IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 18, 2020, accepted June 26, 2020, date of publication June 30, 2020, date of current version July 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3006003

An Acoustic-Based Recognition Algorithm for the Unreleased Braking of Railway Wagons in Marshalling Yards

YULING YE[1], JUN ZHANG[1], AND HENGDA LIANG[2]

[1]Key Laboratory of Road and Traffic Engineering of the State Ministry of Education, Shanghai Key Laboratory of Rail Infrastructure Durability and System Safety, College of Transportation Engineering, Tongji University, Shanghai 201804, China
[2]Department of Logistics Planning and Control, SAIC Volkswagen Automotive Company Ltd., Shanghai 201805, China

Corresponding author: Jun Zhang (zjvicky@tongji.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1201403, and in part by the Scientific Research Project of China Railway Shanghai Group Company Ltd., under Grant 2018005.

ABSTRACT The abnormal braking of wagons is a challenging safety problem for operators at railway marshalling yard. This paper develops an acoustic-based technology to detect the unreleased braking of wagons during the uncoupling operation. Experiments have been conducted to collect the acoustic waves of wagons abnormal braking, as well as the background sounds like train whistling and wheel vibration. Before data collection, a wayside recording system and an experimental train composed of 5 different wagons have been prepared in the marshalling yard. The recognition algorithm consists of fast Fourier transform (FFT), feature extraction, template matching and support vector machine (SVM) classification. Based on the sample data of different acoustic waves, the FFT is firstly performed to obtain the frequency spectrum from original time-domain signals. Then the major spectrum features of different sounds are carefully extracted for SVM training through a newly-devised algorithm, where the features include the spectrum center, spectrum flux, energy peak and corresponding frequency. During the SVM training, classifiers are designed under the one-against-one strategy to guarantee the recognition accuracy. Given a test data, at most 3 SVM classifiers will be activated according to the decision matrix of template matching. Meanwhile, rules have been made to regulate the classification result considering different activation cases. Finally, a case study of all 12 sound categories has been performed to illustrate the application of proposed algorithm. Results show that the acoustic-based recognition algorithm is indeed reliable to identify wagons unreleased braking, with the global warning accuracy over 98%.

INDEX TERMS Railway safety, unreleased braking, acoustic recognition, feature extraction, SVM classification.

I. INTRODUCTION

In recent years, kinds of intelligent information technologies have been applied in the fields of railway management and maintenance, such as locomotive rescheduling [1], conflicts detection [2], and traction fault diagnosis [3]. Specially, wagons unreleased braking is a challenging safety problem for operators at railway marshalling yard, which is always a threat to workers' personal safety and infrastructures lifecycle. In order to guarantee the safety of uncoupling trains on the hump platform, the demand for an automatic monitoring and warning system of the unreleased braking is increasing. Therefore, researches on relevant recognition methods are of great significance in improving the quality of safety detection during wagons movement.

The unreleased braking of wagons, also known as a kind of brake faults or abnormal braking, is affected by different factors. Aside from human errors in daily uncoupling and maintaining, mechanical failure is the main influence factor. Under a conventional air braking system, it is found that the propagation of brake application is slow for long and heavy-hauled freight trains [4]. Specially, Zhang *et al.* [5] pointed out that the increased inertia and vertical force of heavy wagons would cause a difficulty in releasing brake operation. Due to the disequilibrium of braking force, the brake slack adjuster would get locked and come into a brake seizure [6].

The associate editor coordinating the review of this manuscript and approving it for publication was Jesus Felez.

Furthermore, a freight train is usually a mixed marshalling of different wagons, including the flat car, tank car, open-top car, box car, etc., while the pneumatic brake valves of different wagons may have the different releasing performance and operation quickness [7].

During trains uncoupling on the hump platform, the sound from wheelsets will become different when occurring an unreleased brake, due to a conversion from rolling friction to sliding friction [8], [9]. Generally, the unreleased braking will cause wheel tread damages [10] together with the noise pollution [11], [12], and may lead to wagons dragging or derailment in different degrees [13], which makes it necessary to develop methods to recognize the unreleased braking proactively. Therefore, the major objective of this study is to propose a recognition method for the unreleased braking of uncoupling wagons in the marshalling yard. The main contributions of this research are summarized as follows.

