Faculdade de Engenharia da Universidade do Porto

CONVOLUTIONAL NEURAL NETWORKS APPLICATION FOR THE CLASSIFICATION OF POWERTRAIN BOOMING NOISE ON INTERNAL COMBUSTION ENGINE PASSENGER CARS

Pedro Miguel Gomes Leite

up201505887@fe.up.pt

Mestrado Integrado em Engenharia Mecânica

Dissertation for Master's Degree

Supervisors:

Prof. António Ramos Silva

Eng Claudio Colangeli

March 2021

Abstract

On the automotive industry, for the purpose of meeting customers' expectations of comfort, great attention must be paid regarding the sound quality of the vehicle. Among the vast scope of sounds which can be heard inside the cabin of Internal Combustion Engine (ICE) passenger cars, the booming noise is considered a very annoying acoustic phenomenon. The assessment of the impact that this low frequency Noise, Vibration and Harshness (NVH) problem has on passengers' comfort often relies on subjective evaluation methods.

Throughout this dissertation, a method for evaluating the powertrain booming sensation in end-of-line applications was analyzed with the intention of creating a tool that can be used by the NVH engineer, minimizing the need of subjective evaluation methods. This method consists in the application of Neural Networks (NN) techniques such as Convolutional Neural Networks (CNN) to perform classification tasks on the spectrogram images of the raw sound signal experienced by the driver, in accelerating conditions. In order to generate the database for training the networks, modifications on digital twins of three cars were done to induce the presence of the booming sensation associated with the Engine Firing Rate (EFR). This was done in a fully automated way, creating a large variety of scenarios.

Different CNN architectures were tested for the *Yes*/*No* classification of booming, reaching test accuracies around 95.5% on the best architecture. Later, these tests were also conducted for the *YesSevere*/*YesMild*/*No* classification of booming, in which the accuracies dropped around 2%.

Resumo

Na indústria automóvel, com o objetivo de atender às expectativas de conforto dos clientes, grande atenção deve ser dada à qualidade sonora dos veículos. Dentre a vasta gama de sons existentes no interior dos automóveis com motor de combustão interna, o "booming" é considerado um fenómeno acústico muito incomodativo. A avaliação do impacto que este ruído de baixa frequência tem no conforto dos passageiros é dependente de métodos de avaliação subjetivos.

Ao longo desta dissertação, foi desenvolvido um método para a avaliação da sensação de "booming" associado à cadeia cinemática para ser utilizado em aplicações de fim de linha. Desta forma, foi criada uma ferramenta que poderá ser utilizada por um técnico especializado, minimizando a necessidade de métodos subjetivos de avaliação. Este método consiste na aplicação de técnicas de redes neuronais convolucionais para realizar tarefas de classificação nas imagens correspondentes aos espectrogramas dos sons detetados pelo condutor, em condições de aceleração. Para gerar a base de dados a ser utilizada pelas redes neuronais, foram realizadas modificações aos sons gravados de três carros (gémeos digitais). Estas foram feitas para induzir a presença de "booming". A criação da base de dados foi feita de uma forma automática, criando uma grande variedade de cenários.

Para a classificação entre *Sim*/*Não* de "booming", foram testadas diferentes arquiteturas de redes convolucionais, tendo-se atingido precisões de cerca de 95,5% na melhor arquitetura. Posteriormente, a mesma metodologia foi aplicada para a classificação entre *SimSevero*/*SimLeve*/*Não*, tendo-se registado uma queda de cerca de 2% nas precisões.

Acknowledgements

The work developed throughout this thesis, as all my personal and professional achievements represent a build-up of knowledge and experiences over many years, achieved with the help of several individuals to whom I would like to express my acknowledgement.

First, to Faculdade de Engenharia da Universidade do Porto, where I spent a lot of my time during the past 5 years. I would like to address all the professors and staff, especially professor António Ramos Silva, who was my supervisor and gave me insightful points to improve my work.

To Siemens PLM Software, in the person of my supervisor Eng. Claudio Colangeli who provided the means for this work and was always available to help me at any time I needed.

I would like to leave a special acknowledgement to all friends I made over the last 5 years at FEUP, especially to Francisco Afonso and Raúl Silva, with whom I share a friendship since our first academic year. They were always there for me and the memories we created together will stay with me for the rest of my life.

To my roommates Bruno Ribeiro, Fernando Fernandes and João Ferreira, with whom I shared a house during my years at FEUP turning out in quite unique adventures.

To my family, Maria Gomes, Vítor Leite and Andreia Leite for the role model that they are, and also for the unconditional support, affection, comprehension and love.

Contents

List of figures

List of tables

Nomenclature

ICE – Internal Combustion Engine

NVH – Noise Vibration and Harshness

EFR – Engine Firing Rate

TPA – Transfer Path Analysis

SQE – Sound Quality Equivalent

RPM – Revolutions Per Minute

AI – Artificial Intelligence

ML – Machine Learning

NN – Neural Network

CNN – Convolutional Neural Network

RNN – Recurrent Neural Network

1 INTRODUCTION

The present document was composed by Pedro Miguel Gomes Leite, on the first semester of the 2020/2021 academic year, as his dissertation work paper for obtaining a Master's degree on Mechanical Engineering - Automation. It was a project developed in result of a partnership between Faculdade de Engenharia da Universidade do Porto and Siemens PLM Software. The student was supervised by Prof. António Ramos Silva as his advisor at Faculdade de Engenharia da Universidade do Porto and Eng Claudio Colangeli as his advisor at Siemens PLM Software.

1.1 Framework

As society becomes more aware of the impact that prolonged exposure to noise has in human health, one major concern has been to reduce the amount of sound emitted by man-made machines. One common environmental noise source is traffic. In this context, both the exterior and interior sound of a car poses as a crucial factor in the development process of a vehicle, not only to meet health requirements, but to align with customer expectations of comfort. Due to this, a lot of effort is put into improving interior acoustics of vehicles.

Noise Vibration and Harshness (NVH) is defined as the study and modification of the noise and vibration characteristics of vehicles in the whole frequency range, and its effects on passengers' perception. A common methodology of dealing with noise and vibration in cars is Transfer Path Analysis (TPA). It adopts a system analysis approach, where excitations produce responses through the transfer function of the system, from the noise and vibration source to the receiver. This allows the study of the possible ways of energy transfer from the various sources of excitation to a given target location, thus providing valuable insight into the mechanisms responsible for the problem (de Ponseele 2012).

Even though noise and vibration can be objectively measured, the harshness evaluation requires other tools since it is considered a subjective quality, meaning the human sensation behavior changes from person to person, on a geographic and even on a more personal standpoint. Because of this, converting customer perception into objective measurements and to correlate them is often very challenging for NVH engineers.

Despite the increasing electrification and hybridization of vehicles, the study of NVH problems on conventional Internal Combustion Engines (ICE) is still prevalent on our current time. In this category of vehicles, the engine is the system having the most impact on the overall acoustic scene (Siano and Panza 2017).

Booming noise is an NVH issue which occurs at low frequencies, often caused by driveline resonances which get excited by the main engine orders (Wellmann et al. 2011). The evaluation of this phenomenon on end-of-line applications usually depends on subjective evaluations made by a jury test.

To perform anomaly detection, Machine Learning (ML) techniques are commonly used. Recent progresses on a field of ML called CNNs have been achieved, especially in computer vision tasks. This is allowing the application of these models to a variety of fields.

1.2 Objectives

With the main objective of applying NN models for the classification of simulated NVH defects, the following tasks were performed along the way:

- Gaining insight in the state of the art of data-driven methods applied in the NVH domain, and in NVH engineering in general;
- Gaining skills in digital processing and ML;
- Gaining skills in full vehicle sound synthesis techniques;
- Further developing a booming sound simulator prototype to synthesize vehicle sounds with different booming scenarios, in a fully automated way, creating a diverse database;
- Using a NN approach to perform classification tasks in a variety of scenarios.

1.3 Structure

This work is divided into eight chapters. Chapter 1 gives the reader an introduction about the topics covered throughout the dissertation, contextualizing them in the field of Mechanical Engineering and in the current market. It also covers the main objectives of the dissertation.

Chapter 2 allows the reader to obtain information about the booming noise, its general importance nowadays in the automotive industry and the analysis methods used to study this phenomenon. Information about ML algorithms, focusing on the CNNs is also given. This chapter also presents the state of the art of the booming noise, relating it with ML.

Chapter 3 is about the vehicle sound decomposition, starting by presenting the experimental conditions of the used cars' data. Then, the data was analyzed to assess the most predominant orders for decomposition. Finally, details about the decomposition process are also present.

Chapter 4 exposes information about the vehicle sound synthesis approach. Details regarding the simulation of the booming noise in the sound synthesis are also addressed.

Chapter 5 focuses on the implementation of CNNs for the classification of the booming noise. Different architectures are explored, and the results obtained are exposed.

Chapter 6 presents a discussion of the obtained results.

Chapter 7 concentrates in making conclusions regarding the work developed and the experimental results presented, as well as possible future topics to explore.

2 STATE OF THE ART

2.1 Booming noise

Booming noise is one of the various noises which can be heard inside of a car. It occurs at low frequencies, below 200 Hz (Shin et al. 2009). Usually it is perceived as an annoying acoustic feature by the vehicle's passengers, who define it as being like a continuous bass drum roll, a distant thunder sound or a deep resonant sound like an explosion (Zhen et al. 2011). This subjective phenomenon was revealed to affect more harshly people from Asian countries comparing to western countries in general (Toi, T, Isoyama, H, Ogu, Y, Saito, H, Hoshino, H, Ishikawa, M, et al. 2004). However, the booming experience even changes from person to person, making it difficult for NVH engineers to convert customer perception into objective measurements and to correlate them. There are different origins of booming noise in a vehicle (Mansinh et al. 2012), namely:

- Engine excitation;
- Excessive radiation from intake/exhaust systems;
- Road excitation;
- Wind fluctuation;
- Etc...

When coupled with the passenger's cabin resonance, all of these cause the rise of pressure inside the car's cabin, provoking discomfort on the passengers. Among all of these, the engine is the system having the most impact on the overall acoustic image in a vehicle powered by an internal combustion engine (Siano and Panza 2017). Therefore, the analysis of the booming noise caused by the engine excitation is very relevant for meeting costumers' expectations of comfort.

An engine downsizing trend has been noticed, resulting from a demand for more fuel-efficient powertrains and reduction of CO_2 emissions. The technologies emerging from this increase engine torsional excitations from the engine firing rate (Wellmann et al. 2011), which can be calculated by the expression (1):

$$EFR = \frac{N}{60 \cdot \varepsilon} \cdot N_{cyl}\ [Hz] \quad (1)$$

Where EFR is the engine firing rate and ε is equal to 1 for a two-stroke engine and 2 for a four-stroke engine. N is the engine rotational speed in RPM (Revolutions Per Minute) and N_{cyl} represents the number of cylinders.

Torsional Vibration

Torsional vibrations are angular vibrations of an object, generally a shaft along its axis of rotation, and are created by time-altering torques. These fluctuations are superimposed on the steady rotational speed, causing its variation within a rotation cycle. In figure 1 it can be seen an example of a velocity profile of a 4-cylinder engine during a run-up test. Looking closely, 100 RPM fluctuations can be observed.

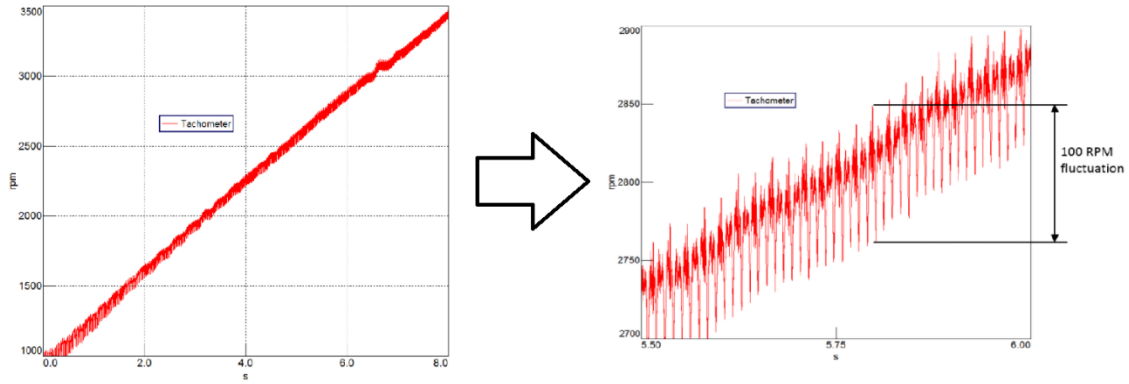


Figure 1, time vs. RPM of an engine run-up, courtesy of Siemens Industry Software.

In an ideal power transmission system using rotation parts, not only smooth torques lead to constant speed, but also the rotating plane of both ends of the transmission is the same. However, this is not the case for ICE, in which the engine generates rough torques, leading to an increase in torsional vibration of the crankshaft.

Torsional vibration can cause various problems such as:

1. Decrease in durability of components;
2. Discomfort induced by vibrations that propagate to places like seats and pedals;
3. Noise problems like gear whine and booming;
4. Synchronization problems causing reduced performance and reduced fuel economy.

When engine torsional vibrations propagate through the transmission, they excite the vehicle's driveline. Structurally sensitive frequencies along the driveline may amplify these phenomena and create NVH related problems, like mentioned in item 3. In figure 2, typical driveline NVH problems are exposed.

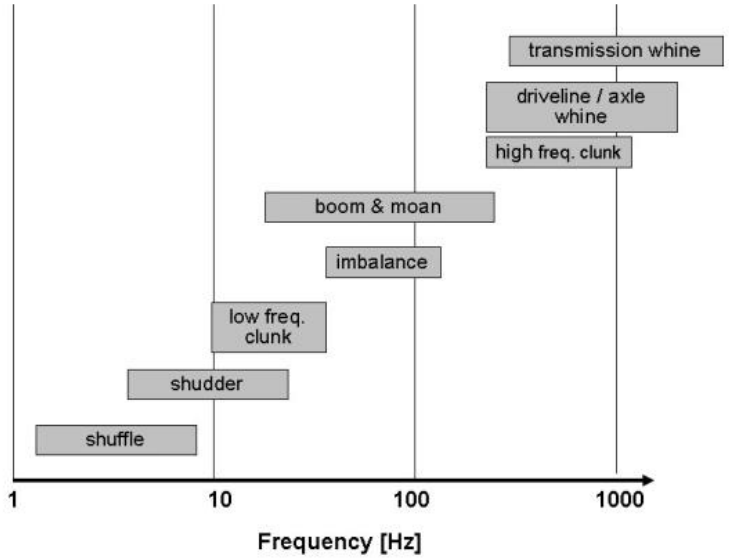


Figure 2, typical driveline NVH problems, as presented by Wellmann et al. in (Wellmann et al. 2007).

According to (Wellmann et al. 2007), driveline torsional vibration at the rear axle cause reaction forces of the axle housing mounts, which can induce noticeable driveline booming noise.

However, the excitation from secondary components of the engine revolution also introduce noise in the cabin through engine mounts (Matsuyama and Maruyama 1998). Thus, several structural paths can cause booming noise through engine excitation, creating resonance effects and a rise of acoustic pressure on the passenger cabin.

It is very important knowing the critical structural paths when trying to reduce a noise problem inside the car. For this analysis, TPA methods are usually applied. A review of TPA methods can be found in the literature (de Ponseele 2012), and some basic notions about this topic is presented next.

Analysis methods for booming noise - Transfer Path Analysis

The noise heard inside the vehicle cavity is the sum of the noises transmitted through multiple paths, from the sources to the receiver. The identification of where a certain noise comes from is not a simple challenge since the overall sound perceived by the passengers depends on the level and nature of the sources and on the way the structural and airborne loads find the target, through the vehicle (de Ponseele 2012). The typical TPA system is consisted of:

- **Sources**, which creates the input loads into the system. Typical vehicle sources include the vibration of the engine, tailpipe noise, intake noise, road inducted vibrations and radiated noise from vibrating panels;
- **Transfer path**, which characterize how noise and vibration goes from the sources to the defined targets;

- **Targets**, which represents the location where it is desired to study noise and vibration. Targets are typically acoustical such as the acoustic pressure perceived by the passengers but can also be vibrational.

Transfer paths can be considered either airborne or structureborne. In case of structureborne paths, the source imparts forces through a mechanical attachment to the structure. In figure 3, it can be seen a scheme of the propagation of structureborne noise from a source to the target. On the other hand, in a case of airborne paths, the source imparts forces acoustically. In figure 4, it can be seen a scheme of the propagation of airborne noise from a source to the target.

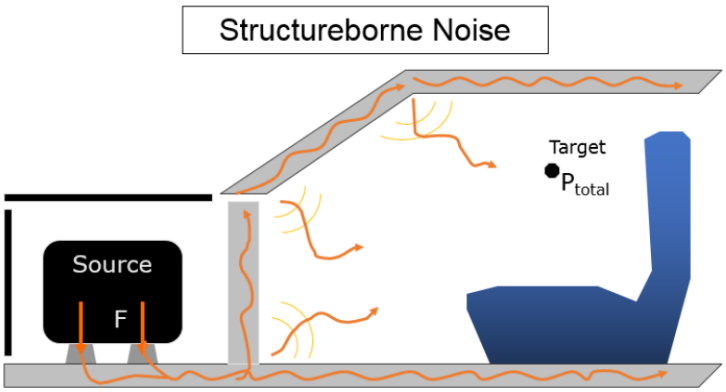


Figure 3, Structureborne noise, courtesy of Siemens Industry Software.

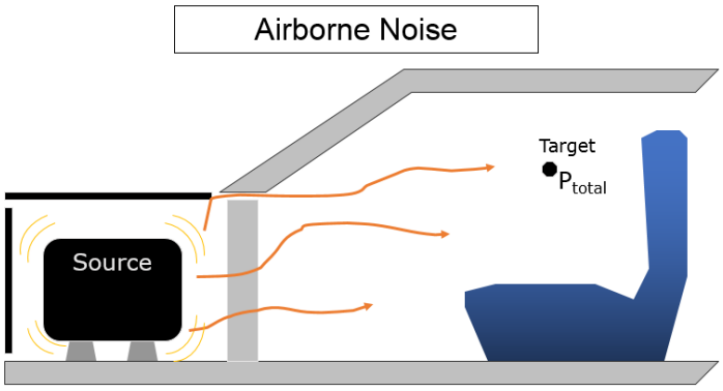


Figure 4, Airborne noise, courtesy of Siemens Industry Software.

The booming noise is considered a structureborne noise. In the TPA analysis, the booming noise SPL (Sound Pressure Level) can be expressed by (2) (Wu et al. 2019):

$$p_k(\omega) = \sum_{i=1}^{n} FRF_{ik}(\omega) \cdot F_i(\omega), \qquad i = 1,2,\dots,n \quad (2)$$

Where:

- $p_k(\omega)$: interior noise at target k;
- $FRF_{ik}(\omega)$: frequency response function of structural paths i;
- $F_i(\omega)$: structural load at path location i;
- $FRF_{ik}(\omega) \cdot F_i(\omega)$: contribution of structural paths i;
- n: number of structural paths.

A car is a complex assembly of many subsystems and so the noise energy is transmitted to the target through multiple structures and spaces. Thus, the frequency response function for each transfer path can be expressed by (3) (Wu et al. 2019):

$$FRF_{ik}(\omega) = \prod_{j=1}^{m} FRF_{ikj}(\omega), \qquad j = 1,2,\dots,m \quad (3)$$

Where $FRF_{ikj}(\omega)$ is the frequency response function of structural paths i in subsystem j and m is the number of subsystems.

In figure 5, a sketch map of vibroacoustic flow in a vehicle is shown to ilustrate the iteration between the multiple subsystems present in a car.

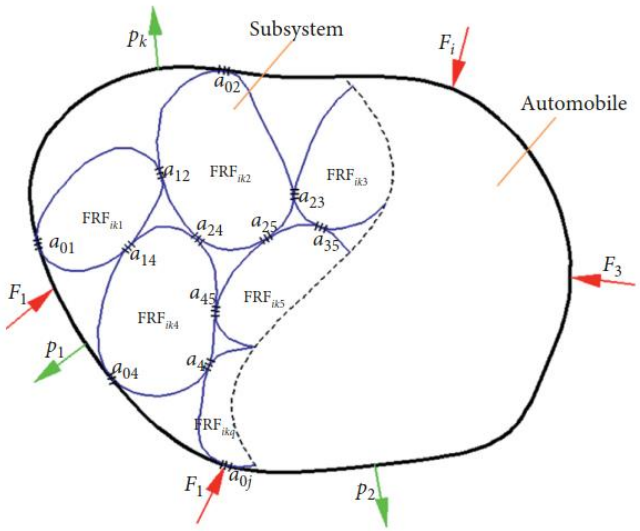


Figure 5, sketch of different subsystems within a car, as presented by Wu et al. in (Wu et al. 2019).

Firstly, in order to analyze the generation mechanism of booming noise, the identification of the transfer paths by which the vibration energy enters the vehicle cavity should be done through Experimental TPA. This well-established approach allows the estimation and ranking of individual noise or vibration contributions. By opting for method, in (Wu et al. 2019) the vibroacoustic energy in the transfer processes was quantitatively analyzed, for the study of booming noise in a minivan. It was concluded that the driveline torsional resonance carries large amounts of vibration energy to the rear suspension and causes a violent pitching vibration of the rear axle. This vibration energy is then transferred to the body and cavity, leading to the experience of booming noise by the passengers.

