

Article

Correlation Analysis and Verification of Railway Crossing Condition Monitoring

X. Liu * and V. L. Markine *

Section of Railway Engineering, Delft University of Technology, 2628 CN Delft, The Netherlands

* Correspondence: Xiangming.Liu@tudelft.nl (X.L.); V.L.Markine@tudelft.nl (V.L.M.)

Received: 25 August 2019; Accepted: 23 September 2019; Published: 26 September 2019

Abstract: This paper presents a correlation analysis of the structural dynamic responses and weather conditions of a railway crossing. Prior to that, the condition monitoring of the crossing as well as the indicators for crossing condition assessment are briefly introduced. In the correlation analysis, strong correlations are found between acceleration responses with irregular contact ratios and the fatigue area. The correlation results between the dynamic responses and weather variables indicate the influence of weather on the performance of the crossing, which is verified using a numerical vehicle-crossing model developed using the multi-body system (MBS) method. The combined correlation and simulation results also indicate degraded track conditions of the monitored crossing. In the condition monitoring of railway crossings, the findings of this study can be applied to data measurement simplification and regression, as well as to assessing the conditions of railway crossings.

Keywords: railway crossing; condition monitoring; condition indicator; correlation analysis; weather impact; numerical verification

1. Introduction

Railway turnouts are essential components of railway infrastructure and provide the ability for trains to transfer from one track to the other. In the meantime, a gap between the wing rail and nose rail in the crossing panel (Figure 1b) introduces a discontinuity in the rail. As a result of trains passing through, the high wheel–rail impact due to the high train velocity causes this type of crossing to suffer from severe damage such as cracks (Figure 1c) and spalling (Figure 1d), and the service lives of some railway turnouts are only 2–3 years. In order to improve the maintenance of the crossing and prolong service life, it is better to perform maintenance in a predictive way by developing a structural health monitoring (SHM) system for railway crossings [1].

In order to obtain information on damage detection, localization and condition assessment in SHM systems, it is important to get insight into the performance of the structures. In recent years, SHM has drawn increasingly more attention in the railway industry. D. Barke and W.K. Chiu reviewed the major contributions of condition monitoring in regards to wheels and bearings [2]. Based on digital image correlations, D. Bowness et al. measured railway track displacement using a high speed camera [3]. The axle box acceleration (ABA) system has been widely applied in the condition monitoring [4] and damage detection [5,6] of railway tracks. However, most of the contributions of SHM are based mainly on normal tracks. Z. Wei et al. have applied the ABA system in railway crossing damage detection [7]. However, as a special and vulnerable component in the railway track system, the study on crossings in terms of condition monitoring are still limited.

In the existing studies, the performance analysis of crossing has been based mainly on numerical approaches. For instance, finite element (FE) wheel-crossing models have been applied to calculate plastic deformation and frictional work [8], to simulate the distribution of stresses in the crossing nose [9] and to predict the fatigue life of a crossing [10]. Also, multi-body system (MBS) vehicle-crossing models

have been used for general train-track interaction analysis [11], track elasticity analysis [12], crossing geometry optimization [13–15] and so on. Due to restricted track access, high costs and time consumption, field measurements have mainly been used for numerical model validation [9,16]. The numerical models are usually developed according to a certain hypothesis with a focus on specific problems. However, for damage detection and assessments of crossing conditions, the numerical approach alone is not enough, and monitoring the conditions of in-service railway crossings is highly necessary.

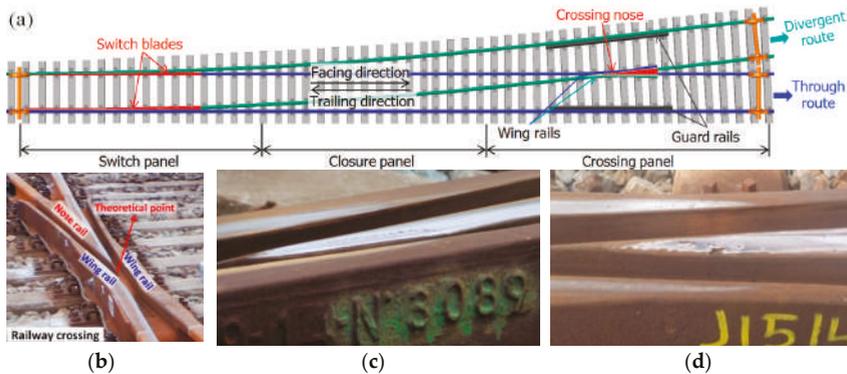


Figure 1. (a) Standard left-hand railway turnout with a 1:9 crossing (drawn by X. Liu); (b) crossing panel on site (shot by V.L. Markine); (c) plastic deformation with cracks (shot by X. Liu); (d) spalling (shot by X. Liu).

In real life, the wheel–rail contact in a crossing can be affected by many factors. Some factors are related to the train track system, such as train type [17], velocity [18], axle load [18,19], wheel–rail friction [18], crossing geometry [18,19], track alignment [19], track elasticity [12] and so on. Some factors are related to the crossing environment, such as the contaminants on the rail [19–21] and rail temperature variation [22,23]. All these factors, especially those introduced by the environment, make the measurement data noisy and the crossing condition cannot be clearly shown [24]. In order to properly analyze the measurement data for monitoring the crossing condition, the first step is to figure out the influence of the above mentioned factors on the performance of the crossing.

In this study, the influence of train track system-related factors is minimized through data selection and a filtering process. Specifically, train type, velocity and the bogie number are restricted to a certain scope. In order to estimate the influence level of the external factors (such as the weather condition), a correlation analysis using Pearson’s correlation coefficient, which is usually applied to quantitatively evaluate the correlation strength between two variables, is performed. The correlation analysis results are verified using a vehicle-crossing model developed using the multi-body system (MBS) method. In this model, the weather changes are modelled according to changes in the properties of the affected track elements. The correlation between the weather condition and the dynamic responses of the crossing provides the foundation for long-term measurement data regression, which will be applied in the crossing degradation assessment procedure. In addition to weather factors, the correlation strengths between the dynamic responses of the crossing are also analyzed, which can be applied to provide guidance for the selection and post-processing of the measurement data and to improve the efficiency of analyzing a large amount of data.

The paper is organized as follows. The condition-monitoring procedure of a railway crossing, including the crossing instrumentation, is presented in Section 2. The indicators applied for the crossing condition assessment are briefly introduced in Section 3. The correlation analysis, including the dynamic responses and weather variables, are given in Section 4. In Section 5, the mechanisms of the weather effects are analyzed and verified through numerical simulation. Finally, in Section 6, the conclusions based on the correlation analysis are provided and further applications for the degradation procedure description of the monitored crossing are discussed.

2. Railway Crossing Condition Monitoring

In this section, monitoring the condition of a railway crossing is discussed. The crossing instrumentation and a brief procedure for processing the measurement data are described.

2.1. Crossing Instrumentation

The monitored crossing in this study is a cast manganese steel crossing with an angle of 1:9, which is the most commonly used crossing for Dutch railway tracks (more than 60% [25]). As part of a double crossover, the crossing is mainly used for through-facing routes (Figure 1a). This railway line is mainly used for passenger transportation with a velocity of passing trains up to 140 km/h. The crossing is instrumental for using the system that has been introduced, and has been actively used in previous studies [1,17,19,26]. An overview of the crossing instrumentation is given in Figure 2.

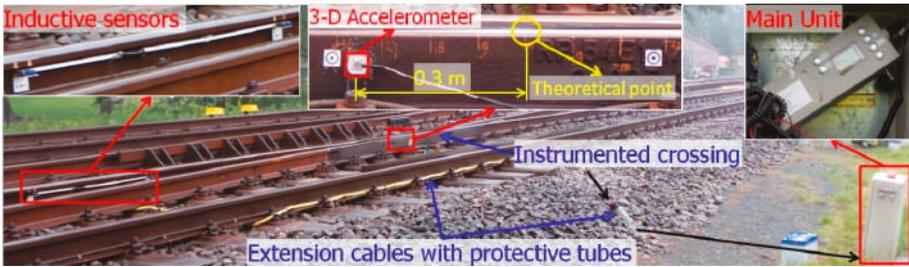


Figure 2. Overview of the crossing instrumentation.

The main components of this device are a 3-D accelerometer attached to the crossing rail, a pair of inductive sensors attached to the rails in the closure panel and the data logger (main unit) installed on the outside of the track. The inductive sensors are used for train detection and the initiation of the measurements, as well as for train velocity determination. All of the sensors are connected to the data logger for data storage and basic analysis of the data. The measurement range and sampling frequency of the acceleration sensor are 500 g and 10 kHz, respectively. The main measured data are the 3-D acceleration responses (i.e., a_x , a_y and a_z) of the crossing due to the passing trains.

An example of the vertical acceleration response in a time domain due to one passing train with 12 wheelsets is shown in Figure 3a. It can be seen from this figure that the time and location of each wheel's impact on the crossing can easily be obtained from the acceleration responses. The region where most of the wheel impact is located is defined as the fatigue area (Figure 3b), which can be used for assessing crossing conditions based on a large amount of data.

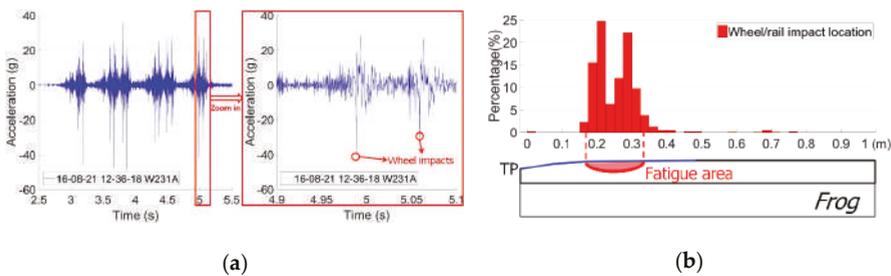


Figure 3. Examples of the output of crossing instrumentation. (a) Vertical acceleration response due to one train's passage; (b) wheel impact location distribution.

2.2. Measurement Data Selection and Processing

The crossing monitored in this study was in a new state at the beginning of the observations. In order to reduce the influence of vehicle variations, the measurement results considered here were restricted to one type of train, namely the VIRM (double-deck) trains that pass with a velocity of around 140 km/h. Moreover, the accelerations caused only by the first bogie were considered. Thus, the remaining uncertainties in the measured data mainly coming from the environment (e.g., the weather). Depending on the amount of monitoring data, the measurement results will be analyzed on three different levels, namely,

- the dynamic response due to the passage of a single wheel;
- the results of multiple-wheel passages from one monitoring day; and
- the statistical results from multiple monitoring days.

An example of vertical acceleration responses in different levels is shown in Figure 4.

The response due to single wheel passages was directly obtained from the measured time domain signal (Figure 4a). The distribution of the maximum impact acceleration from each passing wheel constituted the results of multiple wheel passages (Figure 4b). For the statistical results from multiple monitoring days, each point represented the average value of the impact vertical accelerations of the recorded passing wheels from one monitoring day (Figure 4c). It can be seen that each wheel passed the railway crossing differently. Based on a single wheel's passage it is difficult to assess the performance of the crossing. Yet, some conclusions on wheel–rail interaction can still be drawn based on these data. The statistical analysis based on multiple passing wheels was more applicable for assessing the condition of the railway crossing.

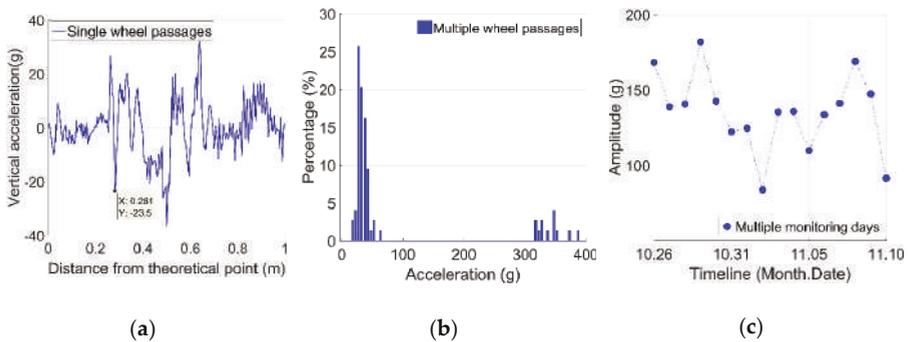


Figure 4. Example of measured vertical acceleration responses: (a) from single-wheel passages; (b) from multiple-wheel passages from one monitoring day; (c) from multiple monitoring days.

3. Condition Indicators

In this section, the indicators for assessing a crossing's condition are briefly described. These indicators are calculated based on the transition region and consist of the irregular contact ratio, 3-D acceleration responses and the fatigue area. To demonstrate the condition analysis procedure, some typical examples of the measurement results from the monitored crossing are presented.

3.1. Transition Region

The transition region of a crossing is the location where the wheel load is transferred from the wing rail to the nose rail (or vice versa, depending on the traveling direction). In practice, the wheel–rail contact points in the crossing can be recognized by looking at the shining band on the rail surface. An example of such a band on the monitored crossing is given in Figure 5 and denoted by the red triangle areas.

3.3. 3-D Acceleration Responses

For the monitored crossing, the regular and irregular contact wheels showed dramatic differences in the 3-D impact acceleration responses (a). For regular passing wheels, the impact vertical acceleration was usually below 50 g, while such impact could be above 300 g for irregular passing wheels. Examples of the 3-D acceleration responses from typical regular and irregular passing wheels are shown in Figures 7 and 8, respectively. In order to better understand the wheel–rail contact, the transition region obtained from field observation (Figure 5) is marked in the figures as a green line on the horizontal axis.

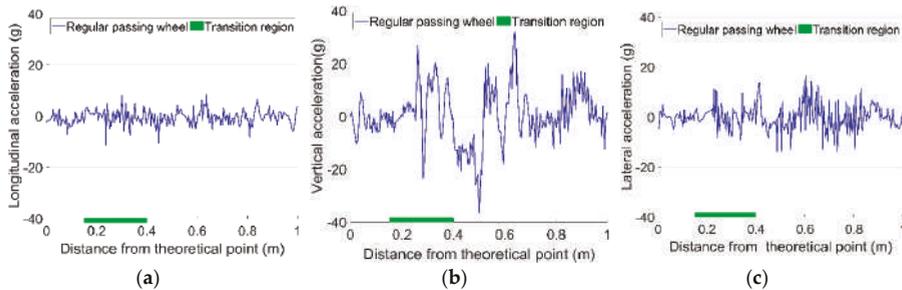


Figure 7. Examples of regular impact acceleration responses due to passing wheels. (a) Longitudinal acceleration; (b) Vertical acceleration; (c) Lateral acceleration.

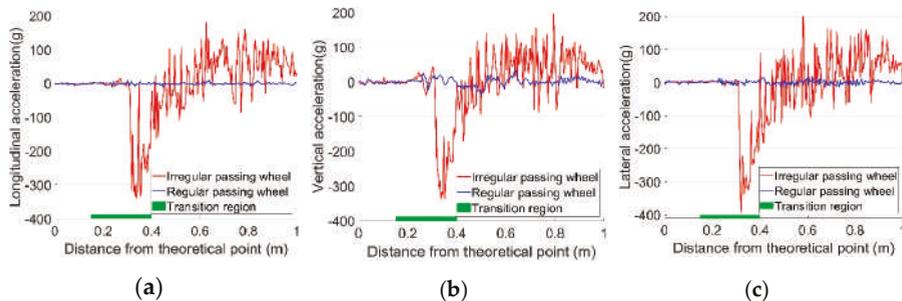


Figure 8. Examples of regular (same as Figure 7) and irregular impact acceleration responses due to passing wheels. (a) Longitudinal acceleration; (b) Vertical acceleration; (c) Lateral acceleration.

It can be seen from Figure 7 that a_y is much higher than a_x for a regular passing wheel, while a_z , meaning that the impact factor (a_z/a_y), is relatively small. It is also indicated that the wheel has two impacts on the crossing, with the first one (22 g) in the transition region and the second one (34 g) after the wheel load transit to the crossing nose rail. Even though the second impact has a higher amplitude, the first one is more damaging, since in the first impact location the nose rail is much thinner than in the second one.

For the irregular passing wheel presented in Figure 8, it can be seen that the impact was located in the transition region, and the accelerations in all three directions were very close to each other (in contrast to the regular passing wheel). Such a strong correlation of the acceleration responses reflects the intense wheel impact on the crossing nose rail and the rough transition of the wheel load from wing rail to the crossing nose rail. The big difference between the two typical wheel–rail impacts gives an example of the violent fluctuation of dynamic response results that can be observed in such crossings.

3.4. Impact Location and Fatigue Area

The impact location is defined as the point where the maximum wheel–rail impact occurs. As described previously, the impact location is restricted within the transition region. For the example given in Figure 4a, the impact location was 0.281 m from the TP.

The fatigue area is defined as the region where most of the wheel impacts are located and is calculated based on multiple wheel passages. In monitoring the conditions of railway crossings, the location and size of the fatigue area reflect the wheel load distribution along the crossing nose. In general, farther impact locations from the TP and wider fatigue areas indicate a better crossing condition.

In practice, to simplify the calculation procedure, the distribution of the wheel impacts due to multiple wheel passages is assumed to be normal distribution, the mean value a is the impact location and the confidence interval $[a - \sigma, a + \sigma]$ is recognized as the fatigue area. An example of the fatigue area of the monitored crossing during a single day is given in Figure 9.

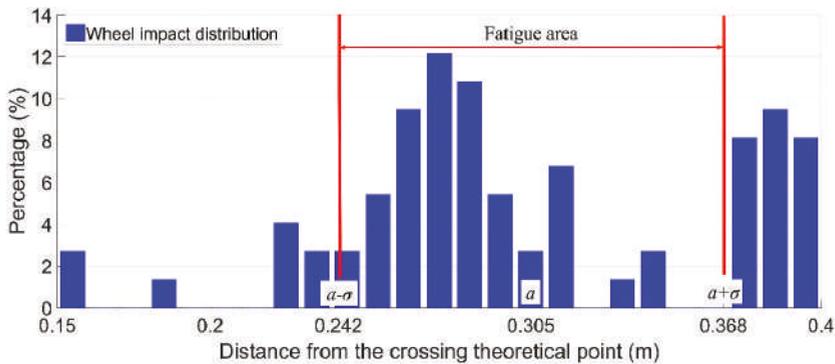


Figure 9. Example of fatigue area calculation.

In this example, the wheel impact location was $a = 0.305$ m, and the standard deviation of the simplified normal distribution was $\sigma = 0.063$ m. Therefore, the fatigue area for the crossing during this monitoring day was between 0.242 and 0.368 m, with a size of 0.126 m. It can be noticed that the calculated fatigue area is not accurate, yet for condition monitoring in the long term, such simplification can provide reasonably acceptable results and highly improve the efficiency of data analysis.

3.5. Results from Multiple Monitoring Days

In order to describe the development of the crossing's condition, the indicators are mainly used as statistical results over multiple monitoring days. An example of the development of vertical crossing acceleration responses as well as an irregular contact ratio is given for a span of 16 days in Figure 10. In this period, no track activities (e.g., maintenance) were performed, and the time frame was relatively too short for the condition of the crossing to degrade; therefore, the crossing condition can be assumed to be stable.

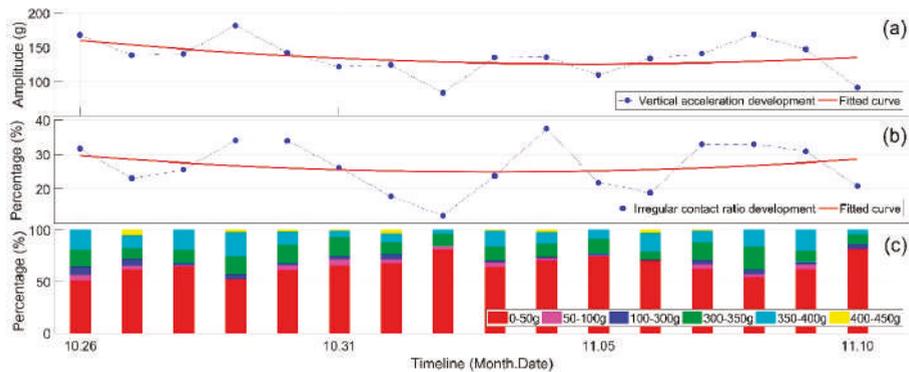


Figure 10. Development of the vertical acceleration responses in the monitored period. (a) Mean value of the vertical acceleration; (b) irregular contact ratio; (c) distribution of the acceleration responses for each day.

From Figure 10a it can be seen that the overall trend of the mean value of the accelerations is relatively stable, while the fluctuations of the responses are quite significant. The vertical accelerations have a minimum value of 84 g and a maximum value of 182 g. Such fluctuations resemble the fluctuations of the irregular contact ratio (Figure 10b). This resemblance will be further studied in the correlation analysis. It should be noted that the irregular contact ratio in the monitored period was above 10%, and for some days even it was higher than 30%, which is much higher than the previously studied 1:15 crossing [1] and reflects the abnormal condition of the monitored 1:9 crossing.

To summarize, the analyzed results have shown the following interesting features:

- a large difference in the dynamic responses from one passing wheel to another;
- a high irregular contact ratio due to multiple wheel passages during a single monitoring day; and
- highly fluctuating acceleration responses, as well as an irregular contact ratio during the short monitoring period.

All these features of the monitored 1:9 crossing indicate quite different performances from the previously studied 1:15 crossing. Investigating the sources of the fluctuation is necessary for a proper assessment of the crossing condition. Also, some condition indicators such as impact acceleration and the irregular contact ratio show possible correlations with each other. Figuring out the relationships between these indicators can help to reduce the amount of the required data, which will improve the efficiency of the post processing of the measurement results. These two questions can be investigated using correlation analysis, which will be presented in the next section.

4. Correlation Analysis

As discussed in the previous section, a high fluctuation was observed in the vertical acceleration responses to the monitored crossing over a short period of time, and was unlikely to be related to structural changes. Considering that the interference factors from the train were minimized by data selection, one possible cause of the fluctuating dynamic responses might have been the continuously changing weather conditions.

4.1. Influence of the Weather

It was discovered in the previous study [24] that temperature variation shows a good correlation with the acceleration fluctuation. In that study, the temperature fluctuation was considered to be the result of the duration of sunshine or precipitation. In order to assess the impact of the weather more accurately, the influences of weather conditions—including mean value of the daily temperature, daily

sunshine and precipitation duration—will be analyzed. Figure 11 shows the fluctuation of crossing vertical acceleration responses with varying weather conditions.

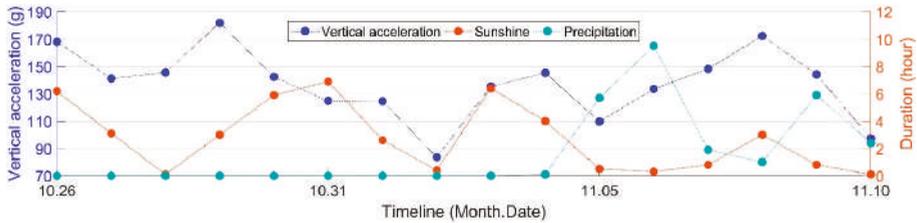


Figure 11. Development of vertical acceleration together with the durations of sunshine and precipitation.

From Figure 11 it can be seen that the fluctuating durations of sunshine showed a similar pattern to the crossing’s vertical acceleration responses. There seems to be connection between these two variables. For durations of precipitation, the connection with the vertical acceleration responses was lower. In order to quantitatively assess the impact of the weather, the correlations between the weather variables and condition indicators must be analyzed.

The weather data are obtained from the Royal Dutch Meteorological Institute (KNMI) [27] in days, and mainly consist of the following items:

- sunshine duration per day (D_s); and
- precipitation duration per day (D_p).

The crossing condition indicators were obtained from the crossing instrumentation, and the statistical results based on multiple monitoring days have been applied. The analyzed indicators include the following parts:

- longitudinal, vertical and lateral acceleration responses (a : a_x , a_y and a_z , respectively);
- an irregular contact ratio (I_r); and
- wheel impact location (L_o) and size of the fatigue area (F_a).

4.2. Pearson’s Correlation Coefficient

In statistics, the linear correlation between two variables is normally measured using Pearson’s correlation coefficient r . For two variables X and Y with the same sample size of n , r can be obtained using the following formula:

$$r_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X\sigma_Y} = \frac{1}{\sigma_X\sigma_Y} \cdot \frac{1}{n} \sum_{i=1}^n [(x_i - \mu_X)(y_i - \mu_Y)] \tag{1}$$

$$X = X(x_1, x_2, \dots, x_n), Y = Y(y_1, y_2, \dots, y_n) \tag{2}$$

where

- $\text{cov}(X, Y)$ is the covariance of X and Y ;
- σ_X & σ_Y are the standard deviations of X & Y , respectively;
- μ_X & μ_Y are the mean values of X & Y , respectively; and
- $E[\dots]$ is the expectation of the given variables

When X is in direct/inverse proportion to Y , then the correlation coefficient is

$$r_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X\sigma_Y} = \pm \frac{\sigma_X\sigma_Y}{\sigma_X\sigma_Y} = \pm 1 \tag{3}$$

If X and Y are independent, then the variable of $(x_i - \mu_X)(y_i - \mu_Y)$ (1) could be a random positive or negative value. In case of a large amount of data ($n \rightarrow \infty$),

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n [(x_i - \mu_X)(y_i - \mu_Y)] = 0 \tag{4}$$

Therefore, the value range of the correlation coefficient is $r_{X,Y} = [-1, 1]$. $r_{X,Y} = \pm 1$ means that the two variables X and Y are perfectly correlated, and $r_{X,Y} = 0$ means that X and Y have no correlation with each other. Otherwise, X and Y are considered partly correlated.

In different research fields, the gradation of the correlation index may have notable distinctions [28]. In some domains such as medicine and psychology, the requirement of the correlation coefficient—that a strong correlation is defined as $|r| \geq 0.7$ —is relatively strict, while in other domains such as politics, $|r| \geq 0.4$ can already be considered a strong correlation. In this study, the structural responses and weather were indirectly associated. The three-level guideline modified from [29] is applied for the correlation strength analysis, as shown in Table 1.

Table 1. The three-level correlation strength guideline.

r	Correlation Strength
$ r < 0.3$	Weak
$0.3 \leq r < 0.5$	Moderate
$0.5 \leq r < 1$	Strong

4.3. Correlation Analysis

In the analysis presented here, the correlations between the dynamic responses of the crossing (a , I_r , L_o and F_a) and the weather-related variables (T_m , D_s and D_p) are studied. The data used for the correlation analysis are from 16 monitoring days (the same as in Figure 10, $n = 16$ in Equation (2)). The correlation within each group of parameters, as well as the cross-correlation between these two groups of parameters, will be analyzed.

The results are presented in Table 2. Nomenclature in the table is presented earlier in Section 4.1. The strong, moderate and weak correlation coefficients are marked with red, blue and black colors, respectively. The correlation results will be analyzed in the different categories presented below.

Table 2. Correlation coefficients for dynamic responses and weather variables.

r	a_x	a_y	a_z	I_r	L_o	F_a	D_s	D_p
a_x	1	0.98	0.99	0.84	-0.30	-0.56	0.43	-0.23
a_y		1	0.99	0.79	-0.37	-0.51	0.36	-0.17
a_z			1	0.85	-0.32	-0.53	0.42	-0.22
I_r				1	-0.09	-0.42	0.40	-0.22
L_o					1	0.36	-0.39	0.14
F_a						1	-0.63	0.38
D_s							1	-0.54
D_p								1

4.3.1. Correlation of the Dynamic Responses

It can be seen from Table 2 it can be seen that the 3-D acceleration responses (a_x , a_y and a_z) are very strongly correlated to each other. The irregular contact ratio (I_r) and the size of fatigue area (F_a) are also strongly correlated with $a(a_x$, a_y and a_z). It can be noted that the correlations between F_a and a are negative, meaning that the increase of a is usually accompanied with the reduction of F_a .

The correlations of the impact location (L_o) with other dynamic responses are not strong, meaning that L_o is relatively independent from the other dynamic response. Some typical correlation results of the dynamic responses (framed in Table 2) are further discussed below.

The very strong correlations of a_x , a_y and a_z ($r \approx 1$) indicate that the 3-D accelerations are synchronously developed. The correlation between a_y and a_z is demonstrated in Figure 12a. Therefore in practice, it is possible to use the accelerations only in one direction (e.g., a_y) to analyze the crossing behavior, which can help improve the efficiency in processing the measurement data.

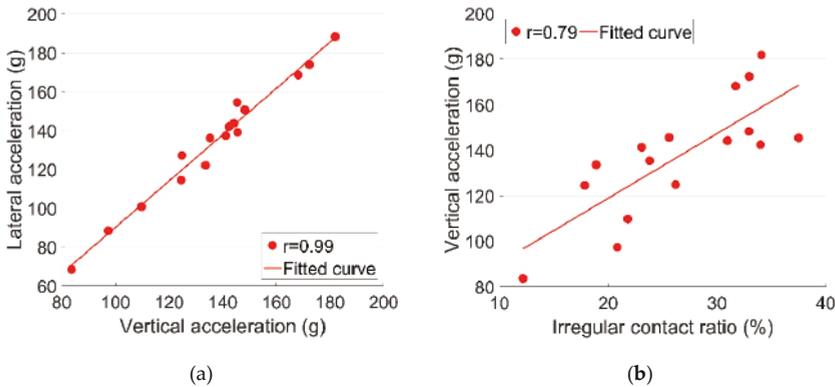


Figure 12. Correlations of the dynamic responses. (a) a_y - a_z ; (b) I_r - a_y .

The strong correlations between I_r and a (Figure 12b) clearly indicate that the high acceleration responses are to a great extent contributed by the high ratio of irregular contact. This phenomenon could have been caused by temporary (not residual) rail displacements due to varying temperature forces in the rail. It has to be noted that all these responses (I_r and a) fluctuated violently, a phenomenon that was likely caused by instable track conditions that were possibly affected by changes in weather conditions. This assumption will be verified later using a numerical model.

Figure 13 shows the correlation between a_y and L_o . The negative result means that when a increased, there was a tendency for L_o to be shifted closer to the crossing's theoretical point, although the moderate correlation strength ($r = -0.37$) indicates that the connection between a and L_o was rather limited. This might have been because the shift of L_o was a long-term effect of rail geometry degradation [1]. However, the rail geometry was unlikely to be changed during the relatively short monitoring period (16 days), so the temporary change of a might not have directly resulted in the shift of L_o .

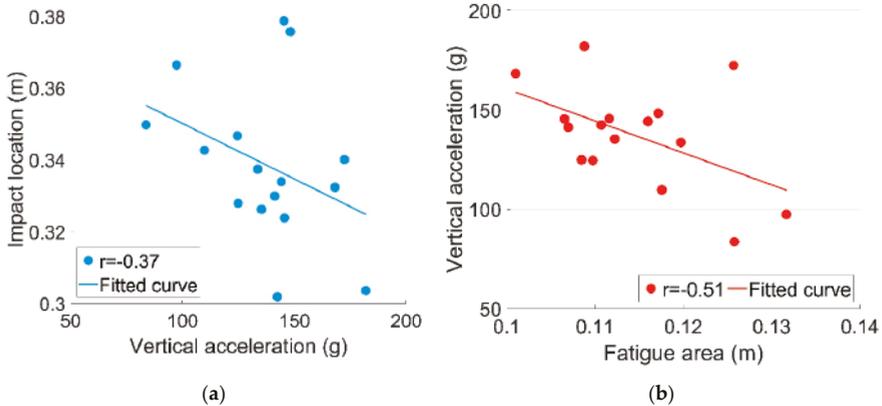


Figure 13. Correlations of the dynamic responses. (a) a_y-L_o ; (b) F_a-a_y .

The correlation between F_a and a_y is shown in Figure 13b. Compared with L_o , F_a was more likely to be reduced due to the increase of a . Combined with the strong correlation between a and I_r , it can be deduced that the impact locations of the irregular contact wheels tended to be centralized, while those of regular contact wheels were decentralized. Such a result confirms that a wider fatigue area will to some extent indicate a better crossing performance.

4.3.2. Correlation of the Weather Conditions

As can be seen from Table 2, the precipitation duration (D_p) had a strongly negative correlation with the sunshine duration (D_s), as shown in Figure 14.

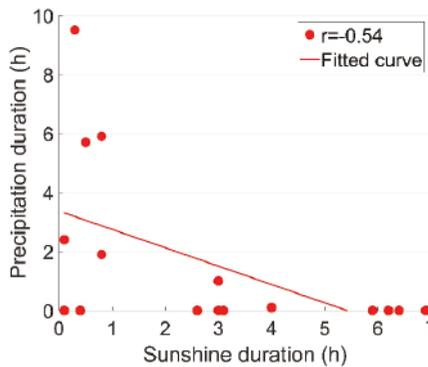


Figure 14. Correlation results between the sunshine and precipitation durations (D_s-D_p).

For the weather variables, D_s and D_p can be considered as two opposite weather conditions. From this point of view, the correlation coefficient of $r = -0.54$ is not very strong. Such results could be explained by the existence of cloudy/overcast conditions, and weather in a single day can switch among sun, rain and clouds/overcast. It can be noticed that in the monitored period, precipitation only occurred in 6 of the 16 days, which to some extent shows the complicity of the weather conditions.

4.3.3. Cross-Correlation between Dynamic Responses and Weather Conditions

According to the correlation results presented in Table 2, the cross-correlations of D_p with the dynamic responses are quite limited, except the moderate correlation with F_a . Meanwhile, D_s was strongly correlated with F_a and moderately correlated with all the other dynamic responses.

The moderate correlation between I_r and D_s is shown in Figure 15a. Such a result can be explained by the fact that an increase in rail temperature due to sunshine causes the displacements in the turnout. Due to these geometrical changes, the wheel cannot pass the crossing normally anymore and results in the increase of the irregular contact. Such a result is consistent with the moderate correlations between D_s and a .

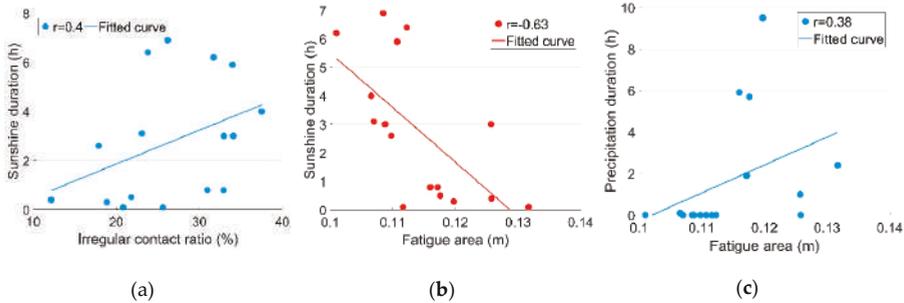


Figure 15. Cross-correlation results between the dynamic responses and weather conditions: (a) I_r - D_s ; (b) F_a - D_s ; (c) F_a - D_p .

The correlation of D_s with F_a was stronger than with the other dynamic responses ($r = -0.63$, Figure 15b), meaning that sunshine-initiated rail displacements were likely to occur primarily in centralized impact locations, which may have increased the likelihood of irregular contact.

An example for demonstrating the influence of sunshine on the dynamic responses of the monitored crossing is given in Figure 16. In this example, there was hardly any sunshine on one day (11.02), and a long period of sunshine on another day (11.03) (Figure 11). It can be seen that on 11.03 (with sunshine), I_r was higher (Figure 16a) and F_a was slightly narrower (Figure 16b). Such results indicate that the temporary effect of sunshine can lead to changes in the crossing performance.

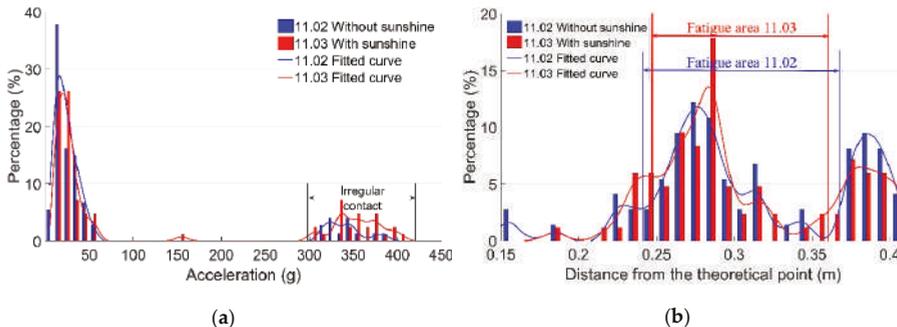


Figure 16. Influence of sunshine on the dynamic responses. (a) Vertical acceleration distribution; (b) fatigue area analysis.

The moderate correlation between D_p and F_a is shown in Figure 15c. Considering that the correlations between D_p and D_s were not very strong, the moderate correlation between the dynamic responses and weather conditions can already indicate a measure of impact. An example of the

measured dynamic responses of the crossing for a day without precipitation (11.04) and a day with precipitation (11.05, Figure 11) are shown in Figure 17.

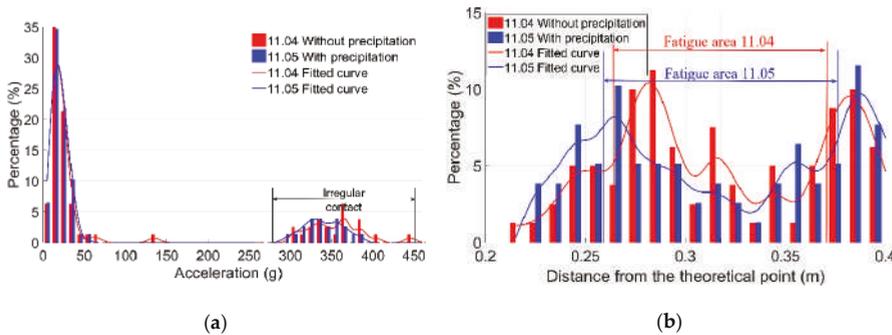


Figure 17. Influence of precipitation on the dynamic responses. (a) Vertical acceleration distribution; (b) fatigue area analysis.

It can be seen in Figure 17 that on the day with precipitation (11.05), I_r was slightly lower than on the day without precipitation (11.04), and F_a was wider. The reason for such results could be that precipitation may reduce the friction coefficient on the rail's surface and make the transition of the wheel load smoother. This assumption will be verified using a numerical model in the next section.

It should be mentioned that the subgrade of the monitored crossing was relatively soft, with canals on both sides of the track. Persistent precipitation could change the property of the subgrade and further affect the dynamic performance of the track. Therefore, the influence of precipitation can be quite complicated.

Based on the correlation analysis, the main conclusions can be drawn as follows:

- The accelerations in all three directions developed synchronously. In monitoring crossing conditions, it is sufficient to use vertical acceleration to represent the 3-D acceleration responses. Through this, the data processing procedure can be simplified.
- The strong correlation between I_r and a_y indicates that irregular contact is likely to result in high-impact accelerations. Such a result confirms that I_r can be used as an indicator for assessing crossing conditions. A high value of I_r indicates a degraded condition of the monitored crossing.
- The high (moderate/strong) correlation results between D_s and the dynamic responses of the crossing clearly indicate the influence of weather. It can be concluded that significant fluctuations in accelerations during a relatively short period are caused by changes in weather conditions. To verify this, a numerical model will be used in the next section.

5. Numerical Verification

In general, solar radiation is one of the major sources of rail thermal force. Depending on the sunshine duration, the associated rail temperature can rise to 40 °C higher than the ambient air temperature [30]. The change in rail temperature will increase the rail stress and amplify lateral displacements in the rail. The lateral displacements will then increase the uncertainty of the impact angle of a wheel in the railway crossing, eventually leading to an increase in the acceleration responses of some passing wheels, as shown in Figure 10.

Precipitation will introduce water to the rail surface that acts as a lubrication layer, which will reduce the friction coefficient in the wheel–rail interface [21]. It has been studied [31] that a low friction coefficient can be helpful in reducing hunting oscillation and, in contrast to sunshine, can reduce the impact angle of a wheel in the railway crossing.

The above-mentioned effects of temperature and friction variation corresponding to sunshine and precipitation are implemented in the multi-body system (MBS) model described below.

5.1. MBS Model Setup and Validation

In order to verify the weather effect hypotheses, a model for analyzing vehicle-crossing interaction developed according to the MBS method (implemented in VI-Rail software) will be used, as shown in Figure 18a. The track model is a straight line with a crossing panel (Figure 18b) situated in the middle of the track. The total length of the track model is 100 m, which allows enough preloading time for the vehicle before it enters into the crossing panel, as well as enough space after the vehicle passes through the crossing. The crossing geometry is defined by the control cross-sections, and the profiles between two pre-defined cross-sections are automatically interpolated using the third-order spline curve. In the track model, the rail is considered to be lumped masses on the sleepers connected with a massless beam. The flexible layers under the rail are the rail bushing that represents the rail pads and clips, and the base busing representing the ballast bed (Figure 18c).