– First, a wayside acoustic acquisition system is presented to detect the unreleased braking of uncoupling trains in railway marshaling yard. The acoustic data of 5 abnormal braking sounds and 7 normal background sounds collected through repeated experiments are stored into the system database, in order to support the acoustic recognition.

– Second, a hybrid algorithm framework of acoustic recognition is proposed as an integration of the off-line data training and the real-time classification. During the algorithm programming, codes of feature extraction, template matching and SVM classification rules are differently designed considering the application context.

– Third, representative case studies have been conducted on real acoustic datasets to validate the performance of the proposed algorithm. The application of this acoustic-based recognition algorithm is competent to solve the problem of abnormal braking detection with a stable performance and a reliable prediction accuracy.

The remainder of this paper is structured as follows. Section II discusses prior studies in brake fault diagnosis, railway noise recognition and acoustics application. Section III introduces the experimental scheme of data collection and processing. Based on fast Fourier transform (FFT), Section IV develops a hybrid recognition algorithm of template matching and SVM identification considering the application context. The detailed case analysis and discussion are presented in Section V and Section VI respectively, both the feasibility and reliability of presented algorithms have been validated. Finally, the study is concluded by highlighting major contributions and possible future research in Section VII.

II. RELATED WORK

Current methods of detecting train brake faults mainly depend on the onboard equipment and sensors. Onboard sensors have been frequently adopted and integrated in the monitoring system to realize the condition diagnosis of freight railway vehicles [14]. In the respect of train braking faults detection, Lonsdale and Wilson [15] suggested using on-board handbrake sensors to monitor the braking operation and give warning signals. Considering the air pressure of pneumatic pipes and air cylinders, Lu and Zhang [16] designed an alarm circuit integrated by pressure switches, air sensors and other devices. Besides the brake cylinder pressure, Aimar and Somà [17] further considered the brake block temperature and developed a prototype of the onboard unit to monitor the brake system. Based on the multi sensor data, Lee [18] presented a contextual air leakage detection based on the idle time and run time of compressor behaviors. Yang *et al.* [19] discussed the detection of pneumatic brake system through the leakage diagnosis algorithm.

Apart from the detection methods by sensing interior mechanical units like the pneumatic pipe and the hand brake, it is also feasible to recognize the abnormal braking through outward manifestations like heat and sound. At the marshalling yard, experienced workers can tell the braking anomalies by visual observation and hearing sensation. Gutierrez and Garrido [20] applied an infrared thermography to analyze the thermal characterization of heated brake rotors under strong frictional stresses. The similar visualization technology using grey projection was applied to identify brake faults by checking the mechanical structure [21].

The acoustic recognition was proved to be an effective approach in train faults diagnosis. Cerullo *et al.* [22] applied the acoustic signal processing techniques to continuously diagnose the train system including the pantographs, rails, wheels, bogies and etc. Pronello [23] carried out on-site acoustic measurements to identify the noise emission of rail traffic, considering the train type, running speed and site configuration. It turned out that the acceleration/deceleration of diesel trains have stronger noise emissions than other conditions. Huang *et al.* [24] designed a uniform rectangular array of microphones to capture the wayside acoustic signal, and improved the diagnosis accuracy for wheel bearing health through an envelope spectrum analysis. Similarly, Jiang [25] demonstrated the possibility and feasibility of using the acoustic energy ratio in detecting fatigue cracks of axles. The railway noise can also be applied in the detection of rail-head roughness and rail structural health. Jones and Packham [26] studied the distribution characteristics between the rolling noise and the train speed by using a microphone mounted near a smooth wheel, where the noise ranges at different speeds can support the identification of rail-head roughness. Janeliukstis and Kaewunruen [27] found that the acoustic sources of railway noise can be separated from each other, and adopted a decision tree classification algorithm to monitor the structural health of railway pre-stressed concrete sleepers.

