Thesis to get the degree of a Diplom-Ingenieur

Condition monitoring of roller bearings in a flywheel application

Autor:
Daniel Schürholz, BSc

Graz, February 2021

Graz University of Technology
Faculty of Electrical and Information Engineering
Institute of Electrical Measurement and Sensor Systems

Referees:
Assoc.-Prof. Dipl.-Ing. Dr.techn. Hannes Wegleiter
Dipl.-Ing. Dr.techn. Bernhard Schweighofer

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 dissertation.

Date

Signature

Contents

Abstract

For machines with very high rotational speed, like flywheels, it is challenging to detect bearing faults. Possible causes of vibrations in a faulty bearing are investigated and impulsive excitations due to spalling of the raceway are found to be the most likely causes. A modeling approach was chosen according to the assumption that impulsive excitations obeying a pseudo-cyclostationary process are most likely to represent the occurring damages. A new and inexpensive method (spark erosion) is proposed for repeatably introducing localized damages in the bearings raceway using a capacitor-discharge process. Two bearings are prepared with the spark erosion technique, one with two outer race faults and one with three inner ring faults. A milling spindle was converted into a bearing-test-rig which is mechanically decoupled from its surrounding by using dampening rubber feet (silent blocks). For capturing the signals a piezo acceleration sensor and a incremental encoder is used. Several state of the art techniques for processing and detecting the faulty bearings signal are examined and tested. For quantifying the effectiveness of each kind of signal processing the *SNR-filter* technique is proposed. To stabilize the signal, *order tracking* is introduced which changes the dependence of the vibrational acceleration signal from (sampling) time to angular position. This reduces the spread in the spectrum resulting from speed variations of the spindle. Additionally, also techniques to get rid of the non random periodic components of the signal are evaluated including an auto-regressive *Linear Prediction* algorithm, *Discrete Random Separation* and *Synchronous Averaging*. Because the structure of several decaying oscillations originating from impulsive excitations are overlapping due to the high rotational speed, a deconvolution technique is applied (*Minimum Entropy Deconvolution*). This results in an approximation of the original excitation signal. Also, a technique called *Spectral Kurtosis* is used to determine in which frequency band the impulsive excitations origin, so that the following reconstruction is only performed from these frequency bands. Also *wavelet denoising* is tested for getting rid of the noisy signal. Then the spectrum of the squared signals envelope is calculated. A thresholding technique for quantifying the usefulness of the processing techniques is introduced by comparing the energy content of the processed and the original signal. Finally, the spectra of the processed vibrational signals are compared to the original signals spectrum, demonstrating a clear difference between healthy and faulty bearings.

Kurzfassung

Bei Maschinen mit sehr hohen Drehzahlen wie Schwungrädern ist es eine Herausforderung Lagerfehler zu erkennen. Mögliche Ursachen für Schwingungen in einem fehlerhaften Lager werden untersucht und impulsive Anregungen aufgrund von Abplatzungen in der Laufbahn als die wahrscheinlichsten Ursachen ermittelt. Es wird ein Modellierungsansatz gewählt, der davon ausgeht, dass impulsartige Anregungen, die einem pseudo-zyklostationären Prozess gehorchen, am ehesten die auftretenden Schäden repräsentieren. Es wird eine neue und kostengünstige Methode (Funkenerosion) vorgeschlagen, um mit Hilfe einer Kapazitätsentladung wiederholbar lokale Schäden in die Laufbahn einzubringen. Zwei Lager werden mit der Funkenerosionstechnik präpariert, eines mit zwei Außenringfehlern und eines mit drei Innenringfehlern. Eine Frässpindel wird zu einem Lagerprüfstand umgebaut, der durch dämpfende Gummifüße (silent blocks) mechanisch von der Umgebung entkoppelt ist. Zur Erfassung der Signale wird ein Piezo-Beschleunigungssensor und ein Inkrementalgeber eingesetzt. Es werden verschiedene dem Stand der Technik entsprechende Verfahren zur Verarbeitung und Erkennung des Fehlersignals der Lager untersucht und getestet. Zur Quantifizierung der Effektivität einer Signalverarbeitungstechnik wird der *SNR-Filter* entwickelt. Zur Stabilisierung des Signals wird *Order Tracking* angewendet, welches die Abhängigkeit des Beschleunigungssingals von der (Abtast-)Zeit zur Winkelposition ändert. Dies reduziert die Streuung im Spektrum, die durch Drehzahlschwankungen der Spindel entsteht. Dann werden einige Techniken geprüft, um die nicht zufälligen periodischen Komponenten des Signals loszuwerden. Namentlich wurden *autoregressiv Linear Prediction*, *Discrete Random Separation* und *Synchronous Averaging* getestet. Da sich die von impulsiven Anregungen herrührenden Strukturen abklingenden Schwingungen aufgrund der hohen Drehzahl überlagern, wird eine Entfaltungstechnik angewendet (*Minimum Entropy Deconvolution*). Dadurch erhält man eine Approximation des ursprünglichen Anregungssignals. Dann wird eine Technik namens *Spektral Kurtosis* verwendet, um zu bestimmen, in welchem Frequenzband die impulsiven Erregungen liegen, und es dann nur aus diesen Frequenzbändern zu rekonstruieren. Auch *Wavelet-Denoising* wird getestet, um das Signalrauschen zu reduzieren. Dann wird das Spektrum der quadrierten Signalhüllkurve berechnet. Eine Schwellenwerttechnik zur Quantifizierung der Nützlichkeit der Verarbeitungstechniken wird eingeführt, indem der Energiegehalt vom verarbeitete Signal mit dem Originalsignal in verschiedenen Frequenzbereichen verglichen wird. Schließlich werden die Spektren der bearbeiteten Vibrationssignale im Vergleich zu den Originalsignalen gezeigt, die einen deutlichen Unterschied zwischen gesunden und fehlerhaften Lagern sichtbar machen.

1 Introduction

1.1 Motivation

Flywheels can be of advantage in applications where a high power density, a long life cycle, the availability of the raw materials and the certain knowledge of the storage's health status are of importance. Due to the high power density it can be used wherever buffering of power surges is required. Examples might be:

- To bridge the time gap between the occurrence of a blackout and the finished start up of an emergency power generator. In many cases nowadays this is achieved with batteries. Compared to flywheels, batteries have the significant drawback, that the health status and also the stored energy is difficult to determine without charging and discharging the battery at least partially. For flywheels it is more obvious: If they rotate, they have stored energy and the amount of stored energy is easily calculated by $E = I \cdot \omega^2/2$.
- To improve the feed-in quality of renewable energy power plants by smoothing the fluctuation of the available energy quantity. This could reduce the regulatory effort in the supply grid.
- To buffer the energy demand of sudden high loads, like quick chargers for electric vehicles.

In future the energy density will improve with further advanced flywheel materials which inherently will increase the demands on most of the other components which leads to the need of a *conditioning monitoring system*. One part of this system is the surveillance of the health status of the roller bearings which this thesis has its focus on.

1.2 Aims and considered questions

The aim of this thesis is to investigate how the health status of roller bearings can be monitored in the case of a flywheel with the focus on vibration analysis. What is unique in this case of application compared to already existing bearing monitoring cases is the very high rotational frequency of up to 30,000 rpm and the relatively small load on the bearings, because a magnetic bearing primarily does all the heavy lifting. However all the measurements were made on a milling spindle rotating at

similar speeds (15,500 rpm) with approximately the same bearing preload of ~300 N. In the course of this work the following problems were dealt with:

- A test rig was set up using acceleration sensors, a milling spindle, a frequency inverter and a data acquisition system.
- Bearing faults purposely introduced in healthy bearing raceways.
- Promising advanced state of the art signal processing techniques were examined to detect inner and outer ring raceway damages. The most suitable signal enhancement techniques were adapted and implemented, and a proof of concept was made, that the introduced bearing faults can be detected.

2 Basics

Here only raceway damages will be considered, because due to the hardness of the ceramic rollers in the flywheel application, a damage of the rollers is rather unlikely. Also any kind of electric corrosion due to currents flowing over the rollers will not be considered, because the bearing is assumed to be isolating. This might be wrong after long times of operation when some metallic abrasion contaminates the lubricant and therefore increases the conductivity of the lubricant enough to permit a current flow, but even then it is assumed to be small enough to be negligible.

2.1 Sources of vibrations

As stated in [1] sources of vibration can be elastic deformation, surface roughness and waviness.

- Elastic deformation: Due to a discrete number of rollers, the load is distributed among them and the outer and inner ring get mechanically displaced due to elastic deformation when a roller passes by which results in an inevitable source of vibration. (Figure 2.1)
- Geometric imperfections:
  - *Surface Roughness*: Here surface asperities can poke through the lubrication film and lead to a direct contact between the materials of the raceway and the rolling element. The resulting vibrations can be considered as random sequence of small impulses which excite the modes of the mechanical structure. (Figure 2.2)
  - *Waviness*: If surface features have a longer wavelength than the Hertzian contact area, the roller will follow the waviness of the raceway and vice versa which leads to vibrations. (Figure 2.3)
  - *Discrete defects*: These defects can occur due to damage in assembly, contamination of any kind, wrong mounting, poor maintenance or excessive operation. Defects behave in a similar way over their lifetime as can be seen in Figure 2.4. First of all, small particles spall from the raceway. Ideally a spherical roller and the raceway would only touch in a single point, but due to elasticity both structures deform slightly and touch in a small contact area. If the bearing is insufficiently dimensioned, the occurring Hertzian stresses a few tenths of a millimeter under the raceway

surface could be high enough to result in cracks because of material fatigue. When these cracks reach the surface, a chunk of material might break loose, which leaves a dent called a *pitting*. A new pitting is most likely little enough to result in an impulse-like excitation of the whole bearing structure when a rolling element clatters over it. These impacts act like a broadband excitation of the coupled mechanical structure which will ring at its resonant frequencies. Over time the pittings will grow due to the continuous bumps with the hard rollers at their exiting edge and they become *extended spalls*. Subsequently, the edges of the dent are rounded off due to the permanent impacts and the excitation will loose some peakiness. [3, 4]
A rule of thumb if bearing defects are detectable comes from [3], stating, that the Equation 2.1 should hold, which is easily achieved in the case of the milling spindle or the flywheel.

$$n \cdot D > 30,000 \rightarrow n > \frac{30,000}{D} = \frac{30,000}{35.7} \approx 800\,rpm \tag{2.1}$$

n ... rpm
D ... Pitch circle diameter of the bearing in mm

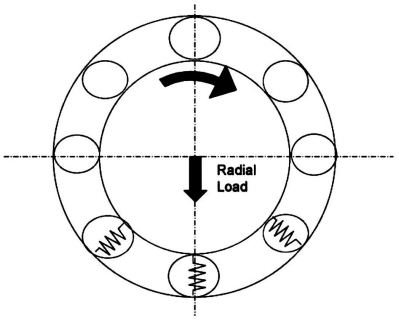

Figure 2.1: Elastic deformation of the bearing elements [1]

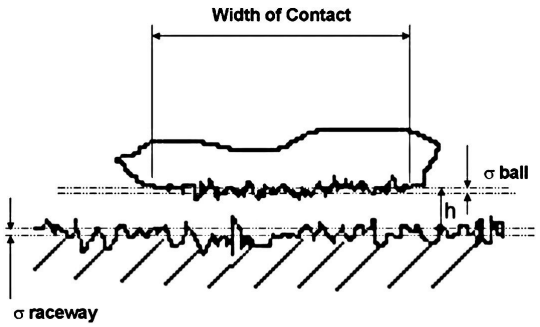

Figure 2.2: Geometric imperfection: Surface roughness due to manufacturing process [1]

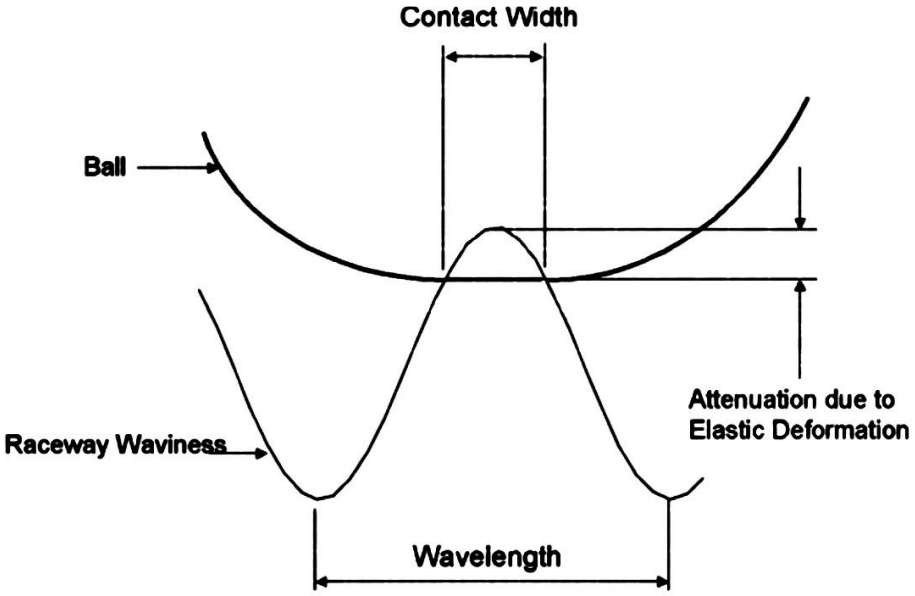

Figure 2.3: Geometric imperfection: Waviness due to the manufacturing process [1]

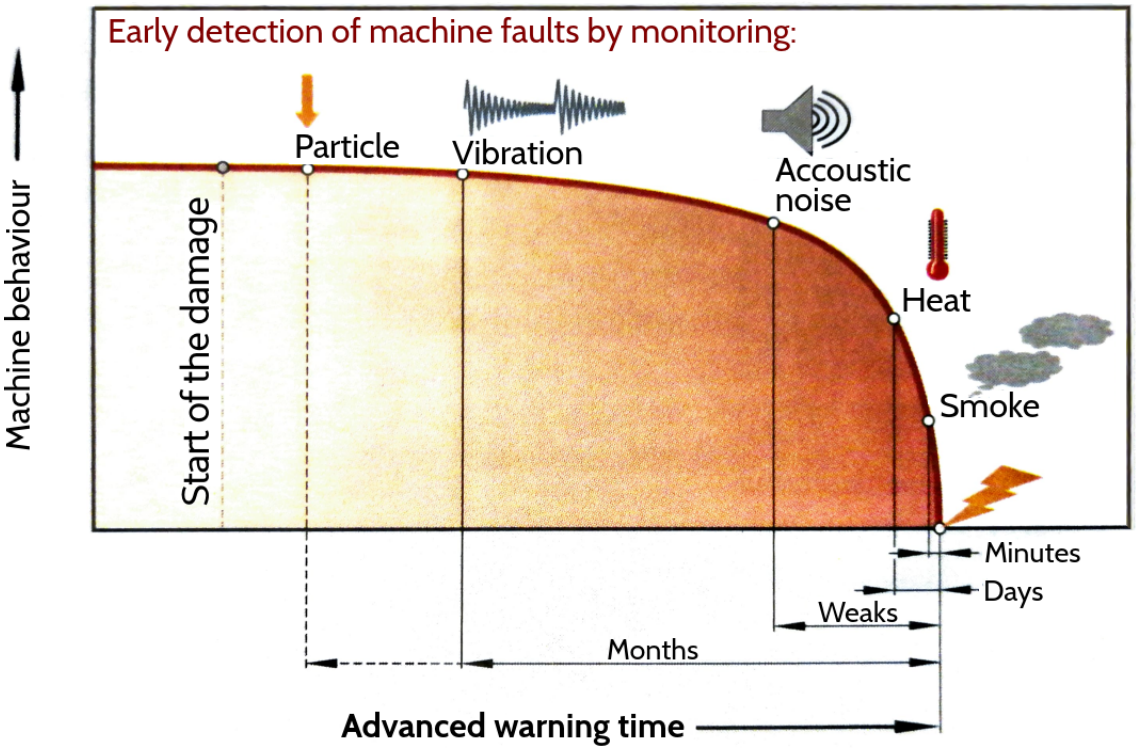

Figure 2.4: Usual development of defects over time. When no analysis of the lubricant is possible, the vibrations are the first signs of damage which can be detected by an acceleration measurement based condition monitoring system. (Adapted from [3])

2.2 Modeling of the Signal

There are three kinematic equations which describe the most relevant frequencies in the case of a raceway damage with neglected slip [4]:

$$FTF = \frac{f_S}{2} \cdot \left(1 - \frac{d}{D} cos(\alpha)\right) \tag{2.2}$$

$$BPFO = N \cdot \frac{f_S}{2} \cdot \left(1 - \frac{d}{D} cos(\alpha)\right) = N \cdot FTF \tag{2.3}$$

$$BPFI = N \cdot \frac{f_S}{2} \cdot \left(1 + \frac{d}{D} cos(\alpha)\right) \tag{2.4}$$

FTF	...	**F**undamental **T**rain **F**requency
BPFO	...	**B**all **P**ass **F**requency **O**uter raceway
BPFI	...	**B**all **P**ass **F**requency **I**nner raceway
N	...	Number of rolling elements in the bearing
f_S	...	Rotational frequency of the shaft
d	...	Diameter of the rolling element
D	...	Pitch circle diameter of the bearing
α	...	Load angle of the bearing

If now an acceleration sensor would be put on the outer ring of a defected bearing one should theoretically measure the signals shown in Figure 2.5. Because the localized fault on the inner ring moves relatively to the sensor and only the radial measurement direction is considered, the acceleration for an inner ring fault signal originates from the impacts being amplitude modulated by the angular position of the shaft. In comparison, the outer ring does not move, therefore the impact of the localized outer ring fault is not modulated.

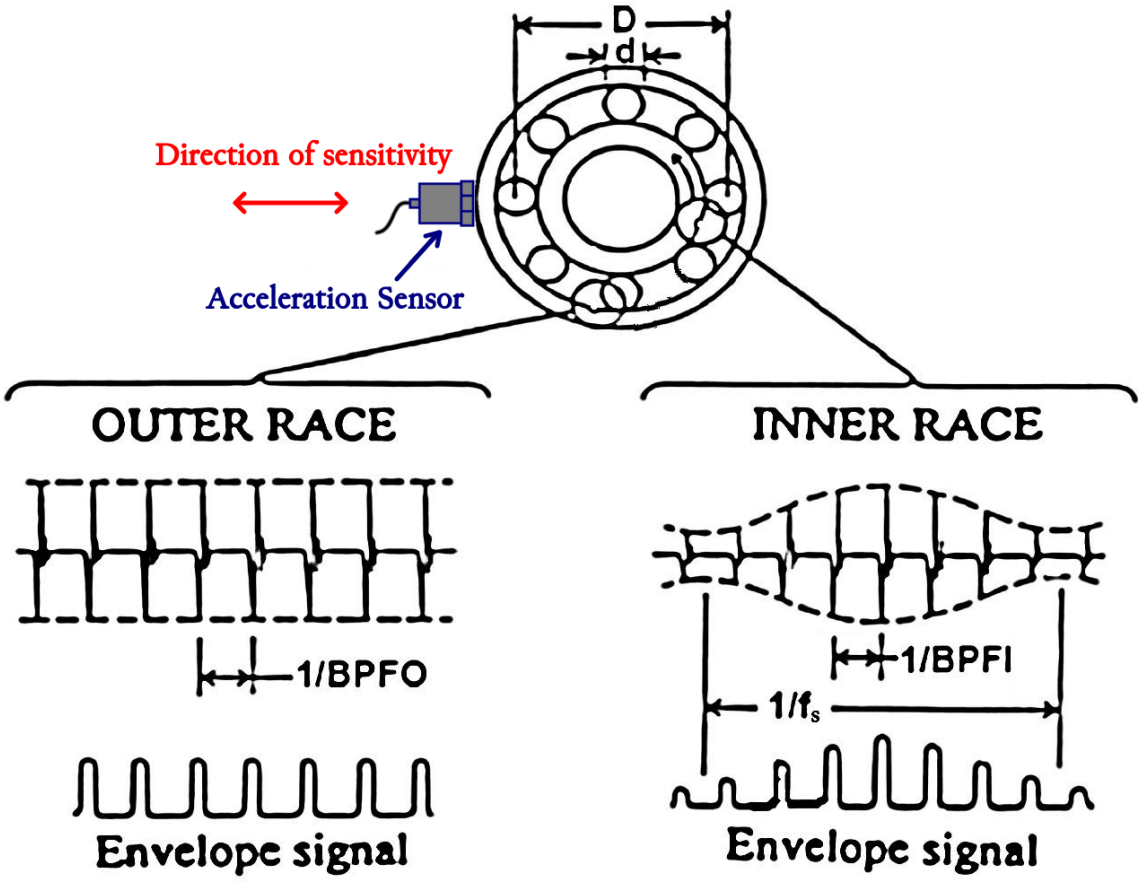

Figure 2.5: Acceleration signal of different defects. Adapted from [2]

2.3 Localized faults

In a roller bearing, the rollers underlie a combination of rolling and (abrasive, unwanted) sliding called *slip*. If properly preloaded, rolling should be by far the major component, but still some slip exists. The amount of slip is not constant, but changes with the variation of the load on a rolling element - hence the variation of load as shown in Figure 2.6 changes the slip in relation to the turning angle of the shaft. Considering slip and the fact, that the system does not have any memory about the position of the rollers and the cage, it is proper to assume, that the time gap between two impacts is considered as random variable. Equation 2.5 shows the older approach to model the moments of impacts as cyclostationary which turned out to be wrong as claimed in [4, 7]. The problem is, that due to the slip the moments of impact are not spaced randomly around a known mean value as assumed in Equation 2.5. Rather the time *duration* between impacts changes due to slip (like in Equation 2.6) where the process is called pseudo-cyclostationary which has the property, that the exact position of the rollers and the cage can never be known not even their "mean value". When in the section 5.7 the squared envelope spectra is chosen over the spectral correlation as final tool to display the measured data, there is however little advantage of considering the time spacing process as pseudo cyclostationary rather than cyclostationary, as can be seen in Figure 2.7 where the difference is marginal enough to be ignored. This is the reason, why for further investigation the process is assumed to be cyclostationary and the squared envelope spectra will be the tool for further investigation.[4]

$$T_i = iT + \delta T_i \tag{2.5}$$

$$\Delta T_i = T_{i+1} - T_i \tag{2.6}$$

i ... i-th impact
T_i ... time of i-th impact
δ ... random deviation
ΔT_i ... time difference between i-th and next impact

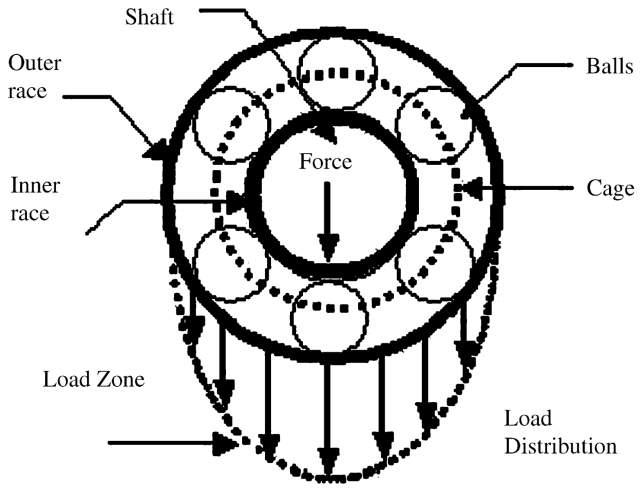

Figure 2.6: Load distribution and bearing components [2]

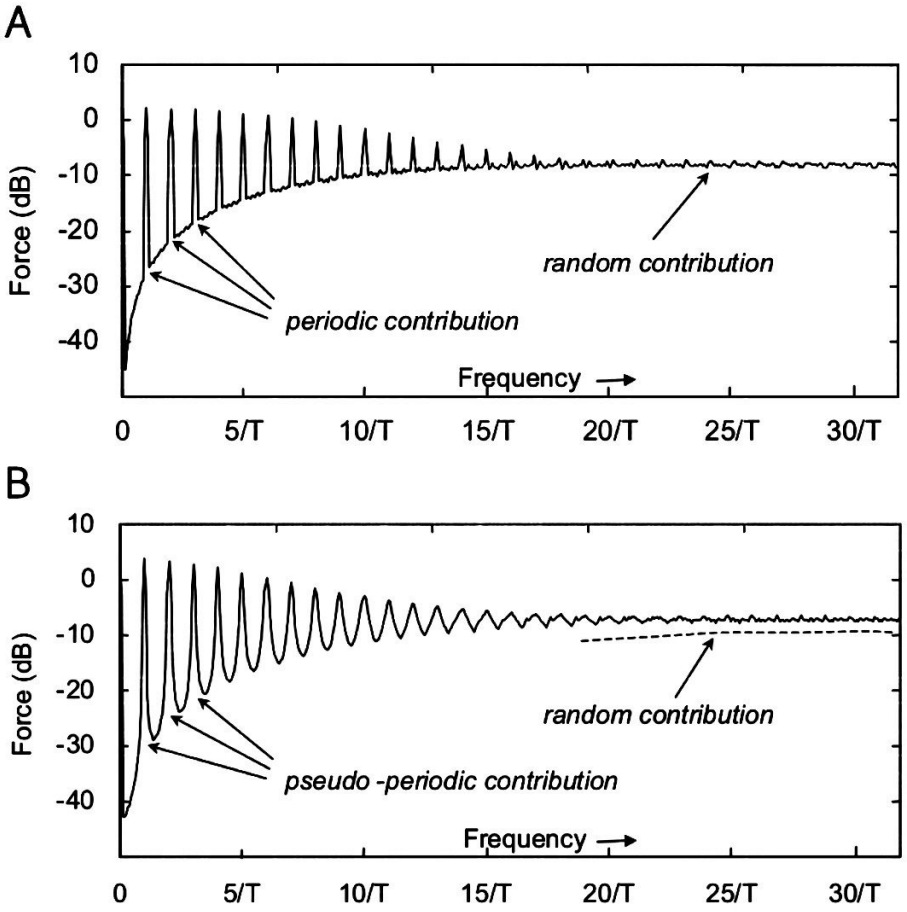

Figure 2.7: *A:* Frequency Spectra comparison of model 1 (Equation 2.5) and *B:* model 2 (Equation 2.6) [4]

2.4 Extended Spalls

Often rollers will have an impulsive bump when they exit an extended spall which is the reason why the envelope analysis is most likely going to detect the fault.