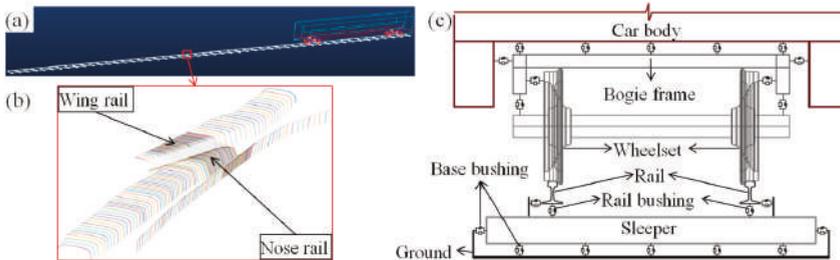


Figure 18. MBS model. (a) Vehicle-track model; (b) crossing profiles; (c) flexible connections in the model.

The crossing model is the same as the monitored 1:9 casted manganese crossing with a rail type of UIC54 E1. The track parameters of Dutch railways [32] applied in the model are shown in Table 3.

Table 3. Track parameters applied in the MBS model.

Track Components		Stiffness, MN/m	Damping, kN·s/m
Rail pad/Clips	Vertical	1300	45
	Lateral	280	580
	Roll	360	390
Ballast		45	32

The vehicle model was developed based on a VIRM locomotive with a total length of 27.5 m comprising a car body, front bogie and rear bogie. In the vehicle model, the car body and bogie frames, as well as the wheel sets, are modelled as rigid bodies with both primary and secondary suspensions taken into account (Figure 18c) [33]. The vehicle travels with a velocity of 140 km/h, the same as in the data analysis measurements. The wheels use a S1002 profile with a wheel load of 10 t. The wheel–rail contact model is defined as the general contact element and uses actual wheel and rail profiles as an input, which allows variable wheel and rail profiles.

The MBS vehicle–track model was validated using the measured acceleration responses from the crossing with the same design and stable conditions. Since the validation simulation was based on ideal track conditions, only the acceleration responses with regular wheel–rail contact were used in the comparison. The selected element for acceleration extraction was the rail with lumped mass (Figure 19a) from the same location as the instrumented accelerometer (Figure 2).

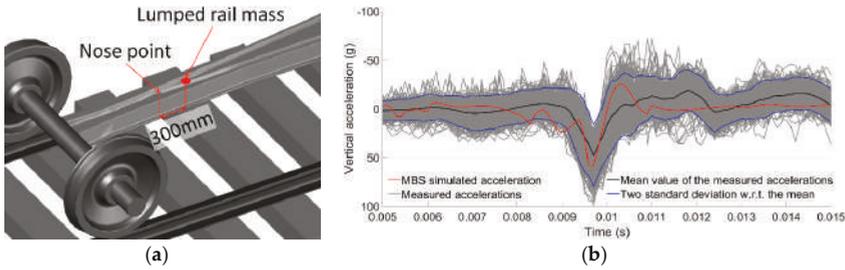


Figure 19. MBS model validation: (a) rail element for acceleration extraction; (b) comparison of MBS simulated acceleration with the measured ones in a time domain.

The validation results are shown in Figure 19b. It can be seen that the simulation results (red line) are quite comparable with the measured accelerations (black line). The magnitude of the simulated vertical acceleration during impact was around 55 g, which is comparable with the mean value of the measured acceleration responses (47 g). Although tolerable deviations of the impact signals exist, the simulation results agree reasonably with the measurements. It can be concluded that the MBS model can catch the main features of the wheel–rail impact during crossing and can be used to analyze crossing performance. Further details about the numerical model development and validation can be found in [34].

5.2. Numerical Analysis

5.2.1. Effect of Sunshine

In the previous study [35], the displacements of a turnout due to the change of the rail temperature were analyzed using a finite element (FE) model. The simulation results indicated that when the rail temperature was increased (from a stress-free temperature) by 40 °C, the turnout rails were laterally displaced up to 4 mm, as shown in Figure 20a. These results are applied in the MBS vehicle-crossing model as the sunshine-initiated lateral displacements. It should be noted that this simulation is based on ideal track conditions. In the case of a degraded track, the temperature-initiated lateral displacements could be amplified.

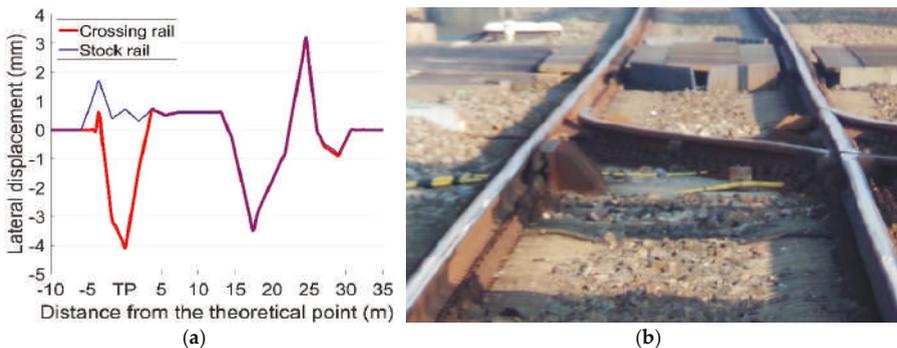


Figure 20. (a) Temperature-initiated rail lateral displacement in FE simulation (adapted from Figure 11.15 in [23]); (b) the monitored crossing.

In order to take the track degradation into account for the degraded track condition, the input lateral rail displacements in the MBS model are assumed to be twice as high as the ideal track condition (with maximum lateral rail displacements of 8 mm). The effect of precipitation is not taken into account

and the friction coefficient of $f = 0.4$ is used. Based on the above assumptions, the vertical accelerations and transition regions of the rail are simulated and presented below.

The calculated transition regions under different track conditions are shown in Figure 21. In the reference condition with no lateral displacement in the track, the sizes of the transition regions for the front wheel and the rear wheel are both 0.031 m [34]. When the temperature-initiated track displacements are taken into account, the transition regions shift closer to the theoretical point and the sizes reduce dramatically to 0.015 m for the front wheel and 0.012 m for the rear wheel. For the degraded track with higher rail displacements, the size of the transition region is only 0.004 m.

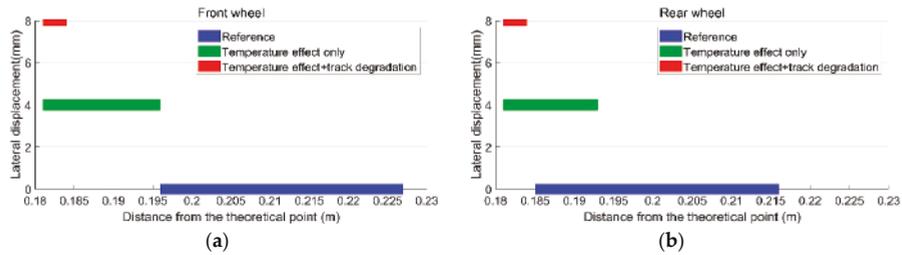


Figure 21. Transition regions of the crossing. (a) Front wheel; (b) rear wheel.

The vertical acceleration response of the rail due to passing wheels is shown in Figure 22. It can be seen that lateral displacement in the rail can result in higher acceleration responses caused by both the front and rear wheels. Combined with the results of the transition region (Figure 21), the simulation results confirm the correlation results (Figure 15a,b) that the long sunshine duration, which will result in a higher temperature in the rail, can lead to a centralized impact location and higher impact acceleration responses at the crossing.

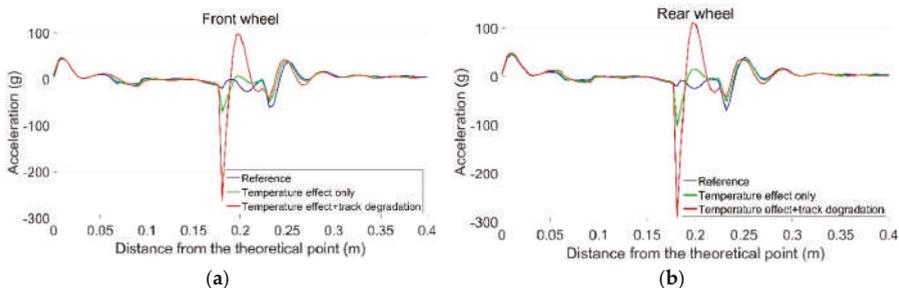


Figure 22. Vertical rail accelerations due to passing wheels. (a) Front wheel; (b) rear wheel.

It can be also seen that with the existence of rail displacement, the acceleration response caused by the rear wheel is higher than that caused by the front wheel from the same bogie. These results indicate that the performance of the rear wheel is not only affected by rail displacement, but also by the passing condition of the front wheel.

In case of a degraded track, higher rail displacements may lead to much higher acceleration responses as a result of both the front and rear wheels. Such impact accelerations (near 300 g) are close to the amplitude of the acceleration responses due to the irregular impacts in the measurements (Figures 16a and 17a). The simulation results prove that the lateral rail displacements caused by increases in rail temperature, in combination with track geometry deviations, can result in high wheel–rail impact accelerations.

5.2.2. Effect of Precipitation

With the influence of precipitation, the friction coefficient (f) in the wheel–rail interface can vary from 0.4 to 0.05 [35]. In this study, the precipitation effect is simulated by a reduction of f . The temperature-initiated rail displacements under ideal track conditions are taken into account. Calculations of rail accelerations resulting from passing wheels are shown in Figure 23.

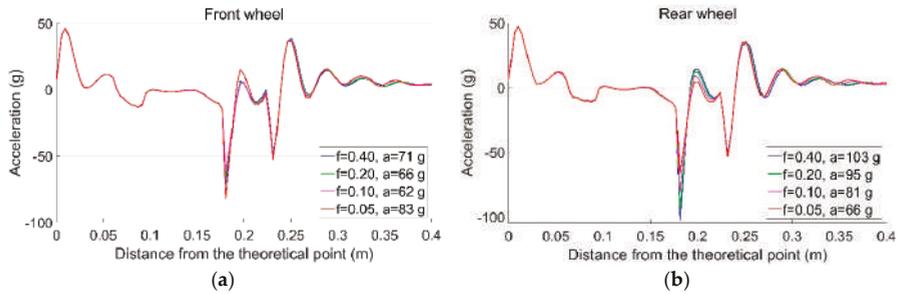


Figure 23. Vertical accelerations of the rail due to the passing wheels. (a) Front wheel; (b) rear wheel.

For the front wheel, when f is reduced from 0.4 to 0.1, the impact acceleration gradually reduces from 71 to 62 g. However, when $f = 0.05$, the maximum impact acceleration is increased to 83 g. Such results show that reducing the friction coefficient is not always helpful for the dynamic performance of the crossing. For the rear wheel, the reduction of f results in a decreased impact acceleration from 103 to 66 g. As discussed previously, the high rail acceleration responses due to the rear wheel are affected by the movement of the front wheel. In this case, the lowered f can help the wheelset return to a balanced position faster due to lower lateral restraint, which reduces the influence of the front wheelset on the rear wheelset from the same bogie.

It can be concluded that the change of f due to precipitation has an influence on the dynamic performance of the crossing, but the effect of a lower f is not always positive. Such results prove the correlation results indicating that an increase of D_p tends to result in lower acceleration responses, but the correlation strength is not high. The moderate correlation between D_p and F_a is also consistent with the simulation results that each wheel passes through the crossing more independently, which leads to less centralized impact locations.

5.3. Discussion

In this section, the MBS model for vehicle–crossing interaction analysis was briefly introduced. Using this model, the sunshine and precipitation effects were simulated as rail displacements and reduced f in the wheel–rail interface, respectively. The simulation results indicate that the rail displacements due to sunshine can lead to an increase in wheel–crossing impact acceleration. Combined with track degradation, such an effect could be highly amplified. Meanwhile, a lower f in the wheel–rail interface due to precipitation might reduce the interaction effect of two wheelsets from the same bogie, but cannot help improve track conditions. Combined with the measurement results, it can be concluded that the monitored crossing was not in the ideal condition, and possessed a certain degree of track degradation that made it more sensitive to changes in weather conditions.

6. Conclusions and Future Work

6.1. Conclusions

In this study, the conditions of a railway crossing were monitored, and the results were presented. The indicators for assessing the conditions of a crossing were briefly introduced. Inspired by the observed connection between vertical acceleration responses of the crossing and variations in the

sunshine duration, correlations of the dynamic responses and weather conditions were calculated. Using the vehicle-crossing MBS model, the influence of weather on the performance of the crossing was verified. The main conclusions of this study can be drawn as follows:

- The strong correlations between the dynamic responses show that the measurement results can be simplified and the crossing conditions can be assessed by only a few indicators (e.g., vertical acceleration, irregular contact ratio and fatigue area).
- The correlation results between the dynamic responses of the crossing and sunshine duration explain the fluctuation of dynamic responses over a short period of time. Such results confirm the temporary influence of weather on the performance of a crossing.
- The correlation results between sunshine duration and precipitation duration, as well as between precipitation duration and the dynamic responses of the crossing, indicate the complexity of the effect of precipitation.
- The simulation results not only verify the impact of weather on the dynamic performance of the crossing, but also indicate that the condition of the track at the monitored crossing was degraded. In cases of track degradation, the influence of weather can be amplified.

In monitoring the conditions of railway crossings, the correlation results among dynamic responses can be used to simplify measurement data. The verified weather effects explain the fluctuation of the dynamic responses over a short time period, which provides the basis for the measurement data regression. It should be noted that although sunshine variation is a short-term effect, the interaction of sunshine with the degraded track can turn this temporary interruption into a permanent track deformation, which will further accelerate the degradation of the track. In monitoring the conditions of railway crossings, the influence of weather can be eliminated through data regression to describe the structural degradation procedure, but the reflected track problem has to draw enough attention. Ensuring good track condition will not only help prolong service life of the crossing, but will also reduce the influence of varying weather conditions.

6.2. Future Work

This study was based on monitoring the conditions of railway crossings. It can be imagined that weather variation might also have an impact on other track sections, especially vulnerable parts such as transition zones, insulated joints, sharp curves, and so on. In the future, the effects of weather on other parts can be further investigated, which will improve the universality of this study and provide broader information for railway track management.

Author Contributions: This article is written by X.L. and supervised by V.L.M.

Funding: This research received no external funding.

Acknowledgments: The first author would like to thank the China Scholarship Council (CSC) for the financial support and Rolf Dollevoet for all the support as the head of Railway Engineering section in TU Delft. Also, the technical support from Ivan Shevtsov (ProRail) and help from Greg Lambert (RailOK, B,V) in the field measurement are highly appreciated.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, X.; Markine, V.L.; Wang, H.; Shevtsov, I.Y. Experimental tools for railway crossing condition monitoring. *Measurement* **2018**, *129*, 424–435. [[CrossRef](#)]
2. Barke, D.; Chiu, W.K. Structural health monitoring in the railway industry: A review. *Struct. Health. Monit.* **2005**, *4*, 81–94. [[CrossRef](#)]
3. Bowness, D.; Lock, A.C.; Powrie, W.; Priest, J.A.; Richards, D.J. Monitoring the dynamic displacements of railway track. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2007**, *221*, 13–22. [[CrossRef](#)]
4. Chudzikiewicz, A.; Bogacz, R.; Kostrzewskim, M.; Konowrockim, R. Condition monitoring of railway track systems by using acceleration signals on wheelset axle-boxes. *Transport* **2018**, *33*, 555–566. [[CrossRef](#)]

5. Sowiński, B. Interrelation between wavelengths of track geometry irregularities and rail vehicle dynamic properties. *Arch. Transp.* **2013**, *25*, 97–108.
6. Tsunashima, H. Condition Monitoring of Railway Tracks from Car-Body Vibration Using a Machine Learning Technique. *Appl. Sci.* **2019**, *9*, 2734. [[CrossRef](#)]
7. Wei, Z.; Núñez, A.; Li, Z.; Dollevoet, R. Evaluating degradation at railway crossings using axle box acceleration measurements. *Sensors* **2017**, *17*, 2236.
8. Wei, Z.; Shen, C.; Li, Z.; Dollevoet, R. Wheel–rail impact at crossings-relating dynamic frictional contact to degradation. *J. Comput. Nonlinear Dyn.* **2017**, *12*, 041016. [[CrossRef](#)]
9. Ma, Y.; Mashal, A.A.; Markine, V.L. Modelling and experimental validation of dynamic impact in 1:9 railway crossing panel. *Tribol. Int.* **2018**, *118*, 208–226. [[CrossRef](#)]
10. Xin, L.; Markine, V.L.; Shevtsov, I.Y. Numerical procedure for fatigue life prediction for railway turnout crossings using explicit finite element approach. *Wear* **2016**, *366*, 167–179. [[CrossRef](#)]
11. Kassa, E.; Andersson, C.; Nielsen, J.C.O. Simulation of dynamic interaction between train and railway turnout. *Veh. Syst. Dyn.* **2006**, *44*, 247–258. [[CrossRef](#)]
12. Wan, C.; Markine, V.L.; Shevtsov, I.Y. Optimisation of the elastic track properties of turnout crossings. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2016**, *230*, 360–373. [[CrossRef](#)]
13. Pålsson, B.A. Optimisation of railway crossing geometry considering a representative set of wheel profiles. *Veh. Syst. Dyn.* **2015**, *53*, 274–301. [[CrossRef](#)]
14. Wan, C.; Markine, V.L.; Shevtsov, I.Y. Improvement of vehicle–turnout interaction by optimizing the shape of crossing nose. *Veh. Syst. Dyn.* **2014**, *52*, 1517–1540. [[CrossRef](#)]
15. Wan, C.; Markine, V.L.; Dollevoet, R.P.B.J. Robust optimisation of railway crossing geometry. *Veh. Syst. Dyn.* **2016**, *54*, 617–637. [[CrossRef](#)]
16. Kassa, E.; Nielsen, J.C.O. Dynamic interaction between train and railway turnout: Full-scale field test and validation of simulation models. *Veh. Syst. Dyn.* **2008**, *46*, 521–534. [[CrossRef](#)]
17. Markine, V.L.; Shevtsov, I.Y. An experimental study on crossing nose damage of railway turnouts in the Netherlands. In Proceedings of the 14th International Conference on Civil, Structural and Environmental Engineering Computing, Cagliari, Italy, 3–6 September 2013.
18. Xin, L.; Markine, V.L.; Shevtsov, I.Y. Numerical analysis of the dynamic interaction between wheelset and turnout crossing using explicit finite element method. *Veh. Syst. Dyn.* **2016**, *54*, 301–327. [[CrossRef](#)]
19. Wan, C.; Markine, V.L. Parametric study of wheel transitions in railway crossings. *Veh. Syst. Dyn.* **2015**, *53*, 1876–1901. [[CrossRef](#)]
20. Arias-Cuevas, O. Low Adhesion in the Wheel-Rail Contact. Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, 2010.
21. Chen, H.; Ban, T.; Ishida, M.; Nakahara, T. Effect of water temperature on the adhesion between rail and wheel. *Proc. Inst. Mech. Eng. Part J J. Eng. Tribol.* **2006**, *220*, 571–579. [[CrossRef](#)]
22. Ertz, M.; Bucher, F. Improved creep force model for wheel/rail contact considering roughness and temperature. *Veh. Syst. Dyn.* **2002**, *37*, 314–325. [[CrossRef](#)]
23. Arts, T.M.H. Measuring the Neutral Temperature in Railway Track During Installation and Use. Master’s Thesis, Delft University of Technology, Delft, The Netherlands, 2011.
24. Markine, V.L.; Liu, X.; Mashal, A.A.; Ma, Y. Analysis and improvement of railway crossing performance using numerical and experimental approach: Application to 1:9 double crossovers. In Proceedings of the 25th International Symposium on Dynamics of Vehicles on Roads and Tracks, Queensland, Australia, 14–18 August 2017.
25. Wegdam, J.J.J. Expert Tool for Numerical & Experimental Assessment of Turnout Crossing Geometry. Master’s Thesis, Delft University of Technology, Delft, The Netherlands, December 2018.
26. Markine, V.L.; Shevtsov, I.Y. Experimental Analysis of the Dynamic Behaviour of Railway Turnouts. In Proceedings of the 11th International Conference on Computational Structures Technology, Dubrovnik, Croatia, 4–7 September 2012.
27. Koninklijk Nederlands Meteorologisch Instituut (KNMI, Royal Dutch Meteorological Institute). 2015. Available online: <https://www.knmi.nl/nederland-nu/klimatologie/daggegevens> (accessed on 10 November 2018).
28. Akoglu, H. User’s guide to correlation coefficients. *Turk. J. Emerg. Med.* **2018**, *18*, 91–93. [[CrossRef](#)] [[PubMed](#)]
29. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; New York University: New York, NY, USA, 1988.

30. ProRail. *Instruction of Procedures at Extreme Weather Conditions RLN00165, Version 002*; ProRail: Utrecht, The Netherlands, 2003.
31. Ohyama, T. Tribological studies on adhesion phenomena between wheel and rail at high speeds. *Wear* **1991**, *144*, 263–275. [[CrossRef](#)]
32. Hiensch, M.; Nielsen, J.C.O.; Verheijen, E. Rail corrugation in The Netherlands—Measurements and simulations. *Wear* **2002**, *253*, 140–149. [[CrossRef](#)]
33. VI-grade GmbH. *VI-Rail 16.0 Documentation*; © 2014 VI-Grade Engineering Software & Services: Marburg, Germany, 2014.
34. Liu, X.; Markine, V.L. Validation and Verification of the MBS Models for the Dynamic Performance Study of Railway Crossings. *Eng. Struct.* (under review).
35. Arts, T.M.H.; Markine, V.L.; Shevtsov, I.Y. Modelling turnout behaviour when achieving a neutral temperature. In *Proceedings of the First International Conference on Railway Technology: Research, Development and Maintenance*, Las Palmas de Gran Canaria, Spain, 18–20 April 2012.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

A Novel Monitoring Approach for Train Tracking and Incursion Detection in Underground Structures Based on Ultra-Weak FBG Sensing Array

Qiuming Nan ¹, Sheng Li ^{1,*}, Yiqiang Yao ², Zhengying Li ², Honghai Wang ¹, Lixing Wang ¹ and Lizhi Sun ³

¹ National Engineering Laboratory for Fiber Optic Sensing Technology, Wuhan University of Technology, Wuhan 430070, China; nqm0723@whut.edu.cn (Q.N.); wanghh@whut.edu.cn (H.W.); lxwang@whut.edu.cn (L.W.)

² School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China; yqyao@whut.edu.cn (Y.Y.); zhyli@whut.edu.cn (Z.L.)

³ Department of Civil and Environmental Engineering, University of California, Irvine, CA 92697-2175, USA; lsun@uci.edu

* Correspondence: lisheng@whut.edu.cn

Received: 23 May 2019; Accepted: 12 June 2019; Published: 13 June 2019

Abstract: Tracking operating trains and identifying illegal intruders are two important and critical issues in subway safety management. One challenge is to find a reliable methodology that would enable these two needs to be addressed with high sensitivity and spatial resolution over a long-distance range. This paper proposes a novel monitoring approach based on distributed vibration, which is suitable for both train tracking and incursion detection. For an actual subway system, ultra-weak fiber Bragg grating (FBG) sensing technology was applied to collect the distributed vibration responses from moving trains and intruders. The monitoring data from the subway operation stage were directly utilized to evaluate the feasibility of the proposed method for tracking trains. Moreover, a field simulation experiment was performed to validate the possibility of detecting human intrusion. The results showed that the diagonal signal pattern in the distributed vibration response can be used to reveal the location and speed of the moving loads (e.g., train and intruders). Other train parameters, such as length and the number of compartments, can also be obtained from the vibration responses through cross-correlation and envelope processing. Experimental results in the time and frequency domains within the selected intrusion range indicated that the proposed method can distinguish designed intrusion cases in terms of strength and mode.

Keywords: underground structure safety; train tracking; incursion detection; ultra-weak FBG; distributed vibration; dynamic measurement

1. Introduction

During the last few decades, the construction of urban subways has developed rapidly worldwide and particularly in China. Aiming to ensure the operational safety of subways, a wide range of research effort has been undertaken in the fields of the subway fires [1–3], structural safety [4–7], and perimeter invasion [8–11]. Among these three fields, structural safety monitoring and perimeter intrusion detection are of more concern than fire monitoring due to the diversification of demand. For long-distance monitoring needs, especially for subway tunnels, distributed fiber-optic sensing technology has been widely recognized as the most promising means of addressing complex needs due to its advantages of large-scale monitoring, high sensitivity, and multiplexing capabilities [12]. For instance, the safety monitoring of subway structures based on Brillouin optical time domain

reflectometry (BOTDR) technology [13] has been reported [14,15]. In addition to BOTDR-based static measurement, distributed fiber-optic sensors for dynamic measurement [16], especially distributed acoustic sensing (DAS) technology [17,18] have been another research hotspot. The use of DAS technology to ensure the safety of subway operations has also attracted widespread attention for both engineers and researchers.

As one of the main concerns for ensuring the operational safety of the subway, tracking the subway trains occupies an important position in the train operation control system; tracking is directly related to train safety and affects the transportation efficiency of the rail transit. In addition to tracking trains in operation, positioning illegal intruders and preventing the risk caused by such intrusion events—which usually occur during the subway outage periods—is another issue worth noting. For the former, Peng et al. [19] reviewed the shortcomings of conventional train positioning techniques and pioneered investigation of the feasibility of train positioning and speed monitoring through Φ -OTDR technology, in which the spatial resolution of the common optic fiber reaches 20 m. He et al. [20] reported a DAS-based method for condition monitoring of the running train, claiming that the train positioning error was less than 20 m. For human intrusion, Catalano et al. [21,22] reported an incursion detection system for railway security using two types of fiber Bragg grating (FBG) sensors, which is apparently only applicable to a limited protection area due to the restricted multiplexing capacity of FBG. In addition, He et al. [23] presented research on railway perimeter safety based on DAS technology, which has a spatial resolution of only 10 m in the reported application scenarios.

Obviously, the dynamic measurement techniques based on DAS provide feasible detection methods for train tracking and human intrusion. Yet, there are still few research efforts on integrated methods for both train tracking and intrusion detection. Compared with DAS technology using common optic fiber, ultra-weak FBG arrays based on the draw tower [24,25] using sensing optic fiber integrates both advantages of fiber optic point sensors and distributed sensors. This technology is a new way to achieve high-precision, fast and wide coverage distributed measurement. Previous research around this technology has focused more on monitoring strain, temperature or strain-based deformation for the object of interest [26–28]. In addition, a multi-parameter measurement system based on an ultra-weak FBG array with sensitive material was proposed in [29]. However, all this research is still limited to static indicators. In fact, ultra-weak FBG array is also adept at performing dynamic monitoring [30,31] in addition to the above positive characteristics usually witnessed in static measurement. The reports in [16,32–34] revealed that the ultra-weak FBG array can not only be used for both static and dynamic measurements, but also has a higher signal-to-noise ratio (SNR) than that of DAS sensors. It is known that higher SNR often leads to better sensing performances such as higher measurement accuracy, faster response time and simpler detection circuit. Therefore, the ultra-weak FBG array is more suitable than DAS when dealing with distributed vibration and other scenarios requiring high-speed measurement.

To eliminate the need for two separate systems, improve measurement efficiency and reduce overall cost, this paper explored the feasibility of addressing train tracking and human intrusion in subway systems using distributed vibration measurement based on the ultra-weak FBG sensing array. The experimental results of identifying running trains and intruders in an actual subway are reported. The sensing and monitoring principles make up the second part of this paper, followed by the details of the design and field arrangement used to validate the proposed method. Finally, the effectiveness on tracking and detecting the objectives of interest is discussed based on the experimental results represented by the responses of distributed vibration of the ultra-weak FBG array.

2. Sensing and Monitoring Principles

Figure 1 illustrates the distributed vibration sensing principle used to detect distributed vibration generated by moving loads, such as trains, intruders and so on. The phenomenon of light interference caused by the reflection signals of adjacent two ultra-weak FBGs is used to detect the vibration of the object of interest. Here, the ultra-weak FBG [35] is regarded as a mirror, and L represents the

distance that causes light interference. In order to ensure a stable optical interference effect and overcome the occurrence of optical interference failure due to the difficulty of matching adjacent ultra-weak FBG caused by, for example, temperature variation drift, ultra-weak FBG in the array uses 3 nm wideband FBG. In addition, since the temperature changes slowly with respect to vibration, the temperature influence is ignored in the demodulation process of the vibration. The spatial resolution of the distributed vibration along the sensing optic fiber is typically determined by the parameter L . The sensitivity and the frequency response of the vibration signal measured by the strain-induced phase variation between two ultra-weak FBGs are improved by the interferometer. Here, Faraday rotating mirrors are utilized in the demodulation process of the ultra-weak FBG array to suppress the polarization effect. Moreover, the 3-by-3 coupler phase demodulation algorithm is used to reconstruct the time domain signal, and restore the phase information of the vibration signal, through which the interrogation of the vibration frequency and amplitude can be realized. Further, the optical time domain reflectometry technique is utilized to achieve vibration localization, and therefore, increasing the length of the optical cable will prolong the sampling interval of the vibration signal and reduce the response bandwidth of the system.

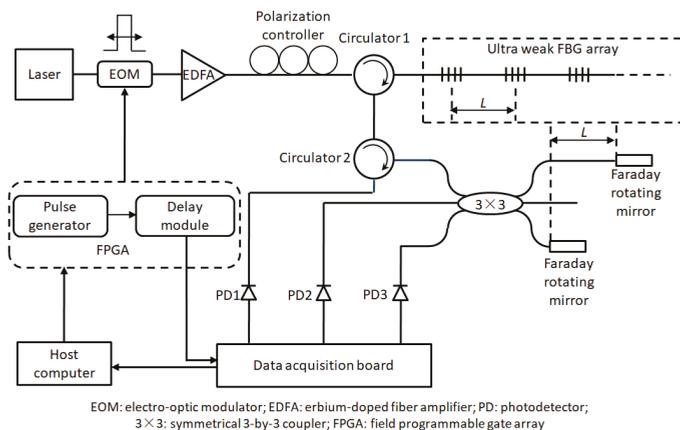


Figure 1. Sensing principle of distributed vibration detection-based on ultra-weak FBG array.

The high sensitivity of large-scale ultra-weak FBGs and the corresponding demodulation system of high speed [36] make the sensing optic fiber particularly suitable for locating structural vibration excited by moving loads occurring within a long-distance range. In addition, the previous study [37] revealed the repeatability of such a sensor represented by strain is around 3.41 nano epsilon. When dealing with train tracking and intrusion detection, either the train or intruder movement can be regarded as a vibration source. Owing to such excitation, the surface waves propagate omni-directionally on the ground. Because the surface wave couples to the track bed and rail track, distributed sensing optic fiber mounted beside the rail track along the subway can detect the vibration generated by a passing train or human footsteps (see Figure 2). The light interference region indicated by the address of ultra-weak FBG can be interrogated with the time- and wavelength-division multiplexing method [38,39], causing each known light interference region along the sensing optic fiber to have a determinate correspondence with the mileage. This also indicates that querying the interference region generated by the distributed vibration excitation is a viable way to track or detect the moving load of interest.

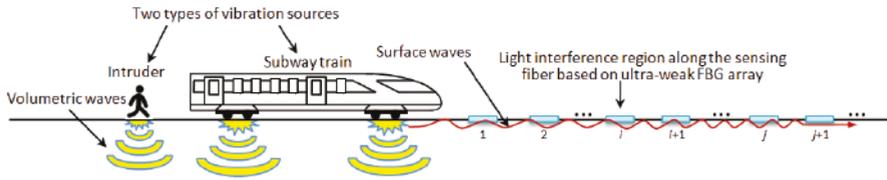


Figure 2. Monitoring principle of capturing the two types of moving loads of interest based on distributed vibration.

Moreover, the speed of the train or intruder can be determined through the τ lag time described in cross-correlation Equation (1) and the known distance between regions i and j as depicted in Figure 2.

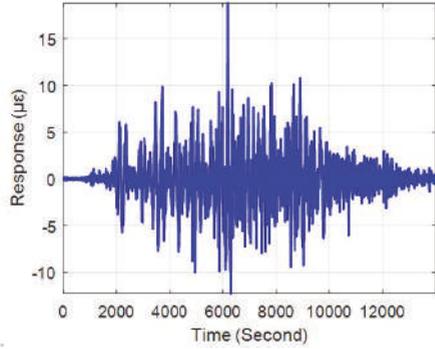
$$R_{S_i S_j}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T S_i(t) S_j(t + \tau) \quad (1)$$

where $S_i(t)$ and $S_j(t)$ represent the vibration response at light interference regions i and j , respectively. The lag time τ is equal to the duration from the regions i to j . Further, through draw-tower grating preparation with five-meter equidistance between adjoining FBGs along the sensing optic fiber, the spatial resolution of the sensing optic fiber discussed in this paper enabled better positioning accuracy of tracking the train and intruder than that of the above-mentioned reports in actual engineering practice.

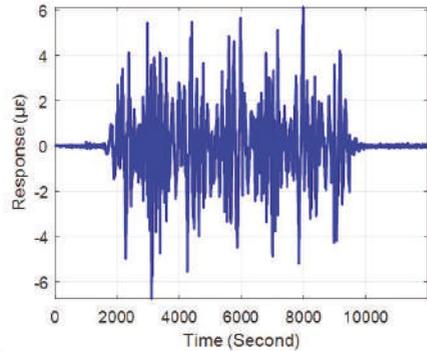
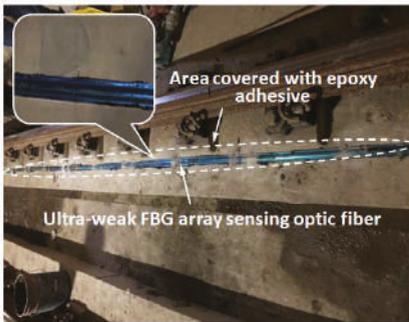
3. Design and Field Arrangement for the Experiments

3.1. Engineering Background of the Trial

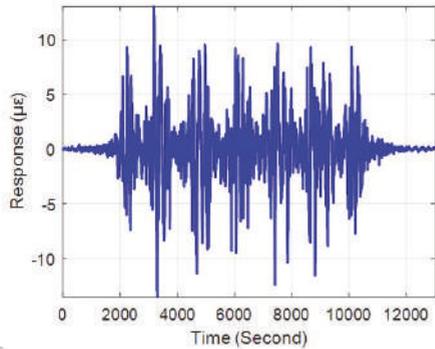
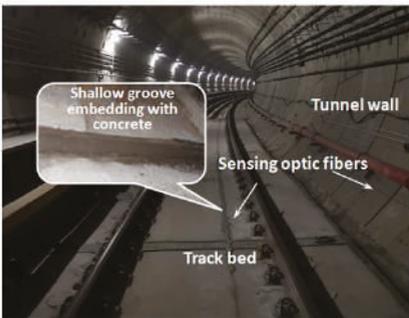
An actual tunnel structure (Wuhan Metro Line 7) was used in this study. Before the operation of the subway, the ultra-weak FBG sensing optic fiber with armored protection using a layer-stranding structure with a loose tube was installed on structure surfaces of the selected tunnel segments, as in the actual layout shown in Figure 3. Here, the research on identifying the two types of moving loads was primarily based on the track bed response. To better obtain the vibration response of the track bed, three methods for fixing the sensing optic fiber to the track bed were tried to evaluate the suppression effect of the disturbance vibration—namely, fixture fixing, epoxy adhesive and shallow groove embedding. The typical vibration responses of a monitoring zone induced by passing trains in each fixing method are shown on the right side of Figure 3. It can be seen that as the coupling constraint between the sensing array and the track bed increased, the amplitude symmetry of the vibration response improved, and the peak regularity of the vibration response associated with train excitation became clear. Therefore, shallow groove embedding was adopted to affix the sensing array to secure a better signal output.



(a)



(b)



(c)

Figure 3. Methods of affixing distributed vibration sensing optic fibers and typical vibration response induced by train: (a) fixture fixing, (b) epoxy adhesive, and (c) shallow groove embedding.

The designed monitoring system can guarantee five kilometers array length and reach multiplexing capacity of 1000 ultra-weak FBGs. As shown in the schematic diagram in Figure 4, the experimental area covered three underground stations with a total length of nearly three kilometers. Due to the spatial resolution of the sensing optic fiber and the specific layout of the tunnel structure, more than

500 vibration regions labeled #1 to #515 along the track bed can be distinguished based on the interrogated address of the light interference. It can be seen from the right side of Figure 4 that in addition to the common track bed structure, the damping track bed was also included in the experimental area. During the trial, the real-time vibration responses with a 1 kHz sampling rate were fully transmitted back to the platform monitoring center and processed by the demodulator and servers. Since the ultra-weak FBG array was fabricated simultaneously in the optic fiber drawing, there was no additional splice in the sensing optic fiber equipped with armored protection, except for the pigtail that needed to be connected to the demodulation instrument.

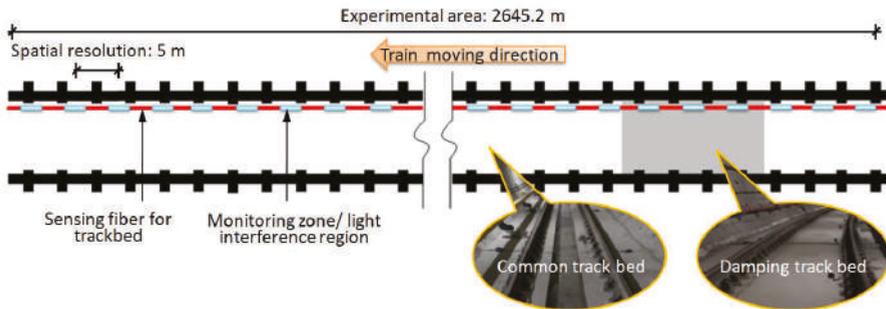


Figure 4. Schematic diagram of experimental area covering different track bed structures.

3.2. Train Tracking Trial

Because the subway line has already been in operation, the distributed vibration responses of the experimental area caused by the train were automatically collected and directly taken as the raw data for the trial. In addition to observing the response caused by the train traveling in the subway tunnel monitored by the sensing optic fiber, the test discussed the identifiability of the sensing optic fiber to the train moving in the opposite direction in the adjacent tunnel. Based on the single point response and overall distribution characteristics, the detection capabilities of the following indicators were discussed in turn: the speed and position of the train, the response difference between the common and damping tracks, and other parameters of the train.

3.3. Intruder Detection Trial

To ensure the safe operation of the subway in the following day, various manual inspections are usually carried out at the subway outage in the early hours of the morning. We conducted the incursion test at this inspection window; this is also the period in which illegal intrusion usually occurs. Based on the specific circumstances and various scheduled tasks, a range in the area of the damping track bed was approved for performing multiple sets of simulated intrusion tests. To minimize cross interference from other simultaneous inspections, the trials were primarily concentrated within a 130 m range of the selected tunnel area. As shown in Figure 5, the trials simulated single and multi-person intrusion scenarios and considered the intrusion patterns of walking and jogging. In each trial, the participant in the simulation test made a round trip within the designed intrusion area.

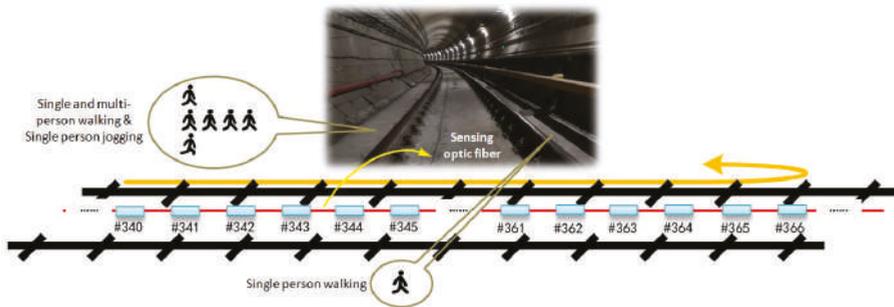


Figure 5. Simulated human intrusion scenarios along the rail track in the selected tunnel area.

4. Results and Discussion

This section reports the characteristics of distributed vibration responses under operating train and simulated incursion conditions, respectively. Feasibility, based on the proposed comprehensive approach concerning train tracking and detecting incursion, was investigated and is discussed. All the following analyses were based on the original output of the distributed vibration responses with no additional techniques adopted to improve the data quality.