The acoustic analysis has also been extensively applied in many other engineering fields. Based on the acoustic data of heavy machinery, Aguilar *et al.* [28] proposed a recognition method to detect risks during the underground mining project. Zhao *et al.* [29] developed an acoustic signal processing system to measure the running state of transformers. Cao *et al.* [30] analyzed the acoustic features during municipal

excavation to protect underground pipelines. Besides, the acoustic recognition can also be applied in the identification of military jets and heavy vehicles [31], [32]. Before acoustic recognition, discrete Fourier transform (DFT) and wavelet transform (WT) are fundamental processing methods to obtain the spectral distribution from original time-domain signals [33], [34]. As to the recognition methods, dynamic time warping (DTW) [35], fuzzy matching [36] and support vector machine (SVM) [37] are major algorithms used in the classification and identification of acoustic data.

From the foregoing literatures, great efforts have been done in train braking faults identification and railway noise recognition for daily operation and maintenance. Current approaches mainly focus on monitoring the braking mode, cylinder pressure and block temperature of brake system for trains running on lines, while few studies have discussed the feasibility of acoustic-based recognition for the unreleased braking of uncoupling trains in the marshaling yard. Compared with onboard sensors, the wayside acoustic recognition has the advantages of non-contact measurement, low cost and easy mounting.

III. DATA COLLECTION

During the daily operation at a railway marshaling yard, the unreleased braking of freight wagons is a random event, which brings about uncertainty and difficulty in collecting the acoustic data of abnormal braking just by installing a wayside sound recording system. Therefore, we have to design experiments to get the acoustic data of this abnormal braking, where the brake valves of different wagons are controlled manually by experienced workers.

A. PRELIMINARY WORK

Before starting experiments, necessary preliminary works have to be done, including the setup of wayside recording system and the preparation of experimental wagons. The wayside acoustic acquisition system is composed of a laptop, a portable power, an external audio adapter, an audio pickup device and transmission cables, as shown in Figure 1. The audio pick-up is used to capture kinds of acoustic signals from freight trains, powered by the portable power supply with 12V DC output. The audio adapter is used to deal with the conversion from electrical signals to digital signals. To guarantee the efficiency of original data acquisition and storage, a simple monitoring system has been developed based on the browser server architecture (B/S).

In the railway marshalling yard, five kinds of wagons are frequently used, including the flat car, tank car, open-top car, box car and JSQ car. Besides, a DF-7C type diesel locomotive is operated to provide the necessary traction force. The wagons and tractor prepared by the marshalling yard are shown in Figure 2.

B. EXPERIMENTAL SCHEME

The experiment was performed in a railway marshaling yard at East Wuhu Station, under the administration of

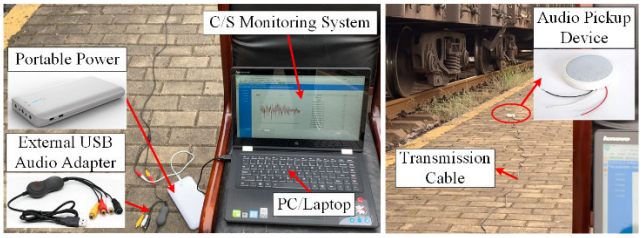


FIGURE 1. The setup of wayside acoustic acquisition system.

FIGURE 2. Experimental wagons and the diesel locomotive.

Shanghai Railway Bureau. In order to reduce the impact to daily uncoupling tasks, 1:30 p.m. to 3:30 p.m. was chosen to be the best experiment period of the day. Meanwhile, the weather should be fine, because the sound of strong wind or heavy rainfall will affect the acquisition quality of acoustic waves. The experiment time is from 10th to 12th April, 2019.

