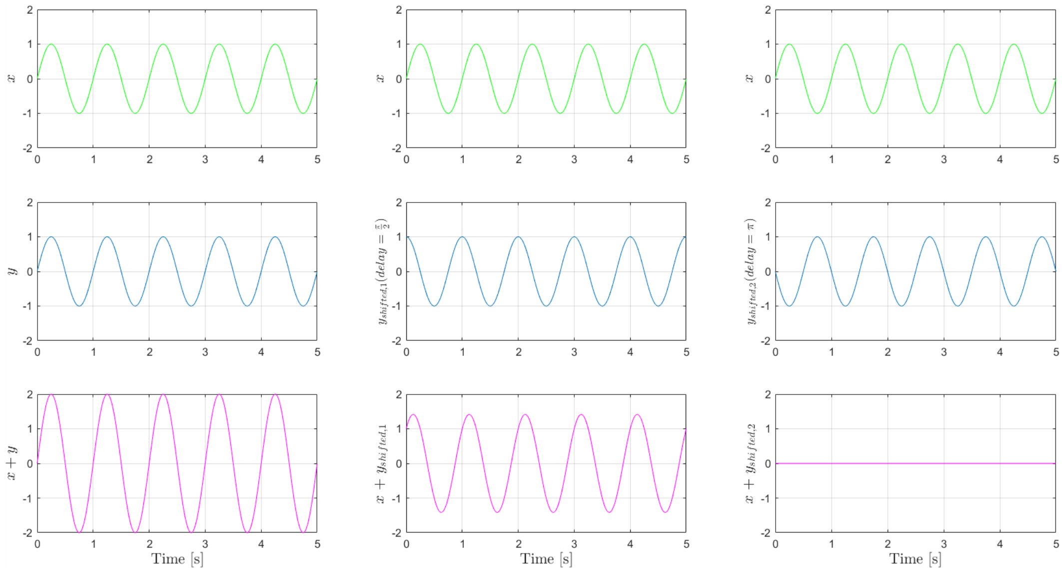


Figure 2.3: Pure tones combination outcomes

Signals present harmonics and sub-harmonics contributions, also called overtones. They are simple sinusoidal tones, whose frequencies are linked by the following relationship (see Equation 2.3):

$$f_n = n \cdot f_1 \tag{2.3}$$

where,

- n is a positive number, defining the harmonic order;
- f_1 is the fundamental frequency of the signal;
- f_n is the frequency of the n^{th} harmonic contribution.

Harmonics are associated to integer values of n, while sub-harmonics with rational ones. They have various amplitude profiles, which affect the overall sound in a different manner depending on the energy content they carry.
Thus, a signal can be usually represented as a sum of N pure sine waves (see Equation 2.4), being N the total number of considered overtones (sound harmonics and sub-harmonics).

$$y = \sum_{i}^{N} [A_i \cdot sin\left(\frac{\omega_i \cdot t}{f_s} + \phi_{0,i}\right)] \tag{2.4}$$

In the following of this dissertation, the choice of the value N is addressed. In Chapter 4, in order to account for real sounds representation, a noise contribution will be added to Equation 2.4. Such component can be considered as a stochastic noise.

Figure 2.4 shows a complex tone (violet wave), resulting from the sum of its fundamental H_1 at 100 Hz, and the two harmonics of order 2 and 3, H_2 and H_3, whose frequencies are respectively 200 Hz and 300 Hz.

The analysis done so far have been conducted in the time domain, as sound is represented as a function of time t. Switching to the frequency domain representation of signals, further analysis can be performed. This topic is discussed in more detail in the next section.

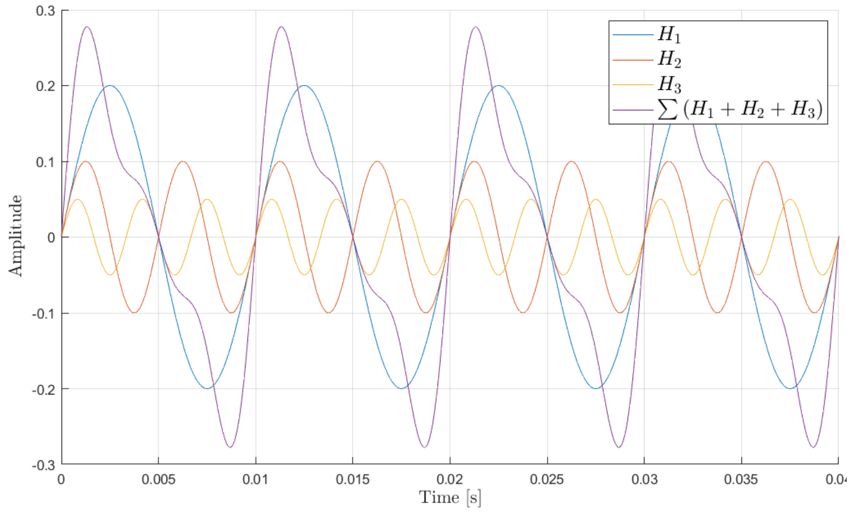


Figure 2.4: Overtones and their sum in the time domain representation

2.2 Sound signal analysis

To get useful information from a signal, different analysis can be preformed, both in time and frequency domain. Until now, all the considerations have been done using the time based representation. By the way, usually spectral analysis is employed to gt further information about the sound structure.

2.2.1 Discrete-time Fourier Transform

The Discrete Fourier Transform (DFT) of a windowed data centered around time mR can be defined as follows[16] (see Equation 2.5):

$$X_m(\omega) = \sum_{n=-\infty}^{\infty} x(n) \cdot w(n - m \cdot R) \cdot e^{-j \cdot \omega \cdot n} \tag{2.5}$$

where,

- $x(n)$ is the input signal at time n;
- w is the window function, also known as envelope, whose length is M;
- ω is the angular frequency of the data, in radiant;
- m is the time bin;
- R is the hop size between successive DFTs, expressed in samples.

DFT can be simply computed in Matlab by means of the Fast Fourier Transform (FFT) algorithm.

With reference to Figure 2.4, the corresponding representation in the frequency domain can be observed in Figure 2.5, where the magnitude in a linear scale, of the FFT, indicated as $|FFT|$, of each simple tone is represented as well as the one of the complex tone.

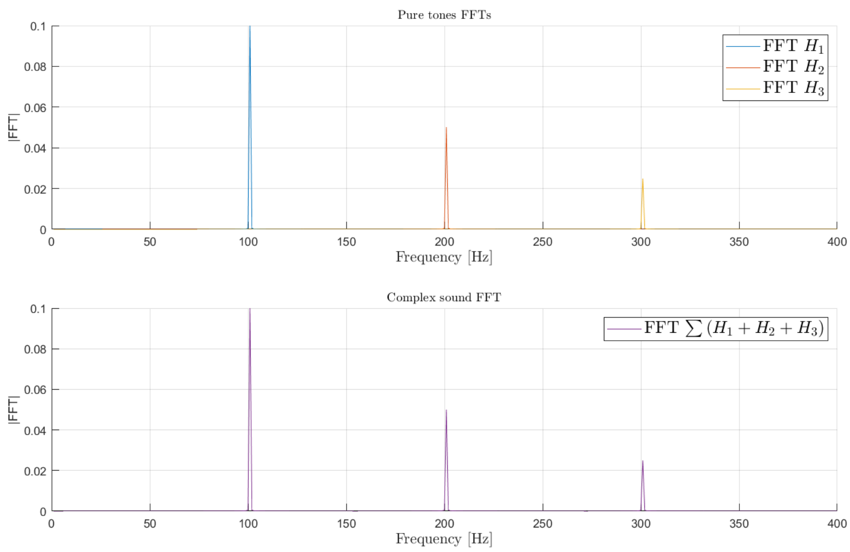


Figure 2.5: Pure tones H_1, H_2 and H_3 VS complex sound $\sum H_1 + H_2 + H_1$. Frequency domain representation

Figure 2.6 shows that the more complex the sound, the more peaks will be visualized in the spectrum.

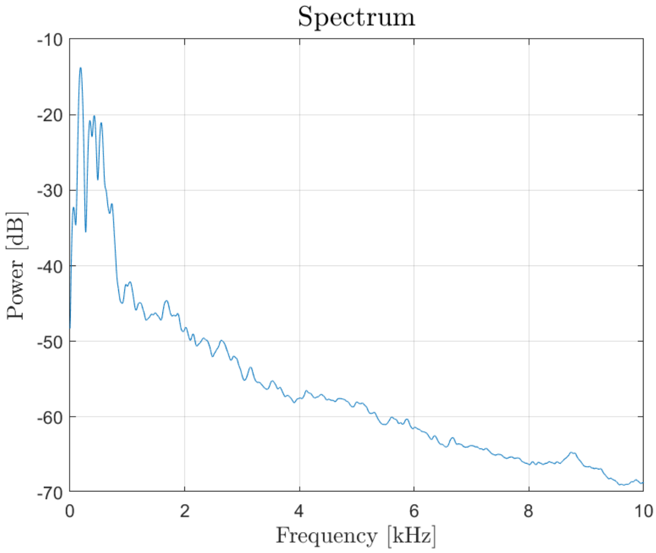


Figure 2.6: Example of spectrum of a complex sound

This kind of representation, referred to as *spectrum* of the signal, allows to easily visualize the frequency content of a signal and so to determine which are the fundamental frequency, the harmonics and sub-harmonics contributions and the power contribution they carry, as the height of the peaks identifies their energetic content.

The magnitude of the FFT can be expressed either in a linear way, that is with a value in between 0 and 1, as in the case of Figure 2.5, or in the Decibel scale, through the following transformations (see Equation 2.6):

$$|FFT|_{dB} = 10 \cdot log_{10}|FFT| \tag{2.6}$$

As exposed in more detail below, the Discrete Fourier Transform represents the basic component of the Short Time Fourier Transform.

2.2.2 Short Time Fourier Transform

The Short Time Fourier Transform[16] (STFT) is representation useful in complex signals analysis to visualize the overtones shape modification over the time.

The STFT of a signal is determined by calculating its DFT at each time bin m, by using a window with duration M and considering an overlap length between consecutive DFTs of R samples. The STFT matrix can be expressed as follows:

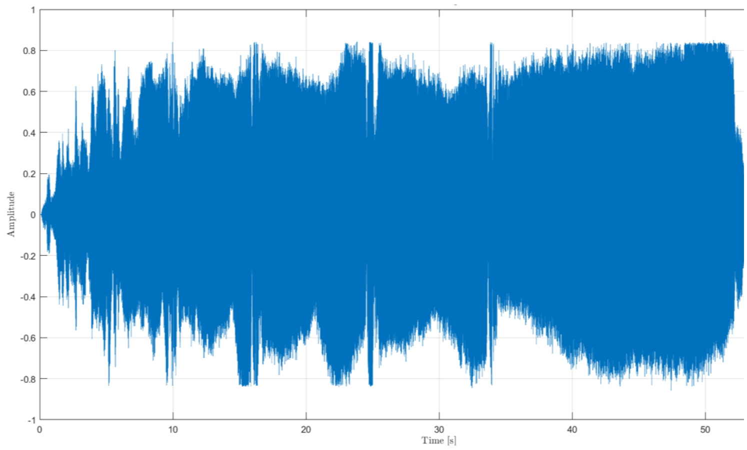
$$X(\omega) = [X_1(\omega), X_2(\omega), X_3(\omega), ..., X_k(\omega)] \tag{2.7}$$

being $X_m(\omega)$ the DFT of the audio frame evaluated in mR.

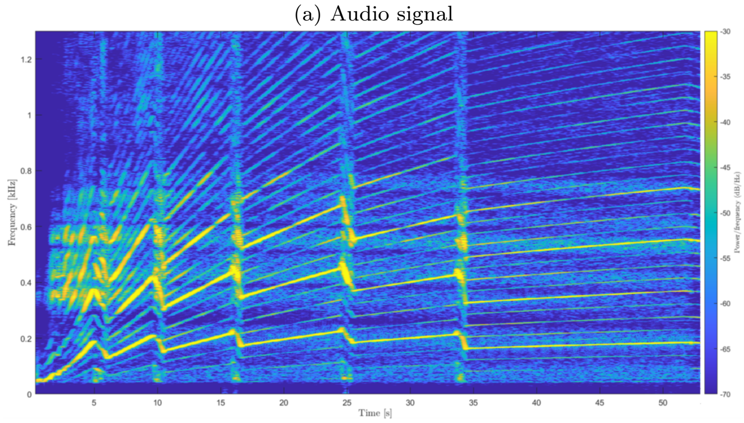
Spectrogram

The spectrogram is the representation of the energy content of the signal. It shows how the frequency content of a signal (displayed on the vertical axis) changes with time (displayed on the horizontal axis). The amount of energy in the signal at any given time and frequency is visualized as a color shade depending on the energy of the signal: intensity increases ad colour becomes lighter. Each color shade is associated with a power value expressed in dB/Hz, that is it's normalized with respect to the frequency.

An example of sound signal (sound available at [17]) along with its spectrogram is presented in Figure 2.7: here, harmonics and sub-harmonics are clearly recognizable and for each of them the energy content can be estimated, simply by looking at the color bar on the right.



(a) Audio signal



(b) Audio signal spectrogram

Figure 2.7: Example of an audio signal along with its spectrogram

2.3 Audio synthesis techniques

Engine sound enhancement is basically performed according to two main approaches, discussed below in this chapter. In the following, the main audio generation methods are briefly explained.

Subtractive synthesis

It works starting from a very complex sound source, rich of overtones, and filtering out the unwanted elements from it, to obtain a desired output track by means of suitable filters.
This results in the elimination of the harmonic contributions we are not interested in, so to make the desired signal standing out[18].
Depending on the filtering process shape, different results are obtained.

Additive synthesis

According to this technique, a complex sound can be produced as the superposition of pure tones, each one generated through an oscillator, as formerly presented by Equation 2.4. The way in which they are combined determines the overall sound complexity and texture.

Frequency modulation

It exploits two oscillators, referred to as modulator and carrier. The carrier oscillator frequency is tuned proportionally to the modulator frequency[18].

Sample-based synthesis

This technique uses a collection of sound samples that are merged together to create a new audio track. It requires lower processing effort with respect to other synthesis approaches, but a huge database of sample recordings is necessary. Samples are recombined as they are provided, without being processed in any way.

Wave-table synthesis

The idea behind wave-table synthesis[19] is to divide a period of a wave into small sound fragments and to store them in cells that are jointed to create the so called waveform lookup table. For each sound wave the same procedure is performed, generating additional tables.
Sound generation involves to scroll among the available wave-tables and within them so to find the audio block to be played back. The direction and the speed at which the samples are played back determines the sound complexity, while the way in which they are chained identifies the newly generated sound timbre.

Granular synthesis

This is a method working with small audio pieces, referred to as *grains*[20]. It is based on the same principles of sample-based and wave-table approaches, but here samples are no longer used directly: they are manipulated and then recombined, modifying their sequence and spacing so to produce a desired sound texture[19]. On the contrary,the other two techniques involve the collection of a huge amount of sound samples which are not processed but only recombined according to the desired output. In particular, sample-based synthesis collects audio pieces and simply recombines them. Instead, wave-table synthesis divides a one-cycle wave into samples and recombines them without applying first an envelope.

The grain envelope choice affects the timbre perception of the synthesized sound, as exposed in more detail in Chapter 5.
In granular synthesis, the sound is synthesized by extracting short sonic grains, generally lasting 10 - 100 ms, from an audio track and combining them according to a certain criterion[21]. Combination can be performed in three ways[19]:

- quasi-synchronous granular synthesis: grains are arranged with regularity;
- asynchronous granular synthesis: grains are distributed almost randomly in time; the resulting sound has a very complex and rich timbre;
- pitch/time-synchronous granular synthesis: grains are chained taking into account the pitch.

Further details concerning granular synthesis theory and implementation will be provided in the following of this dissertation, as the proposed algorithm for engine sound enhancement is based on it.

Physical modeling

According to [19], physical models are mathematical simulations of the behavior of a real sound. Sound is generated according to the model parameters, which account for materials used for the system and the interaction with it.

2.3.1 Engine sound enhancement main approaches

There are mainly two approaches to engine sound enhancement:

- sample-based synthesis;
- tones composition;

Their main drawback can be identified in the huge amount of resources to be managed: in the case of the sample-based technique, a big data set of recorded samples is needed to be able to match all the possible vehicle operational conditions we could have to deal with, while in the case of additive synthesis a huge number of oscillators is required to perform the generation of the harmonics and sub-harmonics of the new signal. Furthermore, tones synthesis from scratch provides an artificial result, that is not suitable to be played in the cabin and outside the vehicle.

In this work, quasi-synchronous granular synthesis has been chosen as method for developing an engine sound enhancement algorithm: its advantage is that the characterization of the vehicle scenarios can be done by dividing a unique audio sample into multiple micro-fragments of sound, modulating them if needed and recombining them at will. This audio generation technique is discussed in more detail in Chapter 5, where the algorithm implementation process is presented too.

Chapter 3

Model of car sound, noise and vehicle parameters

In order to develop an algorithm for engine sound enhancement, firstly a model has been built. Its aim is the identification of the parameters affecting the sound perceived in the cabin of a vehicle equipped with an ICE engine.

3.1 Noise sources analysis

The analysis of the components affecting the perceived engine sound can be conducted according to the *Noise, Vibration and Harshness (NVH)* theory, which deals with in-cabin oscillations and how they are perceived by car passengers[22].

According to the Noise, Vibration and Harshness theory, frequencies can be divided in 3 main ranges:

- *vibration* refers to oscillations at low frequencies, typically below 20 Hz, which are not audible, but they can be felt by human body;
- *harshness* identifies oscillations whose frequencies fall in the range 20 - 100 Hz, where both acoustic and displacement effects can be appreciated[22];
- *noise* refers to oscillations at frequencies above 100 Hz and below 20 kHz, which cannot be physically felt but they are perceived by human hears.

The proposed model takes into account sources of noise both inside and outside the vehicle, most of which are summarized in Table 3.1.

Noise sources

Internal	External
Cylinders configuration and movement	Pavement characteristics
Gear number and engagement	Weather conditions
Throttle valve opening	External sounds
Pressure on the brake pedal	
Items deterioration	

Table 3.1: Noise sources

3.1.1 Internal factors

Concerning the internal sources of noise, they are mainly represented by the mechanical components movement in the engine compartment.
As a parameter affecting the engine sound, the engine cylinders configuration is taken into account. Cylinders pistons don't move contemporary but one at a time, generating a periodic sound where the phases of the combustion process are recognizable. The steps of a four-stroke engine combustion process are shown in Figure 3.1 (available at [23]).