As time passes by the spalls loose their edges and therefore the impacts are much smaller than in the early stages. The signal from smoothed or edgy extended spalls can be considered as a modulated signal as in Figure 2.8 containing a first order cyclostationary component (CS1 - local mean value) and a second order cyclostationary component (CS2 - amplitude modulated noise). The CS1 component can be interpreted as having a periodic mean value with some added noise coming from sources like e.g. rotor imbalance or for machines with gears the meshing noise. The CS2 component comes from the extended spall noise being modulated by the CS1 process. Because the mean value of the CS1 process is periodic it can be approximately predicted and removed (section 5.3) leaving only the CS2 component which can only come from the extended spall. If no modulation takes place because there is too little imbalance or there are no gears in the structure, it will not be possible to detect smoothed extended spalls.[4]

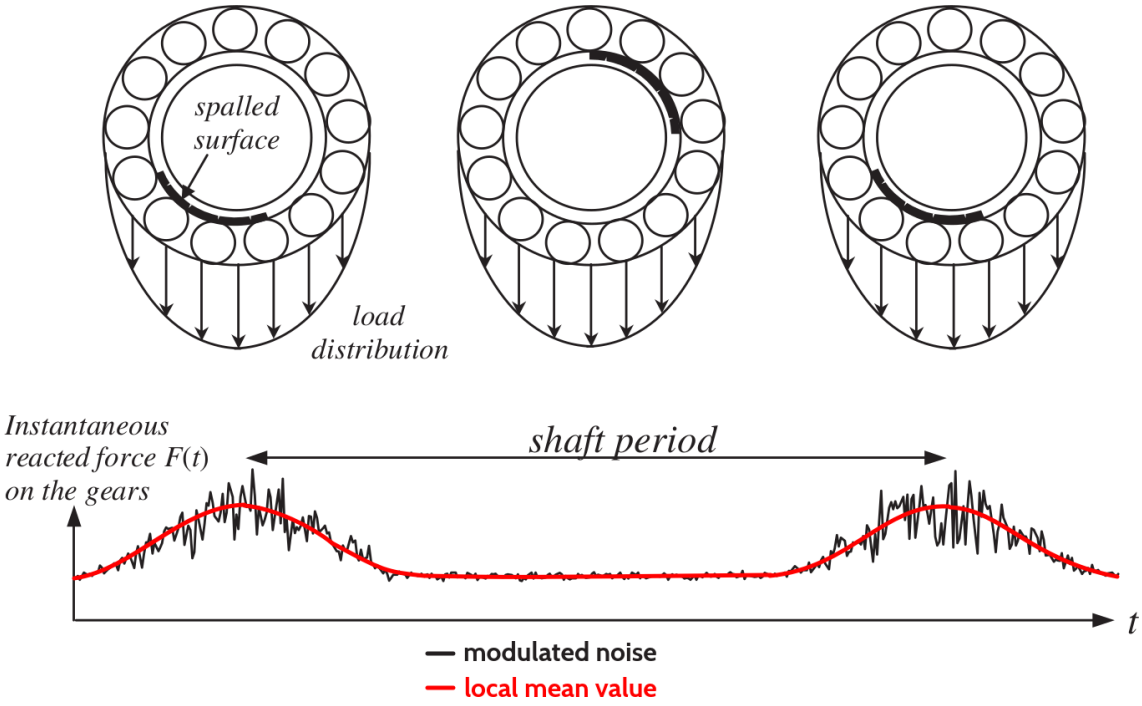

Figure 2.8: Modulating signal from the effect of an extended inner race fault on a gear signal. Adapted from [4]

2.5 Model

With the shown knowledge a signal model can be built as shown in Figure 2.9. The occurring defects are assumed to result in impulses denoted as g_k which are randomly distributed over time because of the processes (pseudo-) cyclostationary nature. The periodic part denoted as d_k describes the deterministic periodic vibrations due to e.g. imbalance of the rotor. The transmission path denoted as h_k describes the mechanical structure transfer function between the source of impact and the accelerometer which measures the vibration signal x_k. The green box called *Filter* contains one of the methods described in section 5.3 which help to get rid of non random periodic signal components. The remaining signal component ε_k denotes the random signal component after it propagated through the mechanical system. The blue box called *MED Filter* contains the algorithm adapted from [8] which is described in section 5.4 and tries to deconvolute ε_k such that the original impulsive excitation signal g_k is regained. The orange box called *Enhancement filter* contains techniques described in section 5.5 and section 5.6 which demonstrate how to remove noise from the signal and how to determine the most impulsive frequency band of the spectra. finally, the pink box called *Squared Envelope Spectrum* represents an algorithm which uses the Hilbert transformation to calculate the squared envelope spectrum which has advantages in terms of interpretability. [4]

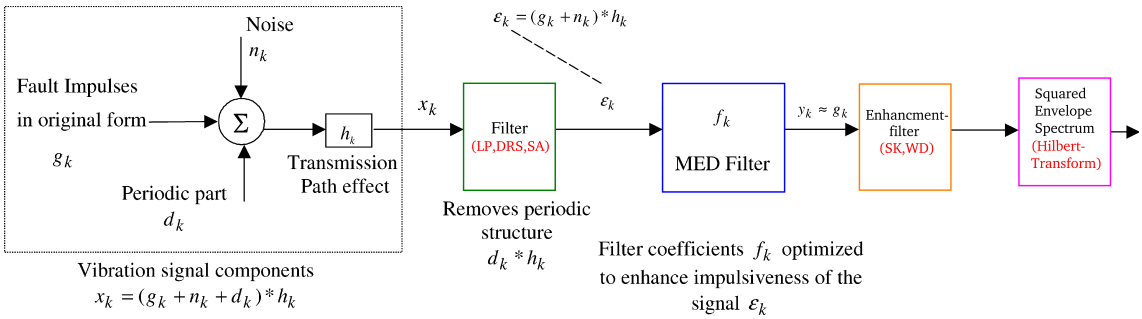

Figure 2.9: Model of the vibrational signal with its conditioning filters [2]

3 Introduction of faults

Because damaged bearings are obviously not readily available for sale, the question arises how to introduce representative damages in the bearings raceway.

3.1 Electric-discharge-machining (EDM)

A theoretically ideal way would be to weaken a healthy bearings ring by eroding a thin protrusion in its backside until in a reasonable time of operation material fatigue would lead to a realistic damage. An advantage of EDM is, that it does not affect the heat treatment of the machined material. Then the bearing should be operated until the damage occurs due to material fatigue. However because neither the know-how nor the machines were available to perform an EDM-weakening of the bearing at that time, two alternatives (section 3.2, section 3.3) were developed and tested to make localized faults.

3.2 Denting the raceway by plastic deformation

For this technique a dent was ground in the outer ring of the raceway, until only little material was left supporting the inner raceway as can be seen in Figure 3.1. Then the idea was to use a center punch to deform the thinned out area of the ring, such that the raceway gets bulged. This technique was not successful at all, because the center punch was softer than the hardened bearing material. Even though the bearing was clamped during the center punch strike very close to the ground spot, the whole ring got deformed into an slightly oval shape (in the order of a few thousands of a mm). The raceway did not bulge in any noticeable way, and no impulsive excitation was measured - therefore this technique is considered to be a fail.

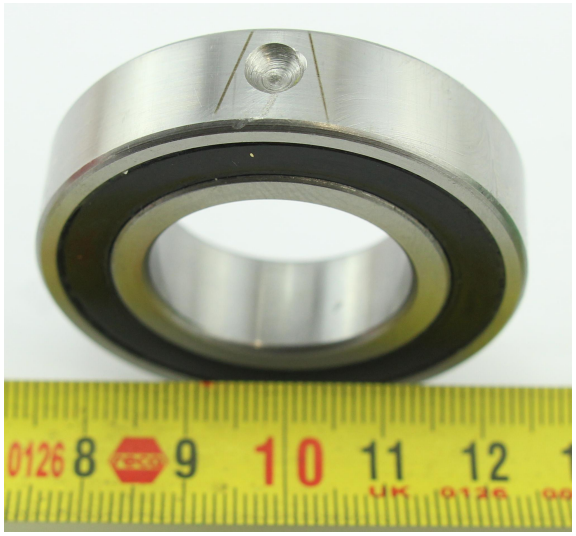

Figure 3.1: Failed attempt to damage the outer raceway: A weak spot was ground into the bearing. Then the weakened spot was tried to be center punched but the whole outer ring most likely warped. It could not be measured with the 10-micron-indicator, but when testing this bearing in the milling spindle it lead to a lot of rattling noise without any impulsive excitation.

3.3 Spark eroding damages directly into the raceway

First of all the bearing were disassembled with the help of a machined aluminum sleeve which matches the outer ring of the bearing and a machined aluminum cylinder matching the bearings inner ring as can be seen in Figure 3.2 and Figure 3.3. This aluminum sleeve is preheated to approximately 70°-100° Celsius and then the bearing is put in. Due to the heated sleeve the outer ring expands until the inner ring could be easily pressed out without much force. It was crucial not to overheat the aluminum, so that the outer ring would not be tempered. Now the bearing components was cleaned from grease, whereby it is important to mention, that some cage composites could get damaged by using the wrong degreaser.
In the next step, to find out the right amount of charge and voltage to make consistent dents with the wanted diameter, a capacitor was used to spark-erode dents into an old bearing ring. The idea came from electric-discharge-machines which use this principle to machine basically any metal regardless of its hardness without tempering it. When the charged capacitor touches the metal, a short-circuit-current-surge evaporates metal from the bearing ring and from the capacitors wire. Like in electric welding it was assumed, that the electrode with positive polarity melts faster, because electrons - which are the major component of charge carriers in this case - accelerate in the sparks electric field and impact into the positive electrode deploying their kinetic energy into it. Therefore the bearing ring was connected to the positive polarity. Also in spark eroding the work piece and tool are submerged in a non conductive fluid which helps to flush away metal residues which are not able to solidify somewhere else on the work piece which is the reason, why the area of the bearing which should be damaged was covered with oil during spark erosion. Distilled water was also tried and works just as fine. The spark eroding setup for testing can be seen in Figure 3.4. The damage only protrudes into the material - as wanted - and does not show any built-up, as was checked with the measurement setup in Figure 3.5. If the indicator needle would move up while pushing it over the spark eroded dents in the surface one could measure the buildup. However no movement of the indicator needle in the upward direction was detected which leaves one with the conclusion that no buildup exists. The result of the erosion test with the chosen parameters can be seen in Figure 3.6. It is advised to wear safety goggles during spark eroding, because tiny oil drops are spattered everywhere.
Then the bearing was reassembled, lubed and then after a 15 minutes run-in period at $f_S = 400\ Hz$ the bearing was tested.

The finally damaged outer- and inner ring can be seen in Figure 3.7 and Figure 3.8.

Figure 3.2: Aluminum sleeve and press-out-cylinder for disassembling the bearing.

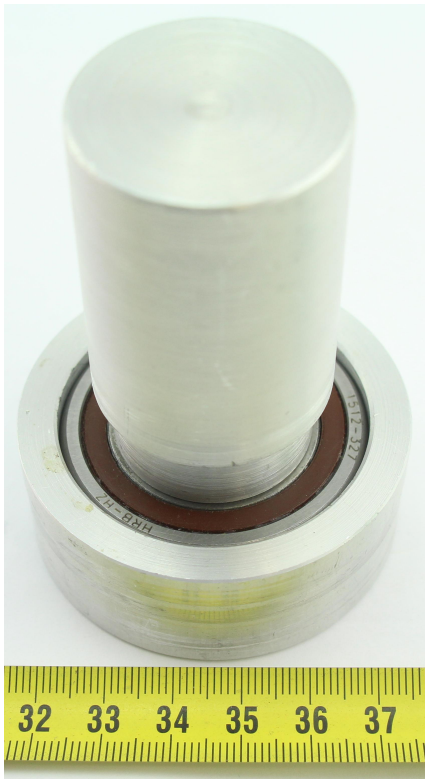

Figure 3.3: Aluminum sleeve and cylinder ready for disassembling the bearing. The sleeve should be preheated to approximately 70°-100° Celsius.

Figure 3.4: Spark eroding test to identify the ideal voltage and capacity values.

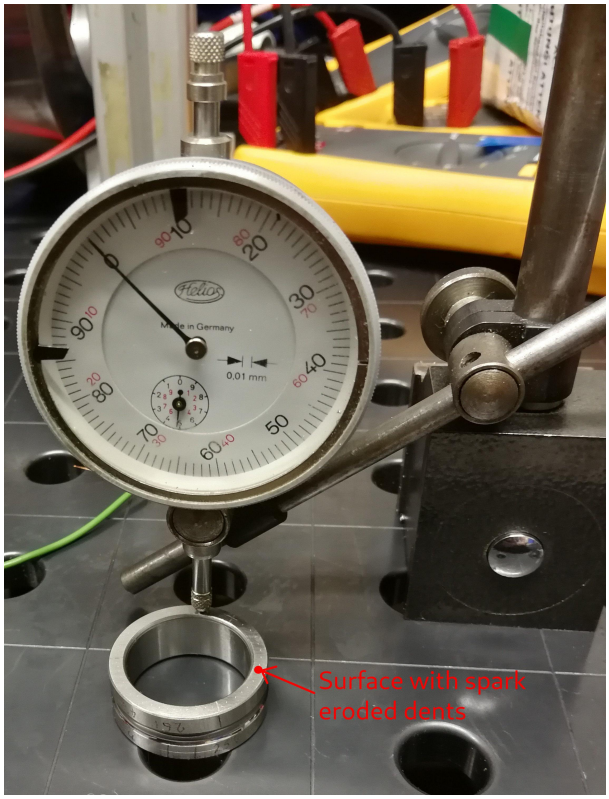

Figure 3.5: Checking if the spark erosion lead to any buildup: No significant signs of buildup were detected.

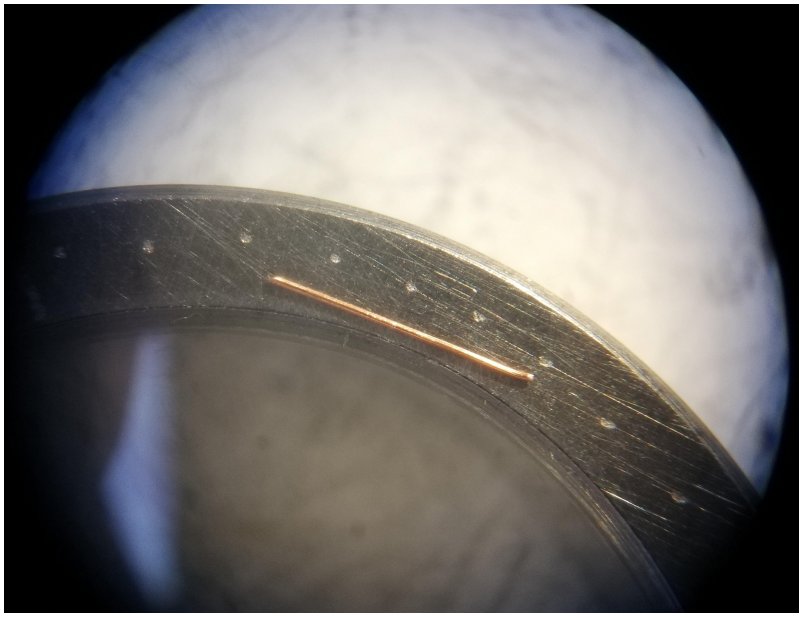

Figure 3.6: Best results of spark erosion using a 47 µF capacitor charged with the power supplies maximum voltage of 32 V. Results might change with capacitors having a different ESR. The copper wire has a diameter of 0.2 mm.

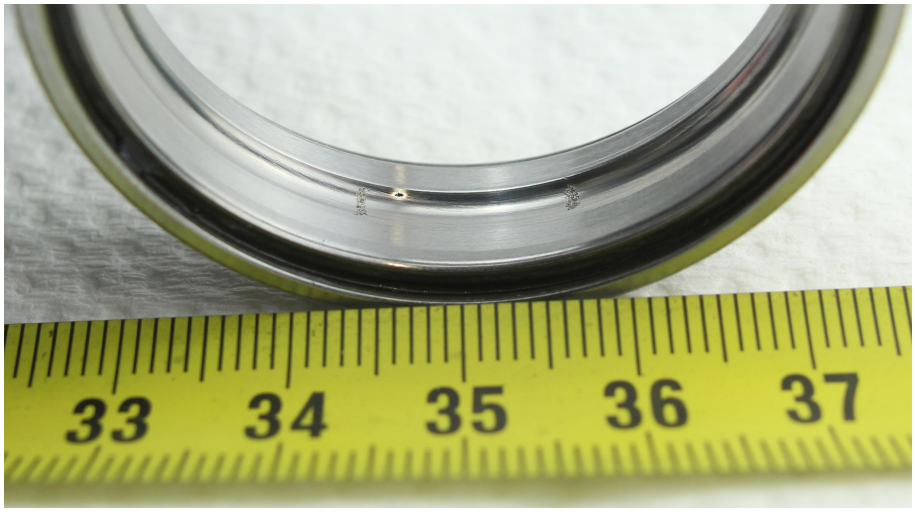

Figure 3.7: Outer ring damage. Two spark eroded spots with parameters: Capacity = 47 µF @32 V.

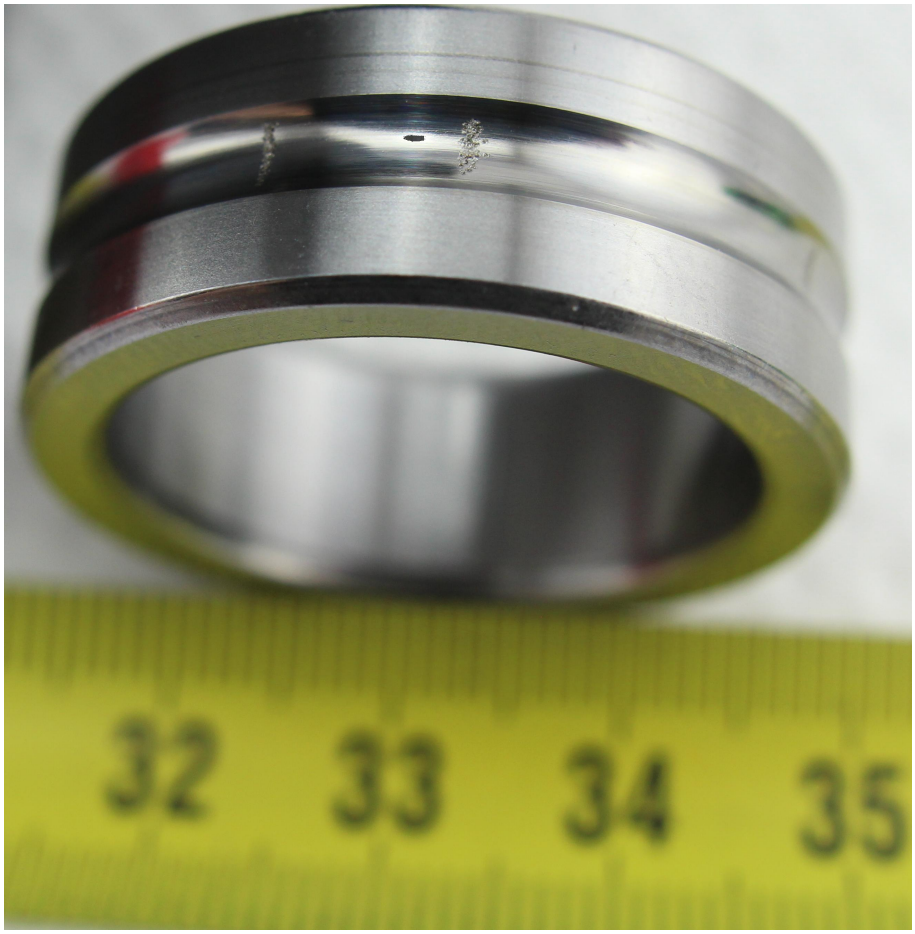

Figure 3.8: Inner ring damage. Three spark eroded spots with parameters: Capacity = 47 µF @32 V.

4 Test rig

The setup of the used test rig can be seen in Figure 4.1. In this chapter the important components of the test rig are described in more detail.

4.1 Data-acquisition-system (DAQ) and frequency inverter (FI)

For the DAQ the *NI-cDAQ-9172* unit from the manufacturer *National Instruments* was used. The measurement module for the acceleration and force measurement was the *NI-9231* module which already has a IEPE-interface. An IEPE-interface means that an IEPE-compatible sensor is already supplied and the signal is amplified by the measurement module. The sample rate of the system is 51.2 kHz with a resolution of 24 Bit.
For controlling the automated test cycles and saving the data to the desired directory on the PC, *NIs* software *LAB-VIEW* was used. To control the *MITSUBISHI-F700* frequency inverter the analog output card *NI-9263* outputs a voltage between 0-5 V. The frequency inverter has an analog input interface which translates the 0-5 V to an output frequency between 0-400 Hz.
Furthermore, the frequency inverter was programmed with the motor parameters and was set to current vector control. This helped overcome issues with slip even though the inverter has no position measurement connected which would make the current vector control very effective. But because the motor was running without any load, it was sufficient to operate the inverter without any feedback of the motor speed. The frequency offset was less then 0.5 % at any speed. An automatic test cycle was implemented which stepped the spindle frequency between 100 Hz and 380 Hz. After reaching the frequency, the motor had a run-in period of approximately 5 seconds, after which the measurement took place.

4.2 Milling spindle

For testing the bearings a water cooled milling spindle which could spin up to 24,000 rpm was modified. Even though the milling spindle might not have needed cooling, because it was operated under no load, the cooling system was still used just to be safe.

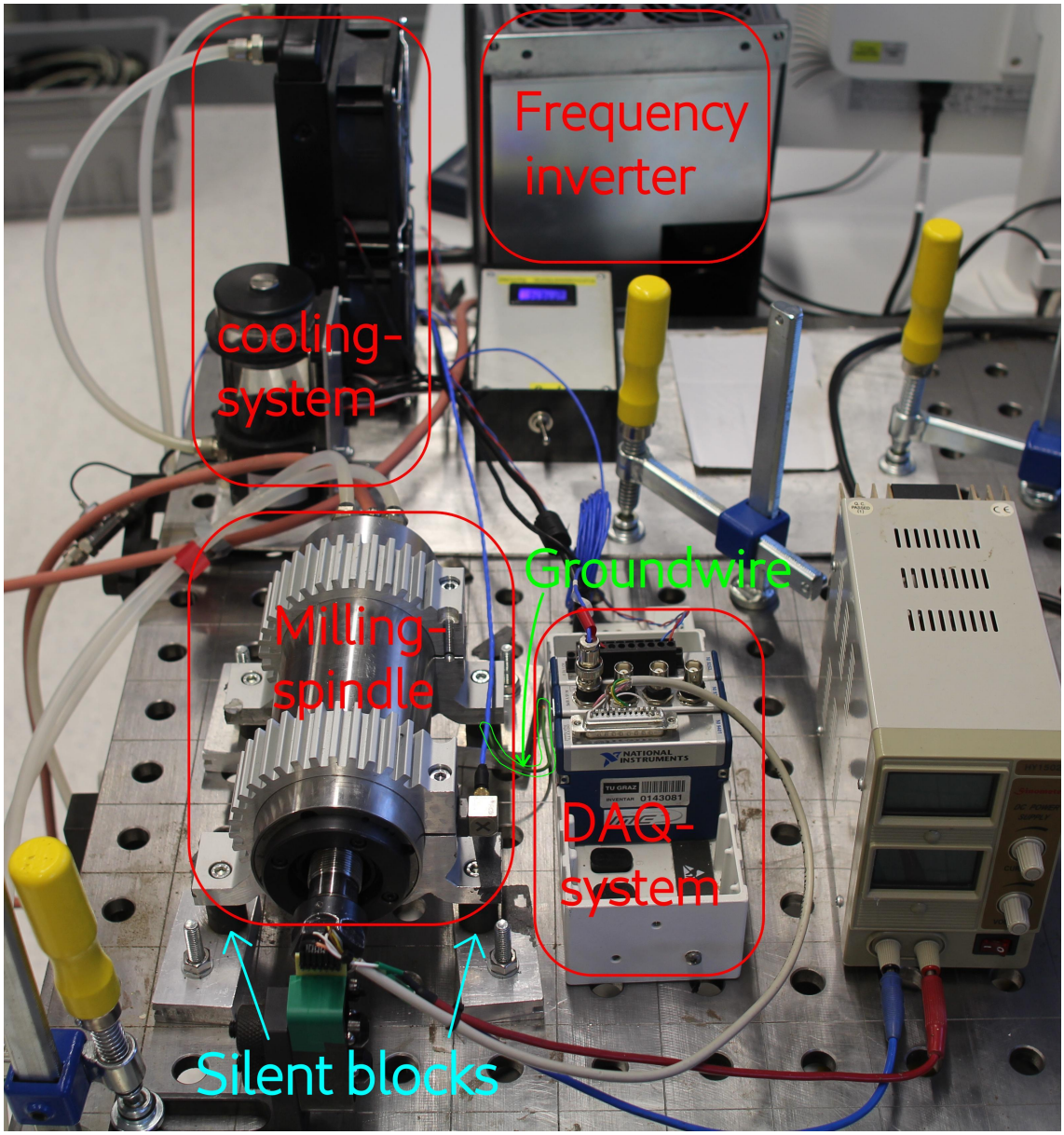

Figure 4.1: Setup of the test rig.

In its original configuration the spindle had two bearings in the front and two bearings in the back of the spindle shaft, as can be seen in Figure 4.2. Because more than one bearing at each support point will lead to unwanted noise, the goal was to remove one bearing in the front and one in the back and to find a way to preload the bearings as wanted. This was done by using a wave spring combined with a preload-adjustment-spacer and a spacer in the back as can be seen in Figure 4.3. By varying the preload-adjustment-spacers thickness the space l_p (shown in Figure 4.3) changes which varies the preload - with the used spacer a preload of approximately 300 N was set. After changing the preload-adjustment-spacer, the front cap has to be tightened down properly to squeeze the wave spring down into the space l_p which results in a preload F_p according to Equation 4.1:

$$F_p = k_{wave} \cdot (l_0 - l_p) \tag{4.1}$$

F_p ... preload force of the bearing
k_{wave} ... spring constant of the wave spring
l_0 ... length of the relaxed spring
l_p ... length of the compressed spring

The milling spindle is mounted to a table with a mounting bracket and silent blocks (rubber feet in Figure 4.1). The silent blocks help to mechanically decouple the motor from the table and allow for a low noise measurement surrounding. Because the silent blocks are isolating the spindle housing from the mounted table, the housing had a floating potential and a stray capacity towards its surrounding. The stray capacity was charged up to grid potential from capacitive currents driven by the frequency inverters switching process. Two input stages of the accelerometer measurement module were most probably destroyed because of the unwanted capacitive currents.
To eliminate the problem, the voltage between the motor housing and the grids ground level was measured - 230 V_{RMS}. Then it was tried to find out, if the inner resistance of the driving source of this currents has a small or a big resistance - hence if it is dangerous to short circuit it to ground or if some other measure has to be taken. For checking the inner resistance of the source a 1 $k\Omega$ resistance was connected to ground and the current was measured. Because the voltage dropped significantly down to a few volts and less than 10 mA_{RMS} current was flowing - which implies a high inner resistance of the interfering source, it was decided, that it is proper to short circuit the housing with the grid-mass using a ground-wire, as can be seen in Figure 4.1. In the subsection 4.3.1 it is further described, how the measurement was influenced by these capacitive inverter currents and how the mounting method of the sensor helped to avoid disturbances.

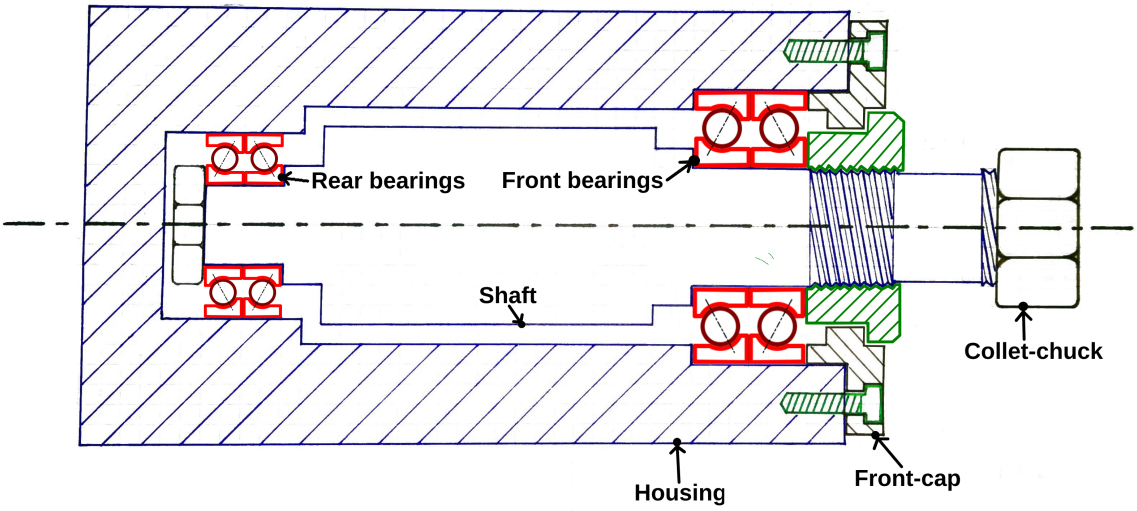

Figure 4.2: Sketch of the original milling spindle.