Analysis methods for booming noise – Spectrogram/waterfall analysis

The spectrogram analysis is a straightforward way to visualize the presence of booming phenomena. It offers a visual representation of the spectrum of frequencies of a signal as it varies with time. There are several ways to generate a spectrogram, but the most common way is by applying the discrete STFT (Short-Time Fourier Transform) to the discrete time signal.

The discrete STFT consists on the application of the FFT (Fast Fourier Transform) on overlapping windowed blocks of the signal. The FFT is an algorithm which performs the DFT (Discrete Fourier Transform) in a computationally efficient manner. The DFT performs a Fourier Transform on a discrete time block (Brunton and Kutz 2017).

The discrete STFT can be calculated through the expression (4):

$$STFT\{x_n\}(m,\omega) \equiv X_k(m) = \sum_{n=0}^{L-1} x_n \cdot w_{n-m} \cdot e^{-\frac{j2\pi kn}{L}} \quad (4)$$

Where:

- x_n: discrete time signal to be transformed
- w_{n-m}: window function, commonly a Hann or Gaussian window
- L: length of the signal

However, the uncertainty principle limits the ability to simultaneously attain high resolution in both time and frequency domains. On one hand, the time series is perfectly resolved in time, but provides no information about the frequency content. One the other hand, if a Fourier transform is applied to the time series, it perfectly resolves frequency content, but provides no information about when in time these frequencies occur (Brunton and Kutz 2017). Different time and frequency resolutions can be obtained by changing the properties of the window function. If a wide window is used, good frequency resolution is achieved at the cost of temporal resolution, while a narrow window has the opposite effect.

Plotting the spectrogram to see the spectral and temporal pattern of time-varying sounds inside of a car's cabin has been usually utilized for the characterization of the booming noise of an accelerating car (Siano and Panza 2017). Typically, on this analysis, the time axis is replaced by the crankshaft's RPM axis to easily see the behavior of the different orders of the engine revolution.

Orders assist in the analysis of noise and vibration of rotating components. Each element of a system contributes to the overall noise and vibration level of that system. Therefore, analyzing the orders assists in the identification of individual rotating components contribution to the overall level of noise and vibration.

In figure 6, it is presented a system composed of two rotating shafts connected by a pulley with ratio 3 to 1.

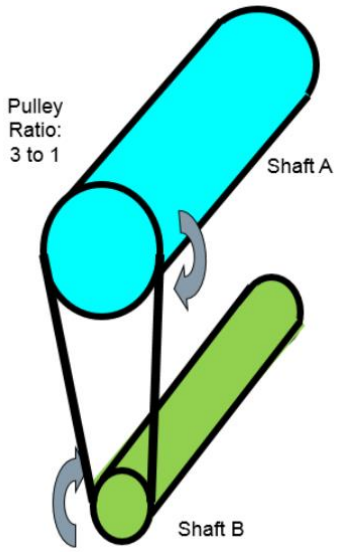


Figure 6, system composed of two shafts and a pulley, courtesy of Siemens Industry Software.

When shaft A rotates at a certain speed with a corresponding frequency, shaft B will rotate at 3 times that speed and frequency, as indicated by the pulley ratio. Then, if shaft A accelerates, e.g. between 600 and 6000 RPM (10 and 100 Hz), shaft B will accelerate between 1800 and 18000 RPM (30 and 300 Hz). When plotting the three-dimensional plot with frequency, RPM and amplitude, a graph like the one on figure 7 can be created, where the orders are shown. In this example, amplitude can be a feature like pressure, acceleration, etc. Thus, the order number is a ratio of events per revolution relative to the first order, and the first order is selected by the RPM range used. If the RPM axis range was between 1800 and 18000 RPM, the first order would refer to the shaft B rotation, and shaft A would be the 1/3 order.

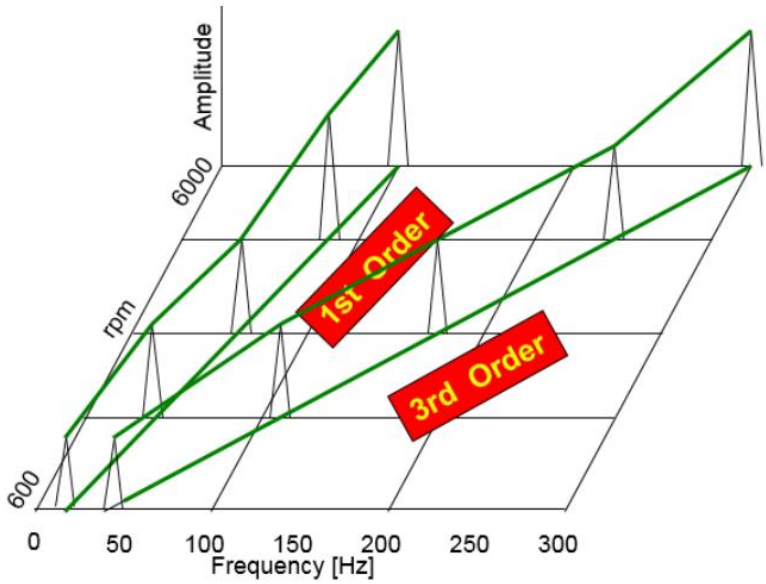


Figure 7, first and third order representation on three dimensional plot, courtesy of Siemens Industry Software.

Taking now the example of a 4-stroke combustion engine. As figure 8 shows, in one cycle, a 4-stroke engine has one combustion event every two revolutions of the crankshaft.

The engine orders are multiples of the rotating frequency of the crankshaft and can be calculated by the equation (5) (Panza 2015):

$$f_0 = k \cdot \frac{N}{60} \ [Hz] \quad (5)$$

Where:

- N: crankshaft rotational speed in RPM
- k: order number
- f_0: frequency of order k

One cycle of a single cylinder of a 4-stroke engine consists of the intake, compression, power and exhaust. Since each stroke up and down corresponds to one revolution of the crankshaft, two revolutions of the crankshaft are needed for all four strokes to occur. Therefore, there is one combustion event for every two rotations of the crankshaft.

Considering a 4-stroke engine with 4 cylinders, it has four combustion events for every two rotation of the crankshaft. This implies that the combustion order is the 2^{nd} engine order.

One way to easily calculate the combustion order of an engine is by using equation (1). The EFR of a 4-stroke and 4-cylinder engine is equal to:

$$EFR = \frac{N}{60 \cdot 2} \cdot 4 = \frac{N}{30} \ [Hz] \quad (6)$$

By equating equations (5) and (6), it can be concluded that the EFR also corresponds to the 2^{nd} engine order.

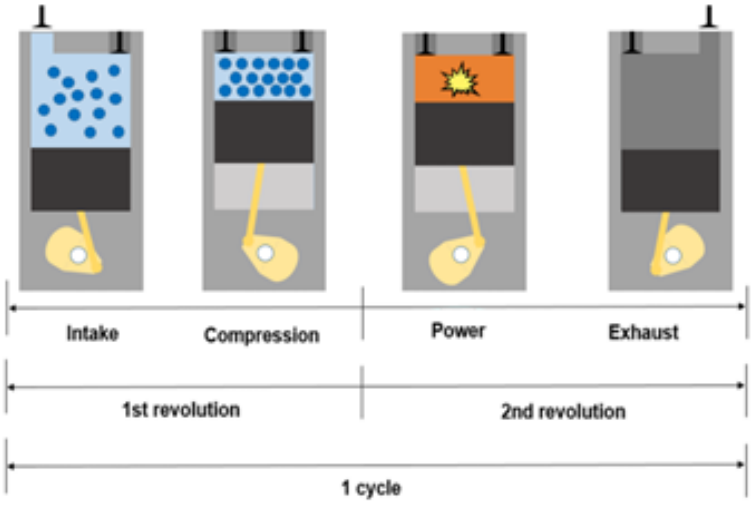


Figure 8, one cycle of a 4-stroke combustion engine, courtesy of Siemens Industry Software.

As mentioned in (Wellmann et al. 2011), there is an increase in torsional excitations from the EFR, which is correlated with booming noise occurring in the passenger's cabin. Thus, analyzing the 2^{nd} engine order is a good approach to assess this phenomenon. In figure 9, it can be seen multiple spectrograms of acoustic pressure levels on different seats from run-up conditions in a diesel passenger's car. These spectrograms were obtained from the raw acoustic pressure levels measured by microphones placed on the left ear position of the passengers. Orders 1,2, 4 and 6 of the engine revolutions are highlighted. As expected, the strongest amplitude peaks correspond to the 2^{nd} engine order, and this is where the booming noise is most likely to occur. It is also worth to note that, in this study (Siano and Panza 2017), there are no significant changes on the spectrums between the driver and the front passenger, as well as both rear passengers. However, as stated by Siano (Siano and Panza 2017), it is difficult to find a direct correlation between the results of the order analysis and the booming phenomenon, as the overall sound pressure level cannot sufficiently express the perceptual feeling.

In order to better understand the level of sound annoyance caused in the car's cabin, some psychoacoustics-based indices can be rather used.

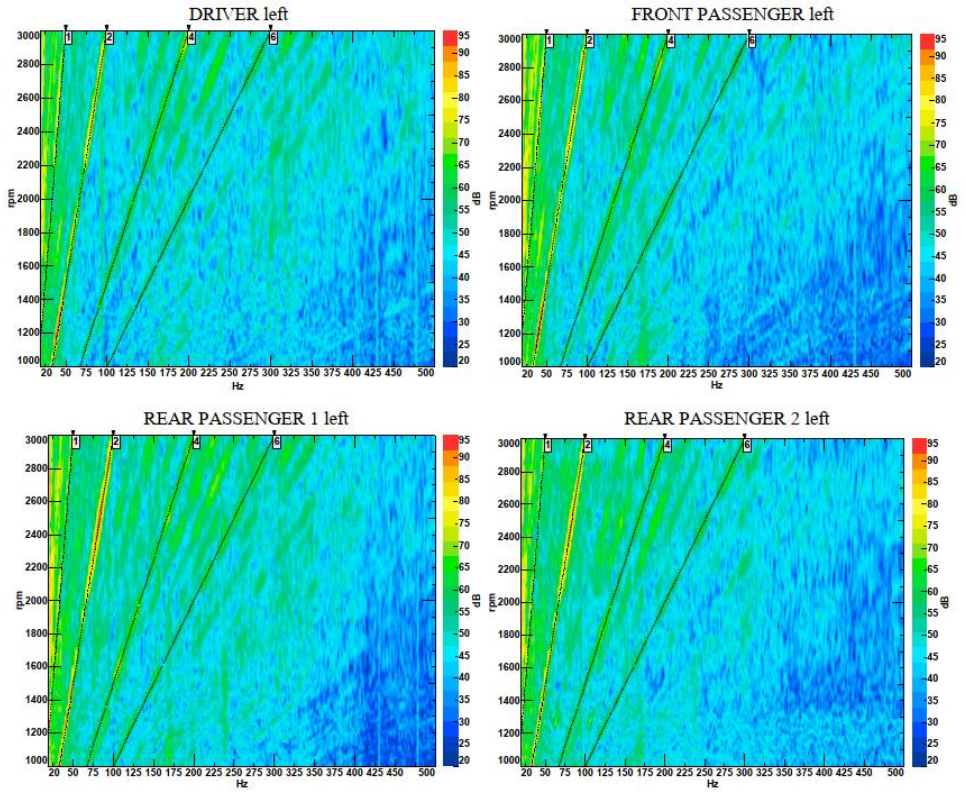


Figure 9, spectrograms in engine run-up condition for front and rear internal microphones, as presented by D. Sianoa and M. A. Panza in (Siano and Panza 2017).

Analysis methods for booming noise – Sound quality analysis

NVH engineers often make use of psycho-acoustic metrics such as loudness, sharpness and roughness for obtaining a desirable frequency balance and sound level in the interior of a vehicle. This

approach is known for accurately describe the passengers' pleasantness or annoyance sensations (Wang et al. 2014).

Regarding the booming noise, loudness is considered the most important metric that accurately describes the sensation (Shin et al. 2009). Loudness represents the auditory perception character related to SPL, frequency content and duration of a sound, and is measured in sones. Several models can be used to calculate loudness, but for non-steady noises like those acquired in vehicle dynamic conditions, the Time-Varying Zwicker Loudness is particularly useful (Siano and Panza 2017; Mitchell 1997).

Many papers such as (Park and Lee 2012), (Hatano and Hashimoto 1996), (Hatano 1999) and (Hatano and Hashimoto 2000) tried to develop a metric which quantifies the booming sensation by using psychoacoustic metrics, having obtained good correlation between their results and subjective jury studies. More information about these papers is found on chapter 2.4.

2.2 Machine Learning

ML is a tool that makes machines mimic the learning process of humans in a way that allows them to transform information into knowledge and is regarded as one of the most influential and powerful technologies in today's world. It is capable of perform a task without the need of receiving explicit instructions to do so, when provided with large datasets. In a certain way, it is a process which programs evolution in an algorithm, building models that give accurate predictions or find patterns in previously unseen data. The widely accepted formal definition of ML as stated by Tom M. Mitchell (Mitchell 1997) is:

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."

The dataset used on a ML problem is usually split into two subsets: training and test datasets. A predictive model or classifier is trained using the training data, and then the test data determine the model's predictive accuracy (Priddy and Keller 2005). In order to reach high accuracies, these algorithms are based on statistics and mathematical optimization which is a process consisting in finding the minima or maxima of a function. One of the most popular optimization algorithms used in ML is called gradient descent.

One way to classify a ML algorithm is by looking at the way it learns. In this aspect, there are three main learning algorithm classes: supervised learning, unsupervised learning and reinforcement learning. The class used during this dissertation is supervised learning. Next, it will be presented the differences between supervised and unsupervised learning.

Supervised learning

In this approach, the algorithm is assisted during the training, meaning the desired response is given for each input. As shown in figure 10, the learning system is exposed to the environment with a set of input features and then predicts a response. The "teacher" knows the desired outputs of the set of input

features, and by comparing that desired output with the response of the learning system, the learning systems updates itself (Priddy and Keller 2005). The goal is to develop a model capable of, in most cases, correctly predict the desired output of a set of data which the model has never seen before.

These models can perform classification and regression tasks. In the former, the data labels are divided into a finite number of classes. For example, the evaluation of a car's acoustic spectrum in the end of a production line can be classified into being acceptable or not acceptable. In the latter, the prediction is on a continuous interval, and its value may differ from the one on the training set labels. For example, the prediction of a car's price based on a variety of characteristics such as mileage, model, engine size, interior style, etc.

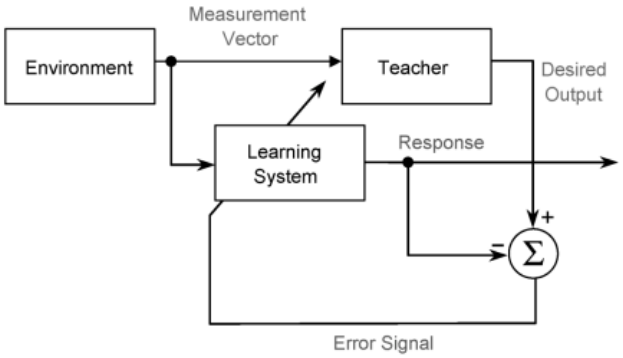


Figure 10, block diagram of supervised-learning model, as presented by Priddy and Keller in (Priddy and Keller 2005).

Unsupervised learning

This class of learning algorithms is similar to the supervised one, as can be seen on figure 11. However, there is not the presence of a "teacher", which knows the desired outputs of the training data, and there is also an extra block named adaptation rule. Based upon the adaptation rule and the system response, the parameters of the learning system are adjusted to obtain the desired performance (Priddy and Keller 2005). This way the algorithm can find a way to structure the data without the need of labeled input data. This process is typically done based on clustering methods, where training examples with alike properties are grouped into the same cluster.

Examples of unsupervised learning algorithms include k-means, hierarchical clustering, principal component analysis and self-organizing maps.

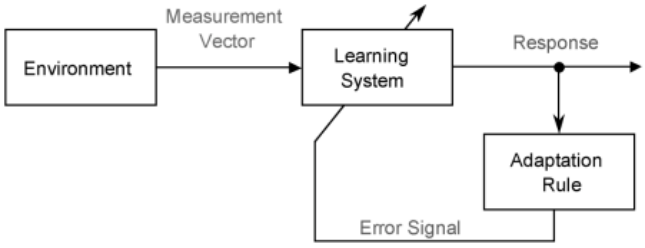


Figure 11, block diagram of unsupervised-learning model, as presented by Priddy and Keller in (Priddy and Keller 2005).

2.3 Neural Networks

NNs are a subset of ML, inspired by the structure of the human brain. In order to achieve this, it uses a multi-layered structed of algorithms which allows an enhancement in terms of power to represent more complex functions. The application of NN architectures like CNNs and Recurrent Neural Networks (RNN) have been applied in a variety of fields such as computer vision, defect detection, audio and speech recognition and bioinformatics. In figure 12, the panorama of CNNS relatively to NN, ML and Artificial Intelligence (AI) is shown.

During this dissertation, 2D-CNNs are applied to classify acoustic spectrum represented on an image. Thus, an introductory explanation of these networks is given next. For the sake of brevity, 2D-CNNs will be referred as CNNs.

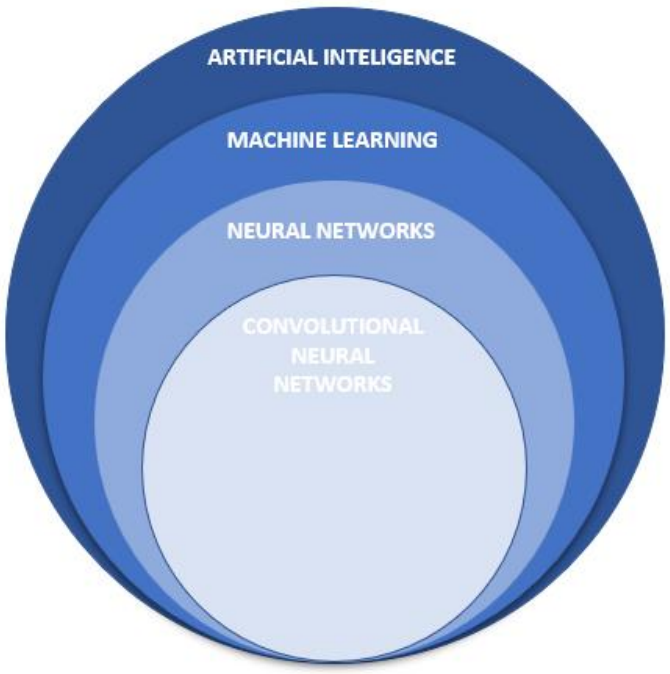


Figure 12, panorama of CNNs relatively to NN, ML and AI.

Convolutional Neural Networks

CNNs are specialized artificial NN for processing data that has a grid-like topology, like time-series data and image data. They were inspired by the organization of animal visual cortex and their name comes from the employment of the mathematical operation of convolution, which is a specialized kind of linear operation (Goodfellow, Bengio, and Courville 2015). These networks have been a dominant method in computer vision tasks since the astounding results presented at the ImageNet Large Scale Visual Recognition Competition (ILSVRC) in 2012 (Russakovsky et al. 2015). However, CNNs can also successfully process 1D data arrays like signals and sequences, including language, and 3D data such as video or volumetric images (LeCun et al. 2015).

The typical architecture of a CNN is composed of three types of layers:

- Convolution layers;
- Pooling layers;
- Fully connected layers;

The first two types of layers, convolution and pooling, perform feature extraction, while the fully connected layer maps the extracted features into final output. In a convolutional layer, a kernel is convolutionally multiplied across the layer's input generating a feature map. After successfully training a CNN, the feature maps of the network will contain characteristics of the inputs which describes their features, leading to a better classification or regression accuracy.

In figure 13, it is possible to observe the process of calculation of a feature map's unit through element-wise product between the input and the kernel, with posterior sum of the obtained matrix values. The kernel will scan all the input according to the stride number defined. This number determines how many "jumps" the kernel does in the input for each value of the output.

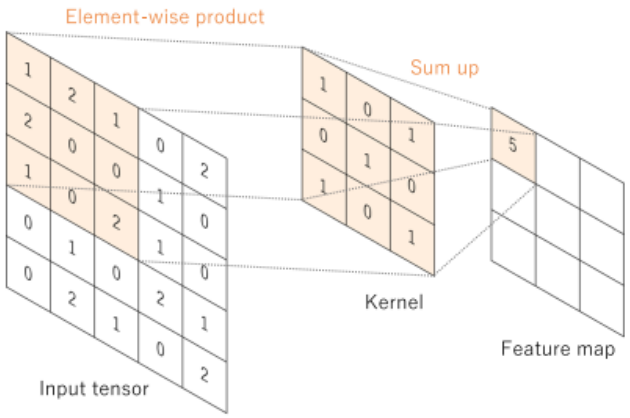


Figure 13, convolution operation for calculation of a unit in the feature map, as presented by Yamashita et al. in (Yamashita et al. 2018).