As mentioned in the literature review, different wagons have different pneumatic brake valves. Therefore, the corresponding sounds of abnormal braking are distinguished from each other. Under the guidance of operators, 5 types of wagons (one for each type) were marshalled into an experimental train together with a DF-7C locomotive. The experimental train moves back and forth on the hump platform as one movement, and each movement will provide two acoustic signals. It should also be noticed that the traction speed is limited between 6 to 8 km/h, which agrees with the normal pushing speed of uncoupling trains on the hump. During every

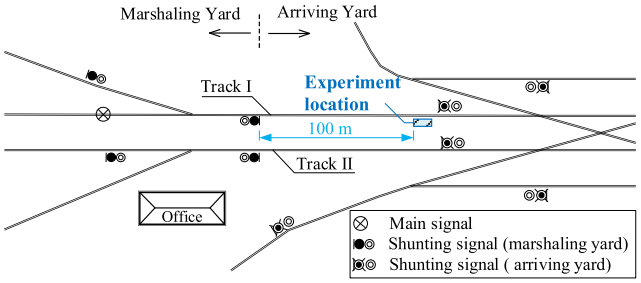


FIGURE 3. The layouts of hump platform area and the selected experiment location.

back-and-forth movement, only one wagon's brake valve is set unreleased manually to simulate the abnormal braking.

Another important consideration is the experimental location. Figure 3 shows the general layouts of the hump platform in the yard, the place where workers uncouple freight trains. Since the experimental train plans to move on Track I, the wayside recording system is installed closed to Track I. The abnormal braking should be identified before the wagon humping into the marshalling yard, which means enough time and distance should be reserved for the recognition procedure. Therefore, the experiment place is located 100 m ahead of the first shunting signal at the marshalling yard, as shown in Figure 3.

In addition to the collection of abnormal braking sounds, acoustic data of surrounding environments should also be collected to avoid errors of misidentification. The background sounds usually come from the adjacent uncoupling trains, the passing-by trains on main lines, the inertia vibration before stop, the roaring of locomotive's engine, the train whistling, the wheel vibration at rail slots, etc. According to the on-site situation, the acquisition of background sounds can be performed at the hump platform without the experimental train.

C. DATA PRE-PROCESSING

The pre-processing of original acoustic data includes a transformation and a selection. Taking an acoustic sample of tank car as an example, the original wave data is a time-domain signal, where the amplitude represents the voltage, as shown in Figure 4(a). Because the sampling frequency is 48000 Hz, the acoustic wave looks rather compact, where the unreleased braking appears at about 9 s to 11 s. To catch more characteristics of acoustic signals, the method of FFT has been applied. Through FFT, the time-domain signal graph can be efficiently transformed into a spectrogram, see Figure 4(b), where the vertical ordinate represents the sound frequency and the darkness represents the energy grade of waveform. It is obvious to find the frequency distribution characteristics of different sounds, where the major frequency distribution of abnormal braking consists of 1000 Hz, 2100 Hz, 3200 Hz, 4400 Hz and 5500 Hz.

An original acoustic sample usually contains two or more sounds, such as the sounds of train whistling and abnormal braking shown in Figure 4(b). Therefore, each kind of sound

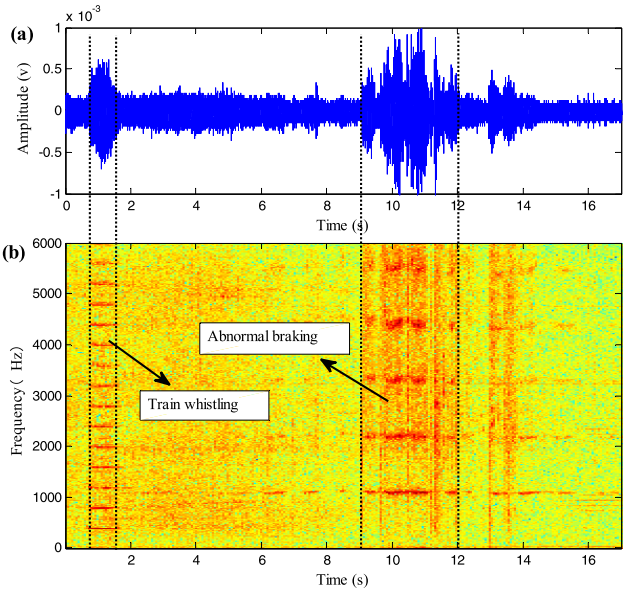


FIGURE 4. (a) The original acoustic signal waveform of a tank car; (b) The corresponding acoustic spectrogram after FFT.