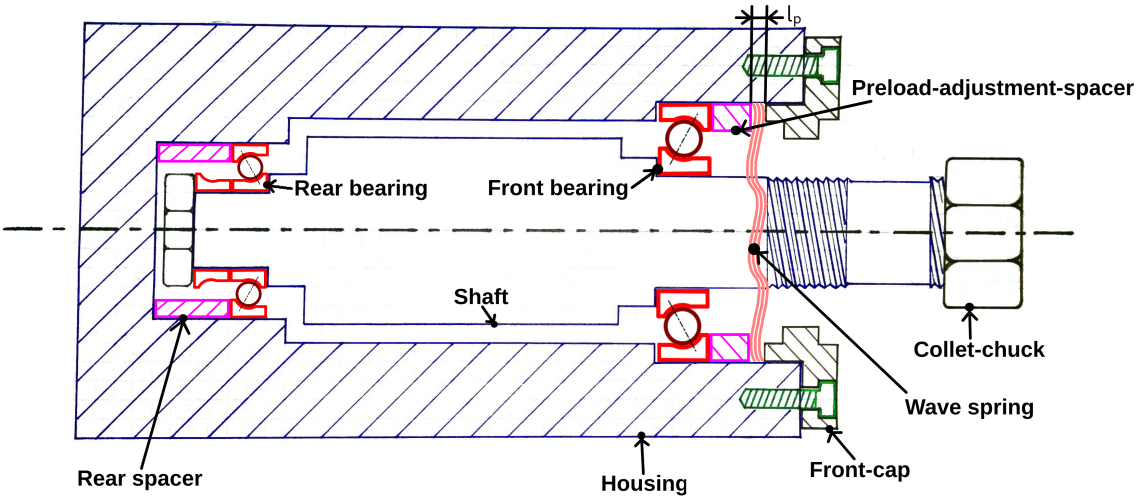

Figure 4.3: Sketch of the modified milling spindle. By varying the preload-adjustment-spacers thickness, the leftover space l_p for the wave spring after tightening down the front-cap changes varying the preload according to Equation 4.1.

4.3 Sensors

4.3.1 Acceleration Sensor

According to the data sheet, the used piezo-based acceleration sensor *PCB-317A41* is suitable up to frequencies of 10 kHz which is assumed to be sufficient to detect enough significant high frequencies of the signal. This sensor might not work in the flywheel application, because it is not specified for the use in vacuum. Even though sensors might be specified as *hermetically sealed*, they often are not specified for the use in vacuum, because vacuum specification also means, that the sensor does not gas out and will therefore not contaminate the vacuum surrounding. In the case of the flywheel it would actually be of no importance if the sensor leaks any gasses as long as it measures correctly, but because the manufacturer won't guarantee its operation as long as it is not specified for vacuum, one has to choose the much more expensive vacuum specified sensors.
To check whether the sensor is within specification, a hammer-tap-test was done. The working principle of the test can be made clear by Newton's second law as can be see in Equation 4.2: Because the hammers excitation force F_H is linked to the acceleration of the sensor a_S via a known mass m_{known}, the transfer function H between the hammers acceleration and the sensors acceleration should be 1 or at least constant if the gain has some offset. The setup of the hammer-tap-test can be seen in Figure 4.4 and its measured results can be seen in Figure 4.5 and Figure 4.6.

$$\mathscr{F}\left(F_H = m_{known} \cdot a_S\right) \rightarrow \frac{a_S\left(f\right)}{\frac{F_H(f)}{m_{known}}} = \frac{a_S\left(f\right)}{a_H\left(f\right)} = H \stackrel{!}{=} 1. \tag{4.2}$$

Looking at Figure 4.6, one can see that the uncalibrated sensor works within the range of 10 kHz quite OK and it should be fine to use: The transfer-function's absolute gain is about 6 dB lower than it should be, but the change of gain is within 1 dB which is good enough. The phase shift with less than 10° deviation within the 10 kHz range is fine too. The gain-level could be adjusted, but this is not important for qualitative measurements.
Two ways of mounting the sensor were investigated. In the first case the sensor is connected with a stud to the motor mounting bracket, but capacitive currents from the frequency inverter lead to a noisy signal, because the current loop was closed over the sensors connection cables mass shielding. Short circuiting these currents within the motor was only partly successful. The second attempt was to mount the sensor isolated. Therefore the sensor is stuck onto the motor bracket using wax. Compared to stud mounting, in theory wax has a little stronger low pass behavior [12]. The mounted sensor can be seen in Figure 4.9. Nevertheless, in this case the mounting was not limiting the high frequencies up to the excited range to any significant extent as can be seen in Figure 4.7 and Figure 4.8, which shows some of the resonance frequencies of the test rig's mechanical structure excited by hammer tapping the motors hull opposite to the sensor. One has to mention, that depending on the

position of the sensor and the strike position of the hammer, different components are excited, and therefore different resonance frequencies might appear. Furthermore, the hammer has a *sinc*-like excitation in the frequency domain, which does not excite the very high frequencies significantly. So this experiment should be considered as demonstration to give an idea where resonance frequencies of the structure might be.

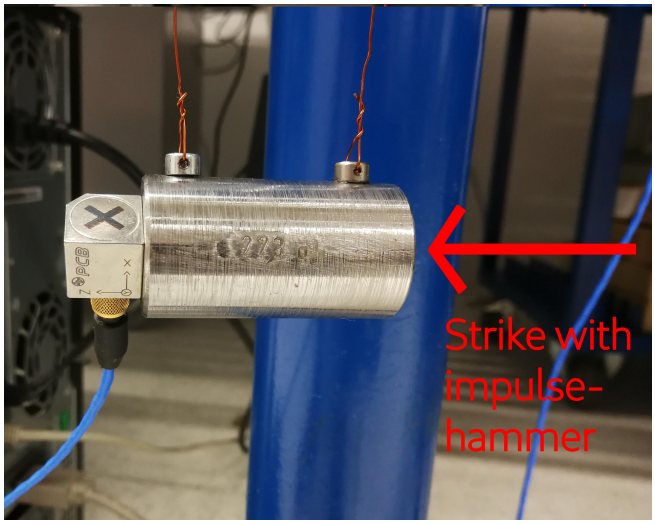

Figure 4.4: Hammer-tap-test for qualitative calibration.

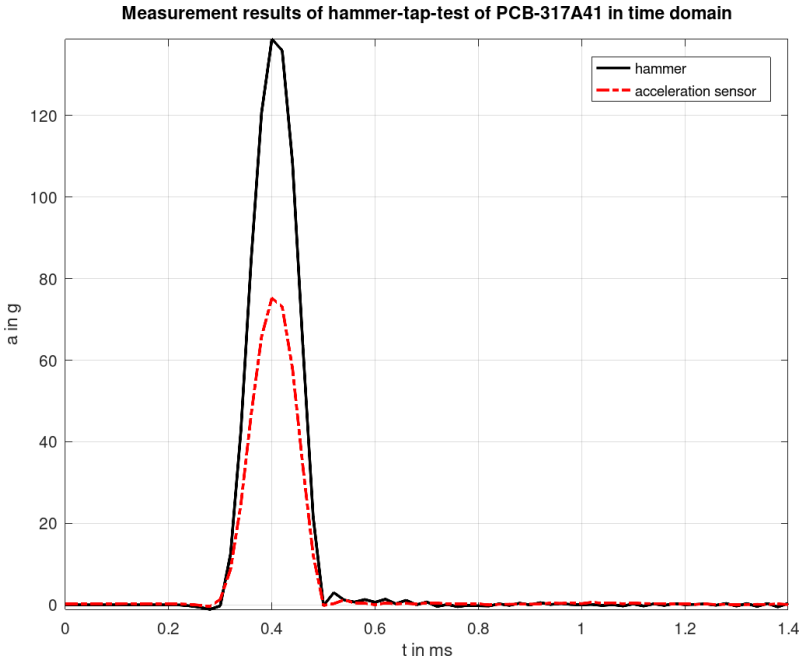

Figure 4.5: Time domain: Sensors acceleration signal compared to hammers calculated acceleration signal according to Newtons second law.

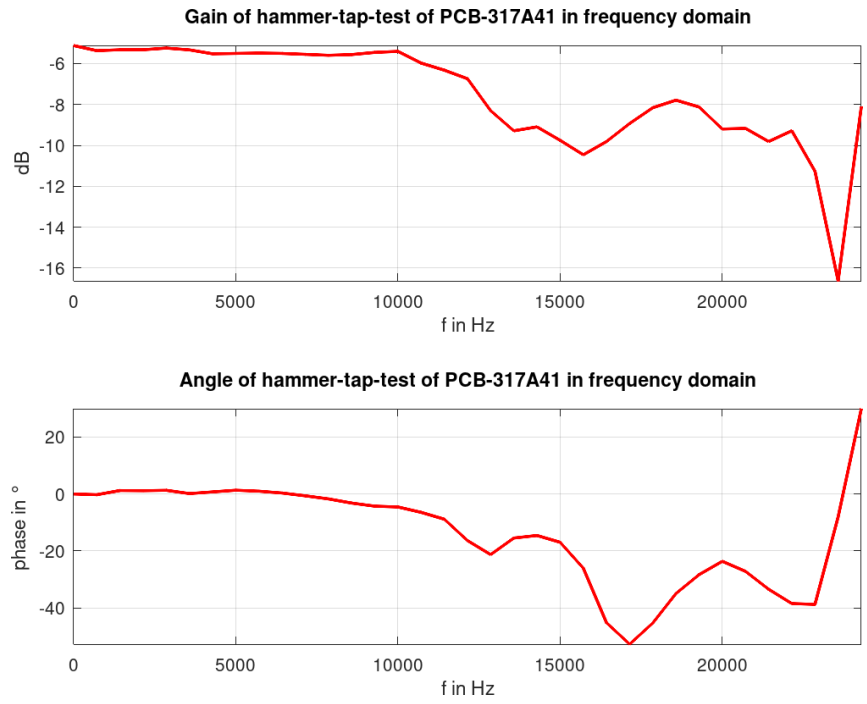

Figure 4.6: Frequency domain: Transfer function of acceleration sensor to hammer according to Newtons second law. Gain should be 0 dB and phase 0° for perfect calibration, but this is good enough.

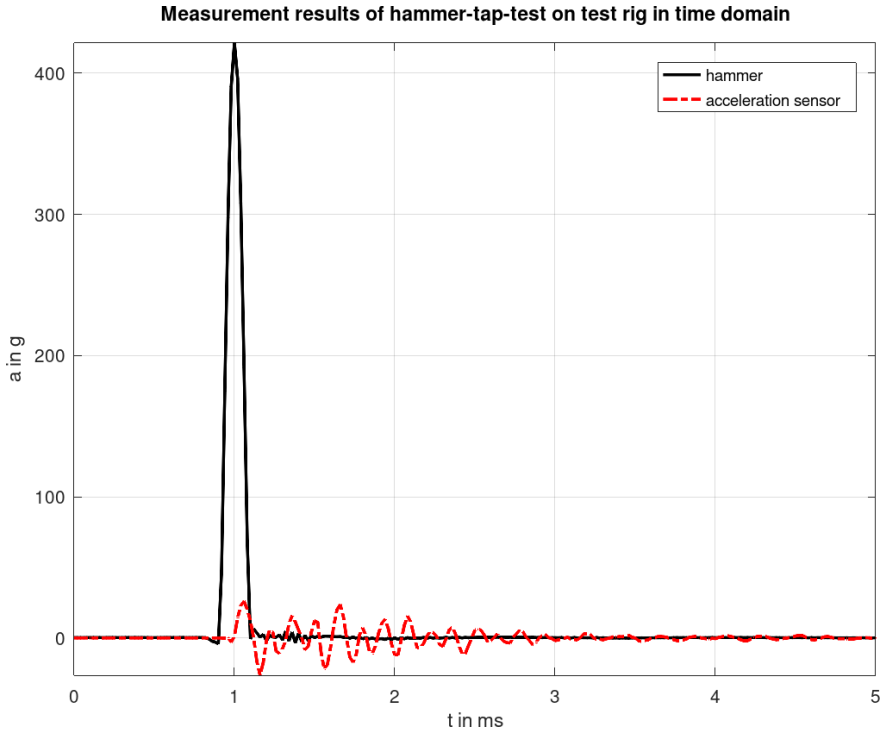

Figure 4.7: Time domain: Sensors acceleration signal compared to hammers calculated acceleration signal according to Newtons second law.

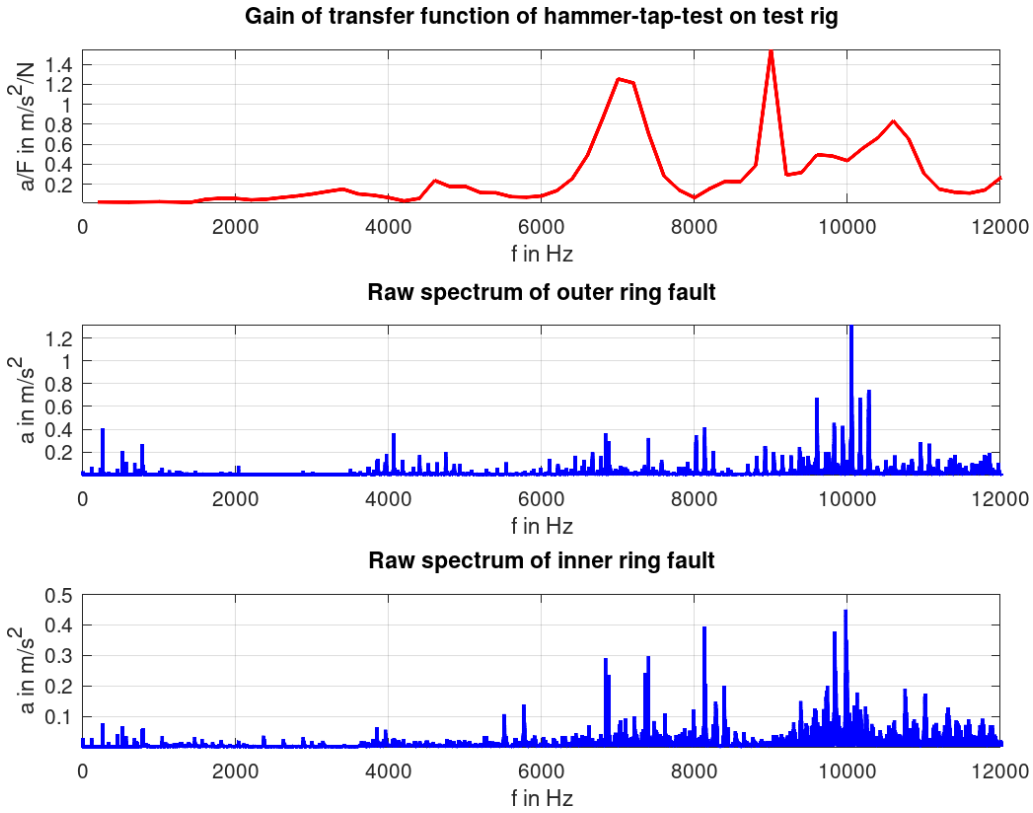

Figure 4.8: Top: Hammer tap excitation on the mechanical structure gives one an idea where the resonances in the transfer path between strike point and sensor are. Middle: Raw spectrum of an outer ring fault: Signal components are more likely in the region, where the hammer tap excitation also shows peaks. Bottom: Raw spectrum of an inner ring fault: Signal components are more likely in the region, where the hammer tap excitation also shows peaks.

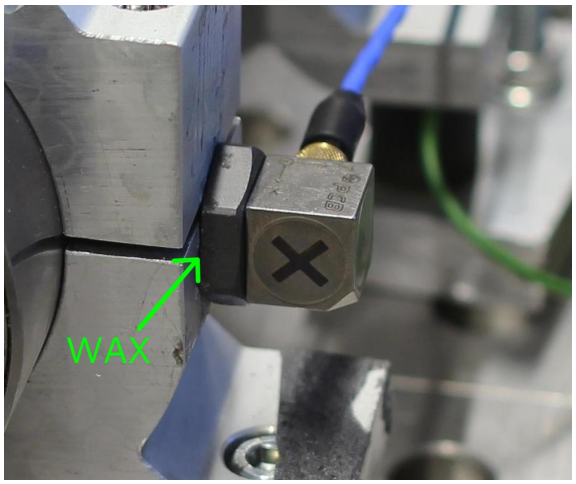

Figure 4.9: Acceleration sensor is stuck onto the motor bracket by using wax.

4.3.2 Rotary encoder

For measuring the position of the rotor the hall-sensor based incremental encoder *AMS-AS5134* is used which has a resolution of 1° up to a speed of $82,000$ *rpm*. The supplied magnet is mounted into the collet chuck with a machined fixture and the encoder-chip is placed a few millimeters above the magnet as can be seen in Figure 4.10. The incremental encoder can now determine the absolute position of the rotor shaft and outputs it via a digital quadrature signal and a reference pin signal to the digital input/output module *NI-9401* which again is connected to NI's LAB VIEW.

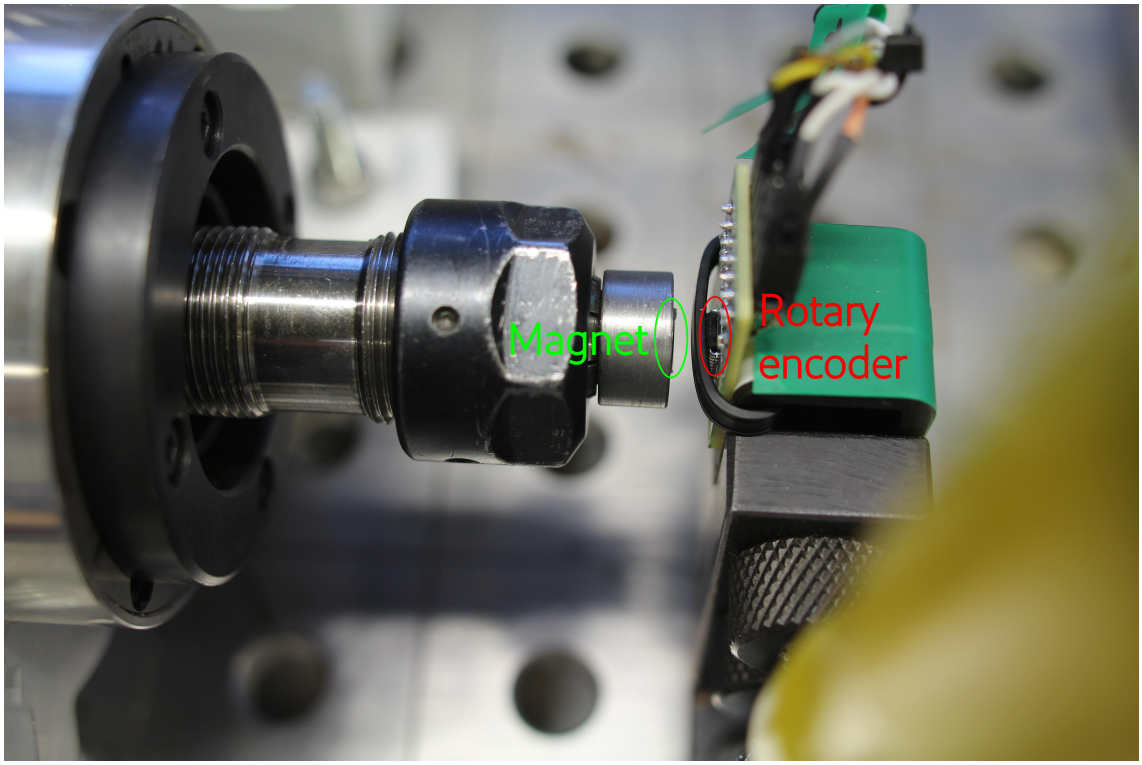

Figure 4.10: Speed sensor and magnet for position measurement.

5 Application of state of the art signal processing techniques

In this chapter the examined state of the art signal processing techniques shown in [4] are explained and a test whether they are useful or not for the milling spindle case is made. In section 5.8 an easy method for automatic detection of faults is proposed. An overview of the tested techniques can be seen in Table 5.1 whereas all signal processing techniques start from the order tracked signal.

Name	Purpose	SNR	
		inner fault	outer fault
original signal	-	0.009	0.056
OT	stabilizing	0.026	0.034
OT+LP	prewithening	0.030	0.034
OT+DRS	prewithening	0.034	0.042
OT+SA	prewithening	0.026	0.033
OT+DRS+MED	deconvolution	0.054	0.051
OT+DRS+MED+SK	filtering	0.054	0.051
OT+DRS+MED+WD	denoising	0.044	0.035
OT+DRS+MED+SK+SES	envelope	0.058	0.056
OT+DRS+MED+WD+SES	envelope	0.057	0.053

OT	... Order Tracking	MED	... Minimum Entropy Deconvolution
LP	... Linear Prediction	DRS	... Discrete Random Separation
WD	... Wavelet Denoising	SA	... Synchronous Averaging
SK	... Spectral Kurtosis	SES	... Squared Envelope Spectrum

Table 5.1: Performance overview of different signal processing techniques. The numbers in two SNR-rows show the values of the ratio between the expected signal power and the overall signal power. High SNR-values are the result of suitable signal processing techniques and therefore the techniques with the highest SNR are highlighted in gray.

The application of the different signal processing techniques which were chosen to be the best performing according to the SNR-benchmark - namely *Order Tracking* (section 5.2), *Discrete Random* Separation (subsection 5.3.2), *Minimum Entropy*

Deconvolution (section 5.4), *Spectral Kurtosis* (section 5.5) and *Squared Envelope Spectrum* (section 5.7) - of section 5.8 are shown in Figure 5.1 and Figure 5.2.

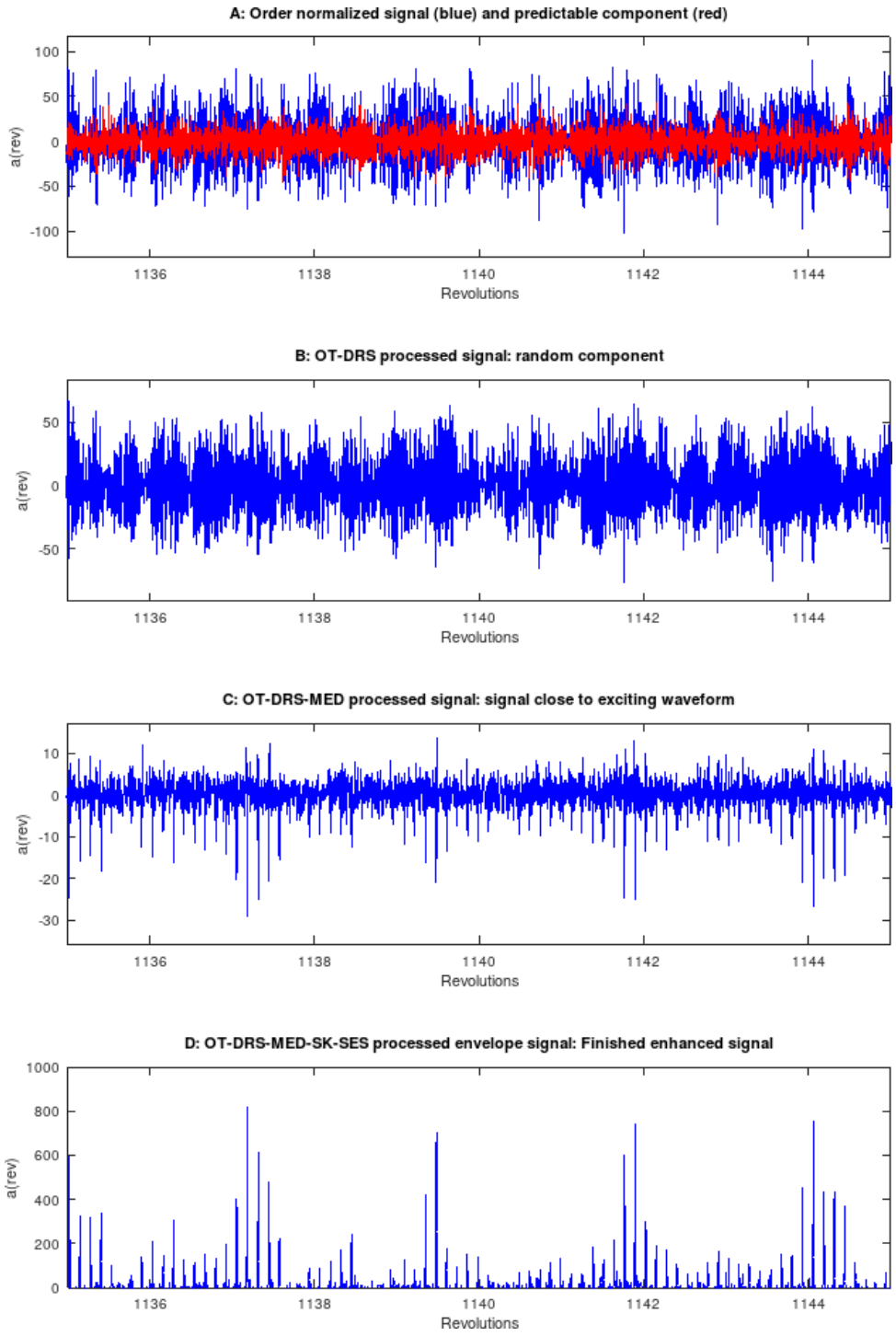

Figure 5.1: Application of different signal processing techniques for an outer race fault. Subplot A: The order tracked signal (blue curve) and the predicted signal component retrieved by order tracking and discrete-random-separation (red). Subplot B: The random signal component calculated by subtracting the predicted component from the order tracked signal. Subplot C: The signal from subplot B after minimum-entropy-deconvolution-filtering. Subplot D: The signal from subplot C after reconstructing it from the optimal frequency band using spectral kurtosis and constructing its squared envelope using the Hilbert transform.

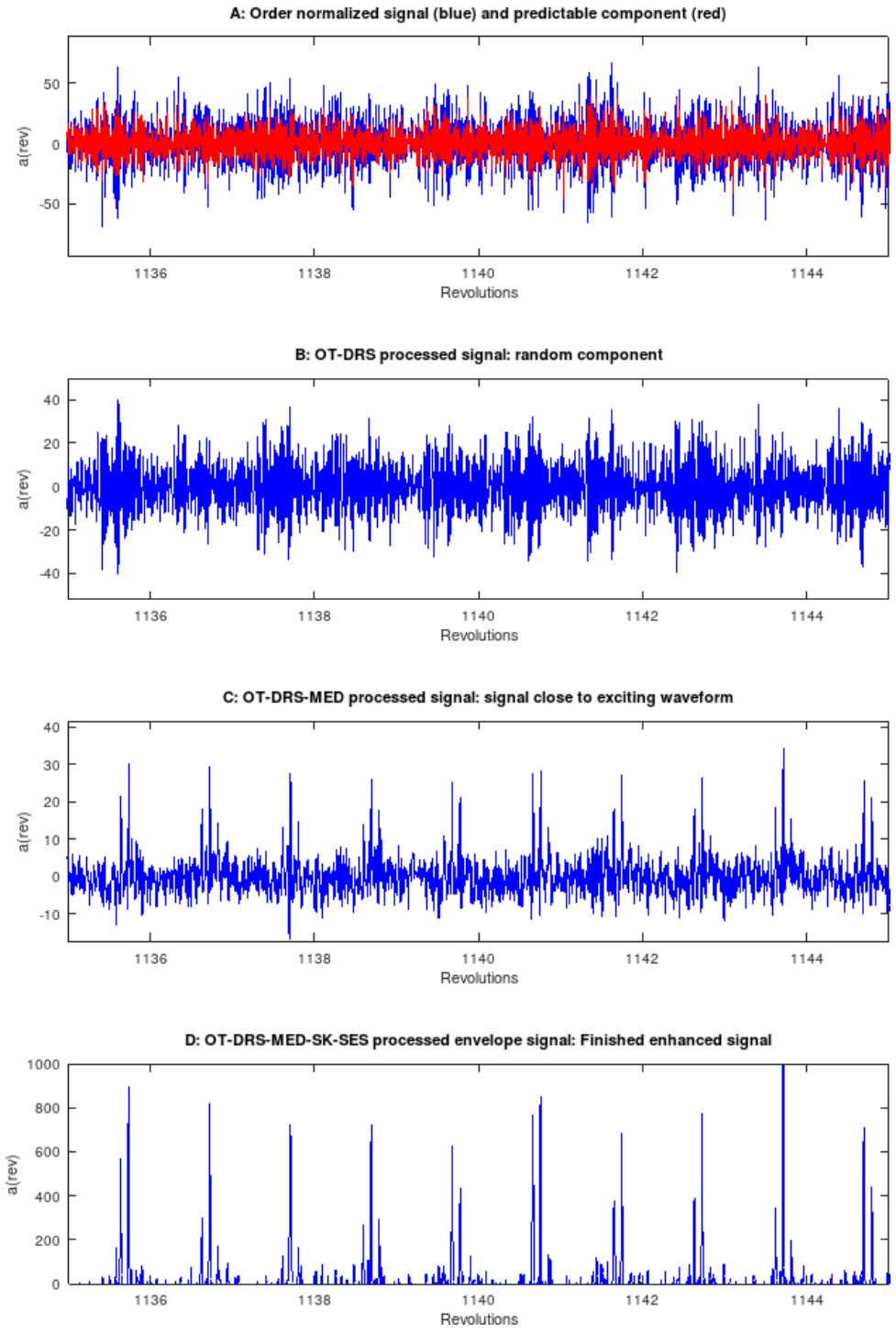

Figure 5.2: Application of different signal processing techniques for an inner race fault. Subplot A: The order tracked signal (blue curve) and the predicted signal component retrieved by order tracking and discrete-random-separation (red). Subplot B: The random signal component calculated by subtracting the predicted component from the order tracked signal. Subplot C: The signal from subplot B after minimum-entropy-deconvolution-filtering. Subplot D: The signal from subplot C after reconstructing it from the optimal frequency band using spectral kurtosis and constructing its squared envelope using the Hilbert transform.

