Marko Pavlić, BSc

Adaptive Pattern Recognition for Train Detection with Fibre Optic Cable

Adaptive Mustererkennung zur Erkennung von Zügen mit Glasfaserkabel

Master's Thesis

to achieve the university degree of

Diplom-Ingenieur

Master Programme: Electrical Engineering

submitted to

Graz University of Technology

Evaluator	Supervisor
Dipl-Ing. Dr.techn. Markus Neumayer	Dipl.Math. Thomas Sigl
Institute of Electrical Measurement and Measurement Signal Processing	ITK Engineering GmbH

Graz, April 2019

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Date	Signature

Abstract

Railway operators are working on concepts to increase the workload on their train network. Knowledge about the position of the trains in their train network is mandatory. Existing systems for positioning, such as balise and axle counters, show drawbacks. The Global Positioning System (GPS) is increasingly used for that purpose but also this approach has some disadvantages with respect to the availability. Therefore, railway operators focus on the research of different techniques for train localisation. One of these techniques is Distributed Acoustic Sensing (DAS). It uses a fibre optic cable mounted along a rail track to measure the vibrations produced by trains in the vicinity of the cable. In this thesis, signal processing methods are used to analyse the DAS measurement data in order to detect and thus localize trains moving on this track. The structure of the DAS measurement data makes it possible to treat the train localization as a classification problem. Different signal analysis tools, namely the Fourier Transformation, Wavelet Transformation and Empirical Mode Decomposition, are used on the measurements to extract features. The first algorithm is a Support Vector Machine (SVM), which classifies between the two classes *train* and *no-train* with a precision of 79.19% and a sensitivity of 96.85%. The influence of different features is investigated, whereby the results listed were achieved with the signal energy as the feature. The amount of labelled data was 1575000. The SVM shows a cross-sensitivity for trains driving on adjacent rail tracks. The second algorithm is an expert system and detects trains using the signal energy. A distinction between trains on the adjacent track and trains on the track under investigation can be made at runtime by using the Fourier Transformation. With this algorithm a precision of 86.38% and a sensitivity of 97.67% was achieved for the classification between the two classes *train* and *no-train*. To improve the results in further work suggestions are given at the end of this thesis.

Zusammenfassung

Bahnbetreiber arbeiten an Konzepten um die Auslastung in ihrem Netzwerk zu erhöhen. Dafür muss die Position der Züge im Schienennetz bekannt sein. Existierende Systeme zur Positionsbestimmung, wie Balisen und Achszähler, sind dafür nicht mehr ausreichend. Das Global Positioning System (GPS) wird für diesen Zweck vermehrt genutzt, jedoch hat dieses Nachteile in Bezug auf die Verfügbarkeit. Deshalb werden von den Bahnbetreibern andere Systeme zur Zuglokalisierung untersucht. Eine dieser Technologien ist Distributed Acoustic Sensing (DAS). Es verwendet ein Glasfaserkabel, welches entlang einer Gleistrasse verlegt ist, um die vom Zug erzeugten Vibrationen in der Umgebung des Kabels zu messen. In dieser Arbeit werden Signalverarbeitungsmethoden zur Auswertung der DAS Messdaten untersucht, um damit Züge, die sich auf dieser Gleistrasse bewegen, zu detektieren und somit zu lokalisieren. Die Struktur der DAS-Messdaten ermöglicht es, die Zuglokalisierung als Klassifizierungsproblem zu betrachten. Verschiedene Signalanalysemethoden, wie die Fourier Transformation, Wavelet Transformation und Empirical Mode Decomposition werden auf die Messdaten angewendet um Features zu extrahieren. Der erste Algorithmus ist eine Support Vektor Maschine (SVM) welche zwischen den Klassen, *train* und *no train*, mit einer Präzision von 79.19% und einer Empfindlichkeit von 96.85%, klassifiziert. Der Einfluss verschiedener Features wird untersucht, wobei die angeführten Ergebnisse mit der Signalenergie als Feature erreicht wurden. Die Ergebnisse beziehen sich auf 1575000 Klassifikationen. Die SVM zeigt eine Querempfindlichkeit gegenüber Zügen auf benachbarten Gleistrassen. Der zweite Algorithmus ist ein Expertensystem und detektiert Züge über die Signalenergie. Mit Hilfe der Fourier Transformation kann zur Laufzeit eine Unterscheidung zwischen Zügen auf der benachbarten Gleistrasse und Zügen auf der untersuchten Gleistrasse getroffen werden. Mit diesem Algorithmus wurde eine Präzision von 86.38% und eine Empfindlichkeit von 97.67% für die beiden Klassen, *train* und *no train*, erzielt. Am Ende werden konkrete Verbesserungsvorschläge gemacht.

Contents

1 Introduction

The main role of railway operators is to ensure the safe and efficient running of trains through their rail network. Therefore, the tracking of trains is mandatory. It is needed to control the train traffic and ensure a safe transportation of large number of people and goods [1]. A fault in the train regulation can lead to an accident. This in turn can cost the lives of many passengers. Trains have a large mass and a very long braking distance due to the low static friction of steel on steel. This is why the rules for railway operations are very strict.
Because of the often large length of the trains it is important to know the beginning and the end of the train, which is referred to as train integrity in the document. The train traffic is designed in such a way that the distance between two trains is always so large that the following train can stop at any time without running into the front train. To ensure this the track is divided into blocks and only one train per block is allowed [2]. Railway operators want to get rid of these fixed long blocks and move to short dynamic blocks. Dynamic means here that the beginning and end of a block can be varied depending on the situation. This is needed to increase the workload of the train traffic. Therefore, precise train localization is required. Required parameters are:

- Position of the train
- Train integrity
- Speed of the train

The control centre needs these parameters to ensure a safe running of trains. Some already used systems for monitoring railway traffic are

- Global Positioning System (GPS)
- Balise and Odometry
- Axel Counter

1.1 Existing Systems

1.1.1 GPS

Global Positioning System (GPS) is a satellite-based positioning system [3]. In figure 1.1 a train receiving signals from three satellites can be seen. The position and speed of a train can be estimated from these signals. The calculated position and speed can then be sent from the train to the control centre by GSM-R which stands for "Global System for Mobile Communications – Rail(way)". It is a communication protocol for voice and data services in railway application [4].
From the needed parameters for tracking, position and speed can be provided by GPS with a sampling time up to 100 ms [5], if the satellite signal is available. In tunnel or thick forest no signal reaches the GPS receiver. Information about train integrity cannot be provided by GPS because trains are only equipped with one receiver on the locomotive. An installation of a second receiver on the tail of each train would be too expensive.

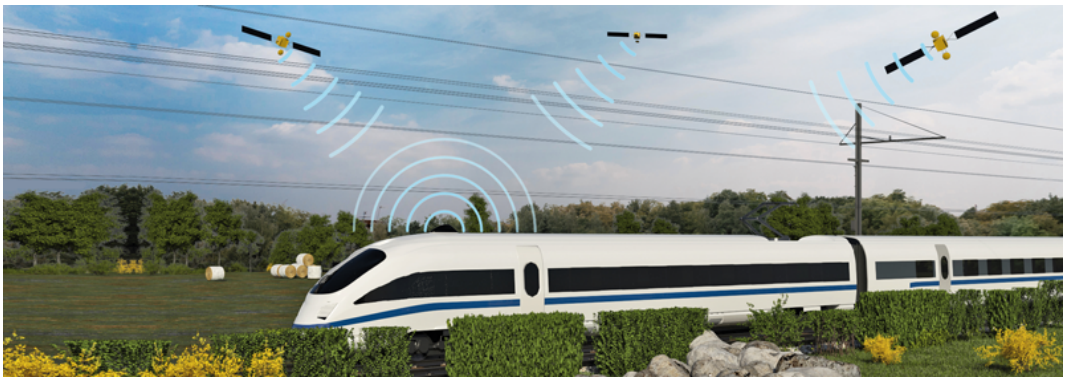


Figure 1.1: Global positioning system (GPS) to estimate the actual position and speed of the train [6].

1.1.2 Balise and Odometry

The other mentioned positioning system are balises [7] in combination with motion sensors counting the wheel rotations. Figure 1.2 shows a balise, which is placed between the rails. It is kind of an electronic beacon providing

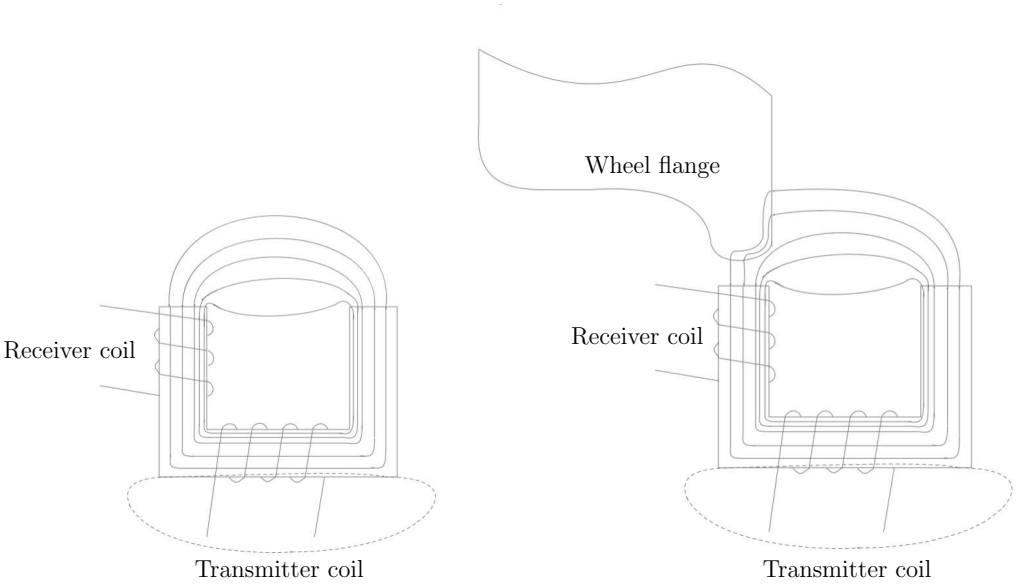
Figure 1.2: Balise system. The balise placed between the rails sends the actual position to the BTM passing over it [7].

the exact position to a system called Balise Transmission Module (BTM). The BTM is the upper yellow part in figure 1.2. It is installed under the train and receives the position when the train passes the balise. For the balise no fixed rate for position updates can be defined because the beacons are placed irregular in the railway network. In urban areas the update rate is usually higher than in rural areas.
Odometry uses data from motion sensors to estimate the position of the object to track [8]. In trains, wheel encoders measure the speed of the train to estimate the position until the next update by a balise.
Balise in combination with wheel encoders do not provide information about train integrity.
The train position and speed are sent to the control centre by GSM-R.

1.1.3 Axle Counter

None of the previous mentioned existing systems can provide information about train integrity. Therefore, axle counters are installed throughout the rail network. In figure 1.3a the installation of such a counter on the track can be seen and figure 1.3b shows the working principle of it [9]. The transmitter coil produces a magnetic flux trough the ferrite core. On the right picture a field deflection can be seen because a wheel flange changed the magnetic path. This results in a voltage change of the receiver coil compared to the case without a wheel on the left picture. The evaluation takes place in a superior system by increasing a counter. Always two of these sensors are coupled together to determine the train integrity. The sensors are usually installed at the beginning and end of a block. Both must count the same number of wheels otherwise it would mean that some type of error occurred. This could be an error of the sensor itself or a train could have lost a wagon in this block. For both types of error, a route inspection must be carried out, which leads to a stop of the train traffic for this track section.
With the known distance between the installation of the two coupled axle counters and the timestamps of the first wheel passing each sensor the speed of the train can be determined. The train position is not measurable with this system because the installation position can change when the sensors are dismounted due to maintenance work on the rails.
Unlike to the previous sensors, this one is not directly communicating with the train. The speed and train integrity are sent over the network cable to the control station.

(a)



(b) Magnetic flux lines without wheel on the left and with wheel on the right.

Figure 1.3: Installation of an axle counter in figure (a) [10] and the working principle of an axle counter in figure (b) [9].

1.2 Fibre Optic Sensing

To get more people to use the train for transportation, railway operators aim to increase the workload in their rail networks without compromising safety.
The rail network can be considered as divided into sections, which are often several hundred meters long and delimited by axle counters and balises. If one of these sections is occupied by a train, no further train may enter this section, which is often also called a block. To increase capacity utilisation, rail operators want to get rid of these long fixed blocks and move to short dynamic blocks.
To achieve this with the systems described above, their number in the

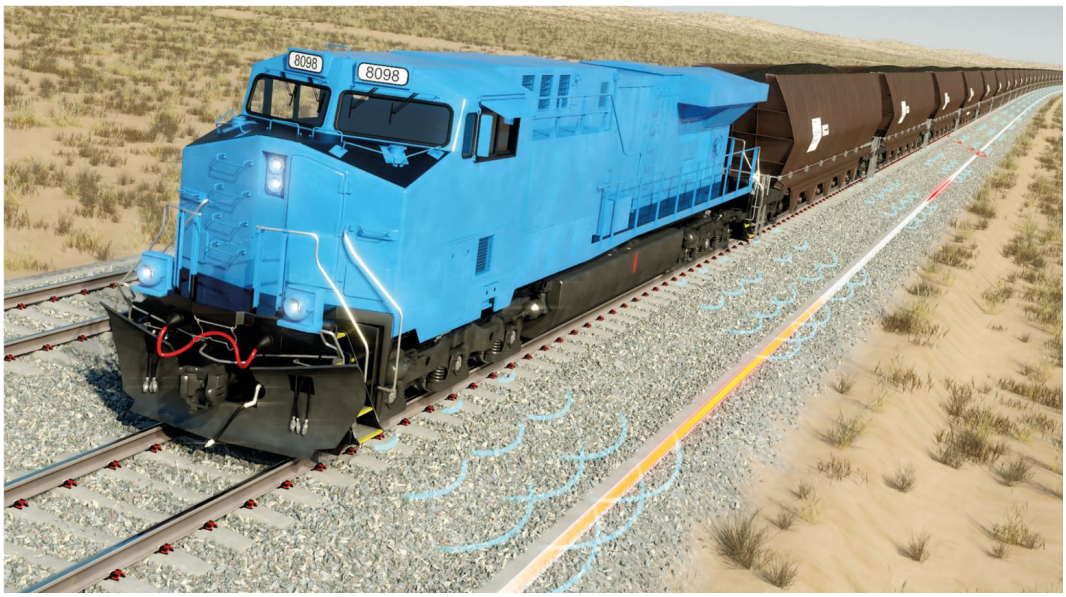


Figure 1.4: Vibrations generated by the train can be measured by a fibre optic cable buried near the track [11]. The sound waves travel from the rail's trough the soil to the cable.

rail network would have to be significantly increased. This would mean not only additional installation costs but also increasing maintenance costs. During maintenance work on axle counters or balises, train operation must be partially stopped. So, an increased number of this systems in the rail

network would lead to more maintenance works and so no increase in capacity utilisation is possible.
That is why railway operators' interest in new systems has increased over the last few years. In this case interest in distributed fibre optic sensing was shown. A fibre is installed in the vicinity of the measurement object and different quantities like strain, temperature, pressure or acoustic can be measured according to which physical effect is utilized [12].
In the oil and gas industry fibre optic sensing is already used for detecting failures or work on pipelines. For railway applications the safety requirements are much higher and so it is still in research in this field [11].
For railway applications the fibre is installed like shown in figure 1.4 in the cable shaft along the track. A decisive reason why this system is in the focus of railway operators is that fibre optic cables are already installed in the cable ducts.
Figure 1.5 shows the principle setup of a distributed fibre optic sensor. A light pulse generated by a laser is travelling through the fibre and is reflected at scatter sites. The reflected light, called backscatter, is travelling back to the detector where the intensity of it is measured as a function of time.
In this thesis a fibre optic sensing technique measuring acoustic waves acting on the fibre is used. It is called Distributed Acoustic Sensing (DAS). Sound waves acting on the fibre change the properties of backscatter.
By using the time difference between sending out the light pulse and

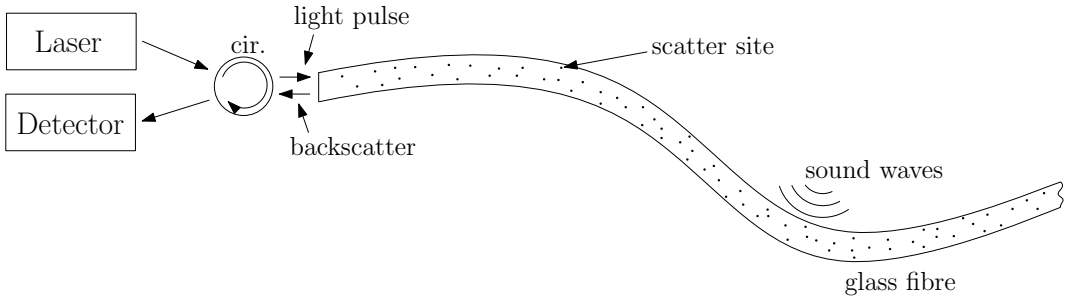


Figure 1.5: Principle setup of a distributed acoustic sensor.

receiving the backscatter, the location of the scatter site is calculated by the detector. This means that the position of the sound waves acting on the fibre can be determined. For every position, with a spatial resolution of 8 m, along the fibre a time evolution of the vibrations acting on the fibre

is measured. The positions are called channels in the further document. A more precise explanation of the measurement system is given in chapter 2.

The signal energy of the measured DAS data is shown in figure 1.6.. On the X-axis the time is plotted and on the Y-axis the distance along the fibre. Here it can be seen that for every position along the fibre a time evolution of the vibrations acting on the cable in this particular position is measured. The colour is proportional to the energy value. The blue areas correspond to low energy whereas the yellow areas correspond to high energy.
It can be seen that it is possible to measure trains along the fibre and from the movement along the fibre, the speed of the train can be calculated. Also, the train length can be determined from the channels occupied by one train at a certain time point. This is indicated with the black arrow for one train. So, in contrast to the existing systems for train localization, with DAS it is possible to get position, speed and train integrity.

One problem with DAS is, that it is not possible to get information about the lateral distance between vibration source and fibre. That is why also trains on neighbour rails are measured. One would expect a constant energy for a train driving with constant speed. But figure 1.6 shows that the energy randomly varies along the channels. The green rectangle exemplary shows channels where the signal to noise ratio (SNR) is low compared to adjacent channels. On the other hand, the system has channels in which unknown vibration sources interfere with the measurements. One is exemplary characterized in the figure by ”other disturbance”.

This work is now to further analyse the DAS data and to develop algorithms that can accurately detect trains. This means:

- Detection of head and tail of the train
  - The measurement quality changes along the channels which makes it important to implement the detection adaptively with respect to the channels.
- Distinguish between the two types of trains
  - Trains on the neighbour rail shall be distinguished from the trains on the test track.