Another important parameter which must be taken in consideration is the presence or not of padding. Padding is an additional layer that can be added to a border of an image. If no padding is used, which can also be called valid padding, the feature map is going to have a size smaller than the input. That is the case

of the example given on figure 13. With this type of padding the values on the middle of the input will get covered more than once, in opposition to the values on the corners which will only get covered by the kernel one time. This results in loss of information and in the shrinking of the feature map. However, if padding is used, which can also be called same padding, the feature map is going to have the same size as the input. In order to achieve this, an additional layer of zeros is added to the borders of the input.

Next, the outputs of the convolution operation are usually multiplied by a non-linearity. The most used functions for this are the sigmoid function and the Rectified Linear Unit (ReLU), and they can be seen in figure 14. These functions allow the introduction of non-linearity into the network, improving its performance.

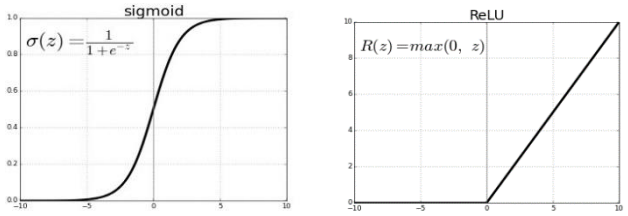


Figure 14, Sigmoid and ReLU functions.

The pooling layer allows the merging of semantically similar features into one, calculating the maximum or the average of all values within the pooling kernel. This results in a reduction of the dimension of the feature maps. Contrarily to the convolution operation, in the pooling layers there are no learnable parameters. The kernel size, stride and padding are the only hyperparameters in pooling operations, similar to convolution operations (Yamashita et al. 2018).

Finally, the output feature maps obtained in the last set of convolution plus pooling will serve as input to a set of fully connected layers which will allow a connection to the output, being it a classification or regression. Each fully connected layer is followed by a nonlinear function, such as ReLU (Yamashita et al. 2018). A sketch of an example of the described typical architecture can be seen on figure 15. Its is important to note that the connections and the architectures can vary.

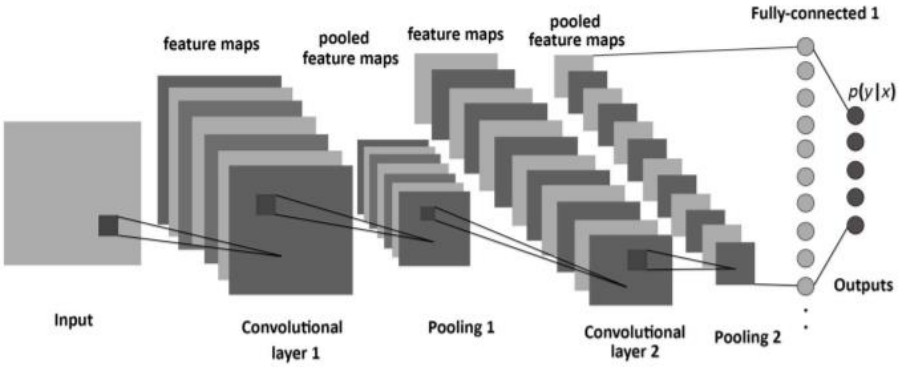


Figure 15, The structure of a CNN, consisting of convolutional, pooling, and fully connected layers, as presented by Saleh Albelwi and Ausif Mahamood in (Saleh Albelwi and Ausif Mahmood 2017).

After having the network architecture, the next step consists in training it. This is a process of finding the kernels in convolution layers and the weights in fully connected layers. These parameters must minimize the difference between the output predictions and the input labels. To achieve this the backpropagation and gradient descent algorithm is commonly used, where the learnable parameters are updated according to the value of the loss function. This function measures the compatibility between the output predictions and the input labels, and depending on the type of problem, the loss function used is different. While on regression problems the mean squared error is typically used, on classification problems it is commonly used the cross-entropy function. The main goal is to minimize the loss, and to achieve this it is usual to look at the gradient of the loss function and take the opposite direction in which the function has the steepest rate of increase with an step size determined by the hyperparameter named learning rate (Yamashita et al. 2018). In figure 16, a two-dimensional visualization of how the gradient descent algorithm works for a single weight is shown. The loss function is represented in the y-axis by the name of cost.

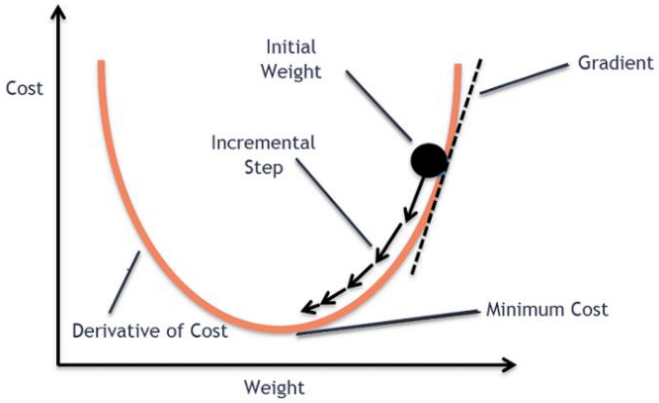


Figure 16, gradient descent implementation for a single weight.

One common practice to increase computation efficiency and to guarantee a more stable convergence towards the global minimum is to use mini-batch gradient descent. This method splits the training dataset into small batches that are used to calculate model error and update model coefficients (Ruder 2017). In addition, there are several other algorithms which can improve gradient descent method, such as Stochastic Gradient Descent with momentum, RMSProp and Adam (Ruder 2017; Qian 1999; Kingma and Ba 2017).

Since the number of weights in the fully connected layers can reach high values, biasing issues are likely to occur. In order to surpass this problem, new regularization techniques such as dropout (Srivastava et al. 2014) can be used. In this approach, a certain percentage of neurons is selected to be cancelled during a forward or backward pass. This removes co-dependency amongst neurons during training which curbs the individual power of each neuron leading to over-fitting of training data.

Moreover, there are other techniques to improve the performance of the CNNs. One of them is batch normalization (Ioffe and Szegedy, n.d.), and it enables a faster and more stable training of the networks. This operation standardizes the inputs to a layer for each mini batch, causing a stabilization of the learning process and considerably reduce the number of training epochs required. One epoch is when the entire dataset is passed forward and backward only once through the neural network.

2.4 State of the art concerning booming and NN

Nowadays the automotive industry is a very competitive business, and meeting costumer's expectations of comfort is imperative for the success of a vehicle on the market. Booming noise is one of the various noises which can be heard inside the cabin of a diesel car, and it is perceived as an annoying acoustic feature. Therefore, a lot of research regarding the attenuation and classification of booming noise has been made.

Several methods have been developed over time to identify and troubleshoot the causes of booming noise (Jha 1976). On a more recent study, Grupta attempted to understand the booming noise by analyzing structural and acoustic modes, achieving the reduction of noise levels as desired (Gupta, Gautam, and Jain 2014). Moreover, on (John Britto et al. 2016), a counter measure for booming was proposed after a root-cause analysis using modal analysis, operational deflection shapes, input point inertance and noise transfer functions. As a result, a methodology for booming noise analysis and countermeasure was framed.

Active and passive control strategies can also be applied for reducing the booming noise. However, passive control methods lead to an increase in cost and weight of the vehicle and result in poor fuel efficiency, so active control methods are predominant in the literature. On (Hasegawa et al. 1992) it was developed the world's first active noise system for production vehicle implementation. Following that, several methods were also proposed for active control of booming noise. Examples of that are presented on (Inoue et al. 2004), (Lee and Nasiri 2007), (Oh, Kim, and Park 2002) and (Kang et al. 2016).

As stated by (Shin et al. 2009) and (Siano and Panza 2017), a passenger in a vehicle is often more exposed to the booming phenomenon related to powertrain excitations than any other causes of booming. This noise occurs when pure or narrow band tones related to the engine firing frequency and its harmonics induce resonance effects on the passenger cabin, causing a prominent rise of the acoustic pressure.

On (Wellmann et al. 2007), the authors pointed the driveline as a significance source of NVH in vehicles, which includes booming, and showed the degree of complexity involved in these issues. They used targeted test, Computer Aided Engineering and hybrid methodologies to solve those challenging problems. Later, part of that team published another paper (Wellmann et al. 2011) focusing on the booming noise caused by the driveline, concentrating on the influence of torsional vibration transmitted via the rear axle as structureborne noise, developing a time-domain TPA process to evaluate that.

As stated by (Park and Lee 2012), almost all the research papers are more interested in the control of A-weighted sound level with a commercial sound level meter rather than the objective evaluation of booming sound quality using subjective parameters. Examples of these parameters are loudness, sharpness, roughness and fluctuation and they are often referred as psychoacoustic metrics. On (Hatano and Hashimoto 1996), a new objective measure called booming level was proposed to quantify the booming sensation of the passengers, based on the loudness of the sound. This model was then modified in (Hatano 1999) by the same authors, in order to achieve a higher correlation with the booming sensation. Later, on (Hatano and Hashimoto 2000), the same authors revised their model again and created the booming index, using the partial and overall loudness in the calculations. Despite this, there are still problems with these models as stated by (Shin et al. 2009).

Another sound quality index for booming noise was proposed in (Park and Lee 2012). Here, the psychoacoustic metrics loudness and sharpness were used as input parameters of an artificial neural network, having obtained a good correlation between the network's output and the averaged subjective rate. In this study, synthesized booming noises were used, where the booming phenomenon was recreated by extracting the 2^{nd} engine order and re-synthesize it using a bell-shaped curve to simulate the booming sensation. The remaining 2^{nd} engine order amplitude content is a constant value.

It is important to note that all the studies pointed until now that deal with the development of a metric to quantify the booming sensation, were done in Asian countries. It was revealed in (Toi, T, Isoyama, H, Ogu, Y, Saito, H, Hoshino, H, Ishikawa, M, et al. 2004) that the booming noise affects more harshly people from Asian countries comparing to western countries in general.

There are more applications of NN on NVH analysis. One example of that is presented on (Stender et al. 2020), where CNNs were applied to detect brake squeal noise on spectrogram images. Very high accuracies were reached by using this approach, which shows the potential for this method in the detection of NVH issues.

CNNs have been a significant tool when it comes to computer vision tasks, across various fields. The recent growth in interest in these networks led to the development of better network architectures. Examples of this include dropout (Srivastava et al. 2014) and batch normalization (Ioffe and Szegedy, n.d.).

2.5 Motivation

Taking into account the powertrain booming noise issue previously mentioned, it is important to point out the necessity of having a reliable booming detection strategy to perform end-of-line noise control in vehicles.

This dissertation was developed as a second stage of a research and investigation project. In the first stage, a prototype was developed to simulate the booming noise on the 2^{nd} engine order and subjective studies were done to assess the classification metrics of the booming noise. In the second stage, further developments were made to the prototype, making it capable of generating a diverse database of booming scenarios, under different conditions, in a fully automated way. Further contributions to the project were made by applying CNNs to classify the synthesized sounds' spectrograms. This will serve as tool to the NVH engineer in end-of-line production of cars, to help assess the existence of booming noise on the produced vehicle, without the need of extensive jury tests.Problem outline

2.6 Problem Outline

This dissertation has the main objective of applying NN models to classify powertrain booming noise on ICE passenger cars. As these models require large databases, and due to the lack of sufficient experimental data, synthesized vehicle sounds were generated from a small set of car sound recordings.

To synthesize these sounds there are two possible approaches. However, due to the nature of the experimental data, a top-down approach has taken to create the Sound Quality Equivalent (SQE) models, which will be then used to synthesize full vehicle sounds. In order to create these models by using the exposed approach, the car sound recordings need to be first decomposed into orders and broadband noise.

Training the NN models requires a variety of booming scenarios on the training databases. To accomplish this, the SQE models were modified to include this aspect and a diverse database was then created.

The problem which arose after the database creation was to find the best NN strategy to implement. While the best approach is to test different models and compare them, time limitations imposed on the conclusion of this dissertation forced the choice of a potential best solution. The different solutions are exposed on the next chapter, as well as the reasons for choosing a particular solution rather than the others.

2.7 Alternatives evaluation

Among the different kind of NN algorithms, the ones selected as having the most potential to reach better results for the problem in question were:

- RNNs;
- 1D-CNNs;
- 2D-CNNs.

RNNs are a powerful and robust tool mainly used for sequential data or time series. Algorithms like the Apple's Siri and Google's voice search use this type of networks, attaining great results. RNNs operate with an "internal memory", remembering important things about the input they received, and in a way adding the immediate past to the present.

Another strategy to evaluate one-dimensional sequences of data is by using 1D-CNNs. These models extract features from the sequential data and map them. Their structure and functioning are very similar to the 2D-CNN, although on 1D-CNNs the kernel only moves in one direction and has one dimension. An extensive analysis of 2D-CNNs is presented on chapter 2.3.

As mentioned in chapter 2.1, an easy way to visualize the booming noise is by looking at the spectrogram images of the full vehicle sounds, using RPM in the Y-axis. Furthermore, spectrogram images have been previously used in the training of 2D-CNNs for the classification of other NVH defects, having obtained good results (Stender et al. 2020). Due to this, a more conservative approach was taken to solve the problem of this dissertation, so the application of 2D-CNNs was chosen.

3 VEHICLE SOUND DECOMPOSITION

In this chapter, the sound decomposition strategy will be exposed. This step is necessary for the creation of digital twins of the measured cars, which will be used for the sound synthesis. Digital twins are digital representations of the cars' sounds.

Before decomposing the sounds, the experimental setup used for obtaining the sounds will be exposed, as well as the analysis of the data to evaluate which orders to extract in the sound decomposition. More details about all this will be presented next.

3.1 Experimental setup

In this dissertation, previously acquired operational measurements from different vehicles were used. This data was obtained as a result of multiple test campaigns done at different times. Several cars were used in these campaigns, although because of data limitations on some cars, the ones used on this dissertation are:

- Ford Focus
- Ford Mondeo
- Opel Vectra

All these vehicles have a 4 stroke ICE with 4 cylinders, which means their EFR is the 2^{nd} engine order, as calculated on chapter 2.1.

The typical goals of these test campaigns include acquiring experimental data for:

- Generation of SQE models of the vehicles;
- Studying of coherence-based methods for the separation of structureborne and airborne noises contribution.

The tests were done on proving grounds, allowing the test of the vehicles on optimal road conditions and uniform asphalt surface. For example, in the Ford Focus case, tests were carried out in the Aldenhoven Test Center, managed by the University of Aachen, in Germany. In figure 17, it can be seen a sketch of this test center. In table 1, it is shown more details about the track.

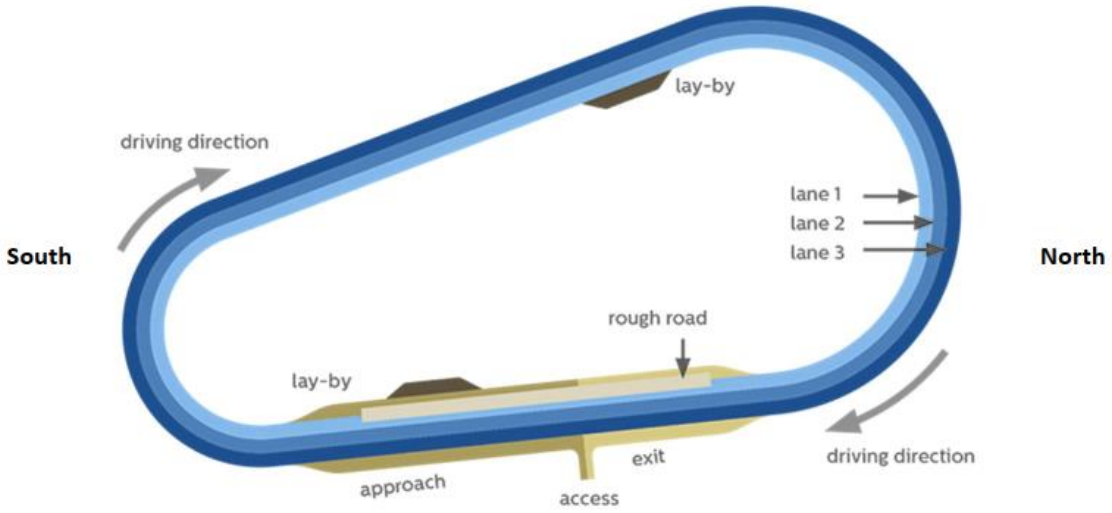


Figure 17, proving ground in Aldenhoven Test Center.

Table 1 – Test track details.

Surface	Asphalt
Length of straight [m]	2 x 400
Length lane 1 / 2 / 3 [m]	2074 / 2097 / 2120
Width lane 1 / 2 / 3 [m]	3.75 / 3.75 / 4
Radius north / south curve [m]	186.5 / 113.5
Maximum free cornering speed in lane 3 north / south curve [km/m]	117 / 96

The typical test matrix includes:

- Runups in every gear;
- Coast down;
- Constant speed;
- Rundown;
- Free-drive recordings.

However, the only data available for this dissertation corresponds to the runup condition on gear 2 with 100% throttle, for every car mentioned. Amongst the many different sensors used, the ones relevant for this dissertation are:

- Microphone recordings in the car cabin with a binaural headset in the driver's seat;
- Tachometer for measuring the rotational speed of the motor.

3.2 Data analysis

After obtaining the data from the different cars, the next step consists in analyzing it to assess the most relevant orders for vehicle sound decomposition. This process was accomplished by using a software develop by Siemens Industry Software called Simcenter Testlab. This program provides a full suite of integrated testing, analytics and modeling tools covering a wide range of test needs.

Taking the example of the Ford Focus, the data acquired by the tachometer can be seen on figure 18. On figure 19 and 20, it is presented the variation of amplitude in time on the microphone placed at the left and right ears of the driver, respectively. Comparing these curves, no major differences can be observed, so only the signals for the left ear microphone will be analyzed. This was expected since on (Siano and Panza 2017) no significant changes were observed on the spectrograms between the driver and the front passenger, as previously shown on chapter 2.1. The same conclusion can be made for the Ford Mondeo and Opel Vectra.

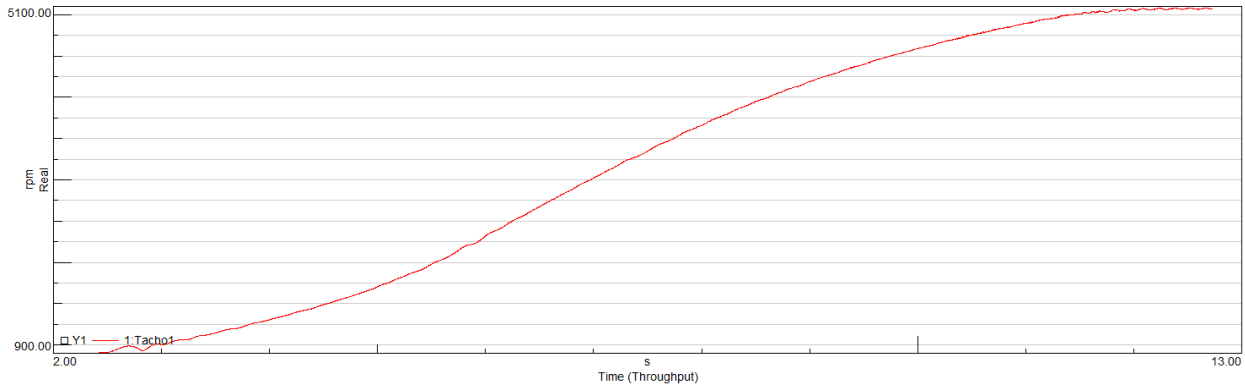


Figure 18, time vs. RPM of Ford Focus.

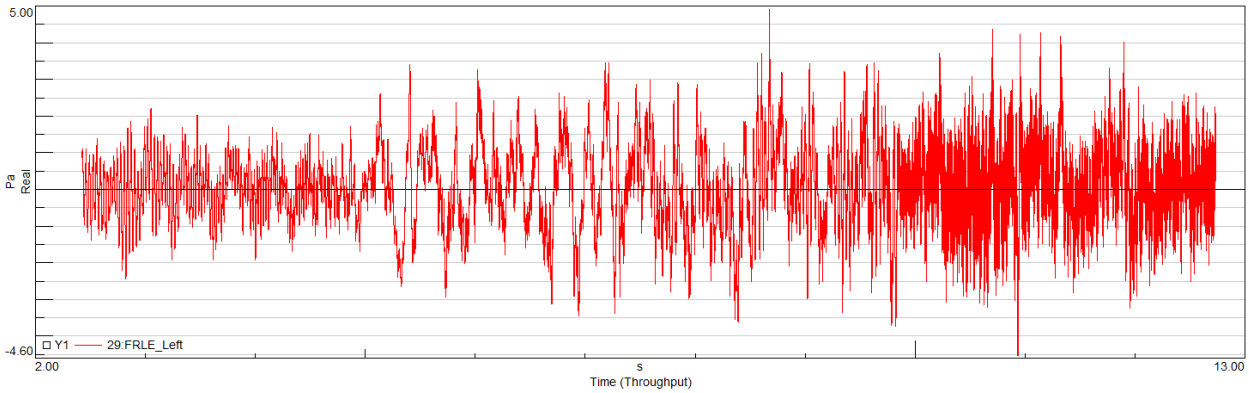


Figure 19, time vs. amplitude of Ford Focus (left side).