TABLE 1. Basic information of selected acoustic samples.

Type	Category of sound source	Sample size	Data size
1	Abnormal braking of JSQ car	10	513 × 812
2	Abnormal braking of flat car	3	513 × 113
3	Abnormal braking of tank car	8	513 × 499
4	Abnormal braking of open car	5	513 × 308
5	Abnormal braking of box car	6	513 × 74
6	Adjacent uncoupling trains	4	513 × 360
7	Wheel vibration at rail slots	8	513 × 37
8	Passing-by trains	2	513 × 375
9	Roaring of locomotive's engine	4	513 × 260
10	Train whistling	10	513 × 147
11	Inertia vibration before stop	4	513 × 57
12	Humping wagons	2	513 × 141

should be selected out into its own categorization for subsequent data training. The selected acoustic samples are listed in Table 1, where the last column is the data size of energy matrix output by FFT. There are 5 types of abnormal sounds of unreleased braking and 7 types of background sounds during normal operation. Meanwhile, the flat car and open-top car have a smaller sample size than other wagons, because some acoustic signals of abnormal braking are submerged in the complicated sounds.

The time domain and frequency domain distributions of representative sound types are shown in Figure 5, including type 1, type 4, type 6, type 8 and type 10. It is obvious that different sound type has different wave form characteristics, represented in the amplitude distribution and frequency distribution. According to the time domain signals, sound type 1, 6 and 8 have higher amplitudes than type 4 and 10, sound type 10 has an instantaneous amplitude peak, while the amplitude distributions of sound type 4 and 9 are more

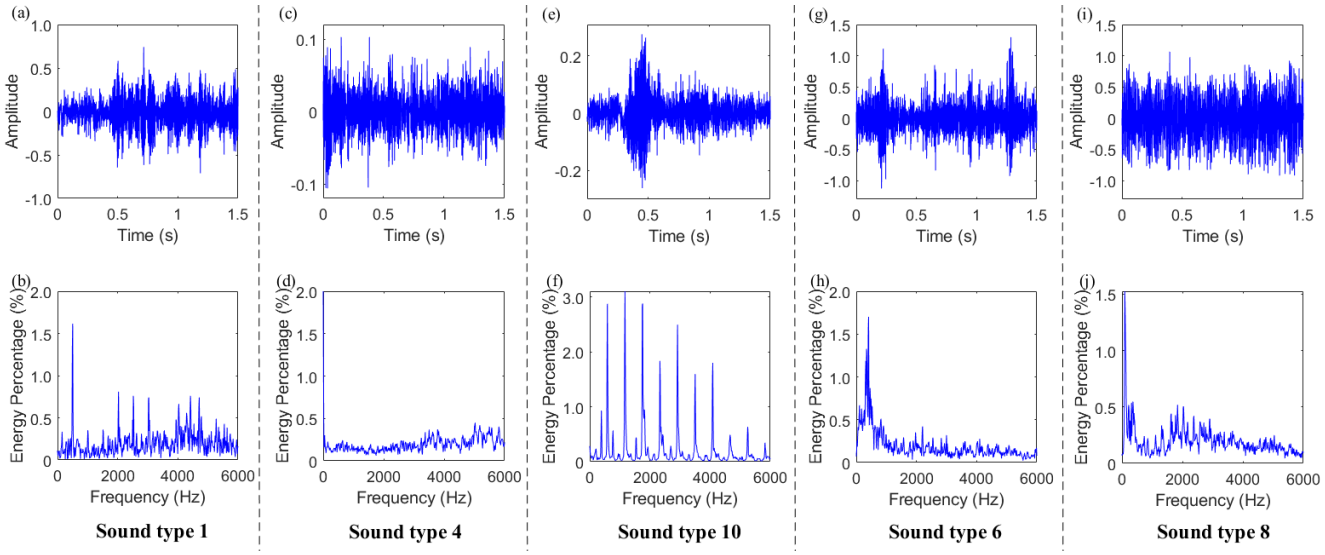


FIGURE 5. Time domain and frequency domain distribution of representative acoustic samples.