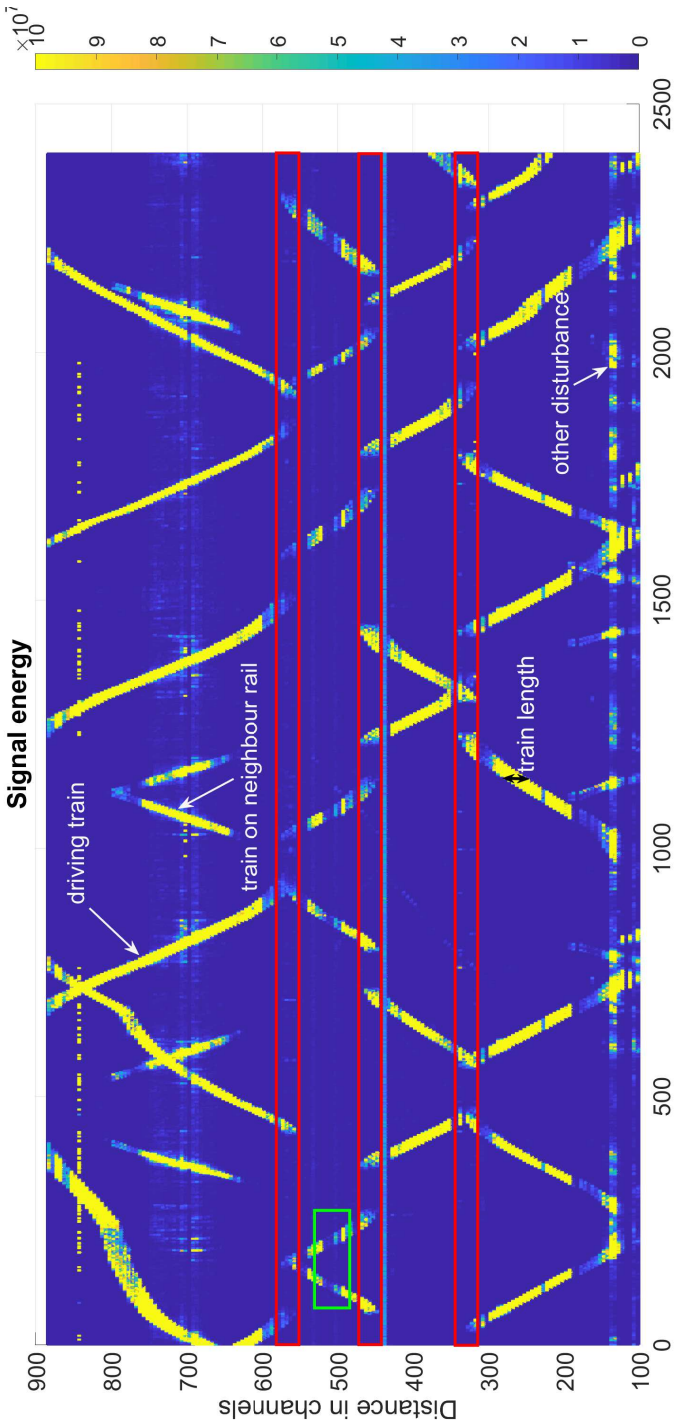


Figure 1.6: Signal energy plotted over distance and time. Red rectangles show the sections where the fibre passes a train station. Standing trains produce no vibrations. The area marked with the green rectangle shows a different SNR than the neighbouring channels. Different types of measured objects are indicated by white arrows. The train length is shown with the black arrow.

For each detected train position, speed and train integrity are the wanted parameters.
From the location of the measurement unit, containing laser and detector, the train parameters can be sent to the control centre or the measurement unit can even be placed directly in the control centre.
A big advantage of DAS is the possibility to measure long distances with a single measurement unit. With some special effort at the installation a measurement length of 175 km was reached by [13].

1.2.1 Test Track Germany

The DAS measurement system used in this thesis is installed in Frankfurt am Main in Germany. A Google maps view of the test track is shown in figure 1.7 with a length of 7.16 km. 6 km of the track are tunnel which makes

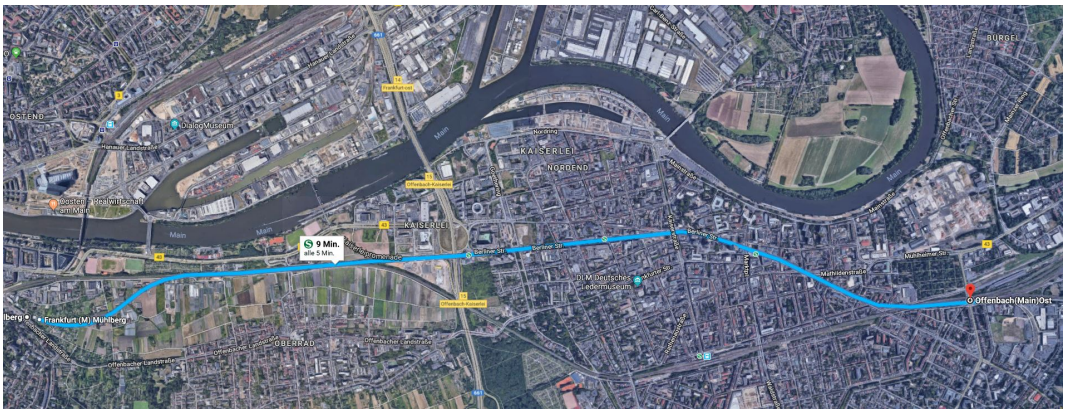


Figure 1.7: A Google Maps satellite picture of Frankfurt am Main with the marked test track for the DAS measurements.

GPS localization impossible. The part outside the tunnel runs parallel to a second track as can be seen in figure 1.8. In this part, the trains driving on the neighbour track can also be measured by the DAS system. These trains should be distinguished from the others with the presented algorithms in this thesis.

Figure 1.8: The part of the track outside the tunnel runs parallel to a second track. In this part also trains of the neighbour track can be measured by the system.

1.3 Approach Types for Tracking Tasks

As already mentioned before, it is a tracking task which has to be implemented. Trains are the targets to track in this thesis and the demanded dimensions of the trains are position, speed and train length.
There are two main ways to approach such a task:

- Model based or
- Signal processing in combination with an expert system.

The first one uses a model to describe the motion of a target to track. In the work of [14] a Kalman-Filter is used as the model to track targets for a radar system. More than one target can be detected at the same time and so each of them needs a separate model. Measurements are used to update the state of the target.

For this thesis the other approach was chosen. In the work of [15] Wavelet Transformation is used to extract wavelet coefficients as features. The features are the input for a supervised classification technique for the detection of myocardial infarction. This is a very similar task because the tracking of trains for a system like DAS is actually a detection problem. This can be seen as a classification between the two classes: *train* and *no-train*.
The dataset of the measurements consists of time signals for every channel. The fibre can be seen as divided into $8m$ long sections, which are the channels. The time signals for each channel are further divided into constantly large buffers and the signal analysis tools are applied on them to extract features. These features are the input for the expert system, which then

classifies the current channel and buffer into one of the two classes: 'no train' and 'train'. Figure 1.9 illustrates this process.
Two algorithms are implemented in this thesis as the expert systems.

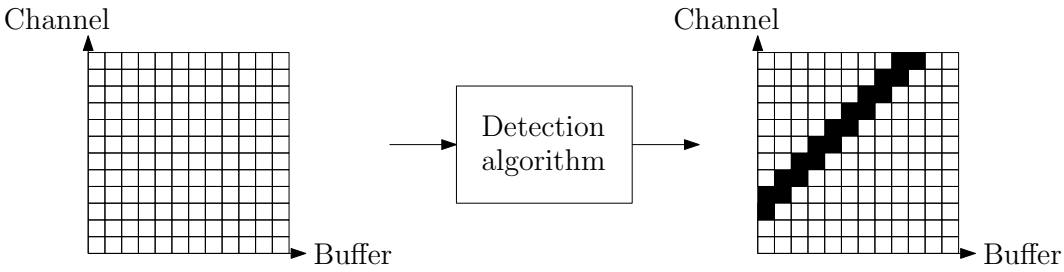


Figure 1.9: Structure for the detection algorithms which receives the features for every channel and buffer and classifies the buffer. White areas were classified as *no train* whereas black areas were classified as *train*.

- Support Vector Machine (SVM)
- Threshold and Pattern Algorithm

The SVM is a machine learning algorithm for supervised classification whereas the threshold and pattern algorithm is using patterns in the frequency spectrum produced by trains.
The most challenging part of the thesis is, that the algorithms must be implemented adaptive in respect to the measurement section because the measurement quality varies. There are some environmental impacts for that:

- Position of the cable duct (beside the track, on the walls of tunnels, ...)
- Cable slack (cable rolled up for maintenance reasons)
- Roads or highways passing near the track
- Echo effects in tunnels.

1.4 Outline

The thesis is structured as follows.
In chapter 2 the measurement system and some fundamentals needed for

the further discussion are described. The detection respectively classification of trains is done in three steps.

- Pre-processing of raw data
- Feature extraction
- Detection/Classification by expert system

The pre-processing is part of chapter 2. In section 2.3 it is described that the frequency spectrum of the measured DAS data contains a lot of train information. Therefore, the Fourier Transformation and Wavelet Transformation are used as analysis tools. As the third tool Empirical Mode Decomposition was chosen. Those transformations are explained in chapter 3.
For the detection two algorithms are implemented, which are discussed in chapter 4:

- Support Vector Machine (SVM)
- Threshold and Pattern Algorithm

These algorithms use the extracted features as an input for the train detection. The second algorithm even classifies between different types of trains. The thesis ends with a conclusion and outlook in chapter 5.

2 Measurement System

The principle setup of the measurement system was shown in figure 1.5 in the previous chapter. A coherent light pulse is injected into the fibre. As the electromagnetic wave travels through the fibre it is scattered into different directions by particles in the medium. A part of the scatter travels back to the detector as backscatter.
Physically, scattering can be divided into two categories, elastic scattering and inelastic scattering [16]. The former is a linear collision process where the photon's energy does not change, whereas the latter is a nonlinear collision process where the photon's energy does change. Elastic and inelastic scattering occur simultaneously in the medium, but with different strengths. Inelastic scattering can further be divided into Stokes and anti-Stokes scattering. Figure 2.1 illustrates the energy levels of scattering. The

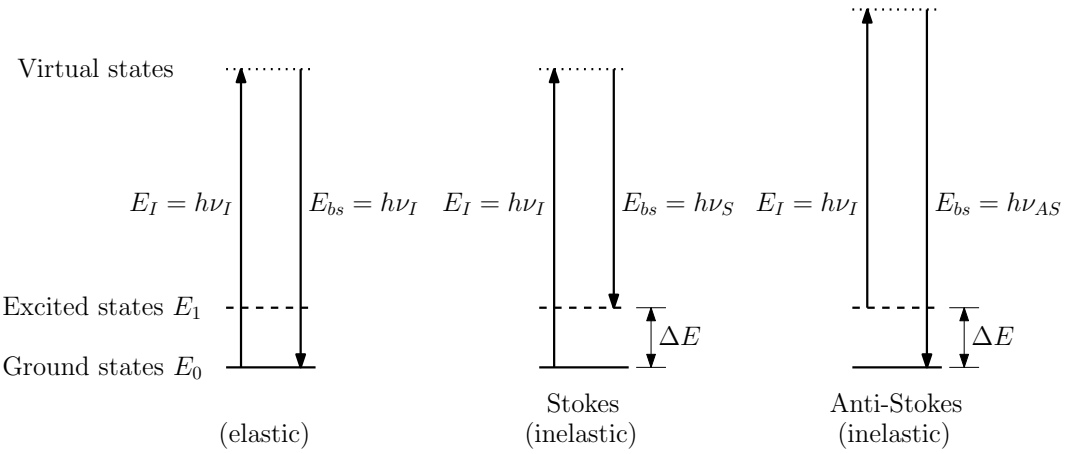


Figure 2.1: Energy level diagram of Rayleigh, Stokes and anti-Stokes scattering.

most molecules in the fibre occupy the ground state E_0 and some the excited state $E_1 = E_0 + \Delta E$.

When an incident photon with energy $E_I = hv_I$, where h is the Planck constant and v_I the frequency of the incident light, is absorbed by the molecule, it transits to a virtual state. Because this state is unstable a photon with Energy E_{bs} is emitted and the molecule transits back to the ground or excited state.
If the energy of the emitted photon is the same as the energy of the absorbed photon it is called elastic scattering and the wavelength of emitted and incident photon are the same according to $\nu_I = \frac{c}{\lambda_I}$, with c the speed of light and λ_I the wavelength of incident light.
Stokes scattering is when emitted photons have less energy than the incident photons which leads to a higher wavelength $\lambda_S = \frac{c}{\nu_S}$ of the backscattered light, respectively anti-Stokes scattering is when emitted photons have more energy than the incident photons and so the wavelength $\lambda_{AS} = \frac{c}{\nu_{AS}}$ of the backscattered light is lower than the one of incident light.
Three different types of backscatter processes can be used to extract the relevant information of the object to measure:

- Rayleigh scattering (elastic)
- Raman scattering (inelastic)
- Brillouin scattering (inelastic)

Their backscatter intensity is dependent on the wavelength, which can be seen in figure 2.2.
All three types occur simultaneously in the fibre but at the detector they are filtered by the wavelength to get the wanted one. The incident light is sent into the fibre with a wavelength of λ_I. Each scattering type is sensitive to different physical parameters.
The intensity of Raman scattering in the anti-Stokes band is dependent on the temperature T of the fibre which can be used to implement Distributed Temperature Sensors (DTS).
The wavelength of Brillouin scattering is dependent on strain ϵ and temperature T. This is used in the oil and gas industry for structural monitoring of pipelines.

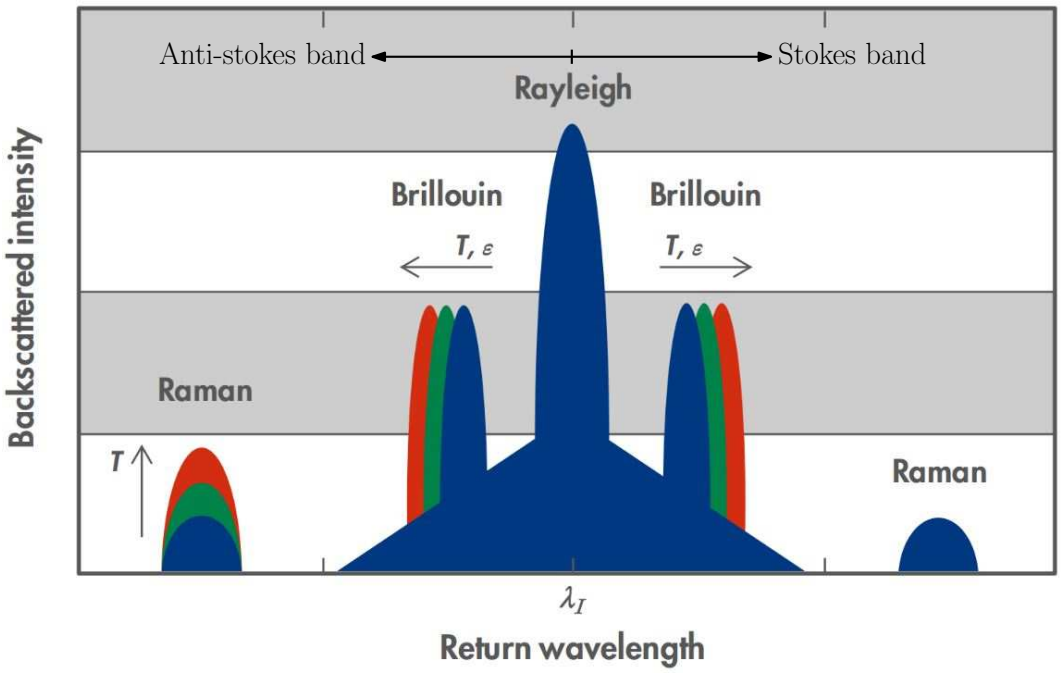


Figure 2.2: Typical scattering spectrum for the different types of scatter [12]. The different colours indicate how the change of temperature T and pressure ϵ affects the respective backscatter.

Rayleigh scattering occurs at the same wavelength as the incident light. The intensity and phase of the backscattered light is dependent on temperature and strain. Slow variations of temperature and background strain can be ignored and so it becomes dominantly dependent of acoustic strain. Distributed Acoustic Sensors (DAS) are based on the principle of Rayleigh Scattering. A phase-sensitive Optical Time Domain Reflectometry, $\phi - OTDR$, is a system which is used for DAS measurements.

2.1 Principle of ϕ-OTDR

In this thesis a dual pulse ϕ-OTDR [12] system with its interrogation scheme shown in figure 2.3 is used. Two highly coherent laser pulses with different

frequencies f_1 and f_2, a pulse width of t_p and delayed by t_G are send through a circulator (to distinguish between forward and backward travelling light) into the fibre. These pulses travel through the fibre and are backscattered at the scatter sites due to Rayleigh scattering.

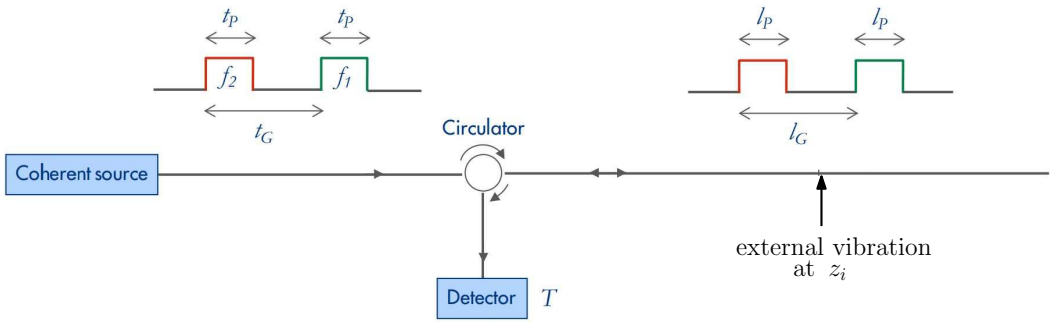


Figure 2.3: Interrogation scheme for a dual pulse ϕ-OTDR system for the use as a DAS [12].

For a better understanding we first consider one pulse travelling through the fibre. The received optical signal is defined by the electric field of the optical wave

$$E(t_a) = A(t_a)e^{j(\omega t_a + \varphi(t_a))} \tag{2.1}$$

that results from the probe pulse undergoing the backscattering process [17]. t_a is the time between launching the pulse and receiving the backscatter with amplitude $A(t_a)$ and phase $\varphi(t_a)$. The probe pulse was launched with the angular frequency ω. $A(t_a)$ and $\varphi(t_a)$ are the summation of all the amplitudes and phases of the scattering centres within the pulse length as illustrated in figure 2.4.

The distance z along the fibre where the backscatter occurred is defined by the time t_a as

$$z(t_a) = \frac{c \cdot t_a}{2n}. \tag{2.2}$$

c is the velocity of light in vacuum and n the refractive index of the fibre medium.

The detector samples the received signal as given in equation (2.1) with the

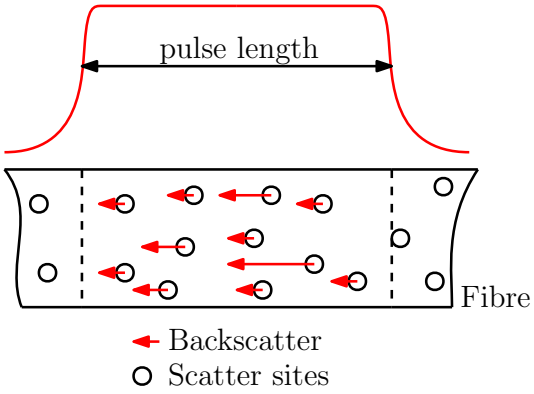


Figure 2.4: The backscattered light measured at the detector is the sum of the contributions of all the scatter sites within the pulse width.

sampling time T. This limits the spatial resolution of the measurements. The fibre can be seen as divided into sections of length

$$\Delta z = \frac{c \cdot T}{2n}. \tag{2.3}$$

These sections are referred to as channels in the thesis. So every sample of signal (2.1) refers to one channel of the fibre. The next launched pulse produces again a signal as given in equation (2.1) and this is then the next sample for every channel. The pulses have a repetition rate of f_s.
In summary, this means that the system measures a time signal with the sampling time $T_s = \frac{1}{f_s}$ for each channel along the cable. This can be written as

$$E(z_i, t) = A_i(t) e^{j(\omega t + \varphi_i(t))} \tag{2.4}$$

with z_i the position along the fibre in spacings of Δz, i the index corresponding to that position and t the time evolution, which is sampled with the sample time T_s. A_i and φ_i are now the amplitude and phase of the scattering sites for position z_i. Acoustic waves acting on the fibre change the amplitude and the phase of the affected scatter sites randomly. If one considers the addition of vectors with random amplitude and phase, the sum could be erased. To get detailed information about the acoustic waves acting on the fibre a dual pulse system has to be used.