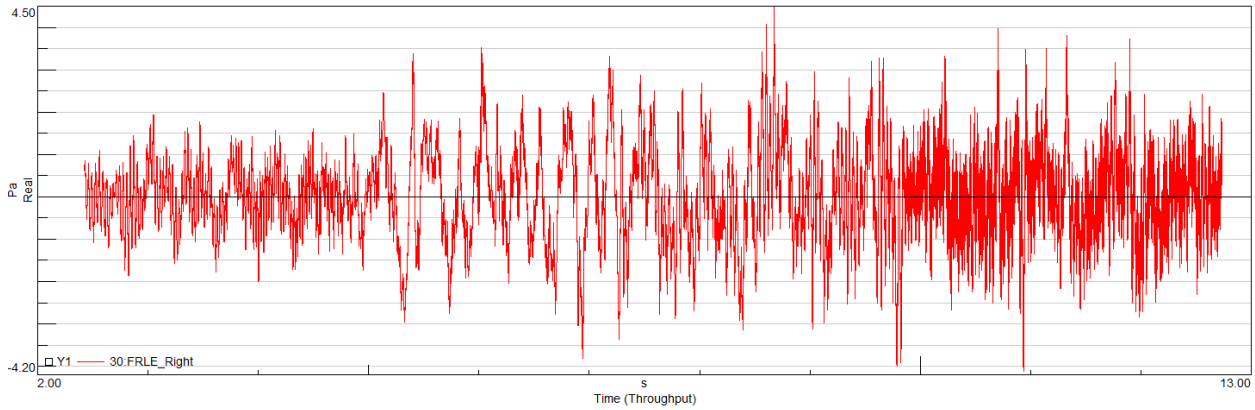


Figure 20, time vs. amplitude of Ford Focus (right side).

Having the experimental data provided by the tachometer and the left ear microphone, it is now possible to plot the spectrograms of all vehicles. These plots are presented on figures 21, 22 and 23, corresponding to Ford Focus, Ford Mondeo and Opel Vectra, respectively. Most of the high amplitude content is located at low frequencies, and the amplitude peaks are distributed on:

- Engine orders represented by straight diagonal lines. It is possible to easily observe the different orders on the spectrograms by moving the black diagonal lines;
- Frequency blocks below 50 Hz. This corresponds to broadband noise, and it is outside the scope of this dissertation.

It is worth to note that the dominant order with the strongest peaks corresponds to the 2^{nd} engine order for all vehicles, especially at low RPM regimes. Moreover, the most significant peaks occur in the Ford Mondeo and Opel Vectra, between 1350 and 2000 RPM, approximately. Therefore, from an initial analysis, this is where it is most likely to observe the presence of booming phenomena in these measurements.

To help this study, several kinds of filters were applied to the sounds, making it easier to isolate certain parts of the spectrograms and listen to the resulting sounds. Firstly, a high-pass filter was implemented to eliminate the low frequency background noise with high amplitudes. Then, order-stop and order-pass filters were implemented to evaluate the influence of the engine orders. From the subjective point of view of the author of this dissertation, there is indeed presence of booming noise on the low RPM range caused by the 2^{nd} engine order in the Ford Mondeo and Opel Vectra. However, further studies must be done to confirm this assumption.

One of the main goals of this data analysis on Simcenter Testlab is to choose the most relevant orders for vehicle sound decomposition, which was accomplished by looking at the spectrograms and select those orders.

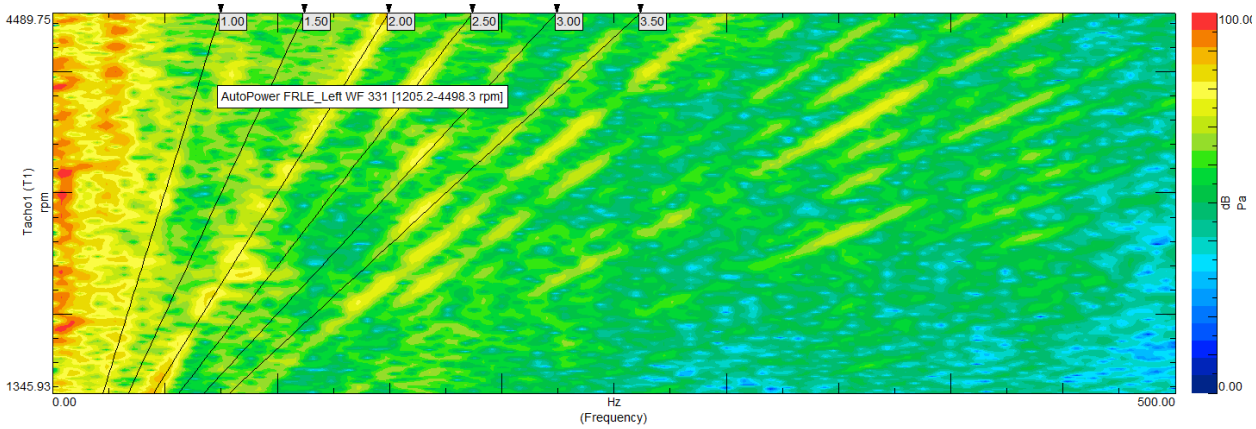


Figure 21, spectrogram of Ford Focus for driver's left ear microphone.

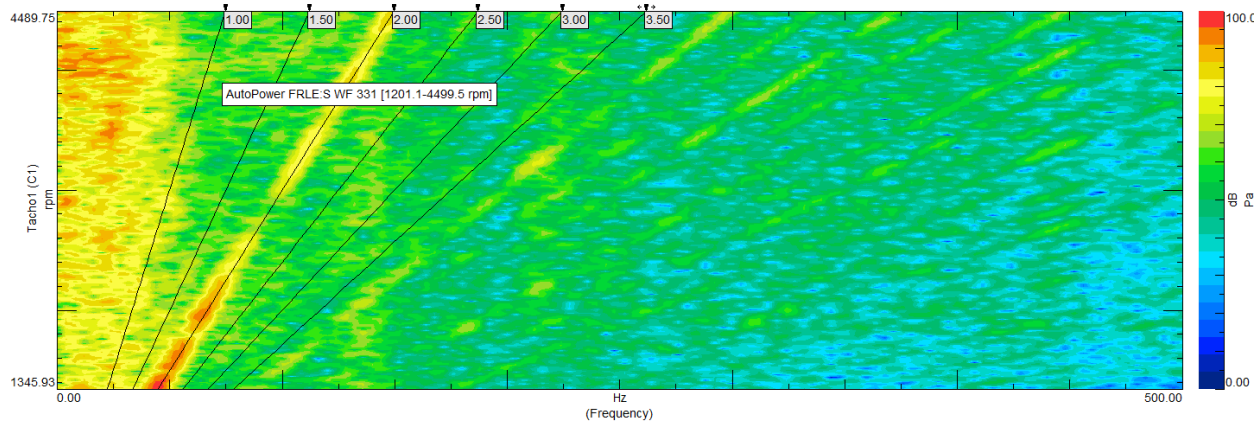


Figure 22, spectrogram of Opel Vectra for driver's left ear microphone.

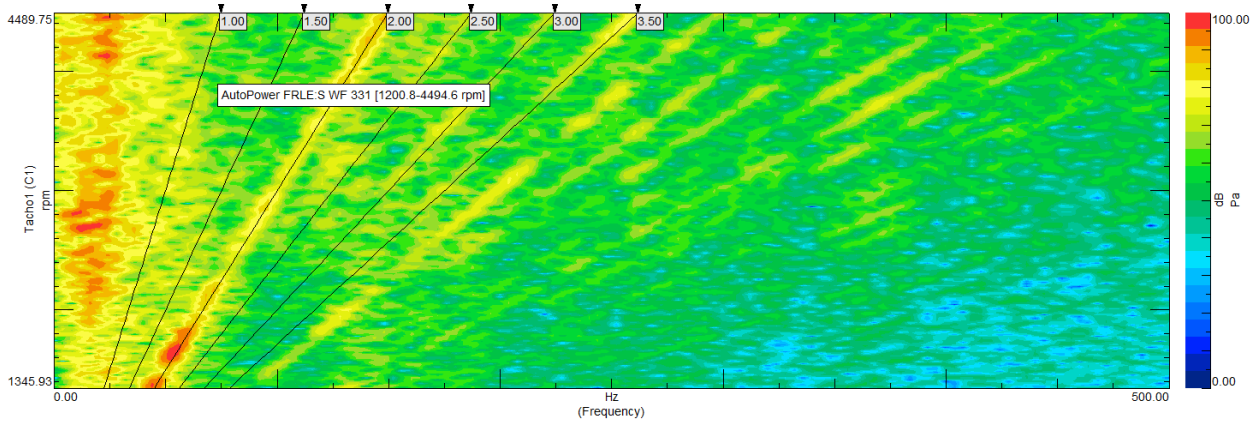


Figure 23, spectrogram of Ford Mondeo for driver's left ear microphone.

3.3 Sound Decomposition

The sound decomposition was done by using a software developed by Siemens Industry Software called Simcenter Vehicle Sound Simulator. This program can be used for several applications, such as:

- Drive vehicles virtual prototypes and frontload NVH studies to the early stages of the design;
- Define a vehicle library for benchmarking and sound target definition;
- Evaluate new components for existing vehicles and compare alternative configurations;
- Evaluate interior and exterior sound design and noise enhancement.

In this dissertation, the program was used in order to create vehicles libraries by decomposing the cars' sounds into orders and broadband noise. For this, an order tracking algorithm based on the Time Variant Discrete Fourier Transform (TVDFT) (Blough, Brown, and Vold 1997) was applied, extracting the orders from the vehicles sounds, and thus separating tonal and broadband noise components.

After the extraction of the selected orders, the results can be exported to Simcenter Testlab. Figure 24 and 25 shows the extracted orders of Ford Mondeo and Opel Vectra, where it was considered the possibility of the occurrence of booming noise at the low RPM range. The high amplitude red curve corresponds to the order 2. On figure 26 and 27, the broadband noise on Ford Mondeo and Opel Vectra are shown.

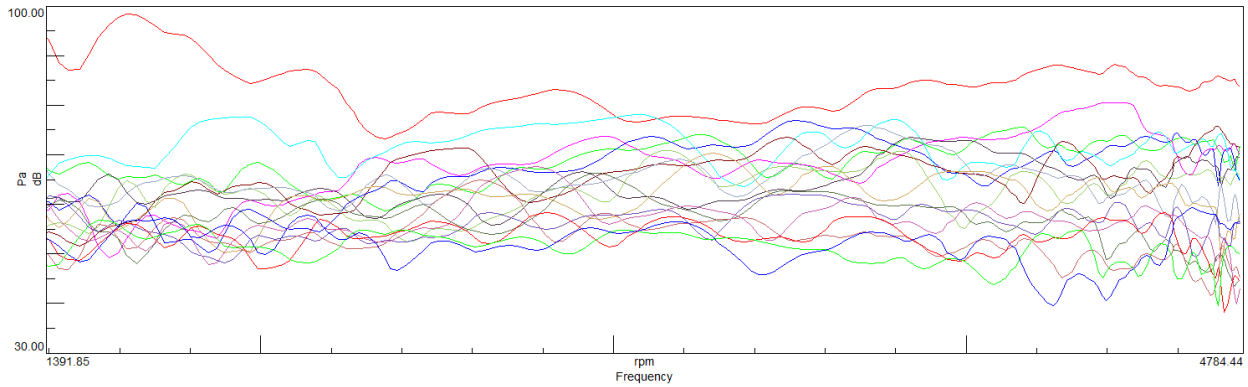


Figure 24, extracted orders of Ford Mondeo.

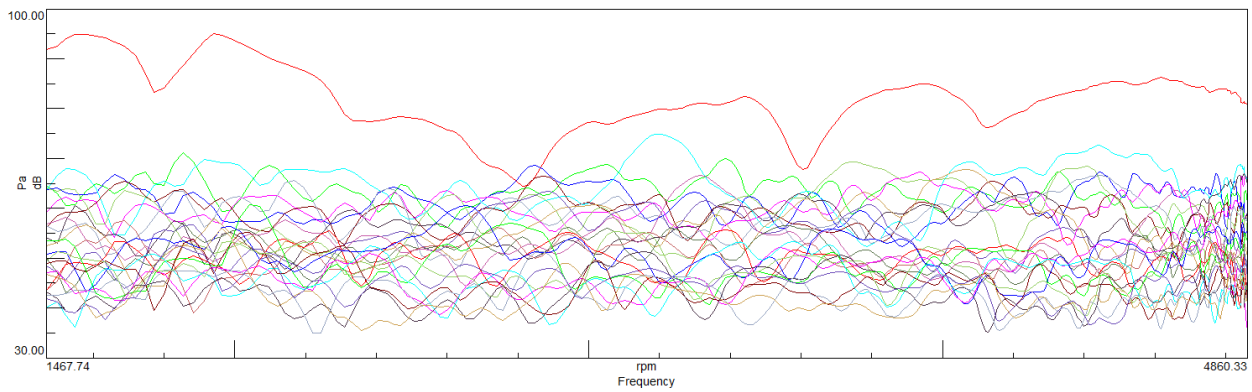


Figure 25, extracted orders of Opel Vectra.

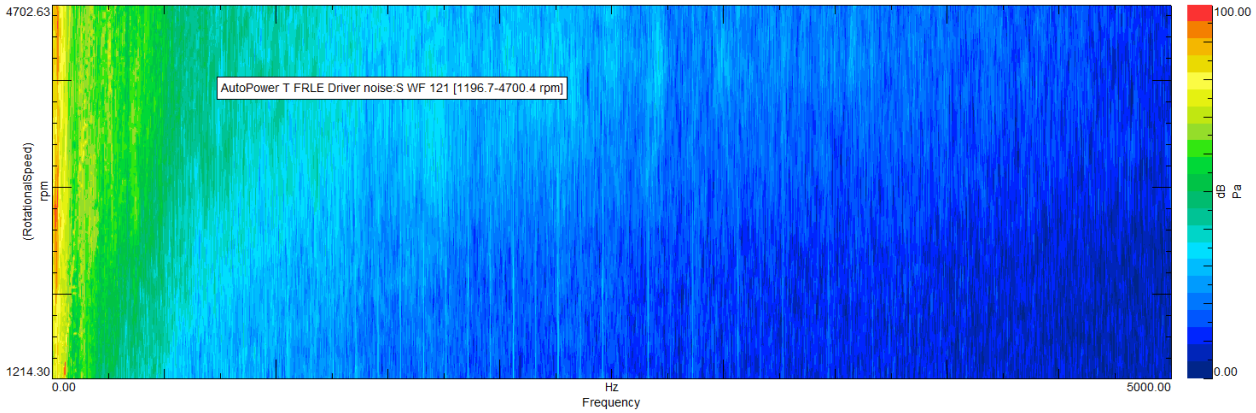


Figure 26, broadband noise in Ford Mondeo.

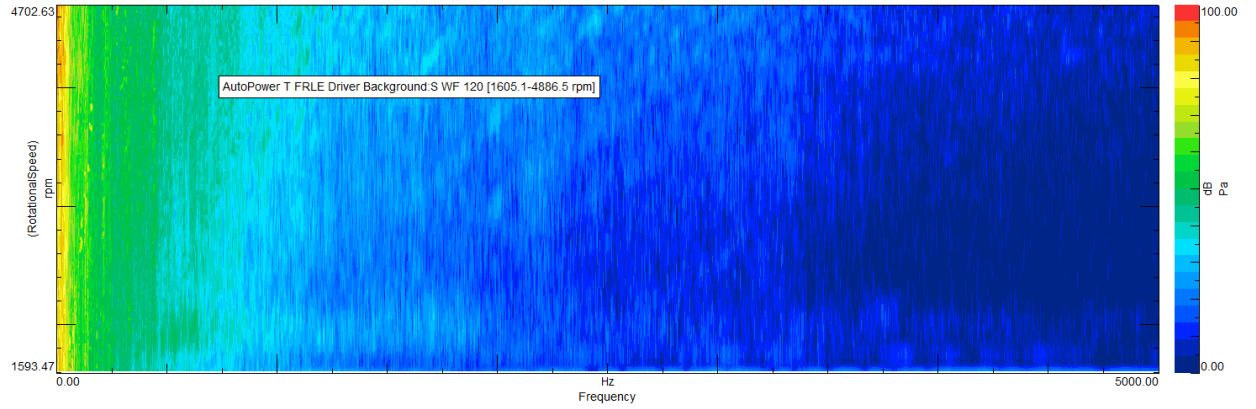


Figure 27, broadband noise in Opel Vectra.

It is relevant to note that in the RPM ranges in which the presence of booming noise is likely to occur, there is a sudden increase on order 2 pressure amplitude, dominating the entire vehicle sound. In these regions, the order 2 curve assumes a bell-shaped curve.

4 VEHICLE SOUND SYNTHESIS

Having the vehicle sounds of runup conditions on gear 2 and 100% throttle decomposed into orders and broadband noise, the next step consists in using the decomposed sounds for the creation of SQE models, with consequent vehicle sound synthesis for the generation of a large database to train ML models. Since the goal of the networks is to classify booming noise, a variety of booming phenomena must appear on the training sets. In order to accomplish this, booming noise has to be modeled during the sound synthesis. This was achieved by the development of a program in MATLAB 2019 which created SQE models and synthesize vehicle sounds from those models, in a fully automated way, having the possibility of inducing booming noise on the main engine orders. This program is also able to plot the spectrograms of the created sounds, which will be inputted on the CNNs.

4.1 SQE model creation

A Sound Quality Equivalent (SQE) model uses a limited set of experimental recordings and/or numerical simulations in order to synthesize a realistic vehicle sound in operational conditions, forming a digital twin of the car. For that, the sound data is organized in numerical blocks containing the vehicle sounds for several driving conditions. This is exemplified in figure 28, where each block corresponds to a certain condition.

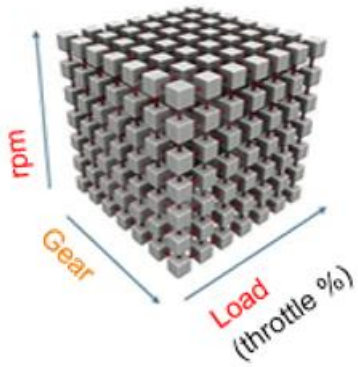


Figure 28, blocks organization on SQE model, courtesy of Siemens Industry Software.

There are two main approaches for the creation of SQE models:

1. Top-down: It is assumed that the vehicle sound is composed by tonal contributors called engine orders and the remaining contributors reveal a broadband behavior in amplitude and a random nature in phase. This approach is based on the application of algorithms for the separation of tonal and broadband noise components from sound recordings of full vehicles

tested in operational conditions. This methodology can be seen on figure 29. The components obtained can then be modified to include, for example, booming noise on the synthesized vehicle sounds;

2. Bottom-up: This approach is based on techniques for combining spectral information about tonal and broadband noise contributions extracted for components tested in benches or numerically simulated.

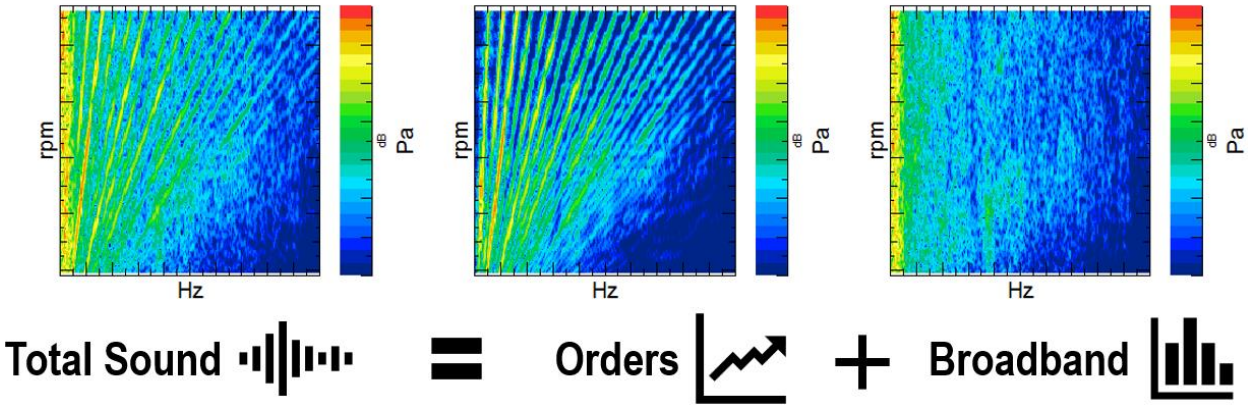


Figure 29, Total sound decomposed into orders and broadband noise, courtesy of Siemens Industry Software.

As already stated in chapter 4.3, in this dissertation the top-down approach has been adopted. One of the main benefits of this approach is that a single measured sound can be decomposed into its component parts without the need for measuring road, wind and powertrain noise separately. Moreover, this method allows to use sound modification tools to vary the contributions of each source and finally to re-synthesize the whole vehicle sound from the separate contributions for any driving condition.

It is also important to point out that the SQE models created for the different cars can only synthesize sounds from gear 2 with 100% throttle, with an RPM range approximately between 1300 and 4600 RPM. This is the case because the measured data obtained corresponds to those operational conditions.