equilibrium than other sound types. Compared with the time domain waveform, the differences are more apparent in the frequency distribution. For example, train whistling sound has a discrete distribution of frequency peaks, where intervals between peaks are almost the same (about 500∼600Hz).

IV. HYBRID RECOGNITION ALGORITHM

A. ALGORITHM DESIGN

The framework of the hybrid algorithm is an integration of off-line data training and real-time recognition, as shown in Figure 6. The devised algorithm has a self-regulated mechanism, where the template matrix and SVM classifiers can be updated with the newly stored data from the database.

The recognition algorithm includes four major steps: FFT processing, feature extraction, template matching and SVM classification. Since FFT is a widely-used method in digital signal processing [38], [39], algorithms of feature extraction, template matching and SVM classification will be mainly introduced, considering the data characteristics and safety requirements.

B. FEATURE EXTRACTION

The frequency of collected acoustic signals is 48000 Hz, namely every second generates 48000 data records. Through FFT processing, the energy matrix contains an average of 16 column vectors per second. For the convenience of subsequent SVM training, the feature extraction is performed to all vectors in the energy matrix of every selected acoustic category.

Affected by various conditions of working temperature, air moisture, trains running speed and load capacity, spectrograms of the same acoustic category may become obviously different. Therefore, the normalization of energy matrix is necessary before feature extraction. The method of Min-Max

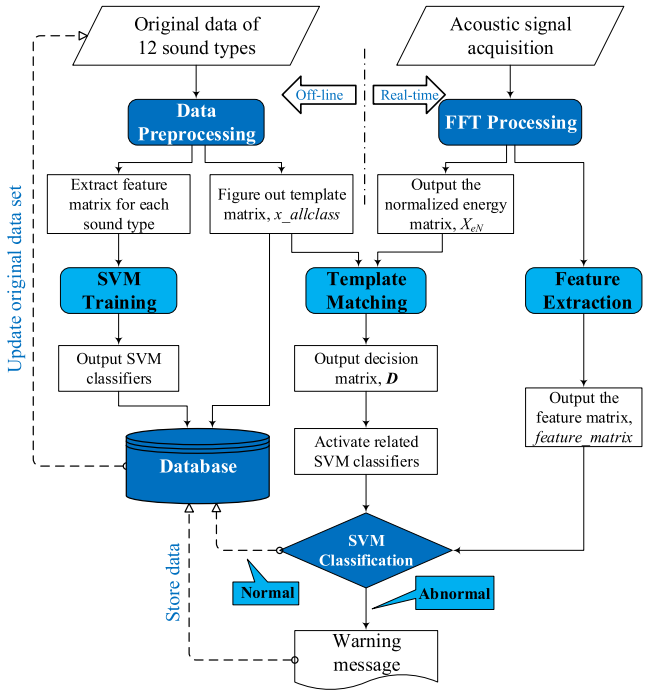


FIGURE 6. The framework and procedure of the recognition algorithm.

normalization has been adopted, as in

$$v'_{ij} = (v_{ij} - \min\{V_j\})/(\max\{V_j\} - \min\{V_j\}) \quad (1)$$

where V_j is j^{th} column vector of an energy matrix, v_{ij} is the i^{th} element of V_j, and v'_{ij} is the normalized value of v_{ij}.

The frequency of energy matrix output by FFT ranges from 0 Hz to 6000 Hz, and it is isometric partitioned into 512 sections. To erase some inherent bias and to enhance computation efficiency, the frequency is re-divided into 60 sections at 100 Hz intervals. Based on the initial analysis of Figure 5,

TABLE 2. Fourteen extracted features from normalized energy matrix.