4.1. Distributed Vibration Responses under Load of Passing Train

Figure 6a depicts a typical visualization relationship between the structural vibration intensity and the space and time. Here, the vibration intensity was represented by color of the figure. A waterfall diagram such as Figure 6a can be used to help analyze the train's running direction, speed change and arrival or departure interval. In the tunnel where the sensing optic fiber arrays were deployed, the train entered the experimental area from #500 monitoring zone in Figure 6a. In this case, a moving train appeared as a diagonal signal pattern where the slope depended on the speed. Here, the diagonal signal pattern highlighted the characteristics of the distributed vibration response caused by the passing train within the monitoring range. Three complete sets of such diagonal signal patterns can be clearly seen in the left part of Figure 6a. Moreover, vibrations generated by moving trains in the opposite direction in the adjacent tunnel were simultaneously detected by the sensing optic fiber, although the vibration intensity was somewhat weak. Further, the process of the train stopping at the station between the diagonal signal patterns can also be observed in the figure. For monitoring zones #250–#500, the region range of the damping track bed structure, can be clearly identified based on the height changes (along the time axis) of the diagonal signal pattern. Due to the large design distance between the tunnel where the experimental areas #250–#500 were located and the adjacent tunnel, the vibration transmitted from the adjacent tunnel becomes invisible in the right part of Figure 6a.

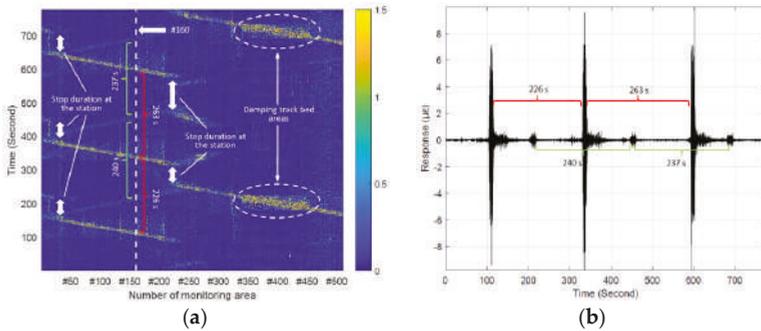


Figure 6. (a) Vibration intensity versus space and time under operating train; (b) original time series of vibrations of monitoring zone #160.

Figure 6b presents a complete vibration response of one monitoring zone in Figure 6a, which quantifies the difference in vibration intensity of the sensing optic fiber caused by the train moving in two adjacent tunnels. In addition, the time interval of the two adjacent trains was approximately four minutes as observed in Figure 6, which was consistent with the planned subway operating timetable. Moreover, compared with the report based on Φ -OTDR [18], the method using sensing optic fiber in this paper did not require the multiple averaging technique to improve the SNR of the original time series of vibrations, which was more efficient for providing location information of the detected object. Therefore, the responses of any two different monitoring zones could be used to determine the train speed. For instance, Figure 7a shows the intensity projections of the two measurement areas on the time axis of Figure 6a, the cross-correlation analysis (see Figure 7b) of the vibration sequences (see Figure 7a) of the two monitoring zones at 650 m apart indicated that the train took 37.51 s to pass through the two selected zones. In this way, the train speed of 62 km/h can be obtained. Moreover, it was found that the amplitudes of these two monitoring areas were different, although the sensing optic fiber and its fixation method were consistent. The reason for this was primarily due to uneven geological properties of the underground structure along the mileage direction of the tunnel and different design curvature along the tunnel line, and the structural stiffness of the shield segments.

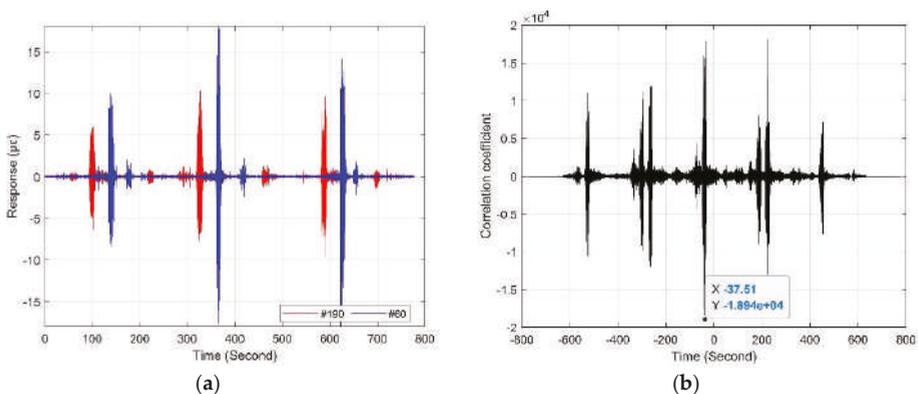


Figure 7. (a) Original time series of vibrations at zones #60 and #190; (b) time lag between zones #60 and #190.

In addition, the vibration response obtained by a particular monitoring area during the passage of the train can reflect some geometric parameters of the train, such as its length and the number of

compartments. The former, length, can be estimated by the calculated speed and the known height changes of the diagonal signal pattern. The latter, number of compartments, can be revealed by the number of peaks or valleys of the envelope signal. Figure 8 shows a typical vibration response of a monitoring zone between #60 and #190 during train passage. The vibration response lasted for about 8.5 s, corresponding to the height change of the diagonal pattern shown in Figure 6a. Based on the obtained average speed of 62 km/h, the calculated train length of 146 m was close to the actual known 142 m. Also, seven envelope peaks and valleys can be recognized in Figure 8 by envelope processing. This envelope result agreed well with the axle impact of the six train compartments.

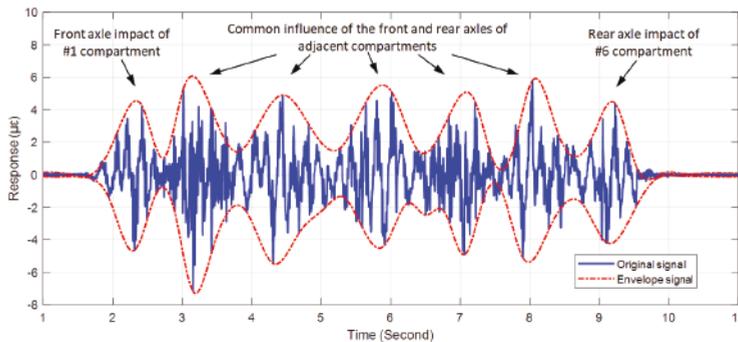


Figure 8. Original time series of vibration and corresponding envelopes during the passage of the train through one monitoring zone.

4.2. Distributed Vibration Response under Footsteps of Intruder

Figures 9 and 10 show the results of the designed human intrusion in the time and frequency domains, respectively. Moreover, the experimental results in the frequency domain for each of the designed cases are depicted two-dimensionally (left) and three-dimensionally (right) in Figure 10. Figure 9 reveals that significant distributed vibration responses generated by walking or jogging as defined in Figure 5 can be detected within the incursion range under both sides of the track. In addition, two diagonal signal patterns in the opposite direction further verified the simulated incursion process represented by round-trip walking or jogging. Furthermore, based on the different slope pattern caused by different speeds of the intruder, it was easy to distinguish the intrusion mode of jogging shown in Figure 9c from the other three intrusion modes of walking. This result was consistent with Figure 10 and was more pronounced in the frequency domain, where the intrusion caused by the jogging shown in Figure 10c led to the maximum fluctuation of the vibration intensity.

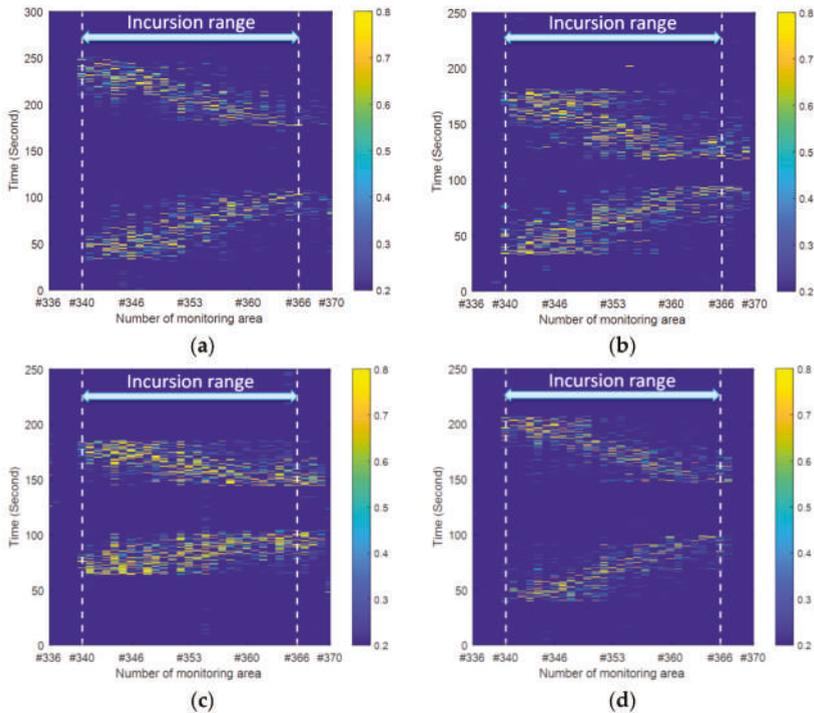


Figure 9. Vibration intensity versus space and time under different intrusion scenarios: (a) single person walking; (b) four people walking; (c) single person jogging; (d) single person walking along the other side.

To further quantify the different intrusion patterns reflected in Figure 9, the effective value represented by the root-mean-square (RMS) of the vibration response signal for each monitoring zone within the intrusion range in the whole test process was calculated and is shown in Figure 11. Here, the effects caused by personnel in the round-trip process outside the intrusion range were not involved in the evaluation. As can be seen from the overall distribution of Figure 11, in addition to the significant difference between jogging and walking intrusion, the dynamic distributed vibration response can distinguish between single and multi-person walking intrusions. Moreover, subtle differences of vibration distribution caused by a single pedestrian intrusion at different distances from the sensing optic fiber can also be observed. Furthermore, Figure 12 quantifies the results represented in Figure 10 by the overall distribution of primary frequency. Here, the frequency value corresponding to the maximum energy of each column represented in Figure 10 was selected as the primary frequency for each monitoring zone. Table 1 further provides the statistical results for the four types of intrusion cases for Figure 12, where cases 1–4 represent a single person walking, four people walking, single person jogging and single person walking along the other side, respectively.

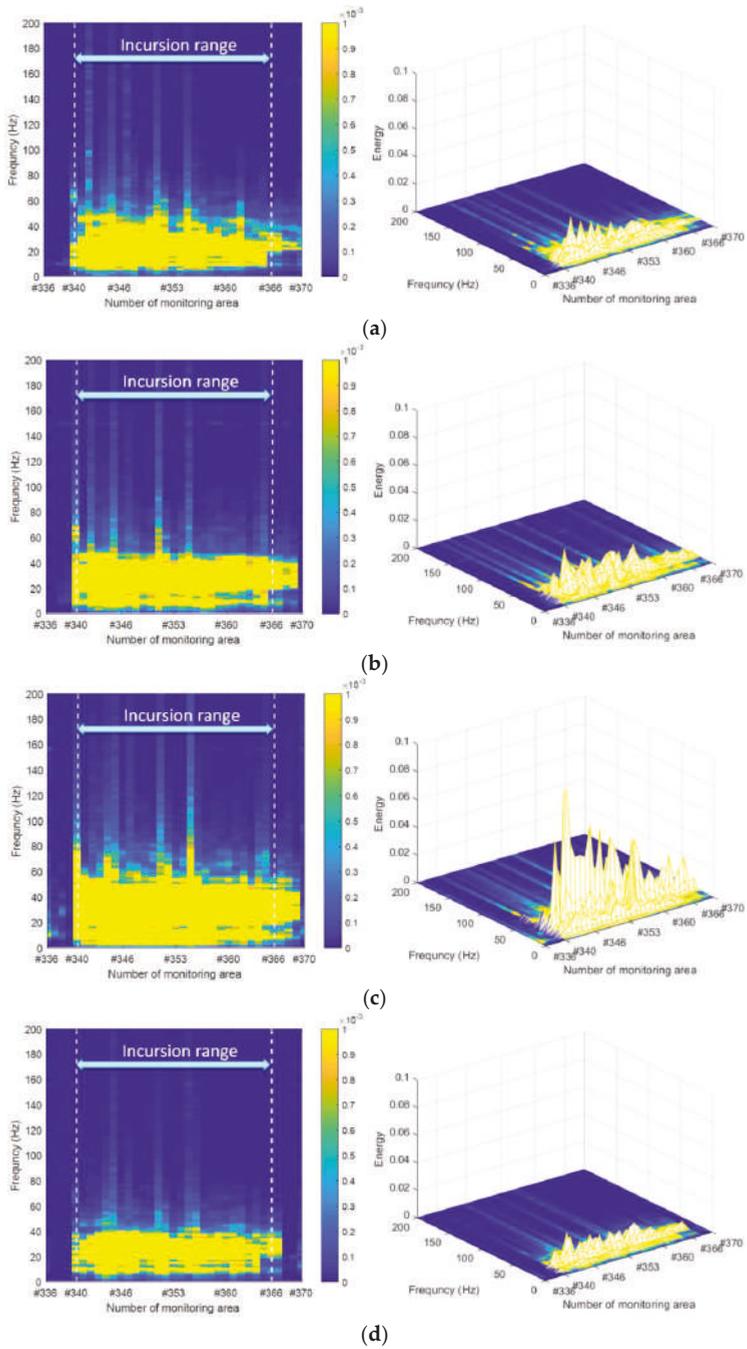


Figure 10. Vibration intensity versus space and frequency under different intrusion scenarios: (a) single person walking; (b) four people walking; (c) single person jogging; (d) single person walking along the other side.

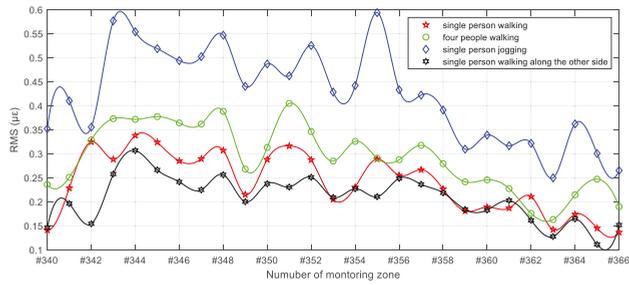


Figure 11. Fitting distribution of effective values of vibration response of incursion range under different simulated intrusion cases.

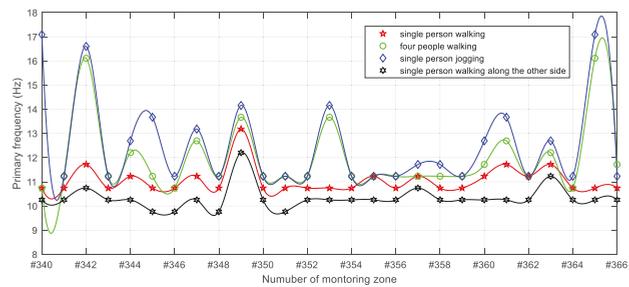


Figure 12. Fitting distribution of primary frequencies of vibration response of incursion range under different simulated intrusion cases.

Table 1. Primary frequency characteristics of the different intrusion cases within the experimental area (unit: Hz).

Comparisons	Case 1	Case 2	Case 3	Case 4
Maximum	13.18	16.11	17.09	12.21
Mode ¹	10.74	11.23	11.23	10.25

¹ The most frequent primary frequency in the monitoring zones of the incursion range.

Since the dynamic characteristics of the structure within the incursion range and the forced vibration mode related to intrusion load frequency and type were different, it can be seen from Figure 12 that the primary frequencies excited by the simulated intrusion were different in the incursion range. However, the similarity of the distribution features in the different intrusion cases shown in Figure 12 can still be observed. That is, the distribution patterns of cases 1 and 4 were closer due to single pedestrian intrusion, while cases 2 and 3 exhibited more broad frequency information under stronger and more complex excitations. The calculated result of the mode values of primary frequency under each case shown in Table 1 further verified this opinion. In addition to the distribution feature, different maximum primary frequencies shown in Table 1, and varied fluctuation strength in Figure 12, also contributed to distinguishing different simulated human intrusions based on the frequency domain results of dynamic distributed sensing of ultra-weak FBG.

5. Conclusions

This study reported an integrated monitoring technology used for ensuring the safety of subway operation, which verified that dynamic distributed measurement based on ultra-weak FBG was a feasible method, suitable for both train tracking and human intrusion detection in an actual engineering application. The analysis based on subway operation monitoring illustrated that the

location, speed, length, and number of train compartments could be determined through the vibration responses and distribution on the track bed. Moreover, the results of the simulated human intrusion performed in the damping track bed area during the subway outage period demonstrated that the sensing optic fiber had the potential to distinguish the strength and pattern of intruders. In view of the available test time and experimental range, the simulated cases of human intrusion were relatively limited and the detection effectiveness in the common track bed was not taken into account; this seems to be less than complete and deserves further attention when conditions permit. However, the advantages determined by the high SNR of ultra-weak FBG, when compared to other distributed sensing technologies based on common optic fiber, make us believe that the proposed method is promising for recovering and identifying signals in more complex modes.

Author Contributions: Data curation, S.L.; Formal analysis, S.L.; Funding acquisition, S.L.; Investigation, Q.N.; Methodology, Z.L., H.W., and L.W.; Project administration, Z.L. and H.W.; Resources, Z.L.; Supervision, L.W. and L.S.; Writing—original draft, Q.N., S.L., and Y.Y.; Writing—review and editing, L.S.

Funding: This research was funded by the National Natural Science Foundation of China, grant numbers 61735013 and 61875155, and the Fundamental Research Funds for the Central Universities (WUT: 2019-III-160CG).

Acknowledgments: The research reported in this paper was supported by the National Engineering Laboratory for Fiber Optic Sensing Technology, Wuhan University of Technology, and the Smart Nanocomposites Laboratory, University of California, Irvine.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Nordmark, A. Fire and life safety for underground facilities: Present status of fire and life safety principles related to underground facilities. *Tunn. Undergr. Space Technol.* **1998**, *13*, 217–269. [[CrossRef](#)]
2. Liu, Z.; Kim, A.K. Review of recent developments in fire detection technologies. *J. Fire Prot. Eng.* **2003**, *13*, 129–151. [[CrossRef](#)]
3. Jiang, D.; Zhou, C.; Yang, M.; Li, S.; Wang, H. Research on optic fiber sensing engineering technology. In Proceedings of the 22nd International Conference on Optical Fiber Sensors, Beijing, China, 15–19 October 2012.
4. Pamukcu, S.; Cheng, L.; Pervizpour, M. Chapter 1—Introduction and overview of underground sensing for sustainable response. In *Underground Sensing. Monitoring and Hazard Detection for Environment and Infrastructure*, 1st ed.; Pamukcu, S., Cheng, L., Eds.; Academic Press: Cambridge, MA, USA, 2018; pp. 1–42.
5. Soga, K.; Kechavarzi, C.; Pelecanos, L.; Battista, N.; Williamson, M.; Gue, C.Y.; Murro, V.D.; Elshafie, M.; Monzón-Hernández, D., Sr.; Bustos, E.; et al. Chapter 6—Fiber-optic underground sensor networks. In *Underground Sensing. Monitoring and Hazard Detection for Environment and Infrastructure*, 1st ed.; Pamukcu, S., Cheng, L., Eds.; Academic Press: Cambridge, MA, USA, 2018; pp. 287–356.
6. Chen, X.; Li, X.; Zhu, H. Condition evaluation of urban metro shield tunnels in Shanghai through multiple indicators multiple causes model combined with multiple regression method. *Undergr. Space Technol.* **2019**, *85*, 170–181. [[CrossRef](#)]
7. Das, S.; Saha, P. A review of some advanced sensors used for health diagnosis of civil engineering structures. *Meas. J. Int. Meas. Confed.* **2018**, *129*, 68–90. [[CrossRef](#)]
8. Huang, J.; Zhang, W.; Huang, W.; Huang, W.; Wang, L.; Luo, Y. High-resolution fiber optic seismic sensor array for intrusion detection of subway tunnel. In Proceedings of the 2018 Asia Communications and Photonics Conference (ACP), Hangzhou, China, 26–29 October 2018.
9. Yu, Z.; Liu, F.; Yuan, Y.; Li, S.; Li, Z. Signal processing for time domain wavelengths of ultra-weak FBGs array in perimeter security monitoring based on spark streaming. *Sensors* **2018**, *18*, 2937. [[CrossRef](#)]
10. Hoentsch, J.; Scholz, S. Reliable detection of security camera misalignment in urban rail environments. In Proceedings of the 2018 Joint Rail Conference, JRC 2018, Pittsburgh, PA, USA, 19–20 April 2018.
11. Catalano, A.; Bruno, F.A.; Galliano, C.; Pisco, M.; Persiano, G.V.; Gutolo, A.; Cusano, A. An optical fiber intrusion detection system for railway security. *Sens. Actuators A Phys.* **2017**, *253*, 91–100. [[CrossRef](#)]
12. Esmailzadeh Noghani, F.; Tofighi, S.; Pishbin, N.; Bahrapour, A.R. Fast and high spatial resolution distributed optical fiber sensor. *Opt. Laser Technol.* **2019**, *115*, 277–288. [[CrossRef](#)]

13. Motil, A.; Bergman, A.; Tur, M. State of the art of Brillouin fiber-optic distributed sensing. *Opt. Laser Technol.* **2016**, *78*, 81–103. [[CrossRef](#)]
14. Hong, C.; Zhang, Y.; Li, G.; Zhang, M.; Liu, Z. Recent progress of using Brillouin distributed fiber optic sensors for geotechnical health monitoring. *Sens. Actuators A Phys.* **2017**, *258*, 131–145. [[CrossRef](#)]
15. Gong, H.; Kizil, M.S.; Chen, Z.; Amanzadeh, M.; Yang, B.; Aminossadati, S.M. Advances in fibre optic based geotechnical monitoring systems for underground excavations. *Int. J. Min. Sci. Technol.* **2017**, *29*, 229–238. [[CrossRef](#)]
16. Liu, X.; Jin, B.; Bai, Q.; Wang, Y.; Wang, D.; Wang, Y. Distributed fiber-optic sensors for vibration detection. *Sensors* **2016**, *16*, 1164. [[CrossRef](#)]
17. He, Z.; Liu, Q.; Fan, X.; Chen, D.; Wang, S.; Yang, G. A review on advances in fiber-optic distributed acoustic sensors (DAS). In Proceedings of the Conference on Lasers and Electro-Optics/Pacific Rim, CLEOPR 2018, Hong Kong, China, 29 July–3 August 2018.
18. Muanenda, Y. Recent advances in distributed acoustic sensing based on phase-sensitive optical time domain reflectometry. *J. Sens.* **2018**, *2018*, 3897873. [[CrossRef](#)]
19. Peng, F.; Duan, N.; Rao, Y.; Li, J. Real-time position and speed monitoring of trains using phase-sensitive OTDR. *IEEE Photonics Technol. Lett.* **2014**, *26*, 2055–2057. [[CrossRef](#)]
20. He, M.; Feng, L.; Zhao, D. Application of distributed acoustic sensor technology in train running condition monitoring of the heavy-haul railway. *Optik* **2019**, *181*, 343–350. [[CrossRef](#)]
21. Catalano, A.; Bruno, F.A.; Pisco, M.; Cutolo, A.; Cusano, A. An intrusion detection system based on the optical fiber technology for the protection of railway assets. In Proceedings of the 18th Conference on Sensors and Microsystems, AISEM 2015, Trento, Italy, 3–5 February 2015.
22. Catalano, A.; Bruno, F.A.; Pisco, M.; Cutolo, A.; Cusano, A. An intrusion detection system for the protection of railway assets using fiber Bragg grating sensors. *Sensors* **2014**, *14*, 18268–18285. [[CrossRef](#)]
23. He, Z.; Liu, Q.; Fan, X.; Chen, D.; Wang, S.; Yang, G. Fiber-optic distributed acoustic sensors (DAS) and applications in railway perimeter security. In Proceedings of the SPIE—The International Society for Optical Engineering, Beijing, China, 11–13 October 2018.
24. Yang, M.; Bai, W.; Guo, H.; Wen, H.; Yu, H.; Jiang, D. Huge capacity fiber-optic sensing network based on ultra-weak draw tower gratings. *Photonic Sens.* **2016**, *6*, 26–41. [[CrossRef](#)]
25. Yang, M.; Li, C.; Mei, Z.; Tang, J.; Guo, H.; Jiang, D. Thousand of fiber grating sensor array based on draw tower: A new platform for fiber-optic sensing. In Proceedings of the Optical Fiber Sensors, OFS 2018, Lausanne, Switzerland, 24–28 September 2018.
26. Bartelt, H.; Schuster, K.; Unger, S.; Chojetzki, C.; Rothhardt, M.; Latka, I. Single-pulse fiber Bragg gratings and specific coatings for use at elevated temperatures. *Appl. Opt.* **2007**, *46*, 3417–3424. [[CrossRef](#)]
27. Ecke, W.; Schmitt, M.W.; Shieh, Y.; Lindner, E.; Willsch, R. Continuous pressure and temperature monitoring in fast rotating paper machine rolls using optical FBG sensor technology. In Proceedings of the 22nd International Conference on Optical Fiber Sensors, Beijing, China, 15–19 October 2012.
28. Yao, Y.; Li, S.; Li, Z. Structural cracks detection based on distributed weak FBG. In Proceedings of the Optical Fiber Sensors, OFS 2018, Lausanne, Switzerland, 24–28 September 2018.
29. Bai, W.; Yang, M.; Hu, C.; Dai, J.; Zhong, X.; Huang, S.; Wang, G. Ultra-weak fiber Bragg grating sensing network coated with sensitive material for multi-parameter measurements. *Sensors* **2017**, *17*, 1509. [[CrossRef](#)]
30. Zhou, L.; Li, Z.; Xiang, N.; Bao, X. High-speed demodulation of weak fiber Bragg gratings based on microwave photonics and chromatic dispersion. *Opt. Lett.* **2018**, *43*, 2430–2433. [[CrossRef](#)]
31. Gan, W.; Li, S.; Li, Z.; Sun, L. Identification of ground intrusion in underground structures based on distributed structural vibration detected by ultra-weak FBG sensing technology. *Sensors* **2019**, *19*, 2160. [[CrossRef](#)]
32. Guo, H.; Qian, L.; Zhou, C.; Zheng, Z.; Yuan, Y.; Xu, R.; Jiang, D. Crosstalk and ghost gratings in a large-scale weak fiber Bragg grating array. *J. Lightwave Technol.* **2017**, *35*, 2032–2036. [[CrossRef](#)]
33. Guo, H.; Liu, F.; Yuan, Y.; Yu, H.; Yang, M. Ultra-weak FBG and its refractive index distribution in the drawing optical fiber. *Opt. Express* **2015**, *23*, 4829–4838. [[CrossRef](#)]
34. Gui, X.; Li, Z.; Wang, F.; Wang, Y.; Wang, C.; Zeng, S.; Yu, H. Distributed sensing technology of high-spatial resolution based on dense ultra-short FBG array with large multiplexing capacity. *Opt. Express* **2017**, *25*, 28112–28122.
35. Liu, S.; Ding, L.; Guo, H.; Zhou, A.; Zhou, C.; Qian, L.; Yu, H.; Jiang, D. Thermal stability of drawing-tower grating written in a single mode fiber. *J. Lightwave Technol.* **2019**, *37*, 3073–3077. [[CrossRef](#)]

36. Li, Z.; Tong, Y.; Fu, X.; Wang, J.; Guo, Q.; Yu, H.; Bao, X. Simultaneous distributed static and dynamic sensing based on ultra-short fiber Bragg gratings. *Opt. Express* **2018**, *26*, 17437–17446. [[CrossRef](#)]
37. Tong, Y.; Li, Z.; Wang, J.; Wang, H.; Yu, H. High-speed Mach-Zehnder-OTDR distributed optical fiber vibration sensor using medium-coherence laser. *Photonic Sens.* **2018**, *8*, 203–212. [[CrossRef](#)]
38. Chuang, W.H.; Tam, H.Y.; Wai, P.K.A.; Khandelwal, A. Time- and wavelength-division multiplexing of FBG sensors using a semiconductor optical amplifier in ring cavity configuration. *IEEE Photonics Technol. Lett.* **2005**, *17*, 2709–2711. [[CrossRef](#)]
39. Luo, Z.; Wen, H.; Guo, H.; Yang, M. A time- and wavelength-division multiplexing sensor network with ultra-weak fiber Bragg gratings. *Opt. Express* **2013**, *21*, 22799–22807. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

Digital Approach to Rotational Speed Measurement Using an Electrostatic Sensor

Lin Li ¹, Hongli Hu ^{1,*}, Yong Qin ² and Kaihao Tang ¹

¹ State Key Laboratory of Power Equipment and Electrical Insulation, Xi'an Jiaotong University, Xi'an 710049, China; ll123xjtu.edu.cn@stu.xjtu.edu.cn (L.L.); mrerr07@stu.xjtu.edu.cn (K.T.)

² State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China; yqin@bjtu.edu.cn

* Correspondence: hlhu@mail.xjtu.edu.cn

Received: 24 March 2019; Accepted: 28 May 2019; Published: 4 June 2019

Abstract: In industrial production processes, rotational speed is a key parameter for equipment condition monitoring and fault diagnosis. To achieve rotational speed measurement of rotational equipment under a condition of high temperature and heavy dust, this article proposes a digital approach using an electrostatic sensor. The proposed method utilizes a strip of a predetermined material stuck on the rotational shaft which will accumulate a charge because of the relative motion with the air. Then an electrostatic sensor mounted near the strip is employed to obtain the fluctuating signal related to the rotation of the charged strip. Via a signal conversion circuit, a square wave, the frequency of which equals that of the rotation shaft can be obtained. Having the square wave, the M/T method and T method are adopted to work out the rotational speed. Experiments were conducted on a laboratory-scale test rig to compare the proposed method with the auto-correlation method. The largest relative errors of the auto-correlation method with the sampling rate of 2 ksps, 5 ksps are 3.2% and 1.3%, respectively. The relative errors using digital approaches are both within $\pm 4\%$. The linearity of the digital approach combined with the M/T method or T method is also superior to that of the auto-correlation method. The performance of the standard deviations and response speed was also compared and analyzed to show the priority of the digital approach.

Keywords: electrostatic sensor; digital approach; rotational speed; correlation algorithm

1. Introduction

In industrial applications, rotational speed measurement is a crucial part for condition monitoring, speed control, and protective supervision of rotation equipment, such as generators, steam turbines, and gas turbines. Various kinds of tachometers based on different mechanisms, such as optical, electrical, and magnetic induction, have been developed and widely used to measure the rotational speed of target objects. W.H. Yeh presented a high-resolution optical shaft encoder to monitor the rotation behavior of a motor [1] and J. N. Lygouras presented a solution for processing the pulses from an optical encoder attached to a motor shaft [2]. W. Lord and R.B. Chatto provided a homopolar tachogenerator with low inertia and noise generation, making it particularly suitable for velocity-control systems using high-performance DC motors as the power actuators [3]. C. Giebler designed a contactless sensor based on the giant magneto-resistance (GMR) effect for position detection and speed sensing [4]. Z. Shi implemented a tachometer using a magnetoelectric composite as a magnetic field sensor which was mounted where the magnetoelectric composites had the highest sensitivity [5]. Considering the operating mode, the rotational speed measurement method can be classified into digital and analog categories. In the analog tachometer output a voltage or current signal proportional to the speed can be used to provide a feedback signal in a closed-loop speed control system [6]. Digital tachometers

have been used over the years, which utilize electronic circuits to measure an average frequency of incoming pulses from an encoder mounted on a shaft [7].

In order to overcome the harsh condition, such as a high temperature, heavy dust environment, the electrostatic method has been used to realize the rotational speed measurement. The electrostatic sensor is adaptable for the speed measurement in various industrial conditions for the advantages of contactless measurement, low cost, simple structure, and easy installation and maintenance. Recently, Y. Yan and L.J. Wang utilized electrostatic sensors and a correlation algorithm to calculate the period or elapsed time and successfully obtained the rotational speed of a rotational shaft [8,9]. The electrostatic method to measure rotational speed utilizes the electrode to induce the electric field generated by carried charges on the shaft. When two materials are touched or rubbed together electrical charge is usually transferred from one to the other [10]. According to the theory of tribo-electric charging, each material has its own surface work function. Then the surface electron transfer will occur in the driven Fermi energy level [11]. The material type determines the work function, which indicates the capability of a material to hold onto its free electrons. Thus, the polarity and quantity of the charge generated due to the triboelectric friction are mainly decided by the material type and surface roughness, and are also affected by its surrounding environment, like temperature and humidity. Thus, if the shaft has a greater work function than the air, the relative motion between the rotational shaft and the air will generate some charge on the surface of shaft due to triboelectric friction. If one of the materials is a good insulator, the charge persists on its surface for a long time, and the effects of the charge transfer are readily apparent [10].

The principle of rotational speed measurement using an electrostatic sensor is shown in Figure 1. According to electrostatic induction theory, when the surface of the shaft carries some charge due to triboelectric friction with the air, it will influence the electrostatic field of its surroundings, thus, the induced charge will be generated on the surface of sensing electrode when it is installed near the shaft. The fluctuation of the induced charge on the electrode generates a current which can be converted into a voltage signal via a current-to-voltage conversion circuit. Additionally, a charge amplifier circuit can be adopted to translate the charge into a voltage signal [12]. The voltage signal collected from the electrode contains a wealth of rotational information, thus processing and analyzing the output signals from the sensor will result in obtaining further information.

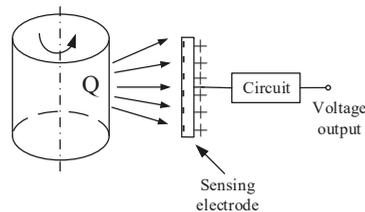


Figure 1. Measurement principle using the electrostatic method.

By now, electrostatic sensors in conjunction with correlation methods, including the cross-correlation method using dual electrostatic sensors and the auto-correlation method using a single electrostatic sensor, have been used to determine rotational speed [8,9,13]. Figure 2 describes the rotational speed measurement system which uses electrostatic sensors and the correlation method. Two or more channels of sensors and the corresponding condition units are connected to an A/D converter. Then a microprocessor system or a computer is needed to execute the correlation algorithm.

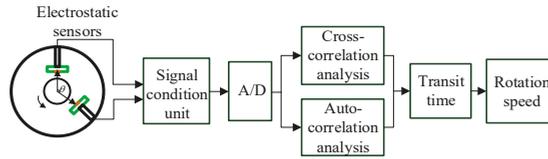


Figure 2. Principle of the rotational speed measurement using electrostatic sensors.

In the time domain, the definition of the cross-correlation function between real power signals $x(t)$ and $y(t)$ is:

$$R_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)y(t + \tau)dt \tag{1}$$

Figure 3 illustrates how to obtain the time-delay between two electrostatic signals using the cross-correlation method. The rotational speed v_c (revolutions per minute, rpm) can be calculated by the sensor angle spacing θ (degree) and the transit time τ (s):

$$v_c = \frac{1}{T} \times 60 = \frac{1}{\frac{360}{\theta} \cdot \tau} \times 60 = \frac{\theta}{6\tau} \tag{2}$$

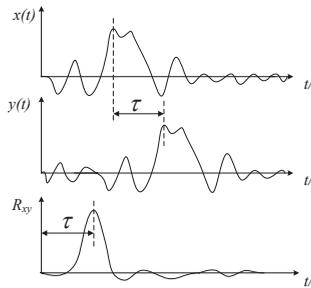


Figure 3. Illustration of the cross-correlation method.

If $x(t)$ and $y(t)$ are the same signals obtained by one electrode, Equation (1) turns out to be the auto-correlation function. With respect to the auto-correlation method, only one channel of the electrostatic signal is needed. The time-delay τ between signal $x(t)$ and signal $y(t)$ is the rotational period T (s) and θ equals 360 degree. Using the auto-correlation algorithm can extract the time of rotation period. After that, the rotational speed v_a (rpm) can be obtained as follows:

$$v_a = \frac{60}{T} \tag{3}$$

Obviously, the correlation method needs to locate the coordinate of the first dominant peak in the waveform of the correlation function, which is influenced by the sampling rate to a great extent. At the same time, the waveforms collected by inducing the signal from a cylinder dielectric sleeve contain complex information and a faint sign of the periodical component. Although the correlation calculation of the waveform has good performance and successfully extracts the elapsed time, the computational accuracy of the correlation method is obviously affected by the sampling rate and signal noise [14].

For the sake of improving the performance of rotational speed measurement via the electrostatic method, this paper proposes an approach to generate a square wave from an electrostatic sensor in order to obtain the rotational speed via digital methods, thus eliminating the influence of the sampling rate and signal noise, and also simplifying the system complexity. In the following article, “square wave” refers to the output waveform from the comparison circuit which generates a pulse

every rotational period. Implementation of a rotational speed measurement system based on this method is presented. Compared to the rotational speed measurement method using an electrostatic sensor in conjunction with correlation, this designation leaves out the AD converter and simplifies the computation code which is more adaptive for the implementation in a microprocessor system.

2. Measurement Principle and Finite Element Simulation

2.1. Measurement Principle

Inspired by the photoelectric method which fixes a strip of a reflection element, this experiment uses a strip of polytetrafluoroethylene (PTFE) stuck to the rotational shaft. The measurement principle is shown in Figure 4. Adopting this designation, the charge generated on the PTFE by the relative rotation with the air will pass the sensor once a revolution, which makes the waveform have a strong sign of periodicity. The electrostatic signal is firstly transformed into a voltage signal. Then, after amplifying, filtering, and comparison, the analog signal will be transformed into a square wave, which is convenient to be connected to a DSP or FPGA system to execute the following rotational speed calculation algorithm.

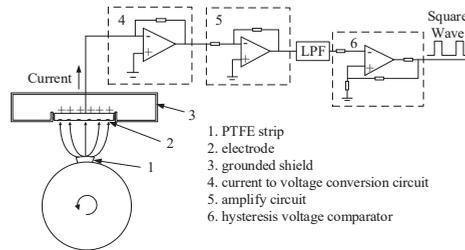


Figure 4. Measurement principle.

2.2. Rotation Speed Computation Algorithm

Usually, three methods are adopted to evaluate the speed based on these square waves: (1) Measuring the elapsed time, commonly termed as the *T* method, which calculates the reciprocal of the duration between consecutive pulses to obtain the frequency; (2) pulse counting, commonly termed as the *M* method, which counts the number of pulses generated within a prescribed period of time; and (3) constant elapsed time, commonly termed as the *M/T* method, is a combination of pulse counting and measuring elapsed time [15–17]. The principles of the three methods are shown in Figure 5.

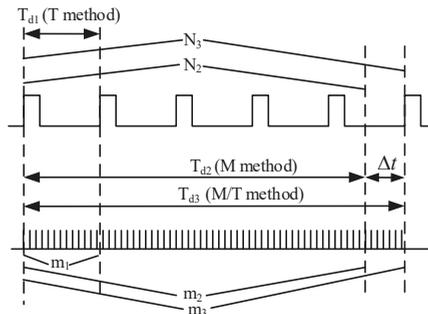


Figure 5. Principles of *T*, *M*, and *M/T* methods.

As seen from Figure 5, the detecting time of the *T* method and *M* method can be obtained according the Equations (4) and (5), correspondingly, where m_1 is the number of clock pulse counting during one

period of square wave, m_2 is the number of pulse counting during the prescribed time, and f_s is the frequency of clock pulse used for counting and timing:

$$T_{d1} = \frac{m_1}{f_s} \quad (4)$$

$$T_{d2} = \frac{m_2}{f_s} \quad (5)$$

Then the rotational speed v_1 (rpm) of the T method can be calculated according to Equation (6). The rotational speed v_2 (rpm) of the M method can be calculated according to Equation (7), where N_2 is the integer number of square wave during the prescribed time:

$$v_1 = \frac{60f_s}{m_1} \quad (6)$$

$$v_2 = \frac{60f_s N_2}{m_2} \quad (7)$$

Different from the M method, which ceases the pulse counting once the prescribed time runs out, the M/T category goes on counting after the prescribed time and stops at the first pulse of rotation after the prescribed time. Thus, the detecting time of the M/T method T_{d3} (s) equals T_{d2} (s) plus Δt . Parameter N_3 is the number of square waves during the detecting time. Parameter m_3 is the number of clock pulses during the detecting time. The detecting time T_{d3} (s) can also be obtained using parameter m_2 and the frequency of clock pulse f_s used for timing, as shown in Equation (8). Thus, the rotational speed of the M/T method is calculated by Equation (9):

$$T_{d3} = \frac{m_3}{f_s} \quad (8)$$

$$v_3 = \frac{60f_s N_3}{m_3} \quad (9)$$

The calculation errors of the three method can be derived according to Equations (10)–(12). Parameters m_1^0 , m_2^0 , and m_3^0 are the ideal pulse numbers needed to perfectly overlap the detecting time. Parameter N_2^0 equals the detecting time divided by the rotation period, which is not an integer in most cases.

$$e_1 = \left(\frac{60f_s}{m_1} - \frac{60f_s}{m_1^0} \right) / \frac{60f_s}{m_1^0} = \frac{m_1^0 - m_1}{m_1} \quad (10)$$

$$e_2 = \left(\frac{60f_s N_2}{m_2} - \frac{60f_s N_2^0}{m_2^0} \right) / \frac{60f_s N_2^0}{m_2^0} = \frac{m_2^0 N_2 / N_2^0 - m_2}{m_2} \quad (11)$$

$$e_3 = \left(\frac{60f_s N_3}{m_3} - \frac{60f_s N_3^0}{m_3^0} \right) / \frac{60f_s N_3^0}{m_3^0} = \frac{m_3^0 - m_3}{m_3} \quad (12)$$

As seen from Equation (10), it can be observed that the error of the T method is low at a high speed (m_1 decreases) and the M method resolution is not high at a low speed (N_2 is not stable). However, considering the frequency of the clock pulse used for timing in this article is 150 MHz, which is significantly greater than the frequency of rotation, the counting errors of the T method and M/T method are extremely small compared to their denominators. The T method and M/T method have absolutely accurate counting numbers of the square wave from circuits 1 and N_3 , correspondingly. The calculating errors of the T method and M/T method mainly result from the counting number of the pulse clock ($m_1^0 - m_1$, $m_3^0 - m_3$). By contrast, regarding the M method, the difference between square counting N_2 and N_2^0 may result in an obvious error.