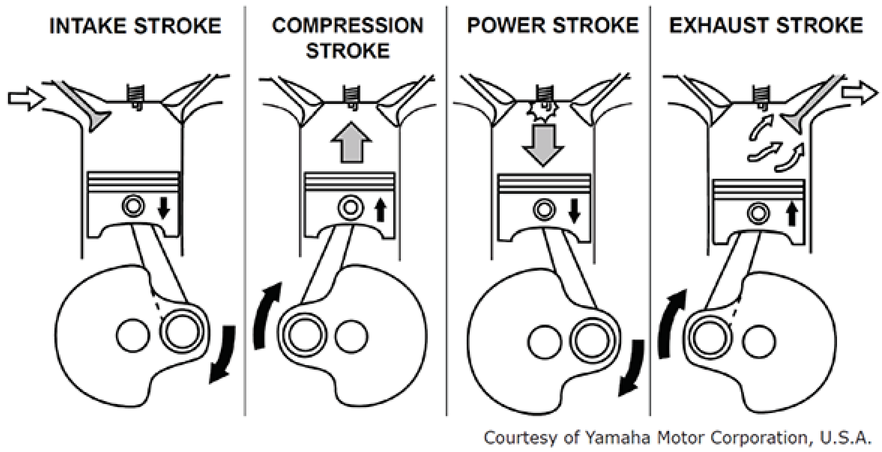


Figure 3.1: Four-stroke engine cylinders

The four-stroke internal combustion engine cycle features two crankshaft revolutions. In particular, it works as follows[23]:

- the starting phase is the intake stroke: the piston moves from the Top Dead Center (TDC) to the Bottom Dead Center (BDC), increasing the chamber volume and allowing the air-fuel mixture to flow in;
- during the second stroke, the mixture is compressed. Both intake and exhaust valves are closed and the piston moves up, completing the first

crankshaft revolution;

- in the power phase, the air-fuel mixture is supplied with the energy needed for the combustion. This increases both the temperature and pressure within the chamber. During this stroke the piston moves to the bottom, producing the power output;

- the latter phase refers to the exhaust valve opening and the movement of the piston from BDC to TDC, closing the exhaust gas out the valve, as the piston reaches the bottom. The remaining exhaust gas is pushed out by the piston as it moves upwards, closing the valve. Then, the cycle repeats.

The engine sound modifies as the number of cylinders and the placement changes. In particular, the higher the number of cylinder, the more overtones characterize the perceived audio, since increasing the number the cylinder makes the number of firings grows and thus the sub-harmonics content is enriched.
Additionally, the piston stroke length has to be taken into account in the sound texture determination. A small stroke length produces a richer sound texture, typical of sports car, while a longer one, typical of trucks, leads to darker and deeper timbres.
Basically, there are three main adopted cylinders configurations, based on different firing orders: the in-line placement, featuring adjacent cylinders, the V-shaped one, where cylinders are divided into two lines and positioned so to form an angle between two adjacent ones and the opposite positioning. Moreover, the V-shaped configuration includes two subcategories, producing

different explosions sequence in the cylinders[24]:

- the *flat-plane crank* setup features a regular positioning, a regular firing order and sounds with a high pitch;
- the *cross-plane crank* is characterized by an more complex structure, irregular firing order, producing a richer, darker and powerful timbre.

Another noise source is represented by the engine, which rotates at a certain speed, depending on the driving operating conditions: the higher the engine speed, the higher the firing frequency and thus the sound pitch, at a fixed gear number.
Mechanical items deterioration, due to aging phenomena, load and internal friction phenomena contributes to the sound perceived in the cabin too.

3.1.2 External factors

They are due to the environment features, like the road status and weather conditions.

Aerodynamic drag force F_a is a longitudinal resistance force against the direction of motion of the vehicle. It increases proportionally to the front section of the vehicle and the vehicle velocity squared[25], as visible in Equation 3.1.

$$F_a = \frac{1}{2} \cdot \rho \cdot S \cdot C_X \cdot v^2 \tag{3.1}$$

where ,

- ρ accounts for the air density;
- C_X is the aerodynamic drag coefficient;
- S is the cross section area;
- v is the vehicle longitudinal velocity.

The noise due to F_a grows proportionally with its intensity.

Depending on the asphalt condition (dry, wet, cobblestones, sand or dirt scenarios) different sounds result.
The noise caused by the contact between tyre and road is due to the rolling of the tyre. This is because the pneumatic has not a perfectly cylindrical shape. The contact area in which the tyre exchanges pressure with the pavement is not represented by a point but it's an extended region. This increases the friction forces and the noise proportionally: the higher the speed, the higher the aerodynamic drag force and so the noise.

The rolling friction[26] is given by Equation 3.2:

$$F_{roll} = \mu_r \cdot m \cdot g = (B_0 + B_2 \cdot v^2) \cdot m \cdot g \tag{3.2}$$

where ,

- μ_r accounts for the rolling friction coefficient, which is affected on three parameters: B_0 that depends on the road surface features, B_2 that depends on the tyre characteristics and v that is the vehicle longitudinal velocity;
- m is the vehicle mass;
- g is the gravitational acceleration, almost equal to 9.81 $\frac{m}{s^2}$;

Atmospheric phenomena, like strong wind and rain, can affect the perceived noise too, since they modify the pavement features, thus the exchange of forces between tyre and road, and they can produce additional unwanted sound impacting on the chassis, as for example when it rains or wind blows.

Furthermore, according to [27], over almost 1.5 kHz the main source of noise is the aerodynamic one, but its influence is appreciated also in the low frequency range, along with the tyres rolling friction contribution too.
Thus, as mentioned in [7], the most of the vehicle noise can be found in the low frequency range, below 1 kHz. By the way, here the most energetic and sound relevant engine harmonics are located too: this means that the noise and the sound to be extracted, made by the engine mechanical components displacement, are overlapped. This is the reason why their separation results to be quite critical.

3.2 Model formalization

Joining the previous analysis with the observation of the spectrogram of an audio recording, along with the vehicle parameters behaviors, allows to end up with some general considerations.

The following pictures (see Figures 3.2, 3.3 and 3.4), reported as examples to better understand the list of observation below, respectively refer to the recordings of an acceleration phase, which spans gears from the first one to the sixth one, a deceleration phase and an idle condition (sounds available at [17]).

Here all the most energetic overtones, that are the one in the low frequency range, are clearly shown by the attached color bar, which permits to estimate the power contribution of the visualized harmonics.

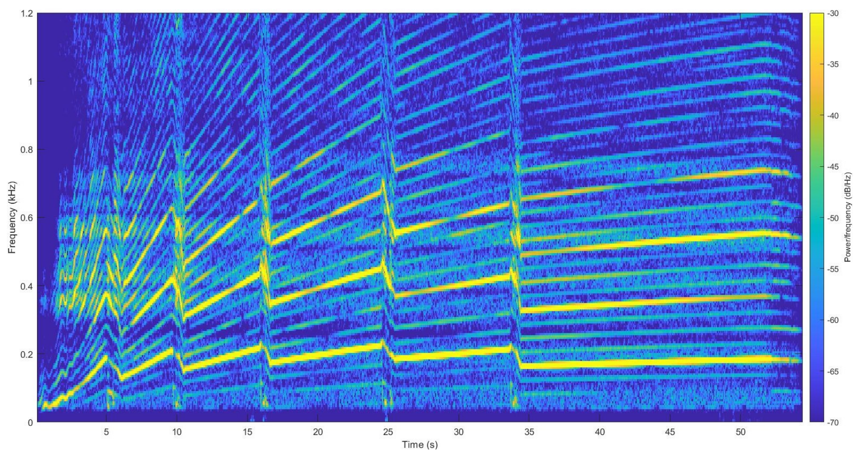


Figure 3.2: Spectrogram of an acceleration phase recording

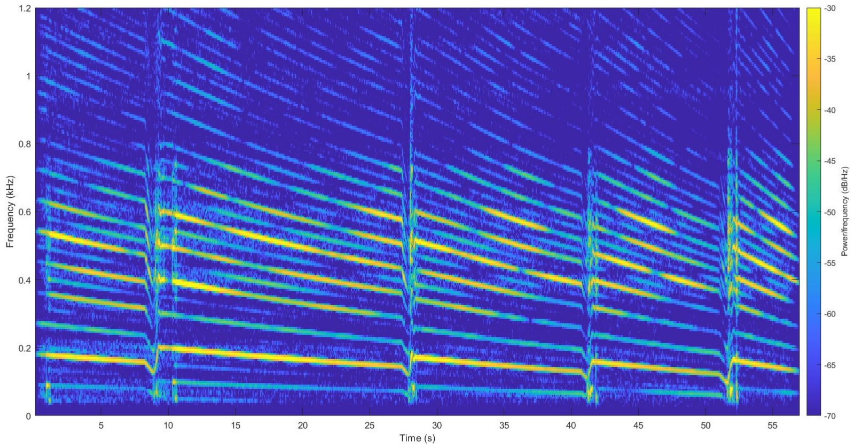


Figure 3.3: Spectrogram of a deceleration phase recording

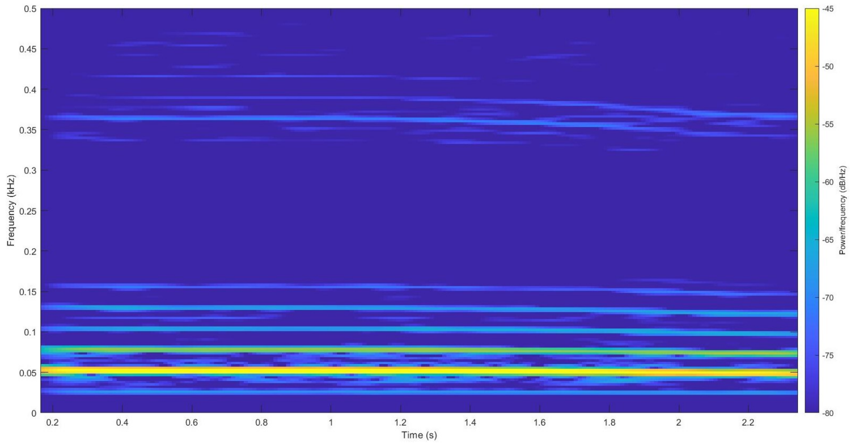


Figure 3.4: Spectrogram of a recording at idle

Some key observations that can be done considering Figures 3.2, 3.3 and 3.4 are listed below.

- The harmonics and sub-harmonics profiles reflect the one of the RPM. This is proved by the relationship[28] linking the RPM value and the corresponding frequency value of each harmonic contribution at idle by means of a proportional relationship (see Equation 3.3):

$$f_{H_k}(t) = k \cdot N_{cyl} \cdot \frac{RPM(t)}{60} \tag{3.3}$$

where,

- $Ncyl$ is the number of cylinder of the considered vehicle;
- RPM is the engine speed at time t, in $\frac{rounds}{min}$;
- k is the index referring to the considered harmonic contribution;
- f_{H_k} is the frequency, expressed in Hertz, at time t, in seconds, characterizing the k^{th} harmonic at time t.

- The higher the velocity, the higher the number of harmonics at an appreciable energetic level; at idle, harmonics content is very poor both in number of appreciable overtones and power intensity with respect to the driving scenario.
  At idle, only the fundamental harmonics carries an considerable energy content, while in driving conditions, other harmonics impact heavily to the sound.
  The main common element is the fundamental harmonics, that is the one that characterizes the sound and that is visible at any operating condition. The other contributions, with their different frequencies and amplitudes content, concur to create the engine timbre.

- The higher the acceleration and the lower the gear, the higher the harmonics slopes: going towards upper gears and lower accelerations, the harmonics behavior becomes flatter.

- The gear shift part is not well characterized, since it represents a zone with a nonlinear behavior; its duration and the harmonics slope depend on the time the driver spends to disengage the gear and engage the new one; considering an automatic gear shifting the time needed for the shift becomes more regular.

- The more energetic harmonic is first, but also the third and fifth ones have a considerable weight on the overall sound. So, typically, the odd ones gives the main contribution to the sound and their energetic content reduces going towards higher harmonics orders.

- Sub-harmonics carry lower energy, but they concur to the creation of a richer and fuller sound.

- The number of sub-harmonics is related to the number of cylinders: given a vehicle with N cylinders, for each integer harmonic H_n (n = 1, 2, ...) there are (N - 1) sub-harmonics of it, which are $\frac{1}{N}H_n$, $\frac{2}{N}H_n$, ..., $\frac{N-1}{N}H_n$. Considering a four-cylinder configuration,there will exist the three sub-harmonics $\frac{1}{4}$, $\frac{2}{4}$ and $\frac{3}{4}$ for each integer harmonics.
  The fundamental harmonics is associated with n = 1.
  These contributions, until the 3^{rd} harmonic, are underlined in Figure 3.5, where the harmonics and sub-harmonics of Figure 3.2 are tracked.

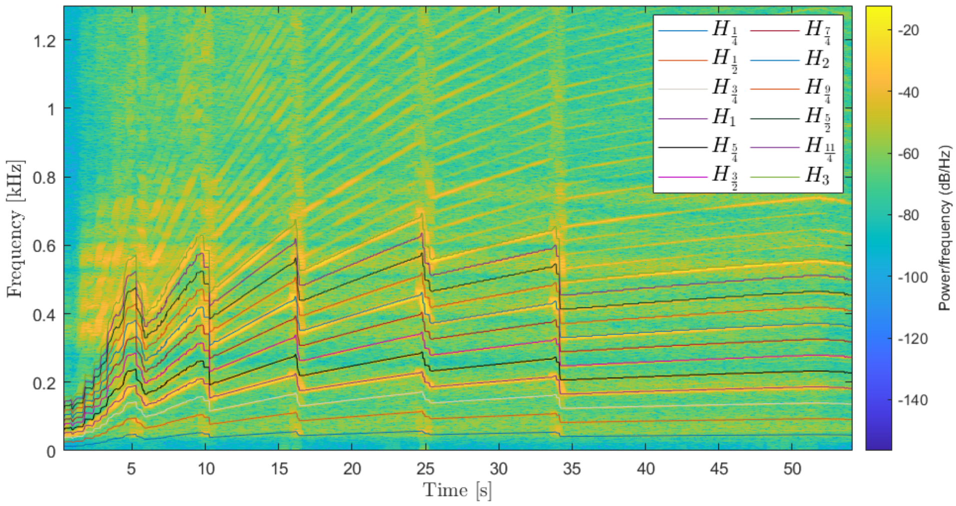


Figure 3.5: Harmonics of Figure 3.2

As the four-stroke engine has two crankshaft revolutions, H_1 represents the frequency of the whole cycle, given by the two rounds. Thus, each of them has a revolution frequency which is a half of H_1. This contribution is the more powerful among the sub-harmonics of the fundamental frequency.

In general, an engine with N_{cyl} cylinders has $\frac{N_{cyl}}{2}$ crankshaft revolutions and the engine revolution frequency[28] can be defined as in Equation 3.4.

$$f_{en}(t) = \frac{f_{H_1}(t)}{2} = \frac{N_{cyl}}{2} \cdot \frac{RPM(t)}{60} \tag{3.4}$$

This model represents the starting point for the engine sound algorithm implementation, presented in Chapter 5.

Chapter 4

CAN messages and audio acquisitions integration

The starting point of this thesis work has been represented by the tentative to perform the audio recordings and CAN messages acquisitions, which aims at obtaining respectively the engine sound and the vehicle parameters profiles.
In order to do this, a setup has been created in the considered vehicle.
The vehicle used for this purpose is the Alfa Romeo Giulia Veloce, having the following features[29]:

- the employed fuel is petrol;
- a power of 280 CV;
- 4 in-line cylinders;
- an automatic transmission with 8 gears.

Being this car a sport one, its sound is quite powerful, thus it's appropriate for engine sound recording.

The desired data have been collected by means of a microphone and an acquisition system from the CAN bus.

4.1 Audio registrations

Various microphones with different features in terms of band width were tested. For engine sound recording, a suitable microphone needs to have a wide range of dynamics so to be able to get all the amplitude oscillations of the audio and a frequency range that can include the most of the harmonic content, especially the fundamental harmonic, which represent the component carrying the most of the energy and identifying timbre of the vehicle. In particular, the microphone shall capture components in the low frequency range, as here the most of the engine noise is located, as reported in [7].

At first an analog microphone with a frequency band of 45 Hz - 20 kHz has been tested, but, due to the limited dynamics it was almost always saturated and an important attenuation has been observed at low frequencies, thus, it was discarded.

Then, a MEMS (Micro-Electro Mechanical System) microphone has been chosen as the one that best fits this application. According to its data sheet available at [30], its main characteristics can be summarized as follows:

- a wide dynamics: it saturates at almost 133 dB SPL, thus it is suitable to get the sound coming from very strong acceleration phases which lead to louder sounds;

- a frequency range equal to 45 Hz - 20 kHz, that is suitable for collecting the fundamental harmonic of the recorded signal, which generally falls around 45 - 50 Hz at idle, that is its minimum value;
- a characteristics presented in Figure 4.1 (available at [30]).

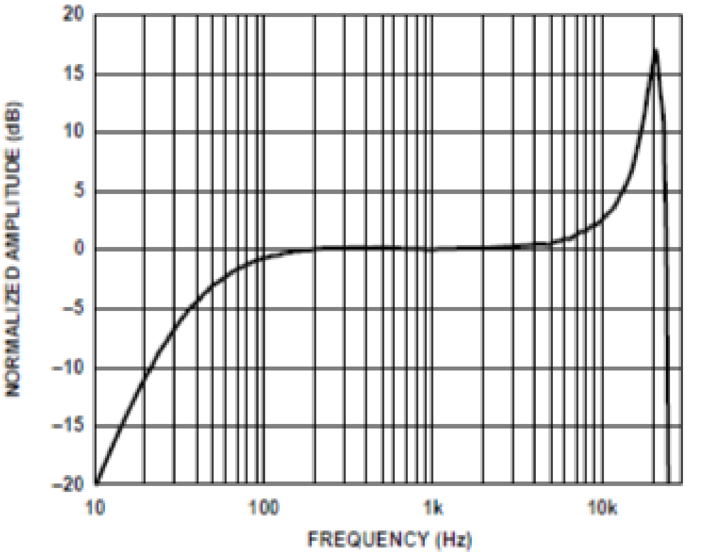
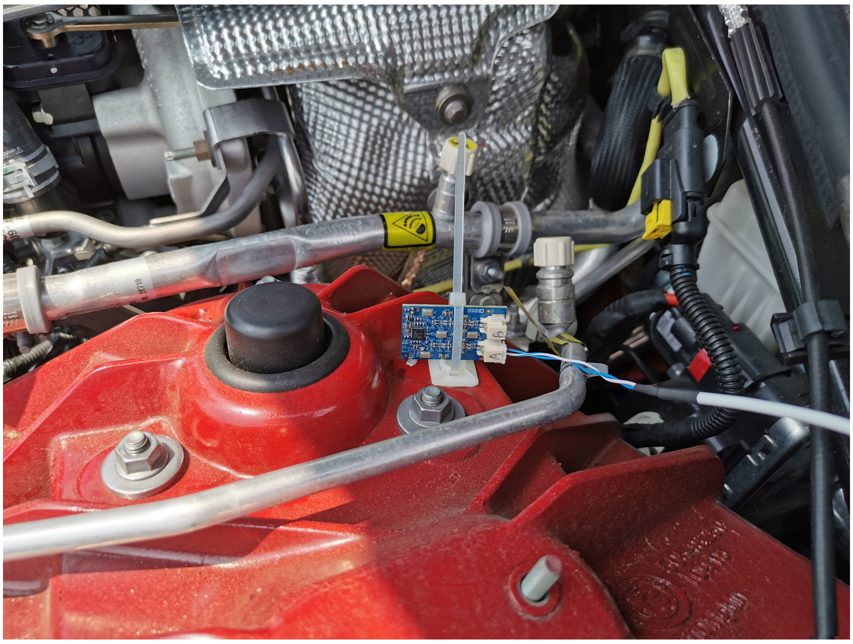


Figure 4.1: MEMS frequency response

Audio recordings have been performed according to two microphone setups, shown in Figure 4.2: the first placement is in the engine compartment, while the second is close to the exhaust pipe.