5.1 Improvement quantification by using the signal to noise ration (SNR)

To quantify the techniques usefulness band-pass-filters were designed, which help to calculate a "benchmark-SNR" (Equation 5.3) by looking for expected ball pass frequencies (Equation 2.3 to Equation 2.2), their harmonics and side band frequencies which contribute to the useful signal component one wants to measure. Two filters were designed, which will be used in further investigation for judging the different techniques regarding their usefulness: one filter for the expected signal frequencies shown in Equation 5.1 and one for the noise components shown in Equation 5.2. The order of the harmonics and the side bands of the filters can always be increased to catch more expected frequencies, but they are based on Equation 2.3, Equation 2.4 and Equation 2.2 which ignore slip. If those frequencies are a little off, due to slip, the filter pass bands and stop bands will not fit to the real frequencies any more. Because the error in estimation scales proportionally to the order of the harmonics and the side bands, it should not be chosen to high. Therefore, only the first 5 harmonics (f_h), the first side band regarding f_S and the first side band regarding the FTF were taken into consideration for construction of the filter $H_1(f)$, as can be seen in Figure 5.3 and Figure 5.4 in subplot B and subplot C.

$$H_1(f) = ([n_{ord} \cdot f_h \pm k_{sb} \cdot FTF,\, n_{ord} \cdot f_h \pm k_{sb} \cdot f_S] * H_{rect}) \cdot H_{crop}(f) \tag{5.1}$$

$$H_2(f) = (1 - H_1(f)) \cdot H_{crop}(f) \tag{5.2}$$

$$SNR = \frac{\int H_1(f) \cdot S(f)\, df}{\int H_2(f) \cdot S(f)\, df} \tag{5.3}$$

$H_1(f)$ ... Filter for expected frequencies
$H_2(f)$ ... Filter for noise
n_{ord} ... order of harmonics
k_{sb} ... order of sidebands
$S(f)$ ... Spectrum of the signal
$H_{crop}(f)$ ... Low-pass filter which crops the spectrum at the desired frequency
H_{rect} ... Rectangular window
f_h ... harmonics of the vibrational signal
f_S ... Rotational frequency of the shaft
FTF ... Fundamental Train Frequency

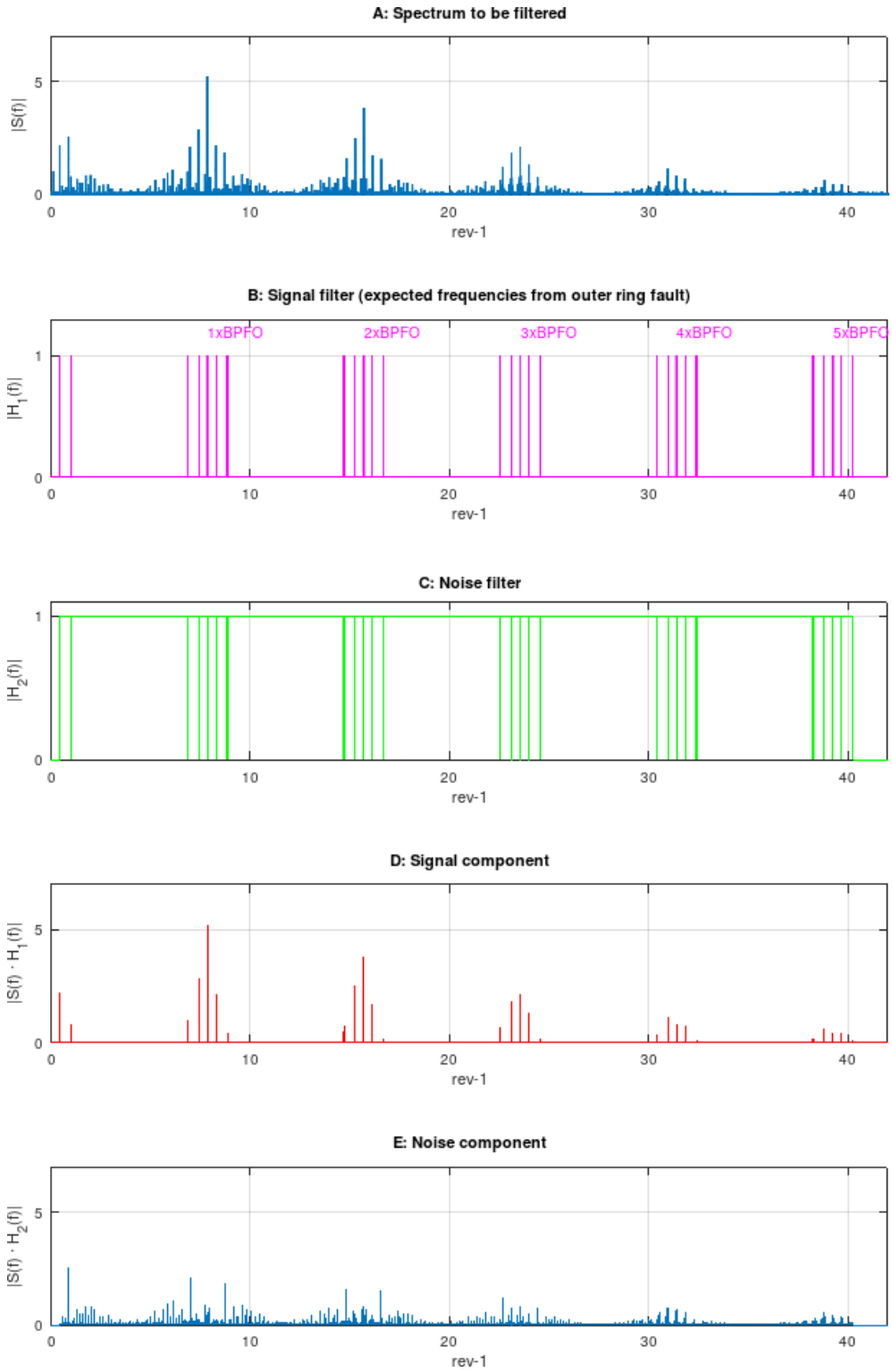

Figure 5.3: Demonstration of how the spectrum of an outer ring fault (subplot A) is split up by the signal filter from Equation 5.1 (subplot B) and the noise filter from Equation 5.2 (subplot C) into the signal component (subplot D) and the noise component (subplot E).

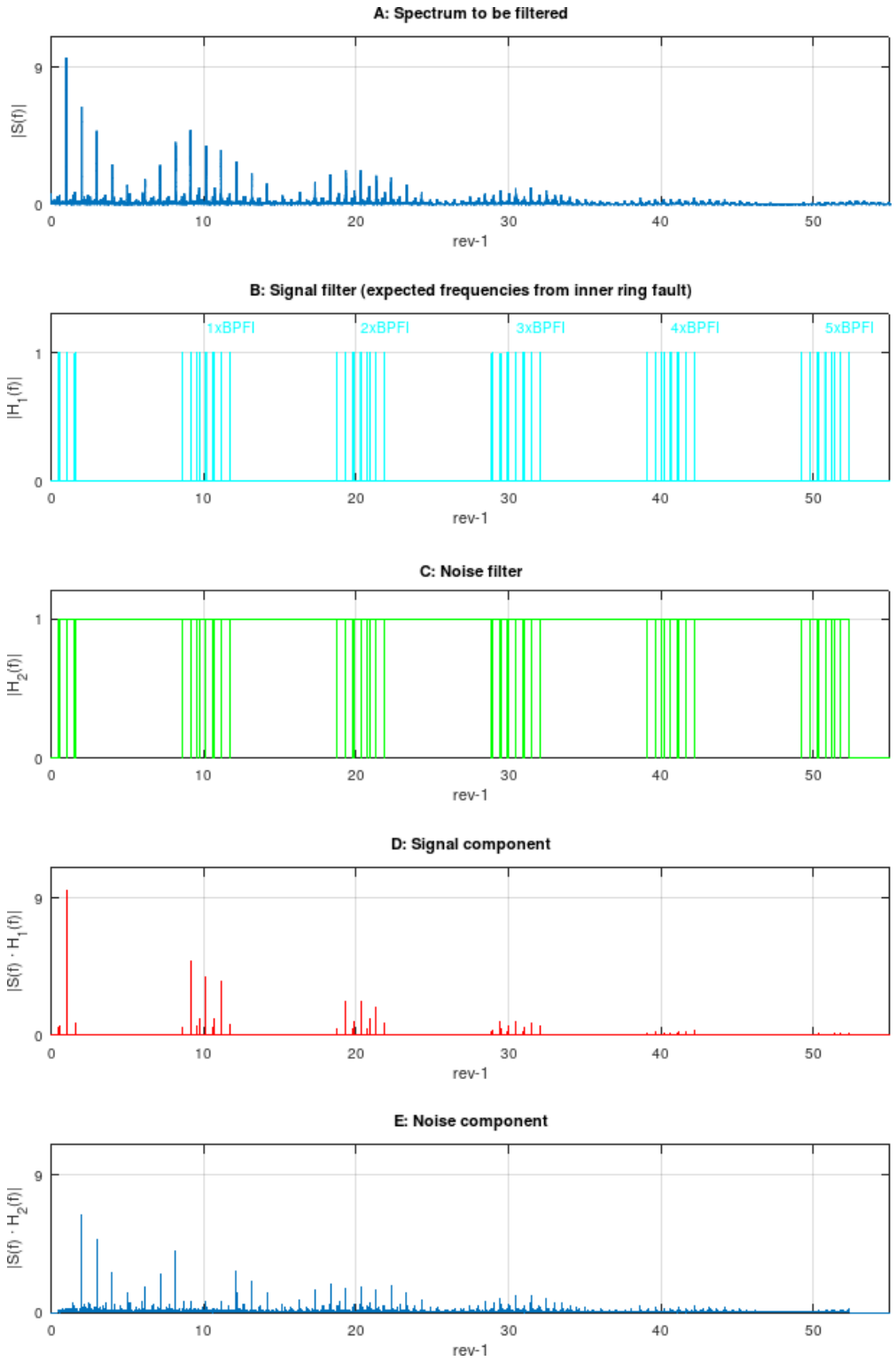

Figure 5.4: Demonstration of how the spectrum of an inner ring fault (subplot A) is split up by the signal filter from Equation 5.1 (subplot B) and the noise filter from Equation 5.2 (subplot C) into the signal component (subplot D) and the noise component (subplot E).

5.2 Order tracking (OT)

The main idea behind OT is to change the dependence of the measured data from time to angular position. This results in the major advantage of getting rid of speed variations during the measurement process and therefore having a less smeared spectrum. Hence the time axis will now show the rotor position instead of time and the frequency axis will show "orders" of shaft speed in $revolutions^{-1}$ instead of Hz. [4]

The technique used in this thesis is called "angular resampling" and a cubic polynomial (called *pchip* in GNU OCTAVE) interpolation is used to generate a polynomial between two sample points. With the interplation polynomial a new sample point with equidistant angle positions is calculated as can be seen in Figure 5.5. Angular resampling can only be properly applied, if a good angular position measurement is available. Therefore, the angular position measurement in the milling-spindle-application was smoothed using a moving average filter which was implemented by using the zero phase filtering technique. Due to zero phase filtering no phase shift was introduced. Because the motor used in the experiments was operated in steady state and has some inertia, the moving-average-filtering technique should provide valid results, since fast changes in speed are improbable. But even if there were some changes, due to the low filter order of 10 the filter would not smooth out significant deviations. As can be seen in Figure 5.7, OT might have improved the raw spectra by shifting the signal energy more towards the ball-pass frequencies and their harmonics which can be seen in slightly rising BPFI's and better SNR for the inner ring fault. In the case of the outer ring fault the SNR decreases but the first three harmonics are increased slightly. However, for further investigations it is necessary to use OT, otherwise the DRS (subsection 5.3.2) and the SA (subsection 5.3.3) might not work properly due to a non-stationary signal. Therefore order tracking is used anyway. The original signal spectrum of the signal can be seen in subplot A of Figure 5.6 and Figure 5.7 whereas the processed spindle signal of this technique in time domain can be seen in subplot B of Figure 5.6 and Figure 5.7.

To proof the concept of order tracking a (frequency modulated) test signal was generated (Equation 5.5, blue curve in Figure 5.8, subplot 2), which has a variation of its spindle frequency (Figure 5.8, subplot 1) according to Equation 5.4. Due to the frequency variation - which equals in principle a fluctuations of rotor speed - one can see the spread of the spectrum in (Figure 5.8, subplot 3) and the spectrum of the order tracked signal (Figure 5.8, subplot 4) where the spread is eliminated. The dependence changes from time to angular position.

$$f_{signal}(t) = f_{average} + \Delta f \cdot sin\left(2\pi f_{modulation} \cdot t\right) \tag{5.4}$$

$$a_{test}(t) = sin\left(2\pi f_{signal}(t) \cdot t\right) \tag{5.5}$$

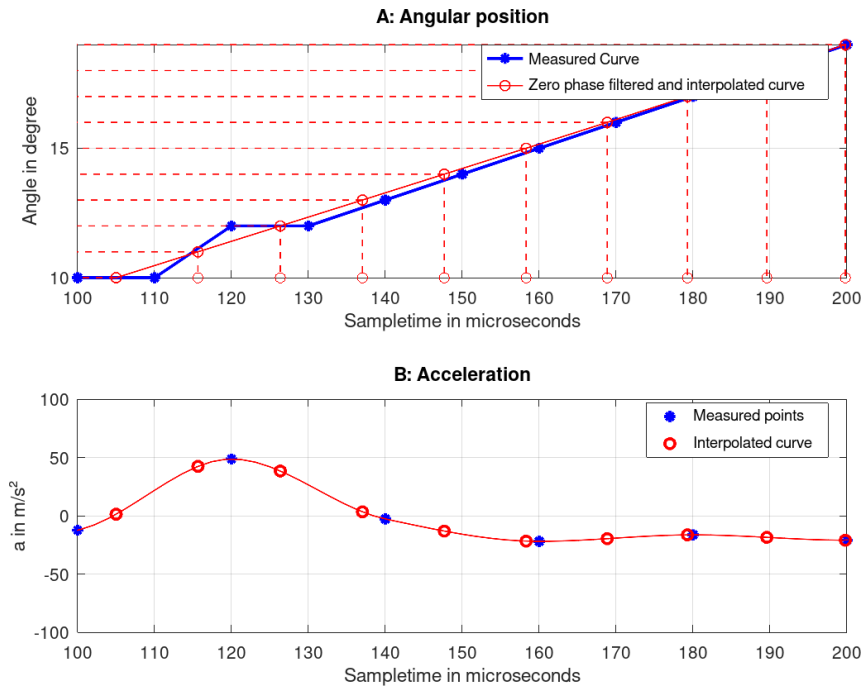

Figure 5.5: *A:* the blue line represents the measured angular position and the red curve is the moving-average-filtered position. *B:* the red curve is an interpolation from the measured (blue) acceleration signal using *GNU OCTAVEs cubic polynomial "pchip"*-interpolation technique.

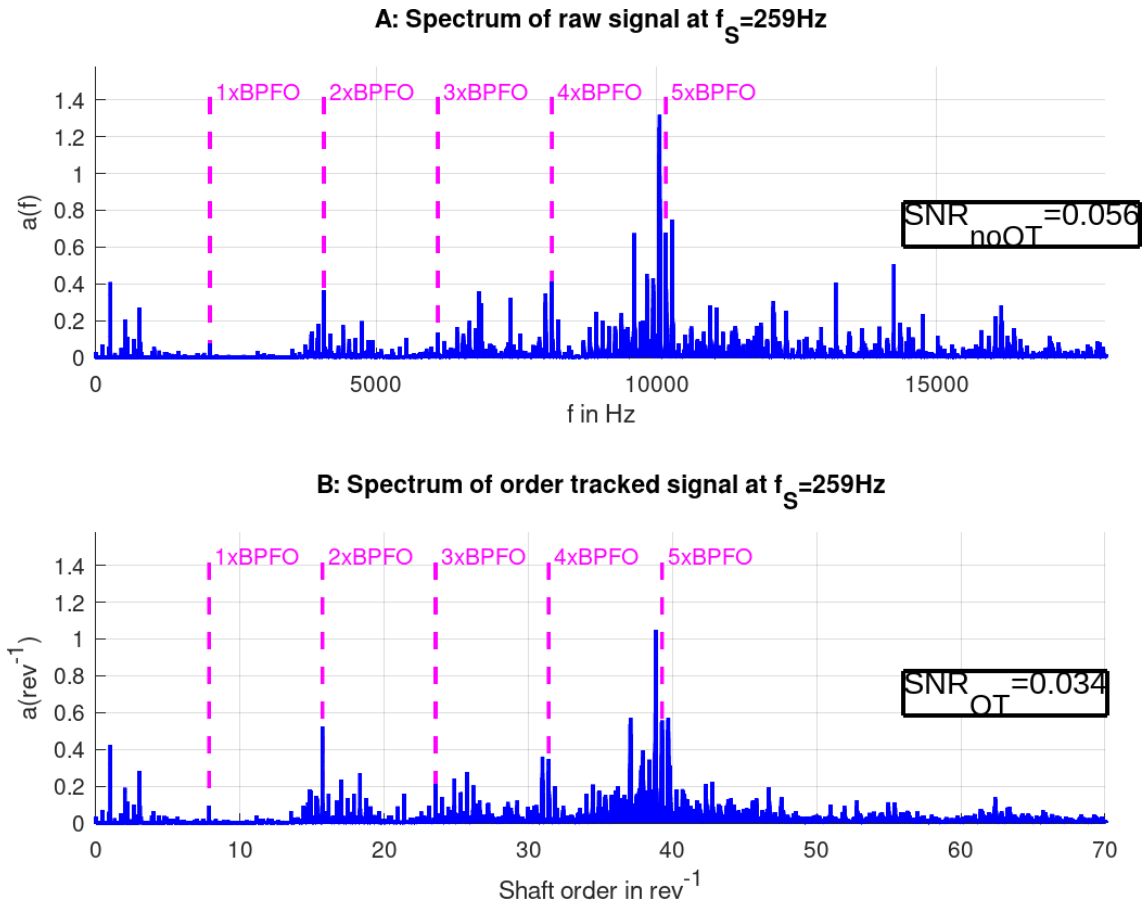

Figure 5.6: Localized outer ring faults: Raw data spectrum (subplot A) vs order tracked spectrum (subplot B). The signal energy of the lower harmonics of the outer-raceway-ball-pass-frequency increases due to order tracking. However the signal-to-noise-ratio (SNR) decreases. Even though the signal-to-noise-ratio decreases, order tracking is still used, because it is a requirement for discrete-random-separation (subsection 5.3.2) or synchronous averaging (subsection 5.3.3) to work properly.

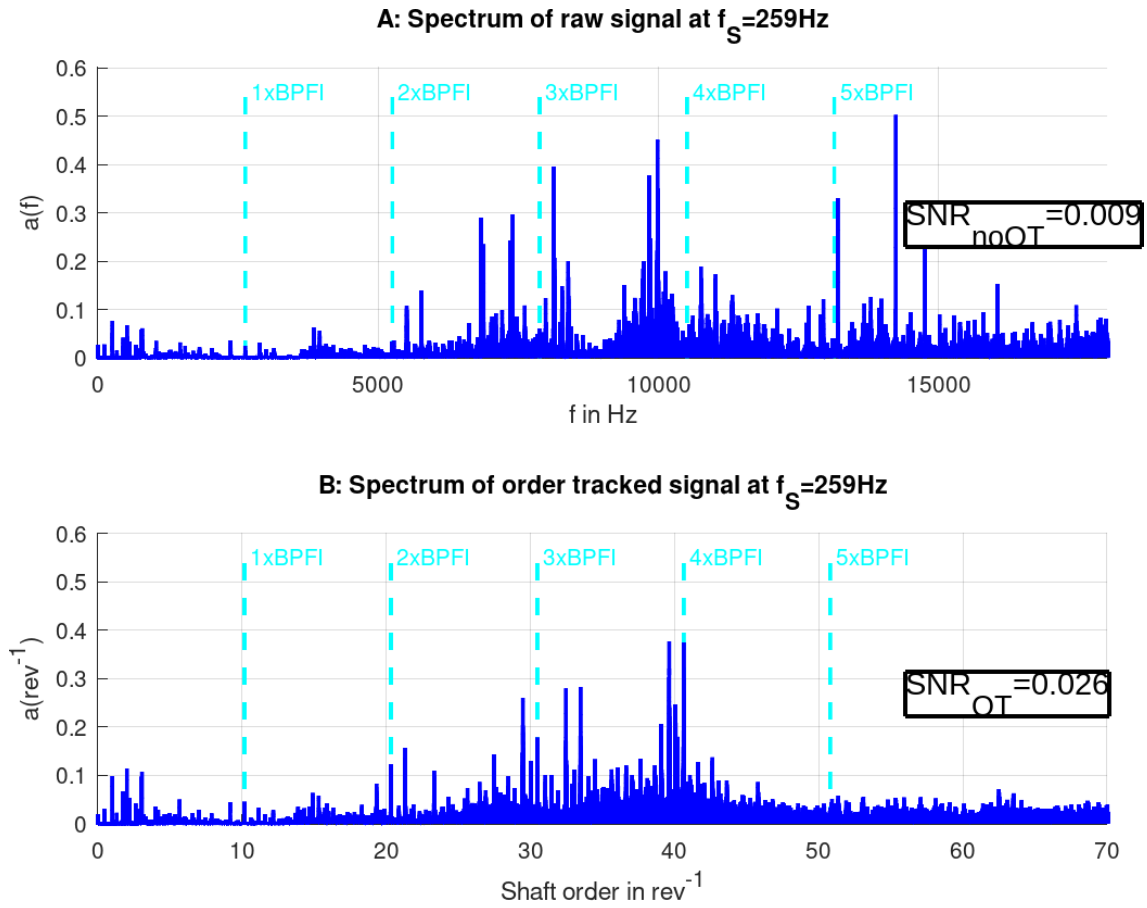

Figure 5.7: Localized inner ring faults: Raw data spectrum (subplot A) vs order tracked spectrum (subplot B). The signal energy of the lower harmonics of the inner-raceway-ball-pass-frequency and the signal-to-noise-ratio (SNR) increases due to order tracking.

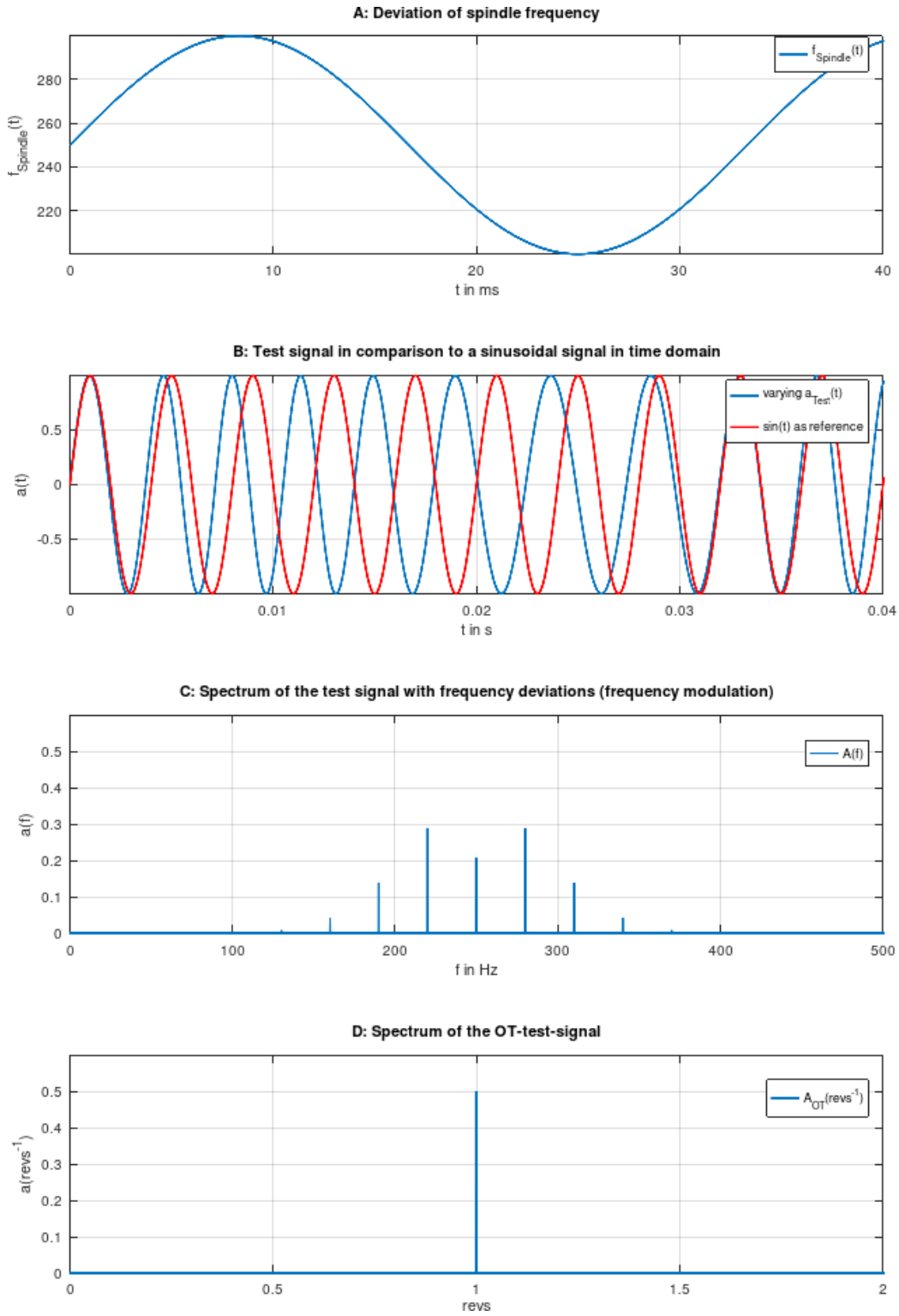

Figure 5.8: Proof of concept for order tracking.
Subplot A: Simulated spindle frequency from Equation 5.4. Subplot B: Comparison between modulated spindle speed vs constant spindle speed. Subplot C: Spectrum of modulated spindle speed without order tracking. Subplot D: Spectrum of order tracked signal. Due to angular resampling, speed variations get canceled out and therefore the sidebands vanish.