4.2 Mission profile definition

After having the SQE model created for a certain car, the following step consists in the definition of the mission profile. Different mission profiles correspond to different runup conditions, and the program generates them randomly from RPM and time condition anchors given as input. In table 2, it is shown an example for the RPM and time conditions inputted in the program.

Table 2 – Example of mission profile anchors.

	Time [s]	**RPM**
Start	0	1050
Finnish	6.5	4570

Using this as input conditions, random velocity profiles can be generated, creating different runup conditions. This will help making a diverse database because when synthesizing the vehicle sounds, the orders and broadband noises in the SQE models will follow these profiles, always generating different vehicle sounds for the imposed conditions. Moreover, using different profiles simulates real conditions, where the runups performed by the driver are always slightly different.

On figure 30 are represented examples of different mission profiles, using the profile anchors in table 2. It is important to note that the Y-axis represents the RPM and the X-axis represents the time.

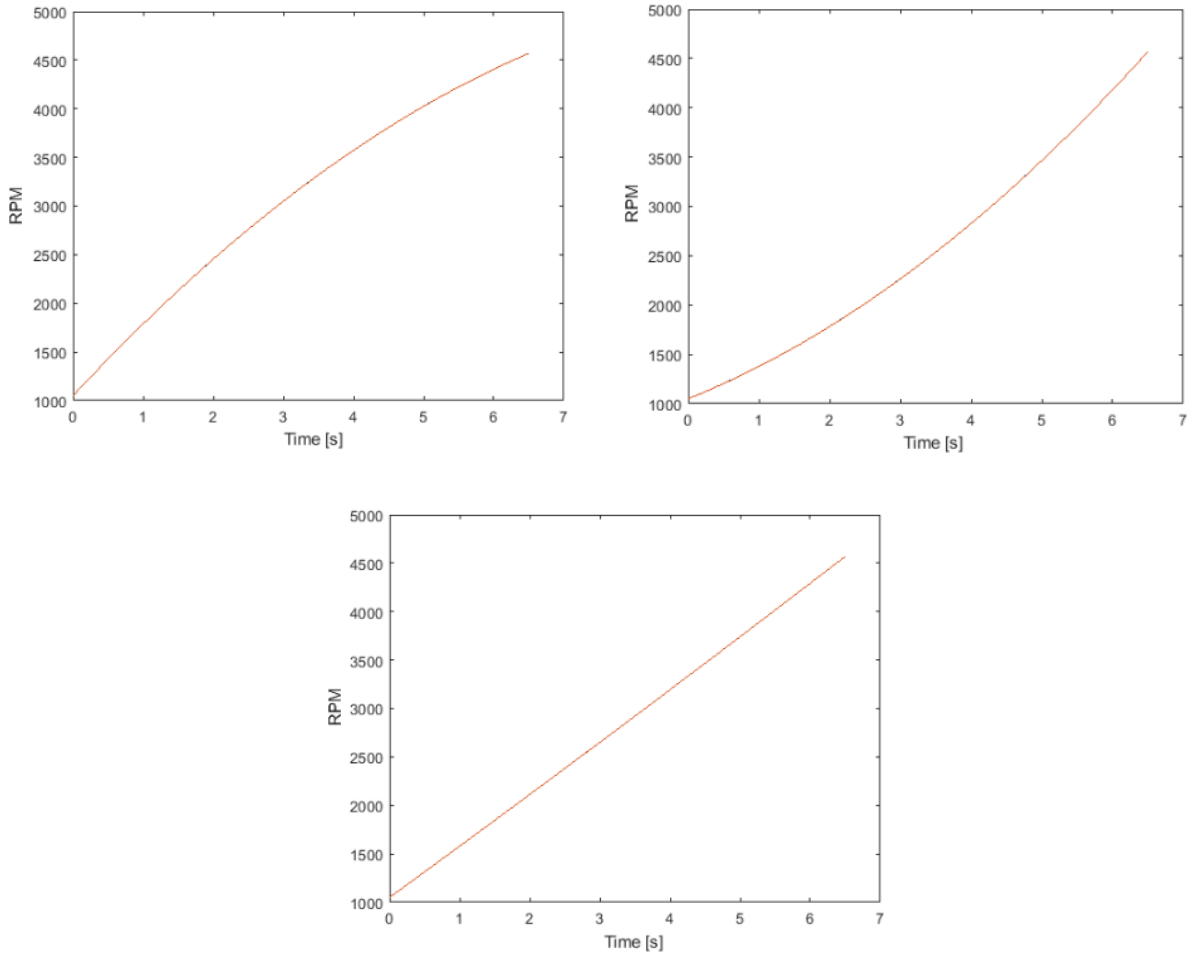


Figure 30, examples of generated mission profiles with profile anchors of table 2.

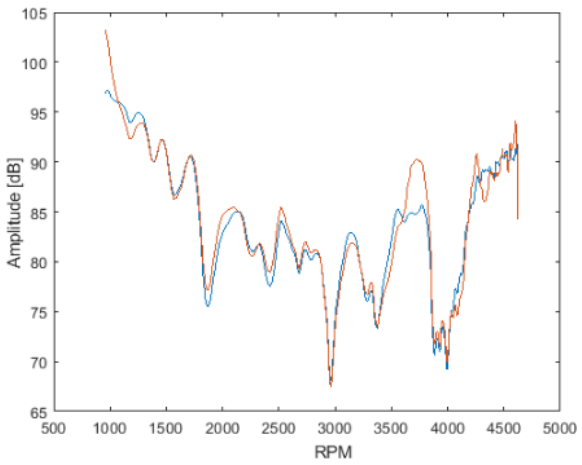
4.3 SQE model manipulation

Now that the SQE models and the mission profiles were created, it is necessary to manipulate the SQE models to include the possibility of generating booming noise on the synthesized vehicle sounds. To achieve this, modifications on the engine orders were made to simulate the booming noise and to create a diverse database. These modifications will be exposed next.

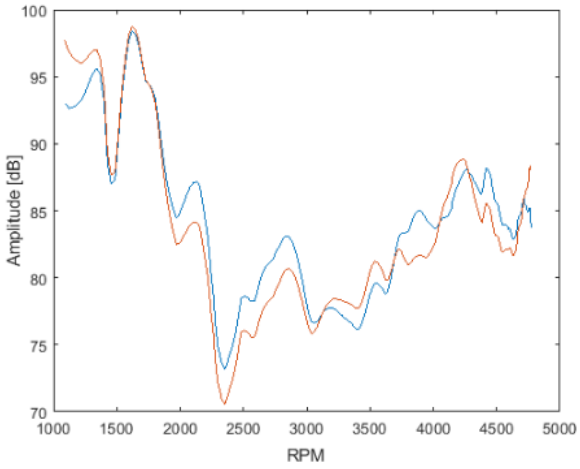
Low frequency random fluctuations

Having in mind the creation of a diverse database, low frequency random fluctuations can be induced on the extracted orders. This also simulates the different values of orders' amplitude that can be obtained on different runups. In figure 31, it is presented different cases of the evolution of the amplitude of order 2 in time. While the blue curve represents the original order, the orange curve stands for the modified order.

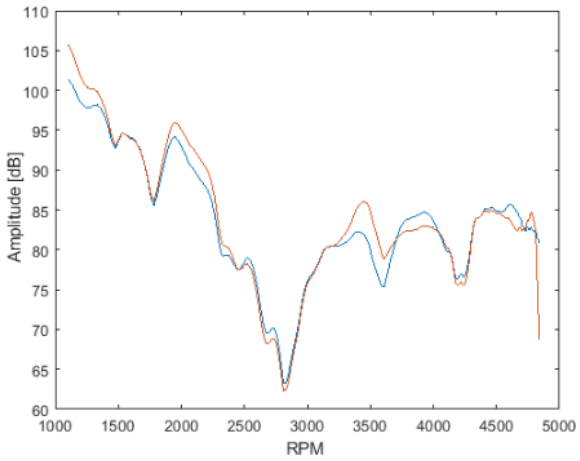
Although some spikes can be noticed in the beginning and/or end of the RPM axis, they can be ignored since the studied RPM range is from 1500 to 4230 RPM, as will be seen later. It is also worth to note that depending on the random fluctuation applied, the modified order can be very similar to the original one, as seen on a) and c), or substantially different, as seen on b).



a) Ford Focus



b) Ford Mondeo



c) Opel Vectra

Figure 31, examples of modified order 2 with low frequency random fluctuations.

Hann window

To model the rise in acoustic pressure on order 2, which can be perceived as booming noise by the passengers when it dominates the total sound pressure, it was introduced a Hann window in the order's curve. This function is defined by having an amplitude of 1 unit, as well as length of N + 1, where N is the number of samples in the X-axis. This is illustrated on figure 32.

The width of the Hann window is defined by dRPM which is a randomly generated number, being then multiplied by a randomly generated gain called G. The range of values of these metrics is shown on table 3. Furthermore, this window will be applied to the RPM range between 1500 and 4230 RPM, which is the studied RPM range. 1500 RPM was arbitrarily defined as the minimum value because below that, the high amplitude content on the order 2 can be associated with the powerfulness of the vehicle, and it is not perceived as a bad sound quality. 4230 RPM was chosen as the maximum value due to synthesis limitations.

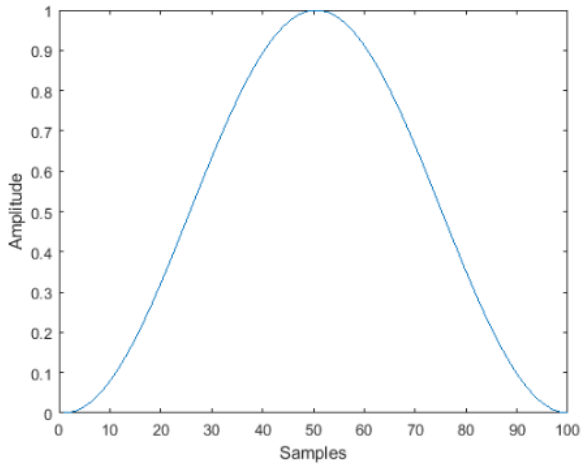
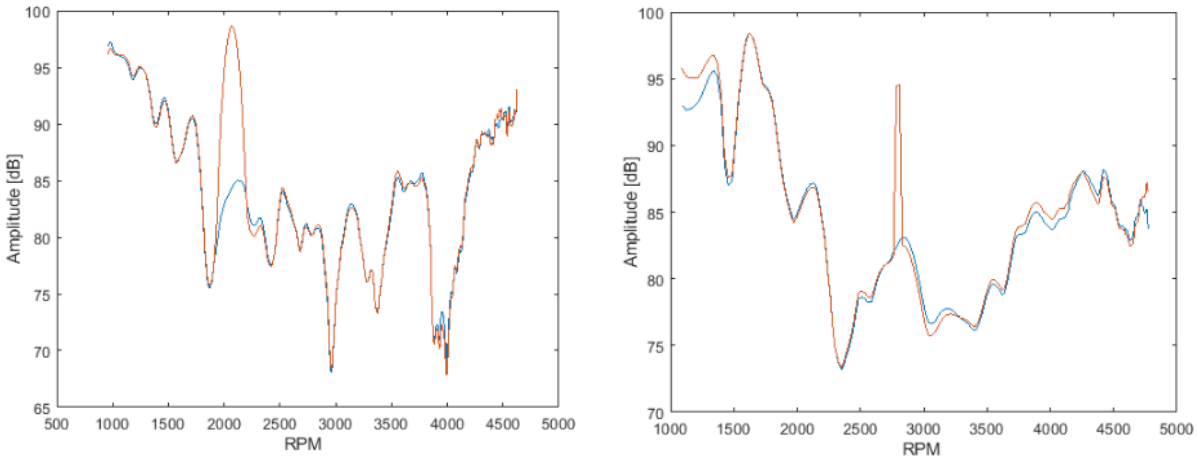


Figure 32, Hann window with 100 samples.

Table 3 – Range of G and dRPM values.

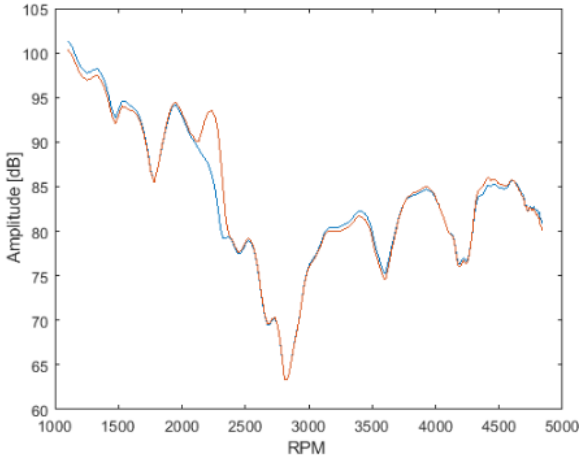
	G	**dRPM**
Minimum value	0	0
Maximum value	1.5	560

In figure 33, several examples of the order 2 manipulation with Hann windows are presented. On a), the Hann window was introduced around 1800 RPM, creating a sudden increase in the pressure amplitude of order 2, which will dominate the total vehicle sound, and will probably induce booming noise. The situation on b) is a bit different because while G has a high value, the dRPM is very low. Most likely, this does not correspond to booming. A phenomenon which recreates a similar curve is, for example, a road bump. On c), a low amplitude window does not change the order profile very significatively.



a) Ford Focus.

b) Ford Mondeo.



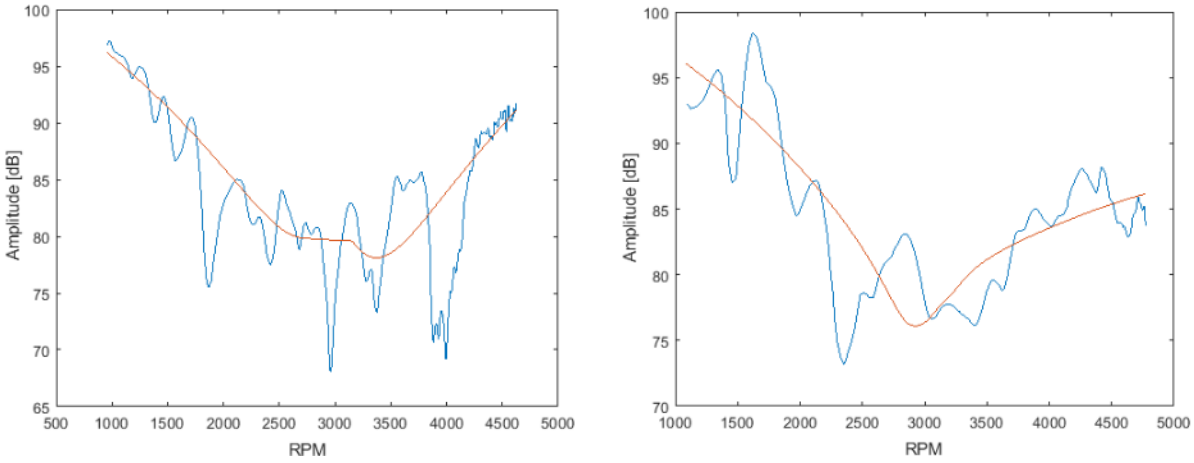
c) Opel Vectra.

Figure 33, examples of order 2 curves after the manipulation to incluce the Hann window.

Smoothing order

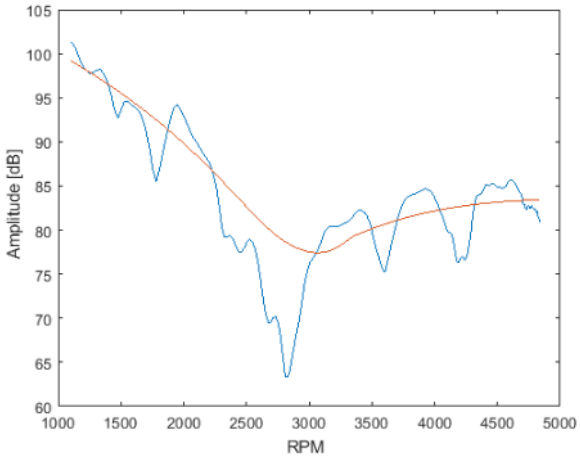
The program has the option to apply a function called smoothing order. This function is capable of smoothing the orders curves, creating more uniform curves. In figure 34 a), it is shown in orange the curve which results from the application of the smoothing function on order 2 of the Ford Focus, while the curves for the Ford Mondeo and Opel Vectra are shown on b) and c), respectively.

By combining the smoothing order function with random low frequency fluctuations, the plots on figure 35 can be obtained. While on a) and b) the smoothed order is significatively changed by the random fluctuations, on c) the order curve is more similar to the one on figure 34 c).



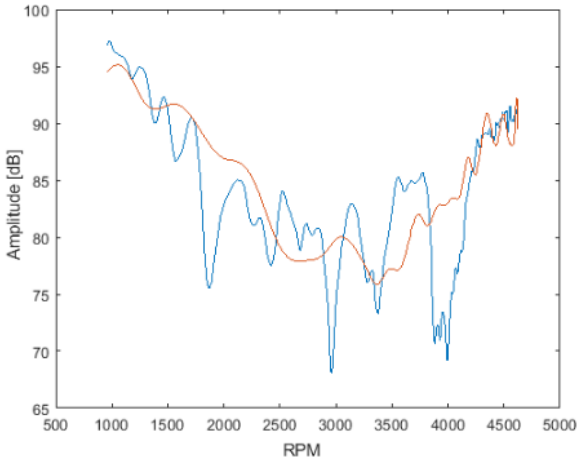
a) Ford Focus.

b) Ford Mondeo.

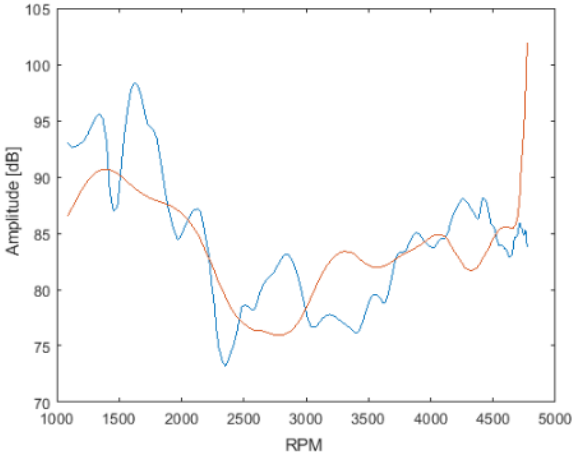


c) Opel Vectra.

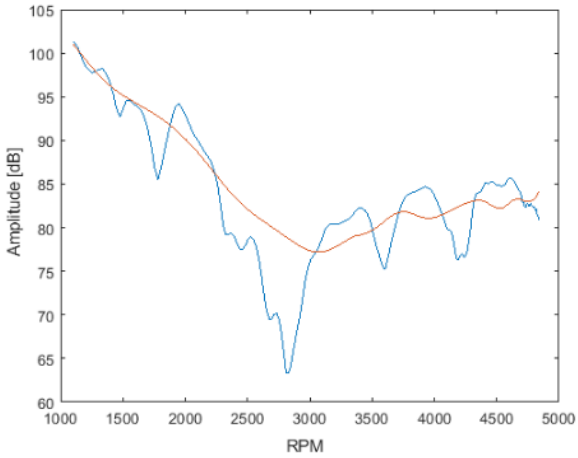
Figure 34, order 2 smoothed curves without random low frequency fluctuations.



a) Ford Focus.



b) Ford Mondeo



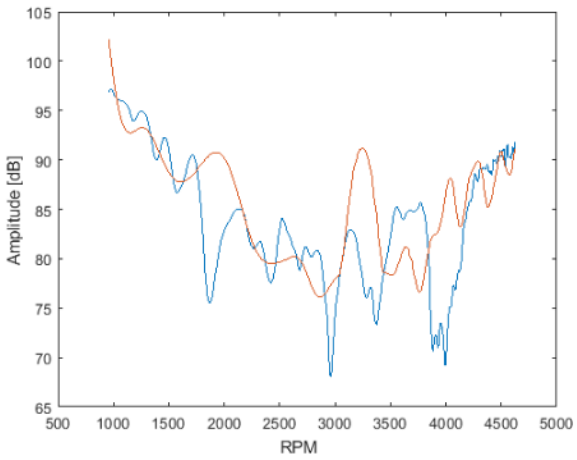
c) Opel Vectra

Figure 35, smoothed order 2 with random low frequency fluctuations.

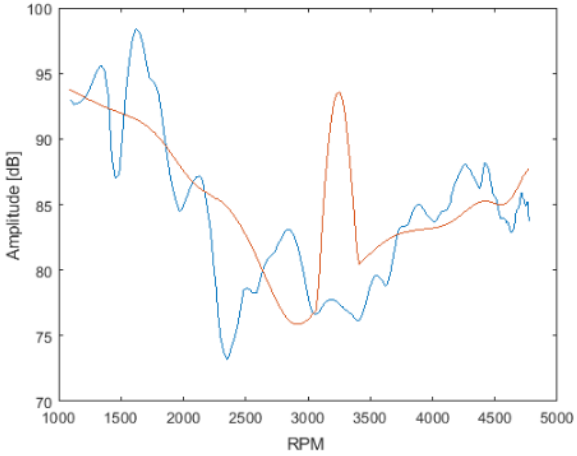
Analyzing these graphs, it can be concluded that smoothing an order with random fluctuations will cause a more homogeneous acoustic pressure profile, as it eliminates the low amplitude dips as well as the majority of the bell-shaped natural phenomena which resembles the Hann window that will be induced. This results in the elimination of the likely occurring booming noise in the low RPM range of the Ford Mondeo and Opel Vectra, as discussed on chapter 4.2. Cases like the one seen on figure 31 a) are also removed, where the random fluctuations accentuate the bell-shaped curve and possibly cause the booming sensation.