(a) Microphone in the engine compartment

(b) Microphone close to the exhaust pipe

Figure 4.2: Microphone positioning

The first one, having the microphone placed in the engine compartment, has produced very weak audio signals. This is due to the value assumed by the *Signal-to-Noise-Ratio* (SNR), whose definition[31] is given by Equation 4.1:

$$SNR_{dB} = 10 \cdot log_{10}\left(\frac{P_{Signal}}{P_{Noise}}\right) \tag{4.1}$$

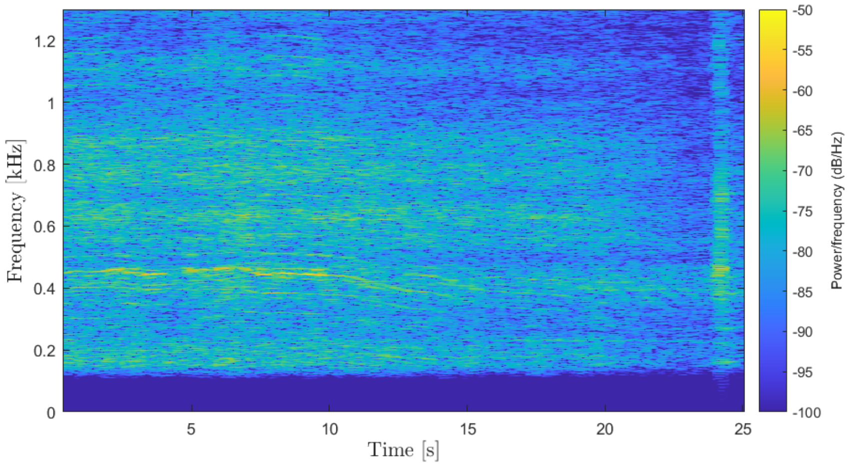
where,

- SNR_{dB} is SNR expressed in dB;
- P_{Signal} identifies the energy of the sound component of interest
- P_{Noise} is the power of the overlapped stochastic noise contribution.

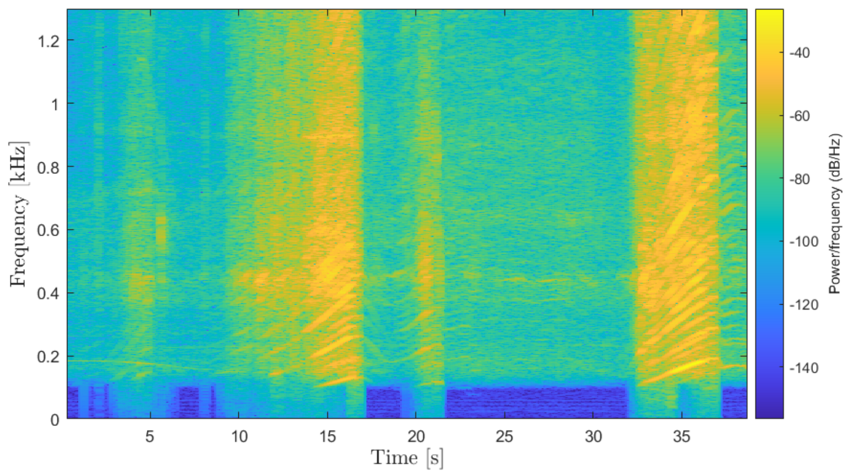
In the faced situation, a low SNR characterizes the recordings. A low SNR means that the noise component has a magnitude comparable with the one of the engine sound, thus the harmonic content below 100 Hz, where for a big part the fundamental is located, is almost canceled and the higher harmonics have low energy, which doesn't make them prevail on the overlapped noise.
Notice that only when high pressure is provided on the gas pedal, i. e. a high acceleration is performed, harmonics and sub-harmonics becomes more energetically appreciable. This doesn't allow to well characterize a full acceleration maneuver spanning all the gears. Thus, the database construction cannot be performed under such setup.

The next step was characterized by the installation of the same microphone close to the exhaust valve. Here, harmonic content is more appreciable as recorded sound is more powerful. By the way the quality of the audio is bad, as it doesn't come from the engine: being the engine sound system designed for reproducing a pleasant audio, it's necessary to get a starting audio with a quality as good as possible. Playing back the audio coming from the exhaust valve cannot lead to a good outcome. Furthermore, dynamics acquisition with the latter setup was impossible, due to the high temperatures in the neighborhood oh the exhaust valve, which damages the microphone, as electronic components performance is affected by the working temperature. The evidences of these problems are reported in Figure 4.3:

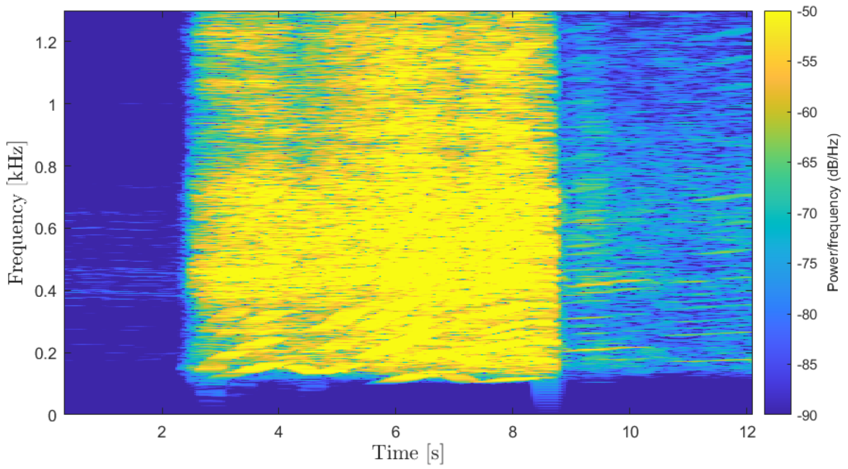
- harmonics are not clearly traceable: in particular, during phases at low or almost null acceleration harmonics are almost not visible as they are covered by noise, while in acceleration phases harmonics increase their energetic content but it generates regions at almost the same power level, where, however, harmonics are not distinguishable from noise;
- in the frequency range below almost 120 Hz the no content is visible, even if the microphones specific assures to get frequencies in this zone, over 45 Hz, and below, unless an acceptable attenuation according to Figure 4.1.



(a)



(b)



(c)

Figure 4.3: Audio recordings spectrograms

Such issues have conducted towards the adoption of already existing audio recordings relative to different vehicles, generally sports car, where sound is more appreciable in terms of energy and quality.
Different sounds have been considered (available at [17]). The variety of tested audio tracks allows to check the correct functioning of the implemented algorithm in different driving scenarios and with different provided timbres.

In the next section, some CAN communication basics are presented, to clarify how vehicles parameters can be acquired from the on-board network. Due to the problems encountered in the audio acquisition, CAN data have not been used for the algorithm implementation. The following will be useful for the future development of this thesis, which will involve the audio and CAN data integration to the algorithm proposed in Chapter 5.

4.2 CAN messages acquisitions

In order to get the vehicle parameters corresponding to the recorded sound, the car CAN network is accessed through the vehicle OBD connector. It is connected to an acquisition system, depicted in Figure 4.4, which allows to get the messages and collect them on the PC.
Such messages are acquired as a stream of data which has to be interpreted, through *parsing*, presented below in this chapter.

The system is implemented according to the block scheme in Figure 4.5.

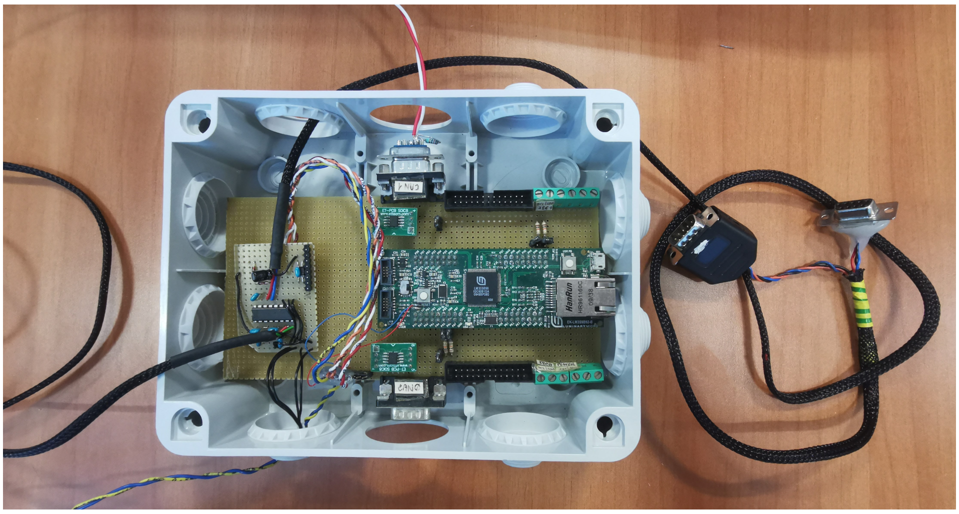


Figure 4.4: CAN messages acquisition system

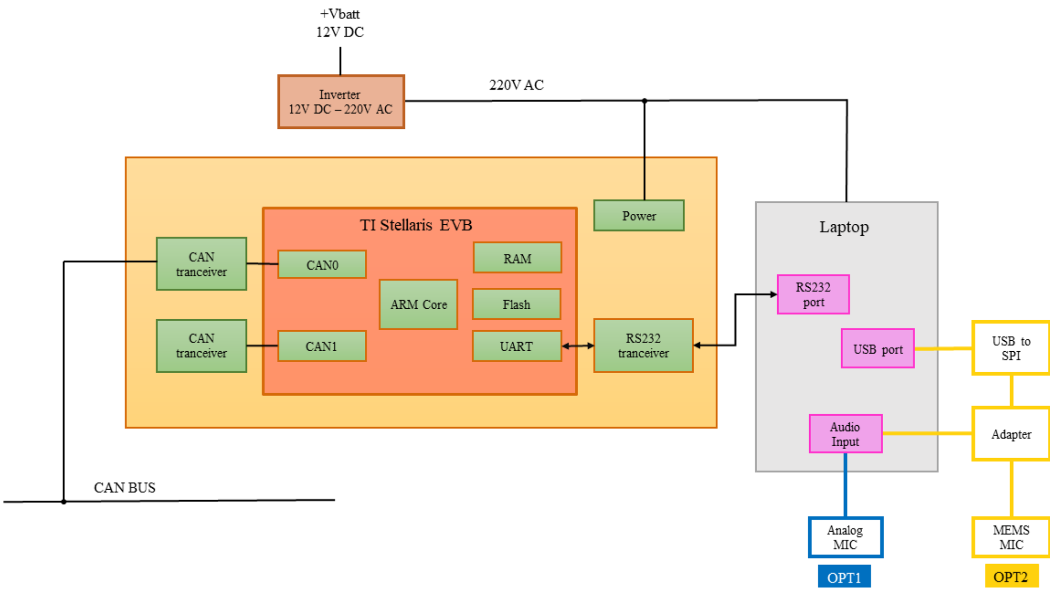
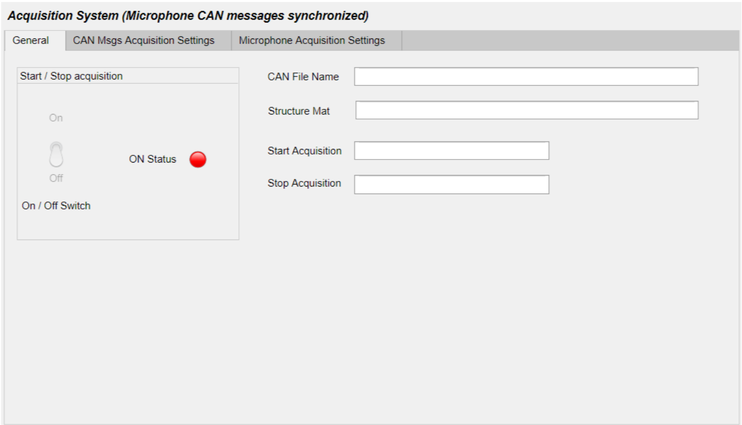


Figure 4.5: CAN messages acquisition system

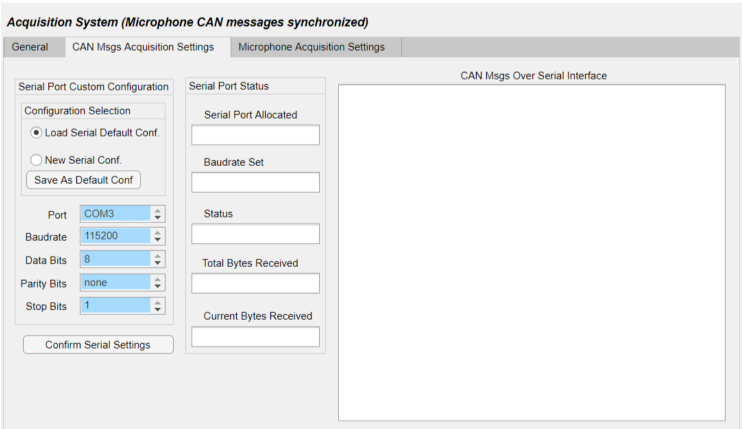
It works as follows: the CAN0 port is connected to the OBD of the vehicle and from the CAN line messages are received. The RS232 transceiver outcome is a stream of data like the following one:

3e2,860884000c03261a

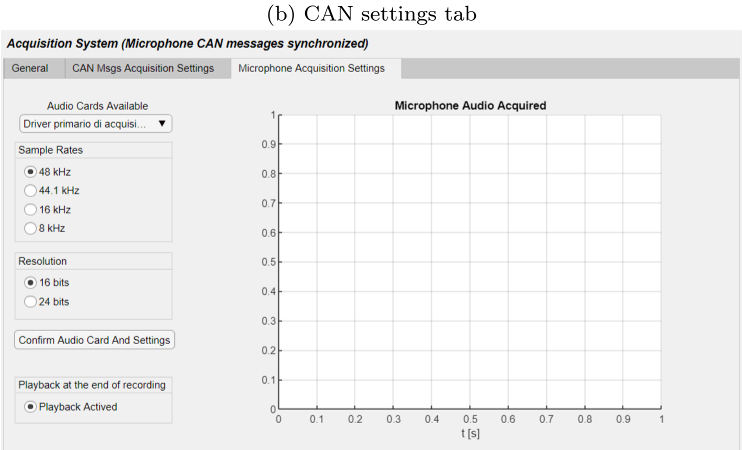
The latter is provided to the laptop, accessed through the DB9 port, on which Matlab is installed. Data are collected by means of a GUI (see Figure 4.6), which permits to modify the microphone settings and the CAN acquisitions parameters and enable the acquisition. The microphone is connected to the laptop too. Two microphones settings options are presented in Figure 4.5: the first one (OPT1), highlighted in blue, uses an analogue microphone, which is directly connected by means of the jack entry, while the second option (OPT2), highlighted in yellow, involves the MEMS microphone, which needs to be mounted on an electronic board for working (USB to SPI block). The latter is connected to the USB port of the laptop on one side and to the jack entry through an adapter on the other one.



(a) Enable/disable acquisition tab



(b) CAN settings tab



(c) Microphone settings tab

Figure 4.6: Matlab GUI for audio and CAN messages recording

4.2.1 CAN basics

The CAN bus, standing for Controlled Area Network, is a communication system between electronics components, typically involved in real-time applications. It was originally developed by Robert Bosch for automotive applications, but currently employed in other fields, such as industrial, aerospace and military ones[32].
CAN bus allows the interconnection between different Electronic Control Units (ECUs) within the vehicle, like ABS, TC system and ESC.

An example of network organization, involving the CAN bus and different ECUs is proposed in Figure 4.7.

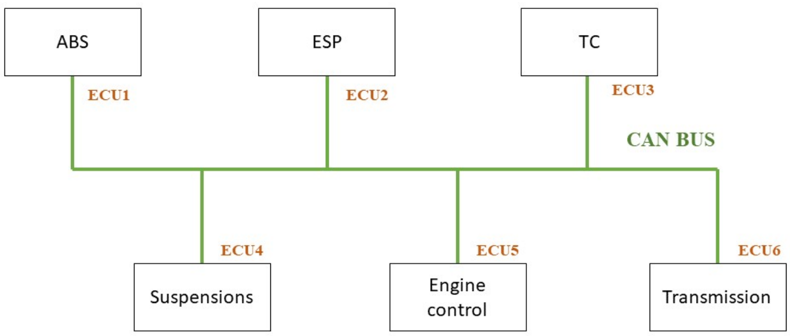


Figure 4.7: CAN organization

Advantages introduced by the CAN communication are several[33]:

- real-time availability of vehicle status information;
- simplicity, as communication is performed only by means of the CAN bus line, thus wiring, errors and costs are kept at bay;
- full centralization, as it offers a unique access for performing collection of

data (like gas pedal position, gear engagement, fuel consumption, engine speed, vehicle velocity and so on) and diagnostics;

- efficiency, as CAN messages travel according to the assigned identifier, which states the priority level.

Data transmission is based on a dominant and recessive bits (respectively, logical 0 and logical 1), according to an inverse wired logic: when ECUs send both 0 and 1 levels, the bus is available for communication. When no zeros are sent, communication is disabled[33].

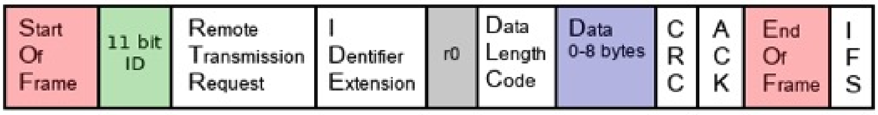
4.2.2 CAN message structure

Frames, also called messages, perform the actual transmission of data. They are associated to an identifier, which states their priority. Messages can be in two formats: short frame format (11 bits) and extended frame format (29 bits)[34].

Four types of CAN messages, or frames, can be distinguished:

- **Data frame**

  It is used to send data over the network. It comprises the following fields, reported in Figure4.8a and Figure 4.8b (both available at: https://www.allaboutcircuits.com/).



(a) Standard format

(b) Extended format

Figure 4.8: CAN data frame structure

- **Remote frame**

  It is used by an ECU to ask for data to another node in the over network.

- **Error frame**

  It is sent whenever a frame is received in an incorrect format.

- **Overload frame**

  It is used to prevent CAN network to be overloaded.

4.2.3 Parsing phase

The acquisition system provides a stream of data which doesn't show directly the information about can messages, thus it has to be converted into an understandable shape. Its translation, referred to as *parsing*, is explained in the sequel of this chapter.

Note that each string like this one contains not only one CAN message, but more that one. Their separation and identification is made by means of specific tables, one for each CAN message.

CAN messages tables

In order to translate the stream of data, a map given by the car maker is needed. This map collects different information related to each signal we can transmit over the network. The layout table has a structure like the one in Figure 4.9.