So back to the dual pulse system where the backscatter of both pulses travels back to the detector and interfere. The resultant backscatter at the detector is the sum of two terms, one produced by pulse 1 (green pulse in figure 2.3) and the other produced by pulse 2. In the following derivation the index 1 always refers to pulse 1 and the index 2 to pulse 2. The backscatter intensity evolution I over time for location z_M can be expressed as:

$$I(z_i, t) =< E(z_i, t) \cdot E(z_i, t)^* > \qquad \text{with} \tag{2.5}$$

$$E(z_i, t) = E_0 \sum_{m=1}^{N_2} e_m \, e^{j(\varphi_m(t)+2\pi f_2 t)} + E_0 \sum_{m=N_G}^{N_G+N_1} e_m \, e^{j(\varphi_m(t)+2\pi f_1 t)} \tag{2.6}$$

according to the work of [18]. Here E_0 is the initial amplitude of the probe pulse, N_2 and N_1 are the number of scattering centres within the pulse length l_p of the second and first pulse, N_G is the number of scattering centres within the gauge length l_G, e_m and φ_m are the stochastic amplitude and phase of scattering centre m, f_j is the frequency of j-th pulse and t is the time evolution. This all corresponds to position z_i.

Consider now that an external vibration acts on the fibre somewhere between the two pulses as shown in figure 2.3. This would change the backscatter generated by pulse 1, since the region of external vibrations was passed by the first pulse whereas the scattering of pulse 2 remains unchanged.
The work of [19] shows that external vibrations acting on the fibre change the optical path l_o of the entire disturbing range by physical elongation and by changing the refractive index of the fibre. This phase, defined by Θ_i, is an accumulation of all the phases in the entire range. The relation between phase and optical path is given by

$$\Theta_i = \frac{2\pi}{\lambda_I} l_o. \tag{2.7}$$

Since the optical wavelength of the incident light λ_I is very small, a small change in the optical path results in a large phase change. The information of the vibration is thus in the phase of the backscattering signal.

Taking into account the external vibration between the pulses and doing some algebra, the electric field from equation (2.6) can be written as:

$$E(z_i, t) = A_{i2}(t)e^{j(2\pi f_2 t + \Psi_{i2}(t))} + A_{i1}(t)e^{j(2\pi f_1 t + \Psi_{i1}(t) + \Theta_i(t))}. \tag{2.8}$$

The backscatter intensity I from equation (2.5) is

$$I(z_i, t) = A_{i1}(t)^2 + A_{i2}(t)^2 + 2A_{i1}(t)A_{i2}(t)cos(2\pi\Delta f t + \Theta_i(t) + \Psi_{i1}(t) - \Psi_{i2}(t)) \tag{2.9}$$

with $\Delta f = f_1 - f_2$. The four quantities $A_{i1}(t)$, $A_{i2}(t)$, $\Psi_{i1}(t)$ and $\Psi_{i2}(t)$ are a result of the specific scattering profile in the absence of external vibration. Their values are stochastic and hard to determine. The external vibration changed the phase of the backscatter produced by pulse 1 and is taken into account in equation (2.9) by $\Theta_i(t)$. i refers to the channel number, which is the position along the fibre.
The backscatter intensity $I(z_i, t)$ is the quantity measured by the detector. Now the vibration information is in the phase $\Theta_i(t)$. [18] proposed a phase demodulation method to determine an estimation $\Theta_i'(t)$ from equation (2.9). The result of the method is

$$\Theta_i'(t) = \Theta_i(t) + \frac{\Psi_{i1}(t) + \Psi_{i2}(t)}{2}. \tag{2.10}$$

$\Theta_i'(t)$ is the time evolution of the phase corresponding to the position z_i. It holds the information of the vibrations acting on the fibre.
The measurements in this thesis are provided by the company *OptaSense*. They use an measurement unit as depicted in figure 2.3 with a post processing after the detector to provide the phase $\Theta_i'(t)$.

Summary of the measurement system:

- Fibre length: 7.16 km
- Spatial resolution: 8 m
  - Number of channels: 895
- Sampling time within a channel: $T_s = 0.4$ ms
- Measured Quantity: Phase $\Theta_i'(t)$ separately for every channel
  - i is the channel number
  - The unit of the quantity is radian but with an unknown gain.

2.2 Pre-processing of Raw Signal

The whole dataset provided consists of 895 channels, which corresponds to a cable length of 7.16 km. For every channel several hours of data was measured with a sampling time of $T_s = 0.4$ ms. In figure 2.5 the raw data for channel 255, which corresponds to position $z = 2.04$ m along the fibre is shown. Three trains are measured here. The provided quantity by the measurement system is an estimation of the phase of backscatter $\Theta_i(t)$ from equation (2.9), which changes when external vibrations act on the fibre.
The second term on the right hand side of equation (2.10) contains the

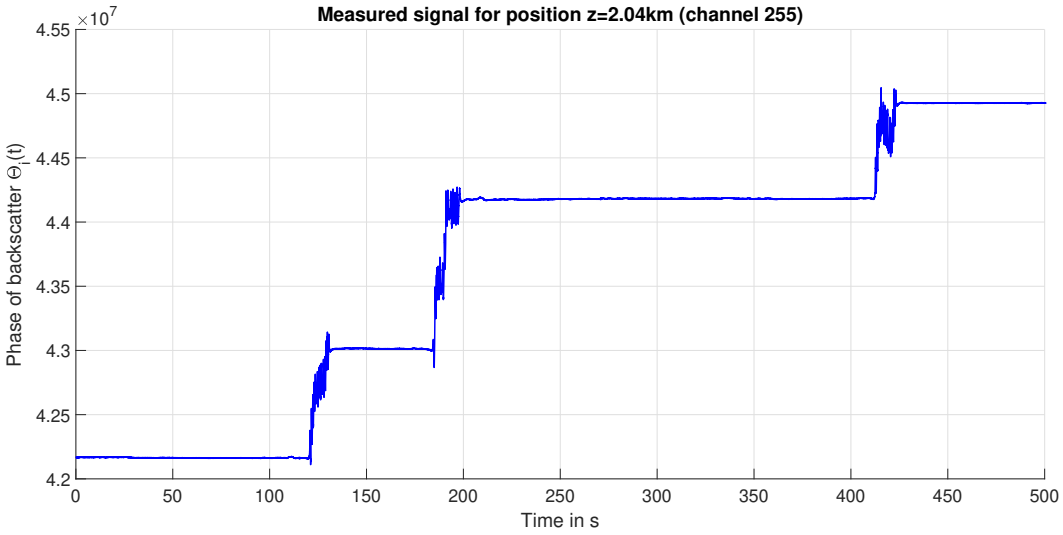


Figure 2.5: Measured phase of backscatter in a distance of 2.04 km from the fibre beginning showing signal jumps. Jumps do not contain vibration information so they are filtered out.

summation of the phases of all scattering sites within the two pulses in absence of external vibrations. In ideal these two values can be seen as constant but in the works of [18] and [20] it is mentioned that in real these parameters can change due to environment temperature fluctuations, the inevitable laser frequency drift and other reasons. The change of the parameters $\Psi_{i1}(t)$ and $\Psi_{i2}(t)$ thus describes the jumps of the phase seen in figure 2.5.
The information of the external vibration is contained in $\Theta_i(t)$ in equation

(2.10). Thus, the jumps do not contain any information and must be filtered out.
Therefore the measured signal $\Theta_i'(t)$ for every channel i is divided into buffers of 1250 samples what corresponds to 0.5 s due to the sampling time $T_s = 0.4$ ms. To filter out the jumps the mean of each buffer is subtracted from the buffer. This leads to the signal shown in figure 2.6.

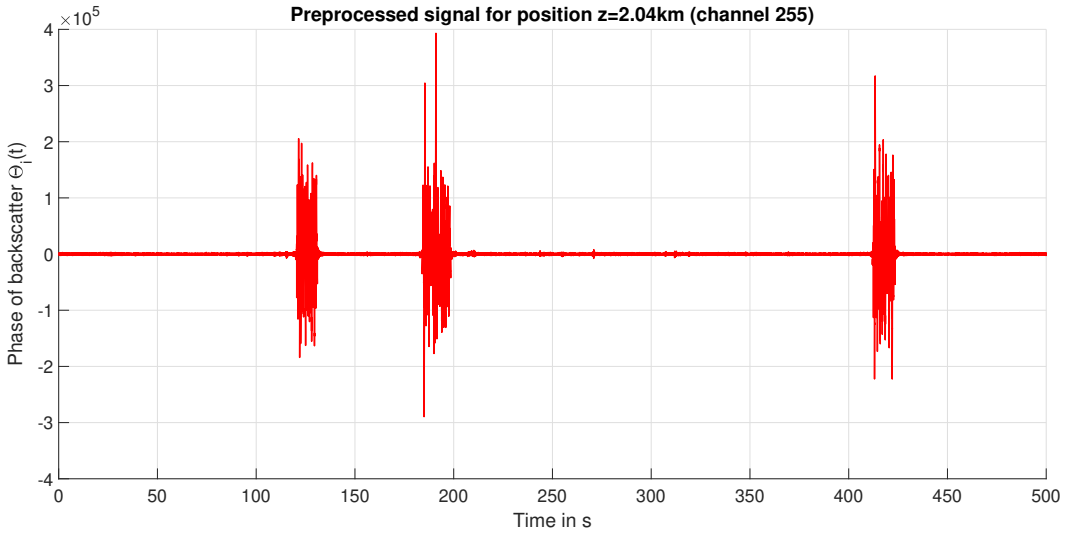


Figure 2.6: Pre-processed phase of backscatter in a distance of 2.04 km from the fibre beginning.

2.3 Dominant Frequencies of a Trainload

In the previous section it has been addressed how vibrations can be measured by using the Rayleigh scatter effect. The following section is about how these vibrations due to trains arise.
The passage of a train consists of a number of similar events, each with individual delay times which are dependent of the train dimensions. On the test track only trains of the production class DB 423/433 and class DB 430 move. In figure 2.7 such a train can be seen.
In chapter 1 the problem of trains on nearby rails was stated. It is known

Figure 2.7: Deutsche Bahn (DB) train of the class 423/433 driving on the test track [21]

that on these rails an ICE3 moves. So, in the dataset for this thesis three possible trains can occur. The dimensions of them are listed in table 2.1.
For the trains on the track under test the dimensions of the first two types in the table refer to figure 2.8. For the ICE3 the dimensions refer to figure 2.9. For the chosen data-based approach it is important to understand the

Type	L_{a1}	L_{a2}	L_{b1}	L_{b2}	L_{train}
-	m	m	m	m	m
DB 423/433	2.2	2.7	14.89	15.46	67.4
DB 430	2.2	2.7	14.89	14.894	68.3

Type	L_a	L_b	L_{c1}	L_{c2}	L_{train}
-	m	m	m	m	m
ICE 3	2.5	17.375	25.835	24.775	200.84

Table 2.1: Train dimensions for all three trains present in the measurements. First two trains are the trains on the track which is measured and the dimensions are referred to figure 2.8. The third train is the train on nearby rails with dimensions referred to figure 2.9.

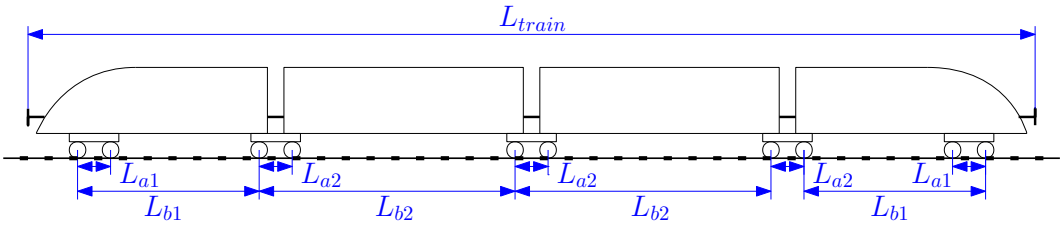


Figure 2.8: Important geometrical train parameters for the trains driving on the rail under test. This refers to trains of type DB 423/433 and DB 430 from table 2.1.

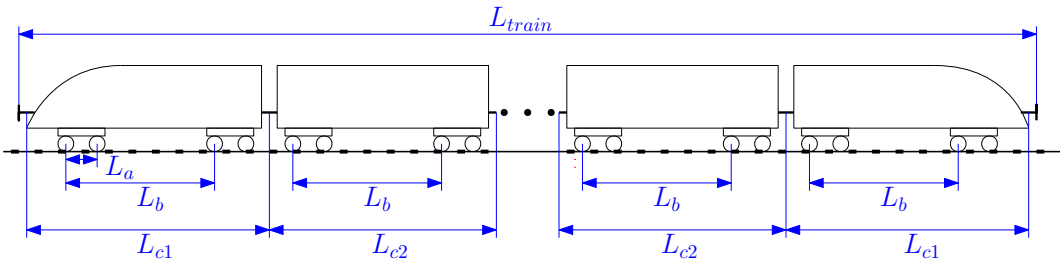


Figure 2.9: Important geometrical train parameters for the train (ICE3) driving on nearby rails. This refers to trains of type ICE 3 from table 2.1.

vibrations a moving train on rails generates so that the right signal analysis tools can be applied on the measured signal.
The following derivation is only done for class DB 423/433 but is analogous to the other trains.

The work of [22] starts the derivation by considering a single wheelset moving with the speed v, its effect can be represented as

$$P(t) = P_{wheelset}\delta(t - t_k) \tag{2.11}$$

where $P_{wheelset}$ is the nominal loading of a wheelset, which is considered constant and often given in kN. $\delta(t)$ is the Dirac function and $t_k = \frac{z_k}{v}$ with z_k the position of the wheelset along the track. The next load is delayed by a distance of L_{a1} (see figure 2.8) and is added to equation (2.11)

$$P(t) = P_{wheelset}\left[\delta(t - t_k) + \delta(t - t_k - \frac{L_{a1}}{v})\right]. \tag{2.12}$$

This has to be continued until all the wheelsets are covered. A train of the DB class consists of four wagons with two different lengths. The whole train load for a train of the class DB 423/430 is given by equation (2.13).

$$\begin{aligned} P(t) = P_{wheelset}\Big[&\delta(t - t_k) + \delta(t - t_k - \frac{L_{a1}}{v}) + \delta(t - t_k - \frac{L_{b1}}{v}) \\ &+ \delta(t - t_k - \frac{L_{b1} + L_{a2}}{v}) + \delta(t - t_k - \frac{L_{b1} + L_{b2}}{v}) \\ &+ \delta(t - t_k - \frac{L_{b1} + L_{b2} + L_{a2}}{v}) + \delta(t - t_k - \frac{L_{b1} + 2 \cdot L_{b2}}{v}) \\ &+ \delta(t - t_k - \frac{L_{b1} + 2 \cdot L_{b2} + L_{a2}}{v}) + \delta(t - t_k - \frac{2 \cdot L_{b1} + 2 \cdot L_{b2} + L_{a2} - L_{a1}}{v}) \\ &+ \delta(t - t_k - \frac{2 \cdot L_{b1} + 2 \cdot L_{b2} + L_{a2}}{v})\Big]. \end{aligned} \tag{2.13}$$

The Fourier transform of it is

$$\begin{aligned} P(f) &= \int_{-\infty}^{+\infty} P(t) e^{-j2\pi f t} dt \\ &= P_{wheelset}\, e^{-j2\pi f t_k} \big(1 + e^{-j2\pi f (L_{a1}/v)} + e^{-j2\pi f (L_{b1}/v)} + e^{-j2\pi f ((L_{b1}+L_{a2})/v)} \\ &+ e^{-j2\pi f ((L_{b1}+L_{b2})/v)} + e^{-j2\pi f ((L_{b1}+L_{b2}+L_{a2})/v)} + e^{-j2\pi f ((L_{b1}+2\cdot L_{b2})/v)} \\ &+ e^{-j2\pi f ((L_{b1}+2\cdot L_{b2}+L_{a2})/v)} + e^{-j2\pi f ((2\cdot L_{b1}+2\cdot L_{b2}+L_{a2}-L_{a1})/v)} \\ &+ e^{-j2\pi f ((2\cdot L_{b1}+2\cdot L_{b2}+L_{a2})/v)}\big). \end{aligned} \tag{2.14}$$

Trains as shown in figure 2.8 can be coupled together. When considering a number of $N_c > 1$ trains coupled together equation (2.14) becomes

$$\begin{aligned} P(f) &= P_{wheelset}\, e^{-j2\pi f t_k} \big(1 + e^{-j2\pi f (L_{a1}/v)} + e^{-j2\pi f (L_{b1}/v)} + e^{-j2\pi f ((L_{b1}+L_{a2})/v)} \\ &+ e^{-j2\pi f ((L_{b1}+L_{b2})/v)} + e^{-j2\pi f ((L_{b1}+L_{b2}+L_{a2})/v)} + e^{-j2\pi f ((L_{b1}+2\cdot L_{b2})/v)} \\ &+ e^{-j2\pi f ((L_{b1}+2\cdot L_{b2}+L_{a2})/v)} + e^{-j2\pi f ((2\cdot L_{b1}+2\cdot L_{b2}+L_{a2}-L_{a1})/v)} \\ &+ e^{-j2\pi f ((2\cdot L_{b1}+2\cdot L_{b2}+L_{a2})/v)}\big) \big(1 + \sum_{n=1}^{N_c-1} e^{-j2\pi f n (L_{train}/v)}\big). \end{aligned} \tag{2.15}$$

Figure 2.10 shows the amplitude spectrum of excitation by a train of type DB 423/433 moving with a speed of $v = 20\,\mathrm{m\,s^{-1}}$ as given in equation (2.15). It shows the ideal frequency spectrum of the effect of a train on the track. This is what is wanted and should actually be measured by the DAS system. Unfortunately, the fibre optic cable is not directly attached to the rail and therefore the sound waves have to be transmitted via the soil to the fibre. What is not modelled here is the influence of the soil on the wave propagation. In the work of [23] it was shown that the magnitude of vibration decreases proportionally with the frequency. No information about the composition of soil and track was given for the test track so this damping can not be modelled here. In section 3.1.1 the ideal frequency spectrum will be compared to the calculated one from the measurements.