The response speed can also be determined from the principle. Among them, the *T* method has the fastest response speed, which enables outputting a result every period. The *M/T* method and *M* method generate a result based on the statistical average principle, which gives them a relatively slow response speed. In summary, the *T* method is more adaptive for the dynamic measurement of variable rotational speed and the *M/T* method or *M* method is more adaptive for constant speed measurement or mean values of a certain time. Considering the response time and accuracy simultaneously, this article utilizes the *M/T* method and the *T* method to deal with the square wave from the measurement circuits.

2.3. Finite Element Simulation

A simulation using a strip object with evenly distributed charge was conducted utilizing COMSOL software (the COMSOL Group in Stockholm, Sweden) to imitate how the rotationally charged strip influences the induced charge on the electrostatic sensor. The model is shown in Figure 6 and the simplified two-dimensional schematic of the simulation is illustrated in Figure 7. The strip object is a 7.64 degree arc with a radius of 15 mm, which is placed tightly to the surface of the metal shaft. The length of the strip is 20 mm. The radii of metal shaft and outer shielding are 15 mm and 30 mm, correspondingly. The charge is evenly distributed on the surface of the object using the Surface Charge Density setting in COMSOL. The surface charge density is set to be 0.025 C/m², so the total amount of charge on the strip is 1 μC. An electrode 20 mm long and 2 mm wide is placed 17 mm away from the central axis with the same *z* coordinate of strip. In the simulation, the strip rotates around the central axis by controlling the angle with respect to the positive axis *x*, beginning at −180 degree and stopping at 180 degree with a step size of one degree. The electrostatic field can be described by the Poisson equation and its corresponding boundary conditions:

$$\begin{cases} -\nabla \cdot (\epsilon_0 \epsilon_r \nabla \varphi) = \rho \\ V_E = 0, V_B = 0 \end{cases} \quad (13)$$

where V_E is the potential of electrode, V_B is the potential of the shielding, and ρ is the space charge density.

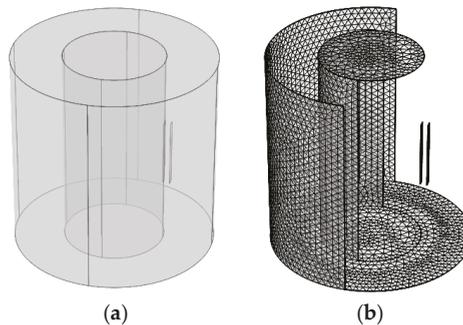


Figure 6. (a) Structure of the model, and (b) a mesh of the simulation model.

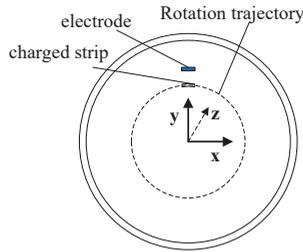


Figure 7. Simplified schematic.

Figure 8a depicts the induced charge on the electrode when the rotation of the strip begins farthest from the electrode, then passes by the electrode, and finally returns to the beginning position. The amount of induced charge on the electrode reflects the ‘far-near-far’ rotational process. The variation of charges on the electrode generates current i , which can be calculated by:

$$i = \frac{dq_e}{dt} \tag{14}$$

where q_e represents the amount of induced charge on the electrode. By calculating the difference of the induced charges, the current can be obtained and is shown in Figure 8c. When the strip is far from the electrode, the current is very weak and can be regarded as 0. When the strip rotates adjacent to the electrode, the current becomes larger. As seen from Figure 8c, the derivation of the induced charge contains thorns and wobbles. This can be explained due to the discretization and unavoidable computation error of the finite element simulation, the curve of induced charge is not smooth enough (shown in Figure 8b), thus leading to the thorns of the derivative curve. To acquire a more optimal result a moving average is applied to smooth the data and the result is shown in Figure 8d.

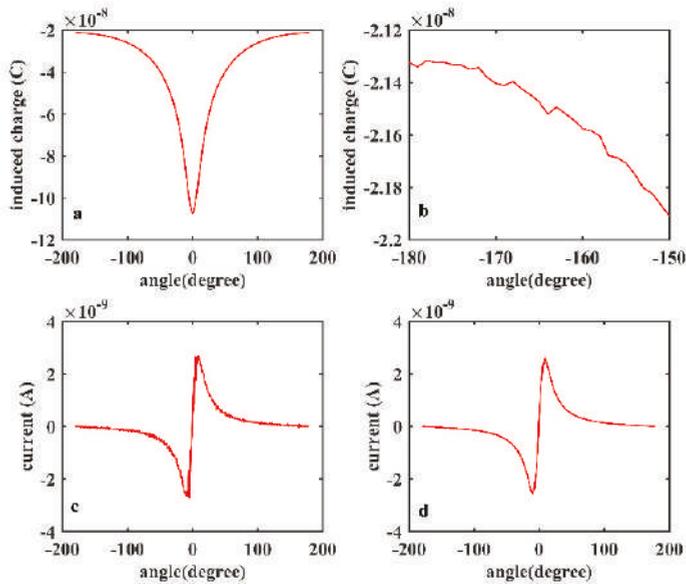


Figure 8. (a) Simulation result of induced charges; and (b) partial drawn of the induced charge. (c) Derivation of the induced charge; and (d) smoothed curve of the current.

Then, we use the rotation of the electrode to replace the rotation of shaft, so that the shaft is relatively at rest, as shown in Figure 9. According to the superposition principle of the electric field, the amount of induced charge Q on the electrode with a displacement of angle α can be calculated by Equation (15). Function $q(\theta)$ is the amount of charges on the location of angle θ . Parameter α indicates the rotated angle and is also a reflection of time. Function $f(\theta-\alpha)$ means when a unit charge is $\theta-\alpha$ degrees away from the electrode, the amount of induced charge on the electrode generated by this unit charge is $f(\theta-\alpha)$. Thus, the total amount of induced charge can be calculated by integrating over θ from -180° to 180° .

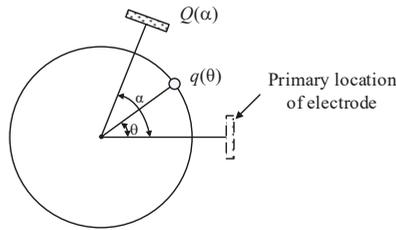


Figure 9. Schematic of electrode rotation.

$$Q(\alpha) = \int_{-180^\circ}^{180^\circ} q(\theta)f(\theta - \alpha)d\theta \tag{15}$$

$$Q(\alpha) = \int_{-\infty}^{+\infty} q_{circle}(\theta)f_r(\theta - \alpha)d\theta \tag{16}$$

As shown in Figure 10, Equation (15) illustrated by Figure 10a with an integral range of $[-180^\circ, 180^\circ]$ can be transformed into Equation (16) illustrated by Figure 10b. In Figure 10b, the values of $f_r(\theta)$ when θ is out of range $[-180, 180]$ are zeros and the waveform of function $f_r(\theta)$ is the same as in Figure 8a, which is only different in amplitude. The induced charge on the electrode can be regarded as a weighted mean value of the contribution of the charge in a sensitive area. Meanwhile, Equation (16) is a convolution operation between the charge distribution function $q_{circle}(\theta)$ and the function $f_r(\theta)$, thus $f_r(\theta)$ can be regarded as a filter function. The low pass filter property can also be obtained from [18]. Obviously, function $f_r(\theta)$ is influenced by the rotational speed ($f_r(\theta_0 + \omega t)$). The cutoff frequency of $f_r(\theta)$ increases with the speed (narrow in the time domain, broad in the frequency domain). Through the analysis, the electrostatic electrode in this case of application can be regarded as a low pass filter which adaptively adjusts its cut-off frequency. Thus, the waveform of the signal mainly contains a low frequency component if the electro-magnetic interference is well shielded, which helps to explain the signal obtained in the experimental part.

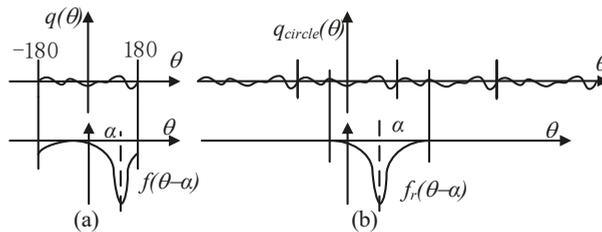


Figure 10. Explanation of Equations (15) (a), and Equations (16) (b).

3. Hardware Design

3.1. Sensor Board

The sensor board shown in Figure 11 contains an electrode and a current-to-current to voltage conversion circuit. A surface-tinned copper strip 20 mm long and 2 mm wide is utilized as the electrode. The electrode is connected to the current voltage conversion circuit, which is built on an LMP7721 (Texas Instruments in Dallas, TX, USA) amplifier with an extremely low bias current of 20 fA maximum. Figure 12 illustrates the schematic of the circuit. The feedback resistor consists of two $10^8 \Omega$ resistors connected in series, which determines the transimpedance gain. In actual application, a feedback capacitor is needed to guarantee the stability of the circuit by inhibiting the high frequency noise.



Figure 11. Sensor board.

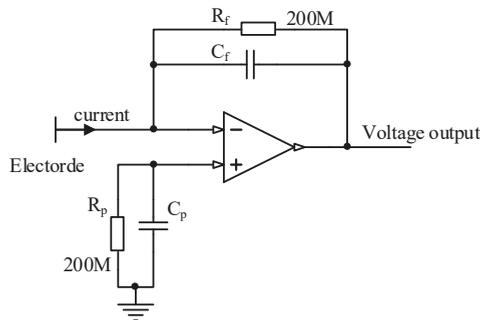


Figure 12. Schematic of current-to-voltage circuit.

The relationship between the output voltage and the input current from the electrode can be calculated according to Equation (17). Thus, the sensitivity of the circuit is 0.2 V/nA. The purpose of the balance resistor and capacitor is to make the impedance of the two inputs equal, thus, the bias current of the amplifier generates no additional offset voltage on the output.

$$U_o = R_f i_e \quad (17)$$

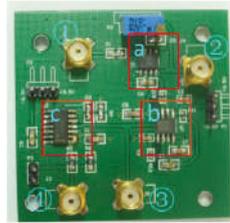
When the electric field near the electrode varies with the rotation of the charged strip, a small current signal will be generated and transformed into a voltage signal via the feedback resistance on the amplifier. The voltage output of the sensor board is collected by the condition unit via a shielded cable to avoid electromagnetic interference in the space.

3.2. Signal Condition Unit

The condition unit in our experiment is shown in Figure 13, which has amplifying, filtering, and comparing circuits designed to generate a square wave. Four connectors are placed on the board.

Connector 1 is used to connect to the output of the sensor board. Connectors 2 and 3 are used to observe the result of amplifying and filtering correspondingly. The square waveform from the comparing circuit is transmitted to a DSP chip via connector 4 or the pin headers nearby.

The amplifying circuit uses the same amplifier chip with the sensor board to meet the performance requirements. The voltage gain of the amplifier can be adjusted by a slide rheostat. Then a third-order Butterworth low pass filter with Sallen-Key topology is used for filtering and inhibiting the noise. The passband frequency is 400 Hz and the stopband frequency is 2.4 kHz. A smooth waveform improves the stability of the square wave. Finally, a hysteresis comparator is utilized to transform the waveform into a square wave which is connected to the DSP board for speed calculation.



Markers ①②③④ refer to four SMA connectors.

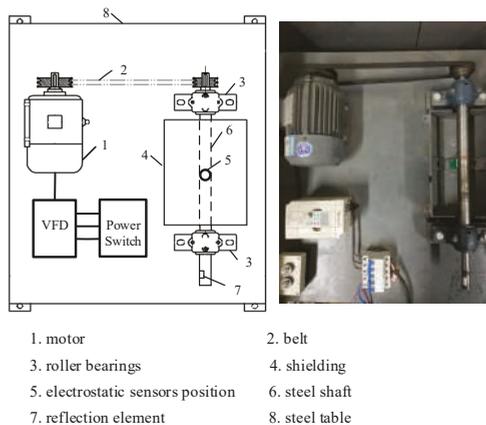
- a. amplifying circuit
- b. filter circuit
- c. hysteresis comparator

Figure 13. Signal condition unit.

4. Experiment Results and Discussion

4.1. Experiment Conditions

A laboratory-scale test rig is designed and built for rotational speed measurement. Figure 14 shows the schematic of the test rig. An external power supply connects to a variable-frequency drive (VFD) via a power switch. The torque of the motor is translated to the shaft via a belt. Thus, the rotational speed of the shaft can be adjusted by the VFD. The shaft is made of steel and supported by two roller bearings with a belt pulley mounted on its side. The middle part of the shaft is surrounded by a grounded cylindrical metal shielding. As shown in Figure 15, a strip of PTFE about 2 mm wide and 20 mm long, the lengthwise direction of which is parallel to the axial direction of shaft, is glued tightly on the shaft. The sensor board is set on the inner wall of metal shielding via a copper pillar, thus the electrode is under the central axis of the shaft and the trajectory of the strip. The copper pillar is utilized to adjust the distance between the sensor and the shaft. In order to inhibit the vibration of the rig, the steel table was screwed to the ground via an expansion screw.



- 1. motor
- 2. belt
- 3. roller bearings
- 4. shielding
- 5. electrostatic sensors position
- 6. steel shaft
- 7. reflection element
- 8. steel table

Figure 14. Schematic and photograph of the test rig.



Figure 15. Photograph of the sensor and strip.

Experiments were conducted on the rig using the same dimension parameters as the simulation. The rotational speed of shaft was adjusted from 300 rpm to 3200 rpm with an increment of 100 rpm via the VFD. To make a comparison between the digital approach and the correlation method, each point was measured five times. Meanwhile, five values of the T method and M/T method transformed from the DSP for each point were saved for analysis. A photoelectric reflection digital tachometer with an accuracy of $\pm 0.05\%$ of the reading plus 1 rpm was used to provide a reference speed in our experiment. The ambient temperature was controlled between 20 °C and 24 °C and the relative humidity was kept between 55% and 65%. The square wave was connected to the external interrupt pin to contend with the square wave immediately. The code of realizing the T method and the M/T method were programed and written into the DSP board separately to test the measurement performance. The DSP transmitted the measurement results of the T method and the M/T method to computer via RS232 serial communication. In the experiment, the prescribed time of the M/T method is set to be 1 s.

Seen from the principle of the correlation method, it can be found that the auto-correlation method can be regarded as a particular case of the cross-correlation method, which leaves out the influence of the installation angle error, the distance differences of the two electrodes to the shaft, and the differences between two channels' circuits. These factors make the accuracy of cross-correlation method not as good as the auto-correlation method. Meanwhile, the cross-correlation method needs two channels of circuits, which is not consistent with the setting in this experiment. Thus, the experiment only makes a comparison between digital approaches and the auto-correlation method.

4.2. Signals

The proposed approach utilizes the electrostatic sensor to induce the charge on the strip of PTFE, which obtains a strong periodic signal. Figure 16 shows the signals before and after filtering, which contain evenly distributed waveforms similar to the simulation result (Figure 8d). The filtered waveform obviously has a higher signal-to-noise ratio. The high signal-to-noise ratio and the strong periodicity helps to improve the stability of the square wave transformed from the signal, which is very important for the rotational speed calculation based on the square wave. Figure 17 shows the square waveform generated by the hysteresis comparator. In order to illustrate the wave clearly, Figure 17 only shows 0.5 s of the signal.

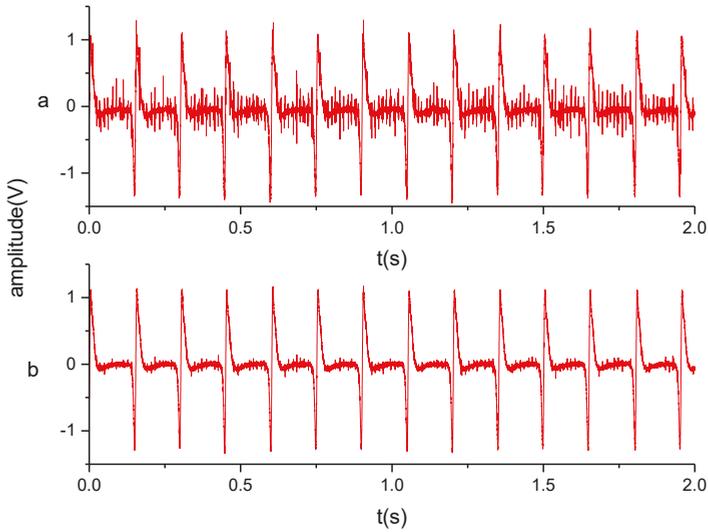


Figure 16. Input (a) and output (b) of the filtering circuit.

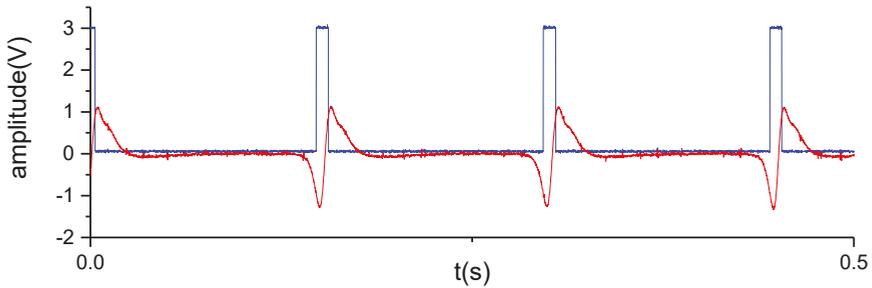


Figure 17. Analog input and digital output of the comparator circuit.

4.3. Accuracy

The mean values of the measurement results for the T method and the M/T method are plotted in Figure 18. Their relative errors are compared with the photoelectric reflection digital tachometer and are listed in Table 1. The linearity of the T method and M/T method are about 0.81‰ and 1.31‰, correspondingly. The measurement results are highly consistent with those of the photoelectric tachometer. Meanwhile, the differences between the T method and the M/T method are hardly discernible by eye.

As seen from the principle, the proposed digital method needs no sampling via an analog-digital converter, while the sampling rate is an important factor that determines the accuracy of the method based on the correlation algorithm. In order to make a comparison between these two methods, the analog signals are also collected at different sampling rates. The auto-correlation functions are calculated using the filtered analog signal. Figure 19 shows the auto-correlation of an example collected at the rotational speed of 400 rpm using a sampling rate of 2 ksp/s.

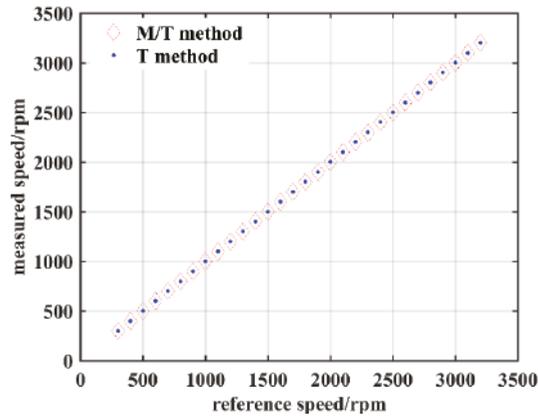


Figure 18. Measurement speed.

Table 1. The relative errors of the measurement points.

Reference Speed (rpm)	Measured Speed (rpm)		Relative Error (%)	
	M/T	T	M/T	T
300	300.50	300.11	1.67	0.36
400	399.97	400.45	-0.08	1.12
500	501.04	501.87	2.08	3.74
600	600.63	601.08	1.05	1.80
700	701.09	701.39	1.56	1.98
800	799.12	799.26	-1.10	-0.92
900	900.10	899.38	0.11	-0.69
1000	1000.02	1001.35	0.02	1.35
1100	1099.65	1100.79	-0.32	0.72
1200	1200.06	1199.55	0.05	-0.38
1300	1300.30	1301.22	0.23	0.94
1400	1400.83	1401.27	0.59	0.91
1500	1501.81	1501.34	1.21	0.89
1600	1601.33	1601.60	0.83	0.99
1700	1700.71	1699.35	0.42	-0.38
1800	1799.75	1802.47	-0.14	1.37
1900	1900.12	1899.11	0.06	-0.47
2000	2000.34	2001.50	0.17	0.75
2100	2100.87	2101.57	0.41	0.75
2200	2201.38	2202.34	0.63	1.06
2300	2298.58	2300.50	-0.62	0.21
2400	2400.52	2404.21	0.22	1.75
2500	2497.95	2503.25	-0.82	1.30
2600	2600.32	2599.51	0.12	0.19
2700	2701.59	2698.14	0.59	-0.69
2800	2802.13	2802.10	0.76	0.75
2900	2899.71	2901.81	-0.10	0.62
3000	3000.15	3001.87	0.05	0.62
3100	3100.52	3098.77	0.17	-0.40
3200	3202.58	3202.23	0.81	0.70

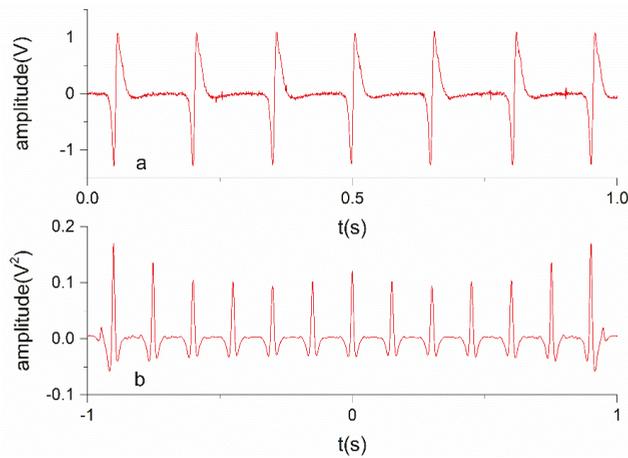


Figure 19. (a) Signal collected at the rotational speed of 400 rpm. (b) Auto-correlation of the signal.

The waveform in Figure 20 is a partial drawing of the part in Figure 19b. As shown in Figure 20, by detecting the first peak after 0 s, the period T of the rotation can be obtained. It can be observed that the waveform near the first peak after 0 s is very smooth, which benefits confirming the accurate and stable value of the period. However, due to the discretization of the data series, the obtained period T will be a time length away from the actual time of the rotation period with a significant probability despite the auto-correlation method confirming the nearest time point to the ideal time point. Moreover, when the signal contains an obvious level of noise or a weak periodicity, the waveform near the peak of the auto-correlation function will be fluctuant, which impairs the result’s accuracy.

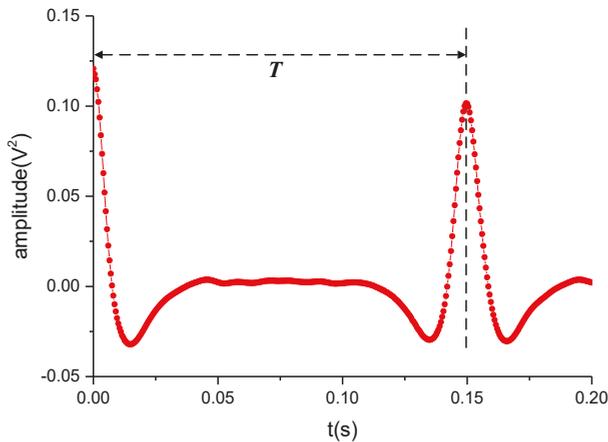


Figure 20. Partial enlarged drawing of Figure 17b.

In order to show the accuracy of this digital method, signals are collected at the sampling rate of 2 ksp/s and 5 ksp/s and analyzed using the auto-correlation method. Relative errors of the two methods are plotted and compared in Figure 21. Figure 21 shows the measurement errors of the digital approaches and the auto-correlation method at the sampling rate of 2 ksp/s and 5 ksp/s. It can be seen from Figure 21 that the digital approaches have better accuracy and the relative errors obtained using 5 ksp/s are smaller than those sampled at 2 ksp/s. The auto-correlation method is apparently influenced by the sampling rate. Meanwhile, the accuracy of auto-correlation method has the tendency to increase

with the rotational speed, which has been explained in 14. The linearity of the auto-correlation method sampled at 2 ksp/s and 5 ksp/s are 3.17% and 1.33%, respectively, which are significantly greater than those of the M/T method and the T method.

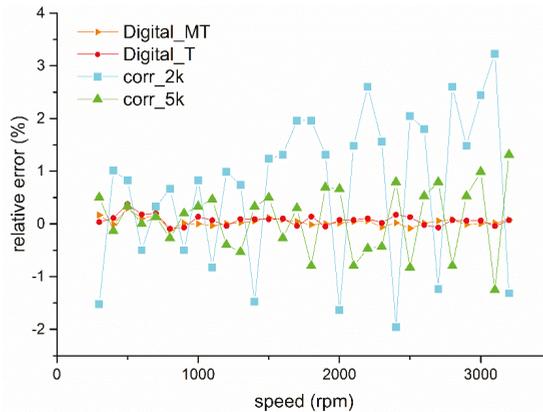


Figure 21. The absolute value of the relative error.

4.4. Standard Deviation

In order to research the robustness of the proposed method, the standard deviations of each measurement point are listed in Table 2. As seen from the table, regarding digital approaches, the M/T and T methods both have significantly small standard deviations. The standard deviations of the auto-correlation method in Table 2 contains a number of zeros and some other values, which can be easily understood from the principle. For example, using 5 ksp/s (Δt is 0.2 ms) to collect the signal of 1400 rpm, the period of which is about 214 times that of Δt , the calculated rotational speed by finding the first peak of the auto-correlation function will be some discrete value calculated by $60/[(214 \pm n)\Delta t]$ ($n = 0, 1, 2, \dots$), like 1395.35, 1401.87, or 1408.45. There are two factors affecting the standard deviations of the auto-correlation method: (1) If the variation of rotation speed is not obvious enough to change the location of the first peak on the auto-correlation function, the measurement results will remain unchanged; and (2) if the locations of the first peak in the auto-correlation function differs one or two sampling intervals from each other due to signal differences, the obtained rotational speeds will show obvious fluctuations.

The standard deviations of the M/T and T methods in Table 2 are all within 1 rpm. Meanwhile, the standard deviations of the M/T method are much smaller than those of the T method. As seen from the principle, the M/T method can be regarded as a mean value of several consecutive T methods. Due to the high response speed of the T method, it is more sensitive to the variation of rotation, which makes its standard deviations greater than those of the M/T method. The minor standard deviations of the M/T and T methods mainly arise from the slight fluctuations of the actual rotation state, which is probably related to the unsteady output rotational speed of the motor and the slippage of the belt on the sheave. With respect to the digital approaches, no matter the M/T method or T method, both have very little spread in the measured speed.

Table 2. The standard deviations of the measurement points.

Reference Speed (rpm)	Standard Deviation (rpm)			
	<i>M/T</i>	<i>T</i>	2k	5k
300	0.055	0.175	1.755	2.047
400	0.013	0.292	2.641	2.250
500	0.013	0.397	4.652	1.867
600	0.032	0.274	1.651	2.161
700	0.023	0.458	3.437	1.472
800	0.097	0.142	2.367	2.592
900	0.091	0.115	3.011	1.206
1000	0.068	0.378	0	2.790
1100	0.011	0.146	0	0
1200	0.044	0.190	0	0
1300	0.015	0.284	0	2.525
1400	0.051	0.038	0	0
1500	0.010	0.027	0	0
1600	0.013	0.265	0	4.674
1700	0.017	0.190	0	0
1800	0.019	0.467	0	2.542
1900	0.010	0.065	0	0
2000	0.027	0.284	0	0
2100	0.029	0.234	0	0
2200	0.051	0.006	18.069	0
2300	0.016	0.368	0	7.878
2400	0.051	0.281	0	0
2500	0.050	0.415	0	0
2600	0.048	0.155	0	0
2700	0.018	0.364	0	0
2800	0.069	0.082	0	0
2900	0.039	0.301	0	0
3000	0.052	0.131	0	0
3100	0.012	0.327	0	0
3200	0.094	0.345	46.747	0

4.5. Response Time

The response time of each approach can be determined from their principles and data process procedures. Regarding the *M/T* method, the response time is decided by adding extra time to the prescribed time. Regarding the analog method, usually the sampling length should be predetermined. Thus, the time needed to acquire one measurement result is nearly confirmed. Even if the auto-correlation method self-adaptively adjusts the sampling length according to the nearest obtained rotational speed, the response speed is still not as fast as the *T* method for the reason that the auto-correlation method needs at least two periods of rotation to achieve the correlation calculation. Moreover, data collection and processing also consume a certain amount of time.

By contrast, the *T* method can output a result every rotational period for the reason that the counter in the DSP can work independently from the code and the DSP only needs to perform an easy computation of the counter number and serial communication. Experiments were conducted to test the capability of the *T* method to measure the variable speed. The motor was adjusted by the VFD output frequency to work at three stages: acceleration by increasing the frequency from 0 Hz to 20 Hz over 5 s, 4 s of constant speed, and deceleration by decreasing the frequency from 20 Hz to 0 Hz over 3 s. Figure 22a shows the 256 acquired rotational speeds via the *T* method, which successfully monitors the acceleration and deceleration processes. The rising and decline curves are not perfectly straight lines because the acceleration of the shaft is not absolutely constant. It can be observed that at a constant frequency, the measurement results are of good stability.

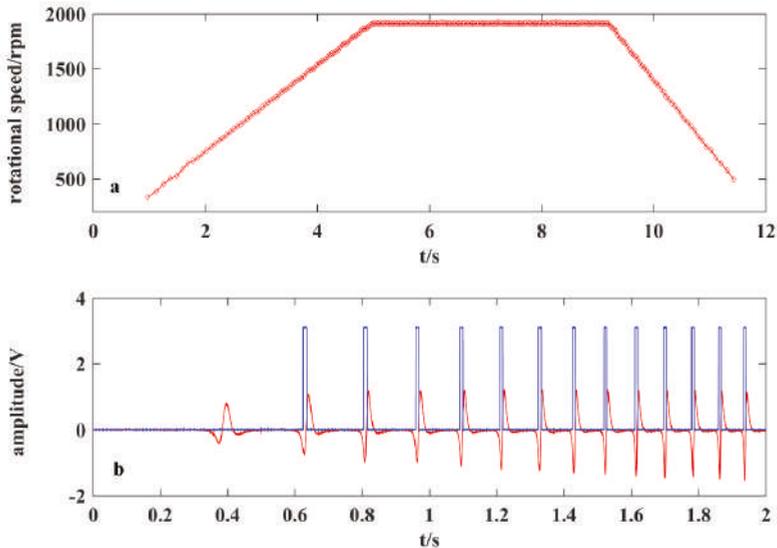


Figure 22. (a) Measured rotational speed by the T method; and (b) waveform during acceleration process.

Figure 22b shows a part of the waveform during the acceleration stage. The acceleration and deceleration process can be clearly observed through the interval variations of square waves. The waveform near 0.5 s has not been transformed into a square wave because a limited electrostatic charge is generated on the strip at low speed, thus making the signal unsuitable for the following square wave generation circuit. With the increases of rotational speed, the amount of charge rises and then becomes stable because of the dynamic balance reached between the natural discharge and recharge. The signal amplitude changes as the unbalanced charge increases or decreases at a low speed, which makes the comparison voltage appear at different positions relative to the waveform. This phenomena limits the application of this method in measuring low variable speed.

5. Conclusions

The work in this paper dedicates to find a more effective approach to cooperate with electrostatic sensors to improve the performance of rotational speed measurement. The proposed approach utilized the electrostatic sensor to induce a charge on a strip PTFE, which obtained a strong periodical signal. Simulation results also described the expected waveform when a strip of charges rotates near an electrode. By adopting a suitable signal condition unit, a square wave, the frequency of which was equal to that of the rotational speed, has been obtained. Having the square wave proportional to rotational speed, the M/T method and T method were adopted to calculate the speed in a DSP system. Experiments were conducted to compare the digital approaches with the auto-correlation method. Through experimental analysis, several conclusions can be summarized as follows:

1. Accuracy: Compared with the auto-correlation method, the M/T method and T method both have an obviously higher accuracy. The linearity of the M/T method and T method are about 0.81% and 1.31%, correspondingly, which are much better than those of the auto-correlation method sampled at 2 ksps (3.17%) or 5 ksps (1.33%). Due to the signal discretization, the auto-correlation method can only obtain some discrete values. Improving the sampling rate, calculation quantity, and storage space, the hardware cost will also increase correspondingly.

2. Robustness: The auto-correlation method has a stable performance in some measurement points and also has some obvious standard deviations, which resulted from the signal discretization. However, the M/T method and T method obtained particularly small standard deviations among all

the measurement points, both within 1 rpm. The M/T method acquired more stable results than the T method due to differences of their respective principles.

3. Response speed: The proposed approach combined with the T method has the fastest response speed. The correlation method and M/T method have relatively slower response speeds. Experiments also shows that the T method is capable of detecting the variable speed.

Indeed, having the square wave related to the rotational period, the M/T method can be adopted for constant speed measurement or a mean value of rotational speed during a certain time and the T method can be employed for dynamic measurement of variable rotational speed. In actual programming, the M/T method and T method can be written into one piece of a DSP or FPGA simultaneously. An FPGA is more recommended to deal with the square wave for its property of parallel processing and high code execution efficiency.

There are several factors limit the application of this method working at a low speed. The amount of charge on the strip is unstable and the response time is poor at low speed. Further studies can be conducted to deal with these issues by adopting an electret material, adding adaptive numbers of strips and electrodes, and improving circuit properties.

Author Contributions: This work has been done in collaboration between all authors. Data curation, L.L.; Funding acquisition, Y.Q.; Investigation, H.H.; Methodology, L.L. and H.H.; Validation, K.T.; Writing—original draft, L.L.; Writing—review & editing, K.T.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 51777151, the National Key R and D Program of China, grant number 2016YFB0901200, the Shaanxi Provincial Key Technologies R and D Programme, grant number 2016GY-001 and the Open Research Fund of State Key Laboratory of Rail Traffic Control and Safety, grant number RCS2017K006.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yeh, W.H.; Bletscher, W.; Mansuripur, M. High resolution optical shaft encoder for motor speed control based on an optical disk pick-up. *Rev. Sci. Instrum.* **1998**, *69*, 3068–3071. [[CrossRef](#)]
2. Lygouras, J.N.; Lalakos, K.A.; Ysalides, P.G. High-performance position detection and velocity adaptive measurement for closed-loop position control. *IEEE Trans. Instrum. Meas.* **1998**, *47*, 978–985. [[CrossRef](#)]
3. Lord, W.; Chatto, R.B. Alternatives to Analog DC Tachogenerators. *IEEE Trans. Ind. Appl.* **1975**, *IA-11*, 470–478. [[CrossRef](#)]
4. Giebeler, C.; Adelerhof, D.J.; Kuiper, A.E.T.; Van Zon, J.B.A.; Oelgeschläger, D.; Schulz, G. Robust GMR sensors for angle detection and rotation speed sensing. *Sens. Actuators A Phys.* **2001**, *91*, 16–20. [[CrossRef](#)]
5. Shi, Z.; Huang, Q.; Wu, G.S.; Xu, Y.H.; Yang, M.; Liu, X.J.; Wang, C.; Ma, J. Design and Development of a Tachometer Using Magnetolectric Composite as Magnetic Field Sensor. *IEEE Trans. Magn.* **2018**, *54*, 4000604. [[CrossRef](#)]
6. Robinson, C.E. Analog Tachometers. *IEEE Trans. Ind. Gen. Appl.* **1966**, *IGA-2*, 144–146. [[CrossRef](#)]
7. Prokin, M. Extremely wide-range speed measurement using a double-buffered method. *IEEE Trans. Ind. Electron.* **1994**, *41*, 550–559. [[CrossRef](#)]
8. Wang, L.; Yan, Y.; Hu, Y.; Qian, X. Rotational Speed Measurement through Electrostatic Sensing and Correlation Signal Processing. *IEEE Trans. Instrum. Meas.* **2014**, *63*, 1190–1199. [[CrossRef](#)]
9. Wang, L.J.; Yan, Y.; Reda, K. Comparison of single and double electrostatic sensors for rotational speed measurement. *Sens. Actuators A Phys.* **2017**, *266*, 46–55. [[CrossRef](#)]
10. Lowell, J.; Rose-Innes, A.C. Contact electrification. *Adv. Phys.* **1980**, *29*, 947–1023. [[CrossRef](#)]
11. Li, J.; Wu, G.; Xu, Z. Tribo-charging properties of waste plastic granules in process of tribo-electrostatic separation. *Waste Manag.* **2015**, *35*, 36–41. [[CrossRef](#)] [[PubMed](#)]
12. Hu, Y.H.; Zhang, S.; Yan, Y.; Wang, L.J.; Qian, X.; Yang, L. A Smart Electrostatic Sensor for Online Condition Monitoring of Power Transmission Belts. *IEEE Trans. Ind. Electron.* **2017**, *64*, 7313–7322. [[CrossRef](#)]
13. Wang, L.J.; Yan, Y.; Hu, Y.H.; Qian, X.C. Rotational Speed Measurement Using Electrostatic Sensors and Correlation Signal Processing Techniques. In Proceedings of the 2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Minneapolis, MN, USA, 6–9 May 2013; pp. 224–227.

14. Lin, L.; Xiaoxin, W.; Hongli, H.; Xiao, L. Use of double correlation techniques for the improvement of rotation speed measurement based on electrostatic sensors. *Meas. Sci. Technol.* **2016**, *27*, 025004.
15. Ohmae, T.; Matsuda, T.; Kamiyama, K.; Tachikawa, M. A Microprocessor-Controlled High-Accuracy Wide-Range Speed Regulator for Motor Drives. *IEEE Trans. Ind. Electron.* **1982**, *IE-29*, 207–211. [[CrossRef](#)]
16. Hachiya, K.; Ohmae, T. Digital speed control system for a motor using two speed detection methods of an incremental encoder. In Proceedings of the 2007 European Conference on Power Electronics and Applications, Aalborg, Denmark, 2–5 September 2007; pp. 1–10.
17. Hace, A. The Improved Division-Less MT-Type Velocity Estimation Algorithm for Low-Cost FPGAs. *Electronics* **2019**, *8*, 361. [[CrossRef](#)]
18. Wang, L.J.; Yan, Y. Mathematical modelling and experimental validation of electrostatic sensors for rotational speed measurement. *Meas. Sci. Technol.* **2014**, *25*, 115101. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

Multi-Factor Operating Condition Recognition Using 1D Convolutional Long Short-Term Network

Zhinong Jiang ^{1,2}, Yuehua Lai ¹, Jinjie Zhang ^{1,2}, Haipeng Zhao ¹ and Zhiwei Mao ^{1,*}

¹ Key Lab of Engine Health Monitoring-Control and Networking of Ministry of Education, Beijing University of Chemical Technology, Beijing 100029, China; jiangzn@mail.buct.edu.cn (Z.J.); lyhlaibit@163.com (Y.L.); zhangjinjie@mail.buct.edu.cn (J.Z.); 18810272291@163.com (H.Z.)