Name	Message	Multiplexing	Start	Length	Byte order	Value type	Initial value	Factor	Offset	Minimum value	Maximum value	Unit	Value target	Comment
RPM	MSG1	-	12	10	Motorola	Unsigned	-5	0.2	-10	0	10000	1/min		
TVO	MSG1	-	27	15	Motorola	Unsigned	0	0	0	0	100	%		
Gear	MSG1	-	32	4	Motorola	Signed	40	2.7	21	0	9			Current gear
Speed	MSG1	-	51	6	Motorola	Unsigned	1	1	0	0	300	km/h		Vehicle speed

Figure 4.9: Table for CAN messages - 1^{st} example

Exploiting the table and map associated to each message, we can extract all the signals it carries. Each CAN message provides information on more than one vehicle variable that can be measured. As an example a certain message can carries the following signals: RPM, oil indication, gear shift, water temperature. For each signal, some parameters are defined:

- name, which is how the signal is called;
- message, which is the CAN message identifier;
- multiplexor;
- byte order convention, either Motorola or Intel;
- offset, which is a shift to be applied to the sequence in the parsing process.

Motorola and Intel refer to two ways for ordering bits. Depending on the chosen method, different translations results at the end of the process.

Then for each CAN message a layout map is provided, so to encode it. An example, referring to 4.9 is shown in Figure 4.10.

	7	6	5	4	3	2	1	0
0	7	6	RPM 5	4	3	2	1	0
1	15	14	13	12	11	10	TVO 9	8
2	23	22	21	20	19	18	17	16
3	31	30	29	28	27	26	25	24
4	39	38	37	36	Gear 35	34	33	32
5	47	46	45	44	43	42	41	Speed 40
6	55	54	53	52	51	50	49	48
7	63	62	61	60	59	58	57	56

Figure 4.10: Map for CAN messages - 1^{st} example

Here we can see all the signals reported in the previous table. As each message has its own table and map layout, another message could be associated to another table, like the one below in Figure 4.11.

Name	Message	Multiplexing	Start	Length	Byte order	Value type	Initial value	Factor	Offset	Minimum value	Maximum value	Unit	Value target	Comment
GasPedal	MSG2	16	0	3	Intel	Unsigned	0	0.2	1	0	1	%		
Torque	MSG2	-	16	8	Intel	Unsigned	4	0	0	-800	3000	Nm		
Oil	MSG2	-	41	7	Intel	Signed	40	2.7	-50	0	1			Oil indicator

Figure 4.11: Table for CAN messages - 2^{nd} example

The corresponding map is reported in Figure 4.12.

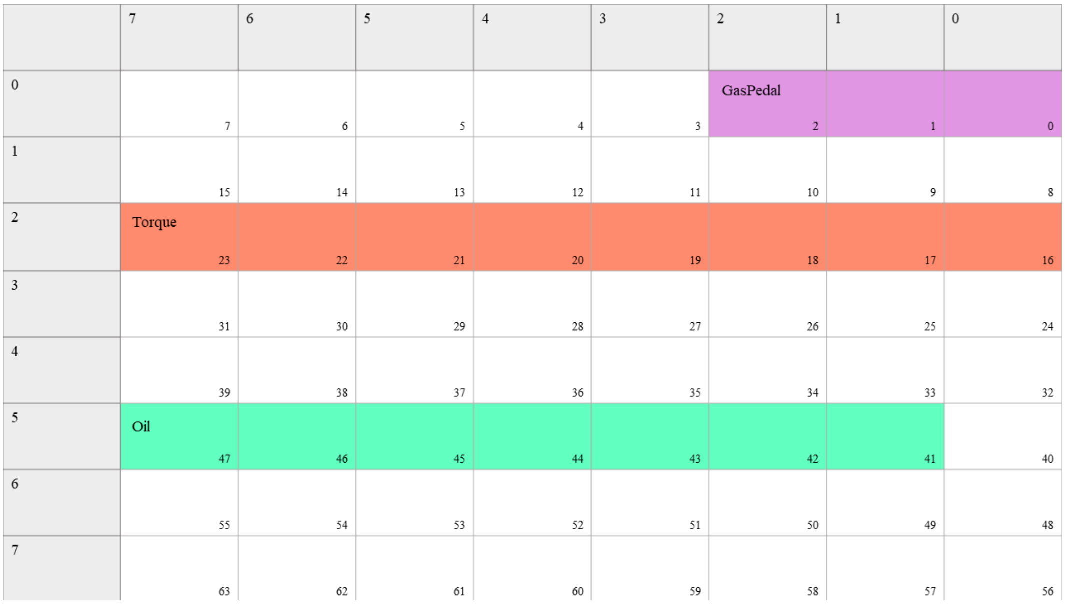


Figure 4.12: Map for CAN messages - 2^{nd} example

Parsing is the translation of a stream of data to get information from it. It consists in identifying the location of each CAN message of interest so to extract it. It involves the shift of the bit sequences according to the layout map by means of masks.

A string of code like the one presented in Figure 4.13 performs this kind of operation, which has to be repeated for all the signals of interest:

```
for i = 1 : 8
    CAN1_bytes(:, i)=(CAN1_char(:,I * 2 - 1) - '0') * 16 +
                 (CAN1_char(:,i * 2) - '0');
end

CAN1_RPMsignal = (bitshift(bitand(CAN1_ bytes(:, 3), 63), + 3) * 4 +
               bitshift(bitand(CAN1_bytes(:, 4), 224), - 5)) * 0.1 + 3;
```

Figure 4.13: Messages parsing with mask application

where 63 is the adopted mask, while 3 and 5 are the shifts to be applied. 0.1 is the factor and 3 the offset. 3 and 4 are the non-empty bit referred to this signal and that have to be shifted.

An example of CAN acquisition outcome after the parsing process is proposed in Figure 4.14. Here, four vehicle parameters are reported: vehicle speed, engine speed, gas pedal position and number of the engaged gear.

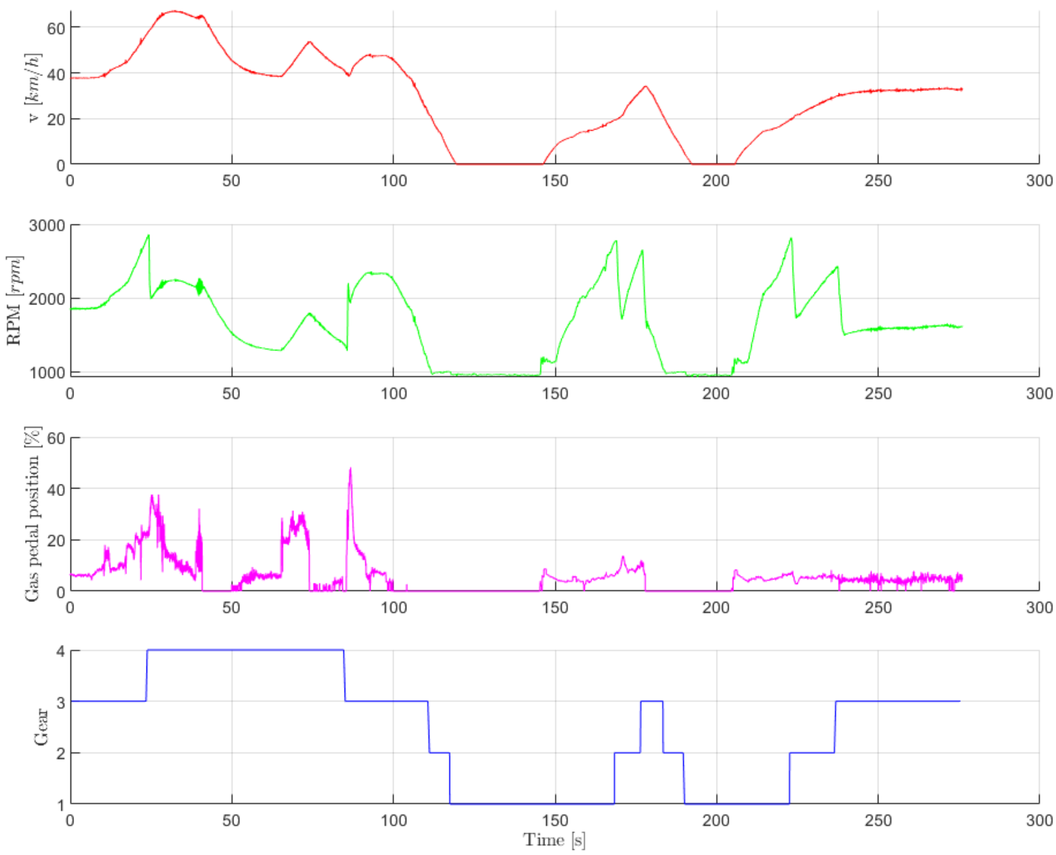


Figure 4.14: CAN messages after parsing

Chapter 5

Algorithm implementation

An approach to the engine sound generation has been proposed in [28]. Starting from it, and taking into account the model developed in Chapter 3, this thesis work provides an algorithm for engine sound enhancement and customization.

The implemented procedure, developed in MATLAB:

- performs the harmonics extraction from a noisy audio signal, whose corresponding vehicle parameters profiles are known;
- offers the possibility to customize the sound so to achieve a specific sound texture;
- granulates the resulting signal, creating a database of small audio samples and rearrange them to synthesize a new engine sound according to a different vehicle parameters profile: this process is referred to as granular synthesis.

The algorithm works partially offline and partially in real-time:

- the analysis of the recorded audio along with the corresponding vehicle parameters is performed on a computer, as well as the harmonics extraction, customization and granulation phases;
- the synthesis part is done in real-time, as it requires as inputs the new vehicle parameters coming from the acquisition and parsing of the CAN messages, got from the CAN bus line during the driving.

In the following of this chapter, all those steps are explained in detail.

5.1 Harmonics extraction phase

This step aims at eliminating the noise included in the recording so to make the engine sound emerge.

The flow of the harmonics extraction algorithm is shown in Figure 5.1.

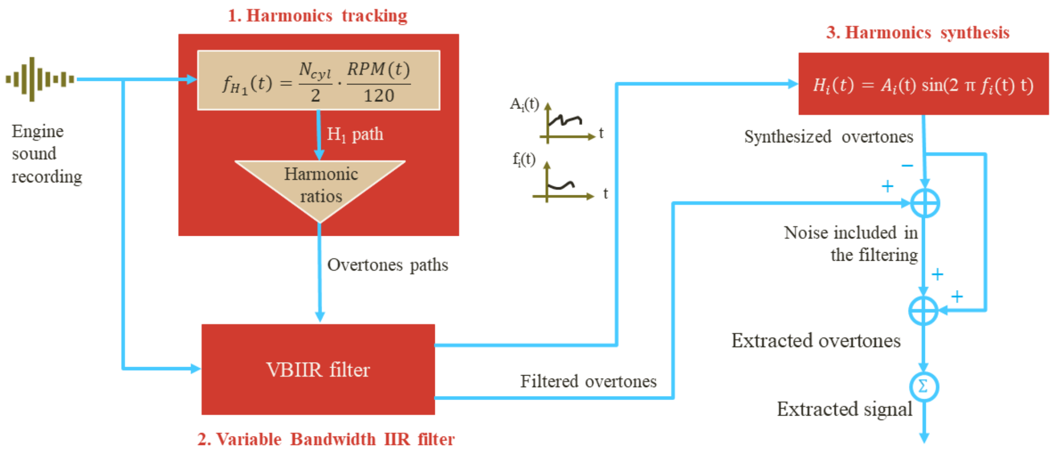


Figure 5.1: Harmonics extraction algorithm

As mentioned in Chapter 2, a real sound signal is made by a part featuring a regular behavior, identifying the harmonic content, and another one having a random shape, defined as stochastic noise contribution.

Thus, it can be written as the sum of those two elements (see Equation 5.1), as also proposed in [35]. The latter contribution cannot be exactly estimated, as it is a random noise source, but it can be foreseen by taking into account the considerations in Chapter 3.

The purpose here is to divide the audio signal harmonic part from the noise one.

$$y_{real}(t) = y_{harmonic}(t) + y_{stochastic}(t) \tag{5.1}$$

Recalling Equation 2.4, Equation 5.1 can be equivalently written as:

$$y_{real}(t) = \sum_{i=\frac{1}{N_{cyl}}}^{n_H} [A_i \cdot sin(\omega_i \cdot t + \phi_{0,i})] + y_{stochastic}(t) \tag{5.2}$$

where,

- i represent the considered harmonic (or sub-harmonic);
- n_H is the higher harmonic order to be included;
- N_{cyl} is the number of cylinders of the vehicle from which the recording is obtained.

The isolation of the harmonic content is required so to avoid to include the noise in the manipulation of the audio signal. The audio resulting from the overall algorithm must be as much as possible noise free: if the noise would be included in the signal generation, when the synthesized audio is played on the vehicle, the perceived noise would be doubled, because there would be both the noise coming from the audio and the actual noise, due to the factors showed in the previous chapter.

Fundamental frequency reconstruction

In order to extract the engine harmonics from the noisy signal, so to divide $y_{harmonic}(t)$ and $y_{stochastic}(t)$, a procedure working in the frequency domain has been implemented.

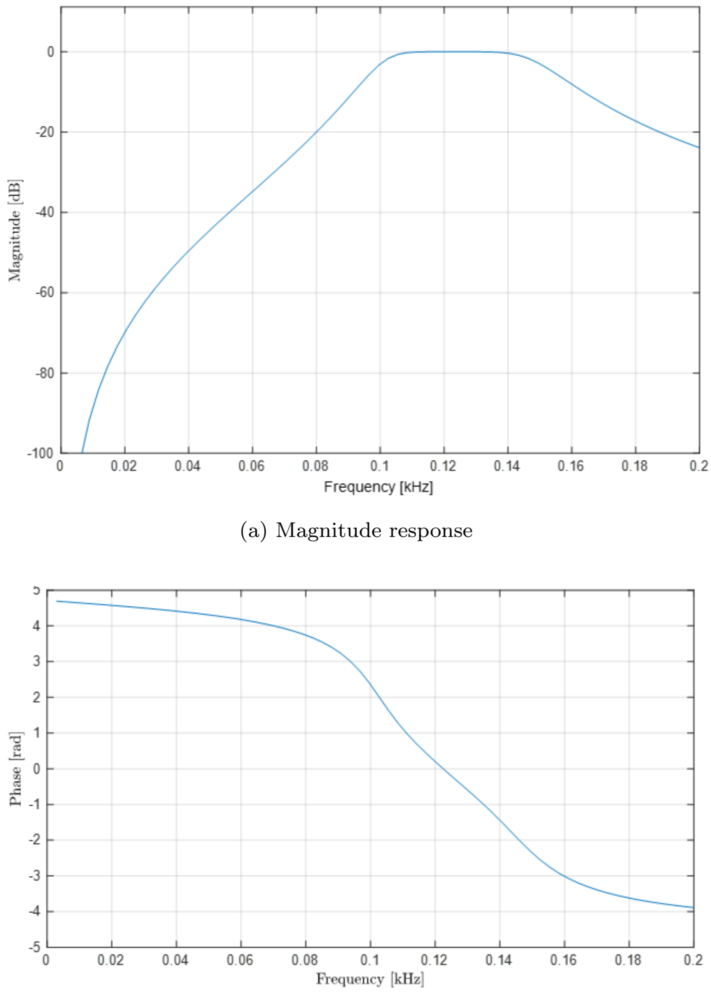
It involves firstly the localization of the fundamental frequency and then its extraction. The other harmonics trajectories are reconstructed starting from the fundamental one, considering the relationship between the first harmonics and the overtones.

The fundamental frequency tracking is based on the engine speed profile associated to the analyzed recording: recalling that the engine speed and the

fundamental frequency trends are linked by Equation 3.3, being the RPM behavior known, the fundamental frequency path is easily determined.
Being this a proportional relationship, it is clear that the fundamental frequency shape will follow the engine speed one, as discussed in Chapter 3.
Then, by exploiting Equation 2.3, the other overtones profiles are reconstructed too: all these trends are the reference for determining the parameters which govern the succeeding filtering process, analyzed in the following of this section.

Filtering process

A generic band-pass filter is such that the frequency components allowed to pass and be visible in the output are the ones within the defined bandwidth, while the outer ones are cut out, so to extract only a certain harmonic component and neglect the other ones. In Figure 5.2, the shape of a simple band-pass filter with band 100 - 150 Hz and order 6 is depicted in terms of magnitude and phase response.



(a) Magnitude response

(b) Phase response

Figure 5.2: Band pass filter response (band 100 - 150 Hz and 6^{th} order)

For the harmonics extraction purpose, a filter with a flat shape in the pass-band region, strong side lobes attenuation and minimum phase delay is suitable for letting only harmonics part to pass, neglecting the noise and preserving as much as possible the phase relations. Furthermore, it shall be able to track the overtones variations overt the time.

To fit these requirement, a *Variable Bandwidth Infinite Impulse Response (VBIIR) filter*[36] has been firstly tested. This is an Infinite Impulse Response (IIR) band-pass elliptic filter with tuneable parameters depending on the fre-

quency: this means that its center frequency and order are updated depending on the paths determined at the previous step. The pass-band ripple, the stop-band attenuation and the band length are, instead, defined a priori.
For what concerns the filter editable parameters, some considerations can be done:

- *Order*

  The tuning of the order is needed to deal with the increment of the frequency: what happens is that as higher frequency values are considered, the probability for instability to occur, at a fixed filter order, increases. Thus, depending on the value of the filter center frequency at a given point in time, the order is varied, so to prevent instability problems: this means that, as frequency rises, the filter order has to be reduced[28].
  In short, variability on the order prevents the filter to become unstable as frequency grows more and more.
  Practically, the order is reduced depending on the order of the current harmonic or sub-harmonic to be extracted rather than on the current frequency value.

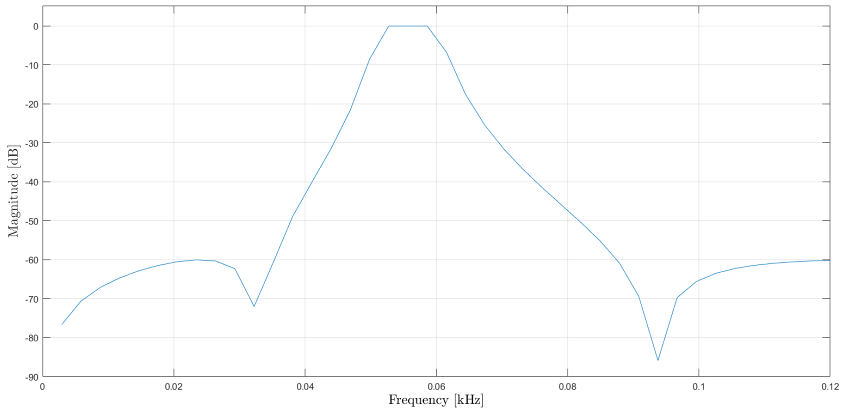
- *Center frequency*

  The center frequency is tuned so to follow the trajectories just computed.