5.3 Removing predictable periodic components

In this section the tested methods for removing discrete predictable components are described and their results are presented. The signal fed into the prediction-removal-algorithms is the order tracked signal from section 5.2. It is important to mention, that the Discrete Random Separation-method (subsection 5.3.2) was chosen as reference technique for further enhancement techniques because it performed the best regarding the SNR. If other aspects like left over signal energy are more important, the SA method might be preferred.

5.3.1 Linear prediction - signal whitening

Linear prediction is based on an auto-regressive model described in Equation 5.6 to Equation 5.8 from [4]. The predicted value $p[k]$ is subtracted from the measurement data $x[k]$ as shown in Equation 5.7 and as a result the error term $\varepsilon[k]$ contains the remaining random signal components from the impacts. This way by fourier transforming Equation 5.7 one gets the transfer function described in Equation 5.8 which helps filtering the predictable component from the measured signal. The result of a simulated case where sines and impulses are added is shown in Figure 5.9. The position of the filter in the signal model is shown in Figure 2.9. Ideally the leftover signal should consist only of the random impulse excitation convoluted with the transmission path's transfer function.[4]

The filter order P was changed between 1 and 200 and the filter which left the highest kurtosis in the random signal was chosen. As can be seen in Figure 5.10 and Figure 5.11 linear prediction worsened the SNR compared to the original signal making this technique useless when it comes to signals with very little predictable components to remove.

$$p[k] = -\sum_{n=1}^{P} a[n] \cdot x[k-n] \tag{5.6}$$

$$\varepsilon[k] = x[k] - p[k] = x[k] + \sum_{n=1}^{P} a[n] \cdot x[k-n] \tag{5.7}$$

$$X(f)A(f) = \varepsilon(f) \rightarrow A(f) = \frac{\varepsilon(f)}{X(f)} \tag{5.8}$$

$p[k]$	...	predicted signal value
$x[k]$	...	real signal value
$a[n]$	...	auto-regressive coefficients
$\epsilon[k]$	...	prediction error = random signal component
$X(f)$	...	spectrum of the real signal
$\epsilon(f)$	...	spectrum of the random signal component
$A(f)$	...	filter transfer function to get random components

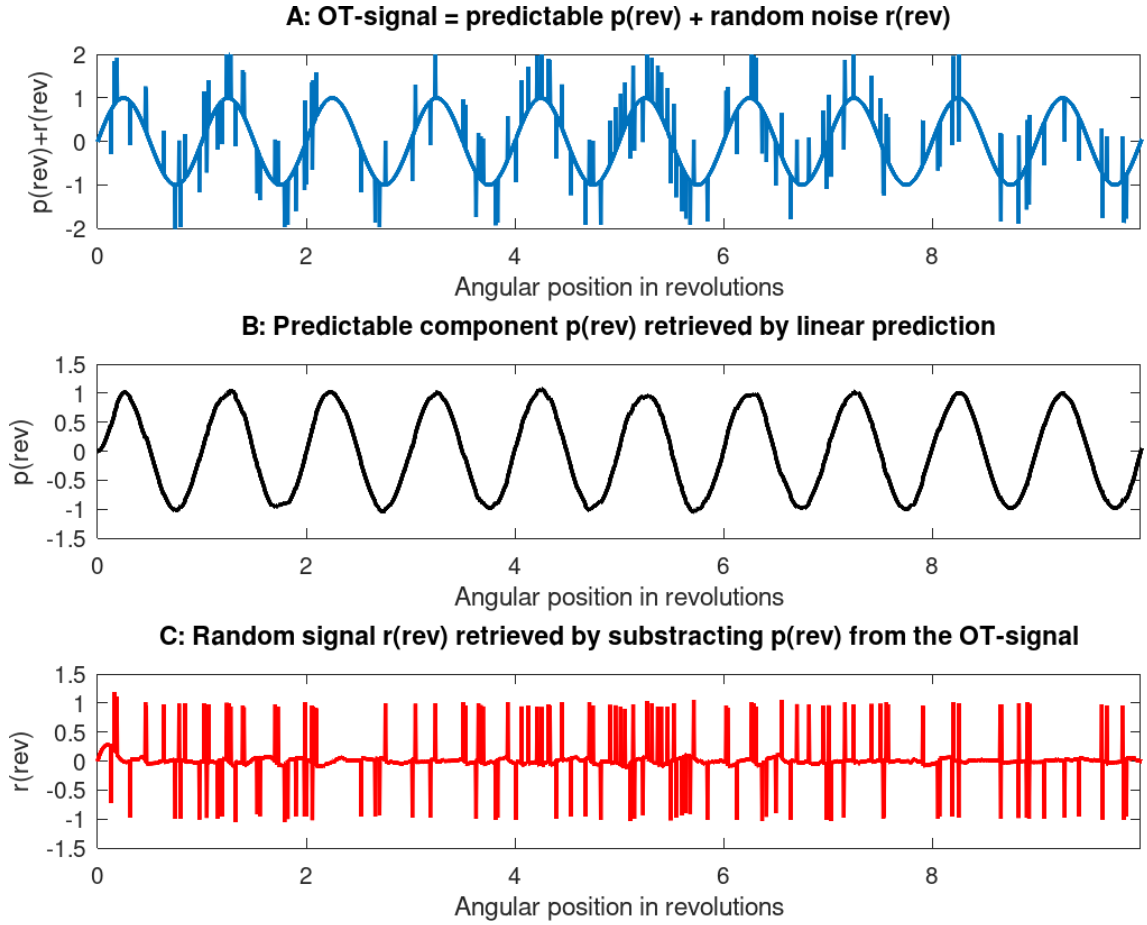

Figure 5.9: Proof of concept for the linear prediction algorithm

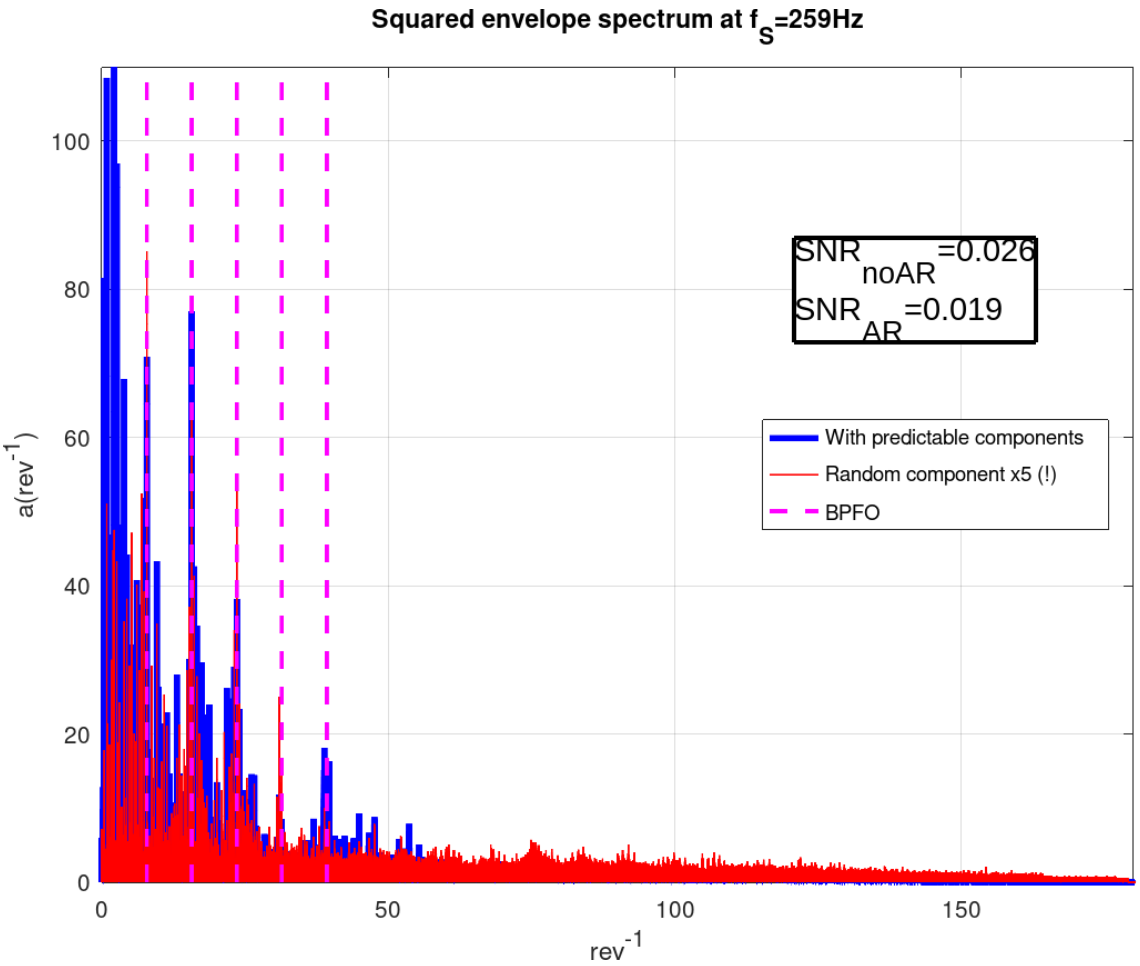

Figure 5.10: Spectrum of a signal enhanced with Order-Tracking, linear prediction and the Squared-Envelope-Spectrum for an outer race fault.

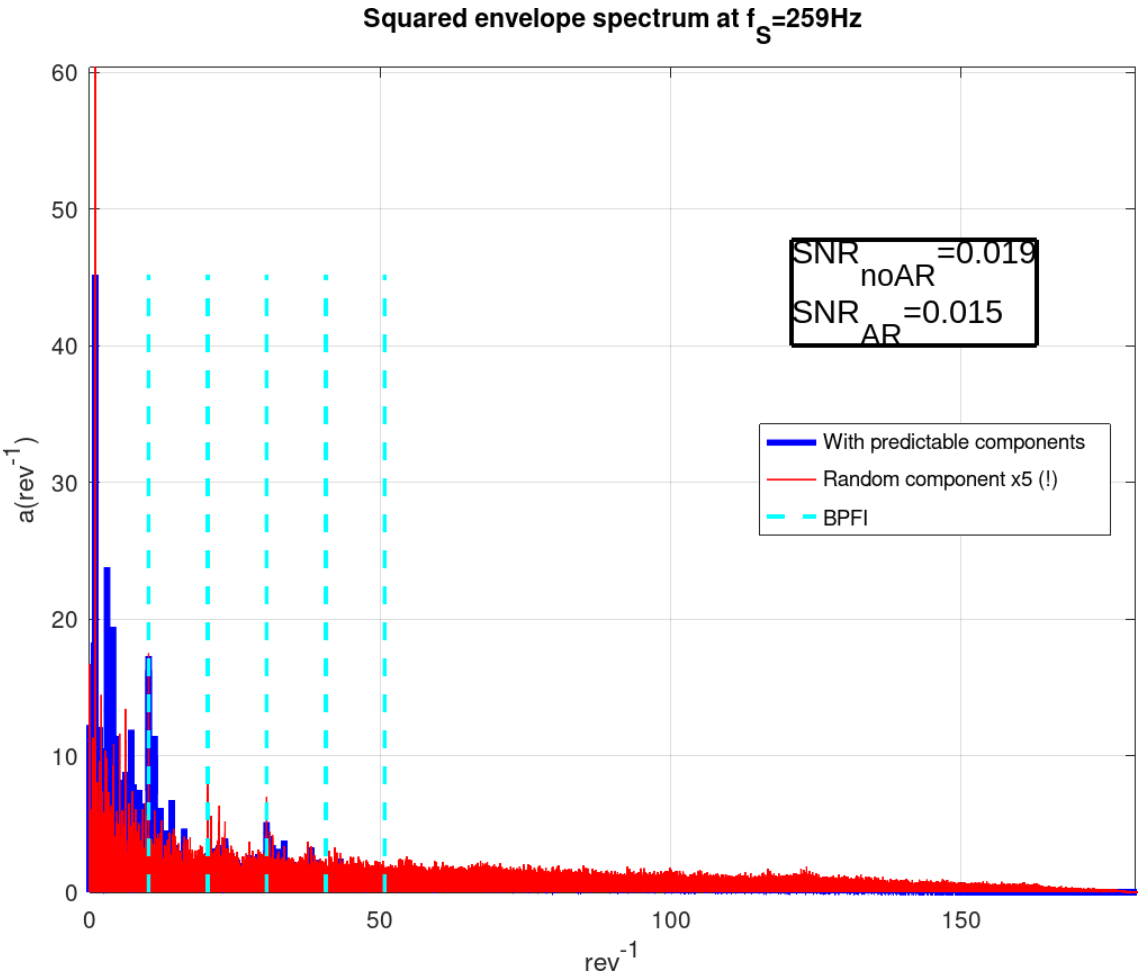

Figure 5.11: Spectrum of a signal enhanced with Order-Tracking, linear prediction and the Squared-Envelope-Spectrum for an inner race fault.

5.3.2 Discrete Random Separation (DRS)

The *Discrete Random Separation* technique can be used if the tracked signal is very stable. In the case of the flywheel or the milling spindle, this is true for stationary operation with OT. The basic idea is, that a signal $y[t]$ is split up into blocks which are windowed by the Hanning window. Then the Fourier transform of the current block $Y(f)$ and its consecutive block $Y(f)_\Delta$ is calculated. The transfer function is then calculated according to Equation 5.9. The expectation operator in Equation 5.9 means, that the numerator and denominator of the transfer function should be calculated multiple times for different segments of the measured signal data, and then the mean value at each frequency should be calculated. By averaging over multiple blocks of signal data, the noise gets canceled out with an increasing number of segments.

Equation 5.9 is now the filter function for gaining the predictable signal component of the raw data. If the signal is now filtered with this transfer function, only random components should be left over. It is recommended, that the size of the filter spans at least 10-20 periods of the signal with the lowest frequency to be removed. A major advantage of this technique compared to the SA (subsection 5.3.3) is that it can also remove side bands of predictable signals. The reason is, that this technique does average over blocks in the frequency domain (compared to SA which averages over blocks in time domain). Because the blocks are windowed in time domain (less leakage) and then transformed into the frequency domain, they contain all frequencies ,whether they are multiples of the shaft frequency or not. Another trick to improve the technique is to perform synchronous sampling when building the filter because then one has less leakage effect. It is worth mentioning, that DRS is the only technique which actually improves the SNR regarding periodic-component-removal-algorithms. The absolute value of the transfer function helps to get rid of phase shifting properties of the transfer function, but then the filter becomes acausal. The position of the filter in the signal model is shown in Figure 2.9. The resulting processed spindle signal of this technique in time domain can be seen as the blue curve in subplot 1 and subplot 2 of Figure 5.1 and Figure 5.2. [4]

$$H(f) = abs\left(\frac{\mathscr{E}\left\{Y(f)Y(f)^*_\Delta\right\}}{\mathscr{E}\left\{abs\left(Y(f)\right)^2\right\}}\right) \tag{5.9}$$

$Y(f)$ ... Windowed data block (segment)
$Y(f)_\Delta$ ... Consecutive windowed data block shifted in the time domain by Δ
$\mathscr{E}$ ... Expected value - mean value over multiple signal blocks

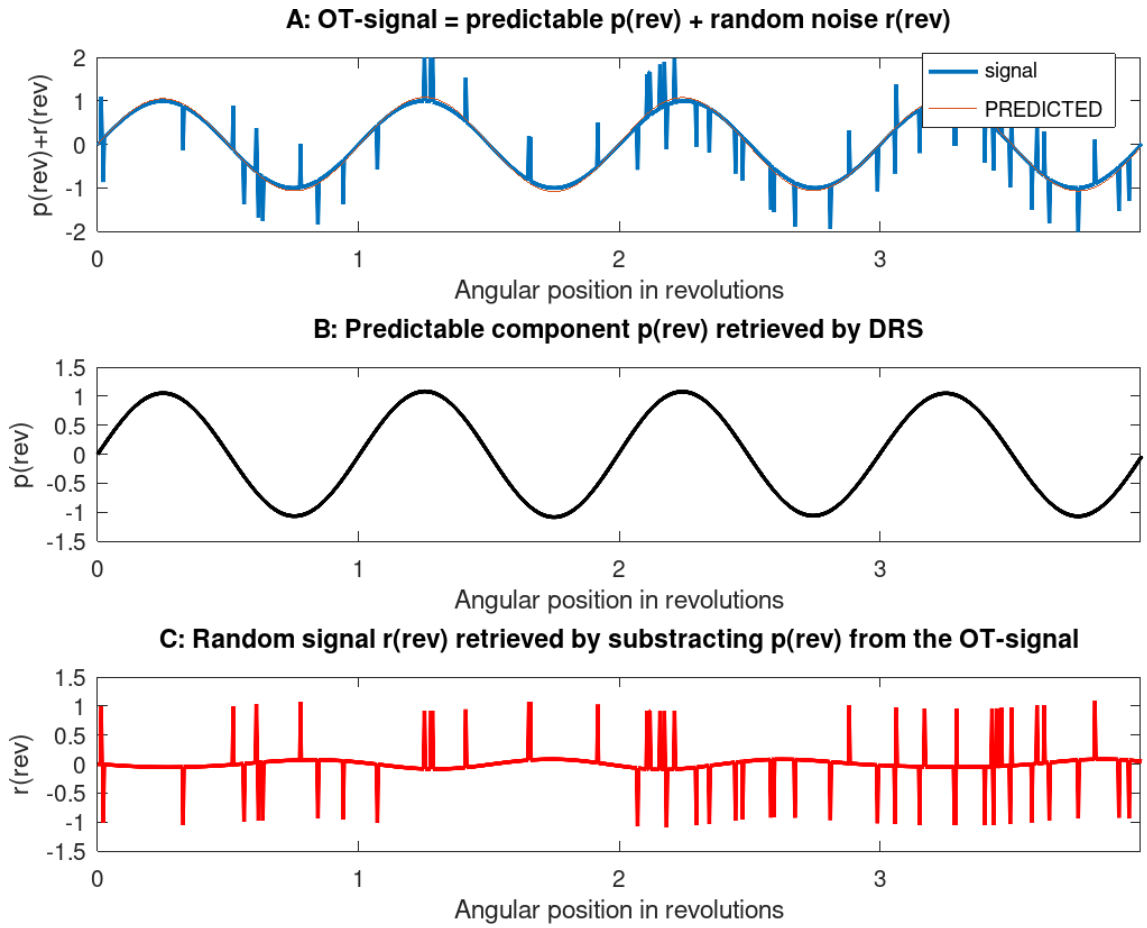

Figure 5.12: Proof of concept for the Discrete-Random-Separation algorithm

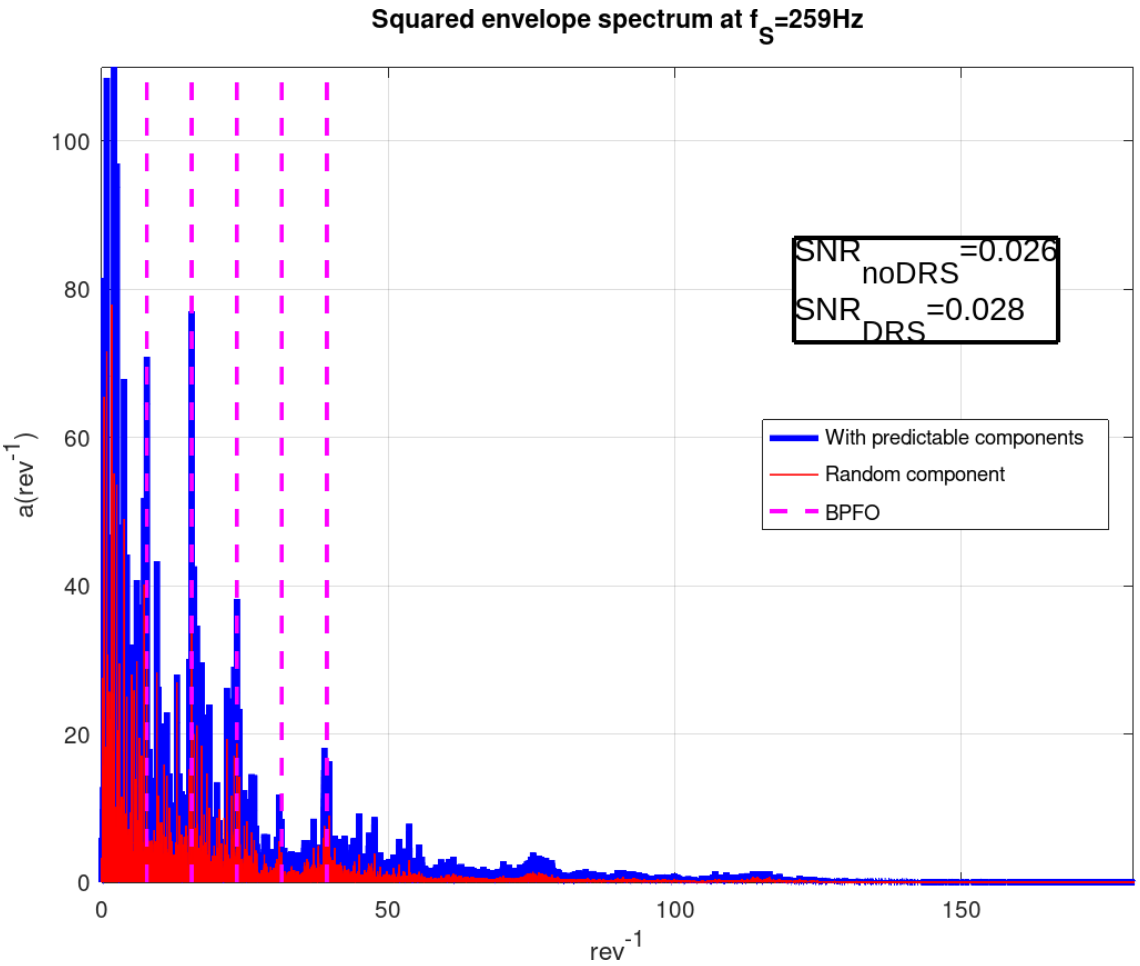

Figure 5.13: Spectrum of a signal enhanced with Order-Tracking, Discrete-Random-Separation and the Squared-Envelope-Spectrum for an outer race fault.

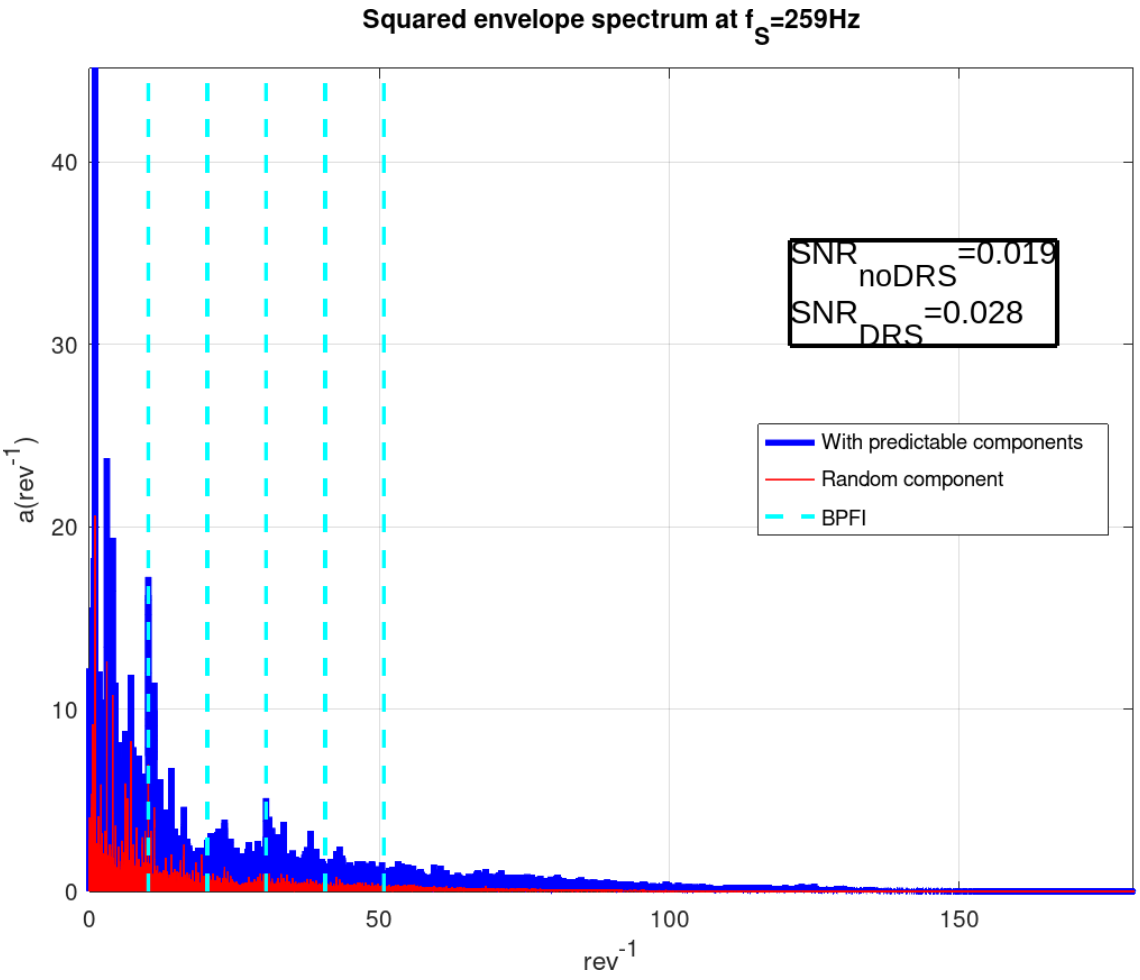

Figure 5.14: Spectrum of a signal enhanced with Order-Tracking, Discrete-Random-Separation and the Squared-Envelope-Spectrum for an inner race fault.

5.3.3 Synchronous averaging (SA)

Synchronous averaging (SA) is used for getting rid of the predictable non-random components of the signal which can be seen for a simulated case in Figure 5.15. SA (similar to DRS) requires a very stable signal where no deviation of frequency is allowed. Hence OT has to be done before SA. For synchronous averaging, the order tracked signal is divided into blocks whose length contains an integer multiple of rotor revolutions. These blocks are then averaged which cancels out any randomness and leaves only the predictable component of the signal. Because the impacts are a random process, they will be canceled out in the averaging process. This SA-signal then gets subtracted from the original order tracked signal (Equation 5.10). This technique has the advantage of being easy to calculate and no big matrices have to be inverted - hence making the computational costs low, but a downside is, that the SA-signal has to be calculated for every different frequency. The position of the filter in the signal model is shown in Figure 2.9, but it should be mentioned, that this kind of operation cannot be considered as a real filtering operation, because it is a subtraction.[4]

$$Sig_{random} = Sig_{OT} - Sig_{SA} \tag{5.10}$$

In the case of the milling spindle and the flywheel, there are no gears or other mechanical components involved which majorly might contribute with periodic predictable signal components. Therefore in this case this technique shows no improvement in the spectrum due to little periodic disturbances (Figure 5.16, Figure 5.17). In our case the only thing which gets filtered out is most probably signals resulting from the rotor imbalance. Figure 5.1 and Figure 5.2 show the improved spectra. A drawback of this technique is, that it will not necessarily get rid of side bands of predictable components, because they do not have to be integer multiple frequencies of the shaft frequency.