² Beijing Key Laboratory of High-End Mechanical Equipment Health Monitoring and Self-Recovery, Beijing University of Chemical Technology, Beijing 100029, China

* Correspondence: maozhiwei@mail.buct.edu.cn

Received: 18 October 2019; Accepted: 10 December 2019; Published: 12 December 2019

Abstract: For a diesel engine, operating conditions have extreme importance in fault detection and diagnosis. Limited to various special circumstances, the multi-factor operating conditions of a diesel engine are difficult to measure, and the demand of automatic condition recognition based on vibration signals is urgent. In this paper, multi-factor operating condition recognition using a one-dimensional (1D) convolutional long short-term network (1D-CLSTM) is proposed. Firstly, a deep neural network framework is proposed based on a 1D convolutional neural network (CNN) and long short-term network (LSTM). According to the characteristics of vibration signals of a diesel engine, batch normalization is introduced to regulate the input of each convolutional layer by fixing the mean value and variance. Subsequently, adaptive dropout is proposed to improve the model sparsity and prevent overfitting in model training. Moreover, the vibration signals measured under 12 operating conditions were used to verify the performance of the trained 1D-CLSTM classifier. Lastly, the vibration signals measured from another kind of diesel engine were applied to verify the generalizability of the proposed approach. Experimental results show that the proposed method is an effective approach for multi-factor operating condition recognition. In addition, the adaptive dropout can achieve better training performance than the constant dropout ratio. Compared with some state-of-the-art methods, the trained 1D-CLSTM classifier can predict new data with higher generalization accuracy.

Keywords: diesel engine; condition recognition; CNN; LSTM; adaptive dropout

1. Introduction

A diesel engine is a kind of internal combustion engine that converts thermal energy into mechanical energy. It plays an important role in the field of national defense, in the chemical industry, in the marine industry, for nuclear power, and so on. Once a diesel engine fails, it not only causes economic losses directly or indirectly in terms of the shutdown of equipment, but it may also threaten the personal safety of users [1,2]. To enhance the availability of the diesel engine, it is imperative to monitor the engine condition and detect early faults. However, the detection of faults and the diagnosis of diesel engines [3] are not simple tasks due to the complex structure and fickle working conditions. If the operating conditions are not considered in detection and diagnosis activities, it is likely to lead to false alarms or missed detection [4,5]. With the information of operating conditions, the engineering applicability of a fault detection and diagnosis method [6–8] can be improved to avoid fatal performance degradation and huge economic losses at an early stage. Unfortunately, most fault detection methods are carried out under stable operating condition to avoiding variable

operating conditions. Therefore, condition recognition is an important and urgent task in practical engineering applications.

In a diesel engine, the flywheel is attached to the crankshaft, and they rotate together. They convert the reciprocating motion of the piston into the rotational motion of the crankshaft, which outputs torque for the driving of the car and other power-driven mechanisms. Therefore, the operating conditions of a diesel engine can be determined by two parameters: load and the rotation speed of the crankshaft. The load is the output torque of the engine through the flywheel. However, the multi-factor operating conditions of a diesel engine are difficult to measure in many situations, such as for the power systems of vehicles, propulsion devices of ships, and other dynamic equipment. Therefore, the demand for automatic recognition of multi-factor operating conditions is urgent.

During the operation of a diesel engine, the corresponding status information can be obtained by using vibration analysis [9], oil analysis [10], thermal performance analysis [11], and visual inspection. Vibration is an intrinsic mechanical phenomenon, and the vibration signals contain rich information about the diesel engine's status; thus, vibration monitoring is a powerful tool for condition recognition, as well as fault detection and diagnosis. In this paper, we aim at recognizing the multi-factor operating conditions of a diesel engine based on vibration signals.

Thanks to the development of computing calculation power and powerful signal processing techniques, the recognition tasks based on vibration signals made great progress. At present, some recognition algorithms based on vibration signals exist, and most of them focus on designing various handcrafted features, fusing multiple features and training different classifiers. In Reference [12], the Hilbert spectrum entropy, which combines the Hilbert spectrum and information entropy, was proposed for the pattern recognition of diesel engine working conditions. In Reference [13], the frequency domain features of vibration signals were extracted for back propagation (BP) and radial basis function (RBF) neural network training to recognize the cylinder pressure. In Reference [14], based on the cylinder head vibration signals measured under stable operating conditions, an engine cylinder pressure identification method using a genetic algorithm with BP neural network was proposed. In Reference [15], combustion evaluation parameters were extracted using time–frequency coherence analysis and the cylinder pressure could be estimated based on the parameters and an RBF neural network. In Reference [16], the measured signal was converted into a crank angle degree signal using the rotational speed monitored by magnetic pickup sensors. Then, a real-time engine load classification algorithm was proposed based on an artificial neural network.

Most pattern recognition studies focused mainly on single-factor conditions or recognition under stable operating conditions. For single-factor conditions, the number of categories is generally no greater than five. In practical engineering applications, a single factor cannot describe complex operating conditions, and this drawback results in ambiguous boundaries among different operating conditions. As for multi-factor operating conditions, as the number of operating conditions increases, so does the complexity of condition recognition. Simultaneously, as the vibration signals are random, transient, and cyclostationary, and as the corresponding feature extraction requires rich domain knowledge, it is difficult to extract sensitive characteristics of significant importance for multi-factor operating condition recognition.

Over the last few years, with the development of deep learning, many researchers exploited deep neural networks (DNNs) as the feature extractor and classifier [17,18]. Benefiting from the powerful feature extraction ability of neural network, especially convolution neural networks (CNNs) [19], these approaches and their variations exhibit good performance in the related tasks. In Reference [20], time domain and frequency domain feature representations were selected to form a vector to act as the input parameters of a CNN. The trained CNN classifier could diagnose the fault patterns of a gearbox with outstanding performance. In Reference [21], the vibration signals of rolling bearings were analyzed using continuous wavelet transform to get time–frequency representations in grayscale. Then, all compressed time–frequency representations were taken as the input for CNN training, and the trained CNN classifier could identify the faults of rolling bearings with strong generalization

ability. In Reference [22], a deep convolutional neural network of up to 38 layers, which could provide high classification accuracy, was proposed for gas classification. For CNN applications with vibration signals, there are different approaches to network input. In other words, the CNN is taken as a classifier, and the input of the CNN is mainly based on other feature extraction methods. At the same time, state-of-the-art CNN models have several parameters, which leads to problems related to storage, computation, and energy cost. In addition, recurrent neural networks (RNNs) and long short-term networks (LSTMs) [23,24] were validated in terms of their performance on one-dimensional (1D) signals. In Reference [25], a CNN and a fully connected neural network were both incorporated into a deep neural network framework to improve LSTM. The framework outperformed the original LSTM for the early diagnosis and prediction of sepsis shock. In Reference [26], an end-to-end model combining a CNN and RNN was proposed for the automatic detection of atrial fibrillation. Compared to the state-of-the-art models evaluated on standard benchmark electrocardiogram datasets, the proposed model produced better performance in detecting atrial fibrillation. The ideas in References [25,26] are very good references for multi-factor operating condition recognition based on vibration signals.

Therefore, a multi-factor operating condition recognition algorithm is proposed herein based on a 1D CNN and LSTM. In the proposed neural network framework, the 1D CNN was designed to extract local features of vibration signals through 1D convolution, and the LSTM was designed to describe the temporal relationship between local features. The contributions of this paper are summarized as follows:

1. A multi-factor operating condition recognition method is proposed using a 1D convolutional long short-term network (1D-CLSTM). As far as we know, this is the first study to combine a 1D CNN and LSTM to recognize operating conditions based on a time series of vibration signals;
2. Considering the particularity of engine vibration signals, batch normalization (BN) is introduced to regulate the input of some layers by fixing the mean value and variance of input signals in each convolutional layer;
3. Adaptive dropout is proposed for improving the model sparsity and preventing overfitting;
4. The designed 1D convolutional long short-term network (1D-CLSTM) classifier can achieve high generalization accuracy for recognizing multi-factor operating conditions.

The rest of this paper is organized as follows: Section 2 presents the test bench of a diesel engine and the experimental data acquisition. Section 3 introduces the technical background for the 1D CNN and LSTM. Section 4 describes the designed 1D-CLSTM and the flowchart of the multi-factor operating condition recognition algorithm. Section 5 shows the training performance of the designed 1D-CLSTM classifier, with generalizability verification, a performance comparison with different methods, and a training performance comparison with different dropout ratios. Finally, conclusions and future prospects are presented in Section 6.

2. Experiment and Vibration Signal

2.1. Test Bench of Diesel Engine

For data acquisition, a four-stroke diesel engine numbered TBD234 (produced by Henan Diesel Engine Industry Co. Ltd., Luoyang, China) was used and tested in different operating conditions. The parameters of the diesel engine are shown in Table 1.

As shown in Figure 1, 12 acceleration sensors were arranged on the surface of corresponding cylinder heads to monitor the status information of the diesel engine in the running state. The vibration signals formed the basis for the multi-factor operating condition recognition of the diesel engine. Moreover, an eddy current sensor was arranged on the flywheel to collect the information of rotating speed. In addition, a hydraulic dynamometer was connected with the output end of the diesel engine to adjust the load.

Table 1. Parameters of TBD234 diesel engine.

Item	Parameter
Number of cylinders	12
Shape	V-shaped 60°
Firing sequence	B1-A1-B5-A5-B3-A3-B6-A6-B2-A2-B4-A4
Rating speed	2100 rev/min
Rating power	485 kW

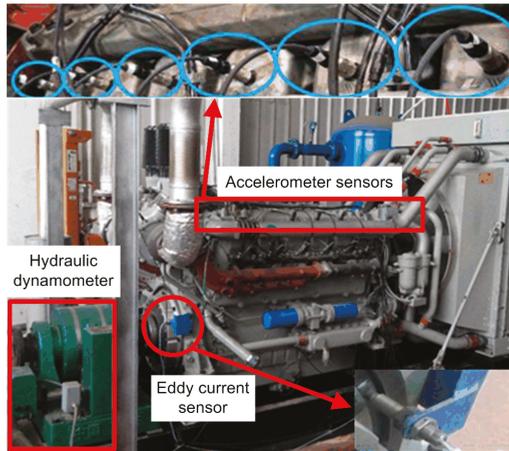


Figure 1. Test bench of the diesel engine.

All signals were measured using an online condition monitoring system (OCMS) at a sampling frequency of 51.2 kHz per channel in all tests, and the results were saved to a server through Ethernet transmission. The structure diagram of the OCMS of the diesel engine is shown in Figure 2.

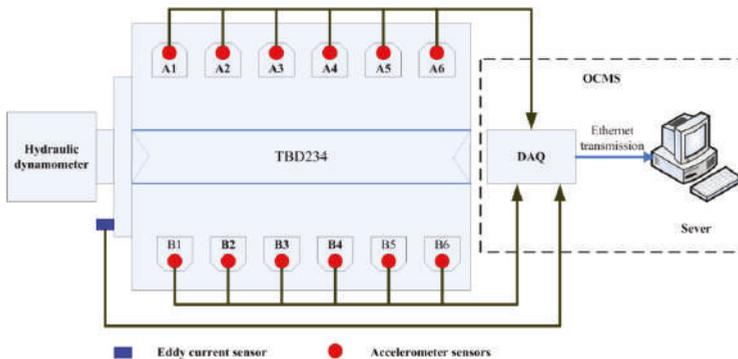


Figure 2. Structure diagram of the online condition monitoring system (OCMS) of the diesel engine.

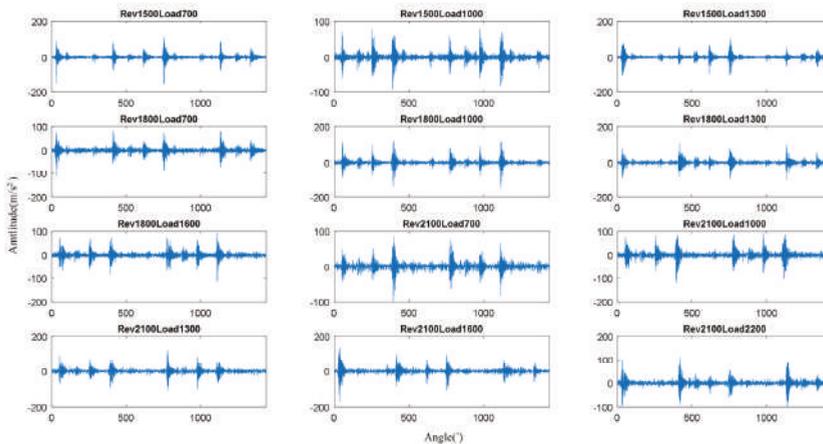
2.2. Experimental Data Acquisition

To extract vibration data under different operating conditions, the engine was run at different levels of operating conditions. The representative operating conditions are listed in Table 2.

Table 2. Operating conditions of the diesel engine.

No.	Rev (rpm)	Load (N·m)	No.	Rev (rpm)	Load (N·m)
1	1500	700	7	1800	1600
2	1500	1000	8	2100	700
3	1500	1300	9	2100	1000
4	1800	700	10	2100	1300
5	1800	1000	11	2100	1600
6	1800	1300	12	2100	2200

Through the OCMS, vibration signals of different operating conditions could be measured. The vibration signals of 12 different operating conditions are shown in Figure 3.

**Figure 3.** Vibration signals of 12 different operating conditions.

The signals in Figure 3 represent two complete periodic vibration signals, with a certain cyclic fluctuation in the angular domain. When fire combustion and closing of the intake valve and exhaust valve occur, an obvious excitation response is produced in the corresponding phase. Due to the different ignition phase points of different cylinders, the corresponding combustion excitation occurs at different positions. As the amplitude of the vibration signal features large randomness, the vibration signal of a diesel engine can be considered a non-periodic and non-stationary signal. This characteristic of the vibration signal greatly increases the difficulty of multi-factor operating condition recognition.

3. Technical Background

In this study, a deep neural network framework is proposed based on a 1D CNN and LSTM for multi-factor operating condition recognition. For the vibration signal in the form of a time series, a 1D CNN was adopted to extract local features of vibration signals through a 1D convolution kernel. Then, an LSTM was adopted to describe the temporal relationship between local features through a memory unit and gate mechanism. In this way, the combination of the 1D CNN and LSTM could perform well for the analysis of vibration signals.

3.1. 1D CNN

A typical CNN [19] contains three types of network layers: a convolutional layer, pooling layer, and fully connected layer. Some excellent variants of CNN were proposed, such as LeNet-5 [27], AlexNet [19], and VGG-16 [28]. The image recognition ability of these CNN variants is outstanding,

and they achieved remarkable results. In CNNs, the receptive field, weight sharing, and pooling can greatly reduce the complexity of the network.

It was proven that a 1D CNN can be applied to the time series analysis of sensor data. In 1D CNNs, features can be extracted from segments through 1D convolution, which is a weighted sum operation between the weight matrix and the vibration data in each segment, with the addition of the overall bias. Every convolution extracts a feature from a local receptive field, and the window of the convolution kernel slides across the entire input sequence with a fixed step to achieve all features. The weight sharing exists to maintain the weights of the convolution kernel in the sliding process. As shown in Figure 4, the size of the i -th convolution kernel is shown, featuring weights (w_{i1}, w_{i2}, w_{i3}) in a 1×3 format, with the bias left out for clarity. The corresponding feature vector $F (f_{i1}, f_{i2}, f_{i3}, \dots, f_{i(n-2)})$ can be obtained from the input signal $X (x_1, x_2, x_3, \dots, x_n)$ with one step of the convolution kernel.

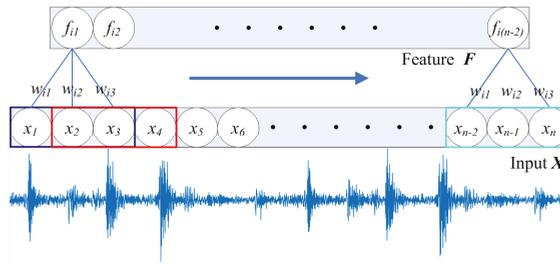


Figure 4. Temporal convolution.

Mathematically, this can be expressed as shown in Equation (1).

$$f_{ij} = \Phi \left(b_i + \sum_{k=1}^m w_{ik} \cdot x_{j+k-1} \right), \tag{1}$$

where m is the size of the convolution kernel, f_{ij} is the output of the j -th neuron of the i -th filter in the hidden layer, Φ is the activation function, and b_i is the overall bias of the i -th filter.

Convolution kernels of different sizes can extract features of different granularity [29]. Usually, the first convolutional layer may only extract some low-level features, and more complex features can be extracted from low-level features by stacking network layers.

As the pooling operation can maintain the variance of the translation, rotation, and scale, the pooling layer is set following each convolutional layer to retain the main features. Meanwhile, it can reduce the number of parameters to prevent overfitting and improve the generalizability of the model. In a pooling layer, the features obtained from the activation function are cut into several regions, and the maximum/average values can be taken as the new features to realize dimension reduction. By repeating operations as described above, features can be extracted continuously to improve the generalizability of the CNN.

Enough sensitive important features can be extracted by alternating convolutional and pooling layers, and the fully connected layers can map the distributed feature representation to the sample markup space. Finally, the output layer with a softmax activation function is used for classification.

3.2. LSTM

A recurrent neural network (RNN) is a kind of neural network which can be used for sequential data analysis, while the LSTM is a specific kind of RNN. Compared with a traditional RNN, a memory cell and gating mechanism are introduced to deal with the existence of gradient disappearance and gradient explosion during the training of long sequences. The gating mechanism can be used to

control the transfer state, which is designed to remember the important information and forget the unimportant information. The memory cell of an LSTM is shown in Figure 5.

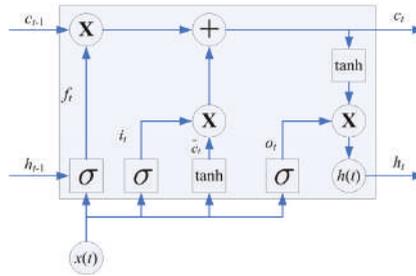


Figure 5. Memory cell of a long short-term memory network (LSTM).

As shown in Figure 5, the memory cell of an LSTM is made up of an input gate, output gate, and forget gate. The sigmoid activation function is used in the forget gate to control the weight of information that can be passed, whereas the tanh activation function is used in the input gate to deal with the input at the current sequence position, and the sigmoid activation function is used in the output gate to update the output based on the results of the input gate and forget gate. Mathematically, the parameters of the LSTM can be updated as shown in Equation (2).

$$\begin{aligned}
 i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i); \\
 f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f); \\
 o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o); \\
 \tilde{c}_t &= \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c); \\
 c_t &= f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t; \\
 h_t &= o_t \cdot \tanh(c_t).
 \end{aligned} \tag{2}$$

where x_t is the input of a sequence, c_{t-1} is the last state, and h_{t-1} is the output of the last memory cell. The state c_t and output h_t of the current memory cell can be obtained after parameter update calculation.

4. Methodologies

In this section, the 1D-CLSTM is firstly constructed for multi-factor operating condition recognition, and then adaptive dropout is proposed. Moreover, the flowchart of the multi-factor operating condition recognition method is introduced.

4.1. 1D Convolutional Long Short-Term Network

4.1.1. Overall Architecture

As described above, the features extracted by different neural networks have different characteristics. The 1D CNN can obtain the features of a receptive field through convolution, but the temporal relationship of the vibration signal is ignored as a result of the size of the convolution kernel. As for the LSTM, a temporal relationship can be described through the memory cell and gating mechanism. Therefore, the multi-factor operating condition recognition algorithm 1D-CLSTM is proposed based on a 1D CNN and LSTM. In the proposed neural network framework, the 1D CNN was designed to extract local features of vibration signals through 1D convolution, and the LSTM was designed to describe the temporal relationship between local features. The overall architecture of the 1D-CLSTM is shown in Figure 6.

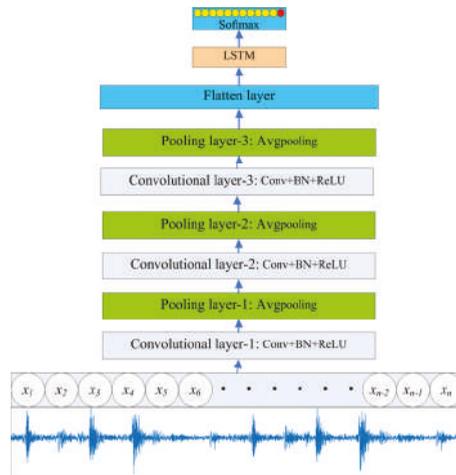


Figure 6. Overall architecture of the one-dimensional convolutional long short-term network (1D-CLSTM).

4.1.2. Architecture Design

According to the sampling frequency of the monitoring system and different operating conditions introduced in Section 2, a signal segment with a length of 4096 can be determined to contain all the information in a cycle. The crankshaft of a four-stroke diesel engine rotates 720 degrees to complete a cycle, which means complete energy conversion. Therefore, the minimum receptive field can be defined as a degree in the angular domain. Moreover, the size of the CNN filter in the first convolutional layer can be set to an odd number greater than $4096/720$.

Considering the particularity of a vibration signal, which is a non-periodic and non-stationary signal, BN [30] is vital for regulating the input of some layers by fixing the mean value and variance of input signals of each convolutional layer, through which the features can maintain the same distribution in the training process of the 1D-CLSTM. Upon increasing the number of layers in a neural network, the decreasing convergence rate often leads to gradient explosion or gradient disappearance, and BN is an excellent solution. Therefore, the convolution is followed by BN in each convolutional layer. In all convolutional layers, the rectified linear unit (ReLU) activation function is adopted, and BN occurs in front of the ReLU activation function. In other words, the results of BN are the input of the ReLU activation function. The ReLU activation function makes the output of some neurons equal to 0, which results in sparsity of the network, thereby reducing the interdependence of parameters and alleviating the occurrence of the overfitting problem. The average values of features obtained from the ReLU activation function are taken as the new features to realize dimension reduction in a pooling layer. The designed 1D-CLSTM begins with a sequence input, after which the features can be extracted by alternate convolutional layers and pooling layers.

A complete periodic signal contains different sequential excitation responses; thus, the sequence length processed by the LSTM can be determined according to the degree of excitation responses in the angular domain. When the degree of an excitation response in the angular domain is 15, the number of LSTM units can be chosen to be greater than $720/15$. Following the final pooling layer, there is a flattening layer to reshape the tensor as the input of the LSTM with 73 units. In order to accelerate the convergence process of 1D-CLSTM training, adaptive dropout is applied. Finally, the output layer with a softmax activation function is used for multi-class classification. The structural parameters of the 1D-CLSTM are shown in Table 3.

Table 3. Structural parameters of the one-dimensional convolutional long short-term network (1D-CLSTM).

No.	Network Layer	Size of Convolution Kernel	Stride	Output Dimension
1	Input layer	-	-	4096 × 1
2	Convolutional layer-1	11	1	4096 × 32
3	Pooling layer-1	3	2	2047 × 32
4	Convolutional layer-2	13	1	2047 × 64
5	Pooling layer-2	3	2	1023 × 64
6	Convolutional layer-3	15	1	1023 × 128
7	Pooling layer-3	3	2	511 × 128
8	Flatten layer	-	-	73 × 896
9	LSTM (two layers)	-	-	73
10	Softmax	-	-	12

4.1.3. Adaptive Dropout

Dropout is widely used for improving model sparsity and preventing overfitting in model training. The learning process of the 1D-CLSTM for multi-factor operating condition recognition is an iterative one. On account of the mutual influence among interconnected neurons, every iteration is a greedy search, whereby we find the best connections. That is, a connection may be unimportant due to the existence of some others, but it becomes important once the others are removed. Therefore, the adaptive dropout ratio is proposed to deal with this problem.

The most popular Bernoulli dropout technique [31] can be applied to neurons or weights. Assuming the input of a weight or neuron as X , the output as Y , the dropout probability as $P(\alpha)$, and the weight matrix as W , each neuron is probabilistically dropped at each training step, as defined in Equation (3).

$$Y = (X \cdot P)W. \quad (3)$$

Each weight in the weight matrices is probabilistically dropped at each training step, as defined in Equation (4).

$$Y = X(W \cdot P). \quad (4)$$

Usually, the dropout ratio α is constant for generating random network structures (for example, 0.5). However, the model capacity is constantly changing within the 1D-CLSTM training. Therefore, the dropout ratio needs to be adaptive to the current network. Neurons or weights are dropped temporarily during training and dropped forever after pruning to solidify the network structure. Compared with the original network structure, the parameters of the current network become sparse after pruning, and the dropout ratio should be reduced.

Assuming that the connection between the input layer and output layer is fully connected, the number of connections can be calculated as shown in Equation (5).

$$C_i = N_i N_o. \quad (5)$$

Since dropout works on neurons, taking C_{io} as the original network and C_{ic} as the current network, the dropout ratio α can be adjusted according to Equation (6).

$$\alpha_c = \frac{\alpha_o N_o}{(N_o + 1)} \sqrt{\frac{C_{ic}}{C_{io}}}, \quad (6)$$

where α_c represents the dropout rate of the current network, and α_o represents the dropout rate of the original network.

4.1.4. Implementation

The loss function, which measures the degree of difference between the predicted value and actual value, is a non-negative real value function. A smaller loss function denotes better robustness of the model. Cross-entropy is frequently used for loss calculation in neural network training, as shown in Equation (7).

$$\text{loss} = -\sum_{i=1}^n y_i \log(y_{i-}), \quad (7)$$

where y_i represents the predicted value, y_{i-} represents the actual output, and n is the number of training samples.

In the training of the 1D-CLSTM designed for multi-factor operating condition recognition, the learning rate was set to 0.001. Through iterative calculation, the loss of 1D-CLSTM decreased continuously and eventually became stable. Then, the weight of 1D-CLSTM was fixed, allowing the 1D-CLSTM classifier to be used for multi-factor operating condition recognition.

To make the training of the 1D-CLSTM model more efficient and achieve better performance, the training techniques described below were introduced.

Mini-batch gradient descent. Considering the huge calculation in network training, a batch sample was adopted in the training process, and the batch size was set to 128. The batch sample strategy uses less memory and achieves a faster training speed than full batch learning. Compared with stochastic gradient descent, mini-batch gradient descent is more efficient. Compared with batch gradient descent, mini-batch gradient descent can achieve robust convergence to avoid local optimization. Therefore, mini-batch gradient descent was taken as the optimizer to minimize the loss and adjust the weights in the designed 1D-CLSTM.

Early termination. In the process of model training with the training set, the performance of the model is also evaluated with the validation set. The validation error decreases in the beginning as the training error decreases. After a certain number of training steps, the training error still decreases, but the validation error no longer decreases. Therefore, early termination can act as a regulator and effectively avoid overfitting of the model. Once the validation error stops decreasing, the early termination of model training can be enforced in the training of the 1D-CLSTM.

4.2. Multi-Factor Operating Condition Recognition

To determine the multi-factor operating condition information of a diesel engine, a condition recognition method using 1D-CLSTM is proposed. Firstly, acceleration sensors were used to monitor the status information of a diesel engine under different operating conditions. Considering the characteristics of the vibration signal, some performance improvement techniques were adopted in the 1D-CLSTM, such as BN, ReLU activation function, adaptive dropout. Moreover, mini-batch gradient descent and early termination were adopted in the training of 1D-CLSTM to achieve a fast training speed and avoid overfitting of the model. Accordingly, the 1D-CLSTM could be trained using supervised learning. After training, the trained 1D-CLSTM classifier could be used for the classification of multi-factor operating conditions. The flowchart of the multi-factor operating condition recognition method is shown in Figure 7.

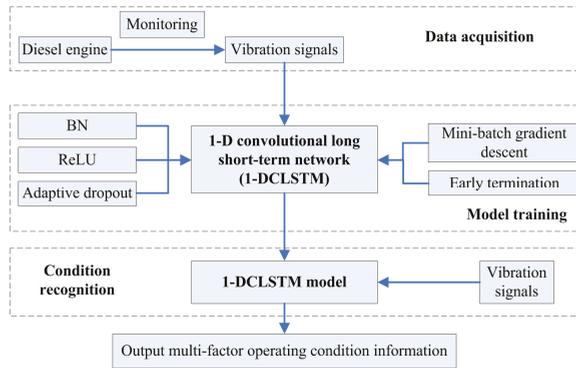


Figure 7. The flowchart of the condition recognition method.

5. Experiments

According to the flowchart shown in Figure 7, the training performance of the designed 1D-CLSTM is presented below. After training, the performance of 1D-CLSTM using vibration signals for multi-factor operating condition recognition was evaluated. Moreover, the vibration signals measured from another kind of diesel engine were applied to verify the generalizability of the proposed approach. Finally, the results of the proposed approach for multi-factor operating condition recognition were compared to other classification algorithms to verify that the designed 1D-CLSTM with strong generalizability could provide higher classification accuracy. The 1D-CLSTM model was written using Python 3.6 with TensorFlow and run on Window 10 with an NVIDIA Quadro P6000.

5.1. Training Performance of the Designed 1D-CLSTM

The vibration signals were in the form of a time series, used as the input data for training the designed network, with a total of 7200 samples. The whole dataset was randomly divided into two sets: 80% for training and 20% for validation. In other words, the training set had 5760 samples, and the validation set had 1440 samples. With the continuous iterative training of 1D-CLSTM, the losses of the training set and validation set decreased as the number of epochs increased, as depicted in Figure 8. On the contrary, the accuracies of the training set and validation set continuously improved, as depicted in Figure 9. According to the early termination, the model training stopped when the loss of the validation set stopped decreasing. The training of 1D-CLSTM stopped at the 63rd epoch when the cross-entropy of the validation set was 0.01913 and the accuracy of the training set was 0.9953. Therefore, the corresponding 1D-CLSTM classifier is a desired classification model for multi-factor operating condition recognition.

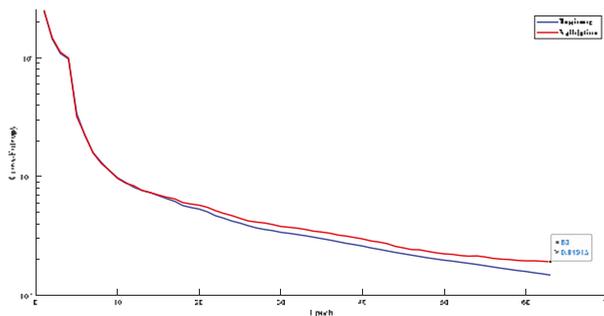


Figure 8. Losses of training set and validation set.

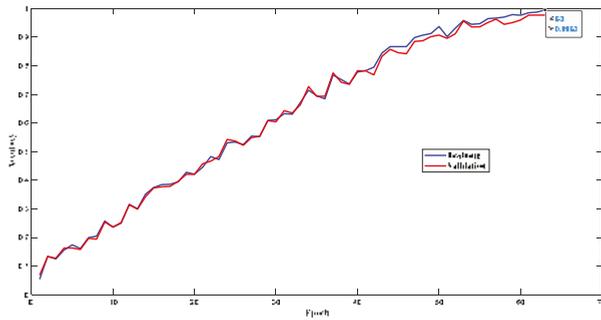


Figure 9. Accuracies of training set and validation set.

A confusion matrix, which contains information about actual and predicted classes, was used to describe the generalizability of the 1D-CLSTM classifier [32]. The testing set had a total of 1200 samples, with 100 samples for each operating condition. The confusion matrix for the testing set is shown in Figure 10.

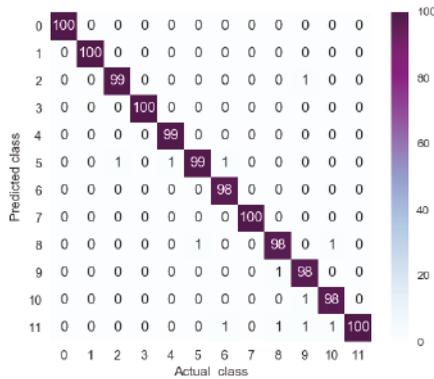


Figure 10. Confusion matrix for testing set.

The elements in row i and column j of the confusion matrix represent the number of times the j class was identified as the i class. Therefore, only the diagonal elements denote correct recognition. It can be seen from Figure 10 that only 11 samples out of 1200 were misclassified. Therefore, the designed 1D-CLSTM can classify multi-factor operating conditions with an accuracy of 99.08%.

5.2. Comparison of Training Performance with Different Dropout Ratios

The convergence process in model training is an important factor for achieving a classifier with excellent performance. Dropout serves as an effective approach to improve the model sparsity and prevent overfitting in model training. To find the best connections in the designed 1D-CLSTM, a suitable dropout ratio was very important. Adaptive dropout, due to its flexibility depending on network capacity, is able to maintain the balance between model performance and model sparsity. To check the effect of adaptive dropout, training accuracy curves of different dropout ratios were plotted, as shown in Figure 11. According to the early termination, the model training using adaptive dropout stopped at the 63rd epoch, and the comparison of training performance with different dropout ratios was conducted within 63 epochs.

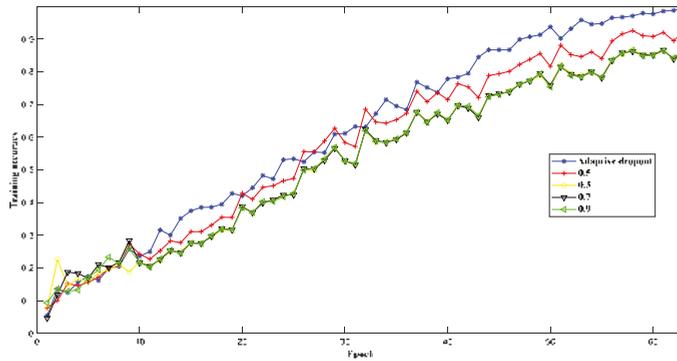


Figure 11. Training accuracy curves with different dropout ratios.

It can be seen from Figure 11 that the training performance using adaptive dropout was best; thus, adaptive dropout can improve the training performance to achieve the desired model.

5.3. Comparison Analysis

To validate the performance of the designed 1D-CLSTM, the proposed method was compared with the following baseline methods:

1. The k-nearest neighbor (kNN) algorithm, which works with a multi-domain feature set [33]. Based on the multi-domain feature set, the kNN algorithm is more suitable than other statistical learning methods.
2. The support vector machine (SVM), which works with a multi-domain feature set. SVM is a kind of generalized linear classifier that can be used for supervised learning.
3. The 1D LeNet-5, which is a convolutional network that has the same network layers as LeNet-5, i.e., two convolutional layers and two fully connected layers. The corresponding structural parameters are listed in Table 4.
4. The 1D AlexNet, which is a convolutional network that has the same network layers as AlexNet, i.e., five convolutional layers and three fully connected layers. The corresponding structural parameters are also listed in Table 4.
5. The 1D VGG-16, which is a convolutional network that has the same network layers as VGG-16, with 1D convolution kernels adopted. The corresponding structural parameters are also listed in Table 4.
6. A traditional LSTM, which has two layers and 32 LSTM units in each layer.

In Table 4, *s* represents the stride, and the convolution is followed by BN in each convolutional layer.

Table 4. Structural parameters of the 1D-CLSTM.

1D LeNet-5	1D AlexNet	1D VGG-16	
Conv1 [1,11] × 64, s = 1	Conv1 [1,11] × 32, s = 1	Conv1 [1,3] × 16, s = 1	Conv9 [1,3] × 128, s = 1
AveragePooling1 [1,3], s = 2	MaxPooling1 [1,3], s = 2	Conv2 [1,3] × 16, s = 1	Conv10 [1,3] × 128, s = 1
Conv2 [1,13] × 128, s = 1	Conv2 [1,5] × 64, s = 1	MaxPooling1 [1,2], s = 2	MaxPooling4 [1,2], s = 2
AveragePooling2 [1,3], s = 2	MaxPooling2 [1,3], s = 2	Conv3 [1,3] × 32, s = 1	Conv11 [1,3] × 256, s = 1
FC1 (1024)	Conv3 [1,3] × 128, s = 1	Conv4 [1,3] × 32, s = 1	Conv12 [1,3] × 256, s = 1
FC2 (512)	Conv4 [1,3] × 128, s = 1	MaxPooling2 [1,2], s = 2	Conv13 [1,3] × 256, s = 1
softmax	Conv5 [1,3] × 128, s = 1	Conv5 [1,3] × 64, s = 1	MaxPooling5 [1,2], s = 2
-	MaxPooling3 [1,3], s = 2	Conv6 [1,3] × 64, s = 1	FC1 (1024)
-	FC1 (1024)	Conv7 [1,3] × 64, s = 1	FC2 (512)
-	FC2 (512)	MaxPooling3 [1,2], s = 2	softmax
-	softmax	Conv8 [1,3] × 128, s = 1	-

For multi-factor operating condition recognition, the class domains of operating conditions are likely to overlap with each other. Our goal was to develop a multi-factor operating condition recognition method that can achieve high generalization accuracy. Therefore, the same vibration data were used for the training and testing with the above methods, and the corresponding model performance is shown in Table 5.

Table 5. Performance comparison. SVM—support vector machine.

Learning Model	Generalization Accuracy (%)
1D-CLSTM	99.08
LSTM	74.12
kNN with a multi-domain feature set	92.18
SVM with a multi-domain feature set	94.91
1D LeNet-5	94.43
1D AlexNet	97.54
1D VGG-16	98.01

It can be seen from Table 5 that the generalization accuracy of the proposed method was the best. This shows that the 1D-CLSTM learns to predict new data with higher accuracy than other machine learning models and avoids overfitting. In addition, the trained 1D-CLSTM classifier can be used as a good initializer for similar tasks of transfer learning (https://github.com/Larrylyh/Condition_Recognition).

5.4. Generalizability Verification

To verify the generalizability of the proposed approach, the designed 1D-CLSTM was applied to a diesel engine with 20 cylinders (V20DE), which is shown in Figure 12.



Figure 12. The diesel engine with 20 cylinders.

The vibration data under different operating conditions, which are listed in Table 6, were measured.

Table 6. Operating conditions of V20DE.

No.	Rev (rpm)	Load (kN·m)
1	600	0
2	1100	17.7
3	1500	22.6
4	1500	26.6
5	1500	28.3

Generally, the data measured from different engine types vary greatly, and the 1D-CLSTM classifier would need to be trained before use. The test set of V20DE contained 2101 samples, and the corresponding confusion matrix is illustrated in Figure 13. As depicted in Figure 13, 32 samples out of 2101 were misclassified, and the corresponding accuracy was 98.48%.

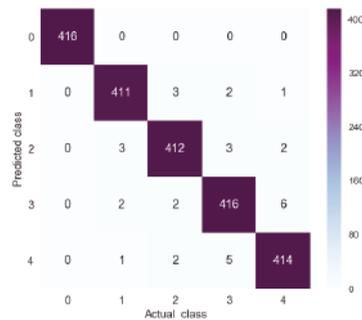


Figure 13. Confusion matrix.

6. Conclusions

In this study, an effective approach was proposed for multi-factor operating condition recognition using a 1D convolutional long short-term network. The proposed method was capable of monitoring and automatically recognizing multi-factor operating conditions based on the vibration signal measured on engine cylinder heads. Moreover, the measured vibration signals no longer needed a complex feature extraction process for condition recognition. Subsequently, adaptive dropout was proposed for improving the model sparsity and preventing overfitting in model training. The experimental results proved that the designed 1D-CLSTM classifier is indeed ideal for multi-factor operating condition recognition with high generalization accuracy. At the same time, adaptive dropout could achieve better training performance than a constant dropout ratio. In addition, this method has the potential for application in real-time scenarios because the implementation of the 1D-CLSTM classifier is simple. Last but not least, the trained 1D-CLSTM classifier can be used as a good initializer for similar tasks of transfer learning. In the future, new studies will be conducted on the transition period between the defined operating conditions to obtain a model that can identify continuous operating conditions. Moreover, continuous operating condition recognition can be the basis of fault detection or diagnosis under variable operating conditions.

Author Contributions: Data curation, Z.M.; funding acquisition, J.Z.; project administration, Z.J. and J.Z.; supervision, Z.J.; writing—original draft, Y.L.; writing—review and editing, Y.L., H.Z., and Z.M.