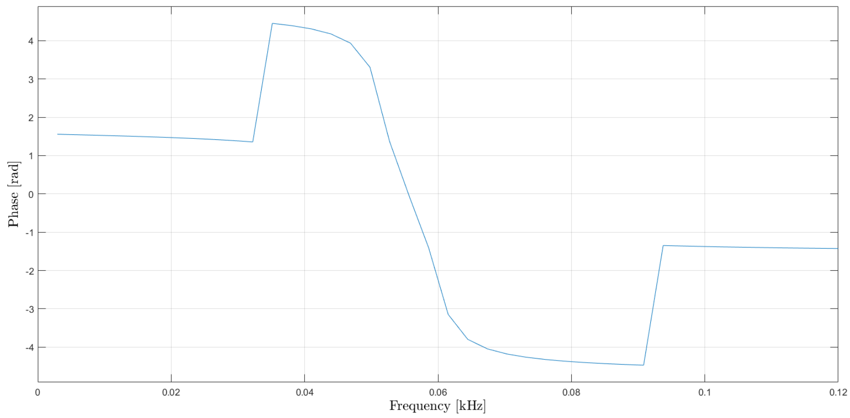
- *Band length*

  The band length is chosen so to collect the whole considered harmonic (or sub-harmonics) and part of the contributions in its neighborhood, isolating only the overtone of interest. The role of the noise part collected by the filter around each overtone in the algorithm is discussed below in this chapter, when dealing with the harmonics and sub-harmonics synthesis.

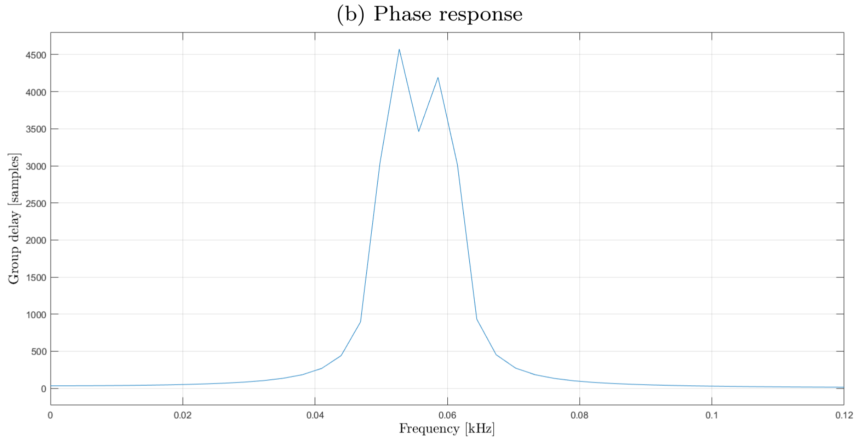
The filter magnitude, phase and group delay characteristics are reported in Figure 5.3, considering a center frequency of 52 Hz and a filter order equal to 6.



(a) Magnitude response



(b) Phase response



(c) Group delay

Figure 5.3: Variable Bandwidth IIR filter response (center frequency: 52 Hz, filter order: 6)

As easily observable, the magnitude presents a flat shape in the pass-band region and very steep edges.

Concerning the phase, the introduced delay is very low.

Instead, the behavior of the group delay presents two peaks around the considered center frequency. Its shape cannot be described by means of a constant or 1^{st} degree polynomial function. Additionally, its value is not negligible, as with the chosen settings of ripple and band length, varying the order and the center frequency it is generally of few thousands of samples.

As mentioned in [37], the phase delay P represents the time delay of each harmonic component of the sound, while the group delay D identifies how the rate of change of the phase with frequency. The difference between them can be understood by considering their definitions (see Equations 5.3 and 5.4):

$$P(f) = -\frac{\phi(f)}{f} \tag{5.3}$$

$$D(f) = -\frac{d\phi(f)}{df} \tag{5.4}$$

being f the frequency and $\theta(f)$ the phase response of the filter.

A nonlinear phase delay characteristics, as the one in Figure 5.3c, leads to a lag in the amplitude envelope of the filtered signal, and thus in the power content carried by a certain harmonic contribution. Graphically, it results is not phased harmonics, which don't respect the harmonic relationships between them.

In order to deal with the nonlinear group delay issue, a second filtering process is performed at the end of the first one. This aims at minimizing the shift occurring in the output of the first filtering stage, as suggested in [38].

Thus, Figure 5.1 can be updated as follows (see Figure 5.4):

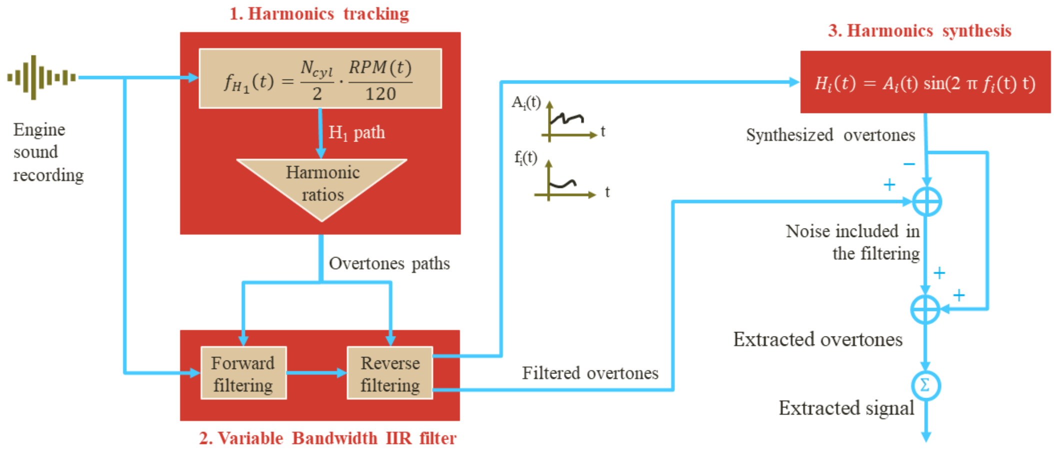


Figure 5.4: Harmonics extraction algorithm structure modification

The first filtering is represented by the block "Forward filtering" in Figure 5.4. This second filtering (identified by the "Reverse filtering" block) consists in passing the outcome of the first filtering stage inversely in the same filter.

However, the result of this procedure doesn't consist in harmonics contributions in phase, so distortions between harmonics occurs. Thus, to deal with such issue, the amplitude and frequency information, collected through the filtering, are exploited.

Recalling Equation 5.1, each harmonic or sub-harmonic is then synthesized in the time domain from scratch, using to the corresponding amplitude and frequency arrays determined at the previous filtering step: the result is a series of harmonics which are in phase.

Notice that, due to the group delay shape, the information about the amplitude is not exact but it is affected by a shift, as formerly presented. Thus, a minimum error affects the envelope of the newly synthesized waves. However, this error is assumed to be negligible.

Furthermore, in order to obtain an extracted signal closer to the real acquired one in terms of timbre characterization, the stochastic noise part, involved in the filtering process outcomes, is added on top of the newly generated harmonics. This consists in the noise component captured by the filter around each overtone, since the filter side lobe attenuation is not $-\infty$ dB and the band has a certain width.

In order to compute the disturbance component captured by the filter, for each overtone the difference between each filtered harmonic and the corresponding synthesized one is performed. Then, this quantity is added on top of the corresponding overtone.This step contributes to increase the realism of the overall sound, making it less artificial from the acoustic point of view. On these contributions, obtained after the addition of the noise part on each of them, sound customization is offered. This process is presented below in this chapter. The comparison of the original signal and the synthesized harmonics part, until the 4^{th} integer harmonic and considering the sub-harmonics too, is shown in Figure 5.5.

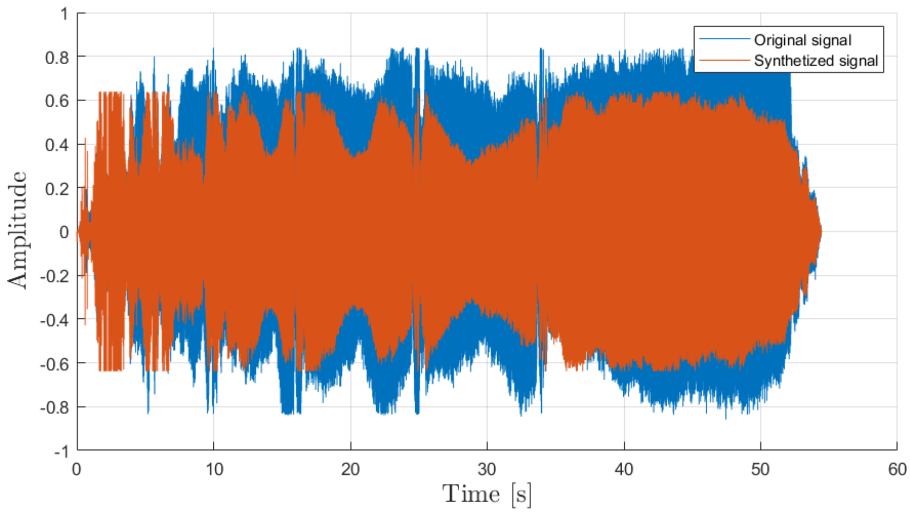


Figure 5.5: Original signal VS synthesized one

Concerning the number of contribution to include in the new signal, indicated as n_H in Equation 5.2, the more harmonics are present, the more full will be the sound. However, being the sound of interest in the low frequency range mainly, as previously analyzed, it is sufficient to cover a frequency range whose upper limit is around 1 - 1.5 kHz.

The harmonics contributions are separately shown in Figure 5.6, while the sum of the synthesized signal and the stochastic noise in comparison with the original sound is reported in Figure 5.7.

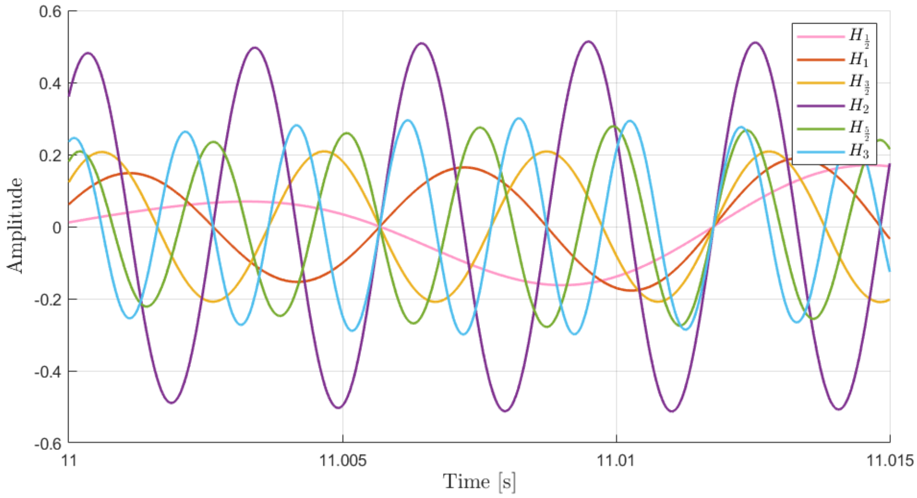


Figure 5.6: Synthesized harmonics

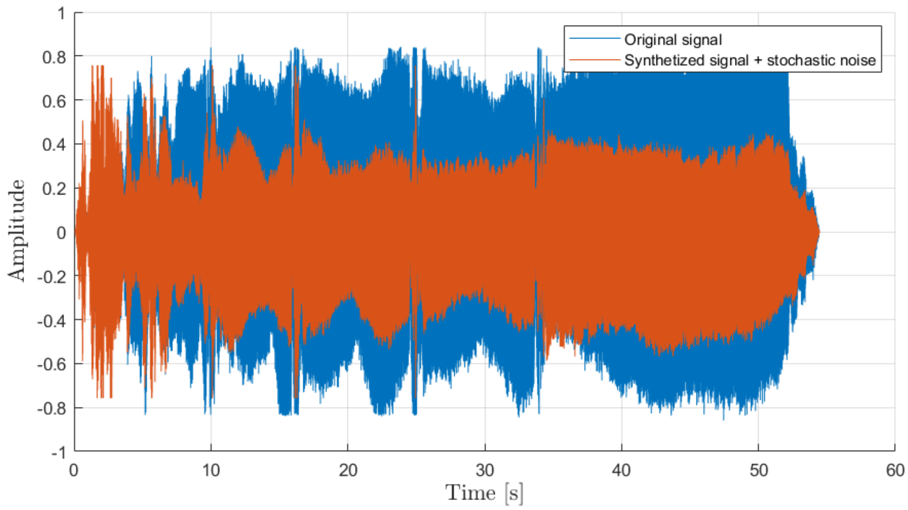


Figure 5.7: Original signal VS synthesized one + stochastic noise

In the following (Figure 5.8 and Figure 5.9), the differences between the spectrogram of the original audio track and the one just obtained can be appreciated: as clearly observable, the noise content in between harmonics and sub-harmonics is reduced by the filtering process. This make overtones more appreciable and sharper, as noise is kept at bay.

Being the order lower for higher order harmonics, the amount of noise captured by the filter becomes bigger. Instead, at low frequency, this quantity is smaller as the order is bigger. Basically, reducing the order, a bigger noise contribution passes to the output. This effect is visible in Figure 5.9, where the noise contribution in between overtones is more appreciable at higher frequencies than at lower ones.

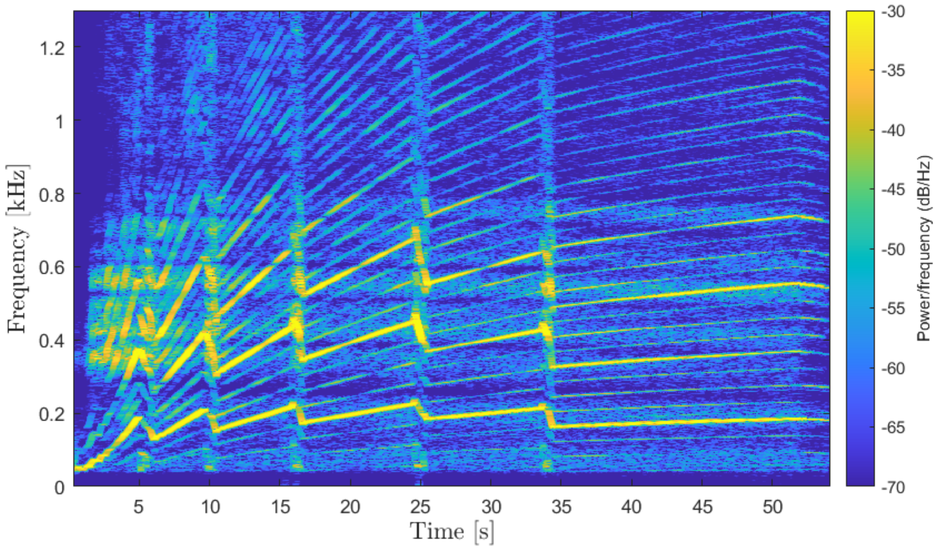


Figure 5.8: Original signal spectrogram

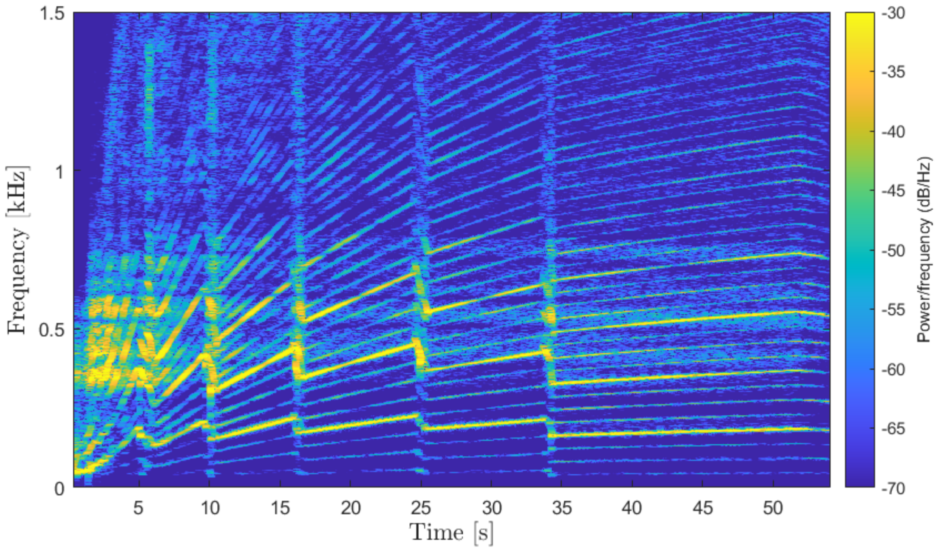


Figure 5.9: Extracted signal spectrogram

Playing the resulting audio, it's very close to the original one and the difference is only represented by the minimization of the external disturbance from the environment surrounding the vehicle. This signal is the starting point for the following step of the engine sound enhancement algorithm, which consists in performing the so called granular synthesis, that involves the audio granulation, the modulation of the obtained blocks and the overlap and add processing so to obtain a new audio signal with a certain set of requirements, given in terms of vehicle parameters.

5.2 Granular synthesis

Granular synthesis[21] is a technique for sound generation and modulation that works by recombining micro audio samples, referred to as *grains*, extracted from an audio track.

Nowadays, it represents an innovative technique in the acoustic field for producing unique and particular sounds, even in real-time application. It is widely employed in video games and, more generally, in the virtual reality world for the synthesis of the acoustic effects and animations. It's diffusion is currently growing, since resulting sounds are very realistic and a huge amount of different sound textures can be created.

A grain is characterized by two main parameters, the *envelope* and the *content*[20], which respectively refer to the amplitude shape of the grain and the actual sonic contribution.

Depending on the manipulations grains are subjected to and the way they are reordered, different sound textures can be created: as an example, granular synthesis can be exploited to perform time-stretching of the original audio they are extracted from, altering the sound duration and letting the pitch unmodified, or, vice versa, to obtain pitch-shifting[39].

They can be arranged randomly or according to some criteria. With regard to this thesis, a new grains sequence is established according to a new provided vehicle parameters profile, that includes the engine speed, the gear and the acceleration information over the time.

The granular synthesis steps are outlined in Figure 5.10.

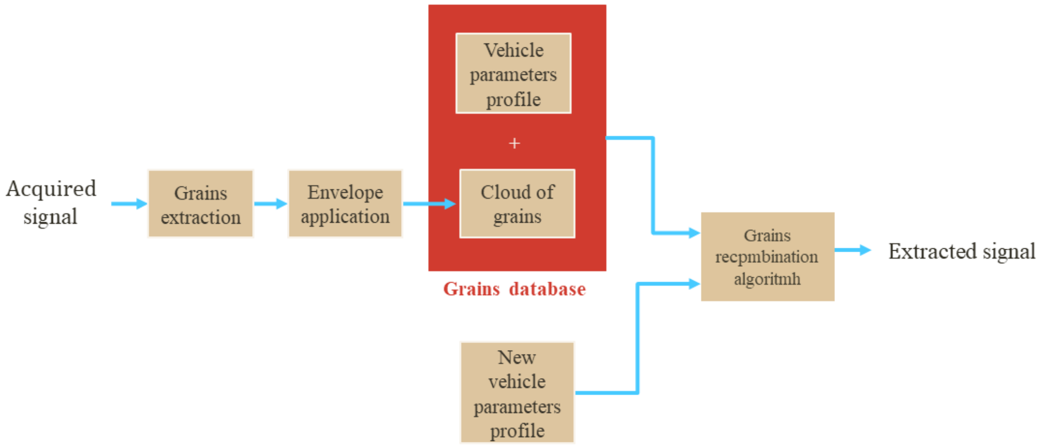


Figure 5.10: Granular synthesis steps

5.2.1 Grain envelope

Envelope refers to the manipulation performed on grains amplitude[20], affecting their shape especially at the boundaries.