#	Features	Unit	Symbol
1	Spectrum center	Hz	f_{ave}
2	Spectrum flux	--	f_{sum}
3	Energy percentage of 1st peak	%	f_{1p}
4	Energy percentage of 2nd peak	%	f_{2p}
5	Energy percentage of 3rd peak	%	f_{3p}
6	Energy percentage of 4th peak	%	f_{4p}
7	Energy percentage of 5th peak	%	f_{5p}
8	Energy percentage of 6th peak	%	f_{6p}
9	Average frequency of 1st peak	Hz	f_{1a}
10	Average frequency of 2nd peak	Hz	f_{2a}
11	Average frequency of 3rd peak	Hz	f_{3a}
12	Average frequency of 4th peak	Hz	f_{4a}
13	Average frequency of 5th peak	Hz	f_{5a}
14	Average frequency of 6th peak	Hz	f_{6a}

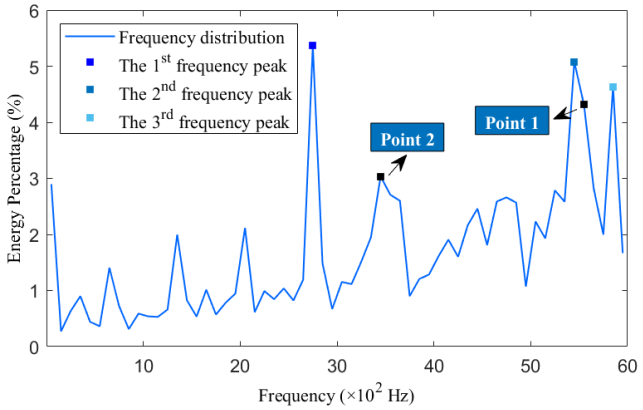


FIGURE 7. The extraction of frequency peaks considering minimum interval.

six frequency peaks and corresponding energy distribution are enough for the recognition of different sound types in the marshaling yard. Features need to be extracted are listed in Table 2.

During the extraction of frequency peaks, 200 Hz is set to be the minimal interval between peaks. The purpose of setting this interval is to guarantee the equilibrium of extracted peaks. As shown in Figure 7, the top three peaks have been marked in the frequency distribution, and their intervals are greater than 200 Hz naturally. When extracting the 4th frequency peak, point 1 should be ignored because it is adjacent to the 2nd peak, and point 2 would become the 4th peak. The major steps for feature extraction is programmed in Matlab R2018a, as shown in Algorithm 1.

C. TEMPLATE MATCHING

The purpose of template matching is to activate major SVM classifiers for real-time collected acoustic data during the time period. To guarantee the reliability and timeliness, the recognition of abnormal braking should depend on the frequency distribution of energy spectrum during a very short time period, like 0.5 s. As discussed before, each frequency distribution corresponds to one column vector in the energy matrix, and there are about 8 vectors during 0.5 s. Through the SVM training of extracted feature matrix, SVM classifiers can be obtained. Assuming that there are n classifiers, then every feature data of one column vector should invoke all the n classifiers to identify its category, and the computation

Algorithm 1 Feature Extraction Algorithm

Input File location, file_path
Energy matrix output by FFT, X_e
Original frequency vector, f_o
Combined frequency vector, f_c