Funding: This work was supported by the National Key Research and Development Plan of China (Grant No. 2016YFF0203305), the Fundamental Research Funds for the Central Universities of China (Grant No. JD1815), and the Double First-Rate Construction Special Funds (Grant No. ZD1601).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Kouremenos, D.A.; Hountalas, D.T. Diagnosis and condition monitoring of medium-speed marine diesel engines. *Lubr. Sci.* **2010**, *4*, 63–91. [[CrossRef](#)]
- Zhiwei, M. Research on Typical Fault Diagnosis and Unstable Condition Monitoring and Evaluation for Piston Engine. Ph.D. Thesis, Beijing University of Chemical Technology, Beijing, China, 2017.
- Porteiro, J.; Collazo, J.; Patiño, D.; Míguez, J.L. Diesel engine condition monitoring using a multi-net neural network system with nonintrusive sensors. *Appl. Therm. Eng.* **2011**, *31*, 4097–4105. [[CrossRef](#)]
- Xu, H.F. New Intelligent Condition Monitoring and Fault Diagnosis System for Diesel Engines Using Embedded System. *Appl. Mech. Mater.* **2012**, *235*, 408–412. [[CrossRef](#)]
- Xu, X.; Yan, X.; Sheng, C.; Yuan, C.; Xu, D.; Yang, J. A Belief Rule-Based Expert System for Fault Diagnosis of Marine Diesel Engines. *IEEE Trans. Syst. Man Cybern. Syst.* **2017**, *99*, 1–17. [[CrossRef](#)]
- Kowalski, J.; Krawczyk, B.; Woźniak, M. Fault diagnosis of marine 4-stroke diesel engines using a one-vs-one extreme learning ensemble. *Eng. Appl. Artif. Intell.* **2017**, *57*, 134–141. [[CrossRef](#)]

7. Shen, H.; Zeng, R.; Yang, W.; Zhou, B.; Ma, W.; Zhang, L.; Projecton Equipment Surport Department, Army Military Transportation University. Diesel Engine Fault Diagnosis Based on Polar Coordinate Enhancement of Time-Frequency Diagram. *J. Vib. Meas. Diagn.* **2018**, *38*, 27–33.
8. Li, Y.; Han, M.; Han, B.; Le, X.; Kanae, S. Fault Diagnosis Method of Diesel Engine Based on Improved Structure Preserving and K-NN Algorithm. In *Advances in Neural Networks-ISNN 2018*; Springer: Cham, Switzerland, 2018.
9. Liu, Y. Research on Fault Diagnosis for Fule System and Valve Train of Diesel Engine Based on Vibration Analysis. Ph.D. Thesis, Tianjin University, Tianjin, China, 2016.
10. Wu, T.; Wu, H.; Du, Y.; Peng, Z. Progress and trend of sensor technology for on-line oil monitoring. *Sci. China Technol. Sci.* **2013**, *56*, 2914–2926. [[CrossRef](#)]
11. Tang, G.; Fu, X.; Shao, G.; Chen, N. Application of Improved Grey Model in Prediction of Thermal Parameters for Diesel Engine. *Ship Boat* **2018**, *5*, 39–46.
12. Li, H.; Zhou, P.; Ma, X. Pattern recognition on diesel engine working condition by using a novel methodology-Hilbert spectrum entropy. *J. Mar. Eng. Technol.* **2005**, *4*, 43–48. [[CrossRef](#)]
13. Ji, S.; Cheng, Y.; Wang, X. Cylinder pressure recognition based on frequency characteristic of vibration signal measured from cylinder head. *J. Vib. Shock* **2008**, *27*, 133–136.
14. Liu, J.; Li, H.; Qiao, X.; Li, X.; Shi, Y. Engine Cylinder Pressure Identification Method Based on Cylinder Head Vibration Signals. *Chin. Intern. Comhuslion Engine Eng.* **2013**, *34*, 32–37.
15. Chang, C.; Jia, J.; Zeng, R.; Mei, J.; Wang, G. Recognition of Cylinder Pressure Based on Time-frequency Coherence and RBF Network. *Veh. Engine* **2016**, *5*, 87–92.
16. Syed, M.S.; Sunghoon, K.; Sungoh, K. Real-Time Classification of Diesel Marine Engine Loads Using Machine Learning. *Sensors* **2019**, *19*, 3172. [[CrossRef](#)]
17. Yoo, Y.; Baek, J.G. A Novel Image Feature for the Remaining Useful Lifetime Prediction of Bearings Based on Continuous Wavelet Transform and Convolutional Neural Network. *Appl. Sci.* **2018**, *8*, 1102. [[CrossRef](#)]
18. Rui, Z.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R.X. Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* **2019**, *115*, 213–237.
19. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012.
20. Zhiqiang, C.; Chuan, L.; Sanchez, R.V. Gearbox Fault Identification and Classification with Convolutional Neural Networks. *Shock Vib.* **2015**, *2015*, 1–10.
21. Yuan, J.; Han, T.; Tang, J.; An, L. An Approach to Intelligent Fault Diagnosis of Rolling Bearing Using Wavelet Time-Frequency Representations and CNN. *Mach. Des. Res.* **2017**, *2*, 101–105.
22. Peng, P.; Zhao, X.; Pan, X.; Ye, W. Gas Classification Using Deep Convolutional Neural Networks. *Sensors* **2018**, *18*, 157. [[CrossRef](#)]
23. Wu, Y.; Yuan, M.; Dong, S.; Li, L.; Liu, Y. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing* **2018**, *275*, 167–179. [[CrossRef](#)]
24. Qing, X.; Niu, Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy* **2018**, *148*, 461–468. [[CrossRef](#)]
25. Lin, C.; Yuan, Z.; Julie, I.; Muge, C.; Ryan, A.; Jeanne, M.H.; Jeanne, M.H.; Min, C. Early diagnosis and prediction of sepsis shock by combining static and dynamic information using convolutional-LSTM. In Proceedings of the 2018 IEEE International Conference on Healthcare Informatics (ICHI), New York, NY, USA, 4–7 June 2018; pp. 219–228.
26. Andersen, R.S.; Abdolrahman, P.; Sadasivan, P. A deep learning approach for real-time detection of atrial fibrillation. *Expert Syst. Appl.* **2019**, *115*, 465–473. [[CrossRef](#)]
27. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
28. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556v6.
29. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlen, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

30. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
31. Nitish, S.; Geoffrey, H.; Alex, K.; Ilya, S.; Ruslan, S. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
32. Visa, S.; Ramsay, B.; Ralescu, A.L.; Van Der Knaap, E. *Confusion Matrix-Based Feature Selection*; MAICS: Metro Manila, Philippines, 2011; pp. 120–127.
33. Yan, X.; Jia, M. A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing. *Neurocomputing* **2018**, *313*, 47–64. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

Comprehensive Improvement of the Sensitivity and Detectability of a Large-Aperture Electromagnetic Wear Particle Detector

Ran Jia ¹, Biao Ma ¹, Changsong Zheng ^{1,*}, Xin Ba ^{1,2}, Liyong Wang ³, Qiu Du ¹ and Kai Wang ¹

¹ School of Mechanical Engineering, Beijing Institute of Technology, Zhongguancun South Street No.5 Haidian District, Beijing 100081, China

² Faculty of Engineering and Information, University of Technology Sydney, Ultimo, NSW 2007, Australia

³ The Ministry of Education Key Laboratory of Modern Measurement and Control Technology, Beijing Information Science and Technology University, Xiaoying East Street No.12, Beijing 100192, China

* Correspondence: zhengchangsong@bit.edu.cn; Tel.: +86-010-6891-8636

Received: 20 May 2019; Accepted: 16 July 2019; Published: 18 July 2019

Abstract: The electromagnetic wear particle detector has been widely studied due to its prospective applications in various fields. In order to meet the requirements of the high-precision wear particle detector, a comprehensive method of improving the sensitivity and detectability of the sensor is proposed. Based on the nature of the sensor, parallel resonant exciting coils are used to increase the impedance change of the exciting circuit caused by particles, and the serial resonant topology structure and an amorphous core are applied to the inductive coil, which improves the magnetic flux change of the inductive coil and enlarges the induced electromotive force of the sensor. Moreover, the influences of the resonance frequency on the sensitivity and effective particle detection range of the sensor are studied, which forms the basis for optimizing the frequency of the magnetic field within the sensor. For further improving the detectability of micro-particles and the real-time monitoring ability of the sensor, a simple and quick extraction method for the particle signal, based on a modified lock-in amplifier and empirical mode decomposition and reverse reconstruction (EMD-RRC), is proposed, which can effectively extract the particle signal from the raw signal with low signal-to-noise ratio (SNR). The simulation and experimental results show that the proposed methods improve the sensitivity of the sensor by more than six times.

Keywords: particle detection; sensitivity; resonance; amorphous core; signal extraction

1. Introduction

Wear is one of the major causes of failure in machine components. The excessive wear of some core parts of machineries, especially for large-scale mechanical equipment, may lead to a poor mechanical performance, which in turn causes enormous economic losses. Therefore, for online monitoring of the wear condition of machineries in order to prevent serious malfunctions, the wear particle detector has demonstrated its value [1–3]. To date, wear particle detectors with different physical principles, including optics, ultrasonics, electronics, and imaging, have been proposed, and the characteristics of the various kinds of sensors are listed in Reference [4]. Among them, electromagnetic wear particle detectors have demonstrated significant advantages in online wear condition monitoring because of their strong anti-interference ability, good temperature stability, and high reliability.

To achieve a better particle detection effect, sensors with different structures have long been objects of study. Flanagan et al. [5] proposed a wear particle detector with a single coil (inner diameter of 6 mm), which identifies particles by the fluctuation of the sensor resonance frequency. Experimental results showed that the sensor could detect iron particles with a diameter of 150 μm . Fan et al. [6]

designed a double-coil wear particle detection sensor. It estimates the size and the material properties of particles by measuring the inductance difference between the sensing coil and the reference coil of the sensor and can successfully detect 100 μm ferromagnetic particles and 500 μm non-ferromagnetic particles. To improve the consistency of the particle detection results, a sensor with planar spiral coils [7] was proposed. The simulation and experimental results showed that the uniformity of the magnetic field in the detection area was greatly improved, however, the sensor could only detect the ferromagnetic particles with a diameter of 700 μm . Further, Hong et al. [8] designed a radial inductive debris detection sensor that consisted of a C-type iron core, a drive coil, and an inductive coil. The experimental results indicated that the sensor could effectively detect a 290 μm ferromagnetic particle in a 20 mm diameter pipe. However, the magnetostatic field was adopted in this sensor, so it could not detect non-ferromagnetic particles. To improve the sensitivity of the sensor, the wear particle detector with a parallel three-coil structure was studied [9–11]. The study demonstrated that the sensor could detect approximately 100 μm ferromagnetic particles and 305 μm non-ferromagnetic particles in a 7.6 mm diameter channel. However, the sensitivity and the detectability are still the main obstacles for the development and application of the wear particle detector. Therefore, some measures have been taken to further improve the sensitivity of the sensor. The most direct and valid approach is adopting micro-channel structures [12]. The typical feature of this kind of sensor is that the diameter of the inner channel is smaller than 1 mm, which reduces the distance between target particles and sensor coils. Du et al. [13] proposed a micro-channel device based on an inductive coulter counting principle to detect metal wear particles in lubricating oil. The device could detect about 50 μm ferromagnetic particles and 125 μm non-ferromagnetic particles. Wu et al. [14] designed a microfluidic chip-based inductive wear particle detection device. For this sensor, the inner diameter of the coil was set to 200 μm , and the experimental results revealed that it could detect ferromagnetic particles with a diameter of 5–10 μm . Although the sensitivity of the sensor was greatly enhanced, the small channel diameter of the sensor greatly limits its application to large-scale machineries. Besides that, Li et al. [15] carried out a study to improve the sensitivity of a single-coil wear particle detector. They innovatively proposed that adding an external capacitor to the sensor coil and making the sensor work in a parallel resonance state could boost the sensitivity of the sensor. Recently, Zhu et al. [16] added a ferrite core to the single-coil wear debris detection sensor for the enhancement of sensor sensitivity. With this method, the sensor could detect 11 μm ferromagnetic particles in fluidic pipes with a diameter of 1 mm under a throughput of 750 mL/min.

The size of the minimum detectable particle and the real-time ability of the sensor are also limited by the noise level of the raw signal and the performance of the particle signal extraction algorithm. Fan et al. [17] presented a joint time-invariant wavelet transform and kurtosis analysis approach to extract the effective particle signal. This method depresses the background noise of a raw signal by a threshold. In this way, the wear particle detection effect is greatly influenced by the environmental noise. Li et al. [10,18,19] adopted the maximal overlap discrete wavelet transform to remove vibration interferences from the raw signal. Luo et al. [20] integrated the resonance-based signal decomposition method and fractional calculus (RSD-FC) to improve the detection accuracy of the sensor. These methods do improve the particle detection effect to a certain degree, but they are only valid when the signal-to-noise ratio (SNR) of the signal is sufficiently high, which generally means higher than 2 dB. Meanwhile, overcomplicated algorithms require a relatively high computational cost, which makes the sensor unsuitable for application to continuous real-time monitoring [21].

To meet the requirements of the high-precision wear particle detector and improve the micro-particle detection effect, a comprehensive method of improving the sensitivity and detectability of the sensor is proposed. Based on the essential features of the sensor, a parallel resonance topology and a series resonance topology are applied to the exciting coil and the inductive coil respectively, to comprehensively boost the sensitivity of the sensor. In addition, the influence of resonance frequency on the sensitivity and effective particle detection range of sensors is studied, which lays the foundation for optimizing the frequency of the magnetic field within the sensor. To further improve the induced

electromotive force, an amorphous iron core is added to the inductive coil. The high permeability and the low hysteresis loss and eddy current loss of the amorphous material contribute to improving the sensitivity and keeping the performance of the sensor under a high-frequency alternating magnetic field. Additionally, to improve the real-time performance of wear monitoring, a quick extraction method of the particle signal, based on a modified lock-in amplifier and empirical mode decomposition, is proposed. This method dramatically reduces the amount of computation of the system and can quickly extract the particle signal from the raw signal with an extremely low signal-to-noise ratio (SNR).

2. Device Description and Measurement Setup

2.1. Sensor Description

The core structure of the proposed wear particle detector is shown in Figure 1. Differing from the conventional wear particle detection sensor, which only includes a coil frame, two reverse exciting coils, and an inductive coil, the proposed particle detector adopts the resonance principle and an amorphous iron core to compressively improve its sensitivity. Based on the features of the sensor, the parallel resonance topology is used for the exciting coil to boost the impedance change of the coil caused by particles. Moreover, the series resonance principle is applied to the inductive coil to improve the induced electromotive force. Therefore, the resonant capacitors C_1 and C_3 are connected to the left and right exciting coils of the sensor in parallel, and the resonant capacitor C_2 is connected to the inductive coil in series. The general working principle of the sensor has been expounded in Reference [22]. In order to achieve the flow requirements of wear monitoring for large-scale machines, the inner diameter of the sensor is set to 7 mm.

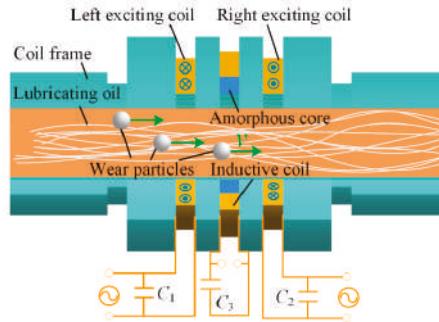


Figure 1. The structure of the proposed wear particle detector.

The metal wear particles passing through the sensor lead to magnetic perturbation of the sensor. More specifically, ferromagnetic particles enhance the local magnetic flux density, while non-ferromagnetic particles decrease the local magnetic flux density [22]. In these cases, the change of the magnetic flux through the exciting coil and the inductive coil can be expressed as (1) and (2), respectively:

$$\Delta\phi_e = \sum \int \Delta B_p(x, y) ds = \Delta(L \times I) \quad (1)$$

$$\Delta\phi_i = K(1 - \lambda)(\phi_{e1} - \phi_{e2}) \quad (2)$$

where, ϕ_e is the magnetic flux through the exciting coil, ΔB_p is the change of magnetic flux density in the sensor caused by particles, L is the inductance of the exciting coil, I is the current through the exciting coil, K is the gain factor of magnetic flux through the inductive coil, λ is the magnetic flux leakage coefficient, which is closely related to the sensor structural parameters, and ϕ_{ei} is the magnetic flux through the i th exciting coil.

The induced electromotive force output by the inductive coil can be expressed as (3), where N_i is the number of turns of the inductive coil:

$$E_0 = -N_i \frac{\Delta\phi_i}{\Delta t} \approx -KN_i(1 - \lambda) \frac{\Delta(L \times I)}{\Delta t}. \tag{3}$$

From the above equation, we can see that for the sensor with certain structural parameters, the magnitude of the induced electromotive force is related to the product of the inductance of the exciting coil and current through the exciting coil, and the gain factor K . Because the change of coil inductance caused by wear particles is extremely weak, one method of improving the sensitivity of the sensor is to enlarge the current variation through the exciting coils, which is closely associated with the impedance change of the exciting circuit caused by particles. Meanwhile, this research proves that a series-resonant inductive coil and an amorphous core can boost the gain factor K . The mechanism of enhancing the sensitivity of the sensor is explained in detail in the following section.

2.2. A Sensitivity Comparison Analysis of the Sensors

To demonstrate the mechanism of sensitivity improvement by the resonant principle and the amorphous core, a sensitivity comparison analysis of the conventional and proposed wear particle detector was conducted. The circuit diagrams of the sensors are displayed in Figure 2a,b, where L_1 and L_2 are the inductances of the exciting coils, L_3 is the inductance of the inductive coil, $C_1, C_2,$ and C_3 are the resonant capacitors for each coil, and the internal resistances of these coils are $r_1 = r_2 = 4.1 \Omega$ and $r_3 = 4.3 \Omega$. For the proposed sensor, as shown in Figure 2b, the resonance condition must be satisfied as Equation (4), where f_0 is the resonant frequency.

$$f_0 \approx \frac{1}{2\pi \sqrt{LC}} \tag{4}$$

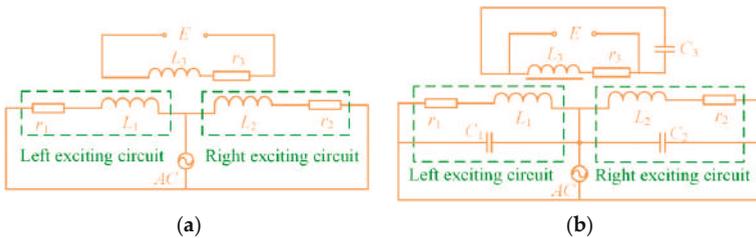


Figure 2. The circuit diagrams of the sensors. (a) The conventional sensor, (b) the proposed sensor.

The impedance change of the exciting circuit caused by particles can characterize the sensitivity of the sensor indirectly. When no particles enter the sensor, the impedance of each exciting circuit of the two sensors, as shown in Figure 2a,b, can be expressed as (5) and (6), respectively. Here, Z_a and Z_b are the impedances of the non-resonant and resonant exciting circuits respectively, $L_q = L_i - M$ is the equivalent inductance of a single exciting coil, L_i is the self-inductance of the i th exciting coil, and M is the mutual inductance between the two exciting coils. Note that, under the resonance state, $1 - \omega^2 L_q C \approx 0$ and $\omega Cr \ll 1$, so it can be obtained that $Z_b \gg Z_a$.

$$Z_a = j\omega L_q + r \tag{5}$$

$$Z_b = \frac{(j\omega L_q + r)}{1 - \omega^2 L_q C + j\omega Cr}. \tag{6}$$

When wear debris gets access to the sensor, the inductance of one of the two exciting coils changes, which further leads to an impedance difference between the two exciting circuits. Taking the ferromagnetic particle as an example, the inductance-change of a coil caused by a ferromagnetic particle with a radius of r_a can be expressed as (7) [23]:

$$\Delta L = \frac{(\sqrt{5}-1)\mu_0\mu_r N^2 r_a^3}{l^2} \tag{7}$$

Here, $\mu_0 = 4\pi \times 10^{-7}$ H/m is the permeability of the vacuum, μ_r is the relative permeability, N is the number of turns of the coil, and l is the width of the coil.

The impedance differences between the exciting circuits of the two sensors, as shown in Figure 2a,b, are given by:

$$\begin{aligned} \Delta Z_a &= j\omega\Delta L \\ \Delta Z_b &= \frac{j\omega\Delta L}{(1-a^2C(L+\Delta L)+iCr\omega)(1-a^2CL+iCr\omega)} \end{aligned} \tag{8}$$

To characterize the sensitivity of the two sensors, the impedance differences between the exciting circuits of each sensor are calculated by MATLAB (MathWorks, USA) and shown in Figure 3. During the calculation, the equivalent inductance of the exciting coils is $L_{q1} = L_{q2} = 270.2 \mu\text{H}$, which is obtained from experimental measurement, the exciting frequency is set to $f_0 = 134.5 \text{ kHz}$, and the corresponding resonant capacitances are $C_1 = C_2 = 5.17 \text{ nF}$. It can be seen that for the sensor with a non-resonance principle, the impedance difference slowly grows with the increase of the particle diameter, and that it is merely 0.41Ω when the diameter of the ferromagnetic particle is $750 \mu\text{m}$. However, for the sensors with resonant exciting coils, the impedance difference rises rapidly with the increase of particle diameter, reaches a peak value (3.99Ω) at the position of r_1 ($528 \mu\text{m}$), and then decreases sharply. Therefore, the obvious impedance difference between the exciting circuits of the proposed sensor signifies that the parallel resonant exciting coil does improve the sensitivity of the sensor to a certain extent. However, the nonlinear characteristics of the impedance difference mean that different sized particles, such as the particles with the diameter of r_p and $r_{p'}$, may lead to the same impedance change, and even the impedance change, caused by the particle larger than r_2 in diameter, turns negative, which means that the large ferromagnetic particle may be recognized as a non-ferromagnetic particle. Therefore, for correctness of the particle detection result, the effective detection range of the proposed sensor is restricted to $(0, r_1)$.

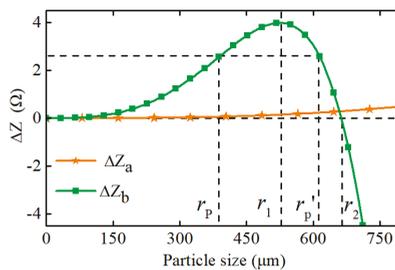


Figure 3. The impedance difference between exciting circuits of different sensors.

To effectively monitor the initial abnormal wear stage of the machinery, some measures must be taken to improve the detectability for micro particles. It is calculated that for the proposed sensor, the resonance capacitance (or resonance frequency) greatly affects the peak position of ΔZ_b . The impedance differences between the two exciting circuits with different resonance capacitors are displayed in Figure 4. It can be seen that with the decrease of the capacitance, the impedance difference curve shifts to the left, which reduces the particle detection range of the sensor to $(0, r_a')$, but enhances the impedance difference between the two exciting circuits caused by micro particles. Therefore, the

smaller resonance capacitance (higher resonance frequency) contributes to the detection of micro wear particles. However, that greatly increases the current through the exciting coils and makes the sensor produce more heat, which is harmful to the reliability of the sensor. Meanwhile, the excessive field frequency increases the magnetic losses in particles, which weakens the detectability for ferromagnetic particles. Considering the above factors, a real well-selecting experiment was conducted, and the results showed that a resonant capacitance of 1nF is appropriate for the detection of ferromagnetic particles. In this situation, the detection range of the sensor was restricted to (0, 300) μm.

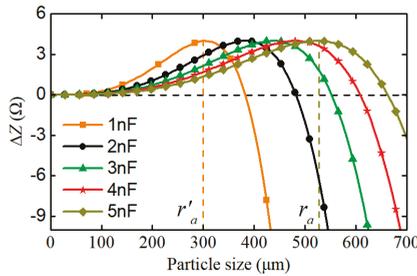


Figure 4. Impedance difference between exciting circuits of the proposed sensor with different resonant capacitors.

The impedance change of exciting coils caused by particles leads to current redistribution, which is one of the key factors of improving the sensitivity of the sensor. Under this circumstance, the current difference between exciting coils, for the sensors shown in Figure 2a,b, can be expressed as (9) and (10), respectively:

$$\Delta I_a = I_0 \left(\frac{1 + \Delta Z_a / Z_a}{2 + \Delta Z_a / Z_a} \right) \left(1 - \frac{1}{1 + \Delta Z_a / Z_a} \right) \tag{9}$$

$$\Delta I_b = I_0 \frac{Z_b}{Z_a} \left(\frac{1 + \Delta Z_b / Z_b}{2 + \Delta Z_b / Z_b} \right) \left(1 - \frac{1}{1 + \Delta Z_a / Z_a} \right). \tag{10}$$

Note here that, when the particle diameter is distributed in the range (0, r'_a), $Z_b > Z_a$ and $\Delta Z_b / Z_b > \Delta Z_a / Z_a$. Therefore, we obtain:

$$\Delta I_b = \frac{Z_b}{Z_a} \left(\frac{1 + \Delta Z_b / Z_b}{2 + \Delta Z_b / Z_b} \right) \left(\frac{2 + \Delta Z_a / Z_a}{1 + \Delta Z_a / Z_a} \right) \Delta I_a \gg \Delta I_a. \tag{11}$$

The combination of (3) and (11) implies that the parallel resonant exciting coil can essentially improve the induced electromotive force. Meanwhile, Equations (2) and (3) indicate that increasing the magnetic flux through the inductive coil is helpful to further enhance the detectability for micro wear particles and boost the sensitivity of the sensor. Therefore, an amorphous iron core is added to the inductive coil. For the inductive coil, the difference in the magnetic flux density between the two exciting coils can be equivalent to a weak external magnetic field H_p , which produces the magnetic flux of the inductive coil. Based on the equation of $B = \mu H, \varphi = \sum \int B ds$, it can be obtained that a ferrite core with a high permeability can boost the external magnetic field and enhance the magnetic flux of the inductive coil. To demonstrate the enhancement effect of the magnetic flux by the amorphous core, a simulation was conducted using the software of COMSOL Multiphysics (COMSOL, Stockholm, Sweden). The simulation parameters used were obtained from the experimental system (illustrated in Section 3). The magnetic fluxes of the inductive coil caused by a 100 μm iron particle for the sensors are displayed in Figure 5. It can be seen that the magnetic flux through the inductive coil of the sensor with the amorphous core increases significantly. In this case, a larger induced electromotive force is produced by the inductive coil.

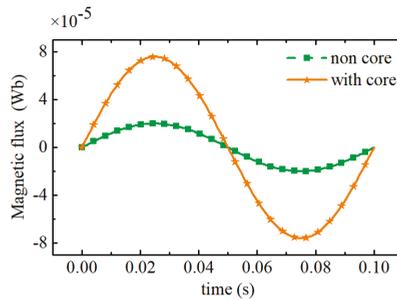


Figure 5. Magnetic flux of the inductive coil.

To further magnify the induced electromotive force caused by particles, the series resonance principle is adopted for the inductive coil and the capacitor C_3 also needs to meet the resonance condition as (4). It is noteworthy that the resonance frequency should maintain a consistent value with the exciting frequency f_0 and the inductive coil can be regarded as a power source. Under the series resonant state, the current through the coil reaches a peak as (12), and the output signal of the sensor can be expressed as (13). The result shows that the series resonant inductive coil magnifies the output signal of the sensor, and the magnification can be comprehensively described as the quality factor of the induction coil. In this situation, the stray capacitance of the coil and the equivalent series resistance of the resonant capacitor cannot be neglected, so it is difficult to directly calculate the quality factor. We measured the quality factor using a digital electric bridge tester (TH2821B) and obtained an approximate value of 3.22, which indicates that the output signal of sensor $E_s \approx 3.22E_0$:

$$I_3 = \frac{E_0}{(r_3 + j\omega L_3 + 1/j\omega C_3)} \approx \frac{E_0}{r_3} \tag{12}$$

$$E_s = I_3(r_3 + j\omega L_3) = E_0 \sqrt{1 + (\omega L_3/r_3)^2} > E_0. \tag{13}$$

Here, I_3 is the current through the inductive coil under the resonant state, and E_0 and E_s are the induced electromotive forces output by the inductive coil and the sensor, respectively.

Consequently, adding an amorphous iron core to the inductive coil and making it work in the series resonance state are two significant methods of further improving the sensitivity of the sensor.

2.3. Particle Signal Measurement Setup

For the proposed sensor, because of the weak inhomogeneity of the magnetic field between the exciting coils, the initially induced electromotive force interference is produced when no particles pass through the sensor. By analyzing the characteristics of the sensor signal, it can be obtained that the real output signal is composed of the effective particle signal, initially induced electromotive force interference, and environmental interference. The real sensor signal can be expressed as:

$$E_s = E_0 \sqrt{1 + (\omega L_3/r_3)^2} = (E(r_a, v) \sin(\omega_1 t + \varphi_2) + E(\Delta)) \sin(\omega_0 t + \varphi_1) + N(t) \tag{14}$$

where, $E(r_a, v) \sin(\omega_1 t + \varphi_2)$ is the effective particle signal, $E(\Delta) \sin(\omega_0 t + \varphi_1)$ is the initially induced electromotive force interference, ω_0 and ω_1 are the angular frequencies of the exciting signal of the sensor and the effective particle signal respectively, and $N(t)$ is the Gaussian noise resulting from environmental interference.

A measurement system for weak signals is crucial for the detection of wear particles. For satisfying the high real-time requirements of online wear monitoring, a new signal extraction method, based on a modified lock-in amplifier (MLIA) and empirical mode decomposition (EMD), is proposed. Compared

with conventional peak-detection (PD) algorithms [17,18,20], the proposed method is much simpler and faster. It can adapt to circumstances with an extremely low signal-to-noise ratio (SNR). Figure 6 shows the block diagram of the signal measurement system. The frequency synthesizer is used to adjust the frequency of the exciting signal to satisfy various monitoring situations. A capacitance matcher is applied to match suitable capacitances for sensor coils. The process of particle signal extraction includes the pre-detection process, preliminary signal extraction, and signal shaping. In the pre-detection process, the raw signal of the sensor is amplified and then filtered by a power frequency filter and an anti-aliasing filter to remove the 50 Hz interference and the high-frequency interference which is generally caused by mechanical vibration of the sensor. For preliminary signal extraction, a modified lock-in amplifier (MLIA) is proposed. In contrast to a conventional lock-in amplifier (LIA), the MLIA adopts two Bessel-type band-pass filters with a center frequency f_0 due to the essential feature of the sensor signal, and the effective particle signal is amplitude-modulated by a sinusoidal signal with a frequency of f_0 . Besides that, to quickly eliminate the initially induced electromotive force interference, a Bessel high-pass filter with a cut-off frequency of 5 Hz was used. Because the extraction effect of the particle signal is relevant to the function of these filters and SNR of the raw signal, to adapt the detection requirement of the particles with different speeds, the raw signal is always under-filtered by these filters. Therefore, some unfiltered Gaussian interference still exists in the particle signal, which lowers the detection effect for particles, especially for particles with a low speed. Hence, the particle signal-shaping method based on the EMD is proposed.

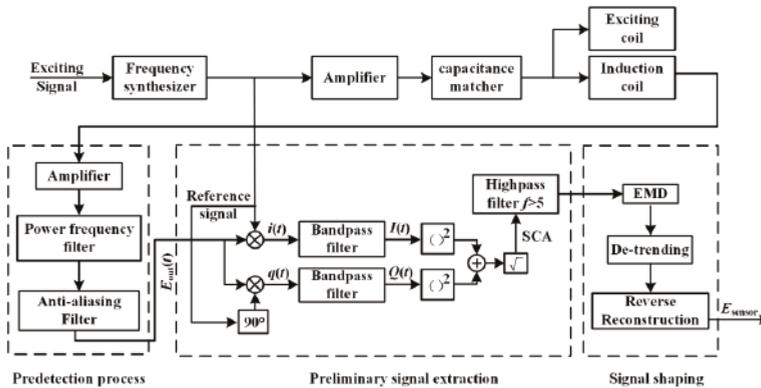


Figure 6. Block diagram of the signal measurement system.

In the procedure for preliminary signal extraction, the reference signal of MLIA is set to $A \sin(\omega_0 t + \varphi_3)$, which has the same frequency as the exciting signal. After that, the raw signal is multiplied by both the reference signal and a signal in quadrature with respect to a reference signal of $A \cos(\omega_0 t + \varphi_3)$. The signals of $i(t)$ and $q(t)$ can be obtained as (15) and (16), respectively. It can be seen that $i(t)$ and $q(t)$ consist of three parts: the amplitude component, high-frequency part (frequency is $2f_0$), and noise sector:

$$\begin{aligned}
 i(t) &= ((E(r_a, v) * \sin(2\pi f_1 + \varphi_2) + E(\Delta)) * \sin(2\pi f_0 t + \varphi_1) + N(t)) A \sin(2\pi f_0 t + \varphi_3) \\
 &= \frac{A}{2} (E(r_a, v) * \sin(2\pi f_1 + \varphi_2) + E(\Delta)) * \cos(\varphi_1 - \varphi_3) - \frac{A}{2} (E(r_a, v) * \sin(2\pi f_1 + \varphi_2) \\
 &\quad + E(\Delta)) * \cos(2 * 2\pi f_0 t + \varphi_1 + \varphi_3) + N(t) * A \sin(2\pi f_0 t + \varphi_3)
 \end{aligned} \tag{15}$$

$$\begin{aligned}
 q(t) &= ((E(r_a, v) * \sin(2\pi f_1 + \varphi_2) + E(\Delta)) * \sin(2\pi f_0 t + \varphi_1) + V(t) + N(t)) * A \cos(2\pi f_0 t + \varphi_2) \\
 &= \frac{A}{2} (E(r_a, v) * \sin(2\pi f_1 + \varphi_2) + E(\Delta)) * \sin(\varphi_1 - \varphi_3) + \frac{A}{2} (E(r_a, v) * \sin(2\pi f_1 + \varphi_2) \\
 &\quad + E(\Delta)) * \sin(2 * 2\pi f_0 t + \varphi_1 + \varphi_3) + N(t) * A \cos(2\pi f_0 t + \varphi_3)
 \end{aligned} \tag{16}$$

After the MLIA's band-pass filters, the high-frequency component and most of the noise interference can be removed. Therefore, the following signals are obtained:

$$I(t) = \frac{A}{2}(E(r_a, v) * \sin(2\pi f_1 + \varphi_2) + E(\Delta)) * \cos(\varphi_1 - \varphi_3) \quad (17)$$

$$Q(t) = \frac{A}{2}(E(r_a, v) * \sin(2\pi f_1 + \varphi_2) + E(\Delta)) * \sin(\varphi_1 - \varphi_3). \quad (18)$$

The estimation of the specific component amplitude (SCA) is given by (19). There are two sectors in the SCA: a sinusoidal component with a frequency of f_1 , which involves the effective particle signal, and a direct component that reflects the amplitude of the initially induced electromotive force interference. Therefore, a Bessel high-pass filter with a cut-off frequency of 5 Hz is used to remove the DC interference component, and the effective particle signal is then obtained as (20):

$$SCA = \sqrt{I(t)^2 + Q(t)^2} = \frac{A}{2}(E(r_a, v) * \sin(2\pi f_1 + \varphi_2) + E(\Delta)) \quad (19)$$

$$E_{sig} = \frac{A}{2}(E(r_a, v) * \sin(2\pi f_1 + \varphi_2)). \quad (20)$$

That the cut-off frequency of the high-pass filter is 5 Hz means that the allowable minimal speed of particles passing through the sensor is $v = f_1 * l = 5 * 11 * 10^{-3} = 5.5 * 10^{-2}$ m/s, and the corresponding allowable minimum quantity of flow is $V = \pi v d^2 / 4 = 0.127$ L/min. Here, l is the outer distance between the exciting coils and d is the inner diameter of the sensor.

Although the modified lock-in amplifier can preliminarily extract the weak particle signal and greatly improve the SNR of the signal, there is still some unfiltered Gaussian interference which influences the accurate judgment of the signal amplitude. Therefore, the signal-shaping method based on the EMD-RRC (empirical mode decomposition and reverse reconstruction) is adopted. EMD is an adaptive time-frequency signal processing method used to decompose non-stationary or nonlinear data into several elementary intrinsic mode functions (IMFs), which contain the local features of the raw signal at different time scales. The detailed decomposition process is stated in [24,25]. The preliminarily extracted particle signal can be decomposed by the EMD method as:

$$E_{sig} = \sum_{i=1}^k c_i(t) + r(t) \quad (21)$$

where, $c_i(t)$ is the i th intrinsic mode function and $r(t)$ is the residual term.

Based on the theory of the EMD, the low-order IMFs contain the high-frequency component of the raw signal, and the high-order IMFs and the residual term represent the low-frequency trend component of the signal. Considering the preliminarily extracted particle signal, in order to eliminate the residual interference, the trend component with a low frequency should be removed first. Hence, a trend component identification method is adopted. In this method, the trend component is identified as [10]:

$$m(t) = \sum_{i=k_1}^k c_i(t) + r(t) \quad (22)$$

where, k_1 is the trend order of IMFs which satisfies:

$$\begin{aligned} \prod_{i=k_1}^k (|\text{Mean}(c_i(t))| - H_T) > 0 \\ \prod_{i=1}^{k_1-1} (|\text{Mean}(c_i(t))| - H_T) < 0 \end{aligned} \quad (23)$$

where, $\text{Mean}(\cdot)$ denotes the mean function, and $H_T = 0.05|\text{Mean}(r(t))|$ is the threshold.

To further eliminate the high-frequency interference, a reverse reconstruction method is proposed to reconstruct the signal of the particle. This method gradually adds lower-order IMFs to the detrended highest-order IMF, which produces a series of reconstruction signals expressed as:

$$E_{\text{rsig}}^j = \sum_{i=k_1-j}^{k_1-1} c_i(t). \quad (24)$$

The best denoising effect means the maximal correlation between the particle signal and an ideal sinusoidal signal. Hence, the synthesized correlation coefficient as (25) is used to evaluate these reconstructed signals and to select the best reconstruction order:

$$\rho_{\text{rsig}}^j = \frac{\text{COV}(E_{\text{rsig}}^j, E_{\text{std}})}{\sqrt{E_{\text{rsig}}^j} \sqrt{E_{\text{std}}}}. \quad (25)$$

Here, $\text{COV}(\cdot)$ denotes the covariance function and E_{std} is an ideal sinusoidal signal.

The array of synthesized correlation coefficients for the different reconstruction particle signals is established as:

$$\rho_{\text{max}} = \max(|\rho_{\text{rsig}}^1|, |\rho_{\text{rsig}}^2|, \dots, |\rho_{\text{rsig}}^j|). \quad (26)$$

Combining Equations (24)–(26), the best reconstruction signal is expressed as:

$$E_{\text{out}} = \sum E_{\text{rsig}}^j * (\text{sgn}(|\rho_{\text{rsig}}^j| - \rho_{\text{max}}) + 1). \quad (27)$$

The signal extraction process is simulated by MATLAB SIMULINK and the signal-to-noise ratio (SNR), as shown Equation (28), is used to evaluate the effect of the proposed signal measurement system. In addition, to illustrate the influence on the signal detection effect by the initially induced electromotive force interference, the signal-to-harmonics ratio (SHR) is defined as (29).

$$\text{SNR} = 10 \log_{10} (P_p / P_N) \quad (28)$$

$$\text{SHR} = \frac{E_p|_{p-p}}{E_0|_{p-p}}. \quad (29)$$

Here, P_p and P_N are the power of the effective particle signal and the noise signal respectively, E_p is the effective particle signal, E_0 is the initially induced electromotive force, and the subscript p-p means the peak-to-peak value.

The simulation is conducted on the condition that the effective particle signal is $E_0 = 5 \times 10^{-5} \sin(2\pi f_0 t)$, SHR equals 1/100, the variance of Gaussian noise is 1e-8, and the signal amplification factor is 100. In this situation, the raw signal of the sensor is demonstrated in Figure 7a, which shows that the particle signal is fully submerged in the interference, and the SNR of the raw signal is as low as -21.37 dB. The preliminarily extracted particle signal is displayed in Figure 7b. It can be seen that the interference component is greatly removed from the raw signal, however, the residual interference still influences the amplitude recognition. In the process of signal-shaping, the preliminarily extracted signal is decomposed into several IMFs and a residual component by the EMD method, as shown in Figure 7c. Based on Equations (21)–(25), the IMF5 and the residual component are regarded as low-frequency trend components and the IMF1 and IMF2 are treated as high-frequency interference. After eliminating all the interference, the reconstructed signal can be obtained, as shown in Figure 7d. It shows that the shaped particle signal has obvious sinusoidal characteristics.