With the same procedure as above the train load for a train of the class ICE3 can be calculated. The ICE3 consists of eight wagons and so the train load

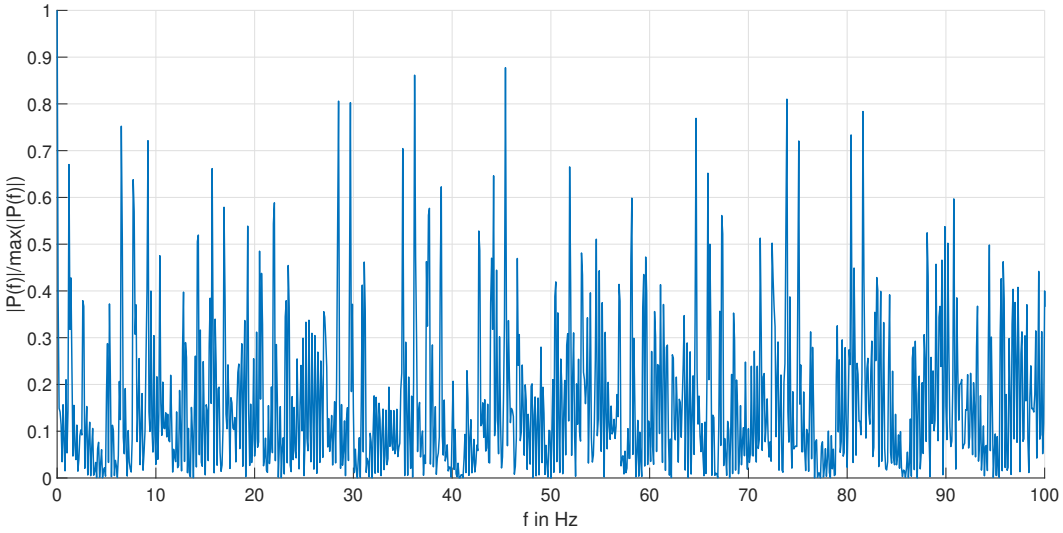


Figure 2.10: Amplitude of the theoretical frequency spectrum P(f) from equation (2.15) produced by a train of type DB 423/433 moving with a speed $v = 20\,\mathrm{m\,s^{-1}}$.

is:

$$
\begin{aligned}
P(f) = P_{wheelset}\, e^{-j2\pi f t_k} \big(1 + e^{-j2\pi f (L_a/v)} + e^{-j2\pi f (L_b/v)} \\
+ e^{-j2\pi f ((L_a+L_b)/v)}\big) \cdot \big(1 + e^{-j2\pi f ((L_{c1})/v)} + e^{-j2\pi f ((L_{c1}+L_{c2})/v)} \\
+ e^{-j2\pi f ((L_{c1}+2\cdot L_{c2})/v)} + e^{-j2\pi f ((L_{c1}+3\cdot L_{c2})/v)} + e^{-j2\pi f ((L_{c1}+4\cdot L_{c2})/v)} \\
+ e^{-j2\pi f ((L_{c1}+5\cdot L_{c2})/v)} + + e^{-j2\pi f ((L_{c1}+6\cdot L_{c2})/v)}\big) \big(1 + \sum_{n=1}^{N_c-1} e^{-j2\pi f n (L_{train}/v)}\big).
\end{aligned}
\tag{2.16}
$$

In figure 2.11 from [22] the typical frequency ranges for excitations by the train dimensions are shown in blue. A moving train also produces contributions in the higher frequency range which are dependent on the track or ground characteristics. In figure 2.11 their frequency range is marked in red and green. Unfortunately, for this thesis there is no information about the composition of track or soil provided. So the focus will lie on the vibrations generated by the train dimensions in the low frequency ranges, which are marked in blue in figure 2.11.

Also other works like [24] and [25] use the vibration characteristic induced

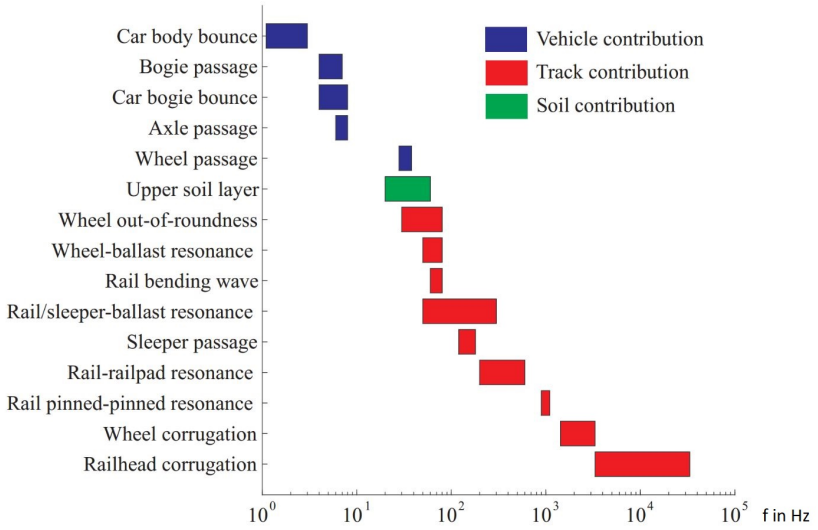


Figure 2.11: Main contribution of dynamic vehicle/track and soil interactions in the frequency spectrum [22].

by the trainload in the frequency domain for their tasks. Because of the large amount of information in the frequency range of the measured signal the Fourier Transform and the Wavelet Transform where chosen as analysis tools. A third tool, called Empirical Mode Decomposition (EMD), which is often used in classification tasks together with the Fourier and Wavelet Transform like in [26], is also used. It is a method for analysing nonlinear and non-stationary data [27].

3 Feature Extraction

The following chapter deals with the feature extraction. In section 2.2 it was mentiond that the measurements for each channel are divided into buffers. The transformations discussed in this chapter are applied on these buffers. This is called block transform because of applying the transformation on short blocks of an actually long signal [28].
The buffer length was chosen to be $B = 1250$ samples. With the sampling time $T_s = 0.4\,\text{ms}$ the buffer corresponds to $0.5\,\text{s}$. In the following the buffer is denoted by $x[m]$ with $m = 0, 1, ..., B-1$.

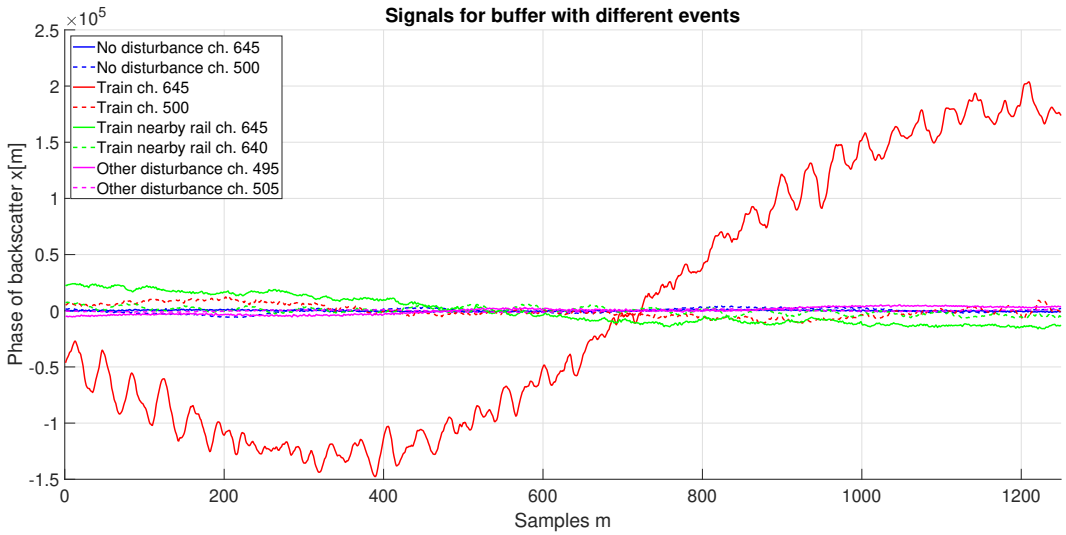


Figure 3.1: Pre-processed phase of backscatter in different channels at the same time. The section corresponds to a buffer. Solid and dashed lines in the same colour always show the same event but in a different channel. It can be seen that the measurements differ along the track.

In figure 3.1 the signal $x[m]$ for different buffers is shown. These buffers show different events occurring along the fibre. The events shown are:

- No disturbance: no vibrations
- Train: vibrations produced by the trains to detect
- Train nearby rail: vibrations produced by the trains on nearby rails
- Other disturbances: vibrations produced by other sources
  - Other vehicles driving in the surrounding of the fibre (i.e. above the tunnel).
  - Construction areas.

In chapter 1 it was shown that the energy value varies along the cable due to environmental impacts. This can also be seen in this plot. The solid lines of one colour always show the same event as the dashed line in the same colour but in a different channel. Here it can be again seen that the environment of the cable impacts the quality of the measurements.
What can be seen is that it is not possible to distinguish the events from the time signals shown in figure 3.1, except for the one plotted with a solid red line where the measured phase of the backscatter is high enough. Note that the phase of the backscatter is the quantity, which contains the information of the vibrations acting on the fibre.

The idea is now to use the analysis tools in this chapter to calculate features for every buffer with whom it is possible to distinguish between the events. In chapter 4 these features are going to be used as an input for the expert systems.

In this chapter all the analysis tools are exemplary applied to the signals shown in figure 3.1.

3.1 Fourier Transformation

The Discrete Fourier-Transformation (DFT) transforms a finite signal into spectral coefficients. She is defined as

$$X[k] = \sum_{m=0}^{B-1} x[m] e^{-j2\pi mk/B} \qquad k = 0, 1, ..., B-1. \tag{3.1}$$

where $x[m]$ is here the signal of one buffer with length B. k is the number of the spectral component. A sampling time of $T_s = 0.4\,\text{ms}$ and a buffer length of $B = 1250$ samples results in a frequency resolution of $\Delta f = 2\,\text{Hz}$. With the DFT it is possible to extract the frequency components present in the signal. More information about the DFT can be looked up in [28].
In MATLAB the DFT is calculated with the function $fft()$ which uses a Fast Fourier Transform algorithm. For the buffers with the different events shown at the beginning of this chapter their DFT is shown in figure 3.2. Compared to the ideal frequency spectrum shown in figure 2.10 the impact of the environment on the wave propagation can be seen. The impact is a combination of the frequency proportional damping property of the soil and the contributions of the track shown in red in figure 2.11. This impacts where not modelled due to lack of information.
For a better representation the red solid line is not shown fully, but the exact amplitude is not important here either. More important are the frequencies to the peaks.
Figure 3.3 shows two frequency ranges in more detail. In both of them it can be seen that the amplitude spectrum of signals from trains (red and green lines) is higher compared to the other signals (blue and magenta lines), even for the measurements in dashed lines, which were referred to lower SNR, a difference is noticeable. The red and green amplitude spectrum in the frequency range of the upper plot of figure 3.3 can be related to the contributions by the train dimensions according to figure 2.11, while the amplitude spectrum of the frequency range in the lower picture can be related to the contributions of the track.

This shows that trains can be distinguished from other disturbances and noise by the DFT. It is also possible to distinguish between the different train types because the amplitude spectrum in figure 3.3 is dependent of the

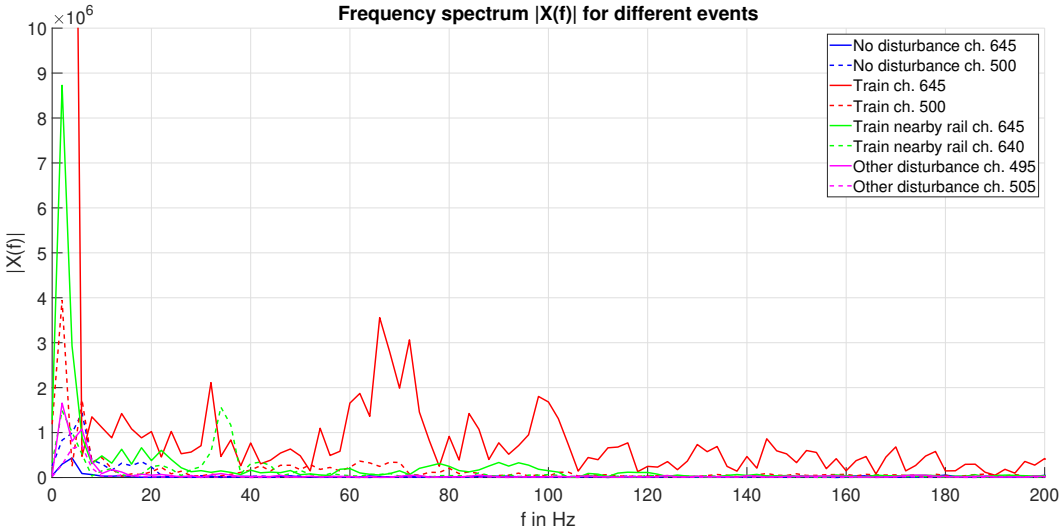


Figure 3.2: Single-Sided amplitude spectrum of x[m] for different events.

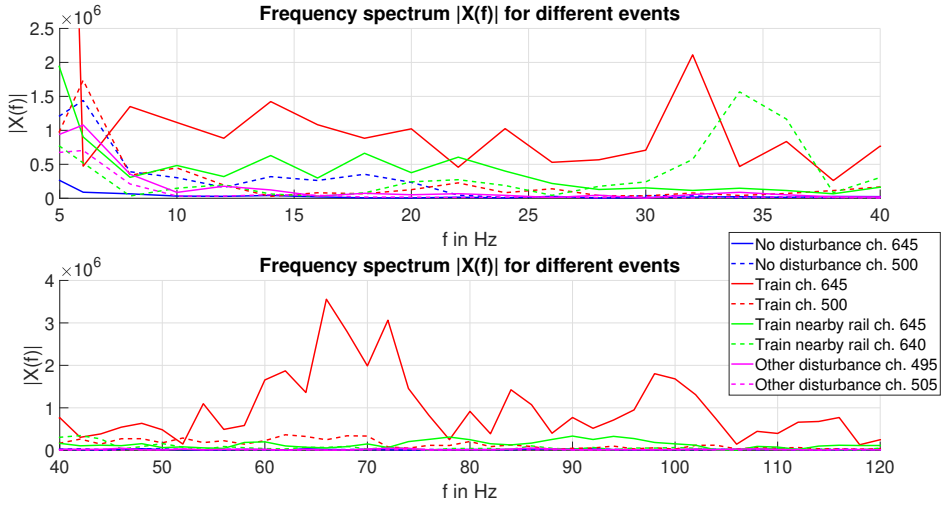


Figure 3.3: Single-Sided amplitude spectrum of x[m] for two specific frequency ranges. DFT of trains show the expected behaviour according to figure 2.11.

train dimensions and the train speed. This is further discussed in the next subsection because this information will be needed for the second algorithm in chapter 4 to distinguish between train types.

3.1.1 Dominant Frequencies in the Measurements

In chapter 2 the theoretical frequency spectrum of a moving train was presented. The signal shown in figure 3.4 is the phase of backscatter for a whole train of type DB423/433 passing channel 650. This signal contains of 24 buffers to cover a whole train passing channel 650.
To calculate the appropriate theoretical frequency spectrum by equation

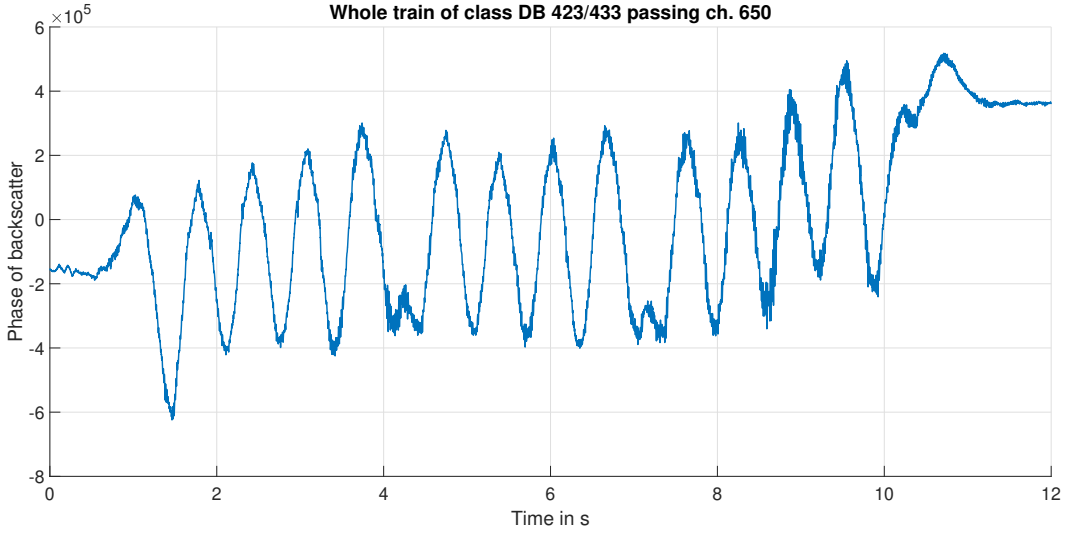


Figure 3.4: Phase of the backscatter in channel 650 produced by a passing train of class DB 423/433.

(2.15), the speed of the train is needed. The velocity of this train can be determined from the movement of the train along the fibre. The signal energy of the measurements for channels $590 - 670$ is shown in figure 3.5 for the length of 2000 buffers. The trains in this plot move nearly with a constant velocity. The head and tail of the train corresponding to figure 3.4 are marked with the red lines. The red lines were calculated by setting a threshold value for the energy. This was done for the channels $620 - 670$

because the train slows down a bit in the lower part of the plot. Note here that the determined head and tail are not correctly determined but that does not affect the speed estimation.

When taking the mean of both slopes the speed can be estimated and

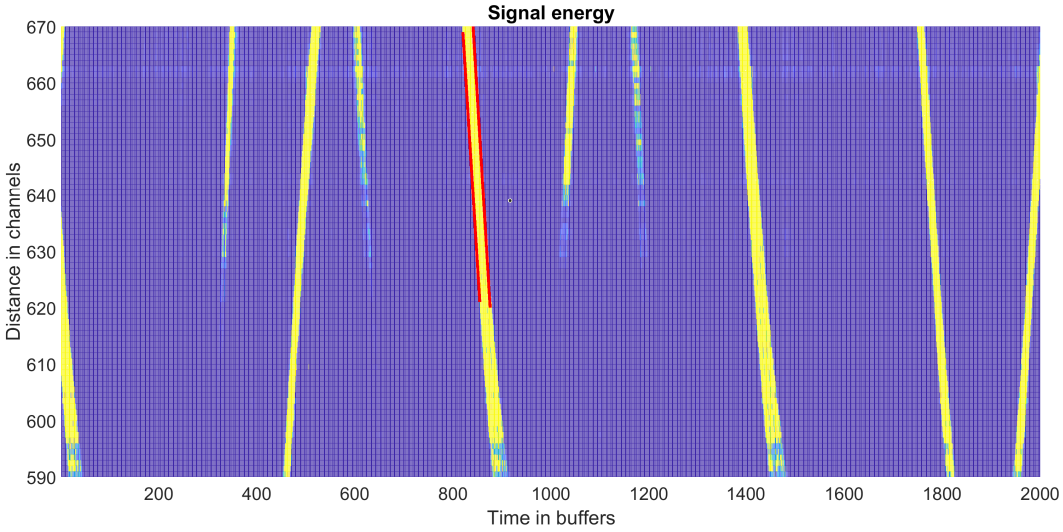


Figure 3.5: Signal energy for a small part of the fibre (channel 590 to 670). The signal in figure 3.4 is taken from channel 650 for the train marked with the red lines. The beginning and end of the train are marked with those lines. By calculating the slope of this lines, the train speed was estimated.

results in $v = 23.68\,\mathrm{m\,s^{-1}}$.
By inserting this value in equation (2.15) with the number $N_c = 3$ of trains coupled together, which is known from the reference provided to the measurements, the blue spectrum in figure 3.6 can be calculated.

From the signal in figure 3.4 the DFT of the measurements is calculated and added to figure 3.6 as the red curves. Both curves are normalized by the maximum value in the depicted frequency band. Compared to figure 3.3 the frequency resolution is better here because the number of samples for the calculation of the DFT is higher. This allows a better analysis in the low frequency range. But the buffer length of 0.5 s cannot be increased because this would decrease the update rate of position information. For the second algorithm in chapter 4 buffers are stored when a train is detected to get more precise information with every buffer.

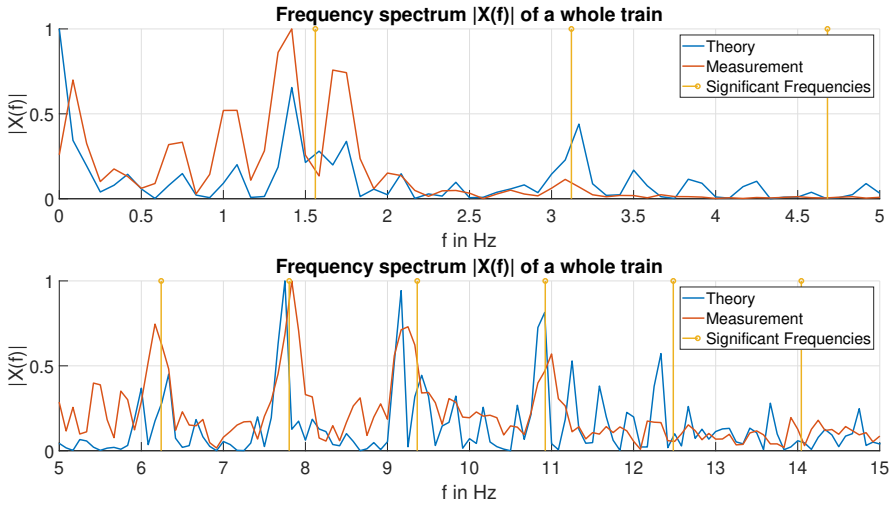


Figure 3.6: Theoretical frequency spectrum for the train type DB 423/433 as given in equation (2.15) compared to the Fourier Transformation of signal in figure 3.4. The yellow lines show the significant frequency components referred to the wagon length.