This is needed to have transitions between grains in the recombination process as gradual as possible. Thus, after they are extracted, grains are windowed, which means an envelope is applied. Depending on the imposed amplitude shape modification, different grains can result and so the timbre is more or less preserved. Basically, envelope application generates the *fade-in* and *fade-out* effects at the grains boundaries.

As the grain shape follows exactly the one of the superimposed window[20], it is clear that the envelope governs both the grains duration and amplitude. Windows typical configuration is named *ASR envelope*[20] (see Figure 5.11), which stands for the Attack, Sustain and Release phases of the overall envelope duration:

- **Attack time**: it determines the slope of the window rising part. Here the grain amplitude is modified according to the steepness of such interval. The smoother the slope, the more natural the transition between the previous and the current grains will appear.
- **Sustain time**: it refers to the region where the grain amplitude is unmodified with respect to the original shape it has in the starting audio sample. The sustain time is crucial for preserving the sound timbre: if the grain is quite long but, on the contrary, the sustain phase is very short, it is equivalent to have a very short grains, so in a pulse effect which doesn't allow to appreciate the timbre.
- **Release time**: it determines the slope of the window in the falling. The bigger the slope, the rougher will be the transition to the next grain in the sequence.

The sum of these three contributions provides the envelope length, corresponding to the grain duration. This parameter, and especially the proportion between the three phases, establishes how distinguishable the contents of the grain are: the higher the duration and the longer the sustain phase, the more the sound timbre is preserved. Thus, when going towards smaller duration, the grain sound is perceived mainly as a click, that is the timbre information is no longer intact[20]. This means that envelope parameters are crucial for granular synthesis and their choice must be performed carefully so to achieve an optimal signal reconstruction.

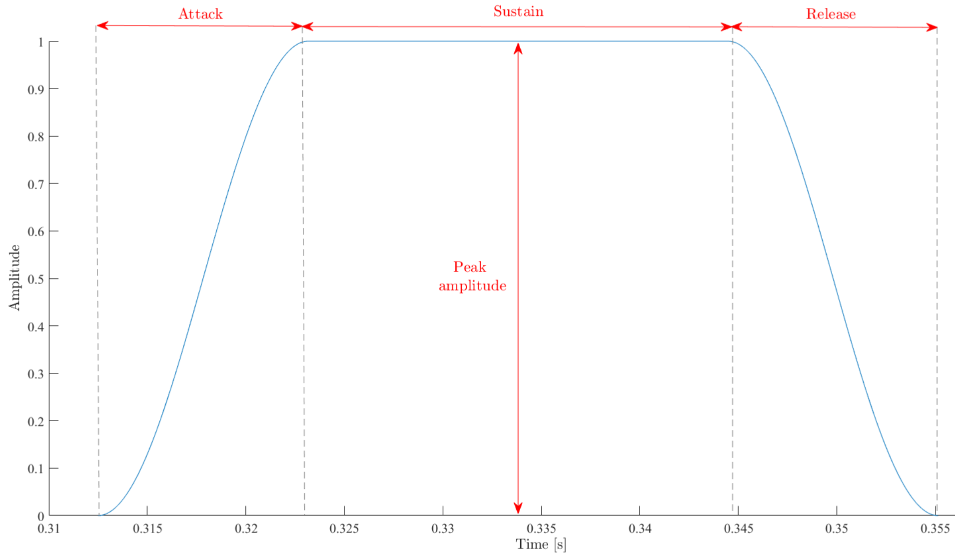


Figure 5.11: Window structure

An example of how the grain shape changes after windowing is provided in Figure 5.12: as clearly visible, the window application result is a smoothing effect at the grain boundaries since its amplitude fall to zero at the edges, due to the side attenuation introduced by the envelope.

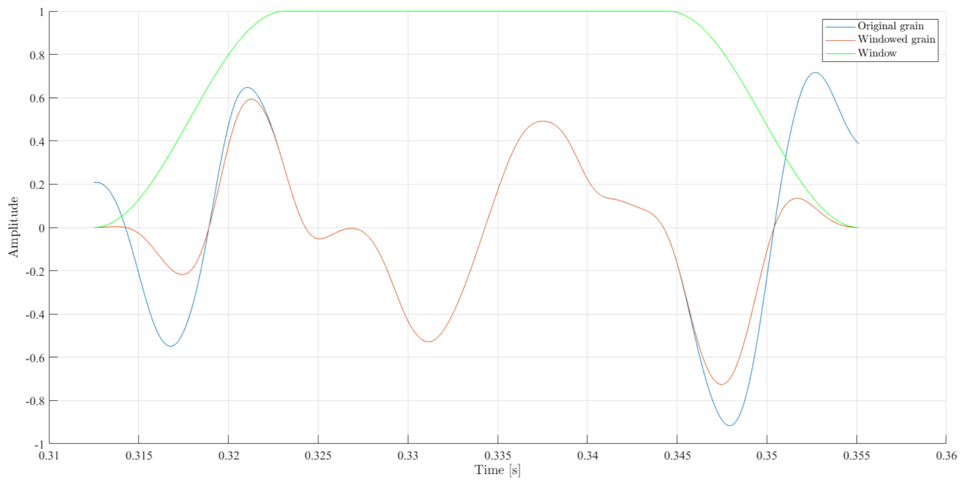
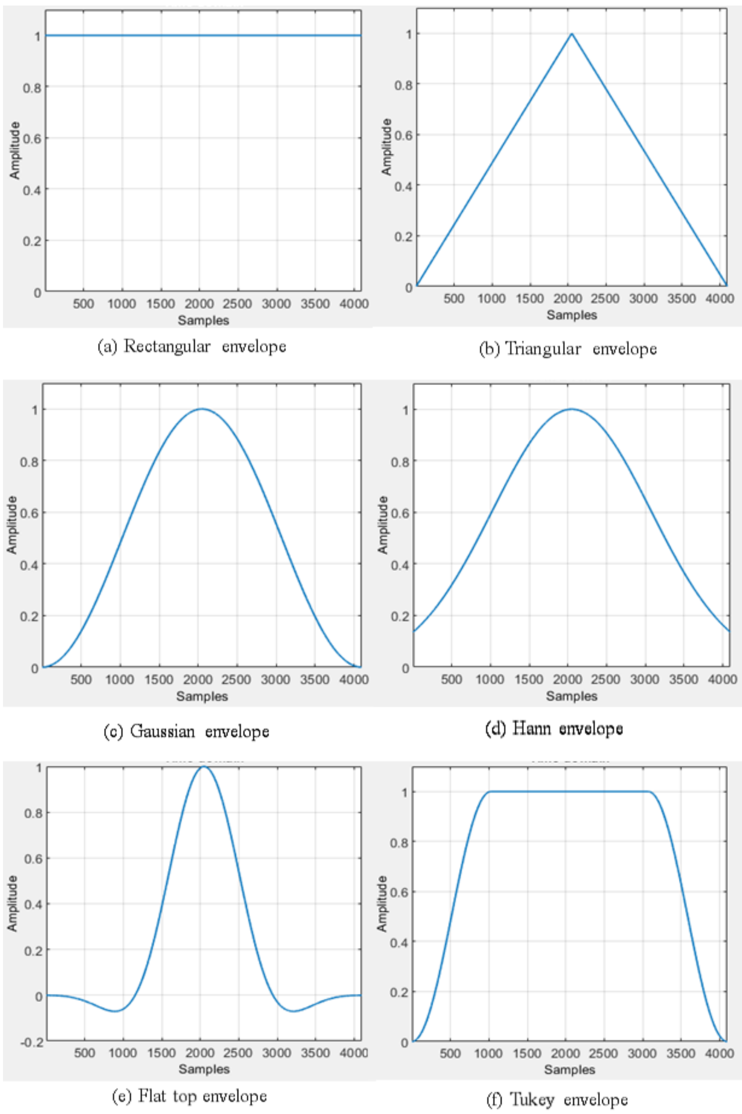


Figure 5.12: Examples of grain windowing.

Some possible envelopes are reported in Figure 5.13. The most suitable for sonic content preservation is an envelope with trapezoidal profile, as the sustain phase has a width that allows to maintain a considerable part of the grain unchanged from the amplitude point of view, preserving the acoustic fingerprint.



(a) Rectangular envelope (b) Triangular envelope (c) Gaussian envelope (d) Hann envelope (e) Flat top envelope (f) Tukey envelope

Figure 5.13: Examples of different envelope shapes.

5.2.2 Grains extraction

Mathematically, grains extraction and windowing can be expressed by Equations 5.6 and 5.5:

$$Grains = \sum_{i=1}^{N} g_i \tag{5.5}$$

with

$$g_i = x((i-1) \cdot L + 1 : i \cdot L) \cdot win(L) \tag{5.6}$$

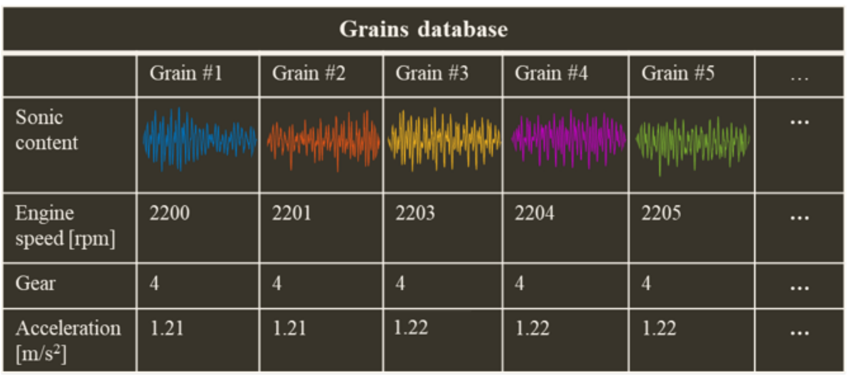
where,

- $Grains$ is the grains matrix, belonging to $\mathbb{R}^{L,N}$, as grains are arranged one after the other along the columns: N is the number of grains to be extracted, whose length is L;
- g_i the i^{th} extracted grain, with $i = 1, 2, ..., N$;
- x represents the original input from which grains are derived, that is the sound resulting from the harmonics extraction phase and the sound customization, discussed below in this dissertation;
- win is the applied envelope, which can be, for example, one of those proposed in Figure 5.12.

The outcome of this step is a collection of grains with the corresponding amplitude profiles and a set of associated vehicle state variables. In this way, each grain is characterized by its set of vehicle parameters. Such data set is the starting point for the recombination phase, which is further explained in the continuation of this chapter.

The richer the data set, the more accurate the synthesized signal and the higher the number of possible combinations of grains. By the way, even if not all the vehicle operating conditions are covered, that is some holes are in between two grains, the more similar grain to the needed one in terms of RPM at a given gear value can be considered, without producing a big error. Thus, if the database is quite rich, a missing grain can be approximated with one of the two closest ones with a negligible error. Generally, the acoustic difference between a recording at a certain RPM value p_1 and one at $p_1 + \mathrm{k}$, where k is a very small variation in the engine speed, is not appreciable, thus we can avoid to generate a new grain, which requires in additional computational effort.

Eventually, if the data base in not full enough, grains can be created according to the needed vehicle parameters set from the existing ones, performing frequency modulation of the most similar grains. The resulting database structure reflects the one in Figure 5.14.



Grains database						
	Grain #1	Grain #2	Grain #3	Grain #4	Grain #5	...
Sonic content						...
Engine speed [rpm]	2200	2201	2203	2204	2205	...
Gear	4	4	4	4	4	...
Acceleration [m/s^2]	1.21	1.21	1.22	1.22	1.22	...

Figure 5.14: Grains database structure

5.2.3 Grains recombination and audio artifacts

Grain recombination can be conducted either randomly or according to a predefined criterion. In this thesis work, the quasi-synchronous granular synthesis has been selected, driven by the incoming vehicle parameters: each grain, when extracted and windowed, is associated with a set of parameters (engine speed, acceleration, gear), so a matrix with N columns, where N is the number of extracted grains, and M rows, where M corresponds to the number of considered vehicle parameters (three in this case) plus one, that is the value of the grain amplitude profile (referred to as sonic content in Figure 5.14).

Then, depending on the current set of values RPM-acceleration-gear, grains reordering is performed: given a set of new parameters, a research in the grain matrix is performed to find the grain that better matches the requirements in terms of vehicle driving condition. The higher N, the higher the probability to find a correspondence in the database. If the database doesn't contain the needed grain, it can be either synthesized starting from the most similar one in terms of RPM at a given gear value, that is one of the two adjacent stored grains, or it can be approximated with the latter. Additionally, a modulation in the power content can be performed so to get the desired value and a more accurate vehicle parameters profile tracking. As presented in Chapter 3, the acceleration influences the energy content of the harmonic. This concept will be exploited in the following for the discussion of an additional customization method.

Grains are arranged one after the other with a partial overlap, in order to minimize the artifacts of recombination, perceived as unpleasant sound effects in the synthesized signal.

As formerly analyzed, *cross-fade* is an essential step towards a good synthesis product. Joining this with the partial overlap of the windowed grains, a softer transition is assured when merging them together.

OLA[40], standing for OverLap and Add, involves the partial overlap of each grain with the previous one. Such overlap is a fixed quantity. It aims at compensating the discontinuities in amplitude between two grains that have to be chained.

The drawback of such algorithm is that it doesn't preserve the phase relation between one grain and the following one, creating discontinuities in the newly synthesized signal .

OLA can be replaced with WSOLA (Waveform Similarity OverLap and Add) in order to achieve better transitions: since grains are different, so the sonic content changes from one to the other, the adoption of WSOLA algorithm[40] can lead to an improvement, that is in the minimization of amplitude discontinuities between grains in the synthesised signal[16].

5.2.4 Waveform Similarity OverLap and Add

WSOLA[40] is a technique widely employed in audio time-stretching to minimize discontinuities in the resulting signal. It aims at finding the point in time where the two grains better match in terms of amplitude.

It can be seen as an improvement of the OLA approach: the idea is to no longer have a constant overlap factor, but to introduce a tolerance on it. Basically, WSOLA introduces an interval where the junction point between two grains is contained and must be determined[36]. At each grain addition, the junction point is established according to the cross-correlation between the current grain and the previous one. Cross-correlation[41] measures the analogy between two signals. It can be defined according to Equation 5.7:

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t)y^*(t-\tau)dt \tag{5.7}$$

where,

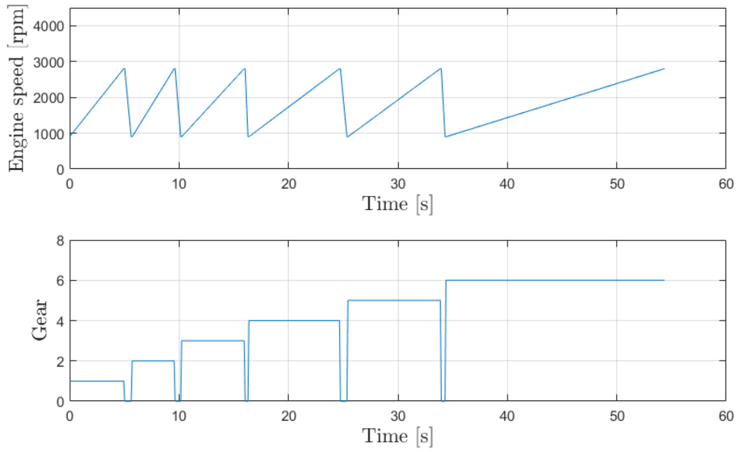
- * is the complex conjugate operation;
- $x(t)$ and $y(t)$ are the two audio samples;
- τ represents the time step.

WSOLA[40] exploits the maximum of the cross-correlation between two grains to determine where to joint them. Cross-correlation is not computed over the whole signal, but its calculation restricted to an interval. Thus, the new grain insertion cannot occur in whatever position along the previous grain, but it can only happen in a predefined range, whose width is defined by the research interval amplitude.
The point in time at which cross-correlation has a maximum identifies where the two grains are more similar: this is where they shall be merged. Performing grains transitions in this way allows to reduce discontinuities in the newly synthesized signal amplitude, that can be clearly observed when applying OLA strategy: discontinuities are evident both in the audio as clicks, which are unpleasant to be heard, and in the sound wave representation with angular points.

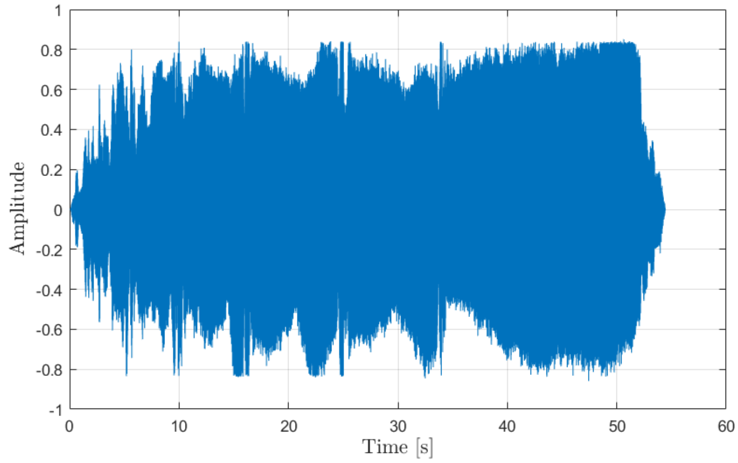
5.2.5 Synthesis outcome

The original signal along with its vehicle parameters profiles and spectrogram are shown below. Some examples of granular synthesis outcome are presented in Figure 5.16, to 5.22, where the original signal and the new synthesized one with the corresponding spectrograms are provided, along with the corresponding RPM-gear-acceleration profiles. The represented scenarios refer to idle, deceleration with gear shift down, acceleration with gear shift up and a mixed profile with gear shift up.

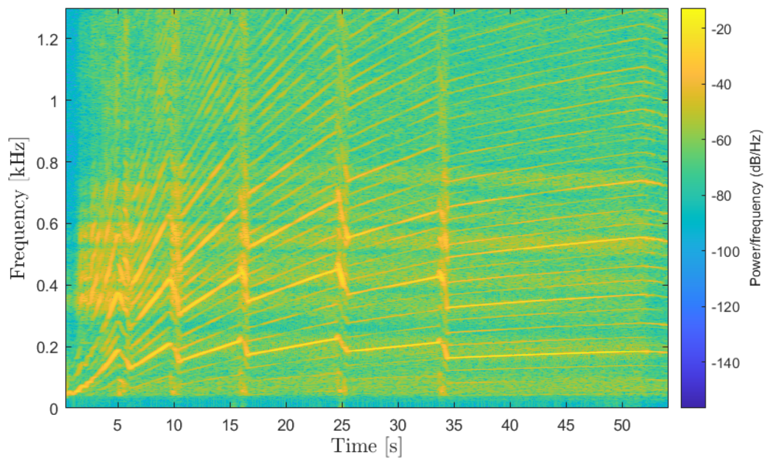
It can be easily noticed in the spectrogram that the harmonics and sub-harmonics follows the new imposed vehicle parameters profiles. Concerning the audio quality, the sound is quite realistic, unless some sporadic artifacts of reconstruction. As previously mentioned, WSOLA is an optimization technique and not an exact one, thus, some remaining artifacts are possible.