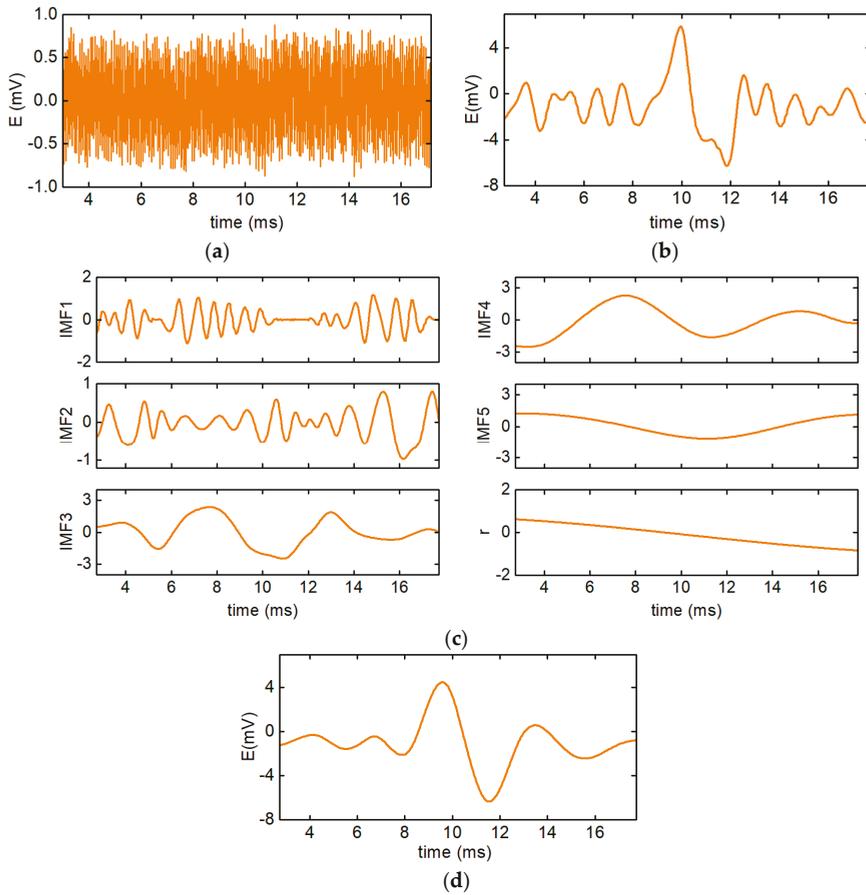


Figure 7. The simulation of the signal extraction process. (a) The raw signal of the sensor, (b) the preliminarily extracted particle signal, (c) the decomposed particle signal, and (d) the shaped particle signal.

To evaluate the validity of the proposed signal extraction and shaping method, the SNR values of the raw signal, preliminarily extracted signal, and shaped signal are calculated and presented in Table 1. The result illustrates that the SNR of the signal is greatly improved, which contributes to boosting the particle detection effect of the sensor.

Table 1. The signal-to-noise ratio (SNR) of signals.

SNR of Raw Signal	SNR of Preliminarily Extracted Signal	SNR of Shaping Signal
-21.37 dB	3.71 dB	13.181 dB

2.4. Analysis of the Computational Cost and Performance of Methods

As wear particles are monitored in real time by an electromagnetic wear particle detector, the computational efficiency of particle signal extraction algorithms and the correctness of detection results are of important concern. Therefore, in this section, a comparative analysis, involving the computational cost and extraction effect of particle signals incurred by the application of RSD-FC

(resonance-based signal decomposition method and fractional calculus) [20], VMD-based method (variational mode decomposition) [26–28], and EMD-RRC (empirical mode decomposition and reverse reconstruction), is presented.

With respect to EMD and VMD, the algorithms decompose raw signals into several sub-signals (modes). However, the implementation of VMD requires first performing a Hilbert transform which involves an EMD process, so the VMD carries on a computational cost higher than the EMD. Besides that, the VMD requires a predetermined number of decomposition level k , which greatly influences its decomposition effect and computational efficiency [28]. Moreover, it's difficult to adjust the value of k for the optimal decomposition effect self-adaptively. The RSD-FC expresses a signal as the sum of a 'high-resonance' component which generally represents the interferences and a 'low-resonance' component which characterizes the particle signal. To achieve this goal, a morphology component analysis needs to be conducted, in which, an iterative optimization algorithm is utilized to update the transform coefficient matrices [20], so the method requires extensive calculations. To evaluate the computational efficiency, the preliminarily extracted particle signal with a sampling time of 1 s, extended from the data of Figure 7b, is processed using different algorithms running on a PC (Intel(R) Core(TM) i7-4720HQ CPU, 2.60 GHz, 8 GB RAM, Windows 10 operating system). For effective detection of wear particles with high speed, the sampling frequency is set to 3000 Hz. The theoretical peak-to-peak value of the particle signal output by the sensor is 10 mV. The performance of the algorithms is evaluated using the mean signal-to-noise ratio (MSNR), mean peak-to-peak value (MPPV), and mean relative amplitude error (MRAE):

$$MRAE = \frac{1}{n} \sum_{i=1}^n \left| \frac{T_i - M_i}{T_i} \right| \times 100\% \quad (30)$$

where, T_i and M_i represent respectively, the theoretical and measured peak-to-peak value of particle signals, and n is the number of samples.

The extraction results of particle signals by RSD-FC, VDM-based method ($k = 7$), and the EMD-RRC are demonstrated in Figure 8a–c, which shows that the residual interferences in preliminarily extracted particle signals are removed to different degrees. The computational time and the performance of the algorithms are displayed in Table 2. It can be seen that all the methods do improve the SNR of signals to a certain degree and the MSNR of the extracted particle signals are higher than 10, which contributes to the effective detection of micro-particles. Furthermore, among these methods, the computational time of the RSD-FC is the longest and reaches to 1.9548 s, which is much larger than the sampling time (1 s). Therefore, it is difficult to guarantee real-time performance of particle detection sensors. Besides that, the correctness of the particle detection results is relatively poor. The MPPV and MRAE of particle signals extracted by the RSD-FC are 9.26 mV and 7.4%, respectively. For the VMD-based method, with the increase of the number of decomposition level k , the computational time rises accordingly. Moreover, comprehensively considering the evaluation indicators, the VMD-based method with $k = 7$ performs best (MSNR = 13.357 dB, MPPV = 9.71 mV, and MRAE = 2.9%). However, in this case, the computational time is 1.4942 s, which is also larger than the sampling time (1 s). While for the proposed EMD-RRC method, the MPPV and the MRAE of signals are 9.68 mV and 3.2%, respectively. Although, they are slightly lower than that of the VMD-based method with $k = 7$, the average computational time is only about 0.83 s which is sufficient to process the data of 1 s long with 3000 samples in real time. In summary, the proposed method is sufficiently fast for on-line application in terms of both computational efficiency and detection quality.

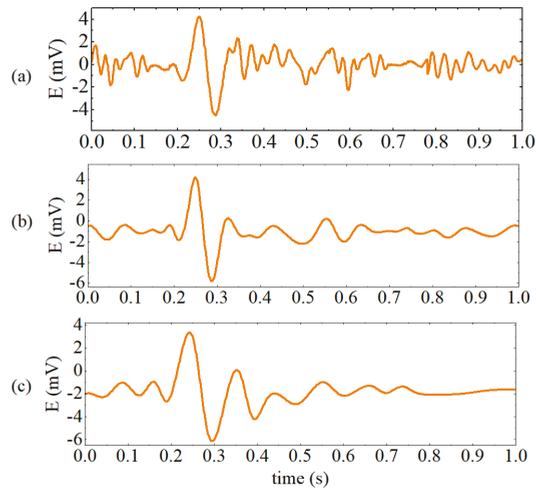


Figure 8. The extraction results of different algorithms applied on the preliminarily extracted signals. (a) resonance-based signal decomposition method and fractional calculus (RSD-FC), (b) variational mode decomposition (VMD)-based method with $k = 7$, (c) empirical mode decomposition and reverse reconstruction (EMD-RCC).

Table 2. Computational times and the performance of different algorithms.

Algorithms	MSNR (dB)	MPPV (mV)	MRAE	Mean Computational Times (s)
RSD-FC	8.854	9.26	7.4%	1.9548
VMD-Based ($k = 4$)	10.347	9.27	7.3%	0.9368
VMD-Based ($k = 5$)	11.755	9.39	6.1%	1.1256
VMD-Based ($k = 6$)	12.982	9.54	4.6%	1.3573
VMD-Based ($k = 7$)	13.357	9.69	3.1%	1.4942
VMD-Based ($k = 8$)	12.793	9.60	4.0%	1.5783
VMD-Based ($k = 9$)	12.584	9.52	4.8%	1.7137
EMD-RRC	13.181	9.68	3.2%	0.8314

3. Experiment

3.1. Experimental System

To verify the improvement of the sensitivity and the detectability of the sensor contributed by the resonance mechanism, the amorphous iron core, and the proposed signal measurement system, the detection efficiencies of the conventional and proposed sensors for wear particles were tested. The complete experimental system, as shown in Figure 9a, consists of the sensor, the excitation and detection unit, which is used to supply the exciting signal and to extract the particle signal, and the data collecting and processing software. The core parameters of the sensors adopted in the experiments are listed in Table 3. Furthermore, some sphere-like iron particles with the diameters of 75, 120, and 150 μm are selected by the scanning electron microscope as target particles, as shown in Figure 9b. The previous experimental research shows [29] that the lubricating oil does not affect the signal of the sensor, so the sensitivity analysis experiments are conducted under an oil-less condition.

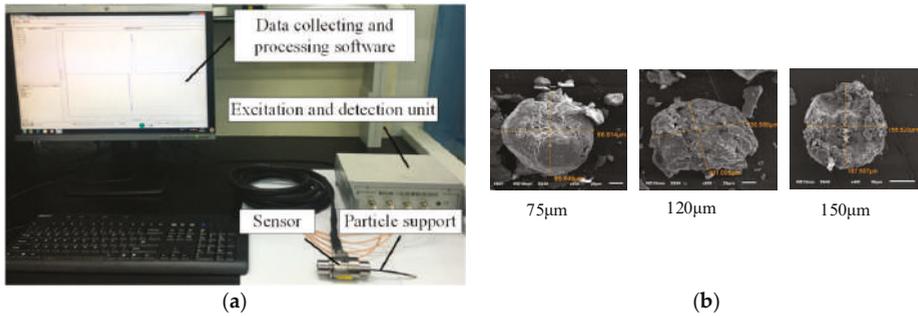


Figure 9. Experimental system. (a) The particle detection system, (b) the selected iron particles.

Table 3. The core parameters of the sensors adopted in the experiments.

Parameters	Conventional Sensor	The Proposed Sensor
Inner diameter of the sensor	7 mm	7 mm
Width of the coils	2 mm	2 mm
Inner diameter of exciting coils	9 mm	9 mm
Number of turns of exciting coils N_e	127	127
Number of turns of exciting coils N_i	110	110
Inner diameter of inductive coil	11 mm	11 mm
Inner diameter of amorphous core	-	9 mm
Outer diameter of amorphous core	-	11 mm
Resonant exciting capacitance C_1, C_2	-	1.0 nF
Resonant inductive capacitance C_3	-	0.63 nF

During the experiment, the measurement data shows that the initially induced electromotive forces of the sensors are about $E_0 = 7.3 \times 10^{-4} \sin(2\pi f_0 t)$ V and the Gaussian noise is very apparent. In this case, the particle signal is totally submerged in the inference. Taking the proposed sensor as an example, Figure 10 shows the raw signal of the sensor caused by a particle with the diameter of 120 μm . Because the particle speed may influence the signal extraction effect to a certain degree, particle detection experiments were conducted when the particle moved at the speed of 3 m/s, 5 m/s, and 8 m/s, respectively. The preliminarily extracted particle signal and the shaped particle signal are shown in Figure 11a,b, respectively. The results indicate that for the preliminarily extracted signals, a better detection is achieved at a higher particle speed. Moreover, after the signal shaping, the residual interference is further removed and the signals of the particle with different speeds can be effectively extracted. The SNR and peak-to-peak values of the particle signals are listed in Table 4, which shows that the proposed particle signal extraction method can greatly enhance the SNR of the particle signals and benefit the detection of micro wear particles. In addition, the peak-to-peak values of the signals are approximately consistent, which means that the signal measurement system has high fidelity.

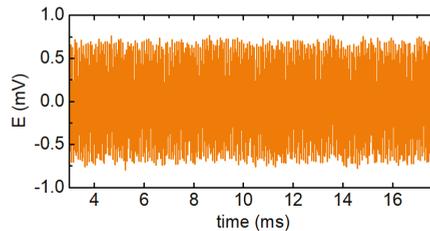


Figure 10. The raw signal of the proposed sensor.

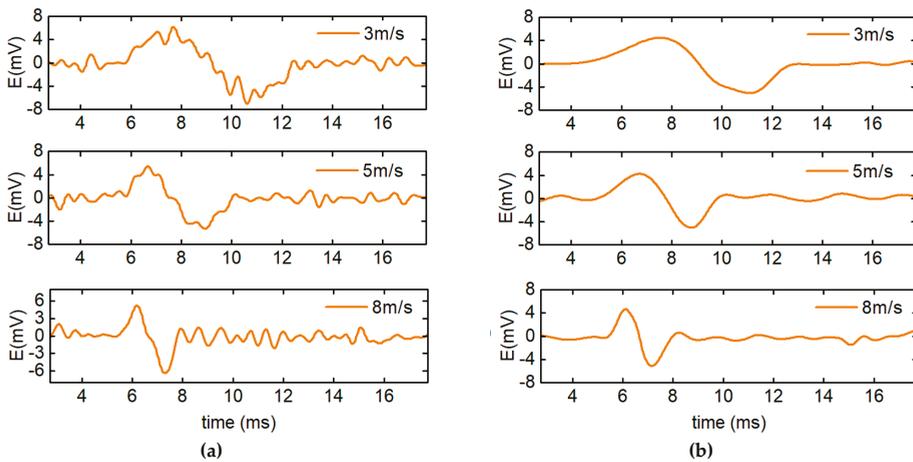


Figure 11. The signal of particles with different speeds. (a) The preliminarily extracted particle signals and (b) the shaped particle signals.

Table 4. The signal-to-noise ratio (SNR) of real signals.

Particle Speed (m/s)	SNR (dB)			Peak-To-Peak Value (mV)
	Raw Signal	Preliminarily Extracted Signal	Shaping Signal	
3	-14.318	9.341	12.933	9.62
5	-15.541	10.366	17.609	9.58
8	-19.917	11.625	15.173	9.55

3.2. Sensitivity Comparison for Ferromagnetic Particle Detection

To illustrate the sensitivity improvement by the proposed methods, both the conventional sensor, as shown in Figure 2a, and the proposed sensor, as shown in Figure 2b, were tested. Figure 12 shows the output signal of the sensors caused by the different sizes of ferromagnetic particles. In the figure, the green curve illustrates the signal output by the conventional sensor, and the orange curve represents the output signal of the proposed sensor, which adopts a resonance principle and an amorphous iron core. It can be seen that, for the conventional sensor, it is difficult to effectively detect iron particles less than 100 μm in diameter and the peak value of the induced electromotive force caused by a 100 μm iron particle is only 0.59 mV. However, for the proposed sensor, the signal amplitude of the particle with the diameter of 75 μm reaches 2.6 mV, which is much greater than that of the conventional sensor.

A comparison analysis of the detection result of the conventional sensor and the proposed sensor with various resonant capacitances is presented in Figure 13. It can be seen that the particle signal output by the proposed sensor is much larger than that of the traditional one, and with the decrease of the exciting capacitance, the sensitivity of the sensor gradually increases. The amplitude of the signal caused by a 75 μm iron particle, when the exciting capacitance equals 1 nF, is 2.6 mV, which is much greater than that under the circumstance of $C_1 = C_2 = 5$ nF (1.06 mV), and the increasing trend tends to be more evident for larger particles. However, excessive reduction of the resonant capacitance leads to a stronger eddy current effect in ferromagnetic particles and increases the current through the exciting coil rapidly, which may weaken the detectability for ferromagnetic particles and greatly reduce the reliability of the sensor. Therefore, a 1 nF resonance capacitance for the exciting coil is finally used for ferromagnetic particle detection.

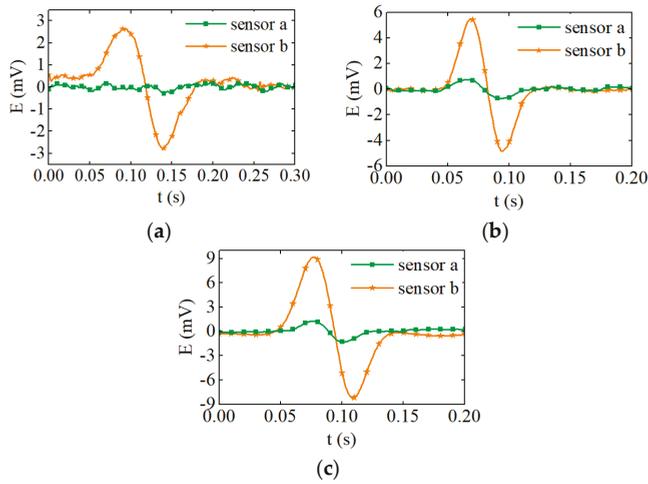


Figure 12. The particle signal output by the sensors. (a) 75 μm , (b) 120 μm , (c) 150 μm .

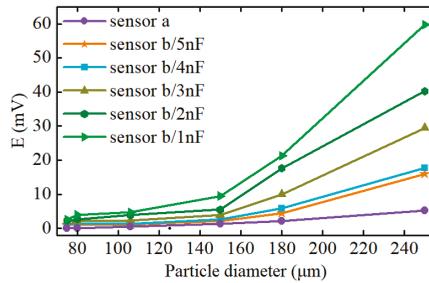


Figure 13. The comparison analysis of sensor sensitivity.

3.3. Wear Monitoring in a Real Oil Environment

To verify the detection effect of the sensor in a real oil environment, the sensor was assembled in the lubrication system with large ferromagnetic wear particles, comprised of 20 particles with a diameter of 80–100 μm , 20 particles with a diameter of 120–150 μm , and 20 particles with a diameter of 150–180 μm . These particles were added into the oil to simulate a serious wear fault of the mechanical equipment. The lubricating oil, including the wear particles, were driven by a pump and cycles through the sensor 20 times. By monitoring the wear particles using the sensor, the size distribution and the number of wear particles were estimated. The statistical result is displayed in Figure 14, which shows that the number of detected wear particles greater than 100 μm in diameter is approximately consistent with the standard value (400). However, the number of iron particles smaller than 100 μm in diameter is slightly more than the standard value. The possible reason for this phenomenon is that some parts of the larger wear particles may stick to the inner surface of the pipeline or be ground down to smaller particles by the blades of the pump during its running process. Therefore, based on the experimental result in a real oil environment, it can be concluded that the sensor can effectively monitor the quantity of the wear particles with different sizes, which helps to estimate the wear state of the mechanical equipment and to prevent mechanical failure caused by serious wear.

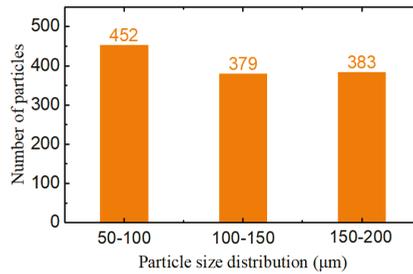


Figure 14. The statistical distribution of the wear particles.

4. Conclusions

The electromagnetic particles' detection sensor is of great importance due to its prospective application in various fields, and the sensitivity and detectability are still major obstacles in the development of wear particle detectors. Therefore, this paper has proposed that the resonance principle, an amorphous iron core, and a new signal measurement system are adopted to comprehensively improve the sensor sensitivity and detectability. Based on the work, the following conclusions are obtained:

- (1) For the three-coil wear particle detector, the parallel resonant exciting coil magnifies the impedance difference between exciting circuits caused by particles. Additionally, the amorphous iron core and the series resonant inductive coil increase the magnetic flux through the coil and enhance the induced electromagnetic force of the sensor, which can comprehensively improve the particle signal more than six times compared to the conventional sensor.
- (2) Under the resonance state, the nonlinear characteristics of the impedance difference between exciting circuits of the proposed sensor mean that the effective particle detection range of the sensor is restricted to $(0, r_a')$.
- (3) Decreasing the resonant capacitance and increasing the exciting frequency can further improve the detection ability for micro-particles, though this reduces the effective particle detection range of sensors.
- (4) By comparing different algorithms, the signal measurement system based on the MLIA and EMD-RRC guarantees the real-time ability for online particle detection and can effectively extract the particle signals from the raw signal with an extremely low SNR (≈ -20 dB). The experimental results indicate that based on the proposed method of improving the sensitivity and detectability, the large-calibre (7 mm) sensor can effectively monitor the initial abnormal wear of the heavy machines.

Author Contributions: Conceptualization, R.J., B.M., and C.Z.; methodology, R.J., L.W., and X.B.; validation, Q.D. and K.W.; writing—original draft preparation, R.J.; writing—review and editing, C.Z.; funding acquisition, C.Z. and L.W.

Funding: This work was funded by the National Natural Science Foundation of China (NSFC) (grant number: 51475044) and Beijing finance found of science and technology planning project (grant number: KZ201611232032).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mabe, J.; Zubia, J.; Gorritxategi, E. Photonic low cost micro-sensor for in-Line wear particle detection in flowing lube oils. *Sensors* **2017**, *17*, 5863. [[CrossRef](#)] [[PubMed](#)]
2. Loutas, T.H.; Roulias, D.; Pauly, E.; Kostopoulos, V. The combined use of vibration, acoustic emission and oil debris on-line monitoring towards a more effective condition monitoring of rotating machinery. *Mech. Syst. Signal Process.* **2011**, *4*, 1339–1352. [[CrossRef](#)]

3. Guo, Z.; Yuan, C.; Yan, X.; Peng, Z. 3D surface characterizations of wear particles generated from lubricated regular concave cylinder liners. *Tribol. Lett.* **2014**, *1*, 131–142. [[CrossRef](#)]
4. Wu, T.; Wu, H.; Ying, D.; Peng, Z. Progress and trend of sensor technology for on-line oil monitoring. *Sci China Technol. Sci.* **2013**, *12*, 2914–2926. [[CrossRef](#)]
5. Flanagan, I.M.; Jordan, J.R.; Whittington, H.W. An inductive method for estimating the composition and size of metal particles. *Meas. Sci. Technol.* **1990**, *5*, 381–384. [[CrossRef](#)]
6. Fan, H.B.; Zhang, Y.T.; Ren, G.Q.; Li, Z.N. Experiment study of an on-line monitoring sensor for wear particles in oil. *Tribology* **2010**, *4*, 338–343. (In Chinese)
7. Kim, B.; Han, S.; Kim, K. Planar spiral coil design for a pulsed induction metal detector to improve the sensitivities. *IEEE Antenn. Wirel. Propag. Lett.* **2014**, *13*, 1501–1504.
8. Hong, W.; Wang, S.; Tomovic, M.; Han, L.; Shi, J. Radial inductive debris detection sensor and performance analysis. *Meas. Sci. Technol.* **2013**, *12*, 125103. [[CrossRef](#)]
9. Hong, H.; Liang, M. A fractional calculus technique for on-line detection of oil debris. *Meas. Sci. Technol.* **2008**, *5*, 55703. [[CrossRef](#)]
10. Li, C.; Liang, M. Extraction of oil debris signal using integral enhanced empirical mode decomposition and correlated reconstruction. *Meas. Sci. Technol.* **2011**, *8*, 85701. [[CrossRef](#)]
11. Zhan, H.; Song, Y.; Zhao, H.; Gu, J.; Yang, H.; Li, S. Study of the sensor for on-line lubricating oil debris monitoring. *Sens. Transducers* **2014**, *7*, 214–219.
12. Zeng, L.; Zhang, H.; Wang, Q.; Zhang, X. Monitoring of non-ferrous wear debris in hydraulic oil by detecting the equivalent resistance of inductive sensors. *Micromachines* **2018**, *3*, 117. [[CrossRef](#)] [[PubMed](#)]
13. Du, L.; Zhe, J.; Carletta, J.; Veillette, R.; Choy, F. Real-time monitoring of wear debris in lubrication oil using a microfluidic inductive Coulter counting device. *Microfluid. Nanofluid.* **2010**, *6*, 1241–1245. [[CrossRef](#)]
14. Wu, Y.; Zhang, H.; Zeng, L.; Chen, H.; Sun, Y. Determination of metal particles in oil using a microfluidic chip-based inductive sensor. *Instrum. Sci. Technol.* **2016**, *3*, 259–269. [[CrossRef](#)]
15. Du, L.; Zhu, X.; Han, Y.; Zhao, L.; Zhe, J. Improving sensitivity of an inductive pulse sensor for detection of metallic wear debris in lubricants using parallel LC resonance method. *Meas. Sci. Technol.* **2013**, *24*, 0751067. [[CrossRef](#)]
16. Zhu, X.L.; Zhong, C.; Zhe, J. A high sensitivity wear debris sensor using ferrite cores for online oil condition monitoring. *Meas. Sci. Technol.* **2017**, *7*, 075102. [[CrossRef](#)]
17. Fan, X.; Liang, M.; Yeap, T. A joint time-invariant wavelet transform and kurtosis approach to the improvement of in-line oil debris sensor capability. *Smart Mater. Struct.* **2009**, *8*, 085010. [[CrossRef](#)]
18. Li, C.; Liang, M. Enhancement of oil debris sensor capability by reliable debris signal extraction via wavelet domain target and interference signal tracking. *Measurement* **2013**, *4*, 1442–1453. [[CrossRef](#)]
19. Liang, M.; Li, C. Enhancement of the wear particle monitoring capability of oil debris sensors using a maximal overlap discrete wavelet transform with optimal decomposition depth. *Sensors* **2014**, *14*, 6207–6228.
20. Luo, J.; Yu, D.; Liang, M. Enhancement of oil particle sensor capability via resonance-based signal decomposition and fractional calculus. *Measurement* **2015**, *76*, 240–254. [[CrossRef](#)]
21. Barrios, M.L.R.; Montero, F.E.H.; Gómez Mancilla, J.C.; Marín, E.P. Application of lock-in amplifier on gear diagnosis. *Measurement* **2017**, *107*, 120–127. [[CrossRef](#)]
22. Jia, R.; Ma, B.; Zheng, C.S.; Wang, L.Y.; Ba, X.; Du, Q.; Wang, K. Magnetic properties of ferromagnetic particles under alternating magnetic fields: Focus on wear particle detection sensor applications. *Sensors* **2018**, *12*, 4144. [[CrossRef](#)] [[PubMed](#)]
23. Zeng, L.; Zhang, H.P.; Teng, H.B.; Zhang, X.M. Novel method for detection of multi-contaminants in marine lubricants. *J. Mech. Eng.* **2018**, *54*, 125–132. [[CrossRef](#)]
24. Camarena-Martinez, D.; Valtierra-Rodriguez, M.; Perez-Ramirez, C.; Amezquita-Sanchez, J.; Rene, R.T.; Garcia-Perez, A. Novel down-sampling empirical mode decomposition approach for power quality analysis. *IEEE Trans. Ind. Electron.* **2015**, *63*, 2369–2378. [[CrossRef](#)]
25. Rai, A.; Upadhyay, S.H. Bearing performance degradation assessment based on a combination of empirical mode decomposition and k-medoids clustering. *Mech. Syst. Signal Process.* **2017**, *93*, 16–29. [[CrossRef](#)]
26. Lahmiri, S. A variational mode decomposition approach for analysis and forecasting of economic and financial time series. *Expert Syst. Appl.* **2016**, *55*, 268–273. [[CrossRef](#)]

27. Liu, Y.; Stone, J.E.; Cai, E.; Fei, J.; Lee, S.H.; Park, S.; Ha, T.; Selvin, P.R.; Schulten, K. VMD as a software for visualization and quantitative analysis of super resolution imaging and single particle tracking. *Biophys. J.* **2014**, *106*, 202a. [[CrossRef](#)]
28. Fan, J.; Zhu, Z.; Wei, L. An improved vmd with empirical mode decomposition and its application in incipient fault detection of rolling bearing. *IEEE Access* **2018**, *6*, 44483–44493.
29. Jia, R.; Ma, B.; Zheng, C.S.; Wang, L.Y.; Du, Q.; Wang, K. Sensitivity improvement method of on-line inductive wear particles monitor sensor. *J. Hunan Univ. Nat. Sci.* **2018**, *45*, 129–137.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Analysis of Satellite Compass Error's Spectrum

Andrzej Felski ¹, Krzysztof Jaskólski ^{1,*}, Karolina Zwolak ¹ and Paweł Piskur ²

¹ Polish Naval Academy, Faculty of Navigation and Naval Weapons, Smidowicza St, 69, 81127 Gdynia, Poland; a.felski@amw.gdynia.pl (A.F.); k.zwolak@amw.gdynia.pl (K.Z.)

² Polish Naval Academy, Faculty of Mechanical and Electrical Engineering, Smidowicza St, 69, 81127 Gdynia, Poland; p.piskur@amw.gdynia.pl

* Correspondence: k.jaskolski@amw.gdynia.pl

Received: 29 May 2020; Accepted: 17 July 2020; Published: 22 July 2020

Abstract: The satellite compass is one of new variants of satellite navigational devices. Is it still treated with caution on International Convention for the Safety of Life at Sea (SOLAS) vessels, but has become popular on the fishing vessels and pleasure crafts. The standard data obtained by such devices suggest accuracy of satellite compasses at a level of about 1 degree, so it seems to be as accurate as gyro or the magnetic equivalent. A changeability of heading errors, especially its frequency spectrum, is analyzed and presented in the paper. The results of comparison of an onboard standard gyrocompass, a fiber-optic gyrocompass (FOG) and a satellite compass in real shipping circumstances have been discussed based on the available literature and previous research. The similar comportment of these compasses are confirmed, however, in real circumstances it is difficult to separate heading oscillations produced by the ships yaw (or helmsman abilities) from the oscillations of the compass. Analysis of the heading oscillations has been performed based on the measurements of the heading indications of stationary compass devices and the devices mounted on the vehicles moving on the straight line (straight part of a road and tram line) to separate the impact of the vessel steering system. Results of heading changeability in the frequency domain are presented based on the Fourier transform theory.

Keywords: satellite compass; accuracy; spectrum analysis; Fourier transform

1. Introduction

It is impossible to navigate a vessel without any directional reference. All movements, no matter for people or for vehicles, in environments such as deserts, seas or air, require direction indicators. The contemporary ship is equipped with a magnetic compass and a gyrocompass as indispensable devices. This refers to all the open-sea ships which must be equipped in compliance with the International Convention for the Safety of Life at Sea (SOLAS). Toward the end of the 20th century, satellite compasses began to trace in a completely new way [1,2] in the form of a specific version of a multiantenna GNSS receiver with the additional option to determine a ship's heading. The most popular is the two-antenna solution, which gives an opportunity to measure two angles: heading and pitch or heading and roll, depending on how it is installed in relation to the centerline of the ship. Three-antenna solutions are also accessible. They allow measuring the full information of the ship's orientation in the space. According to Sperry Marine [3], one of the manufacturers of such devices, it has been designed as a low-cost alternative to conventional spinning-mass and fiber-optic gyrocompasses for application on workboats, commercial fishing vessels, large private yachts, naval patrol boats, and small merchant ships, which are not required to carry a gyrocompass.

The origin of this devices can be found in Very Long Base Interferometry (VLBI)—a radio-astronomical method in which space sources of electromagnetic signals (usually quasar) are collected by multiple radio telescopes distributed on the Earth [4]. On this basis, by means of correlation of

random-type noise registered in the same time in different places (global network), distances between telescopes can be calculated. In the 1960s and 1970s, it was a very efficient method in geodesy, geodynamics etc. on a global scale.

Signals from satellites can be treated in the same manner. A correlation between signals received by an array of antennas, distributed in a specific way, give us an opportunity to calculate direction on the source of the signal (satellite) when the structure of this signals is known. In the case of a Global Positioning System (GPS) satellite, we are working with a 19 cm long electromagnetic wave, so the distribution of receiving antennas can be of around 1 m. By using two receiving systems and utilizing the carrier wave of GPS signals, we are in fact using (RTK) Real Time Kinematic GPS technology. In the simplest version, the two antennas, namely, base (primary) and rover (secondary), are situated along one of the axis of the ship. In the classical version of RTK, the coordinates of the primary antenna should be known; then, a spatial vector between both antennas can be calculated. As we are not interested in very accurate measurements of the antennas' positions and the base distance between both antennas is known (due to the fact that a base line between the two antennas is constant and situated in a constant position referring to the hull of the ship), the distances between each antenna and satellite enables us to determine the angles in two axes (Figure 1, axes X and Z).

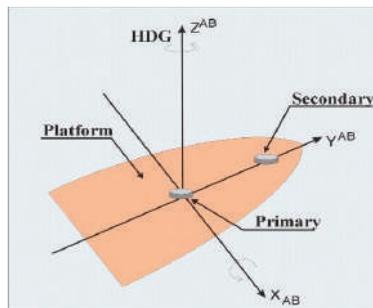


Figure 1. The idea of the use of two antennas to determine the heading of the ship [1].

Such ideas appeared at the end of the 20th century. The numerous publications of Calgary authors, including [5–8], are particularly noteworthy. Before the manufacturers proposed such devices, many researchers used them as reference systems for compass testing, for example [9,10]

Proposals for using the GPS system-derived devices to determine the angles of the spatial orientation of the object, using multiple antennas, appeared earlier. Anthony Evans is the author of significant achievements in this area by using the GPS system to determine ship orientation in the 1980s [11]. He proposed a method of measuring orientation angles through a single antenna that cyclically rotates inside the aircraft fuselage. In 1988, he began experimenting with an 18-channel receiver that used a system of three antennas spaced from 40 cm to 60 cm apart. This satellite compass precursor was tested under a marine conditions on the “USS Yorktown” to determine an accuracy during movement [12]. Parameters such as the duration of system initialization, maintaining its continuity and required accuracy in a real time in a dynamic environment, were examined at the beginning of satellite compasses development. Confirmation of the hypothesis was obtained that multi-antenna GPS receivers, in addition to the positioning ability, are able to determine reliable data regarding a ship’s spatial orientation. Another example is the proposal contained in the patent of 1998 [13]. The authors proposed a compass, which determines its spatial orientation based on a construction with two antennas rotated by a stepper motor until a phase equalization of both antennas occurs.

The first devices available to wider users were introduced in the 1990s. In 1991, Ashtech launched the first multi-antenna GPS receiver: 3DF. Using this system, it was possible to determine the heading,

longitudinal and transverse tilt, as well as the position, using a system consisting of four antennas, one of which served as the base antenna, and three others were supporting the base. Each antenna cooperated with a separate receiver. All of them, using signals from at least four different satellites simultaneously, by measuring the difference in phase, determined the orientation of the antenna assembly in three-dimensional space. The research was continued by scientists from the University of Calgary, who in 1994 conducted tests on the Canadian research ship, “Endeavor”. Four GPS antennas were mounted on the vessel’s helipad. The purpose of the tests was to compare the indications of this system with the Sperry Mark 3 Model C gyro compass available on the vessel [14,15]. Other analyses related to this subject, especially over the optimal configuration of antennas, have also been published in [5,16].

In 1994, Trimble introduced a four-antenna system called TANS Vector [17]. The system performance was based on phase measurements between one of the antennas (main, base) and each of the others, which were treated as slave antennas.

2. Materials and Methods

2.1. Background

Despite many tests confirming the usefulness of multi-antenna GNSS receivers for measurements related to spatial orientation, devices that met the criteria specified in international conventions have been developed no earlier than in 2005. Large errors appeared periodically in all the previous constructions, which were related to the changing satellite constellation parameters. An example of such measurements, taken in 2004 with the Crescent compass installed on the roof of the building is shown in Figure 2.

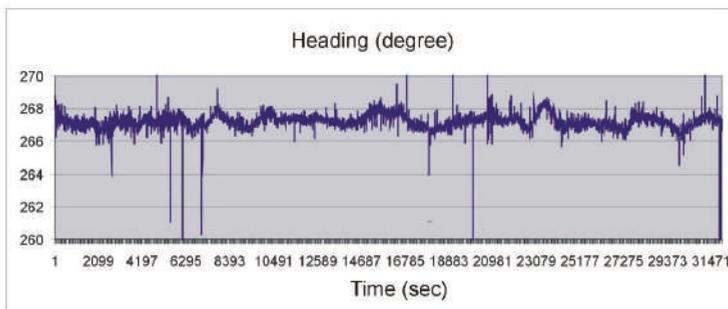


Figure 2. Heading measured with a satellite compass during stationary measurements [2].

The MX 575 compass, which as the first receiver of a certificate allowing it to be used as a heading transmitting devices, was introduced in 2005. It could be used as a backup source of heading information in IMO-compliant (International Maritime Organisation) vessels. One of the important solutions was the use of the MEMS-type (Microelectronic Mechanised System) gyroscope, which stabilizes the indications when “raw” measurements in the radio domain turn out to be temporarily inaccurate. Modern constructions often have a triad of gyroscopes and accelerometers. They are integrated systems, able to continue working for several minutes even in the event of satellite signals disappearance. The dynamic properties of such devices largely depend on the details of an algorithm used for the calculation and filtering of signals. The traditional gyro-compass has an electromechanical sensor, whose center of gravity is shifted relative to the geometric center, and thus behaves like a pendulum oscillating with a period of about 84 min (Schuler period, Schuler tuning). Maximilian Schuler made a proposal in 1923 that gyro compasses lend themselves to particularly successful tuning when the curvature of the Earth is taken into consideration. In this way, the instruments can be made insensitive to the disturbances that are caused by the result of the accelerations of the carriers along the surface of the Earth. According to this requirement, the instruments have to be tuned to an oscillation period

of 84.3 min. Thus, the classical gyrocompass has its own fluctuation with a long period. There are no kinematic problems in satellite compasses, however, the results depend on changes in the satellite constellation and the properties of the measurement-processing algorithm. This is an issue addressed by the authors of this article. Manufacturers commonly describe the quality of such devices by declaring their accuracy based on an average square error or using similar methods. The assumption of white noise may not be true. Dynamic errors are caused by dynamic factors affecting the system, such as vibrations, roll, pitch or linear acceleration. According to [3] ‘this error may have an amplitude and frequency related to the environmental influences and the parameters of the system itself’. However, for implementation in more complex measuring systems, when the fundamental issue is the selection of devices with different error characteristics, the question regarding the error frequency spectrum is important. The basic principle of integrating devices that perform similar functions is to vary the output error rate.

There are two main types of satellite compasses available on the market now: dual-antennas and tri-antennas. The most popular are dual-antenna constructions, which give the opportunity to measure two angles in transverse directions to the base between two antennas (pitch or roll in addition to heading). Designs with three antennas give the opportunity to measure all three angles of orientation of the carrier. Besides the number of antennas, there are devices with a constant distribution of antennas and movable antennas, so the distance between them can be changed by the owner or by the fitter. Additional sensors, commonly made in MEMS technology, are used to stabilize angular measurements. In addition to gyroscopes, these devices are often equipped with accelerometers. These are extremely useful for measuring the heave, which is thought to be important on small hydrographic units. In more extensive systems, there is also an option to include information from magnetic sensors or (and) a barometric measuring element. Possible block-diagram of a standard satellite compass was depicted in Figure 3.

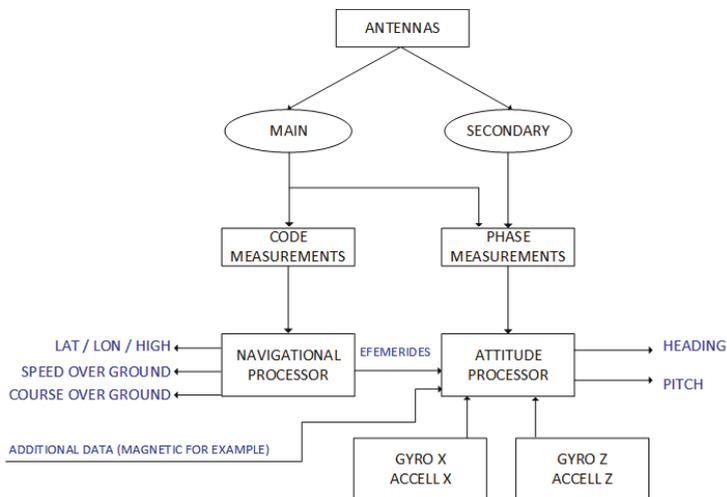


Figure 3. Possible block-diagram of a standard satellite compass. (Source: A.Felski).

Studies published in [18] proved that a satellite compass behaves similarly to a standard gyro or fiber-optic gyro (FOG) on a ship in motion. In the Figure 4 a small, systematic shift of measurements from individual devices can be noticed, however, this is due to inaccuracies in the installation of the satellite compass and FOG for the time of experiments. In general, all compasses seem to show almost the same values, and visible oscillations are probably due to imperfections in the control system and inertia of the ship. The existence of a very low frequency (Schuler tuning) characteristic of a classic

gyro compass (NAVIGAT X) is noticeable. The most changes in the heading presented in the image occur due the behavior of the ship, so they are very similar to each other.

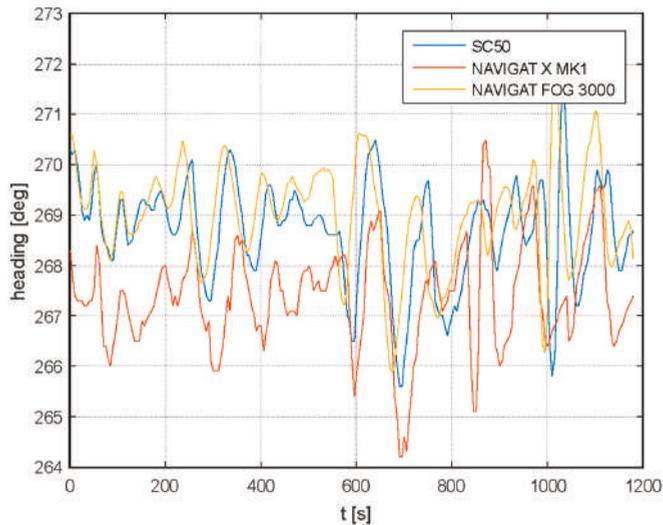


Figure 4. Example of registered headings from a classic gyrocompass (NAVIGAT X) fiber-optic gyro (NAVIGAT FOG 3000) and satellite compass FURUNO SC50 [18].