It can be seen that the position of the high peaks on the frequency axis match between theory and measurement. Equation (2.15) contains the train dimensions given in table 2.1. The works of [25] and [29] show that the most significant frequencies in the vibrations produced by trains correspond to the wagon lengths. This means that for a train of class DB 423/433 the highest peaks occur at frequencies which refer to the lengths L_{b1} and L_{b2}. The frequencies produced by the wagon length can be calculated by $f = \frac{n \cdot v}{L_x}$ with $L_x = \frac{L_{b1} + L_{b2}}{2}$ the mean of both lengths and $n \in \mathbb{N}$ is to take all the harmonics into account. In figure 3.6 the frequency components for $n = 1, 2, 3, 4, 5, 6, 7, 8, 9$ are delineated with the yellow lines. The mean of the two wagon lengths was taken because they just differ slightly and so the frequency difference, when plotting the components for both lengths, would lie within the frequency resolution. In addition, the train always consists of the same number of wagons of both lengths.
It can be seen that some of the frequencies produced by the wagon length fit to the highest peaks of the spectrum whereby a small error is acceptable

because the speed was estimated. The frequency component for $n = 1$ does not fit to the peak. This means that this peak is somehow suppressed by differently induced contributions in the frequency spectrum. That is consistent with the work of [25] and [29] which shows that not every harmonic produced by the wagon length is present.

For the ICE3 the same procedure can be done. Therefore a whole train was extracted from the data and can be seen in figure 3.7. The corresponding theoretical DFT calculated with equation (2.16) and the determined DFT from the measured signal are depicted in figure 3.8. The frequency com-

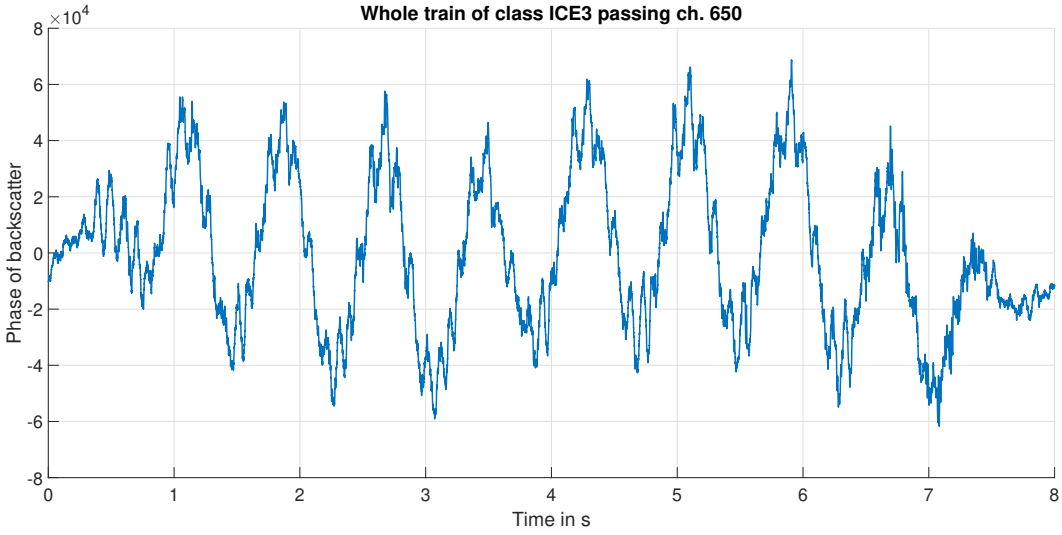


Figure 3.7: Phase of backscatter in channel 650 produced by a passing train of class ICE3.

ponents referred to the wagon length for the ICE3 can be calculated with $L_x = L_{c2}$. Here L_{c2} is taken because the whole train consists of six wagon with that length and just two wagon with length L_{c1}. The blue lines show the theoretical Fourier Transformation for a speed of $v = 30.95\,\mathrm{m\,s^{-1}}$. The red lines show the calculated Fourier Transformation from the measurements depicted in figure 3.7. The frequency components of the wagon length are shown in yellow for $n = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11$. Again, some of them fit to the highest peaks of the spectrum. The works of [25] and [29] also show that not every harmonic component has to be present in the amplitude spectrum. Some can be cancelled out by the generated frequencies of other

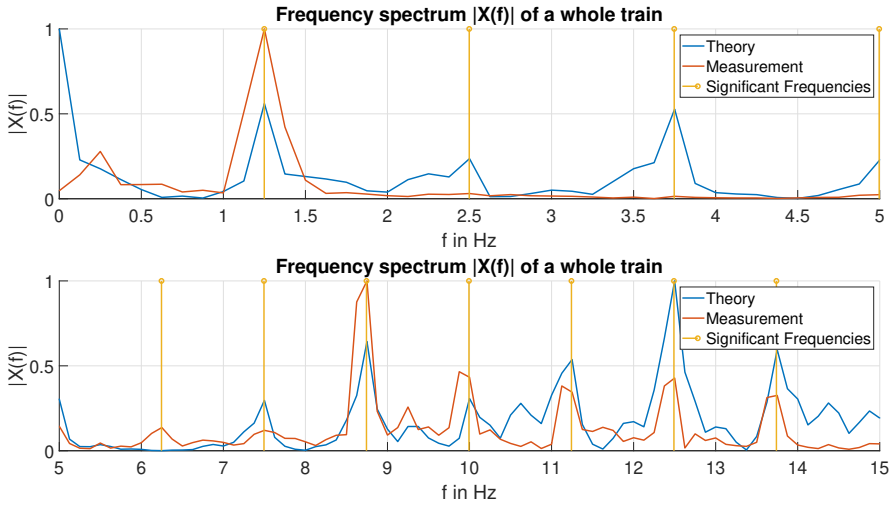


Figure 3.8: Theoretical frequency spectrum for the train type ICE3 as given in equation (2.16) compared to the Fourier Transformation of the measured signal in figure 3.7. The yellow lines show the selected frequency components referred to the wagon length.

train lengths.

3.2 Wavelet Transformation

The second analysis tool to extract features is the Wavelet Transformation [30]. There are two types of Wavelet Transformations:

- Continuous Wavelet Transformation (CWT) and
- Discrete Wavelet Transformation (DWT)

The CWT is defined as

$$CWT(a,b) = \int_{-\infty}^{\infty} x(t)\psi^*_{(a,b)}(t)dt \tag{3.2}$$

where $x(t)$ represents the signal to transform and $\psi^*_{(a,b)}(t)$ is the continuous wavelet, which is

$$\psi^*_{(a,b)}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}) \quad \text{with } a, b \in \mathbb{R}, a \neq 0 \tag{3.3}$$

where ψ is called the mother wavelet and can be seen as the basis function for this transformation. The parameters a and b are the scale and the translation parameter.
The CWT is used for theoretical research whereas the DWT is a discretized form of the CWT used more in practical tasks [28]. The wavelet in equation (3.3) changes for the DWT to

$$\psi^*_{(j,k)}(t) = \frac{1}{\sqrt{2^j}}\psi(\frac{t-2^jkN}{2^j}). \tag{3.4}$$

where $a = 2^j$ and $b = ka = k2^jN$ were substituted and $N = 19$ is the sample interval for the unscaled wavelet for this thesis. The here used unscaled (j=0) Haar-wavelet is shown in figure 3.9. The used discretization is the most common one and is called dyadic sampling.
By increasing j the scale increases by a factor of 2 which means the wavelet is

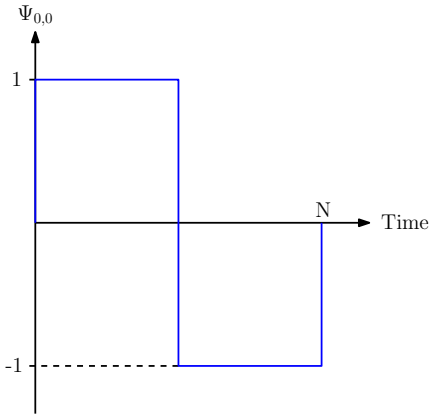


Figure 3.9: Mother Haar-Wavelet.

dilated to the double length in every step. This results in the time resolution getting worse but the frequency resolution better. The time and frequency resolution for different scale values is shown in figure 3.10 with the scales

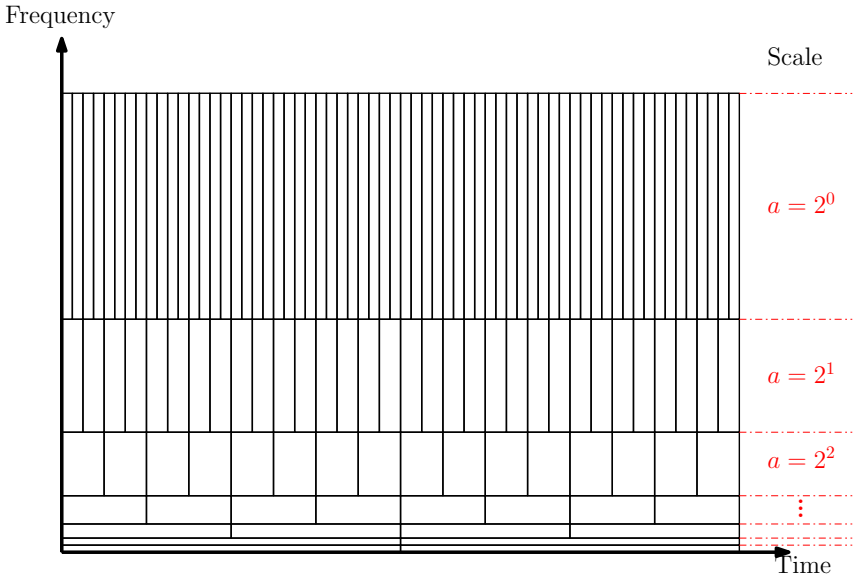


Figure 3.10: Time and frequency resolution with dyadic sampling of the scale and time shift.

in red on the right side.
The discretized Wavelet Coefficients

$$DWT(j,k) = \langle \mathbf{x}, \boldsymbol{\psi}^*_{(j,k)} \rangle \tag{3.5}$$

can be calculated from the scalar product between signal $\mathbf{x}$ and wavelet $\boldsymbol{\psi}^*_{(j,k)}$. Here j defines the scaling (frequency resolution in figure 3.10) and k is the actual position on the time axis in figure 3.10.
A 6-level decomposition is applied to buffers with a length of 1250 samples. The decomposition filter bank can be seen in figure 3.11. Note here that the decomposition level shown in the figure is different from the scaling level j. The relationship between scaling level j and frequency range can be seen in figure 3.10. Decomposition level 1, which is scaling level $j = 0$, was not further used, as no further information gain was assumed in the highest frequency range. This assumption was based on the frequency proportional damping factor of sound waves through soil. Noise prevails in this high frequency range.

So the Discrete Wavelet Transform provides a signal for every decomposi-

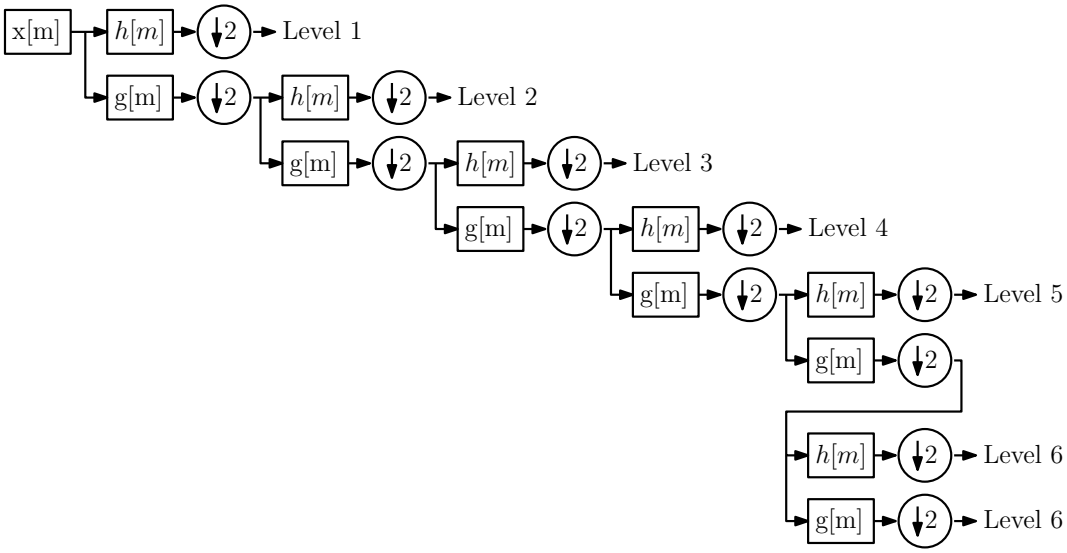


Figure 3.11: Level 6 decomposition filter bank for the wavelet analysis. $x[m]$ is the signal buffer, $h[m]$ is the high-pass filter and $g[m]$ the low-pass filter. After the filter operation the signal is down sampled by a factor 2.

tion level. There is no need to determine the exact time of the individual frequency components within the short buffer size and so the absolute values for each decomposition level are summed up

$$d(j) = \sum_{k} |DWT(j,k)| \tag{3.6}$$

The calculated sum of wavelet coefficients $d(j)$ for the events from figure 3.1 are shown in figure 3.12. High values in the respective scaling level j mean a high proportion of components in this frequency range.
The Wavelet Transformation of signals from trains show high components in the lower frequency spectrum, especially in the scaling levels $j = 5, 6$. These scaling levels refer to the same frequency range as the DFT shown in the top of figure 3.3. They refer to the vibration components produced by the train dimensions.
Note that the solid red line in the image has been cut off to better represent the other lines. It can be seen that trains can be distinguished from other disturbances and noise in scaling level $j = 4, 5, 6$. But compared to the DFT, no visual distinction can be seen between the train types.

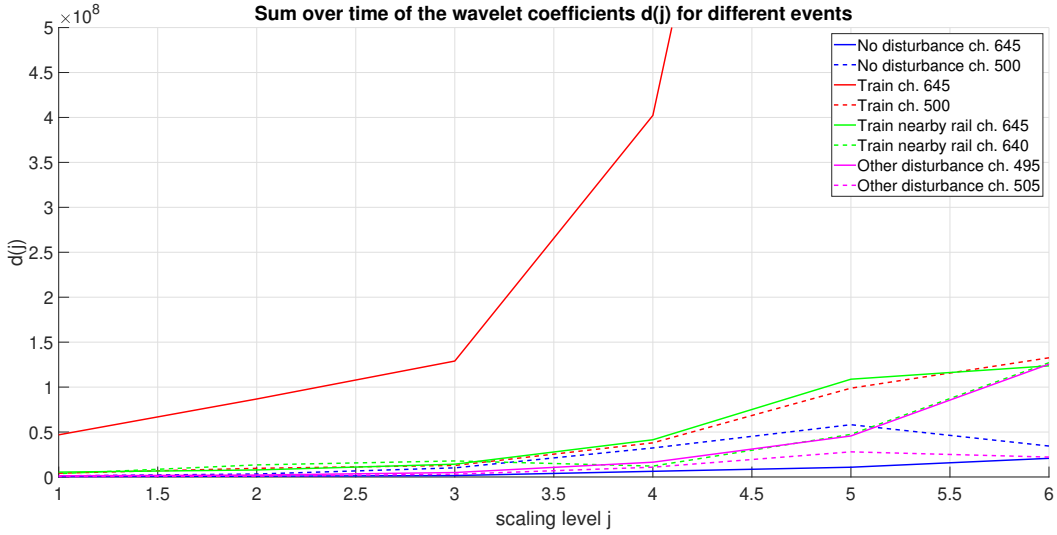


Figure 3.12: Sum over time of the wavelet coefficients $d(j)$ for different events.

3.3 Empirical Mode Decomposition (EMD)

The third analysis tool chosen to extract features is the Empirical Mode Decomposition (EMD). The EMD has no analytical formulation and is therefore defined by an algorithm [31].
For the thesis the algorithm introduced in [32] was implemented. The idea is to decompose a signal into a set of different scaled data sequences and a final residue. These sequences are called intrinsic mode functions (IMF), which have to satisfy the following conditions:

- The number of extrema points and zero-crossings differ by at most one.
- The mean values of the upper and lower envelopes that are defined by the local maxima and local minima should be equal to zero at any point in the signal.

A signal $x(t)$ can be decomposed into

$$x(t) = \sum_{i=1}^{n} q_i(t) + r(t) \tag{3.7}$$

where $q_i(t)$ is the i-th IMF and $r(t)$ is the final residue. The algorithm to calculate the EMD consists of the following steps, which are quoted from [32].

1. Initial setting. Set $i = 1$, $r_{i-1}(t) = x(t)$, and $h_i^o(t) = r_{i-1}(t)$, where i is the current IMF count, $r_{i-1}(t)$ is the residue after $i-1$ IMFs have been extracted, and $h_i^o(t)$ is the i-th candidate IMF generated in the previous iteration.
2. Identify the extrema (the local maxima and local minima) of $h_i^o(t)$.
3. Generate the upper envelope $up(t)$ and lower envelope $low(t)$ from the local maxima and local minima, respectively.
4. Compute the mean envelope $m(t)$ by averaging the upper and lower envelopes: $m(t) = (up(t) + low(t))/2$.
5. Sift mean: $h_i^o(t) = h_i^n(t) - m(t)$, where $h_i^n(t)$ is the new i-th candidate IMF generated in the current decomposition iteration.
6. Review the stopping criterion. If the stopping criterion is not satisfied, then proceed to the next iteration by letting $h_i^o(t) = h_i^n(t)$ and repeating steps 2–6. If the stopping criterion is satisfied, then let the i-th IMF $q_i(t) = h_i^n(t)$ and proceed to step 7.
7. Update the residue $r_i(t)$ by subtracting the obtained IMF function $q_i(t)$ from the previous step: $r_i(t) = r_{i-1}(t) - q_i(t)$, where $r_i(t)$ is the residue after the previous i IMFs have been extracted.
8. Determine whether $r_i(t)$ becomes a monotonic function. If $r_i(t)$ is not a monotonic function, then let $i = i + 1$, $h_i^o(t) = r_{i-1}(t)$, and repeat steps 2–8 to identify the remaining IMFs. If $r_i(t)$ becomes a monotonic function, then let the final residue $r(t) = r_i(t)$ and stop the EMD because no more IMFs can be obtained.

For the stopping criterion the standard deviation bound (SDB) was chosen from the work of [32]. The SDB is given by

$$sd = \sum_{\tau=1}^{B} \frac{(h_i^o[\tau] - h_i^n[\tau])^2}{(h_i^o[\tau])^2} \tag{3.8}$$

where B is the buffer length. The stopping criterion is satisfied when sd is smaller than a chosen value of 0.5. For smaller threshold values the

calculation time was not acceptable.
Equation (3.7) put into the discrete time domain yields to

$$x[m] = \sum_{i=1}^{n} q_i[m] + r[m]. \tag{3.9}$$

According to equation (3.9) the IMF's and residue are always functions of time. To minimize the number of features the sum over time of the absolute value of the IMF's and the residue was taken. For the IMF's this is

$$EMD[k] = \sum_{\tau=1}^{B} |q_k[\tau]| \tag{3.10}$$

with $k = 1, 2, ..., n$ where n is the number of IMF's. For the residue it is

$$EMD[n+1] = \sum_{\tau=1}^{B} |r[\tau]|. \tag{3.11}$$

The EMD, as given in equation (3.10) and (3.11), is shown in figure 3.13 for the events which are analysed in this chapter. Because of the constant stopping criterion from equation (3.8) the decomposition depth k is not always the same. What can be said is that the highest decomposition depth k can be considered as the contribution of the lowest frequency band. However, the frequency bands for an EMD cannot be narrowed down. What was done for all the future steps is to take the six highest values of every calculated EMD. This means that the high frequency components were again neglected. As for the other two features before, also the EMD of buffers containing train measurements differ, albeit very slightly, from the other disturbances and noise.