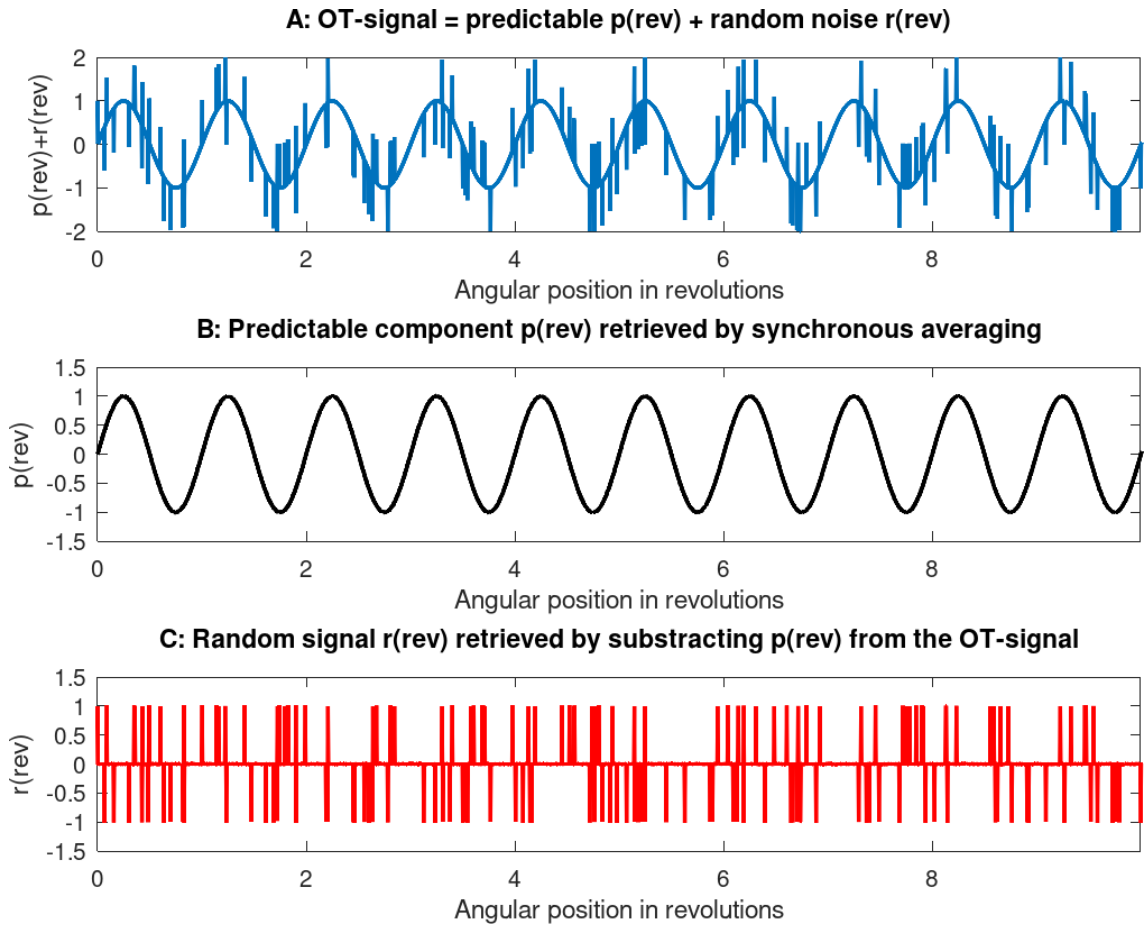

Figure 5.15: Proof of concept for the Synchronous-Averaging algorithm

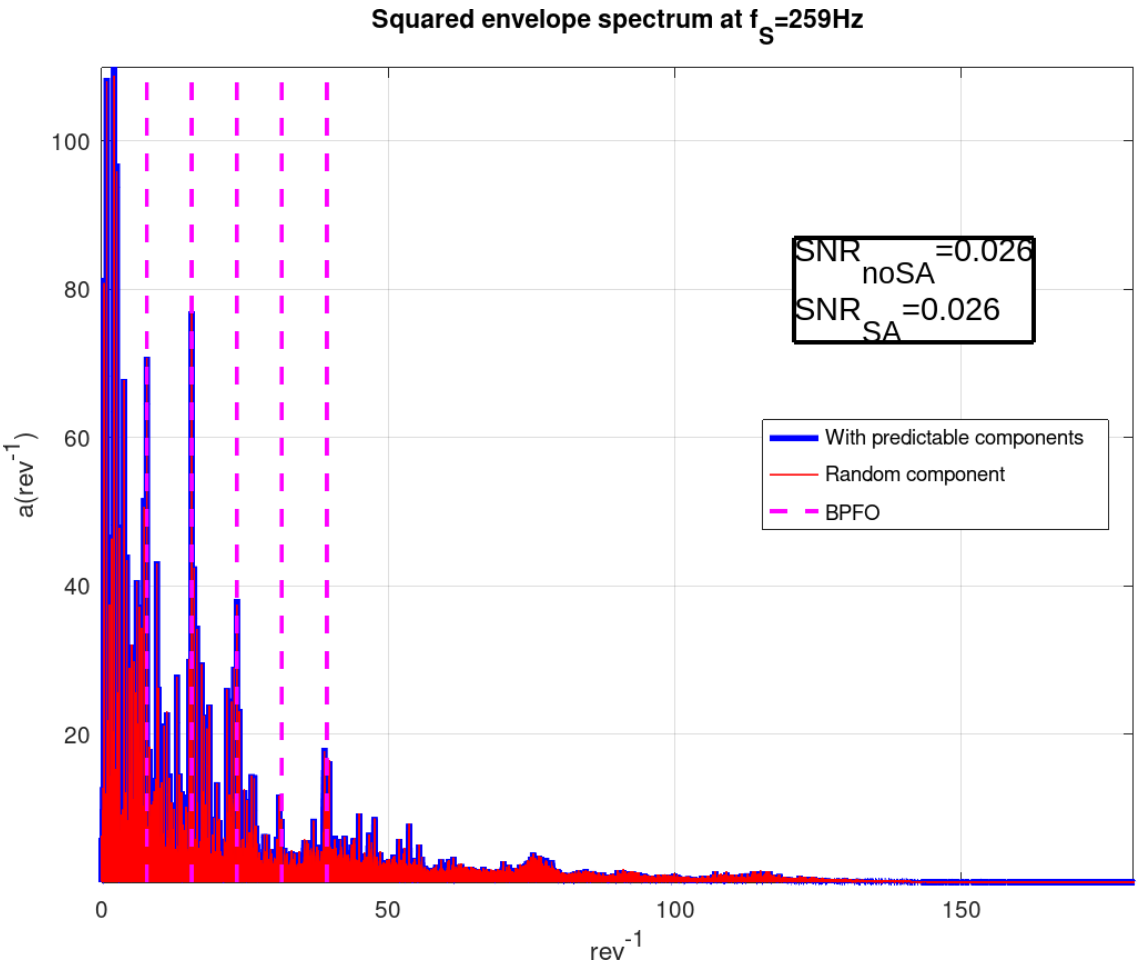

Figure 5.16: Spectrum of a signal enhanced with Order-Tracking, Synchronous-Averaging and the Squared-Envelope-Spectrum for an outer race fault.

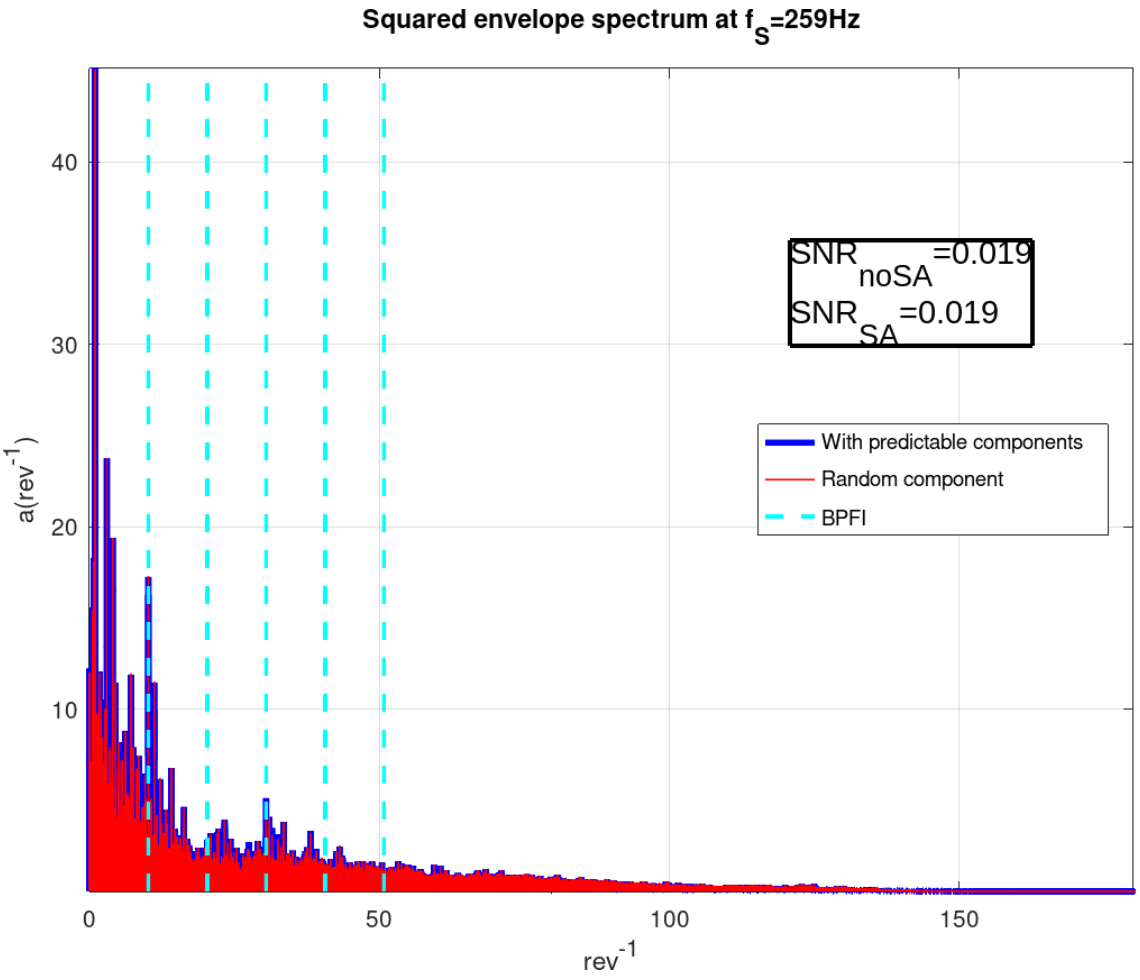

Figure 5.17: Spectrum of a signal enhanced with Order-Tracking, Synchronous-Averaging and the Squared-Envelope-Spectrum for an inner race fault.

5.4 Minimum entropy deconvolution (MED)

The signal fed into the MED-algorithm is the DRS signal from subsection 5.3.2.

The idea of the *minimum entropy deconvolution* (MED) is to reduce the impulse response function's temporal spread to get a waveform closer to its original exciting shape. Because the spread impulse response function is narrowed down, this technique helps to separate overlapping impulses which is especially useful for very high speed applications like in the case of the flywheel where by its nature, impulses might be excited faster, than they could decay. The idea is to find the inverse filter of the transmission paths transfer function which is shown in Figure 2.9 as the box named *MED*. It is assumed, that the periodic component $d_k * h_k$ is already removed by the DRS algorithm in subsection 5.3.2. At first, the excitation signal g_k mixed with the additive noise n_k excites the structure h_k and then gets deconvoluted by the MED filter which leaves us with an approximation of g_k. The *assumption* is, that the excitation is *as impulsive as possible*. The MED-filter is modeled as L-th order FIR-filter f_{MED} (Equation 5.11) and inverts the systems impulse response $h[k]$ so that the exciting impulse is reconstructed as can be seen in Equation 5.12. f_{MED} has a delay l_{delay} which makes the filter causal.

$$y_k = \sum_{l=1}^{L} f[l] \cdot \varepsilon_l[k-l] \tag{5.11}$$

$$f_{MED} * h[k] = \delta[k - l_{delay}] \tag{5.12}$$

$$\frac{\partial y[k]}{\partial f_{MED}[l]} = \varepsilon_l[k-l] \tag{5.13}$$

The objective function is the kurtosis of the MED filtered signal who's first derivative should be zero (Equation 5.14).

$$O_{kurtosis}(f_{MED}) = \frac{\sum_{k=1}^{K} y^4[k]}{\left[\sum_{k=1}^{K} y[k]\right]^2} \Longrightarrow \frac{\partial O_{kurtosis}(f_{MED})}{\partial f_{MED}} = 0 \tag{5.14}$$

Equations Equation 5.11, Equation 5.12 and Equation 5.13 can be combined to the final (Equation 5.15) for the algorithm to optimize.

$$\underbrace{\frac{\sum_{k=1}^{K} y^2[k]}{\sum_{k=1}^{K} y^4[k]} \sum_{k=1}^{K} y^3[k]\varepsilon_l[k-l]}_{b} = \underbrace{\sum_{p=1}^{L} f_{MED}[p] \underbrace{\sum_{k=1}^{K} \varepsilon_l[k-l]\varepsilon_l[k-p]}_{A}}_{f} \Longrightarrow A^+ b = f_{MED} \tag{5.15}$$

Here A represents the autocorrelation matrix with toeplitz structure of the filtered signal and A^+ its pseudo-inverse.

The steps to calculate a MED filter according to the algorithm from [2, 4, 5] are the following:

1. Compute the auto correlation matrix from the signal $\varepsilon[k]$ (Needs only to be calculated once)
2. Assume the initial filter coefficients (The first filter is simply a delayed impulse response $f_{MED,initial} = [0, 1, 0, 0, ..., 0]$)
3. Compute the filtered output signal $y[k]$ from Equation 5.11 with the filter coefficients $f_{MED}[l]$ and the measured signal $\varepsilon[k]$.
4. Compute the column vector b.
5. Solve Equation 5.15 to get the improved filter coefficients.
6. Redo step 3 to 5 until either the change of kurtosis is below a defined threshold, or a maximum number of iterations is reached.

The convergence of the kurtosis towards a maximum level can be seen in Figure 5.18, subplot 2. The software implementation from the MED algorithm in GNU OCTAVE was adapted from [8] and improved in speed by using more matrix operations instead of loops. As can be seen in Figure 5.18, the filter converges after a few iterations and then stops either due to fulfilling the convergence criteria or the maximum number of iterations. In Figure 5.19 one can see a two dimensional surface which represents the sum of the kurtosis for the filtered signal over the filter order $L = N_{MED}$ and the iterations of the algorithm. The highest spot might suggest the optimal filter order, but one may not forget, that this plot only shows the kurtosis. It could happen, that a filter order of 1000 shows the highest kurtosis, but the leftover signal energy after MED-filtering is so small, that the spectrum makes no more sense. The best results in the case of the milling spindle were found to be approximately $L = N_{MED} = 300$. The resulting processed spindle signal of this technique, in the time domain, can be seen as the blue curve in subplot 3 of Figure 5.1 and Figure 5.2. The improvement of the spectra can be seen in Figure 5.20 and Figure 5.21. Comparing all the different techniques, MED filtering is definitely the most powerful technique to enhance the impulses.

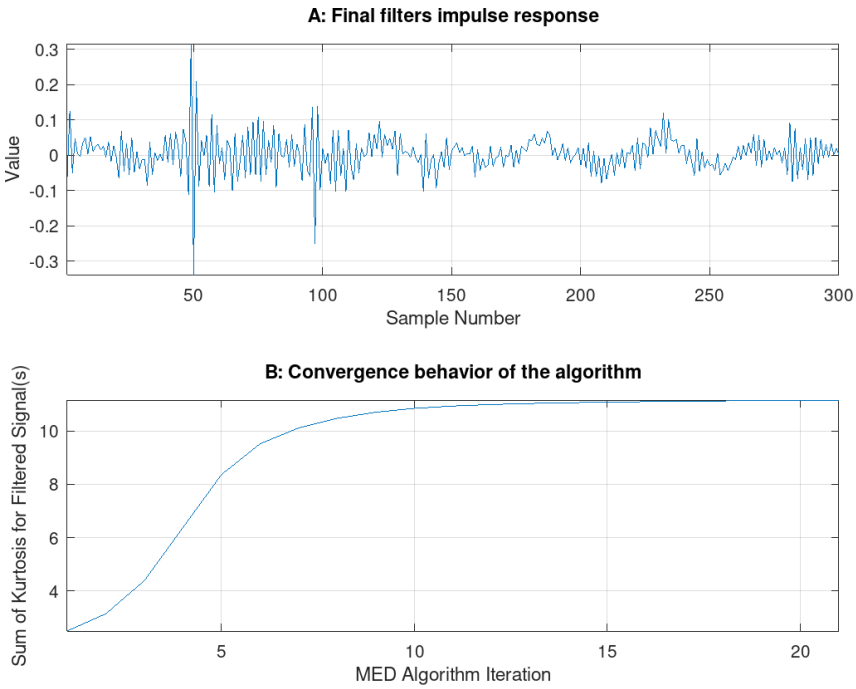

Figure 5.18: Subplot A: Impulse response of the Minimum-Entropy-Deconvolution-filter. Subplot B: Convergence-behavior of the Minimum-Entropy-Deconvolution-algorithm regarding the number of iterations for filter order L=300.

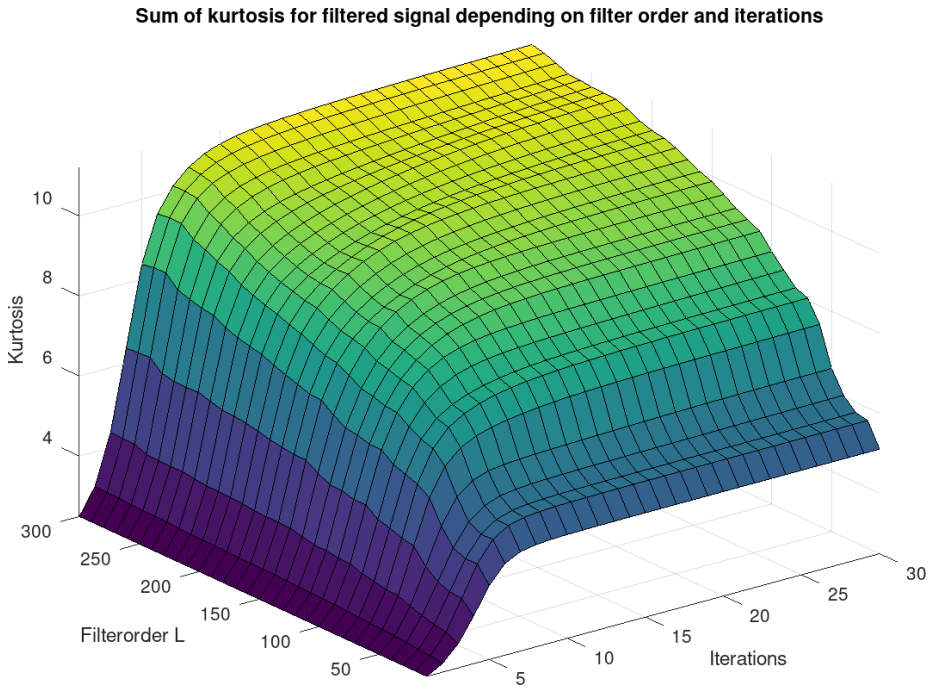

Figure 5.19: Two dimensional surface representing kurtosis over filter order L and iterations of the algorithm.

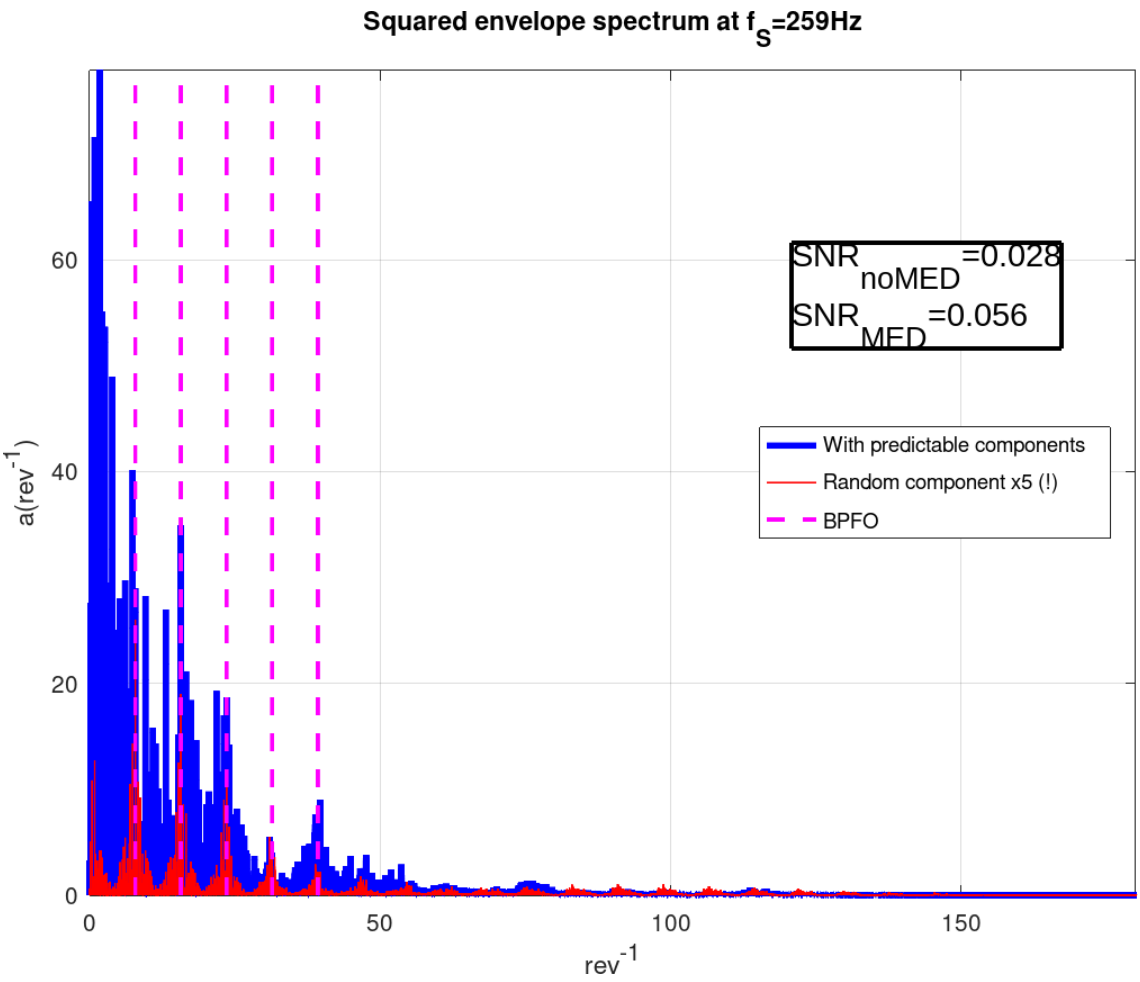

Figure 5.20: Spectrum of a signal enhanced with Order-Tracking, Discrete-Random-Separation, Minimum-Entropy-Deconvolution and Squared-Envelope-Spectrum for an outer race fault.

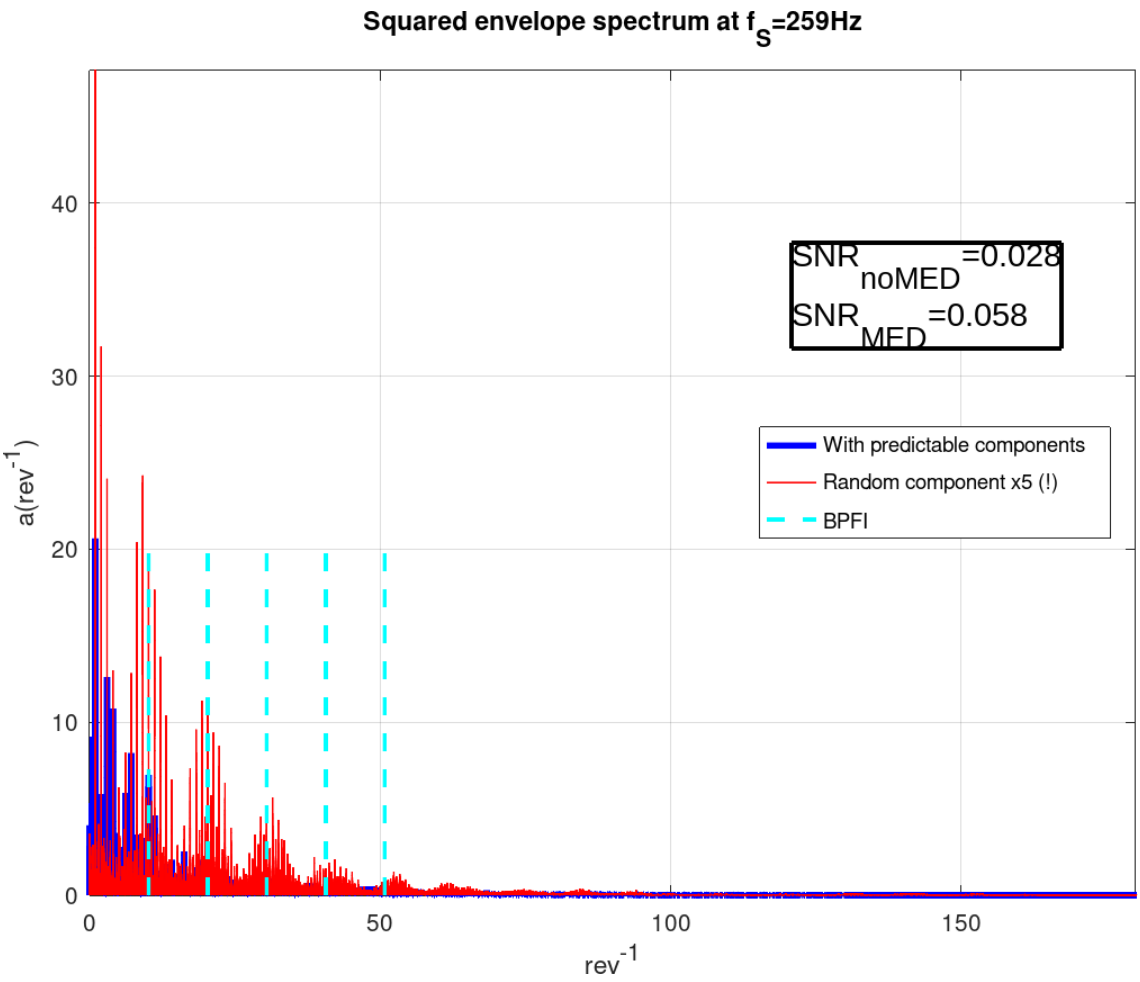

Figure 5.21: Spectrum of a signal enhanced with Order-Tracking, Discrete-Random-Separation, Minimum-Entropy-Deconvolution and Squared-Envelope-Spectrum for an inner race fault.

5.5 Spectral Kurtosis (SK) - Fast Kurtogram (FK)

The signal fed into the spectral-kurtosis-algorithm is the MED-signal from section 5.4.

After MED filtering the question arises, what the most suitable band for demodulation is. This task can be solved by the SK which helps to identify the frequency band containing the most impulsiveness. The SK extends the concept of the kurtosis to the frequency domain. Figure 5.22.*A* shows a simulated bearing fault signal, Figure 5.22.*B* shows the Short-Time-Fourier-Transform (STFT) of the signal and Figure 5.22.*C* shows the kurtosis of the STFT calculated along its time domain and the optimal band pass filter for filtering out the most impulsive frequency band. For a given window length the kurtosis now only depends on frequency - hence the name *spectral kurtosis*. The shape of the spectral kurtosis curve changes with the window length of the STFT which can be shown in the *kurtogram* in Figure 5.23. The value with the highest kurtosis represents the frequency band with the highest impulsiveness and therefore should be taken as the center frequency for further band pass filtering. For efficient calculation of the kurtogram, the *fast kurtogram* (Figure 5.24 and Figure 5.25) was developed, which - instead of calculating the SK for every window length - splits the window lengths up according to a binary tree structure and this way calculates a cascade of spectral kurtoses. The advantage of binary tree structures is, that now filter banks can be used which allow faster computation with only little drawbacks regarding resolution compared to the continuous kurtogram. Instead of using the STFT the wavelet transform and wavelet filter banks can further improve this technique.[4]

In the case of the flywheel, the algorithm to calculate the FK was taken from [9] which uses a *decimated filter bank tree* to compute the wavelet packet transform. The results are shown in Figure 5.24 and Figure 5.25 and suggest, that the optimal band for band pass filtering with the highest kurtosis is at decomposition level 0. Hence, the suggested optimal band pass filter has a bandwidth of 180 Hz and a center frequency of 90 Hz which actually means, that no band pass filtering should be done. The resulting processed spindle signal of this technique in time domain can be seen as the blue curve in subplot D of Figure 5.1 and subplot D of Figure 5.2. This technique is still applied, because in some case it might improve the spectra and for the current case it at least does not worsen it. The unchanged spectra can be seen in subplot 4 of Figure 5.26 and Figure 5.27 which equals exactly the spectra after MED's (section 5.4) and SES's (section 5.7) processing.