The spectral analysis of these measurements proves that low-frequency oscillations dominate, and one can also distinguish oscillations common to all three compasses, i.e., yaw resulting from the characteristics of the ships movement (0.01 Hz).

On the other hand, various frequency bands are not clearly repeated in registrations made with individual compasses. For example, significant differences occur at around 0.008 Hz for the satellite compass (as shown in Figure 5).

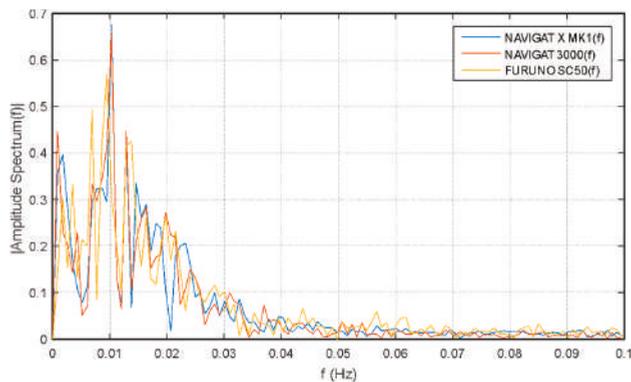


Figure 5. Single-side amplitude spectrum of oscillations presented in Figure 4 [18].

2.2. Devices

The satellite compasses of three different manufacturers were used in the experiments presented in this paper. These are Novatel PwrPak7D, the Advanced Navigation GNSS compass in a low-cost

variant and Furuno SC-50. Basic parameters, based on specifications provided by manufacturers, are presented in Table 1.

Table 1. Basic technical parameters of the three compared satellite compasses used in the experiments [19–21].

No.	Feature	Novatel PwrPak7D-E1	Advanced Navigation	Furuno SC-50
1.	Position accuracy	0.6 m (DGNS)	2.0 m or 0.6 m— (DGNS)	3–10 m (depending on corrections)
2.	Velocity accuracy	0.03 m/s	0.05 m/s	no data
3.	Roll and pitch accuracy	no data	0.4°	no data
4.	Heading accuracy	0.1° (for 2 m base length)	0.2° (base length 620 mm)	0.5° RMS (base length 430.3 mm)
5.	Heave accuracy	no data	5% or 0.05 m (whichever is greater)	no data
6.	Output data rate	Up to 100Hz	Up to 100 Hz	Up to 40 Hz
7.	Base length	adjustable	permanent 620 mm	permanent 430.4 mm
8.	Supported navigation systems	GPS L1, L2	GPS L1 SBAS GALILEO E1 BeiDou B1	GPS L1

Where: (GPS) Global Position System, (SBAS) Satellite Based Augmentation System, L1, L2—frequency of GPS, (RMS) Root Mean Square, DGNS - Differential Global Navigation Satellite System.

Novatel PwrPak7D-E1 is a robust GNSS receiver that combines dual antenna signal and (INS)—inertial navigation system hardware in a single enclosure to provide easy-to-deploy industry-leading position and heading data. In this experiments, two G5Ant-4AT1 models made by Antcom were used. According to the manufacturers, the device is suitable for ground vehicle, marine and air-based systems. Its software takes into account SPAN (synchronous position, attitude and navigation) technology based on GNSS+INS sensors as well as ALIGN software for angular determinations. It uses an OEM7720 receiver card and Epson G320N MEMS (IMU)—inertial measurement unit.

Advanced Navigation is a compact, low weight device, designed for marine and automotive applications, including small-size vehicles. It contains a 9 axis IMU that is integrated with a dual antenna GNSS receiver. Antennas are placed inside a 672 mm enclosure, together with all the signal processing electronic components. This seems to be around 3.25 lengths of the L1 wave of distance between the centers of the antennas. The core of this device is composed of two u-blox M8T GPS modules. The incorporation of all the processing components into the antennas enclosure makes it easy to integrate it even in restricted space conditions. Data are sent through the serial cable or via the (NMEA) National Marine Electronics Association 2000 network. The device is certified to be used on commercial vessels.

Furuno SC-50 is a popular satellite compass for commercial shipping. The large size of its processing unit qualifies it specifically for use on vessels. With its three-fold antenna, it can be useful for surveying ships. Clear display is designed for installation on the vessel's bridge, although data or data recorders can be sent to other devices using the standard marine format, NMEA 0183, or NMEA 2000 with the proper converters. The SC-303 antenna unit applied in tests consists of three antennas in one robust housing with a 650 mm diameter and a distance between the centers of the antennas of 430 mm. This is about 2.25 of the L1 wave length.

In summary, three different GPS receivers with different antennas and different distances between them, as well as different algorithms for angular calculations, were tested. The data were acquired with a data recording rate of 1 Hz.

2.3. Conducted Experiments

In this paper, two kinds of experiments are reported: stationary and dynamic. During the stationary part of experiments, all the antennas were situated on a building roof or bench in a suburban area. The second part of tests was conducted using an automotive vehicle driven directly out of the urban area; however, the horizon was partially by trees. In addition, we used the part of our previous experiments conducted in Gdańsk on tramway routes and published in [22] when the surprisingly low accuracy of the Furuno compass was observed. We intended to verify how significant the influence of the surroundings was on the observed errors in that investigation.

2.3.1. Twenty-Four-Hour Stationary Measurement Experiment

The stationary experiment results displayed the typical characteristics of this type of test. It was performed from 27 April 1300 UTC to 28 April 1200 UTC, 2020 in the suburban area of Gdynia, Poland, in the vicinity of a wall of a one-story building, partially obscuring the sky from the north side. The compasses' antennas were placed 4 m above the ground (Figure 6b). Data were recorded using the NMEA 0183 protocol through the RS-232 and RS-422 serial ports. The satellite constellation was assessed before data registration and the cut off angle of 20 degrees was applied. GPS satellite elevation during the test is presented in Figure 7 and their visibility is presented in Figure 8. The Advanced Navigation compass was set to the stationary variant of measurements. The following settings were applied for Furuno SC50: sampling frequency, 1 Hz; position smooth, 5 s; (SOG) speed over ground smooth, 5 s. A sampling frequency of 1Hz was set for the Novatel compass.



Figure 6. Experimental setup: (a) on a car roof for the automotive tests; (b) on a house roof for the stationary sets. Photo: K. Zwolak.

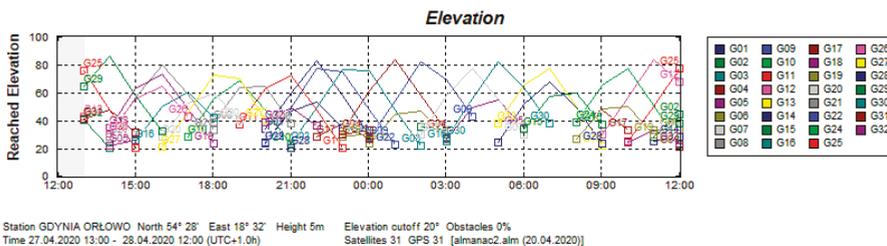


Figure 7. GPS satellite elevations during the experiment on 27–28 April 2020. (Source: Trimble Planning 2.9, 2010).

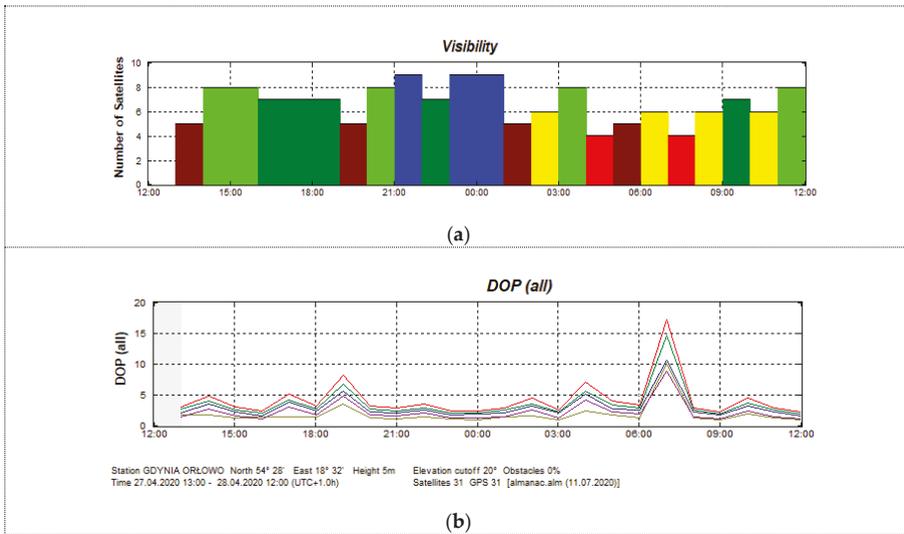


Figure 8. Visibility of satellites (a) and dilution of precision coefficients (b) in the measurement area on 27–28 April 2020. Notes: Visibility (a): red—4 satellites, brown—5 satellites, yellow—6 satellites, dark green—7 satellites, light green—8 satellites, blue—9 satellites. DOP (Dilution of Precision) (b): red—geometrical DOP, green—position DOP, blue—vertical DOP, brown—horizontal DOP, magenta—time DOP. (Source: Trimble Planning 2.9, 2010).

It must be emphasized that the heading determination needs signals from at least five satellites, in contrast to the position determination, which requires four. Only four satellites have been available twice during the data registration (this occurred at about 0300 UTC and 0600 UTC on 28 April). The largest distortions between real time and average heading value in this test were observed for the FURUNO SC50 compass (Figure 9b). The maximum heading distortion for this compass is 2.8 degrees. The offset values fluctuated between -2.0 and $+2.8$ degrees. Similar results were observed for the Advanced Navigation compass with heading distortion values in the range of -2.2 to $+2.7$ degrees (Figure 9a). The lowest values of the exchange rate distortion registered for the NOVATEL PWRPAK 7D-E1 compass, however, during the tests, its antennas were 1.2 m apart, which is twice the distance of the other two cases. The heading distortion in this case varied from -0.7 to $+0.8$ degrees (Figure 9c).

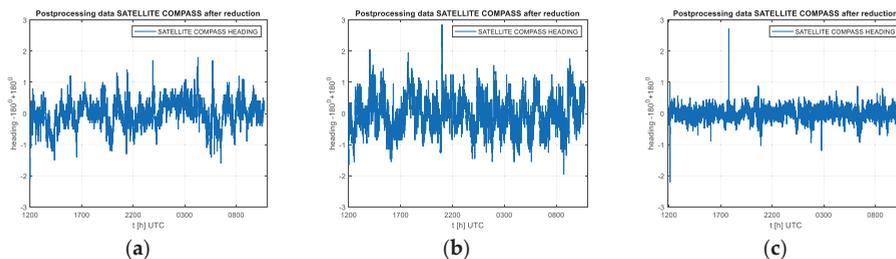


Figure 9. The example subsets of stationary heading registrations by the Advanced Navigation compass (a), the Furuno compass (b) and the Novatel compass (c) (27, 28 April 2020).

The root mean square of the heading distortion for the Furuno compass is 0.6 degrees, for Advance Navigation is 0.4 degrees and for Novatel is 0.2 degrees.

In order to perform the spectrum analysis of the signal in the frequency domain, the presentation of the frequency band in the range above $f = 0.1$ Hz was abandoned due to the negligible variability of the signal amplitude—heading distortion, which is typical for stationary measurements. Stationary registration results are very similar for all three compasses, characterized by a very low frequency of heading changes, falling in the band lower than 0.02 Hz. However, the amplitudes of these changes vary. The maximum value for the Furuno compass was 0.21 degrees, for Advanced Navigation 0.19 degrees, and for the Novatel product, it was only 0.08 degrees. Undoubtedly, this is due to the length of the base line between antennas, but it can be assumed that this is also the result of the different method of filtration or azimuth calculation. Heading registration spectrum for the three compasses are shown in the Figure 10.

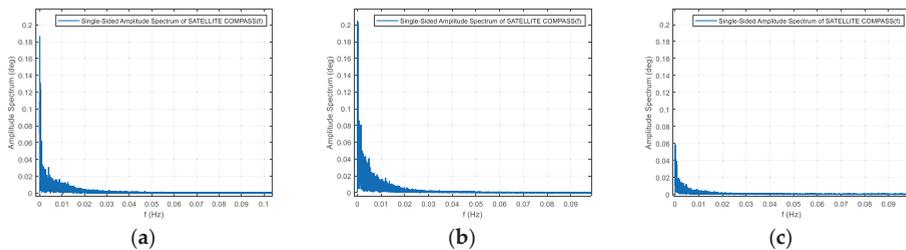


Figure 10. Heading registration spectrum for the three compasses: (a) Advanced Navigation, (b) Furuno, (c) Novatel. Stationary experiment, 27, 28 April 2020.

2.3.2. Automotive Experiments

The analysis of heading oscillations in compass indications in stationary conditions were performed based on calculating distortions from the average value. In this case, the direction of the compass does not matter. In dynamic conditions, the reference direction is needed. Therefore, the tests were conducted in such conditions that the direction of movement of the object was known and determined by natural conditions, i.e., on a straight sections of road or tram track. Knowing the heading of the vehicle during movement, the distortions of individual readings and the oscillations were calculated, treated as corrected measurements, and analyzed using a Fourier analysis. Matlab scripts were written to perform the analysis.

Automotive experiments, with the antennas mounted on the roof of the car, were carried out on a straight section of a rural road with a length of 1550 m and a direction of 342/162 degrees. There are single tall trees in the central part of the test section, along the road, and from the east side, which can occasionally cause interference. This is visible in the Figure 11 in the form of a break in position data registration due to incidental obstruction of the satellite signal. Such a gap is a result of a specific configuration of the satellites during this test. During other tests, similar gaps occurred in other places. Unfortunately, it was not possible to guarantee a repetitive configuration of the satellites, however, these records can be treated as examples of how important and diverse the impact of obstructions on the work of such compasses can be. The devices have options to adjust to the vehicle movement, that is, the Novatel compass has “sampling frequency: 1 Hz” and Furuno has “position smooth, 1 s; SOG smooth, 1 s; sampling frequency, 1 Hz”. In addition, it is worth noting that the compasses have advanced inertial systems for the stabilization of readings, but this did not ensure the complete elimination of rapid changes at the time of appearance of another configuration of the satellites received by the device due to the appearance of obstructions.

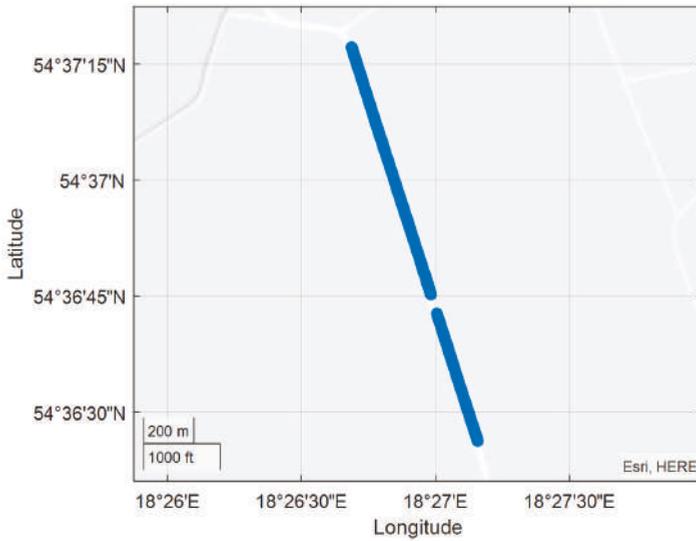


Figure 11. Positions recorded during the road test with the gap in data registration visible on the map view.

Measurements were carried out at speeds of 10, 20 and 30 km/h. Raw heading records for the compasses used in this part of the experiment are presented in Figure 12 for the Furuno compass and Figure 13 for the Novatel compass.

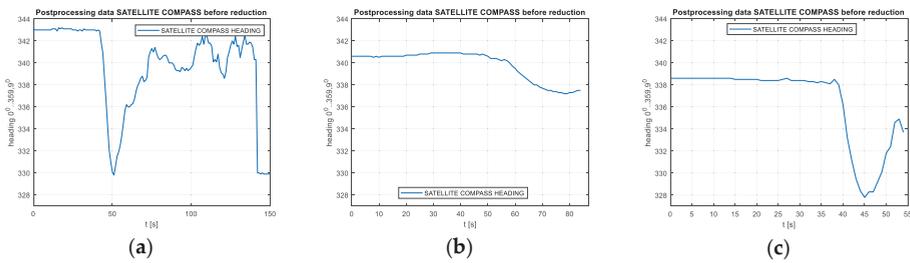


Figure 12. Raw heading records for the Furuno compass for the speeds of 10 km/h (a), 20 km/h (b) and 30 km/h (c).

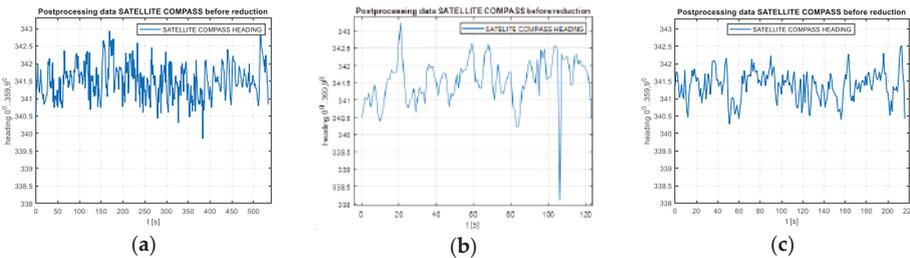


Figure 13. Raw heading records for the Novatel compass for the speeds of 10 km/h (a), 20 km/h (b) and 30 km/h (c).

The frequency spectrum of the signals presented above are plotted in Figures 14 and 15.

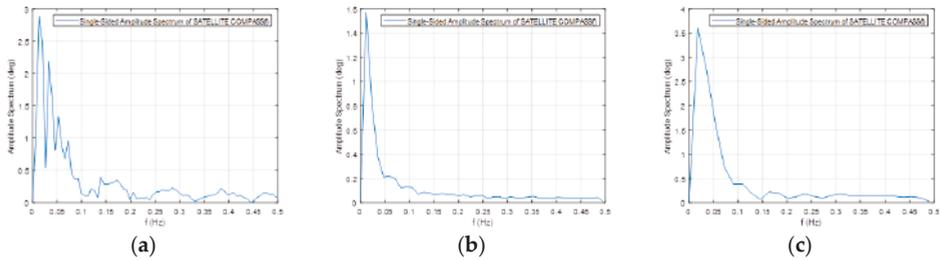


Figure 14. Frequency spectrum for the Furuno compass raw heading records in Figure 12 for the speeds of 10 km/h (a), 20 km/h (b) and 30 km/h (c).

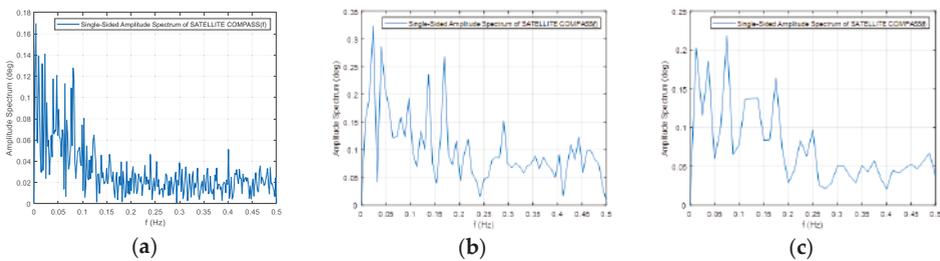


Figure 15. Frequency spectrum for the Novatel compass raw heading records in Figure 12 for the speeds of 10 km/h (a), 20 km/h (b) and 30 km/h (c).

In the context of a reaction to rapid changes in the satellite constellation, there is also a question regarding the impact on the stability of compass indications based on inertial sensors that are able to support the work of the radio (satellite) segment [23]. An example of the behavior of the Advanced Navigation compass in the case of complete obscuring of satellite signals (667 s after the start of registration) is shown in Figure 16. A clear drift of values is observed, which was similar in other tests, although the directions of the drift were different. After approximately 100 s, no information about the heading was reported by the device.

2.3.3. Tests on Tram Rails

The tram experiment was performed on 28 November 2018 along the route indicated in Figure 17 with the use FURUNO SC50 only. The experiment was conducted in Gdansk on a several-kilometer tram rail with variable sky visibility conditions, with the aim of assessing the performance (accuracy) of the satellite compass operation in non-standard terrain conditions. The measuring instrument used in the experiment was placed on a trolley of the DWF 300 series tram and pulled behind a tram [22] above the tram rails axis. The task of this measurements was more complex, and we now use only a small part of this registration made on the rail part, characterized by a constant direction (Figure 17).

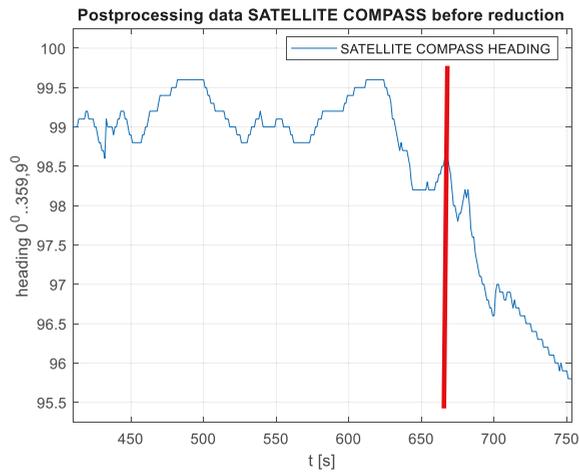


Figure 16. Raw heading value results during the stationary test with the Advance Navigation compass used as an example. The orange line denotes the satellite signals in the moment of being completely obscured.

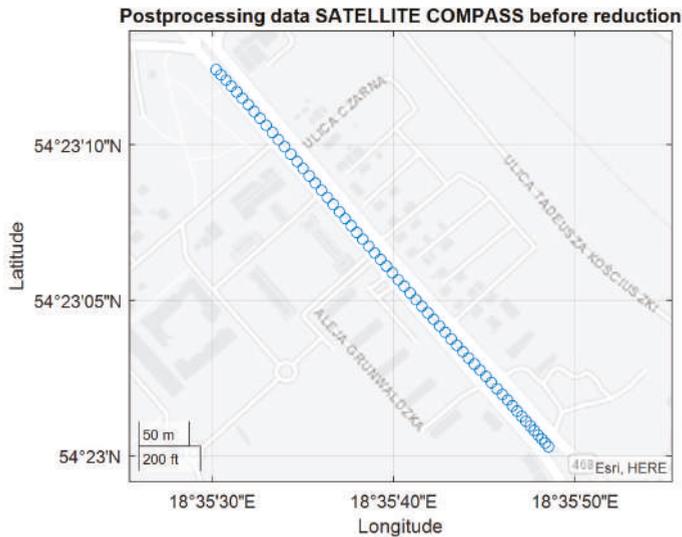


Figure 17. Position registration of the tram from 23:06:25 (UTC) - Universal Time Coordinated to 23:07:25 UTC.

For this analysis, it is important that the tram route ran through an urbanized area and some sky obstructions were observed during the registration. The GPS satellite elevations are presented in Figure 18.

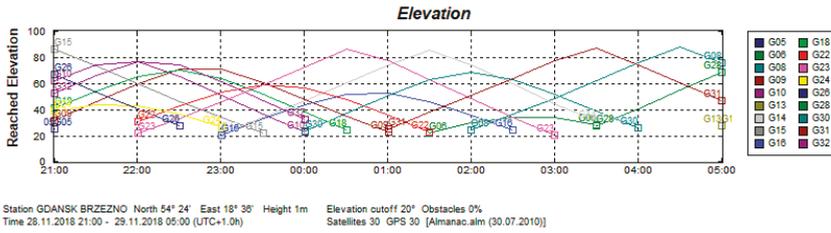


Figure 18. GPS satellite elevations on 28, 29 November 2018. (Source: Trimble Planning 2.9, 2010).

The satellite compass requires a signal from at least five satellites for each antenna to determine vehicle heading. Based on Figure 19, it can be seen that a condition of a visibility of at least five satellites to quantify the heading of the vehicle with the arbitrarily assumption of the elevation cut-off at 20 degrees has been met during the experiment.

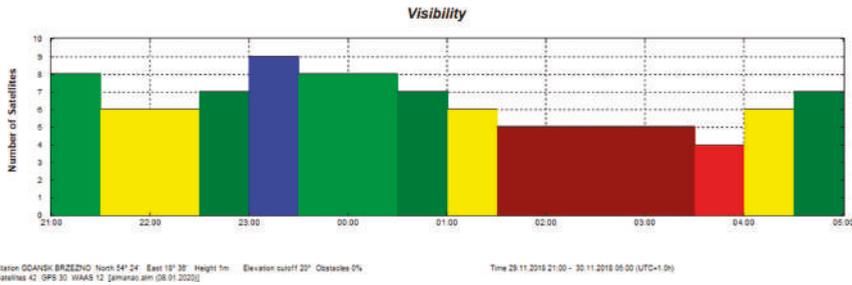


Figure 19. GPS satellite visibility on 28, 29 November 2018. (Source: Trimble Planning 2.9, 2010).

The presented test was started at 2200 UTC with the visibility of six GPS satellites (Figure 19). The problem occurred when the number of visible satellites was reduced to four (from 0230 UTC to 0300 UTC on 29 November 2018). Based on the data registered from 22:59:27 UTC on 28 November 2018 to 03:26:41 UTC on 29 November 2018, the parts of the straight tram rail section have been chosen. Data have been recorded from the 419th second to 479th second of the run, which is from 23:06:25 to 23:07:25, from the position (LAT) Latitude: 54.386780° N, (LON) Longitude: 018.591723° E to the position LAT: 54.383415° N, LON: 18.5968183° E. Heading oscillations for a vehicle on tram rails were observed in the range of -0.8 . to $+0.5$ degrees. The values of the heading distortions and the frequency spectrum of the heading record changes are presented in Figure 20.

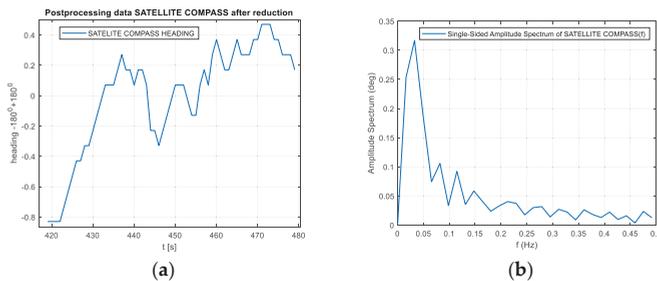


Figure 20. (a) Heading distortions. (b) Frequency spectrum of the heading records. Heading distortions and the frequency spectrum of the heading records from 23:06:25 to 23:07:25, from the position LAT: 54.386780° N, LON: 018.591723° E to the position LAT: 54.383415° N, LON: 18.5968183° E. (Source: K. Jaskólski).

The heading distortion analysis in the time domain in Figure 20a confirms the declared accuracy of the device indications in accordance with the technical specification of the device, which is 0.51 degrees (RMS)—Root Mean Square error. The spectrum analysis of the signal in the frequency domain in Figure 20b differs from that recorded during the stationary tests because oscillations appear at frequencies higher than 0.02 Hz. The maximum amplitude is slightly higher than that during the stationary record.

Another example of a registration on tram rails is shown in Figure 21 and research scores with a few course deviations are shown in Figure 22. In the heading record, there are four observed significant distortions from the track direction, which result from sky obstructions caused by high buildings in the vicinity of rails. The spectrum of this record differs significantly from others, which is undoubtedly caused by these four clearly distorted parts.

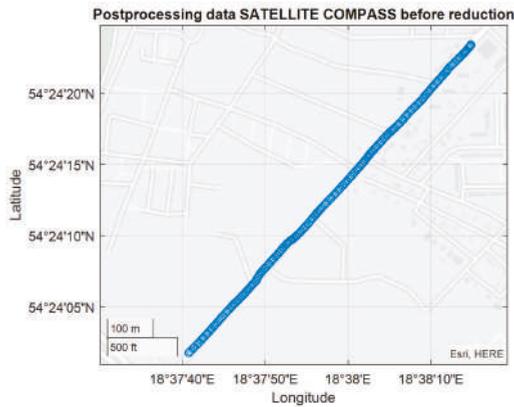


Figure 21. Position registration of the tram from 00:25:17 UTC to 00:30:37 UTC.

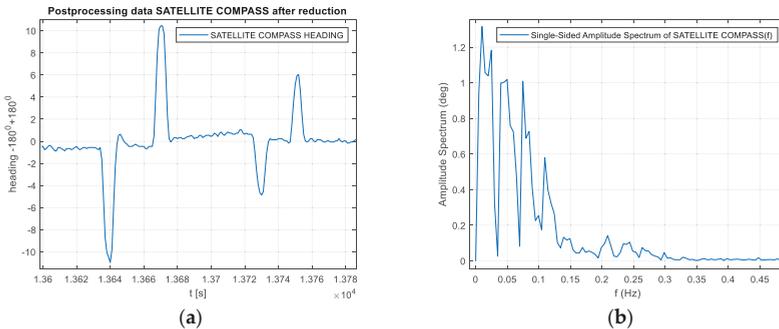


Figure 22. Heading distortions from 00:25:17 UTC to 00:30:37 UTC (a) and a frequency spectrum of heading changes (b).

3. Results

Studies have confirmed the different accuracies of the devices used in the experiment. It can be assumed that compasses with a fixed antenna system (Furuno and Advanced Navigation), built primarily for seagoing vessels, have an RMS error of approximately 0.5 degrees. The Novatel compass with adjustable distance between the antennas was observed to have higher accuracy, but this is obvious due to the fact that a longer antenna base line (over 1 meter) has been used.

It was confirmed that changes in the constellation of satellites accepted for the solution are the reason for the oscillations occurring in heading registrations. Typical frequencies appearing in the error

spectrum are very low, but less than 0.02 Hz for objects in movement. These oscillations are caused by changes in the set of tracked satellites and the satellites included in the calculations. For moving objects, oscillations occur due to changes in the orientation of the vehicle, as well as unpredictable changes occurring as a result of obstructing the satellites by obstacles in the environment.

4. Discussion

The motivation to conduct the research was the authors' experience, published in [18], and especially the results described in the paper [22], in which the results of Furuno compass errors turned out to be surprisingly high. In this paper, the authors were interested in a solution that would be potentially useful on a small floating object used for hydrographic measurements. The issue of the accuracy of the position determined by such a device was purposely not analyzed here, because today it seems to be a trivial issue and comes down to the choice of support service (augmentation) of the GNSS system. This text focuses on the issue of accuracy of the heading, based on the knowledge that the accuracy of the heading measured with each type of compass has different dynamic characteristics. An assessment of a measurement of heading uncertainty on a moving object requires taking into account changes in object orientation angles, because the readings will include both this information and the inaccuracy of the compass or any other gauges. For this reason, three different compasses were tested during static and dynamic experiments. Dynamic experiments, with heading reference data included, were carried out at different speeds on straight sections of tram rails and on straight sections of a roads.

The results of stationary experiments confirm the clear relationship between the antenna base line length (distance between antennas) and heading accuracy. Analyzing the compass documentation, this relationship can be presented in a form of a curve, presented in Figure 23. The curve presented here is obtained by interpolating the sparse data regarding the Novatel compass with the quadratic function. The red square represents the manufacturer's information regarding the Furuno compass, which is consistent with data on Novatel. However, Advance Navigation data deviate from this relationship. The results of stationary tests confirm the declarations of the Novatel and Furuno compass manufacturers, although the Advance Navigation compass tests showed an error of twice the value—0.4 degrees.

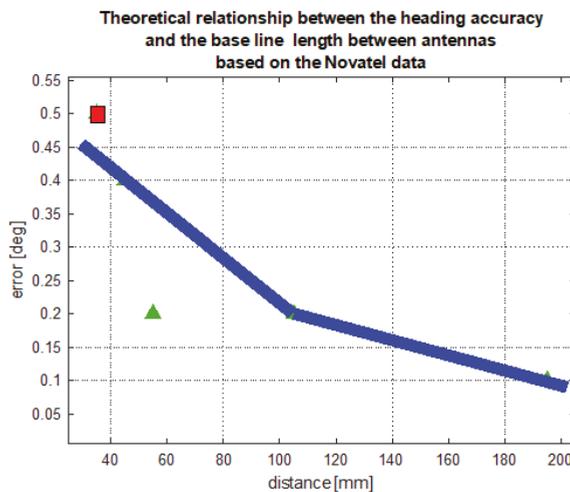


Figure 23. Theoretical relationship between the heading accuracy and the base line length between antennas, based on the Novatel data published in the device documentation (**black curve**). Declared accuracy of the Furuno compass as a function of baseline (**red square**) and of the Advanced Navigation compass (**green triangle**).

For dynamic applications, knowledge of the error spectrum is extremely important. This spectrum is characterized by the dominance of very low frequencies, which is understandable with the use of filtration and inertial sensor support. One can clearly indicate the 0.02 Hz value as the upper limit of these oscillations for all tested compasses. However, in dynamic conditions, higher frequencies appear, which result from slight changes in the orientation of the object during movement. In addition, there are also harmonics, which are in fact the results of rapid deviations of results from previous values. The reason is that despite the solutions using inertial sensors and filtration, these compasses are still sensitive to changes in the temporary constellation of satellites used for calculations. Even during the stationary tests, one can easily identify the relationship between changes in indicated heading values and changes in the satellite constellation. This statement is confirmed by the analysis of the devices' behavior at the time of signal obstructions. It is always associated with a sharp change in the value at the compass output, although the natural change in the set of satellites seems to have less effect on the results than the short-term screening of one or two satellites, which disappear after a few seconds.

The suppositions that appeared in relation to the results presented in [22], that the accuracy of the entire population of the recorded results do not reflect the whole truth regarding the compass, were also confirmed. It is necessary to take into account the effects of appearing obstructions, and the user should be aware of the effects of the nearby obstructions of the operation of the satellite compass.

The mentioned jumps in the heading measurement for a particular compass cannot be clearly determined with regards to time and amplitude. These changes, although similar and occurring almost at the same time in different compasses, are not identical. Sometimes they are slightly shifted in time for different compasses, relative to the change in constellation. This is clearly seen in Figure 9, where synchronous registrations made in the same place are shown, and sharp changes occur at different times. For example, after 1700 and just before 2200, the sharp changes can be observed in all three registrations at almost the same moments, but their amplitudes differ. There is no rule that one particular compass always shows larger changes. These changes can be considered as a typical filtration effect used in the integrated system. This kind of effect also appears in the coordinate indications referred to by the GNSS receiver, which is a separate element of the tested device. However, the code principle of a position measurement results in fundamental differences from the phase principle of angle measurement. Therefore, during measurements, especially in motion, there were several cases observed in which the heading indications were less accurate or even incorrect, while the coordinate values were correct. Such cases occur when, due to the presence of the obstructions, the configuration of the satellites has decreased to four, while setting the heading requires observation of signals from five satellites. Practitioners using such devices should pay attention to this fact. It is common to observe the values of DOP and (HDOP) – Horizontal Dilution of Precision as indicators of the quality of the receiver's work, while this refers to the position and is not true in relation to the heading (Figures 8 and 9). It seems that for applications, such as hydrographic vessels, it would be advisable to propose a new indicator to facilitate the control of heading information quality in satellite compasses.

5. Conclusions

Satellite compasses have been known for about 20 years and certainly now they can be treated as well-developed solutions. They are becoming increasingly popular in a wide range of applications, including shipping and aviation. They are particularly attractive in controlling self-propelled robots and machines. Their small size and light weight make them very attractive for small autonomous vehicles, except for underwater devices.

At the same time, it is obvious that, like any technical solution, satellite compasses have some limitations. The key is the dependence on satellite signals, which, in the case of the occurrence of terrain obstructions, becomes a significant limitation on land or in inland waters, where the near-shore objects often obstruct satellites. This aspect, however, also appeared recently in the context of potential interference with GNSS signals or spoofing. The literature and media report many cases of interference

with satellite signals, even in the sea and in the air. This leads to the question of the legitimacy of the full dependency of navigation systems only on a satellite compass in terms of heading. This aspect is particularly important in the context of planned offshore autonomous vessels, where the risk of interference with satellite signals can cause many complications, not only in the context of the position, but also the heading and effective satellite communication.

When assessing the accuracy of satellite compasses, it should be noted that for several years, they have been constructed as systems supported by inertial sensors (gyroscopes and accelerometers), and now also by other sensors, such as magnetic sensors or those based on pressure. As a result, integrated systems are created, whose measurement properties largely depend on the structure of the data-processing algorithm. This, in turn, takes into account the purpose of the designed system. When assessing a modern satellite compass, one should take into account the branch of applications for which the device was constructed. There are different expectations for the fishing boat and the quadcopter. The requirements may even be different for a ship performing hydrographic measurements in river estuaries, and different for a ship performing similar measurements in the middle of the North Sea or the Gulf of Mexico.

The experiments presented here focused primarily on the spectrum of frequencies that appear in the recorded results of the heading measurements by the mean of satellite compasses. It was confirmed that the reason for the occurring oscillations is the changes in the constellation of satellites accepted for a heading solution. Typical frequencies appearing in the error spectrum are very low, below 0.02 Hz; however, the movement of the object on which they are installed, as well as obstacles causing obscuring satellites result in rapid changes in measurements values, which is evident in the registration of the occurrence of various higher frequencies in the spectrum of heading changes.

For people using such devices, importance should be placed on being aware that the quality of the satellite compass and the quality of the positioning receiver depend on various factors. Therefore, although they are in the same device and many users treat them as one device, one cannot draw conclusions about the quality of the course on the basis of indicators resulting from DOP, which characterize only the positional service. Geomagnetic storms and traveling ionospheric disturbances (TIDs) are also known as sources of GPS positioning quality deterioration. GPS scintillations lead to a range of errors in GNSS due to diffraction [24]. Deep signal fades that appear during small-scale irregularities effect the result in navigation outages [25,26]. Satellite compasses are very popular in high-latitude regions, however, the aurora borealis becomes especially visible in such regions, and the effect of the Earth's ionosphere on GNSS signal propagation (total electron content) is one of the main error sources which limits the accuracy and reliability of GNSS applications [27].

Author Contributions: Conceptualization, A.F. and K.J.; methodology, K.J.; software, K.J. and P.P.; validation, K.J., K.Z.; formal analysis, K.J.; investigation, A.F. and K.J.; data curation, K.J. and K.Z.; writing—original draft preparation, A.F., K.Z., K.J.; writing—review and editing, A.F., K.Z., K.J.; visualization, K.J., P.P.; supervision, A.F., K.Z., K.J. All authors have read and agree to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the K2sea Company, for making the Advanced Navigation satellite compass available for testing.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Felski, A. Exploitative properties of different types of satellite compasses. *Annu. Navig.* **2010**, *16*, 33–40.
2. Felski, A.; Nowak, A. Practical Experiences from the Use of the Satellite Compass. In Proceedings of the ENC Conference, Parthenope University in Naples, Naples, Italy, 4–6 May 2009.
3. Satellite Compass. A Practical Guide to GNSS Transmitting Heading Devices for Marine Navigation. Northrop Grumman. 2005. Available online: <https://www.sperrymarine.com/PageResources/767/NavistarGuide.pdf> (accessed on 13 April 2020).

26. Yasyukevich, Y.; Vasilyev, R.; Ratovsky, K.; Setov, A.; Globa, M.; Syrovatskii, S.; Yasyukevich, A.; Kiselev, A.; Vesnin, A. Small-Scale Ionospheric Irregularities of Auroral Origin at Mid-latitudes during the 22 June 2015 Magnetic Storm and Their Effect on GPS Positioning. *Remote Sens.* **2020**, *12*, 1579. [[CrossRef](#)]
27. Warnant, R.; Lejeune, S.; Bavier, M. Space Weather Influence on Satellite-based Navigation and Precise Positioning. In *Space Weather. Astrophysics and Space Science Library*; Lilensten, J., Ed.; Springer: Dordrecht, The Netherlands, 2007; Volume 344. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Sensors Editorial Office
E-mail: sensors@mdpi.com
www.mdpi.com/journal/sensors



MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34
Fax: +41 61 302 89 18

www.mdpi.com



ISBN 978-3-0365-0427-8