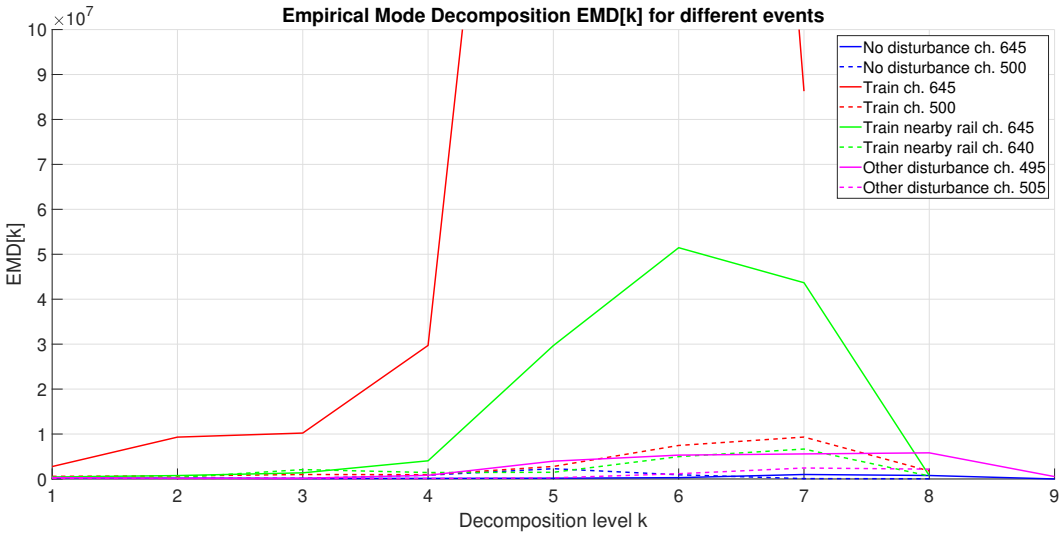


Figure 3.13: Empirical Mode Decomposition of x[m] for different events.

4 Detection Algorithm and Results

After completion of the feature extraction, the last part of the task is now to use the previously described features as input for the expert system. As mentioned in the introduction, the localization of a train when using fibre optic sensing is a detection task.
In this chapter, two algorithms are described as the expert system to detect trains. The general structure of both algorithms was already shown in the introduction chapter but is again shown in figure 4.1. Remember the structure of the dataset given in chapter 2. It consists of channels (every channel belongs to a position along the fibre) and the time signal for every channel is buffered (buffer length: 0.5 s). The features are calculated with these buffers and are the input for the detection algorithm, which can decide between three classes

- *no train* (0)
- *train* (1)
- *ICE3* (2).

The first algorithm can only decide between the first two classes. The second algorithm also classifies for the third class but because no reference is provided for that class it is considered as class *no train* for the evaluation. The output of the algorithm looks like shown in figure 1.9. The buffers marked in black were detected as a train and so the train is localized with position (channel) and timestamp (buffer).
The classification is done by two approaches. The first one is training a Support Vector Machine (SVM) with labelled training data. Unfortunately no reference was provided for the trains on nearby rails (class *ICE3*) so it was not possible to train the SVM for this class. The SVM classifies only into *no train* and *train*.
The second algorithm for classification is an expert system, which was

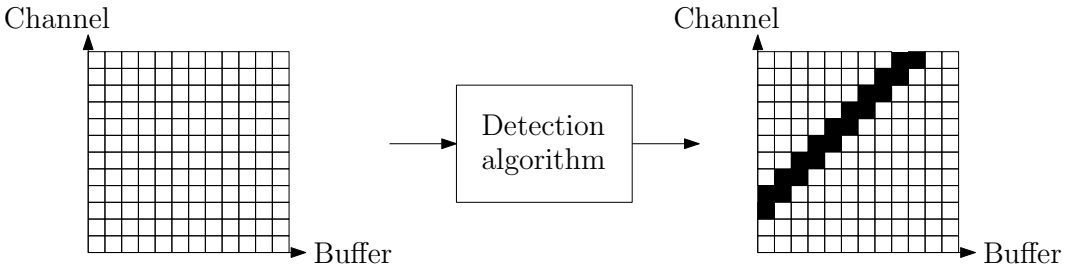


Figure 4.1: Structure for the detection algorithms which receives the features for every channel and buffer and classifies the buffer. White areas were classified as *no train* whereas black areas were classified as *train*.

developed myself. It uses the signal energy and the Fourier Transformation with the information about significant frequencies provided in section 2.3 and 3.1.1. With this second approach it is possible to distinguish between different train types.

Before starting with the algorithms, a suitable reference must first be generated. Unfortunately, the provided reference to the measurements is limited.

4.1 Reference Construction

A reference is given in form of ground truth data. It provides the number N_c of trains, as shown in figure 2.8, coupled together and the corresponding timestamps when this train enters channel 630. For all the other channels no reference was provided. So the first thing to do was to generate a reference for a small part of the dataset in the form like shown in figure 1.9.
For this purpose the signal energy was calculated for every channel and buffer. A low threshold was set for the energy to receive a very rough detection like shown in figure 4.2.
The timestamp of the head of the train is given for channel 630 with the corresponding train length. To get the correct timestamp for the tail of the train in channel 630 the speed of the train is needed. For this reason, only channel sections in which trains move at almost constant speed were selected for the reference. This channel groups are:

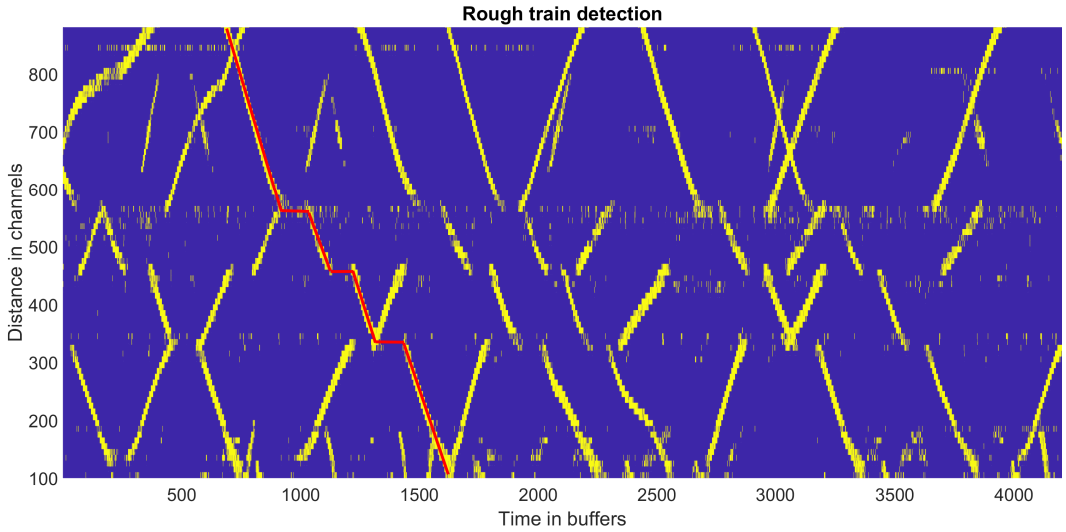


Figure 4.2: Rough detection by setting a threshold for the signal energy. In blue the areas where no disturbance was detected and in yellow the areas where some disturbance was detected. The red line shows the same train through the whole data set.

- channels 190-290
- channels 360-430
- channels 480-540
- channels 590-670
- channels 820-880

In figure 4.3 the chosen sections are shown.
The next step is to estimate the speed of all the trains in this selected channel groups. This was done the same way as already in chapter 3 (figure 3.5) by averaging the slope of head and tail of the train. Note hear that the head and tail for the estimation are not the exact one. The train is determined as much longer as it is in reality. This is because of the rough detection where the threshold for detecting a train is set way to low. But this does not affect the speed estimation.

With the knowledge of the speed and the given train length in form of the number N_c of coupled trains the timestamp for the tail can be calculated. This is shown for the channel section $590 - 670$ for one train in figure 4.4.

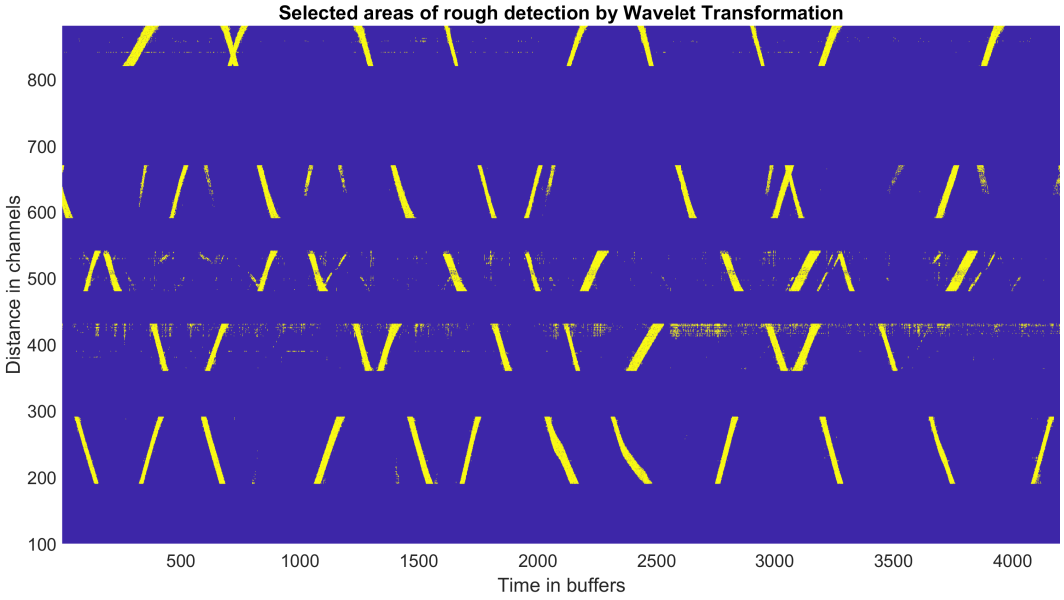


Figure 4.3: From the whole dataset shown in figure 4.2 only the parts where trains move with constant velocity where chosen for determining the reference.

The determined head and tail of this train are shown in green. The dashed lines show the channel and the timestamp of this train given in the ground truth data.

For all the other channel sections the construction of the reference was a bit more difficult because no ground truth is provided. The train length in form of the number N_c of coupled train was extended to the entire dataset by manually tracking the train like indicated with the red line in figure 4.2. The estimation of the speed was done as before. With the speed and length of a train the distance in time between head and tail was calculated. What is missing now is the exact position of the head. This was done by placing the head and tail line, as shown in green in figure 4.5, symmetrically into the rough detection shown in blue.

The whole reference for the selected channel sections can be seen in figure 4.6. Compared to figure 4.3 the channels that were excluded from the reference generation are hidden here. For the trains on nearby rails no ground truth data was given so the class *ICE3* is not part of the reference. So it is a binary reference with the two classes *no train* in blue and *train* in yellow.

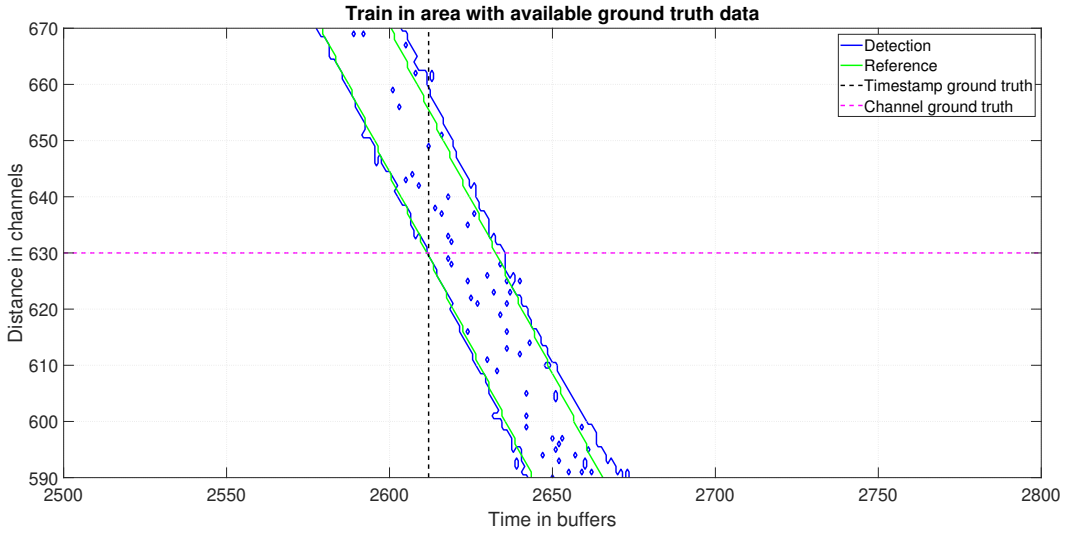


Figure 4.4: Reference construction for one train in the dataset. The channel and the timestamp of the train head given in the ground truth are marked with the dashed lines.

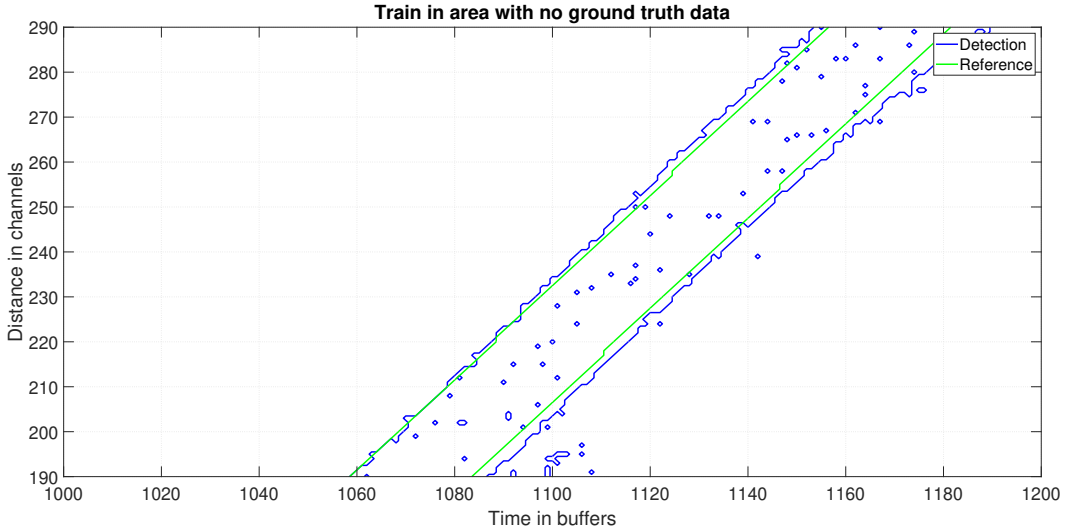


Figure 4.5: Reference construction for one train in the dataset. Here no ground truth was available. The distance in time and the slope of the green lines were determined from the estimated speed and the train length. These lines were placed symmetrically in the rough detection (blue contour).

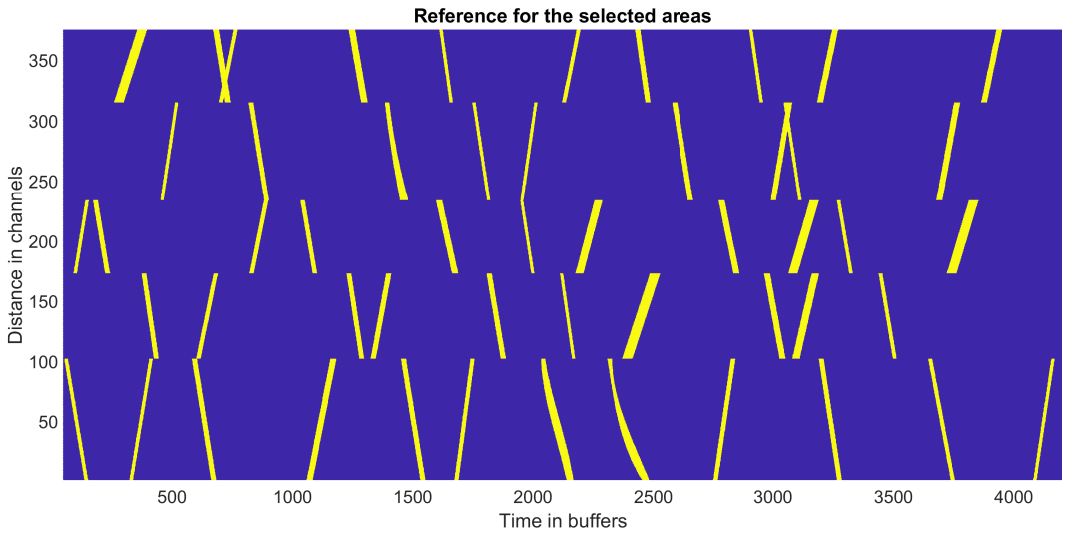


Figure 4.6: The constructed reference for the dataset. It will be used for training the SVM and evaluating both algorithms. The reference contains the two classes *no train* labelled in blue and *train* labelled in yellow. Train station areas were not used for the reference and are hidden in this representation.

4.2 Performance Measure of a Classifier

The easiest way to determine the performance of a classifier is to calculate the accuracy, which is the number of correct classifications divided by the total number of classifications. When the classes in the dataset are not even separated the accuracy is not a good performance measure. In the dataset of the thesis there are much more buffer labelled as *no train* than as *train*. Another performance measure for tasks like this is to calculate the confusion matrix [33]. Each row in a confusion matrix represents one class, while each column represents a predicted class. In this work the reference only contains two classes and so the confusion matrix is a $2x2$-matrix with the classes *no*

train and *train*:

$$\begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix} \tag{4.1}$$

where TN means *true negative* and is the number of correctly classified as *no train*, FP means *false positive* and is the number of wrong classified as *train*, FN means *false negative* and is the number of wrong classified as *no train* and TP means *true positive* and is the number of correctly classified as *train*. With this declaration some other interesting performance measures can be defined. One of them is the accuracy of the positive predictions which is called *precision* and is defined as

$$precision = \frac{TP}{TP + FP}. \tag{4.2}$$

It is typically used in combination with *recall*, which is the ratio of positive classes that are correctly classified. Another name therefore is sensitivity.

$$recall = sensitivity = \frac{TP}{TP + FN} \tag{4.3}$$

Often *precision* and *recall* are combined into one performance measure called $F1$-score. It is the harmonic mean of precision and recall and is defined as

$$F1 = \frac{2}{\frac{1}{precision} + \frac{1}{sensitivity}} = \frac{TP}{TP + \frac{FN + FP}{2}}. \tag{4.4}$$

The harmonic mean gives more weight to low values so a high $F1$-score can only be achieved when both, *precision* and *recall*, are high.