(a) RPM and gear profiles

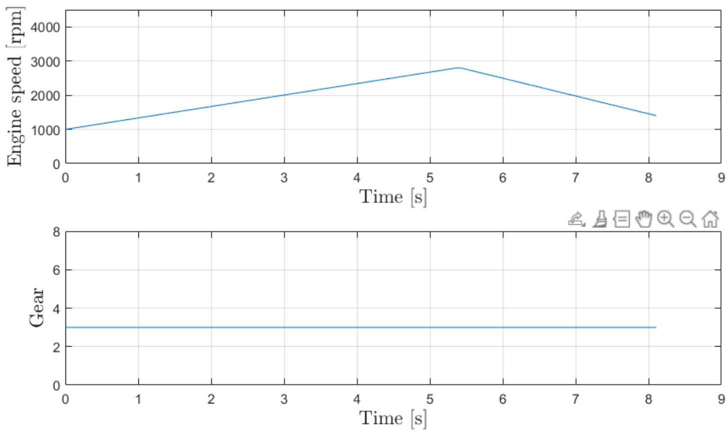


(b) Audio signal

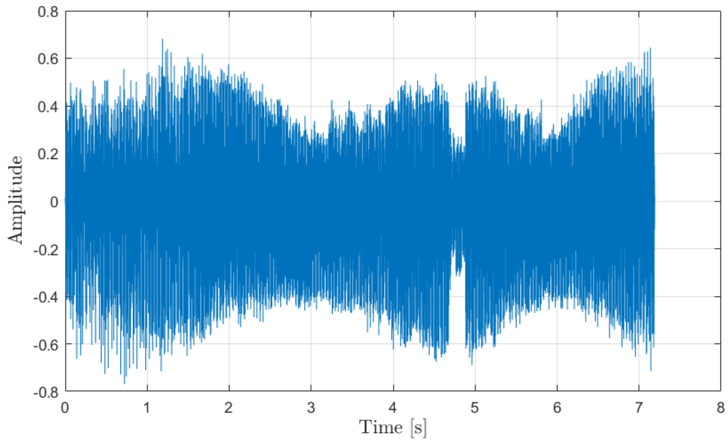


(c) Spectrogram

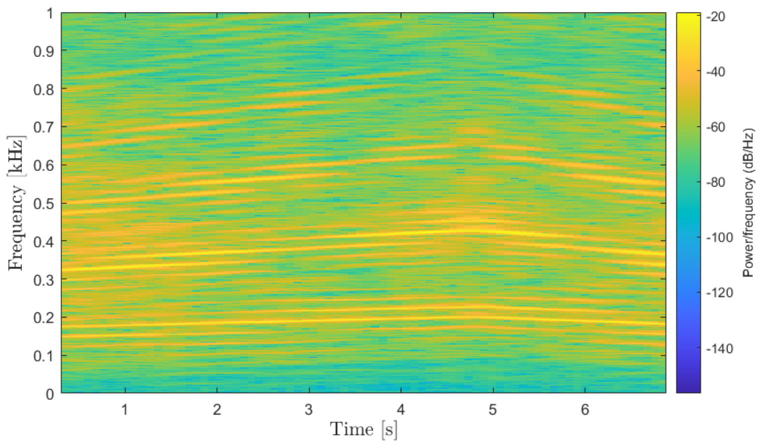
Figure 5.15: Original signal



(a) RPM and gear profiles

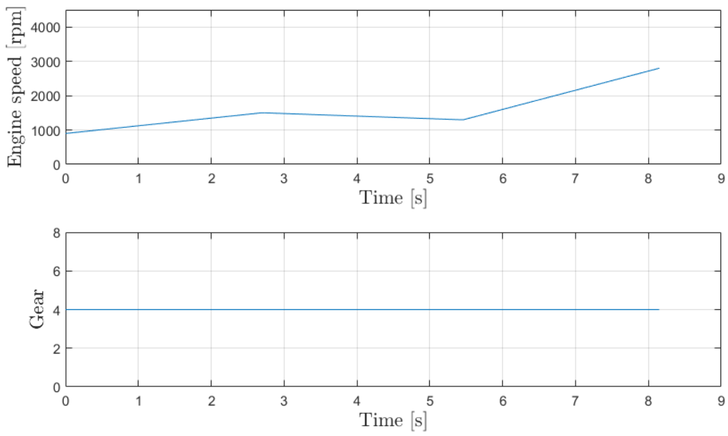


(b) Audio signal

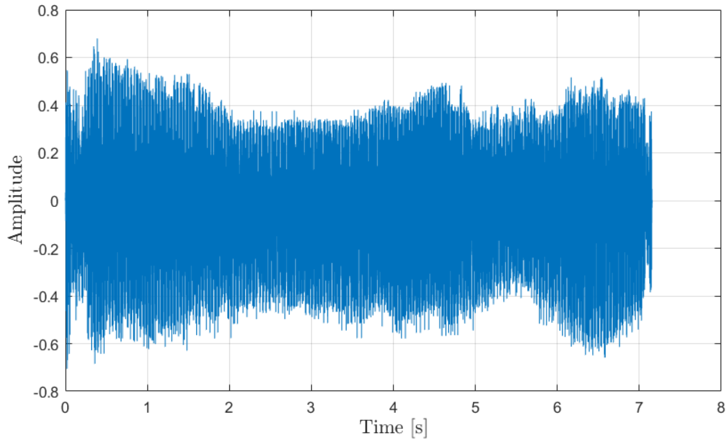


(c) Spectrogram

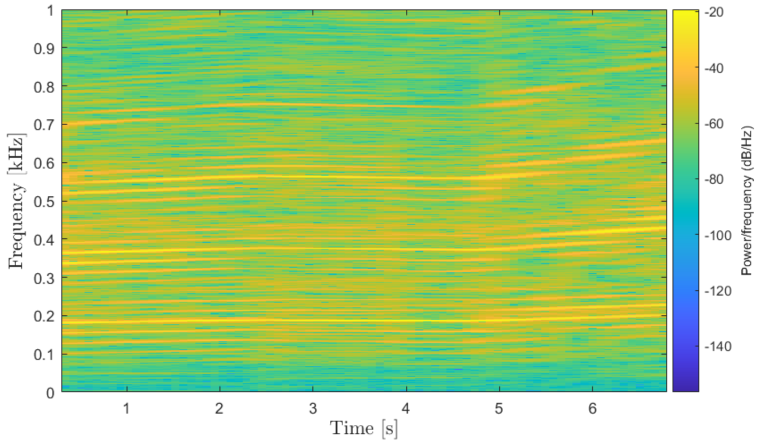
Figure 5.16: Granular synthesized signal - 1^{st} example



(a) RPM and gear profiles

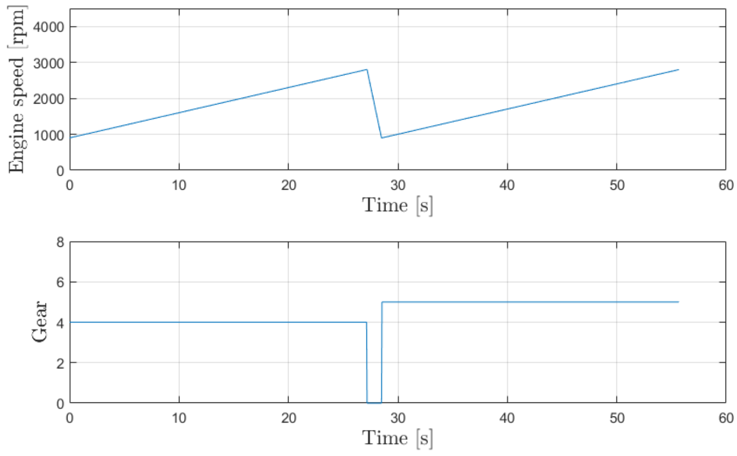


(b) Audio signal

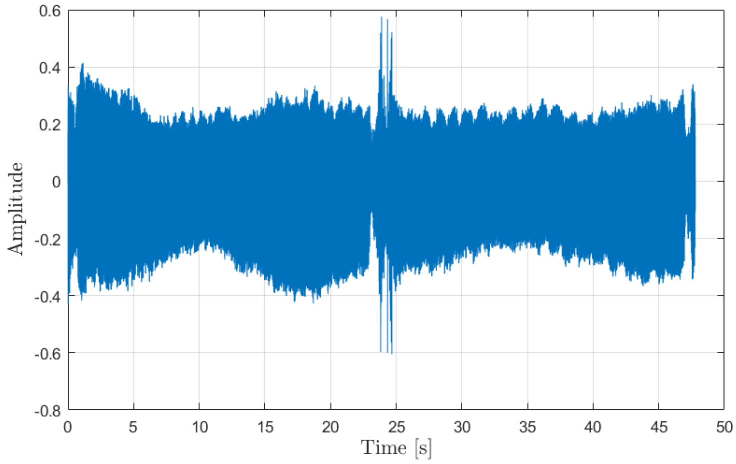


(c) Spectrogram

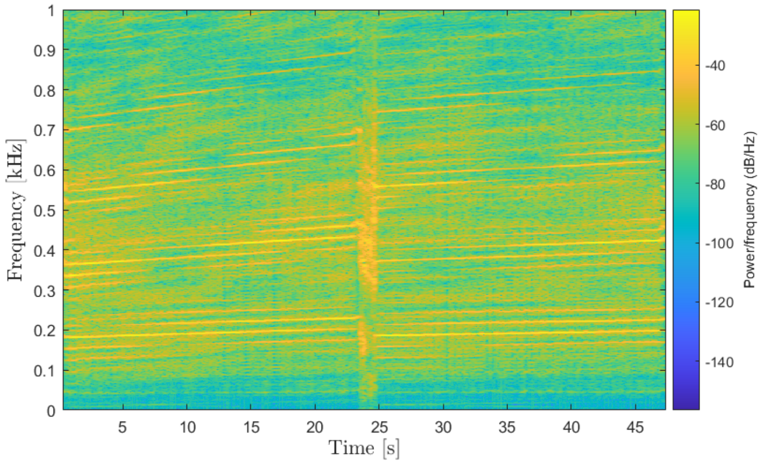
Figure 5.17: Granular synthesized signal - 2^{nd} example



(a) RPM and gear profiles

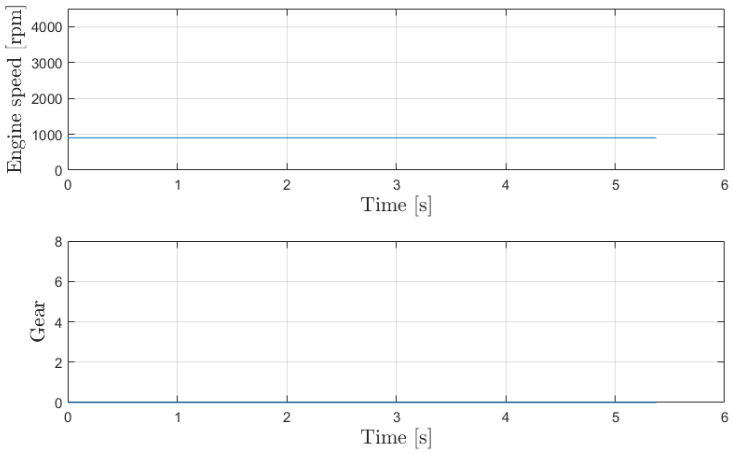


(b) Audio signal

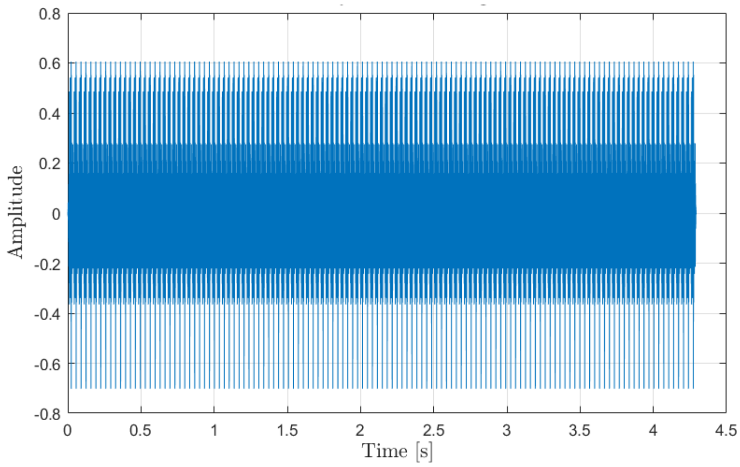


(c) Spectrogram

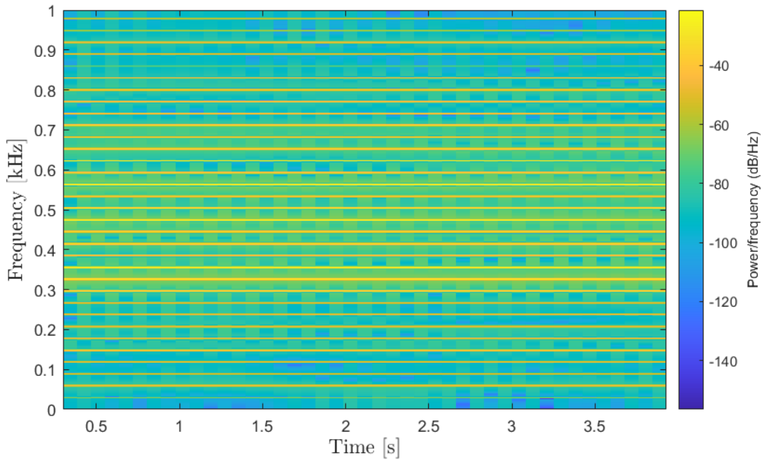
Figure 5.18: Granular synthesized signal - 3^{rd} example



(a) RPM and gear profiles

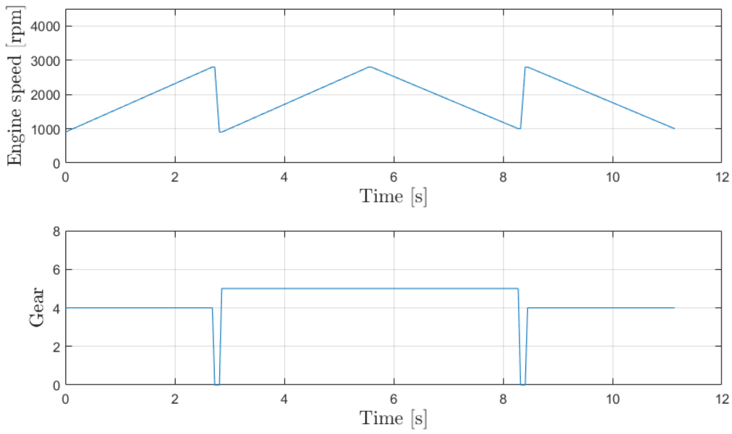


(b) Audio signal

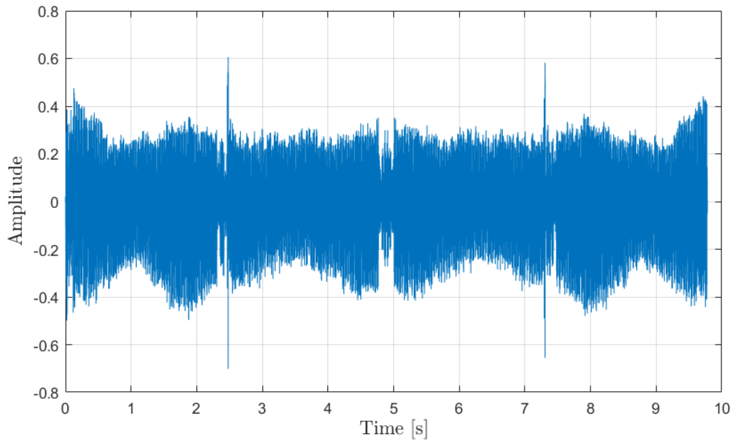


(c) Spectrogram

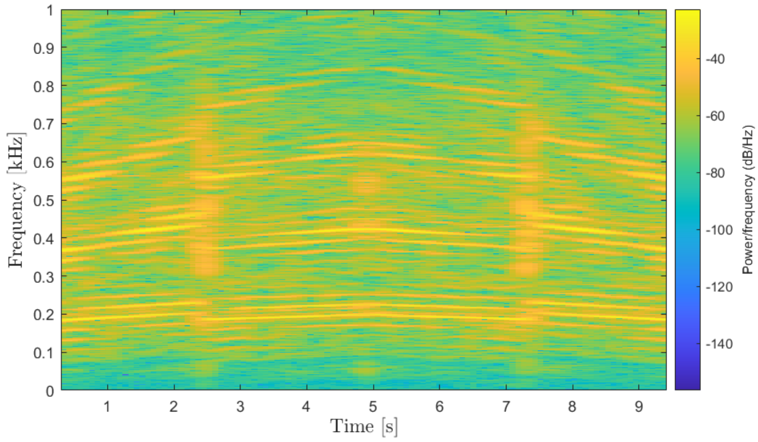
Figure 5.19: Granular synthesized signal - 4^{th} example



(a) RPM and gear profiles

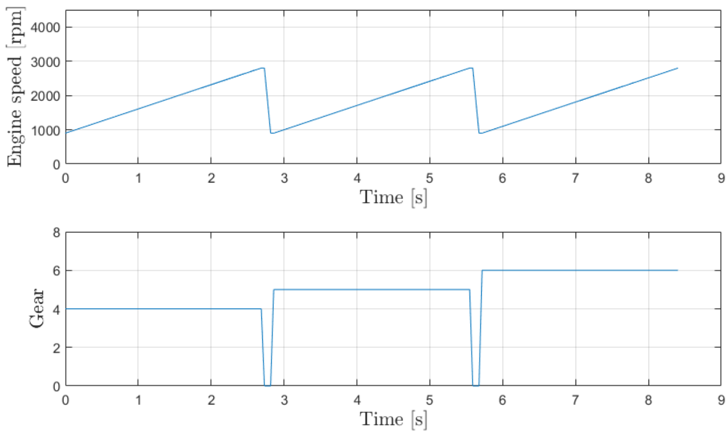


(b) Audio signal

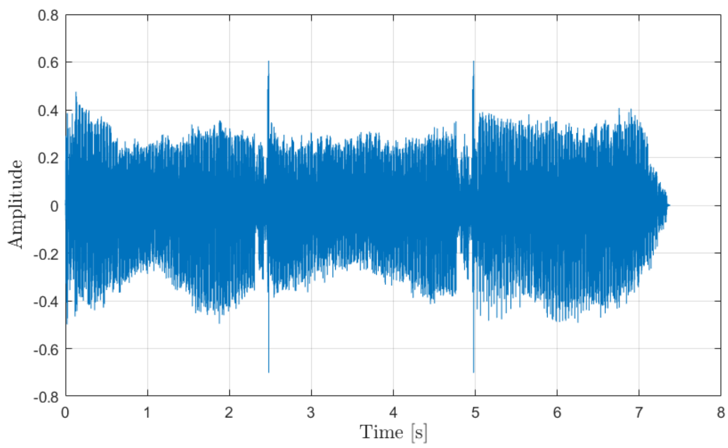


(c) Spectrogram

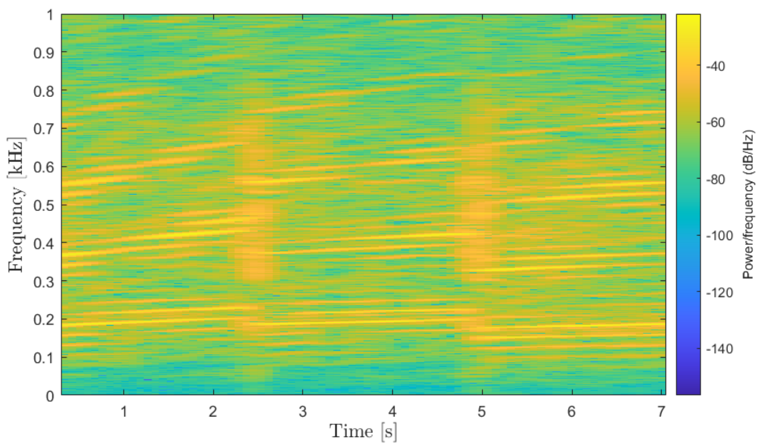
Figure 5.20: Granular synthesized signal - 5^{th} example