Figure 36 shows the results of introducing a Hann window on the smoothed order 2 with random fluctuations. When taking this approach, the induced booming is always prominent, going above the average sound level. This will possibly be advantageous when training the CNNs, making them look for the booming sound at medium and high amplitudes.

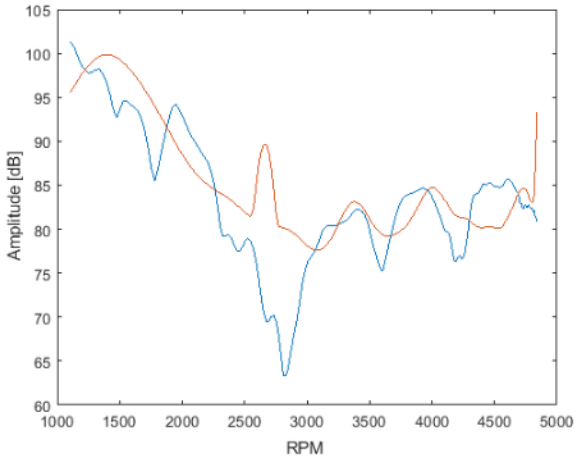
Now that the SQE model has been manipulated to include all the possibilities exposed until now, it is necessary to define a metric to classify the booming noise. As already mentioned, the work on this dissertation is the continuation of a project. On the first stage of this project, subjective studies were done to classify the booming noise. By using the same approach regarding the application of a smoothing function and a Hann window defined by G and dRPM, it was concluded that a window with G higher than 0.5 and dRPM higher than 280 RPM, simultaneously, was perceived as booming by the subjects. These will be the metrics used in this dissertation.



a) Ford Focus.



b) Ford Mondeo.



c) Opel Vectra.

Figure 36, smoothed order 2 with low frequency random fluctuations after the manipulation to include the Hann window.

4.4 Sound Synthesis

For the sound synthesis, the manipulated SQE model from the orders and the SQE model of the broadband noise are used for synthesizing the full vehicle sound under the conditions imposed by the mission profile. First, a synthesis frequency must be chosen considering it has to be higher than two times the value of the highest frequency present on the signal which will be synthesized, according to the Nyquist theorem. Since higher frequency components can be observed on the broadband noise, compared to the orders, the maximum frequency value must be assessed. By doing this study, it was chosen a synthesis frequency of 44100 Hz.

It is worth to note that when synthesizing the broadband noise, a "Narrow Band" synthesis approach (Sarrazin, Colangeli, and Janssens 2013) was taken. While the details of this method are outside the scope of this dissertation, it can be said that it induces random fluctuations on the broadband noise.

On figure 37, an example of synthesized orders and broadband noises can be seen. The blue curve corresponds to the orders, while the orange curve corresponds to the broadband noise. Combining the two sounds, the plot on figure 38 is obtained.

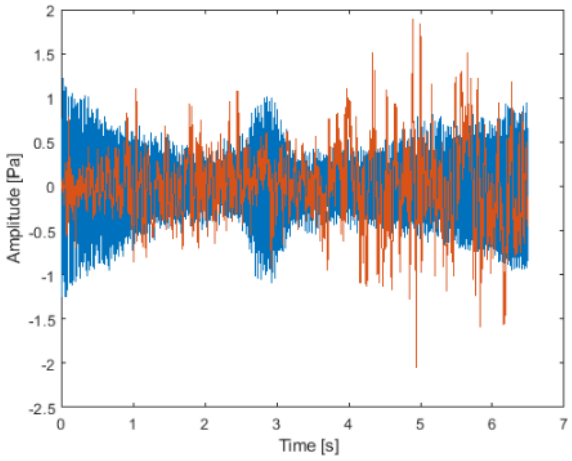


Figure 37, amplitude pressure of synthesized orders and broadband noises.

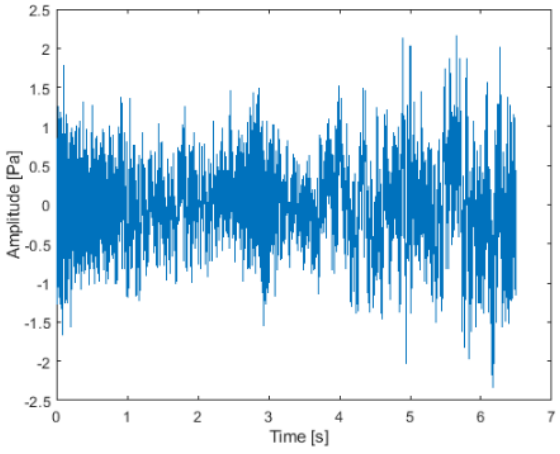


Figure 38, amplitude pressure of synthesized full vehicle sound.

5 CNN IMPLEMENTATION

In this chapter, the synthesized sounds, manipulated to include the smoothing function with random fluctuations and a Hann window, will be pre-processed to create their spectrogram images. Then, the resulting images will serve as input to train CNNs. The training of the networks will be done by using synthesized sounds with smoothing orders because the studies done to create the classification metrics, G and dRPM, also included the smoothing function. Details about the architectures used as well as the hyperparameters and the number of training cases generated will also be exposed.

These networks will be used to perform two types of classification tasks: *Yes*/*No* classification and *YesSevere*/*YesMild*/*No* classification. Further details about these classifications can be found on chapter 6.1. Later, the networks will be tested to try to classify sounds synthesized without the application of the smoothing function. Everything presented in this chapter was done in MATLAB 2019.

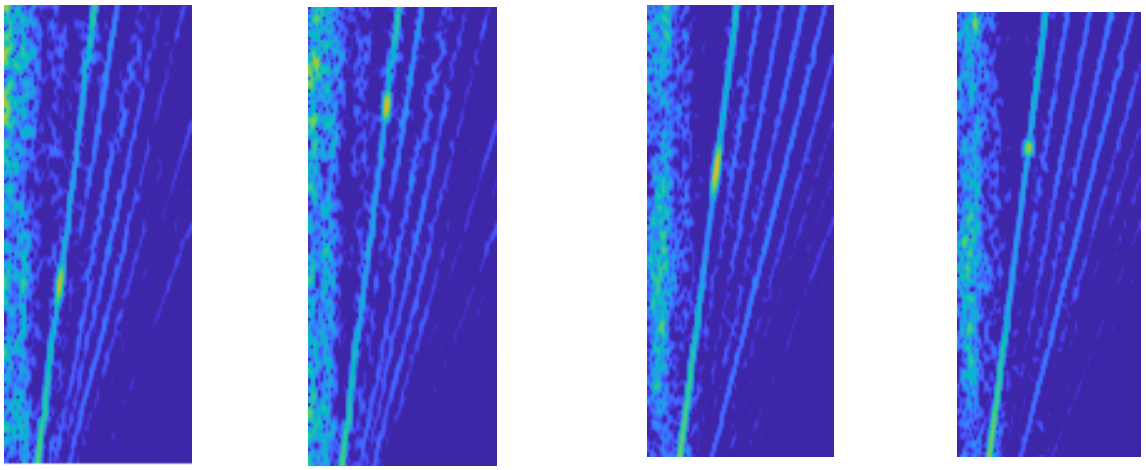
5.1 Pre-processing

Once the vehicle sounds have been synthesized, some pre-processing work must be done in order to transform the time signal data into spectrogram images, which will serve as input to the CNNs.

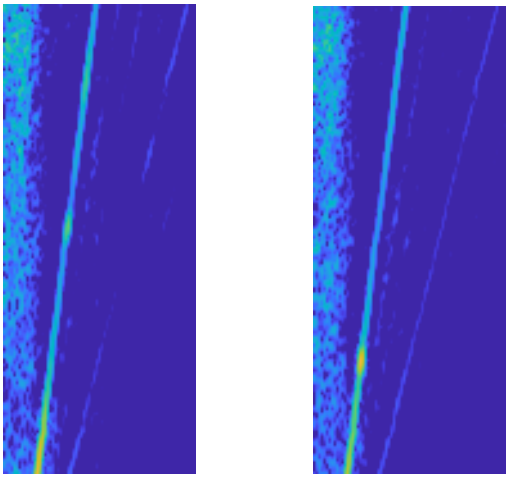
While plotting the spectrogram images, some concerns must be addressed in order to reach better results in the CNNs classification:

1. The Y-axis of the spectrograms must be the engine RPM instead of the time. This will make evident the engine's order 2, where the booming noise is induced;
2. All the spectrograms must have the same RPM and amplitude color scales. This is necessary because the metrics to classify the booming noise were defined in those scales. As already stated, the RPM range is between 1500 and 4230 RPM, with a resolution of 15 RPM which is the RPM resolution of the extracted orders. The amplitude color scale of the spectrograms is between 60 and 105 dB. These values were selected after analyzing the amplitude content of order 2 in all cars;
3. The frequency axis can exclude high frequency components. Since the main orders' profiles are focused on the low frequency end of the spectrum, it is beneficial for the classification algorithms to be fed with that low frequency content, having less data to analyze and derive features from.

In figure 39, examples of the resulting spectrograms are presented. For each car, the image on the left corresponds to the presence of booming noise, while the image on the right does not. As it is exemplified in the Ford Focus spectrogram images, there are many examples where it is difficult for the human eye to correctly evaluate the presence of booming noise by looking at the spectrogram images. Implementing a CNN will be beneficial to better classify this phenomenon.



a) Ford Focus b) Ford Mondeo



c) Opel Vectra

Figure 39, spectrograms of synthesized full vehicle sounds.

Now that it is possible to generate a diverse database for the input data, the next step consists in making sure that the database has an evenly distributed number of cases for each classification class. In a first instance, the classification was between *Yes* or *No*, referring to the presence or not of booming noise. Later, the classification problem was expanded, including *YesMild*, *YesSevere* and *No*. This divides the *Yes* classification into two domains, according to the degree of booming.

As already exposed on chapter 5.3, the conditions for the *Yes* classification of booming are:

$$Yes: \ dRPM > 280 \ \wedge \ G > 0.5$$

All the other cases correspond to a *No* classification. Regarding the *YesMild*, *YesSevere* and *No* classification, the conditions imposed for the classification were arbitrarily chosen by the author of this dissertation. The goal was to evaluate the performance of the networks when an extra class was created. The conditions for the *YesMild* and *YesSevere* classification of booming are:

$$YesMild: \ dRPM > 280 \ \wedge \ 0.5 < G < 1$$

$$YesSevere: \ dRPM > 280 \ \wedge \ 1 < G < 1.5$$

All the other cases correspond to a *No* classification. With this in mind, measures were taken to make sure that evenly distributed databases were created.

5.2 Architectures layout and hyperparameters

As mentioned before, a supervised learning approach was applied, meaning that the networks know the input's labels while training. The CNNs used are very similar to the ones on figure 15, where a set of convolutional and pooling layers are used, and when the last pooling layer is reached, a set of fully connected layers lead to the final outputs. The difference lies on the addiction of a dropout layer after the last pooling layer, and batch normalization after each convolutional layer.

The dropout technique was used with the goal to eliminate overfitting, which was being observed in an initial training phase of the networks. Batch normalization was applied since it is a well establish technique, making the training more stable and allowing the use of higher learning rates, accelerating the training process.

The used activation functions on the convolutional layers and the outputs were the ReLU and Softmax, respectively. The Softmax activation function is similar to the sigmoid function, but instead of producing independent probabilities, the probabilities produced sum up to 1 by design.

Max pooling was defined as the pooling strategy because it selects the brighter pixels of an image. This is ideal for the present study, as the booming noise is associated with high amplitude pressures, which will be represented by brighter pixels on the spectrogram images.

Using CNNs algorithms implies the manipulation of other several parameters, being the learning rate one of them. This parameter can interfere with both training speed and accuracy. If it is too low, it not only causes a lower training speed, because of the slow update of the weights and bias, but it also can get the training stuck in a local minimum instead of the intended global minimum. However, if it is too high, the system may never converge, and the global minimum will not be found. One way to allow high learning rates, but still guaranteeing convergence is to apply a learning rate decay. As the names suggests, the learning rate will decrease throughout the training, according to the learning rate decay period.

There are a multitude of other parameters which can be tuned in order to try to obtain the best accuracy possible. However, considering that extreme values of those parameters are not used, there are no significant changes on the performance of the networks by slightly alter the parameters. This means that after finding a good balance of the parameters, there is no need for further tuning them.

The parameters used are exposed on table 4. The gradient descent method used was the stochastic gradient descent with momentum.

Table 4 – Hyperparameters values.

Momentum	0.9
Learning rate	0.001
Learning rate decay	0.1
Learning rate decay period	10
Mini-batch size	32
Convolution kernel size	3 x 3
Number of filters on convolutional layer *n*	$8 \times 2^{n-1}$
Padding	same
Convolution stride	1
Max pooling kernel size	2 x 2
Max pooling stride	2
Dropout	50%
L2 regularization coefficient	0.0001

Choosing the appropriate number of convolutional layers is an important step to guarantee higher accuracies, bearing in mind that:

1. A low number of layers may lead to a low training accuracy, as the low complexity of the networks might not be able to correctly apprehend all the scenarios;
2. A high number of layers may cause overfitting to the data and increase both the computational cost and time. The overfitting of the networks to training data leads to a good accuracy during the training phase (meaning all the scenarios were learn), although this accuracy will drop considerably on the test phase, when evaluating unseen scenarios, implying that there's a lack of capability to generalize the predictions on unseen data.

This facet will be explored on the upcoming chapters.

5.3 Number of training cases

Before trying to find the suitable number of convolutional and fully connected layers, a study about the number of training cases to generate must be done. For this, a simple architecture was used, with just 3 convolutional layers and 1 fully connected layer, which connects the last max pooling operation with the outputs. This study was done only with the *Yes* and *No* classification.

First, 1250 cases were created, of which 1000 cases were for training and 250 for testing. This means that 80% of the generated was used for training. The distribution of number of cases for each class is evenly distributed, in both train and test sets. In figure 40, it is plotted the accuracy and loss curves, corresponding to the training of the referred case. On the accuracy plot, the light blue curve refers to the evolution of the training accuracy, the dark blue curve to the smoothed training accuracy and the black curve to the test accuracy. On the loss plot, the orange curve corresponds to the evolution of the training loss, while the black curve to the test loss.

In this the training session, 100 epochs were done, as in each epoch all 1000 cases were analyzed. Each iteration corresponds to one pass through a mini batch.

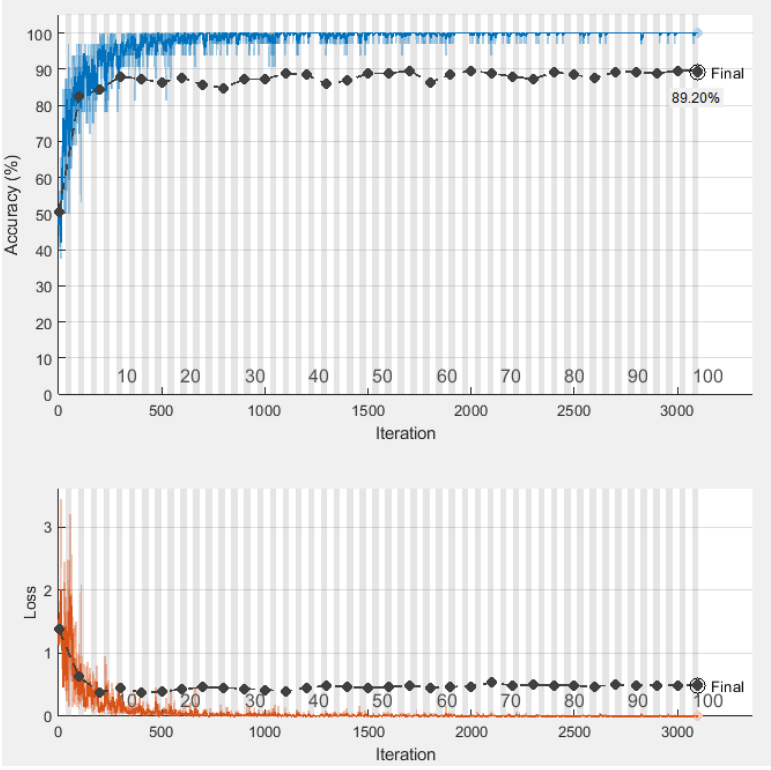


Figure 40, accuracy and loss curves for 1250 cases.

It can be observed that even with the regularization applied, there is still presence of overfitting. Therefore, the next step consists in adding more training data so that the network cannot memorize the input data, bringing the two curves together. For this, 12500 cases were generated, of which 10000 cases were for training and 2500 for testing. As before, this means 80% of the data will be used for training.

Also, the distribution of number of cases for each class is evenly distributed, in both train and test sets. In figure 41, it is plotted the accuracy and loss curves for this case.

In this training, only 10 epochs were done because to do a correct comparison between 1000 and 10000 training cases, the algorithms must do the same number of iterations, as can be seen on the figures presented.

Analyzing the content of figure 41, it is immediate to note that the train and test curves are now closer together. Therefore, the problem of overfitting from before was eliminated. Moreover, the accuracy also increased. Because of all of this, the studies done on the next chapters use 10000 training cases.

It is also worth to note two more aspects:

1. Although the train and test curve are still developing, the number of iterations done is enough to compare the two different cases, as the curves are relatively stable;
2. On both cases, the test accuracy curve presents several periodic valleys. The reason for this lies on the use of mini batches for training. However, this is not relevant since the learning rate has decay, resulting in increasingly smaller valleys with training.

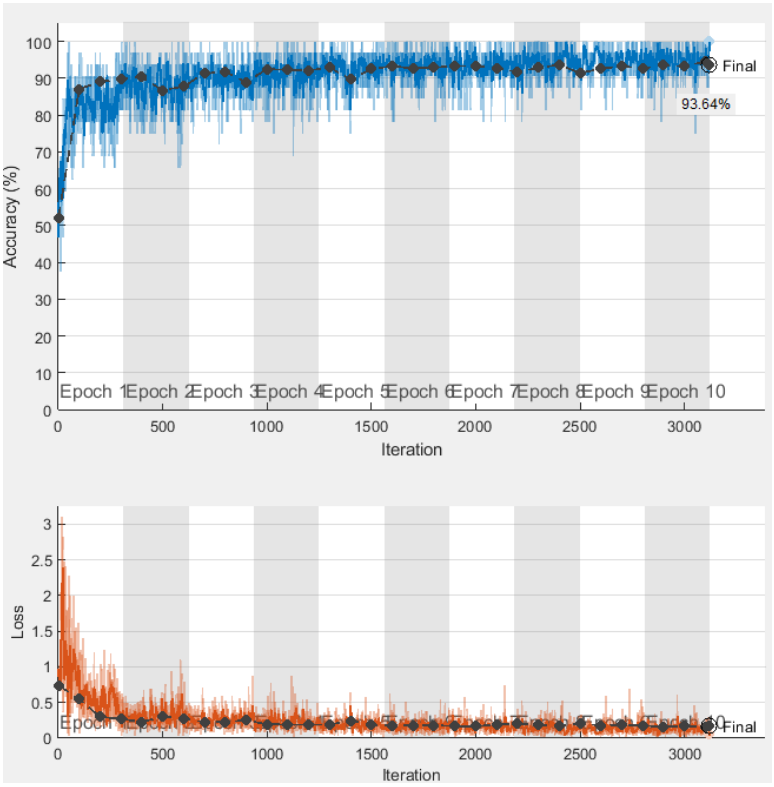


Figure 41, accuracy and loss curves for 12500 cases.

5.4 Test results

With the number of training cases defined, the next phase is to try to find a suitable number of convolutional layers. For this, it was evaluated the influence of increasingly change the number of convolutional layers.

In order to do this study, the mean of the accuracy of 5 tests was calculated for 10 epochs for each case. The number of epochs for each test was 10 because with that number of epochs the networks are stable enough in order to allow a comparison between networks, as can be seen on figure 41. The obtained results are presented on figure 42.