4.3 Support Vector Machine (SVM)

The first algorithm to test is the Support Vector Machine (SVM). A SVM is a very powerful and versatile Machine Learning model [33]. The idea

behind can best be explained by considering a simple binary classification. The decision function therefore is defined as

$$f(\mathbf{x}) = sign(\mathbf{w}^T\mathbf{x} + b). \tag{4.5}$$

In figure 4.7 from [34] an example is shown where the separating hyperplane $(\mathbf{w}^T\mathbf{x} + b) = 0$ is shown. The data points are labelled with $'+'$ for

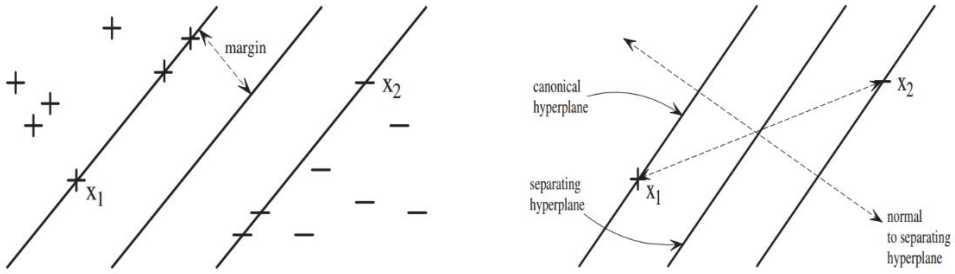


Figure 4.7: Example of a binary classification problem. The hyperplane separates the two classes labelled with $+$ and $-$. x_1 and x_2 are examples of support vectors which are the closest points to the separating hyperplane [34].

$(\mathbf{w}^T\mathbf{x} + b) > 0$ and $'-'$ for $(\mathbf{w}^T\mathbf{x} + b) < 0$. The hyperplane is determined from the training data by maximizing the margin, which means to find the hyperplane that maximizes the distance to the nearest training points. The closest points to the separating hyperplane are called support vectors. After the model is trained, the remaining data is classified by the sign of the decision function.

If the dataset is not linearly separable, a so called **kernel trick** can be used. The main idea is to transform the data into a higher dimensional space where it is linearly separable. This transformation is relatively computation intensive. A kernel is defined as

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j) \tag{4.6}$$

with $\mathbf{x}_i$ and $\mathbf{x}_j$ vectors in the actual space and $\phi()$ the transformation function. The kernel function helps to reduce the computational load by not explicitly computing the new coordinates. A commonly used kernel is the polynomial

kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = [(\gamma \mathbf{x}_i^T \mathbf{x}_j) + r]^d. \tag{4.7}$$

A polynomial kernel of degree $d = 3$ is used for this thesis. More information about that kernel trick can be read in [33].
In MATLAB a SVM can be learned by using the function *fitcsvm()*. It is a binary SVM classifier so it can be only used for classifying between *no train* and *train*. Also the reference in figure 4.6 only contains two classes.
To train the SVM the whole dataset has to be split into a training set and a validation set. For the training data, 20% of the data labelled as *train* and the same amount labelled as *no train* was chosen.

4.3.1 Results

The SVM was used in two ways:

- train a separate model for each channel
- train one model for all channels

The used features to train the SVM are:

- Fourier Transformation
- Wavelet Transformation
- Empirical Mode Decomposition
- Signal Energy.

These are used individually to train the SVM and are then compared to each other. As mentioned above the SVM can only classify between the two classes *no train* and *train* because no reference for the class *ICE3* is available to train the SVM accordingly.

In figure 4.8 the results are shown when training a separate SVM for every channel with 20% of the data which is labelled as *train* and the same amount labelled as *no train*. The corresponding performance values and the distribution of data are shown in table 4.1.
For such a dataset where one label prevails the other, the accuracy is not a good performance measure as was already mentioned previously because

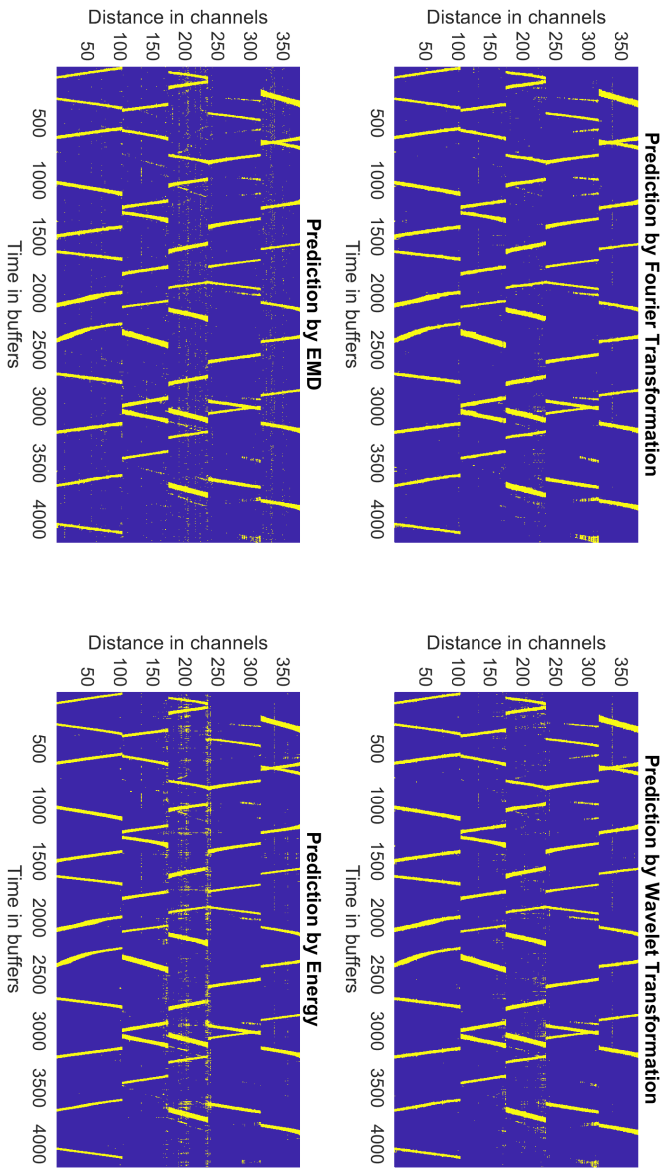


Figure 4.8: Results using a SVM model for each channel, trained with different features. The performance results corresponding to this plots are given in table 4.1.

Feature	Accuracy training	Accuracy validation	Precision	Sensitivity	F1
-	%	%	%	%	%
Fourier	99.9140	98.0264	73.4768	98.4455	84.1480
Wavelet	99.6592	98.2866	76.2427	98.3360	85.8914
EMD	99.8465	98.1796	75.7741	96.6634	84.9535
Energy	98.5077	98.5079	79.1855	96.8521	87.1323

Data distribution	Training	Validation	*train*	*no train*	All
	32568	1542432	82150	1492850	1575000

Table 4.1: Results using a SVM model for each channel. 20% of the data labelled as *train* and the same amount of data labelled as *no train* was used to train the SVM.

95.73% of the data are labelled with *no train* in the reference. The other values play a more important role here. The *recall* in table 4.1 is quite high for all 4 features used. This means that the probability not to detect a train is very low. Whereas the *precision* is not that good. This is because the trains on nearby rails are wrongly detected as *train* for all 4 features, which increases the value of *FP* and so decreases the *precision*. Trains on nearby rails can be recognized in figure 4.8 by the fact that they suddenly appear at around channel 300.
With the *F1*-score the results can be compared and it shows that the energy comes out best as a feature. And generally, it can be seen that compared to the plot from the introduction the detection is better. No spaces in the detection (green rectangle in figure 1.6) and less detected disturbances. But there are still some detections where no train should be in figure 4.8 for all features.

The second way of using the SVM is to train just one model and use it for the whole dataset. The results to this approach are shown in figure 4.9 and table 4.2. This approach delivers horrible results for the Fourier Transformation and the EMD as features. In the introduction it was discussed that the measurement quality varies along the fibre. So, every channel has slightly different properties. This was also seen in the examples discussed

in chapter 3. This variations over the channels lead here to these bad results. Compared to the DFT and the EMD, the results of the Wavelet Transforma-

Feature	Accuracy training	Accuracy validation	Precision	Sensitivity	F1
-	%	%	%	%	%
Fourier	53.921	96.0126	90.6926	7.6506	14.1109
Wavelet	97.6603	97.7768	70.8792	97.3086	82.0173
EMD	50.6356	95.7636	83.844	1.0992	2.1700
Energy	91.6452	98.4545	83.7443	83.9696	83.8568

Data distribution	Training	Validation	*train*	*no train*	All
	32568	1542432	82150	1492850	1575000

Table 4.2: Results using a SVM model for all channels. 20% of the data labelled as *train* and the same amount of data labelled as *no train* was used to train the SVM.

tion and the Energy as a feature are not that bad but still worse than for the approach with a SVM model for every channel separately.

At the end it can be said that it is important to implement the detection adaptive with respect to the channel. The differences in the measurement quality can be due to different reasons. Some examples are:

- Position of the cable duct (beside the track, on the walls of tunnels, ...)
- Cable slack (cable rolled up for maintenance reasons)
- Roads or highways passing near the track
- Echo effects in tunnels.

Information regarding this would be very helpful for future works.

To further improve the results of this first algorithm it is necessary to detect the trains on neighbour rails correctly. This should improve the *precision* and thus also the *F1*-score.

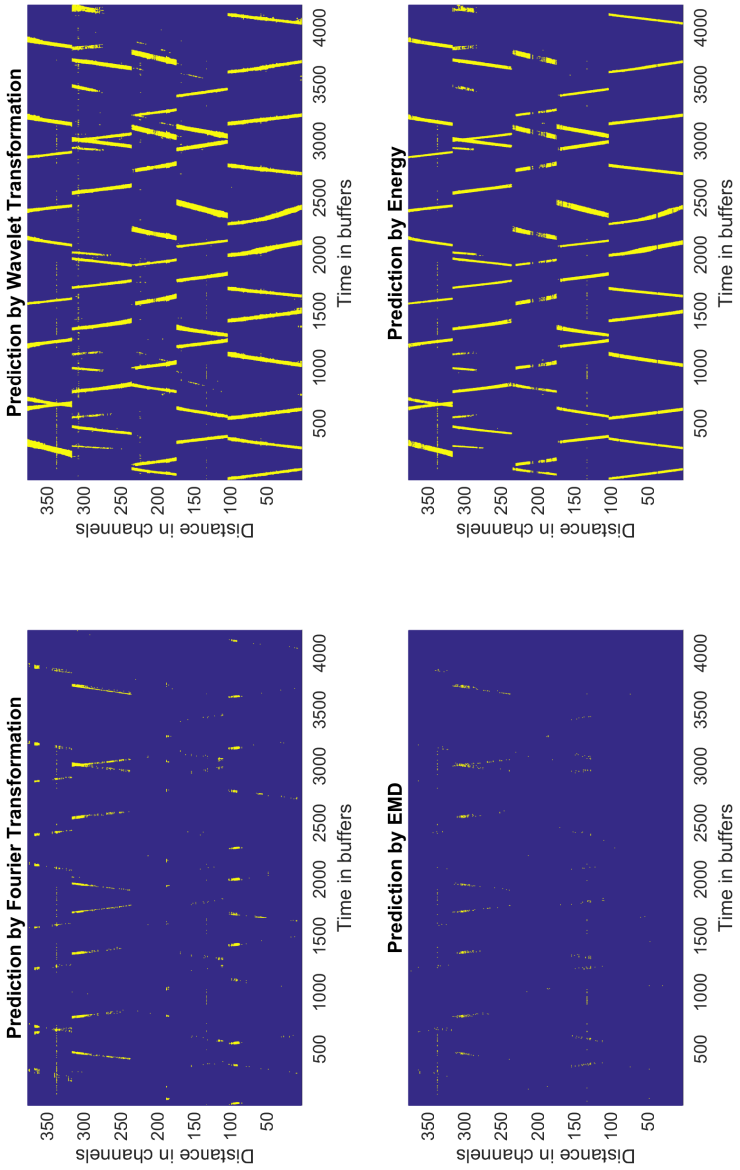


Figure 4.9: Results using a SVM model for all channel trained with different features. The performance results corresponding to this plots are given in table 4.2

4.4 Threshold and Pattern Algorithm

The second algorithm implemented in this thesis is an improvement to the SVM algorithm as it distinguishes between different types of trains. The difference to the previous algorithm is that it is no machine learning algorithm. No labelled training data is needed. It is based on the combination of two features. Train vibrations acting on the fibre are detected when the signal energy exceeds a threshold. The energy threshold for the detection is implemented adaptively. The energy for the detection was chosen because of the best results according to the *F1*-score in the SVM approach.
The detected trains are then classified using the DFT into either a train of type DB423/433 or a train of type ICE3.

The flow graph of that algorithm is shown in figure 4.10. The SVM has shown that the detection works better when looking at each channel separately. This algorithm is also applied on each channel, which is indicated by the parameter i. The parameter j is the actual buffer.
After initialization the first buffer of the actual channel i is loaded, preprocessed and the energy of that buffer is calculated. The threshold for detecting a train is determined by a moving average filter of length 50.
After the energy of the actual buffer is calculated there are three possible paths through the algorithm.

If the energy is less than the threshold and the previous state was *no train*, then the algorithm goes on with the next buffer, but if the previous state was *train*, this means, that the end of a train was detected and so the train length L_{train} can be calculated from the estimated speed v and the time Δt needed to pass the actual channel i by $L_{train} = v \cdot \Delta t$. The end of the train was detected so the state is *no train* and the next buffer is processed. If the calculated train length is less than 10 m, the detected train is discarded. Such a short train makes no sense as the wagon lengths are already longer than this.

The third possible path is when the actual energy level is higher than the threshold which means a train was detected. The input buffer x_{ij} is stored into a new buffer and a Discrete Fourier Transformation is applied on it.

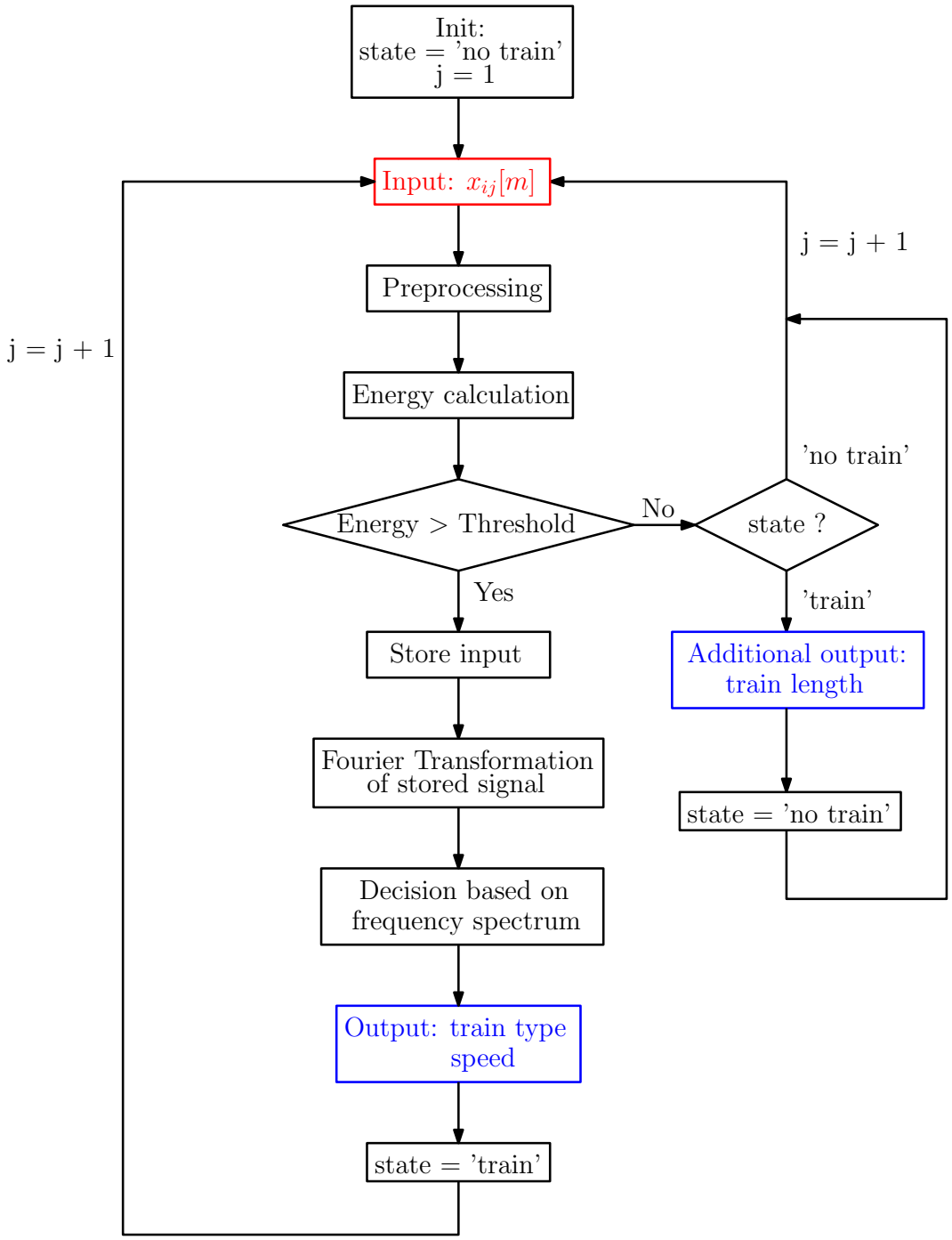


Figure 4.10: Flow graph of the adaptive threshold algorithm. This algorithm is applied to every channel i and j is the actual buffer.

In section 3.1.1 it was talked about the frequency spectrum a moving train produces and also about the most significant frequency components referred to the wagon lengths. This can now be used to distinguish between the two classes:

- Train type DB 423/433 (labelled with *train*)
- Train type ICE3 (labelled with *ICE3*)

Note here that because of the negligible small differences in the wagon lengths of train type DB423/433 and DB430 (see table 2.1) these two types were considered one.
The idea is to compare the calculated DFT of the train from the measurement with the ideal frequency spectrum of a train according to equation (2.15) for type DB423/433 and equation (2.16) for type ICE3. This is done by solving an optimization problem.

An optimization problem can generally be written as: [35]

$$\min_{\mathbf{x}} f(\mathbf{x}) \tag{4.8}$$

subject to

$$\mathbf{c}(\mathbf{x}) \leq \mathbf{0} \tag{4.9}$$
$$\mathbf{c}_{eq}(\mathbf{x}) = \mathbf{0} \tag{4.10}$$
$$\mathbf{A} \cdot \mathbf{x} \leq \mathbf{b} \tag{4.11}$$
$$\mathbf{A}_{eq} \cdot \mathbf{x} = \mathbf{b}_{eq} \tag{4.12}$$
$$\mathbf{lb} \leq \mathbf{x} \leq \mathbf{ub} \tag{4.13}$$

$f(\mathbf{x})$ is the function to minimize by manipulating the optimization variables $\mathbf{x}$. $\mathbf{c}(\mathbf{x})$ and $\mathbf{c}_{eq}(\mathbf{x})$ are the nonlinear constraints functions. $\mathbf{A}$, $\mathbf{A}_{eq}$, $\mathbf{b}$ and $\mathbf{b}_{eq}$ the matrices and vectors respectively to form the linear constraints. $\mathbf{lb}$ and $\mathbf{ub}$ define the lower and upper bound of the optimization variables.

To get the ideal frequency spectrum of a train according to equation (2.15) and (2.16), only the speed of the train is unknown. The last term containing the number N_c of trains coupled together was not used here. So, the speed of train is the only optimization variable.

$$\mathbf{x} = x = v \tag{4.14}$$

The optimization function was chosen to be the Euclidean norm

$$f(\mathbf{x}) = ||z(\mathbf{x}) - y|| \tag{4.15}$$

with $z(\mathbf{x})$ the fitting model function, which is equation (2.15) respectively equation (2.16) and y the calculated DFT of the stored buffers. For the optimization only the frequency range from $0 - 15\,\mathrm{Hz}$ was chosen. This is the range where the train dimensions have the most influence as was shown in figure 3.6. Both, the model $z(\mathbf{x})$ and the DFT y are normalized by the maximum amplitude value in the examined frequency range. To finish the setup of the optimization problem the constraints have to be defined. Lower and upper bound of the optimization variable was set to $v_{min} = 5\,\mathrm{m\,s^{-1}}$ and $v_{max} = 65\,\mathrm{m\,s^{-1}}$. Maximum value is the maximum speed of the ICE3. The DB423/433 has an even lower maximum speed. Minimum value was not chosen to be $0\,\mathrm{m\,s^{-1}}$ because the selected areas of the reference contain no train stations where the trains stop.
So just equation (4.13) is needed for the optimization problem with

$$\mathbf{lb} = \begin{bmatrix} v_{min} \end{bmatrix}, \quad \mathbf{ub} = \begin{bmatrix} v_{max} \end{bmatrix} \tag{4.16}$$

All the other constraints are left empty. In MATLAB the function *fmincon()* was used to solve the optimization problem. This optimization problem was solved separately for both classes, DB 423/433 and ICE3. The smaller value of the function $f(\mathbf{x})$ determined the class.