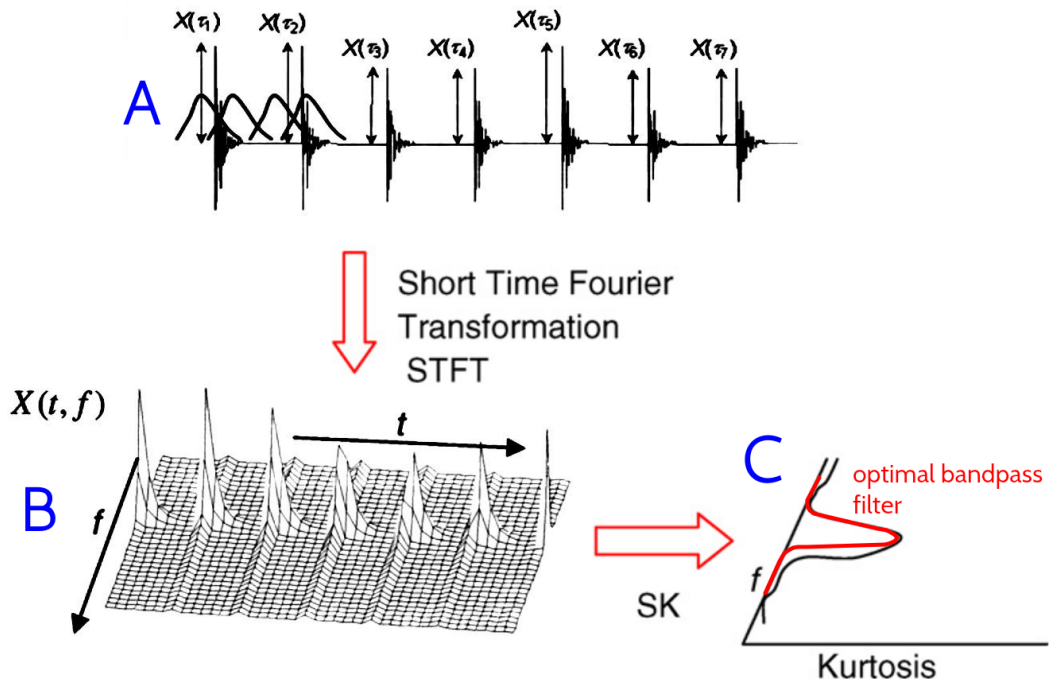

Figure 5.22: Spectral kurtosis shows which frequency band contains the impulsiveness. The red curve in C represents the optimum bandpass filter to enhance the fault signal as good as possible. [4]

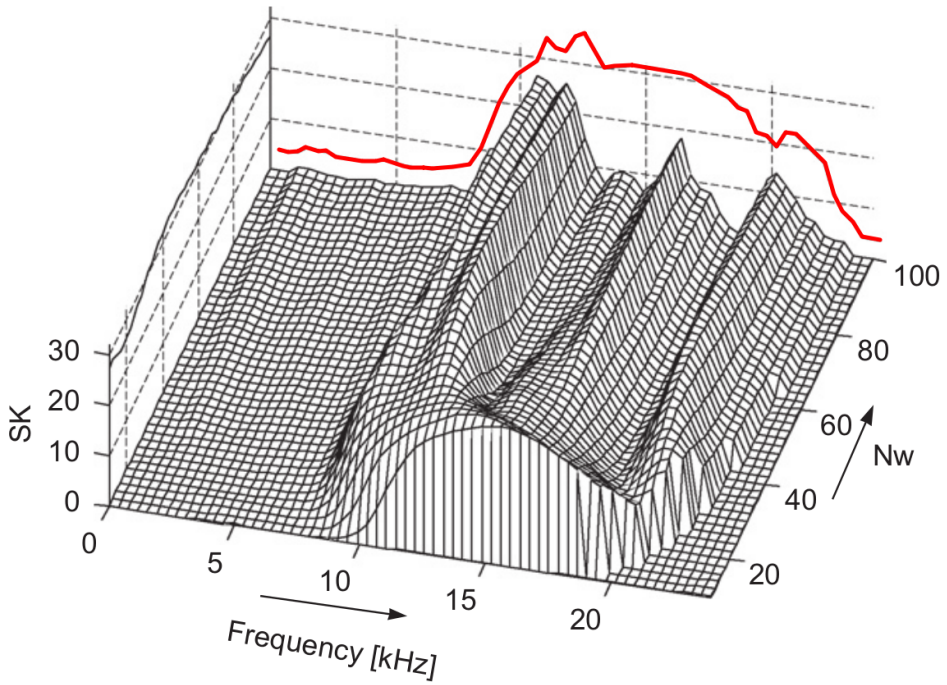

Figure 5.23: Kurtogram for a bearing fault. N_w is the window length of the Short-Time-Fourier-Transform. The projected red curve is the Spectral-Kurtosis for a given window length. [4]

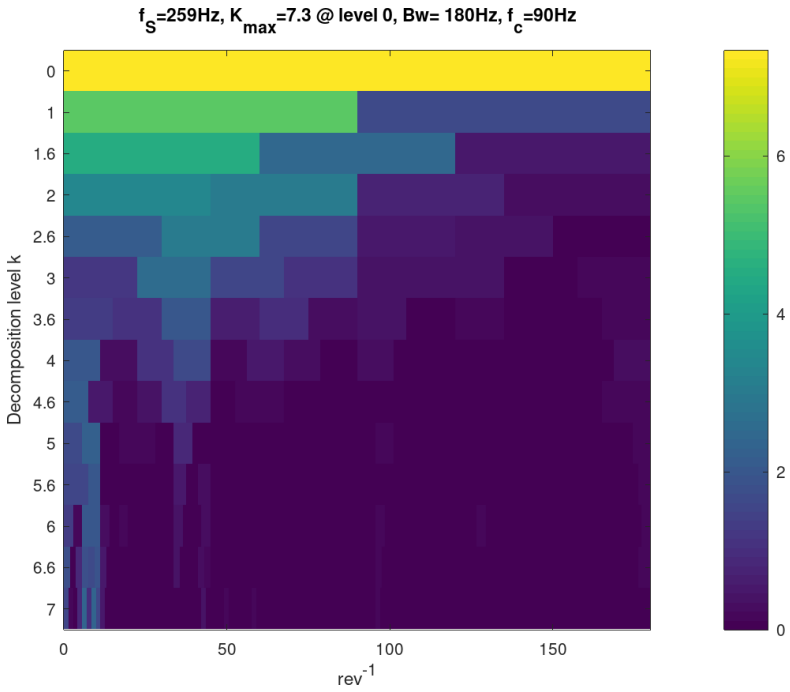

Figure 5.24: Fast kurtogram of the outer race fault. The fast kurtogram represents the amplitude of the kurtosis on a color scale. The algorithm to calculate the fast kurtogram was taken from[9]

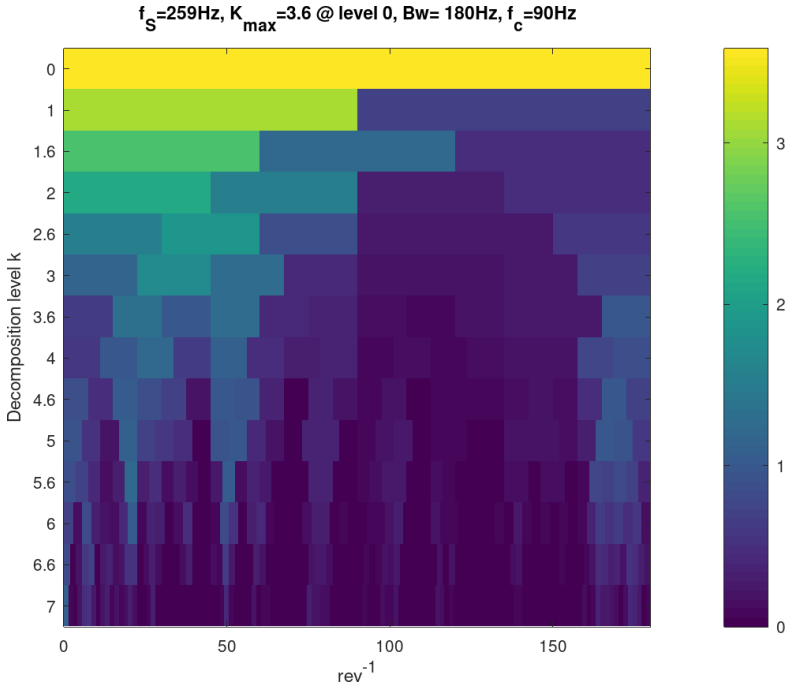

Figure 5.25: Fast kurtogram of the inner race fault. The fast kurtogram represents the amplitude of the kurtosis on a color scale. The algorithm to calculate the fast kurtogram was taken from [9]

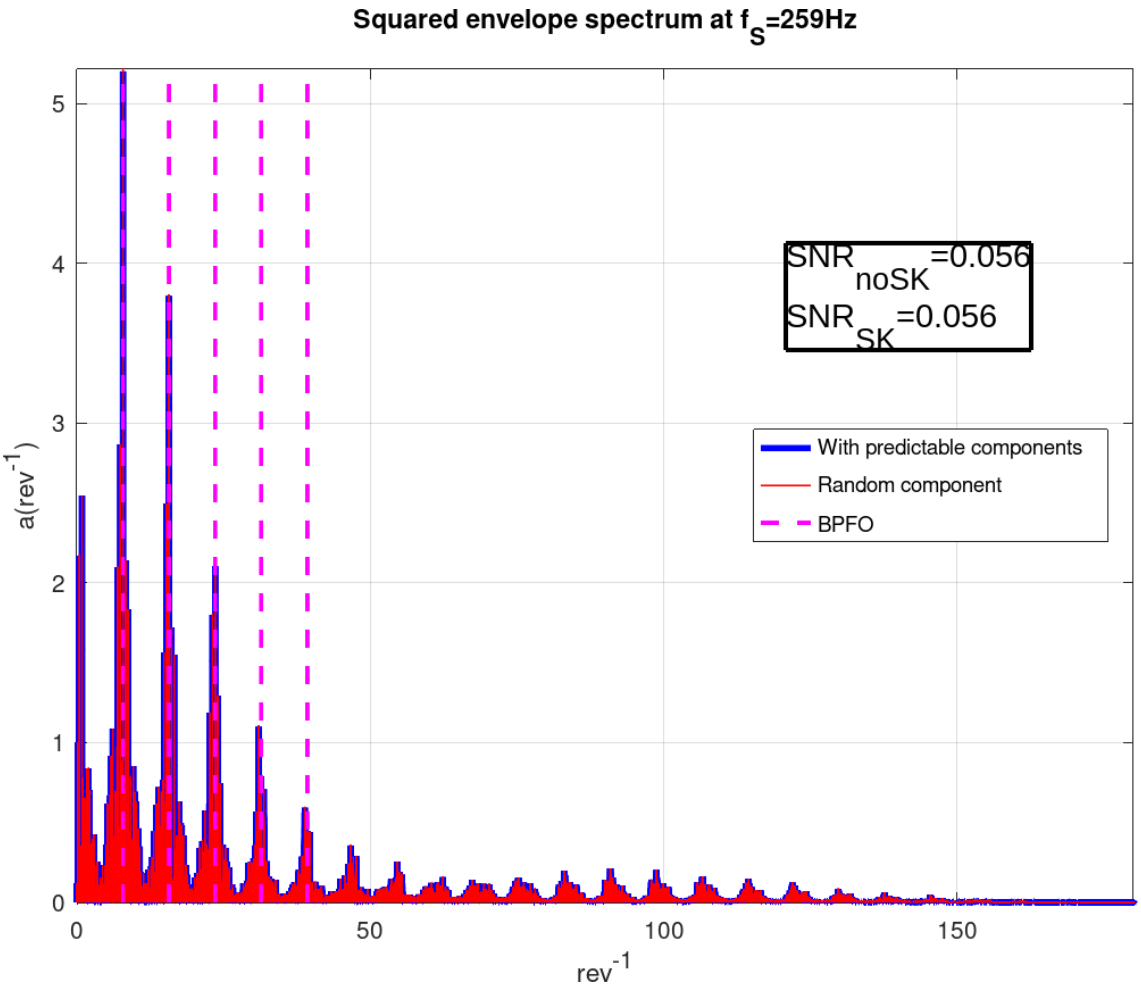

Figure 5.26: Spectrum of a signal enhanced with Order-Tracking, Discrete-Random-Separation, Minimum-Entropy-Deconvolution, Spectral-Kurtosis and Squared-Envelope-Spectrum for an outer race fault.

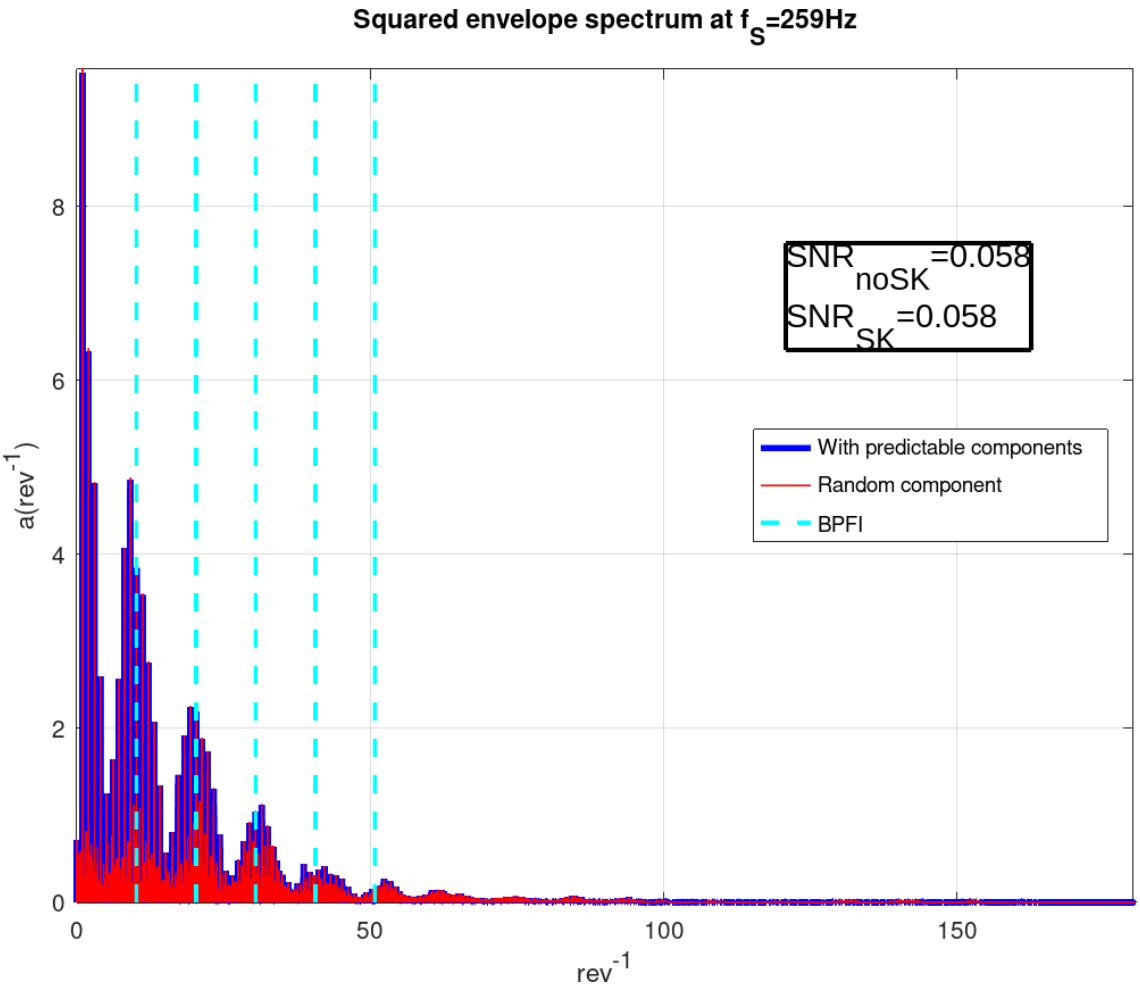

Figure 5.27: Spectrum of a signal enhanced with Order-Tracking, Discrete-Random-Separation, Minimum-Entropy-Deconvolution, Spectral-Kurtosis and Squared-Envelope-Spectrum for an inner race fault.

5.6 Wavelet denoising (WD)

Wavelet denoising is an alternative technique to the spectral kurtosis. The advantage of this denoising technique comes from the fact, that it can separate noise from a signal, even if their frequency spectra are overlapping. The basic approach in this example is, that first the MED filtered signal is wavelet transformed using a filter bank which outputs the wavelet coefficients according to its filter levels. Then these coefficients are manipulated according to the thresholding technique one uses. The three thresholding techniques available in GNU OCTAVE are *hard*, *wiener* and *soft* thresholding illustrated in Figure 5.28. The chosen technique was hard thresholding, because in theory it should better keep the impulsive edges in the signal. Testing showed, that it does return a higher SNR than the other thresholding techniques. A cutout of the denoised wavelet coefficients can be seen in Figure 5.29 and Figure 5.30. For choosing the threshold the SUREshrink algorithm described in [11] was used. The used wavelet was called *db1* in GNU OCTAVE which seemed suitable for this task. Because WD only worsened the SNR this technique was not further investigated. The spectra can be seen in Figure 5.31 and Figure 5.32.

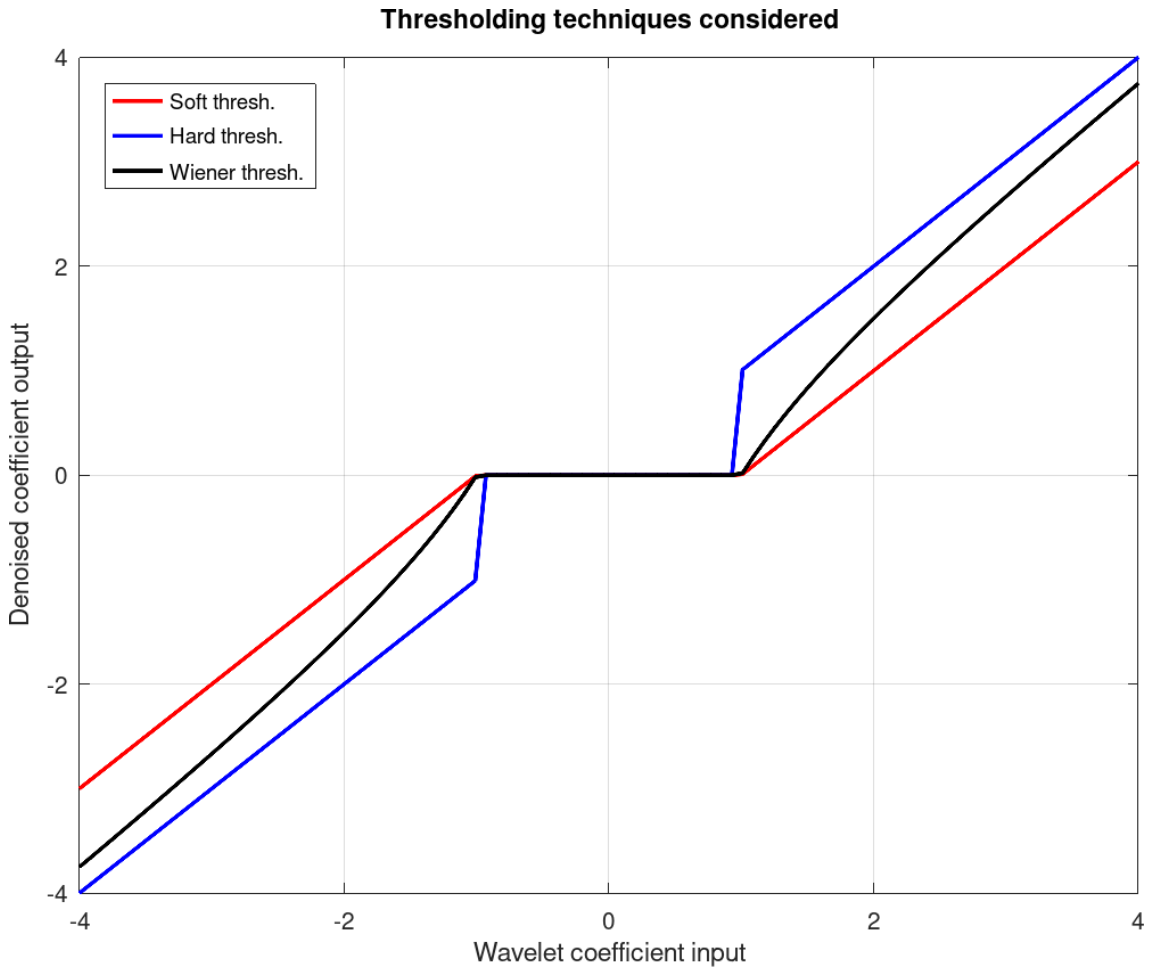

Figure 5.28: Thresholding techniques for wavelet denoising.

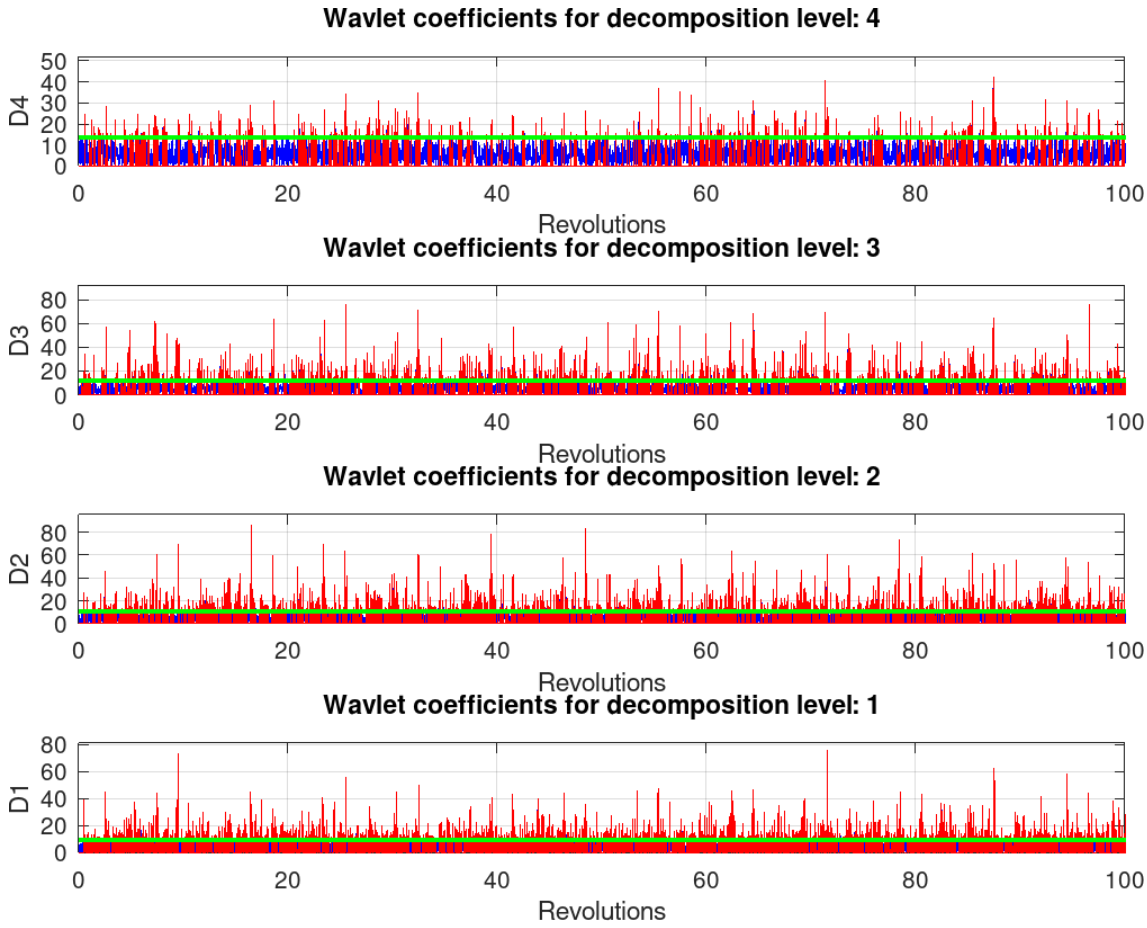

Figure 5.29: Wavelet coefficients for an outer ring fault. Blue: original wavelet coefficients. Red: denoised wavelet coefficients. Green: threshold for denoising.

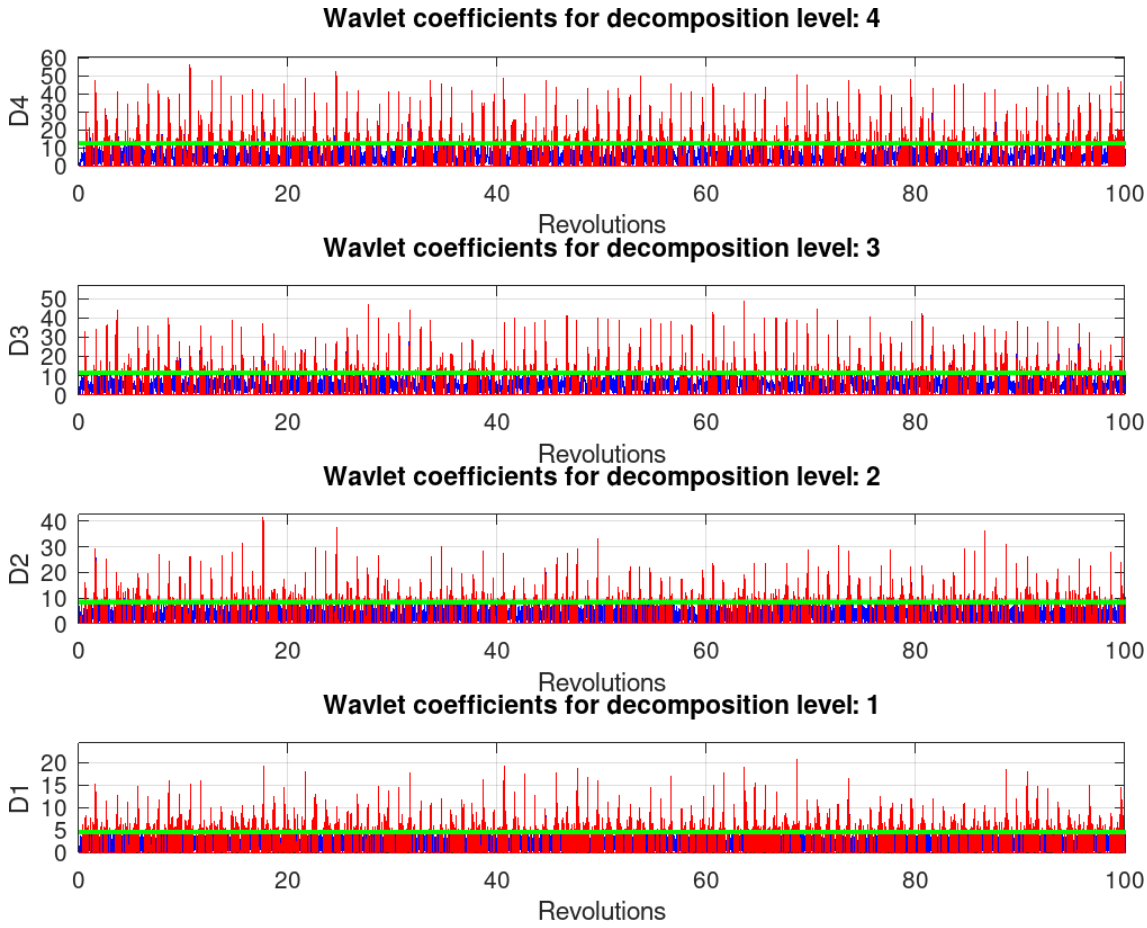

Figure 5.30: Wavelet coefficients for an inner ring fault. Blue: original wavelet coefficients. Red: denoised wavelet coefficients. Green: threshold for denoising.