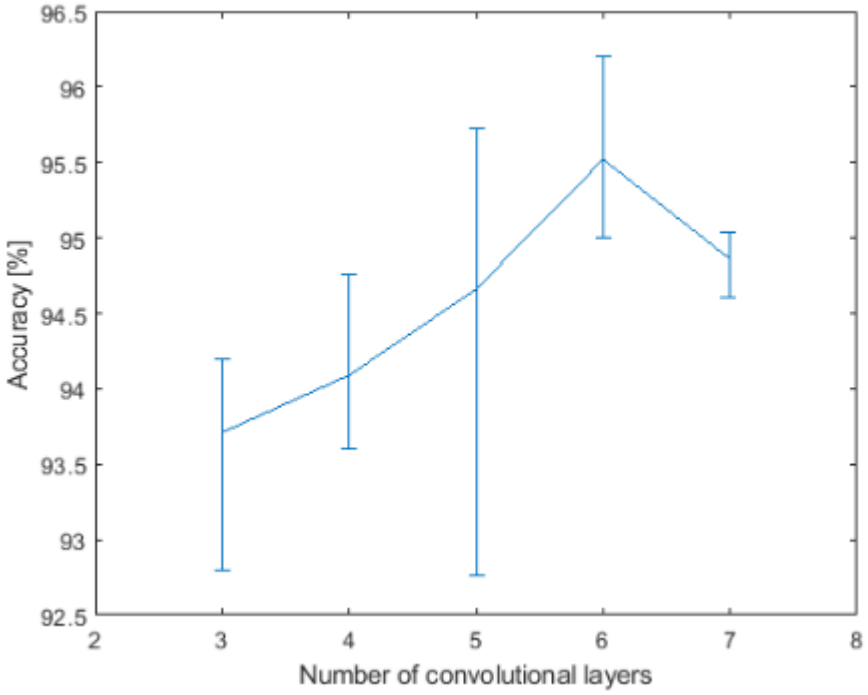


Figure 42, minimum, maximum and mean values of test accuracy represented for each number of convolutional layers.

By analyzing the results, the most promising architecture was revealed to have 6 convolutional layers. The test accuracy mean value presents an increase from three to 6 convolutional layers, decreasing when 7 convolutional layers are used. It is important to note that the accuracy mean value of a certain number of convolutional layers is very close to the accuracy maximum value of the previous layer, except on layer 7, in which the accuracy mean value is close to the accuracy minimum value of the previous layer. This indicates that choosing 6 convolutional layers is the best option for the present case.

Since there are some fluctuations on the accuracy value of the selected architecture and because the test accuracy curve seems to be still increasing, to calculate the final accuracy 30 epochs were done. An example of a test conducted in these conditions is presented on figure 43. It can be seen that there is presence of overfitting, and the test accuracy does not reach higher values than before.

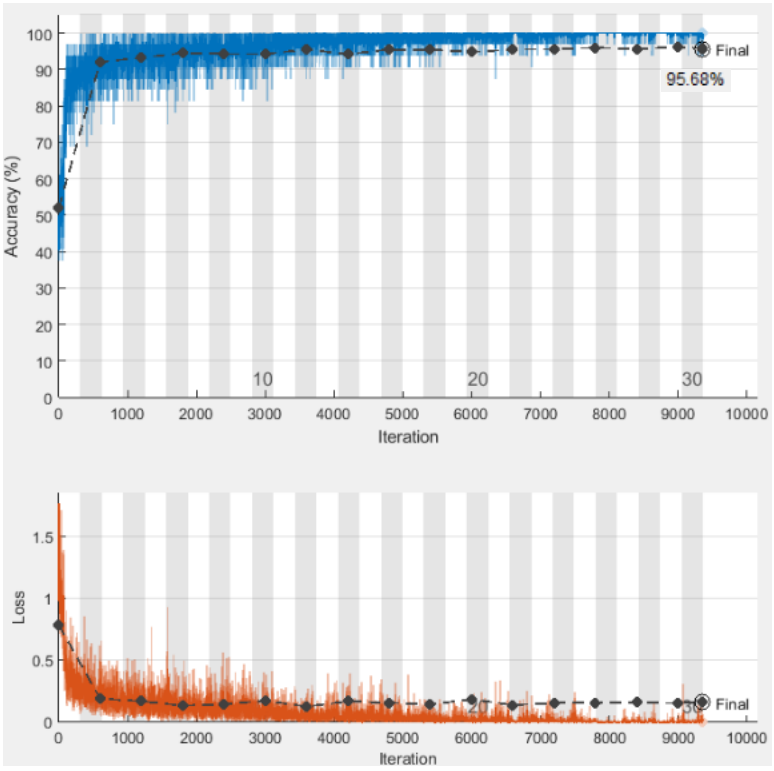


Figure 43, Accuracy and loss curves for the *Yes*/*No* classification with 30 epochs.

As already stated, all the test results presented are for the *Yes*/*No* classification of booming. To evaluate the *YesMild*/*YesSevere/No* classification, a similar strategy was taken. Since this classification focus on the same parameters as the *Yes*/*No* classification, but with a higher degree of complexity, 6 convolutional layers was the training starting point, and the evaluation of higher number of convolutional layers was also done. The obtained results are presented on figure 44.

By analyzing the results, the most promising architecture was revealed to have 6 convolutional layers. As before, since there are some fluctuations on the accuracy value of the selected architecture and because the test accuracy curve seems to be still increasing, to calculate the final accuracy 30 epochs were done. An example of a test conducted in these conditions is presented on figure 45. It can be seen that there is presence of overfitting, and the test accuracy does not reach higher values than before.

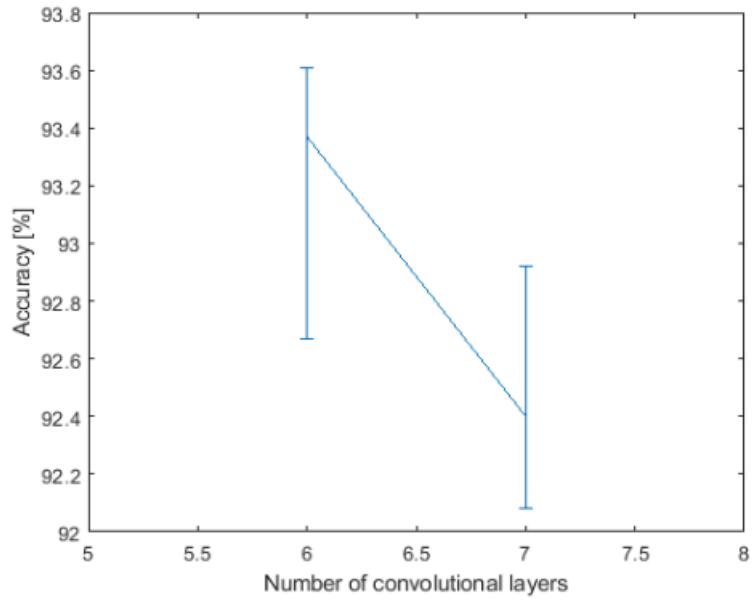


Figure 44, minimum, maximum and mean values of test accuracy represented for each number of convolutional layers.

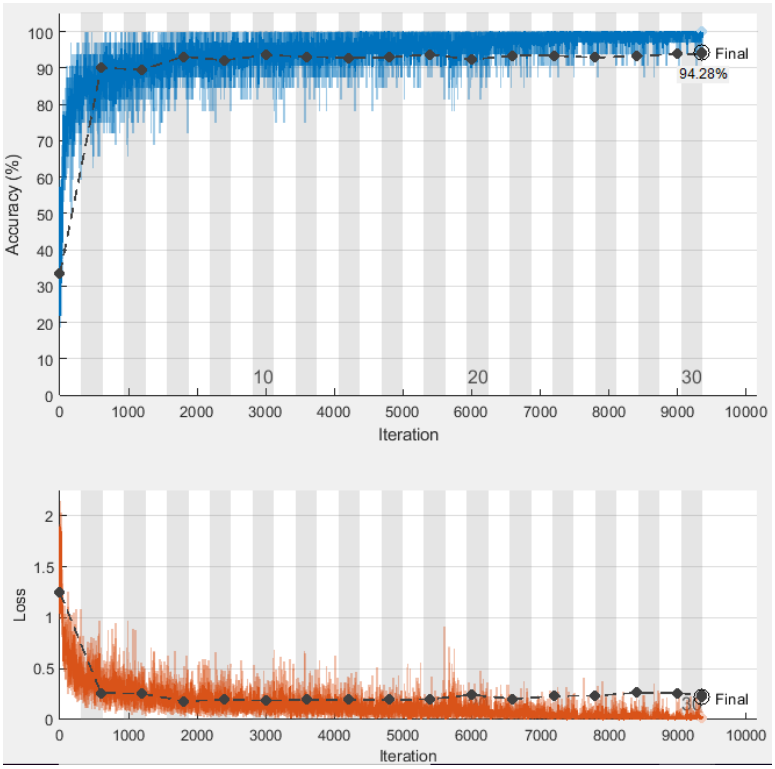


Figure 45, Accuracy and loss curves for the *YesSevere*/*YesMild*/*No* classification with 30 epochs.

5.5 Test results on non-smoothed orders

Until now, both the training and test sets were created using the smoothing function. In this chapter, the results of testing with non-smoothed orders are presented. The networks used for testing were the ones presented on the last chapter, so the training was still done using the smoothing function. The goal is to evaluate the performance of training with smoothed cases and testing with non-smoothed cases. First, for the *Yes*/*No* classification, the results are shown on table 5.

Table 5 – Results of the accuracy on non-smoothed cases for each network.

Network	Number of convolutional layers	Number of epochs	Accuracy on smoothed cases	Accuracy on non-smoothed cases
1	3	10	94,20 %	84,51 %
2	4	10	94,76 %	84,78 %
3	5	10	95,72 %	86,11 %
4	6	10	96,20 %	85,25 %

By looking at network 3, which performed the best on the non-smoothed cases, a study of the accuracies of individual cars was done to try to understand why the accuracies decreased in the non-smoothed cases. The results are shown on table 6. It is immediate to note that the accuracy on non-smoothed Ford Focus cases is substantially higher that the other cases.

Table 6 – Accuracy of network 4 non-smoothed cases for each car.

Network	Accuracy on non-smoothed cases	Accuracy on non-smoothed Ford Focus cases	Accuracy on non-smoothed Ford Mondeo cases	Accuracy on non-smoothed Opel Vectra cases
3	86,11 %	93,37 %	81,33 %	83,73 %

Now for the *YesMild*/*YesSevere*/*No* classification, the same study was done, and the results are presented at table 7.

Table 7 - Results of the accuracy on non-smoothed cases for each network.

Network	Number of convolutional layers	Number of epochs	Accuracy on smoothed cases	Accuracy on non-smoothed cases
6	6	10	93,44 %	70,34 %
7	7	10	92,92 %	78,16 %

By looking at network 7, which performed the best on the non-smoothed cases, a study of the accuracies of individual cars was done. The results are shown on table 8. It is immediate to note that the accuracy on non-smoothed Ford Focus cases is substantially higher that the other cases.

Table 8 – Accuracy of network 7 on non-smoothed cases for each car.

Network	Accuracy on non-smoothed cases	Accuracy on non-smoothed Ford Focus cases	Accuracy on non-smoothed Ford Mondeo cases	Accuracy on non-smoothed Opel Vectra cases
7	78,16 %	90,54 %	69,01 %	75,05 %

5.6 Results validation

In this section, an evaluation of the cases that were misclassified is done. This study will focus on the test databases with smoothing and non-smoothing orders, for the *Yes*/*No* classification of booming. So, the network 4 will be used for the smoothed databases, and network 3 will be used for non-smoothed databases.

On figure 46, it is shown a scatter plot with 30 random misclassified cases of network 4. The horizontal black line represents the boundary condition for G, while the black vertical line represents the boundary condition for dRPM. Most points are located near those boundary conditions, especially close to the orange dot, which is just a reference point. The dots form a “L” shape on the boarders of the top right square because that is the region separating the *Yes*/*No* classification. No critical misclassification was observed. On figure 47, the histogram of number of classes in different RPM ranges is presented. The x-axis represents the RPM where the Hann window starts. More than half of the misclassifications occurred in the low RPM range.

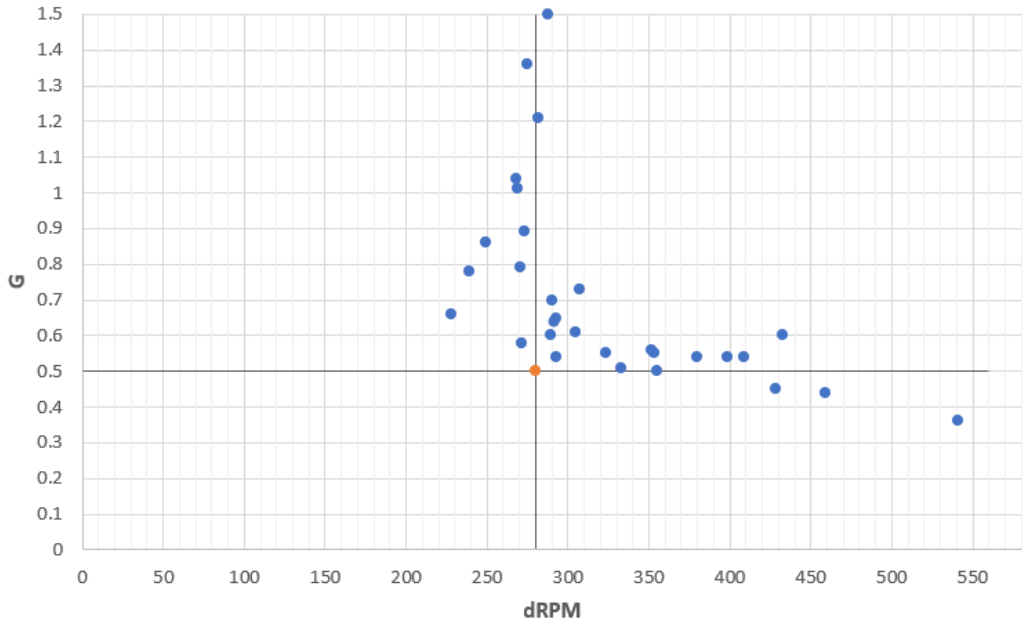


Figure 46, network 4 misclassifications examples.

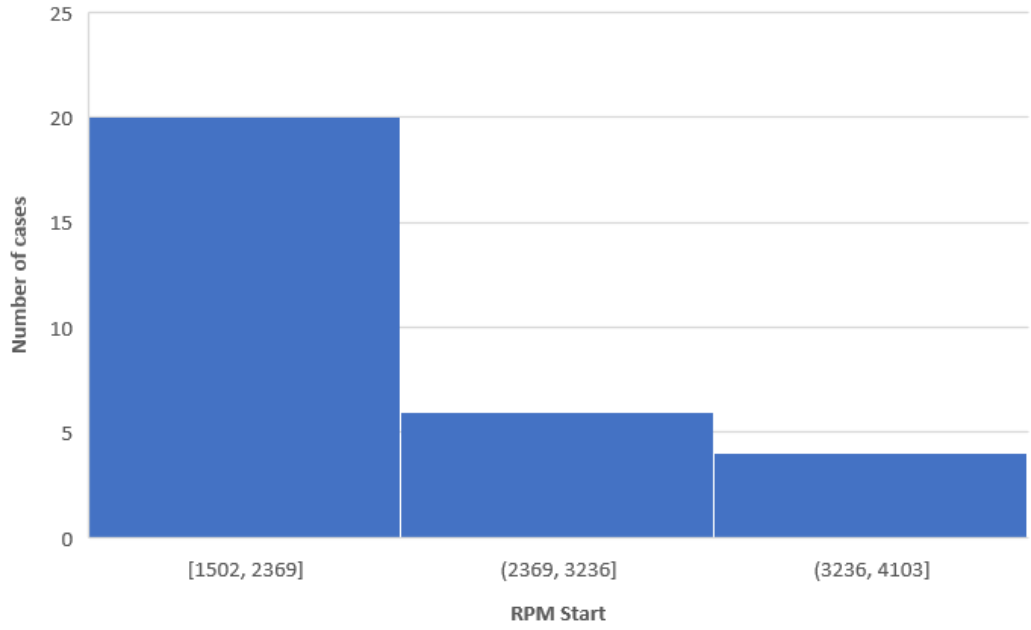


Figure 47, histogram of number of classes in different RPM ranges in network 4.

On figure 48, it is shown a scatter plot with 30 random misclassified cases of network 7. In this case, although there is still a concentration of points around the boundary conditions, the points are much more scattered, and some critical misclassifications were made. On figure 49, the histogram of number of classes in different RPM ranges is presented. Half of the misclassifications occurred in the low RPM range.

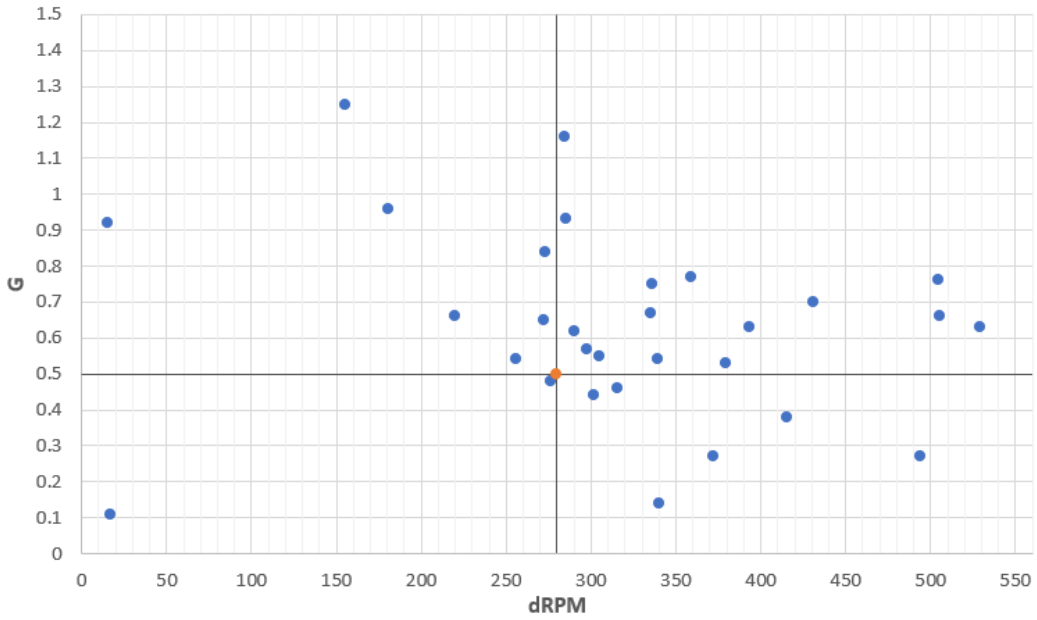


Figure 48, network 7 misclassifications examples.

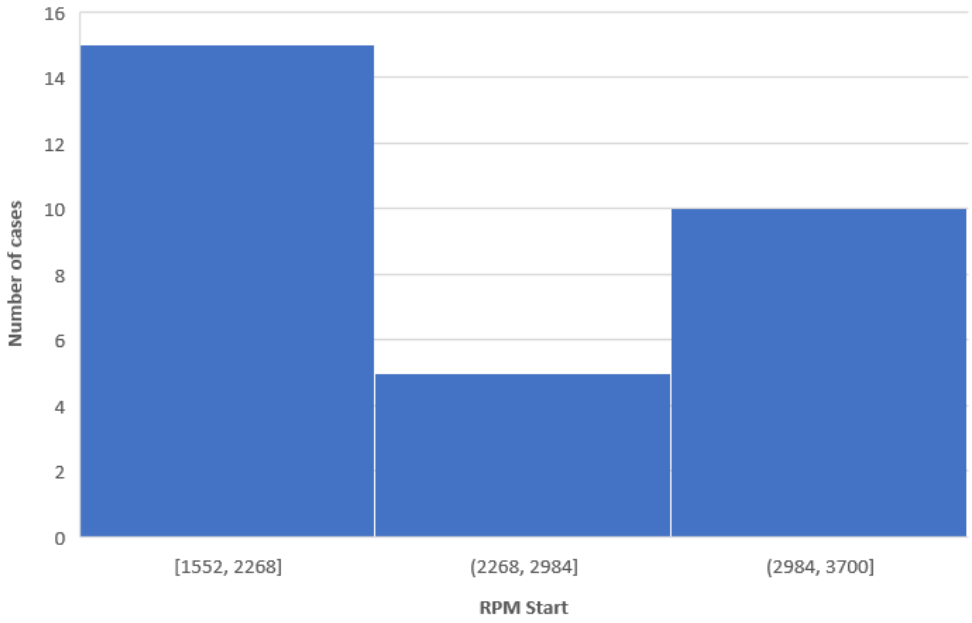


Figure 49, histogram of number of classes in different RPM ranges in network 7.

6 DISCUSSION

In this dissertation, digital twins of three cars were manipulated to induce Hann windows with different amplitudes and lengths in an RPM range from 1500 to 4230 RPM, creating several booming scenarios. On chapter 4.2, was considered the presence of booming in the experimental data of the Ford Mondeo and the Opel Vectra. By analyzing the decomposed orders, high amplitude peaks characterized by bell-shaped curves on the 2^{nd} engine order were observed on the RPM range experienced as booming. On (Siano and Panza 2017), similar conclusions were made about the relation between shape and amplitude of order 2 curve and the presence of booming noise. Other studies like (Park and Lee 2012) also assumed a bell-shape curve as a good indicator of booming noise, since in their study booming noise was synthesized by applying a bell-shaped curve to the engine order 2. Therefore, Hann widows are included on the SQE models to try to resemble that bell-shaped curve.

Other modifications were made to the engine order 2 before synthesizing the sounds, such as including low frequency random fluctuations on the engine orders, as well as using a smoothing function. On (Park and Lee 2012), a simplistic approach has taken while synthesizing the engine order 2. As mentioned, the authors included a bell-shaped on engine order 2 to simulate the booming sensation, but the remaining content of that order has the same pressure amplitude.