(a) RPM and gear profiles

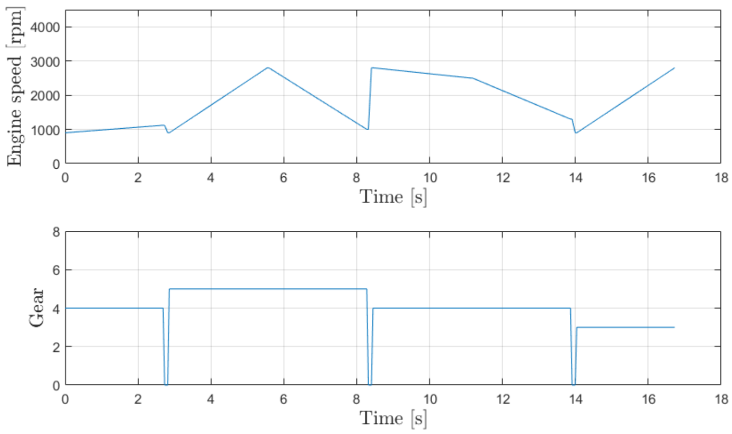


(b) Audio signal

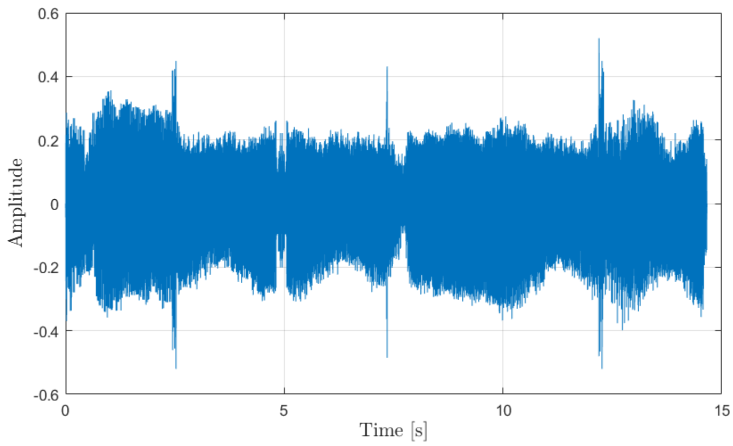


(c) Spectrogram

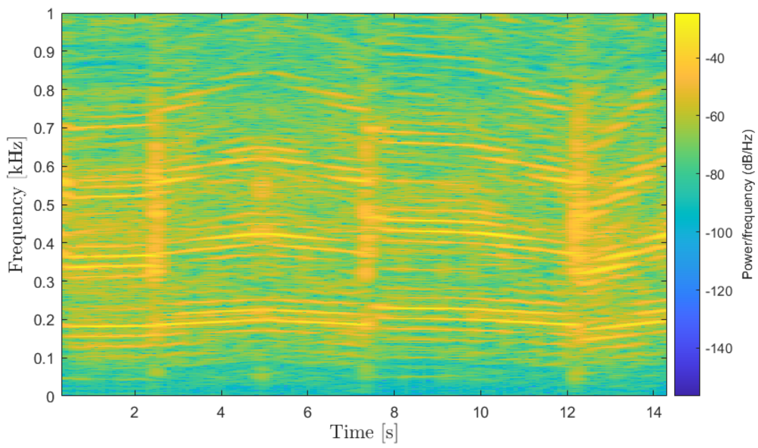
Figure 5.21: Granular synthesized signal - 6^{th} example



(a) RPM and gear profiles



(b) Audio signal



(c) Spectrogram

Figure 5.22: Granular synthesized signal - 7^{th} example

5.3 Sound customization

As mentioned in the introduction of this dissertation, a relevant aspect in the car design is represented by the opportunity to customize some aspects of the vehicle. One of this is the engine sound.
This can be applied not only to electric vehicles, which needs an acoustic system to address the lack of noise issue, but also to other categories of vehicles, which already have a clearly audible sound.

Often, car makers require the possibility to give more emphasis to some engine sound harmonics so to let them prevail upon the others. This allows to enhance the timbre and get a more powerful sound.
Generally, the harmonics to be emphasised are the odd ones, thus the first, the third and the fifth. Also higher order harmonics can be highlighted. This effect is obtained by inserting a tunable gain on each tracked harmonic and sub-harmonic of the original signal, before performing granular synthesis. This is quite easy, as the harmonics decomposition has already been implemented to perform noise elimination and granular synthesis. Furthermore, such components can be either enhanced, by setting a gain bigger than 1, or attenuated, imposing a gain smaller than 1. Default gain value is set to 1, that is no correction is performed on the harmonic energy.

Car perception can be modified by making it sounds like a different one in terms of cylinder configuration.
Another modification on the car sound can be the addition of various sound effects, so to further enrich the perceived audio. Effects include, as an exam-

ple, reverberation, tremolo, echo, equalization and panning.

5.3.1 Cylinder configuration modification

As mentioned in Chapter 3, the number of sub-harmonics depends on the number of cylinders. Thus, one possibility for sound customization is represented by the exaltation of sub-harmonics components which normally feature a low energy content and thus are not visible on the spectrogram of the signal. This aims at perceiving a more full and rich engine sound, coming from a motor with a higher number of cylinders with respect to the actual one. Thus, such modification allows to make a car sounds like a more powerful one having more cylinders.

This is performed by considering the relationship between the sub-harmonics of interest and the fundamental one, stated by Equation 2.3. Concerning the amplitude profile of each new sub-harmonic to be synthesized, it is approximated with the one of the succeeding harmonic or sub-harmonic already determined at the filtering stage.

As reported in Chapter 3, sub-harmonics aim at enriching the sound and making it more full. Their increment makes road users and car passengers perceive a sound that features a vehicle with more cylinders than the actual number, depending on the number of added synthesized contributions.

A further propose concerning the cylinder settings, is to add on top of the newly synthesized signal a reverb so to simulate a cross-plane crank cylinder configuration, featuring a more deep and strong sound with respect to a flat-plane crank one or an in-line cylinder positioning.

5.3.2 Harmonics and sub-harmonics enhancement

Harmonic components energy can be tuned so to enhance or attenuate them. This is done by adding two kind of gains on each contribution: a constant gain over the time and a variable one, which changes following the acceleration profile related to the signal.

Gains aims at modifying the weight of the different harmonics in the overall sound to get a desired outcome.

Basically, a constant multiplication factor, whose default value is 1, is applied on each harmonic (and sub-harmonics). A variable gain is included too: it aims at introducing the dependency of the energy content from the acceleration value in the time, so as acceleration increases, the power intensity grows too. The effect is an additional exaltation of the timbre during acceleration phases.

Exaltation of harmonics is performed according to Equation 5.8:

$$PSD_{H_i}(t) = K_{C,H_i} \cdot K_{V,H_i} \cdot PSD_{0,H_i}(t) \tag{5.8}$$

$$i = \tfrac{1}{N_{cyl}}, ..., N_H$$

$$K_{C,H_i} \in \mathbb{R}$$

$$K_{V,H_i} = \{0, 1\}$$

where,

- $PSD_{H_i}(t)$ is the new power content of harmonic i;
- $PSD_{0,H_i}(t)$ is the original power content of harmonic i;

- K_{C,H_i} is the constant gain on harmonic i;
- $K_{V,H_i} = e^{acc(t)}$ is the variable gain on harmonic i;
- $acc(t)$ is the acceleration profile.

Some examples of spectrogram modification according to the constant gain introduction are presented in the following (see Figures 5.23 to 5.26).

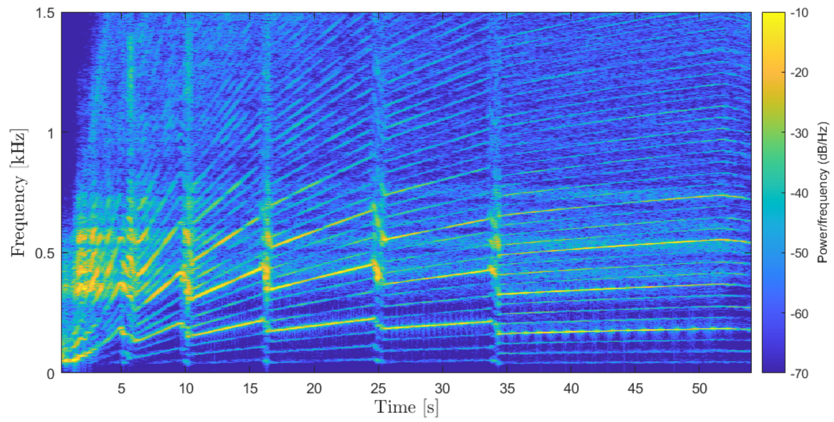


Figure 5.23: Spectrogram with a constant gain equal to 1 (default) on each harmonic and sub-harmonic

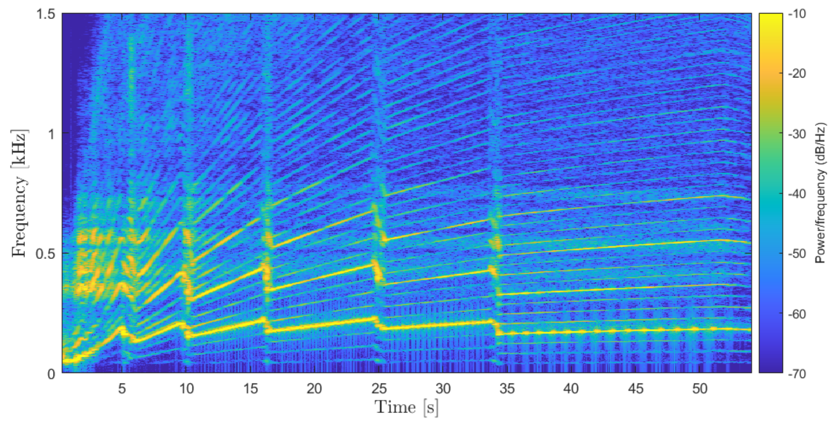


Figure 5.24: Spectrogram with a constant gain equal to 4 on the 1^{st} harmonic and 1 otherwise

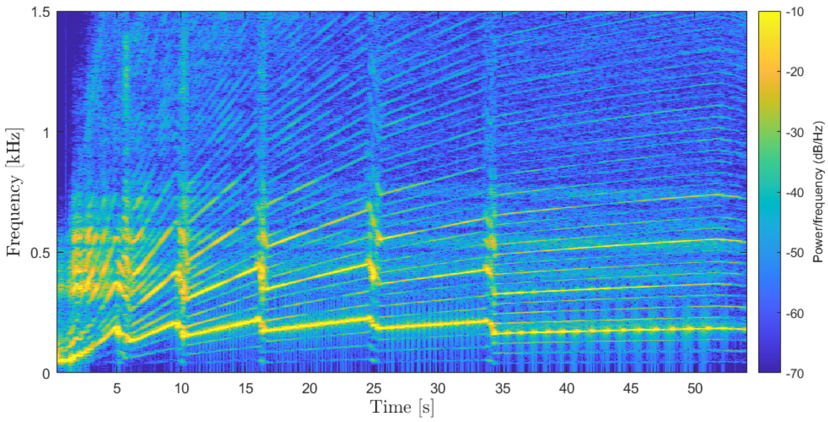


Figure 5.25: Spectrogram with a variable gain proportional to the acceleration on the 1^{st} harmonic and 1 otherwise

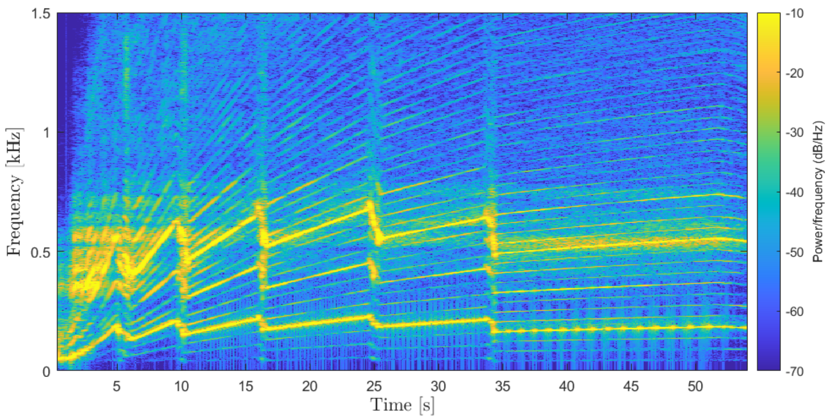


Figure 5.26: Spectrogram with a variable gain proportional to the acceleration on the 1^{st} harmonic, constant gain of 20 on the 3^{rd} harmonic and 1 otherwise

It is clearly observable that, as the gain on the harmonic increases, its energetic content is enriched too, thus, it will have a bigger impact on the overall sound. Concerning the variable gain, it can be noticed that as the harmonic power grows along with acceleration and also the neighborhood of the component is enhanced accordingly, thus, as acceleration increases, the harmonic is thicker and more energetic.

When adopting the variable gain, an envelope on the considered harmonic is inserted. It is created according to the vehicle acceleration profile. Thus, when accelerating more, the sound is further enhanced, letting road and car users to perceive the velocity variations in a more appreciable manner.

Conclusion

The proposed algorithm is able to perform the audio tracking of a RPM-gear-acceleration profile, producing a quite realistic engine sound.

The obtained sounds can be played in the cabin through the vehicle loudspeakers and outside, by means of loudspeakers placed, as an example, in the engine compartment. This improvement is part of the practical implementation phase, which represents a future step for this work.

As the database is built on recordings found on the Internet, a further development could be represented by the integration of the algorithm with real recordings along with CAN vehicle acquisitions, according to the analysis and setup discussed in Chapter 4.

Such improvements have not be performed within this thesis work due to practical issues, represented by the lack of suitable instrumentation and the limited time spent in the company, due to the COVID-19 emergency: the advance of the project has been slowed down by the lock-down period, which has forced to work from home, without possibility to perform activities on board of the vehicle used for building the setup for a huge period of time.

Additionally, concerning the future improvements, the database width can be augmented so to include more vehicles operational conditions and the dependency of the algorithm on other vehicle parameters than the ones considered here can be included, so to perform other manipulations on the newly synthesized sound: this would lead to a more sophisticated engine sound enhancement system which takes into account other vehicles state variables, offering a higher quality audio generation.

Furthermore, other audio effects than reverberation can be considered for sound customization, so to achieve a more original texture, for fitting car users requests.

For the implementation on the vehicle, the definition of the quantities equivalent to acceleration, gear, and engine speed in the thermal vehicles have to be identified for the electric motor. This means that the quantities determining the driving operations in the electric vehicle have to be mapped to adapt the developed algorithm to the electric power train. In an EM, the control can be actuated considering the current and the voltage: by monitoring their variations, the equivalent parameters needed for letting the algorithm work can be obtained.

References

[1] Ishimura K. Singh S. Wakamatsu T. “Introduction to Active Sound Management Technology”. In: DENSO TEN Technical Review Vol. 1 (2018), pp. 23–30.

[2] https://www.evgo.com/.

[3] https://www.hawaiianelectric.com/products-and-services/electric-vehicles/electric-vehicle-basics.

[4] *Slides from Stefano Alberto Malan, Politecnico di Torino, about hybrid electric vehicles, Automotive Control Systems course.*

[5] https://www.skoda-storyboard.com/.

[6] https://www.goodenergy.co.uk/.

[7] David A. Heitbrink and Steve Cable. “Design of a Driving Simulation Sound Engine”. In: (2007).

[8] https://www.electrive.com/.

[9] https://enginevox.com/.

[10] https://automotive.bose.com/.

[11] https://www.autoblog.com/.

[12] https://www.jaguar.co.uk/.

[13] https://www.physicsclassroom.com/class/sound/u11l1c.cfm.

[14] http://hyperphysics.phy-astr.gsu.edu/.

[15] https://physics.info/.

[16] https://www.mathworks.com/.

[17] https://www.zapsplat.com/.

[18] http://digitalsoundandmusic.com/.

[19] http://users.cs.cf.ac.uk/Dave.Marshall/Multimedia/PDF/05_Audio_Synthesis.pdf.

[20] Timothy Opie. "Sound in a Nutshell: Granular Synthesis". In: 1999.

[21] https://granularsynthesis.com/.

[22] Bernd Heißing and Metin Ersoy. "Ride Comfort and NVH". In: (Jan. 2011). DOI: 10.1007/978-3-8348-9789-3_5.

[23] https://www.uti.edu/.

[24] https://www.cjponyparts.com/resources/flat-plane-vs-cross-plane-crankshafts.

[25] *Slides from Mauro Velardocchia, Politecnico di Torino, about longitudinal dynamics, Technologies for Autonomous Vehicles course.*

[26] *Slides from Mauro Velardocchia, Politecnico di Torino, about tyre, Technologies for Autonomous Vehicles course.*

[27] Sciabica et al. "Towards the timbre modeling of interior car sound". In: *Proccedings of the ICAD number15* (May 2009).

[28] Jan Jagla, Julien Maillard, and Nadine Martin. "Sample-based engine noise synthesis using an enhanced pitch-synchronous overlap-And-Add method". In: *The Journal of the Acoustical Society of America* 132 (Nov. 2012), pp. 3098–108. DOI: 10.1121/1.4754663.

[29] https://www.alfaromeo.it/.

[30] https://invensense.tdk.com/.

[31] https://en.wikipedia.org/wiki/Signal-to-noise_ratio.

[32] https://www.inventure-automotive.com/.

[33] https://www.csselectronics.com/.

[34] https://www.ti.com/.

[35] Jan Jagla, Julien Maillard, and Nadine Martin. "Sample-based engine noise synthesis using a harmonic synchronous overlap-and-add method". In: *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on* (Mar. 2012). DOI: 10.1109/ICASSP.2012.6287894.

[36] https://it.mathworks.com/.

[37] https://www.dsprelated.com.

[38] Reto Pieren, Thomas Bütler, and K. Heutschi. "Auralization of Accelerating Passenger Cars Using Spectral Modeling Synthesis". In: *Applied Sciences* 6 (Dec. 2015), p. 5. DOI: 10.3390/app6010005.

[39] https://support.apple.com/it-it/guide/logicpro/lgsife419a20/mac.

[40] Jonathan Driedger and Meinard Müller. "A Review of Time-Scale Modification of Music Signals". In: *Applied Sciences* 6 (Feb. 2016), p. 57. DOI: 10.3390/app6020057.

[41] https://www.allaboutcircuits.com/technical-articles/understanding-correlation/.