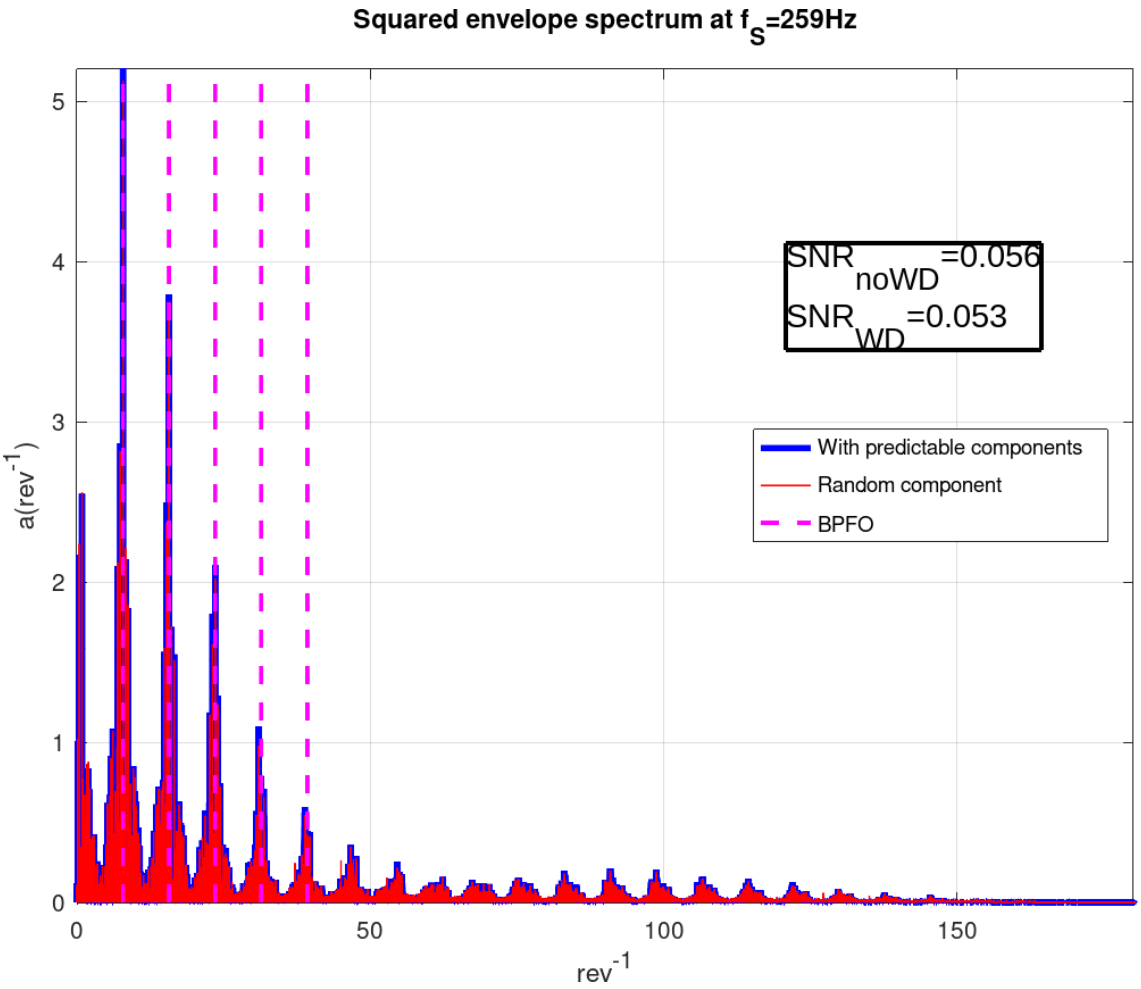

Figure 5.31: Spectrum of a signal enhanced with Order-Tracking, Discrete-Random-Separation, Minimum-Entropy-Deconvolution, Wavelet-Denoised and Squared-Envelope-Spectrum for an outer race fault.

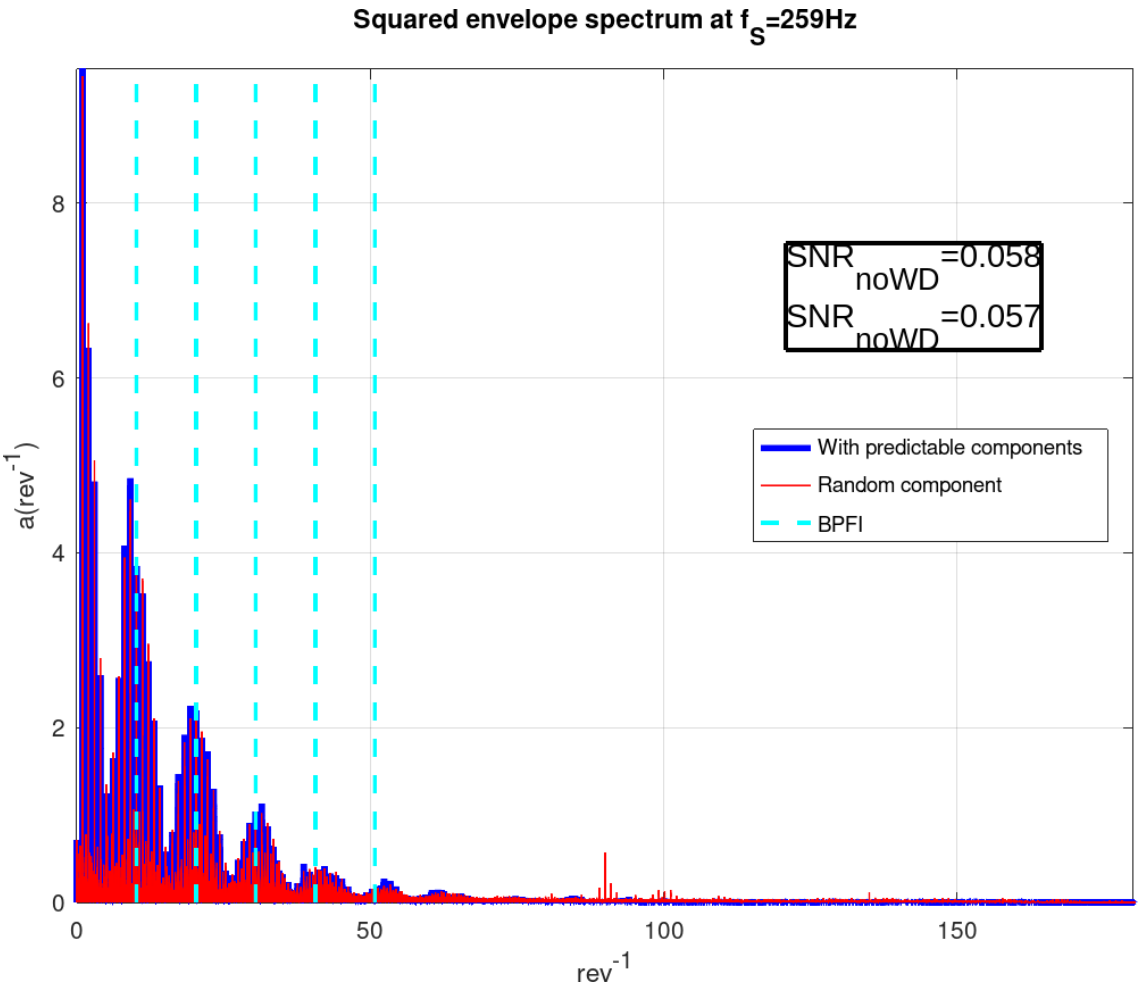

Figure 5.32: Spectrum of a signal enhanced with Order-Tracking, Discrete-Random-Separation, Minimum-Entropy-Deconvolution, Wavelet-Denoised and Squared-Envelope-Spectrum for an inner race fault.

5.7 Squared-Envelope-Spectrum (SES) and Spectral Correlation (SC)

When it comes to visualizing (pseudo-) cyclostationary signals one of the most powerful tools might be the so called *spectral correlation* (SC) shown in Equation 5.16. [4]

$$S_{YY}(\alpha, f) = \lim_{W\to\infty} \frac{1}{W} \int_{\mathbb{R}} \int_{-W/2}^{W/2} R_{YY}(t, \tau)\, e^{-i2\pi(f\cdot\tau-\alpha\cdot t)} dt d\tau \tag{5.16}$$

The variable α is called cyclic frequency and describes a modulation in the time domain which equals a shift in the frequency domain (property of the Fourier transform). This shift in the frequency domain helps to identify certain periodic structures like harmonics, side bands or such, and they become visible as high spots along the ordinary frequency axis, whereas frequencies which are always in the spectrum are lines along the cyclic frequency axis which can be seen in Figure 5.33 and Figure 5.34. However a drawback of the SC is, that it takes long to compute and therefore a second alternative - needing much less computational effort - is the so called Squared-Envelope-Spectrum (SES). As it is shown in the appendix of [4], the integral over the SC along the ordinary frequency axis equals the SES. The nice thing for the SES is, that it has almost no drawbacks regarding the analytic information it offers compared to the SC. According to [4] it offers no drawbacks when it comes to inner race fault of any kind and outer race faults with impulsive excitation, only outer race faults with weak impulsive excitation, e.g. due to heavy wear, might (under bad circumstances) not be detected with the SES which could probably still be detected with the spectral correlation. [4]

As stated in [4] it is of advantage to use the Squared-Envelope-Spectrum instead of a non Squared-Envelope-Spectrum, because this way one will have less artifacts due to the sharp corners which occur due to rectification as can be seen in Figure 5.35. The final result of the squared envelope in time domain can be seen in subplot 4 of Figure 5.1 and Figure 5.2. The final spectrum can be seen in Figure 5.26 and Figure 5.27.

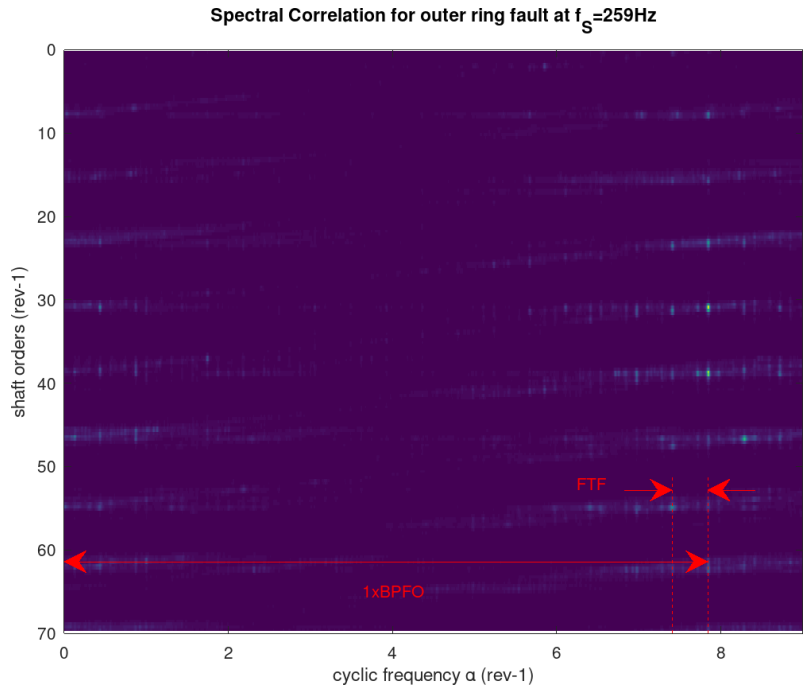

Figure 5.33: Spectral correlation for an outer ring fault (Equation 5.16). One can see, that the Ball-Pass-Frequency-Outer (BPFO) and the Fundamental-Train-Frequency (FTF) of the cage are dominant in this case of fault.

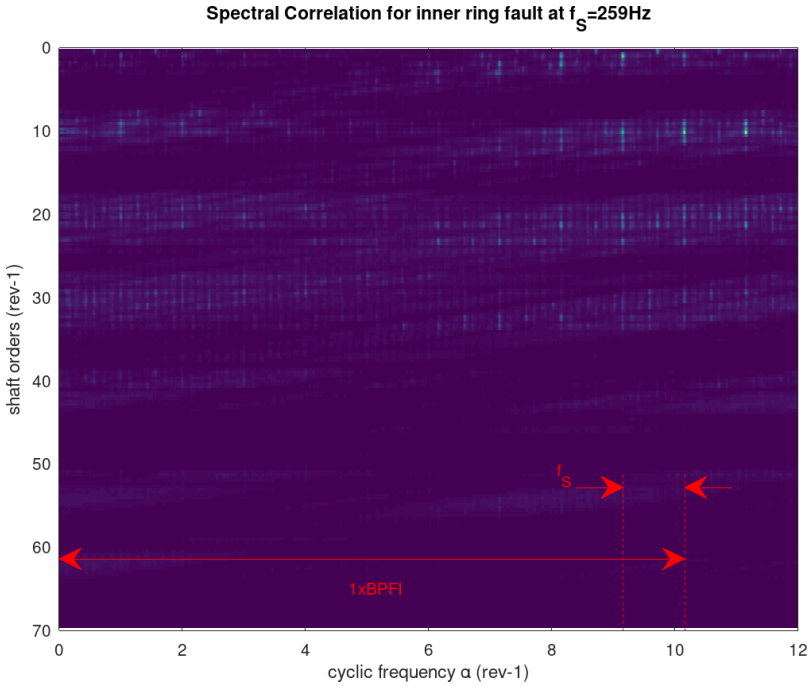

Figure 5.34: Spectral correlation for an inner ring fault (Equation 5.16). One can see, that the Ball-Pass-Frequency-Inner (BPFI) and the spindle frequency f_S is dominant in this case of fault.

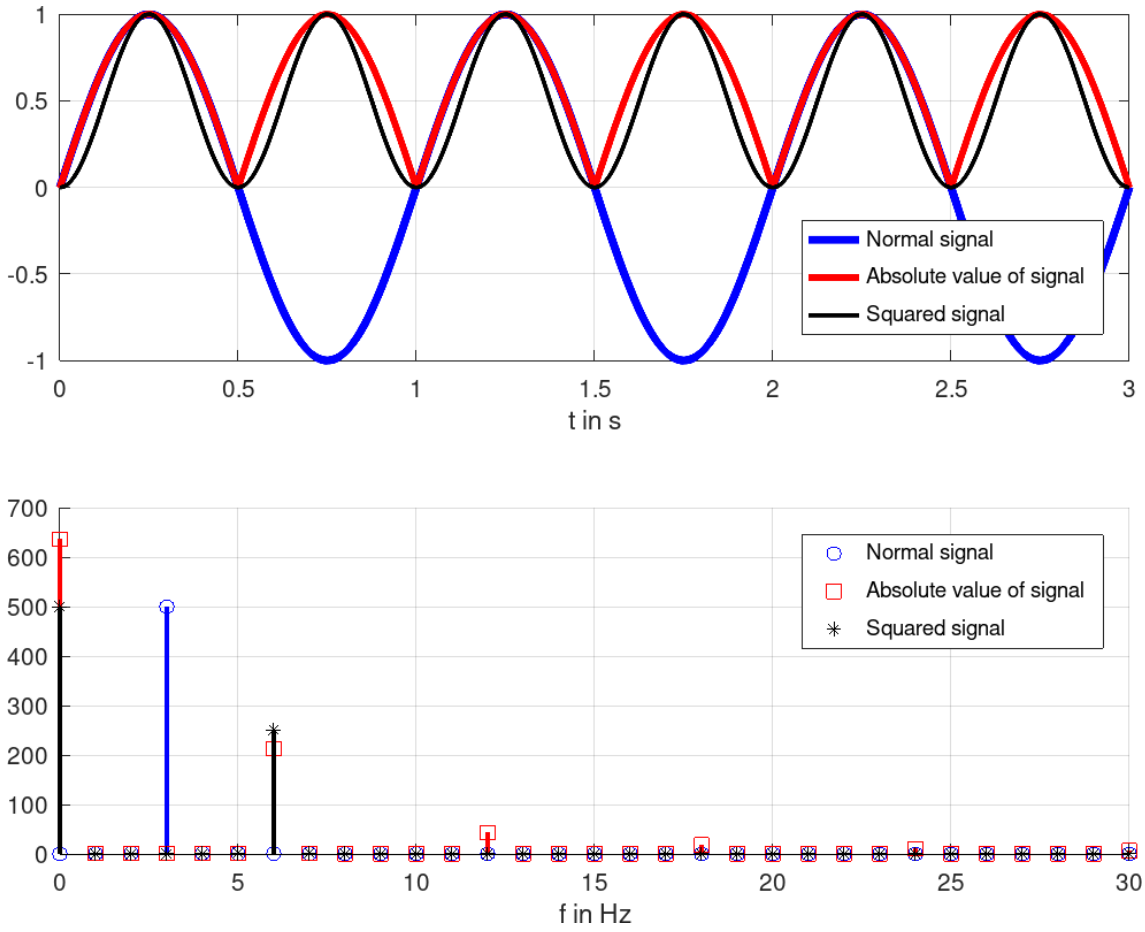

Figure 5.35: Envelope spectrum of normal and squared signal makes a difference when it comes to harmonics.

5.8 Thresholding (SNR-threshold)

After investigating the possible signal enhancement techniques, now the question arises, how an algorithm should determine, whether a bearing is defective or not. If the bearing is found to be defective, the next question is, which component is faulty.

The insight for the proposed technique comes from the comparison of raw spectra and processed spectra. The raw spectra comparing all different unprocessed bearing scenarios, namely *outer ring fault*, *inner ring fault* and *no fault* in a logarithmic diagram can be seen in Figure 5.36. The comparison of the three processed spectra can be seen in Figure 5.37, which shows, that the signal power of the processed signal drops significantly compared to the unprocessed case. Because the signal power of healthy bearings is much lower than the signal power of defected bearings, any variable circumstance like mounting, or the variance between healthy bearings will be ignored. It is obvious, that defects have much more influence than these changing circumstances.

Now in the case of a healthy bearing the signal power decreases due to processing significantly which is not the case for a faulty bearing. In the case of the faulty bearing the signal power increases significantly. A threshold can be set to determine, at which difference of signal power between a processed and an unprocessed signal the bearing is to be considered as healthy or defective. And by looking at the SNR calculated with the technique described in section 5.1 for an outer- or an inner race fault, one might be able to judge whether the bearing has an outer-ring-fault, an inner-ring-fault or both faults, just by comparing the level of the current SNR with a predefined SNR-threshold.

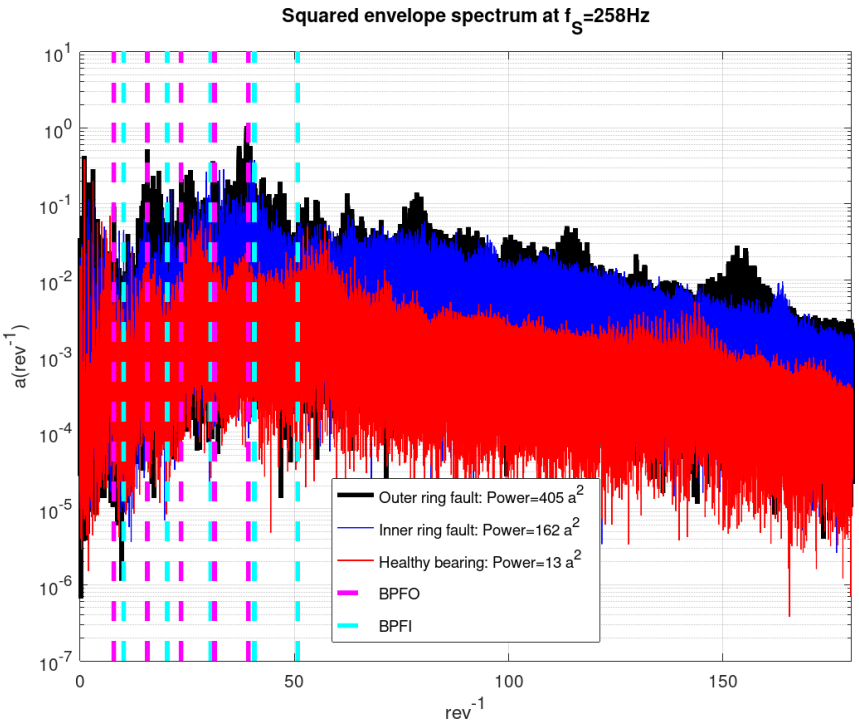

Figure 5.36: Raw spectrum of unprocessed vibrational signals.

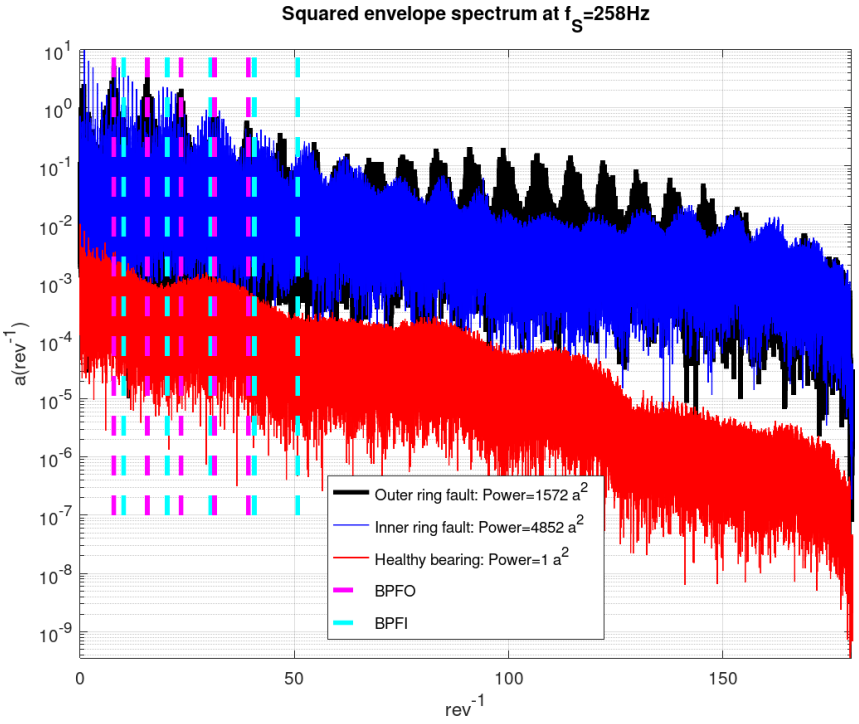

Figure 5.37: Spectrum of Order-Tracking, Discrete-Random-Separation, Minimum-Entropy-Deconvolution, Spectral-Kurtosis and Squared-Envelope-Spectrum processed vibrational acceleration signals.

6 Conclusion and outlook

When it comes to the sources of vibration, it can be shown (Figure 5.36), that in the conducted experiments the damaged bearings result in a much higher vibrational signal power. This leads to the conclusion, that defects like surface roughness, waviness or deviations due to mounting have much less influence, than localized faults.
When it comes to the enhanced spectra, the real difference regarding the difference of signal power becomes obvious - the bearings with a localized fault have a signal power in the order of 10^3 bigger than the healthy bearing as is shown in Figure 5.37. Due to the lack of bearings with non localized faults, it cannot be known if the enhancement algorithm would also enhances those spectra as well as for localized faults.
Tasks still worth investigating regarding the damage of bearings would be:

- Checking which kind of damage occurs due to lack of lubrication.
- Checking if the proposed spark erosion method described in section 3.3 results in similar spectra compared to spectra of real world faults. Also a comparison between the wire-EDM technique described in section 3.1 and the spark erosion method described in section 3.3 should be made. If this was the case, a much cheaper method for damaging bearings generating localized faults in a repeatable manner, would have been found.
- Checking if the spark erosion method described in section 3.3 is also suitable for making spalled areas. It could be done, by damaging a whole segment and then using a sand paper to smooth down the edges. If this was the case, also a much cheaper method for damaging bearings resulting in spalled areas in a repeatable manner would have been found.

Regarding the state of the art signal processing techniques *Order* Tracking (OT), *Discrete Random Separation* (DRS), *Minimum Entropy Deconvolution* (MED), *Spectral Kurtosis* (SK) and *Squared Envelope Spectrum* (SES) have been identified as the processing techniques resulting in the best SNR when applied in combination which can be seen in Table 5.1. The benchmark-test for this statement was the proposed SNR-filter (section 5.1).
For future benchmark tests the signal power of the processed signal should also be considered next to the SNR.

All state of the art techniques should furthermore be tested on a wider variety of defects (because the only defects tested in the course of this thesis were multiple localized defects):

- single faults
- multiple faults with very close spacing
- spalled surfaces with an impulsive excitation
- worn spalled surfaces with or without little impulsive excitation

When comparing the milling-spindle test rig to the flywheel one difference is, that in the case of the flywheel the suspension of the flywheel is assumed to lead to a much less noisier measurement surrounding, than in the case of the milling spindle. Furthermore the acceleration sensor will be mounted closer and more rigidly to the bearing, which will lead to better signal quality compared to the milling-spindle. This will allow for a lower detection SNR-threshold which is described in section 5.8. Finally, a more advanced machine learning algorithm can be developed which is able to detect defective bearings, and learns, how bearings behave over their lifetime and therefore it can try to predict when maintenance will be necessary.

Bibliography

[1] An Overview of Bearing Vibration Analysis, Dr. S. J. Lacey, Engineering Manager Schaeffler UK Limited, FAG

[2] R. B. Randall, J. Antoni, The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis, Mechanical Systems and Signal Processing 21, 2007

[3] Schaeffler, Handbuch zur Gestaltung und Berechnung von Wälzlagerungen - Chapter 9, S. 772, Wälzlagerpraxis vierte Auflage 2015, Vereinigte Fachverlage

[4] R. B. Randall, J. Antoni, Rolling element bearing diagnositcs - A tutorial, Mechanical Systems and Signal Processing 25, 2010

[5] H. Endo, R. B. Randall, Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter, Mechanical Systems and Signal Processing 21, 2007

[6] R. B. Randall, J. Antoni, S. Chobsaard, The relationship between spectral correlation and envelope analysis in the diagnoistics of bearing faults and other cyclostationary machine signals, Mechanical Systems and Signal Processing 15, 2001

[7] J. Antoni, R. B. Randall, Differential diagnosis of gear and bearing faults, ASME Journal of Vibration and Acoustics 124 (202) 165-171

[8] Geoff McDonald. Minimum Entropy Deconvolution Multipack (MED, MEDA, OMEDA, MOMEDA, MCKD), MATLAB Central File Exchange. Retrieved January 15, 2021.

[9] J. Antoni, Fast Computation of the Kurtogram for the Detection of Transient Faults, Mechanical Systems and Signal Processing, Volume 21, Issue 1, 2007, pp.108-124.

[10] J. Antoni, G. Xin, N. Hamzaoui, Fast Computation of the Spectral Correlation, Mechanical Systems and Signal Processing, Elsevier, 2017

[11] Florian Luisier, Thierry Blu, Brigitte Forster and Michael Unser, Which wavelet bases are the best for image denoising?, Biomedical Imaging Group, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland Centre for Mathematical Sciences, Munich University of Technology, Munich, Germany, 2005

[12] Structural Testing Part I: Mechanical Mobility Measurements By Ole Døssing, Brüel & Kjæ