Full vehicle sounds were synthesized using the modified cars' models, and then the resulting sounds were plotted into spectrogram images, with an RPM axis instead of a time axis. This allowed the generation of a diverse database containing different booming scenarios, which were used to train a CNN. On (Park and Lee 2012), NNs were also used to evaluate the booming noise. However, the authors took a different approach. Their pre-processing work before feeding the NNs was to apply algorithms to extract psychoacoustic metrics and then train the networks to calculate a booming index. The work on this dissertation has a simpler pre-processing process since only the spectrograms of the sounds were plotted. Another difference between the two works lies on the fact that, in their work, the jury tests done to create the NNs labels were continuous values characterizing the booming index, turning their problem into a regression problem. In this dissertation, the jury tests only allowed to do a classification task.

This approach was taken even though the results obtained by using CNNs to evaluate spectrogram images have not been as good as the ones obtained for visual images (Wyse 2017). This lies on the fact that machine vision techniques are being applied in order to do machine hearing. Unlike visual images, sounds are not static objects and they arrive as a sequence of air pressure. This can cause many problems for CNNs to analyze spectrogram images. However, the classification metrics used in this dissertation for the assessment of the booming noise are clearly visually present on the spectrograms, when the RPM axis is used. Moreover, CNNs have been applied to classify other NVH problems, such as on (Stender et al. 2020), and good results have been obtained.

The chosen solution was proven to be effective for classifying the simulated booming noise on the spectrogram images, since high accuracies were reached for the classification of *Yes*/*No* and *YesSevere*/*YesMild*/*No*, respectively.

In chapter 6.3, the number of training cases was chosen by observing the overfitting with 1000 training cases, so 10000 cases were studied, and the overfitting disappeared. However, this study was done for only 10 epochs. In chapter 6.4, when 30 epochs were done on the best networks' architectures, a minor overfitting can be noticed. Therefore, if more training data was generated, there is potential for the classification to reach even higher accuracies. This aspect is presented on figures 43 and 45.

When analyzing figures 43 and 45, it can also be observed that high accuracies were reached for both classification tasks, as already stated. The higher accuracy corresponding to the classification of *Yes*/*No* was expected, since the classification of *YesSevere*/*YesMild*/*No* is very similar but with extra boundary conditions of the same nature. Moreover, due to this similar boundary conditions, both classification tasks have the same number of convolutional layers on the optimized network, even though the *YesSevere*/*YesMild*/*No* classification is more complex.

The accuracies of the networks, trained on smoothed orders, drops significantly when testing with non-smoothed databases. This can be explained as being a tunnel vision effect, where the networks are good at classifying a certain type of data, but when small variations are tested, the accuracies decrease substantially. It is worth to note that the drop in accuracy for the Ford Focus case was minimal comparing to the Ford Mondeo and Opel Vectra cases. It can be argued that one reason for that are the higher amplitude pressures on the low RPM range of Ford Mondeo and Opel Vectra, comparing to the Ford Focus. Those high pressures appear brighter in the spectrogram images, making the assessment of the booming noise in low RPM ranges more difficult. The histogram on figure 49 supports this since most misclassifications occurred at a low RPM range. This phenomenon can be observed even with the application of a smoothing function, as seen on figure 47, because the amplitudes at low RPMs are still relatively high. As discussed on chapter 4.2, there are possible occurrences of booming noise at low RPM ranges in Ford Mondeo and Opel Vectra which can also induce errors in the classification of these cars.

One of the reasons the smoothing function was included was to eliminate the natural order non-linearities which occur at low pressure amplitudes, in order to do an effective training of the CNNs. Therefore, one of the causes for the accuracy drop on non-smoothed cases could be the network misclassifying those non-booming order non-linearities. This phenomenon is amplified when the low frequency random fluctuations create cases like the one on figure 31 a).

Sometimes the presence of booming noise at low RPM ranges can be associated with the powerfulness of the vehicle, not being perceived as an annoying acoustic feature. Having this in mind, it can be said that good results were obtained, even in the tests with non-smoothing orders.

7 CONCLUSION

7.1 Conclusions

Several conclusions can be made be made on the research and work developed throughout this dissertation. Firstly, a statement must be made regarding the success accomplished by algorithm implementation with the purpose of booming noise detection/classification, using CNNs as the chosen NNs approach, to evaluate full vehicle spectrogram images.

During this work, it was found that NVH engineering is a complex topic since the human perception of sounds is subjective, which makes it difficult to evaluate the impact that the many different sounds heard inside a car's cabin have in the costumer's comfort. It was also seen that, in vehicles moved by an ICE, the booming noise is one of the most annoying acoustic features, and well define metrics to classify this noise still do not exist.

Throughout this dissertation, an effective way to generate a diverse database of booming scenarios, in a fully automated way, was created. It has been concluded that a good way to simulate the booming noise is by including a bell-shaped curve on the engine order 2 profile. To achieve this, Hann windows with different lengths and amplitudes were added in an RPM range from 1500 to 4230 RPM.

It is acknowledged that, by using modified digital twins to include booming scenarios, there is not the risk of scarcity of data since they can augment synthetically the database. This has been proven to help reaching high accuracies in the booming classification tasks.

The *Yes*/*No* classification of the booming noise reached higher accuracies than the *YesSevere*/*YesMild*/*No*, which was expected since the metrics used to define these classifications are very similar. The high accuracies reached in both cases confirms that the mathematical complexity of the CNN approach is sufficient to correctly classify the booming noise defined metrics.

All the training of the CNNs was done with the application of a smoothing function on the vehicle engine orders, better isolating the booming event. However, when testing with non-smoothed cases, a decrease in accuracy has been noticed, which means that the CNNs can extract order features quite well. It was also concluded that many misclassifications occurred at the low RPM range because of the high-pressure amplitudes in that regime, in both smoothed and non-smoothed cases.

7.2 Future work

In order to continue the work done on this dissertation, the following topics should be considered for future works:

- Develop the SQE models to include more gear and throttle conditions, as well as a higher RPM range. This way, a more diverse database can be created, expanding the classification tasks to a broader spectrum;
- Use the generated database to perform regression tasks. For example, the output of the regression model can give information about the RPM location of the booming event, as well as the parameters G and dRPM;
- Train the networks with the matrices used for generating the spectrograms. The goal is to avoid losing information by turning those matrices to RGB matrices;
- Increase the RPM resolution of the spectrograms. This is beneficial to reach even higher accuracies, by narrowing the grey area of the failed classifications;
- Train networks to do classification of the booming phenomenon not only on order 2 but for any order number;
- Take a hybrid approach to train the networks, combining smoothed with non-smoothed orders;

REFERENCES

Blough, Jason R., David L. Brown, and Håvard Vold. 1997. “The Time Variant Discrete Fourier Transform as an Order Tracking Method.” In , 972006. https://doi.org/10.4271/972006.

Brunton, Steven L, and J Nathan Kutz. 2017. “Data Driven Science & Engineering,” 572.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2015. *Deep Learning*.

Gupta, Gaurav, Rituraj Gautam, and Chetan Prakash Jain. 2014. “Study of Coupling Behavior of Acoustic Cavity Modes to Improve Booming Noise in Passenger Vehicles.” In , 2014-01–1974. https://doi.org/10.4271/2014-01-1974.

Hasegawa, Satoshi, Toshiyuki Tabata, Akio Kinoshita, and Hideki Hyodo. 1992. “The Development of an Active Noise Control System for Automobiles.” In , 922086. https://doi.org/10.4271/922086.

Hatano, S. 1999. “Modification of Booming Level for Higher Correlation with Booming Sensation.” *JSAE Review* 20 (1): 123–25. https://doi.org/10.1016/S0389-4304(98)00052-6.

Hatano, S, and T Hashimoto. 1996. “On an Objective Measure of the Booming Sound Factor —Modification of the Measure for the Spectrum Pattern and the Loudness of the Sound” 27: 85–89.

———. 2000. “BOOMING INDEX AS A MEASURE FOR EVALUATING BOOMING SENSATION,” 5.

Inoue, Toshio, Akira Takahashi, Hisashi Sano, Masahide Onishi, and Yoshio Nakamura. 2004. “NV Countermeasure Technology for a Cylinder-On-Demand Engine- Development of

Active Booming Noise Control System Applying Adaptive Notch Filter." In , 2004-01–0411. https://doi.org/10.4271/2004-01-0411.

Ioffe, Sergey, and Christian Szegedy. n.d. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," 9.

Jha, S.K. 1976. "Characteristics and Sources of Noise and Vibration and Their Control in Motor Cars." *Journal of Sound and Vibration* 47 (4): 543–58. https://doi.org/10.1016/0022-460X(76)90881-6.

John Britto, Vijay Antony, Sudipto Karmakar, Madhan Muthuveeraswamy, and Balasubramanian Natarajasundaram. 2016. "High Speed Booming Noise Reduction in Passenger Car by Application of Cost Optimized NVH Solution." In , 2016-28–0039. https://doi.org/10.4271/2016-28-0039.

Kang, Hyungsouk, TaeYoung Chung, Hyeongcheol Lee, and Hyungbin Ihm. 2016. "Active Booming Noise Control for Hybrid Vehicles." *SAE International Journal of Passenger Cars - Mechanical Systems* 9 (1): 167–73. https://doi.org/10.4271/2016-01-1122.

Kingma, Diederik P., and Jimmy Ba. 2017. "Adam: A Method for Stochastic Optimization." *ArXiv:1412.6980 [Cs]*, January. http://arxiv.org/abs/1412.6980.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–44. https://doi.org/10.1038/nature14539.

Lee, Youn-Hee, and Adel Nasiri. 2007. "Real Time Active Noise Control of Engine Booming in Passenger Vehicles." In , 2007-01–0411. https://doi.org/10.4271/2007-01-0411.

Mansinh, Kumbhar S., Rajesh Bhangale, Mohammad Riyazuddin, and Murali Bodla. 2012. "A Development of Booming Index of Diesel SUV by Using Artificial Neural Network." In , 2012-01–1542. https://doi.org/10.4271/2012-01-1542.

Matsuyama, Souhei, and Shinichi Maruyama. 1998. "Booming Noise Analysis Method Based On Acoustic Excitation Test." In , 980588. https://doi.org/10.4271/980588.

Mitchell, Tom M. 1997. *Machine Learning*. McGraw-Hill Series in Computer Science. New York: McGraw-Hill.

Oh, Shi-Hwan, Hyoun-suk Kim, and Youngjin Park. 2002. "Active Control of Road Booming Noise in Automotive Interiors." *The Journal of the Acoustical Society of America* 111 (1): 180–88. https://doi.org/10.1121/1.1420390.

Panza, Maria Antonietta. 2015. "A Review of Experimental Techniques for NVH Analysis on a Commercial Vehicle." *Energy Procedia* 82 (December): 1017–23. https://doi.org/10.1016/j.egypro.2015.11.861.

Park, Jong Ho, and Sang Kwon Lee. 2012. "Identification of Vehicle Booming Sound and Its Objective Evaluation Using Psychoacoustic Parameters." *International Journal of Vehicle Design* 58 (1): 46. https://doi.org/10.1504/IJVD.2012.045922.

Ponseele, P Van de. 2012. "Source-Transfer-Receiver Approaches: A Review of Methods," 14.

Priddy, Kevin L., and Paul E. Keller. 2005. *Artificial Neural Networks: An Introduction*. Tutorial Texts in Optical Engineering, v. TT 68. Bellingham, Wash: SPIE Press.

Qian, Ning. 1999. "On the Momentum Term in Gradient Descent Learning Algorithms." *Neural Networks* 12 (1): 145–51. https://doi.org/10.1016/S0893-6080(98)00116-6.

Ruder, Sebastian. 2017. "An Overview of Gradient Descent Optimization Algorithms." *ArXiv:1609.04747 [Cs]*, June. http://arxiv.org/abs/1609.04747.

Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, et al. 2015. "ImageNet Large Scale Visual Recognition Challenge." *ArXiv:1409.0575 [Cs]*, January. http://arxiv.org/abs/1409.0575.

Saleh Albelwi and Ausif Mahmood. 2017. "A Framework for Designing the Architectures of Deep Convolutional Neural Networks." *Entropy* 19 (6): 242. https://doi.org/10.3390/e19060242.

Sarrazin, Mathieu, Claudio Colangeli, and Karl Janssens. 2013. "Synthesis Techniques for Wind and Tire-Road Noise," 11.

Shin, Sung-Hwan, Jeong-Guon Ih, Takeo Hashimoto, and Shigeko Hatano. 2009. "Sound Quality Evaluation of the Booming Sensation for Passenger Cars." *Applied Acoustics* 70 (2): 309–20. https://doi.org/10.1016/j.apacoust.2008.03.009.

Siano, D., and M.A. Panza. 2017. "Sound Quality Analysis of the Powertrain Booming Noise in a Diesel Passenger Car." *Energy Procedia* 126 (September): 971–78. https://doi.org/10.1016/j.egypro.2017.08.189.

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," 30.

Stender, Merten, Merten Tiedemann, David Spieler, Daniel Schoepflin, Norbert Hoffmann, and Sebastian Oberst. 2020. "Deep Learning for Brake Squeal: Brake Noise Detection, Characterization and Prediction." *Mechanical Systems and Signal Processing* 149: 107181. https://doi.org/10.1016/j.ymssp.2020.107181.

Toi, T, Isoyama, H, Ogu, Y, Saito, H, Hoshino, H, Ishikawa, M, et al. 2004. “Study of International Relations of Automotive Sound Quality.” In *Proceedings of the Noise and Vibration Forum 2004*, 33–39. paper no. 20044495.: society of automotive engineers of Japan,.

Wang, Y.S., G.Q. Shen, and Y.F. Xing. 2014. “A Sound Quality Model for Objective Synthesis Evaluation of Vehicle Interior Noise Based on Artificial Neural Network.” *Mechanical Systems and Signal Processing* 45 (1): 255–66. https://doi.org/10.1016/j.ymssp.2013.11.001.

Wellmann, Thomas, Kiran Govindswamy, Eugen Braun, and Klaus Wolff. 2007. “Aspects of Driveline Integration for Optimized Vehicle NVH Characteristics.” In , 2007-01–2246. https://doi.org/10.4271/2007-01-2246.

Wellmann, Thomas, Kiran Govindswamy, and Georg Eisele. 2011. “Driveline Boom Interior Noise Prediction Based on Multi Body Simulation.” In , 2011-01–1556. https://doi.org/10.4271/2011-01-1556.

Wu, Yudong, Renxian Li, Weiping Ding, Jan Croes, and Mingliang Yang. 2019. “Mechanism Study and Reduction of Minivan Interior Booming Noise during Acceleration.” *Shock and Vibration* 2019 (August): 1–17. https://doi.org/10.1155/2019/2190462.

Wyse, Lonce. 2017. “Audio Spectrogram Representations for Processing with Convolutional Neural Networks.” . . *Pp*, 6.

Yamashita, Rikiya, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. 2018. “Convolutional Neural Networks: An Overview and Application in Radiology.” *Insights into Imaging* 9 (4): 611–29. https://doi.org/10.1007/s13244-018-0639-9.

Zhen, Jiantie, Chunhui Pan, Ashish Jangale, and Brad Salisbury. 2011. “Identification and Reduction of Booming Noise on a Motor Grader.” In , 2011-01–1729. https://doi.org/10.4271/2011-01-1729.

ANNEX A – BOOMING JURY TESTS

In this annex, it will be exposed some details about the study done on booming perception by a Anastasia Patras (Siemens Industry Software), prior to the realization of this dissertation.

A jury test, in which a group of people rate sounds, is used to ascertain the exact combination of metrics needed to fully understand the perception of a product's sound quality. The goal of the jury test done was to study the perception of booming inside of a car.

Two cars were used in this study, and 36 different sounds were evaluated in gear 2 and 100% throttle conditions. The study was divided into three different tests with the following goal:

- Test 1: Access the amplitude threshold for the booming noise perception;
- Test 2: Access the maximum RPM duration for the booming noise perception;
- Test 3: Access the minimum RPM duration for the booming noise perception.

The study was attended by 26 people. The percentages of people who work in the NVH field and their age is presented on figure 50. The results obtained for the three tests done are presented on figures 51, 52 and 53.

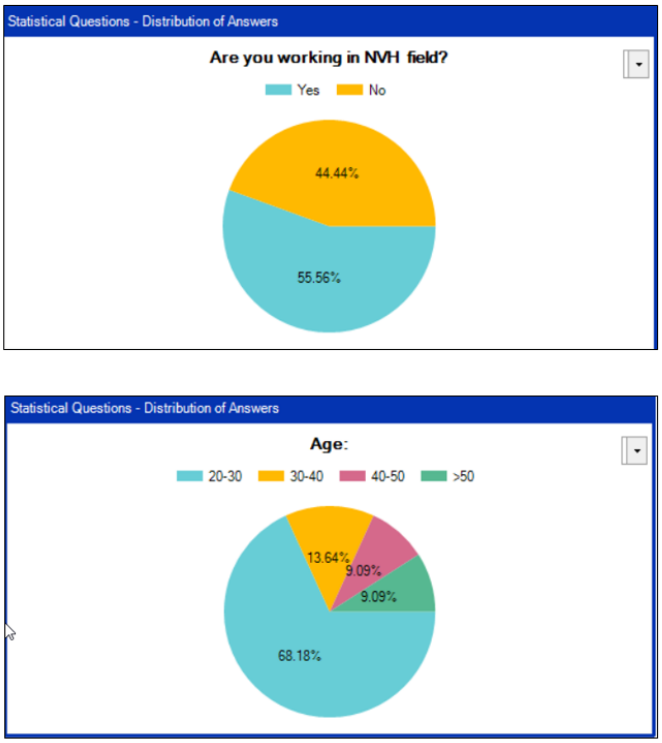
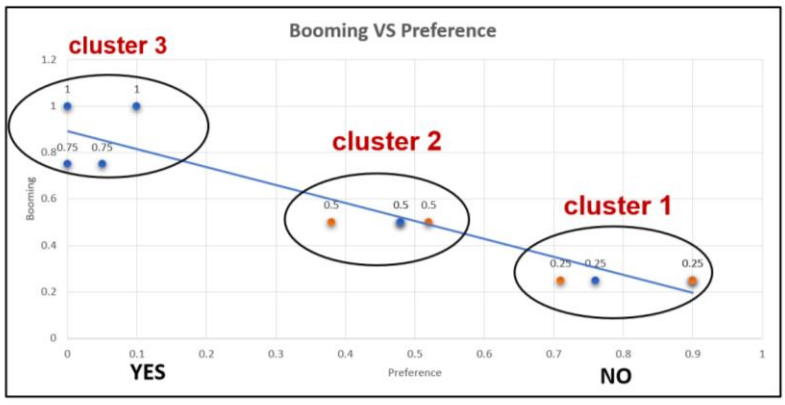
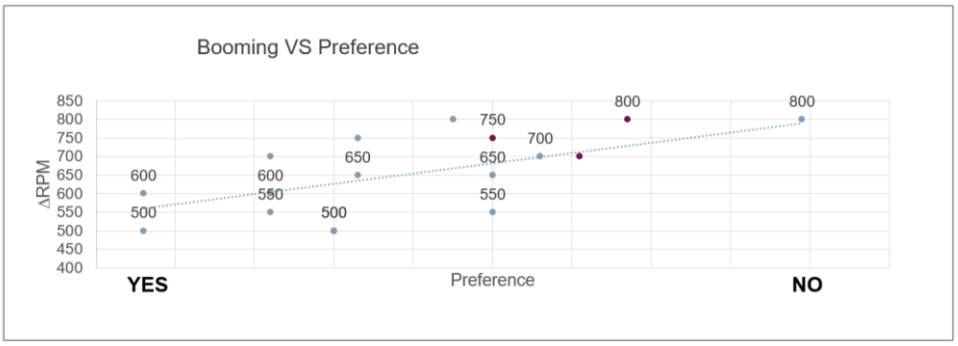


Figure 50, statistics about the participants.



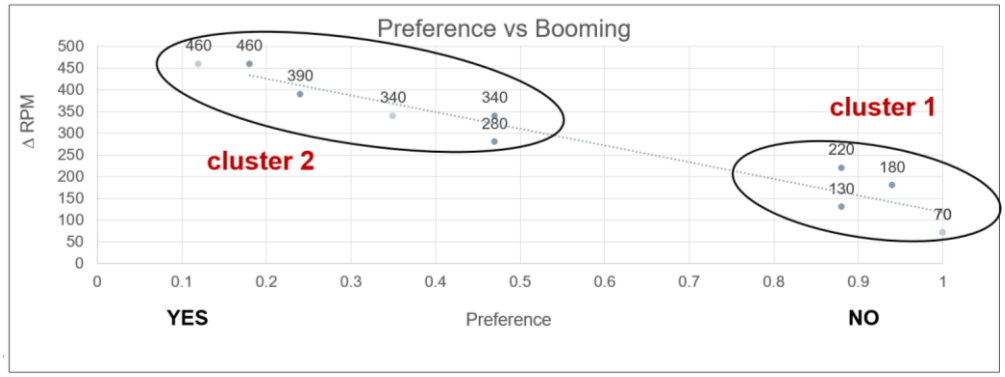
- G > 0.5 is perceived as a booming.

Figure 51, test 1 results.



- This Jury Test 2 was not conclusive.

Figure 52, test 2 results.



- ΔRPM > 280 is perceived as a booming.

Figure 53, test 3 results.