Back to the flow graph in figure 4.10. If the energy in the next iteration is again high enough the new buffer is added to the storage and the optimization is solved again but now with a longer signal with the length of two buffers. This means a higher frequency resolution and so the classification gets better with every buffer added to the storage.

4.4.1 Results

The dataset for testing the algorithm was the same as already for the SVM approach. In figure 4.11 the result for the classification with the threshold and pattern algorithm is shown. In blue the class *no train* is represented, in

green the class *train* and in yellow the class *ICE3*. It can be seen that the ICE3 and DB 423/433 are mostly classified correctly. The ICE3 can easily be recognized because he suddenly appears in the dataset near channel 300. This happens where the railway track leaves the tunnel and then runs parallel to other railway tracks.

To get the same performance measure as for the SVM approach the trains

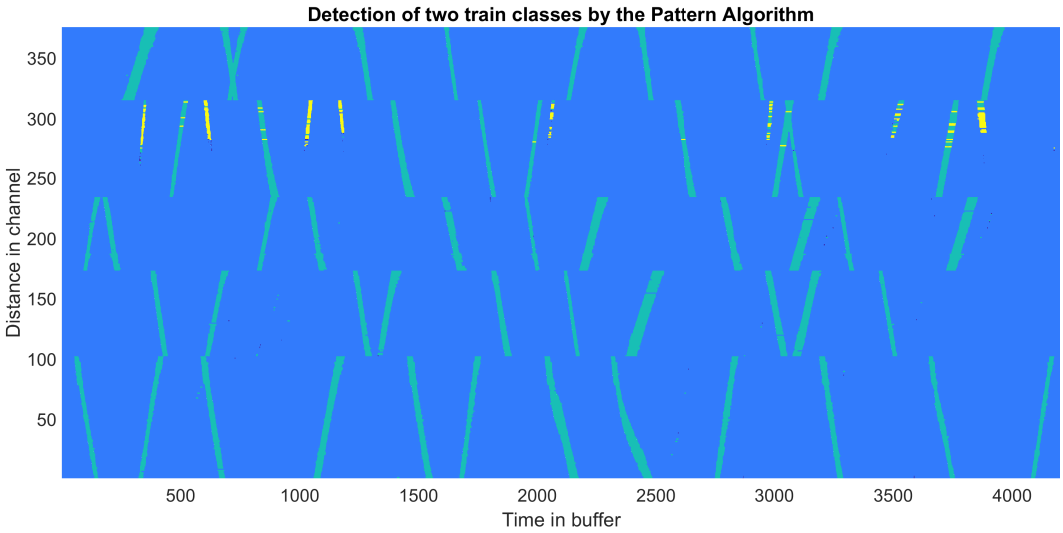


Figure 4.11: Results achieved with the Threshold and Pattern Algorithm. The class *train* is coloured in green, the class *ICE3* in yellow and the class *no train* in blue.

classified as *ICE3* were cleared to class *no train* because the reference in figure 4.6 contains no class *ICE3*. In figure 4.12 the reference and the detection by the threshold and pattern algorithm is shown for two classes. In blue the class *no train* and in yellow the class *train* is depicted. The corresponding performance results are listed in table 4.3 compared to the results from the SVM approach.

The results of the threshold and pattern algorithm are better than the ones achieved with the SVM. The *precision* has been improved because the trains on the neighbour rail were distinguished from the trains under test and so FP in equation (4.2) was decreased.

Also, the sensitivity was slightly improved. When looking at equation (4.3) this means that FN was decreased, which indicates that the tail and head of the trains are better detected.

Figure 4.12: The detection compared to the reference for the Threshold and Pattern algorithm. Class *no train* is coloured in blue, class *train* is coloured in yellow and class ICE_3 was cleared in this figure. The performance results corresponding to this figure is given in table 4.3.

SVM trained channel wise					
Feature	Accuracy training	Accuracy validation	Precision	Sensitivity	F1
-	%	%	%	%	%
Fourier	99.9140	98.0264	73.4768	98.4455	84.1480
Wavelet	99.6592	98.2866	76.2427	98.3360	85.8914
EMD	99.8465	98.1796	75.7741	96.6634	84.9535
Energy	98.5077	98.5079	79.1855	96.8521	87.1323

Threshold and Pattern Algorithm					
Feature	-	Accuracy	Precision	Sensitivity	F1
-	-	%	%	%	%
Energy and Fourier	-	99.0644	86.3789	97.6738	91.6797

Table 4.3: Results of the Threshold and Pattern Algorithm compared to the ones achieved with the SVM, trained for each channel separately.

But the results are not good enough to use this detection algorithm for safety critical tasks in railway applications.

It has to be mentioned here that the reference was also created and thus also contributes into the not optimal results. This algorithm has to be tested with a dataset where also a good and precise reference is provided.

5 Conclusion and Outlook

In the railway transportation there are often still long waiting times. Railway operators are therefore working on concepts to increase the workload on the tracks. What must not suffer as a result is safety. In order to ensure a safe transportation with a higher workload, the position of all trains on the rail network is mandatory. The workload cannot be increased with the existing systems, balise and axle counter, and so the operators want to get rid of these. One system which is used more often is the Global Positioning System (GPS) but it has an unpleasant disadvantage according to the availability (e.g. tunnel, wood, etc.).
That is why railway operators also focus on the research of other systems. One of them is Distributed Acoustic Sensing (DAS) which uses a fibre optic cable to work like a continuous microphone over a range of several kilometres. This makes it possible to capture the vibrations produced by a train. It is assumed to be cheaper as GPS because the system is installed on the track, while tracking with GPS needs every train to be equipped with a GPS-Receiver.

This thesis deals with the investigation of DAS measurement data. The data is analysed with different signal analysis tools and this acquired information is used in algorithms to locate trains.
Other research has shown that the frequency spectrum provides important information when measuring the vibrations produced by trains and that is why the Fourier Transformation and the Wavelet Transformation were chosen as analysis tools. In addition, Empirical Mode Decomposition is used as an analysis tool and also the signal energy is taken into account.
The data analysis shows that the measurement quality depends on the fibre routing. This is what it makes difficult to reliably detect the head and tail of a train.
Another insight was that it is not possible to measure the lateral distance

from the fibre to the vibration source. So also trains on neighbouring rails were measured. These trains differ from the ones on the test track in their dimensions. The work also shows that the frequency spectrum of these two types of trains differ because it is dependent of the train dimensions.
The structure of DAS measurements makes it possible to treat the train localization as a detection problem. For each position along the fibre a separate time signal is measured by the DAS system. For the detection two algorithms are implemented. The analysis tools were applied on small buffers of the whole time signal for each position to extract the features. These features are the input for the detection algorithms, which decides whether this buffer contains train vibrations or not.

The first provided algorithm is a Support Vector Machine (SVM). With a single SVM model for the whole fibre length no good results were achieved because of the fact that the measurement quality differs along the cable due to the way of installation and environmental impacts.
Better results were achieved by training a SVM model for every channel separately. The SVM approach shows a cross-sensitivity for trains driving on the neighbouring rail because no reference for these trains was provided. The evaluation of the algorithm is carried out with the values, precision and sensitivity. They can be calculated by comparing the classified buffers with the buffers from the reference. For the SVM a precision of 79.19% and a sensitivity of 96.85% were reached as the best result. For this result the energy was used as a feature.

A second algorithm, which is a so called expert system, was the next step to improve the results of the SVM. Especially the precision should be improved by the fact that with the second algorithm also a distinction of the train types is possible.
It is based on detecting a train by energy and classifying it by solving an optimization problem using the DFT of the measurements and a model of the train vibrations in the frequency domain. The work includes research on significant frequency components produced by moving trains. It is shown that these frequencies depend on the train dimensions and the moving speed.
With the solving of this optimization problem the second algorithm also classifies the trains on neighbouring rails. Because no reference was available for

them, they were considered as not detected for the evaluation. The precision for this second algorithm is 86.38% and the sensitivity is 97.67%. In the SVM approach the trains on the neighbouring rail were not distinguished what reduced the precision. For the second algorithm this was improved.

The train detection implemented in the work, which determines the head and tail of a train, corresponds to the localisation of the train. When a train is detected in one of the channels along the fibre this is also the current position of the train. The results for precision and sensitivity thus also describe the accuracy of the detection of head and tail of the train.

But there are several improvements which can be done to get better results. First of all, new measurements have to be made and a precise reference has to be provided with this measurements. This would allow a better analysis of the data and it would be possible to improve the algorithms.
Also, more measurements on different tracks with different installations have to be done to get a better understanding of environmental impacts. Therefore, precise references have to be provided, not only for the trains driving on the track but also references regarding the installation of the fibre and information about the trains on neighbour rails.

The idea of railway operators is to use DAS in combination with other sensors such as Global Positioning System (GPS) and Inertial Measurement Unit (IMU) to provide a train localization system without limitations in availability. GPS and IMU work properly most of the time and areas where GPS fails can be covered by DAS. With the additional analysis for the DAS system as mentioned above it could be improved so that it would fulfil the standards for a use in safety critical purposes.

Bibliography

[1] Robert Pascoe and Thomas Eichorn. "What is communication-based train control?" In: *IEEE Vehicular Technology Magazine* 4.4 (2009), pp. 16–21. ISSN: 1556-6072. DOI: 10.1109/MVT.2009.934665 (cit. on p. 1).

[2] Ulrich Maschek. *Sicherung des Schienenverkehrs*. Wiesbaden: Vieweg+Teubner Verlag, 2012. ISBN: 978-3-8348-1020-5. DOI: 10.1007/978-3-8348-2070-9 (cit. on p. 1).

[3] Kunal Maurya, Neelu Jain, and Mandeep Singh. "Reail Time Vehicle Tracking System using GSM and GPS Technology: An Anti-theft Tracking System". In: (2012) (cit. on p. 2).

[4] Kevin Chetty, Qingchao Chen, and Karl Woodbridge. "Train monitoring using GSM-R based passive radar". In: *2016 IEEE Radar Conference (RadarConf)*. IEEE, 2.05.2016 - 06.05.2016, pp. 1–4. ISBN: 978-1-5090-0863-6. DOI: 10.1109/RADAR.2016.7485069 (cit. on p. 2).

[5] M. S. Braasch and A. J. van Dierendonck. "GPS receiver architectures and measurements". In: *Proceedings of the IEEE* 87.1 (1999), pp. 48–64. ISSN: 00189219. DOI: 10.1109/5.736341 (cit. on p. 2).

[6] HUBER+SUHNER. *GPS (GNSS)*. Ed. by HUBER+SUHNER. URL: https://www.hubersuhner.com/en/solutions/railway/applications/train-to-ground-communications/gps-gnss (cit. on p. 2).

[7] railsystem. *Balise*. 2015. URL: http://www.railsystem.net/balise/ (cit. on p. 3).

[8] Mordechai Ben-Ari and Francesco Mondada. *Elements of Robotics*. Cham: Springer International Publishing, 2018. ISBN: 978-3-319-62532-4. DOI: 10.1007/978-3-319-62533-1 (cit. on p. 3).

[9] Yang Jin, Tian Gui-yun, and Gao Yun-lai. "High reliable unilateral inductive axle counting sensor system and applications". In: *2016 IEEE Far East NDT New Technology & Application Forum (FENDT)*. IEEE, 22.06.2016 - 24.06.2016, pp. 227–231. ISBN: 978-1-5090-2664-7. DOI: 10.1109/FENDT.2016.7992030 (cit. on pp. 4, 5).

[10] Frauscher. *Frauscher Axle Counter*. 2005. URL: http://www.selectrail.com.au/references/04-frauscher-axle-counter-acs2000 (cit. on p. 5).

[11] Martin Rosenberger and Andrew Hall. "Distributed Acoustic Sensing as a base technology for railway applications". In: *Signal + Draht* (2016). URL: https://www.frauscher.com/assets/media/Fachartikel/2015_09_SD_Distributed_Acoustic_Sensing.pdf (cit. on pp. 6, 7).

[12] P.G.E. Lumens. *Fibre-optic sensing for application in oil and gas wells*. 2014. DOI: 10.6100/IR769555 (cit. on pp. 7, 17, 18).

[13] Z. N. Wang et al. "Ultra-long phase-sensitive OTDR with hybrid distributed amplification". In: *Optics letters* 39.20 (2014), pp. 5866–5869. DOI: 10.1364/OL.39.005866 (cit. on p. 10).

[14] Yaakov Bar-Shalom, Fred Daum, and Jim Huang. "The probabilistic data association filter". In: *IEEE Control Systems* 29.6 (2009), pp. 82–100. ISSN: 1066-033X. DOI: 10.1109/MCS.2009.934469 (cit. on p. 11).

[15] Saurabh Pal and Madhuchhanda Mitra. "Design of a Form Factor Based Expert System for Detection of Myocardial Infarction Using Supervised Classification Technique". In: *2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies*. IEEE, 28.12.2009 - 29.12.2009, pp. 398–400. ISBN: 978-1-4244-5321-4. DOI: 10.1109/ACT.2009.104 (cit. on p. 11).

[16] Zujie Fang et al., eds. *Fundamentals of Optical Fiber Sensors*. Hoboken, NJ, USA: John Wiley & Sons, Inc, 2012. ISBN: 9781118381717. DOI: 10.1002/9781118381717 (cit. on p. 15).

[17] Leonid B. Liokumovich et al. "Fundamentals of Optical Fiber Sensing Schemes Based on Coherent Optical Time Domain Reflectometry: Signal Model Under Static Fiber Conditions". In: *Journal of Lightwave Technology* 33.17 (2015), pp. 3660–3671. ISSN: 0733-8724. DOI: 10.1109/JLT.2015.2449085 (cit. on p. 18).

[18] Zhou Sha, Hao Feng, and Zhoumo Zeng. "Phase demodulation method in phase-sensitive OTDR without coherent detection". In: *Optics express* 25.5 (2017), pp. 4831–4844. DOI: 10.1364/OE.25.004831 (cit. on pp. 20–22).

[19] T.S.Yu Francis and Yin Shizhuo. *Fiber Optic Sensors*. New York: Marcel Dekker, 2002. ISBN: 0-8247-0732-X (cit. on p. 20).

[20] Andrew B. Lewis et al. "METHOD AND APPARATUS FOR ACOUSTIC SENSING USING MULTIPLE OPTICAL PULSES: Patent Application Publication". GB 2442745B. 2008 (cit. on p. 22).

[21] Armin Schwarz. *Die ET 423 ...* 2014. URL: http://hellertal.startbilder.de/bild/deutschland~triebzuege~br-423-433/369685/die-et-423-332-6--423.html (cit. on p. 24).

[22] Georges Kouroussis et al. "Review of Trackside Monitoring Solutions: From Strain Gages to Optical Fibre Sensors". In: *Sensors (Basel, Switzerland)* 15.8 (2015), pp. 20115–20139. DOI: 10.3390/s150820115 (cit. on pp. 26, 28, 29).

[23] G. Kouroussis, D. P. Connolly, and O. Verlinden. "Railway-induced ground vibrations – a review of vehicle effects". In: *International Journal of Rail Transportation* 2.2 (2014), pp. 69–110. ISSN: 2324-8378. DOI: 10.1080/23248378.2014.897791 (cit. on p. 27).

[24] Sheng-Huoo Ni, Yan-Hong Huang, and Kuo-Feng Lo. "An automatic procedure for train speed evaluation by the dominant frequency method". In: *Computers and Geotechnics* 38.4 (2011), pp. 416–422. ISSN: 0266352X. DOI: 10.1016/j.compgeo.2011.02.003 (cit. on p. 28).

[25] Georges Kouroussis and Olivier Verlinden. "ESTIMATION OF RAILWAY VEHICLE SPEED USING GROUND VIBRATION MEASUREMENTS". In: (2014) (cit. on pp. 28, 37, 38).

[26] J. R. Jimenez and H. R. Wu. "Empirical Mode Decomposition as a tool for data analysis". In: *2011 6th IEEE Conference on Industrial Electronics and Applications*. IEEE, 21.06.2011 - 23.06.2011, pp. 2538–2543. ISBN: 978-1-4244-8754-7. DOI: 10.1109/ICIEA.2011.5976020 (cit. on p. 29).

[27] Norden E. Huang. "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis". In: (1997) (cit. on p. 29).

[28] Alfred Mertins. "Signaltheorie: Grundlagen der Signalbeschreibung, Filterbänke, Wavelets, Zeit-Frequenz,Analyse, Parameter- und Signalschätzung". In: (2013) (cit. on pp. 31, 33, 40).

[29] Shen-Haw Ju, Hung-Ta Lin, and Jeng-Yuan Huang. "Dominant frequencies of train-induced vibrations". In: *Journal of Sound and Vibration* 319.1-2 (2009), pp. 247–259. ISSN: 0022460X. DOI: 10.1016/j.jsv.2008.05.029 (cit. on pp. 37, 38).

[30] Patrick Jahnke. "Machine Learning Approaches for Failure Type Detection and Predictive Maintenance". In: *Technische Universität Darmstadt* (2015) (cit. on p. 39).

[31] Gabriel Rilling, Patrick Flandrin, and Paulo Goncalves. "On Empirical Mode Decomposition and its algorithms". In: *IEEE-EURASIP workshop on nonlinear signal and image processing. Grado, Italy* (2003). ISSN: 1070-9908 (cit. on p. 43).

[32] Pei-Yin Chen, Yen-Chen Lai, and Ju-Yang Zheng. "Hardware Design and Implementation for Empirical Mode Decomposition". In: *IEEE Transactions on Industrial Electronics* 63.6 (2016), pp. 3686–3694. ISSN: 0278-0046. DOI: 10.1109/TIE.2016.2531018 (cit. on pp. 43, 44).

[33] Aurelien Geron. *Hands-On Machine Learning with Scikit-Learn & TensorFlow: Concepts, Tools, And Techniques to Build Intelligent Systems*. United States: O'Reilly Media, 2017. ISBN: 978-1-491-96229-9 (cit. on pp. 52, 53, 55).

[34] Colin Campbell and Ying Yiming. "Learning with Support Vector Machines". In: (2011) (cit. on p. 54).

[35] Richard Byrd, Jean Gilbert, and Jorge Nocedal. "A trust region method based on interior point techniques for nonlinear programming". In: *Mathematical Programming* (2000), pp. 89–149. URL: https://doi.org/10.1007/PL00011391 (cit. on p. 62).

A. Abbreviations

BTM	Balise Transmission Module
CWT	Continuous Wavelet Transformation
DAS	Distributed Acoustic Sensor or Distributed Acoustic Sensing
DFT	Discrete Fourier Transformation
DTS	Distributed Temperature Sensor
DWT	Discrete Wavelet Transformation
EMD	Empirical Mode Decomposition
FN	False Negative
FOS	Fibre Optic Sensor or Fibre Optic Sensing
FP	False Positive
GPS	Global Positioning System
GSM-R	Global System for Mobile Communications-Railway
IMF	Intrinsic Mode Functions
IMU	Inertial Measurement Unit
OTDR	Optical Time Domain Reflectometry
SDB	Standard Deviation Bound

SNR	Signal to Noise Ratio
SVM	Support Vector Machine
TN	True Negative
TP	True Positive

B. List of Figures

C. List of Tables