Output Feature matrix of X_e, *featureMatrix*

Load f_o and f_c from the file_path
Normalize Xe: XeN = (mapminmax(Xe',0,1))'
Redivide X_{eN} at 100 Hz intervals, and obtain X_{eNR}
Calculate the energy distribution matrix X_{eD}
[m_2,n_2]= size(X_{eNR}); X_{eD} =zeros(m_2,n_2); Sum=sum(X_{eNR})
for j=1:n_2; X_{eD}(:,j)= X_{eNR} (:,j)/Sum(j); **end**
Extract the spectrum center *f_ave*
Extract the spectrum flux *f_sum*
place=zeros(6,n_2) % store the index of 6 peaks
f_p_*max*=[] % store the frequency of 6 peaks
Find the 1st frequency peak, f_{1p_max} =max(X_{eD})
for j=1:n,
[place(1,j),temp]=find(f_{1p}_*max*(1,j) == X_{eD}(:,j))
end
Find the 2nd frequency peak f_{2p}_*max*
X_{eD1} = X_{eD}; t=1; p=1
for j =1:n_2; X_{eD1}(place(p,t),j)=0
if place(p,t)==1; X_{eD1}(place(p,t)+1,j)=0
elseif place(p,t)==m_2; X_{eD1}(place(p,t)-1,j)=0
else X_{eD1}(place(p,t)-1,j)=0; X_{eD1}(place(p,t)+1,j)=0
end; t=t+1
end; f_{2p}_*max*=max(X_{eD1})
f_p_*max*=[f_{1p}_*max*; f_{2p}_*max*]
for j=1:n
[place(2,j),temp]=find(f_{2p}_*max*(1,j) == X_{eD1}(:,j))
end
Find the 3rd, 4th, 5th and 6th frequency peak in order
f_p_*max*=[f_{1p}_*max*; f_{2p}_*max*; f_{3p}_*max*; f_{4p}_*max*; f_{5p}_*max*; f_{6p}_*max*]
place_fm=place∗100-100 % transform index into frequency
featureMatrix=[*f_ave*;*f_sum*; f_p_*max*;*place_fm*]

load will be expanded eightfold for the data of 0.5 s, which couldn't meet the demand of real-time recognition. Therefore, a template matching process should be performed before invoking SVM classifiers.

For a sound category, its template vector and matching threshold must be determined. The template vector is obtained by averaging the energy matrix of selected acoustic data by row, and the matching threshold is decided by the maximum proximity between every sample vector and the template vector, where the calculation method of proximity is worthy of careful consideration.

Euclidean distance is widely used in measuring the absolute difference between two individuals, and in our study it is defined as

$$d_{pro}^{E}(r,t) = \sqrt{\sum_{i=1}^{m}(r_i - t_i)^2} \tag{2}$$

where r and t represent the real-time column vector and the template vector respectively, r_i and t_i denote the corresponding i^{th} element values, m is the dimension of column vector, which is 513 in this case, and $d_{pro}^{E}(r,t)$ is the proximity between r and t using Euclidean distance.

However, every real-time vector should be compared with 11 template vectors, and we found that sometimes the proximity calculated by Euclidean distance was not consistent with the reality through repeated tests. To guarantee the reliability of template matching, we decide to adopt the cosine distance [40] to measure the proximity, which is defines as

$$d_{pro}^{C}(r,t) = \sum_{i=1}^{m} r_i \times t_i \Bigg/ \left(\sqrt{\sum_{i=1}^{m} r_i^2} \times \sqrt{\sum_{i=1}^{m} t_i^2}\right) \tag{3}$$

where $d_{pro}^{C}(r,t)$ is the cosine distance between r and t, ranging from -1 to 1.

The results of repeated tests had shown that the cosine distance outperforms the Euclidean distance during the template matching. Besides, Pearson correlation coefficient (PCC) can also be used to measure the proximity [41], which is an improvement of cosine distance through a centralized processing as

$$d_{pro}^{P}(r,t) = \frac{\sum_{i=1}^{m}(r_i - \bar{r}) \times (t_i - \bar{t})}{\sqrt{\sum_{i=1}^{m}(r_i - \bar{r})^2} \times \sqrt{\sum_{i=1}^{m}(t_i - \bar{t})^2}} \tag{4}$$

where $d_{pro}^{P}(r,t)$ is the proximity between r and t via PCC, $\bar{r}$ and $\bar{t}$ denote the mean value of r and t respectively.

Algorithm 2 shows the devised codes of template matching, where major steps are listed as follows.

–For an energy vector from real-time FFT processing, calculate the proximity between the vector and every template vector in turn.

–Compare the proximity values and threshold values, and select top three categories with the nearest proximity.

Algorithm 2 Template Matching Algorithm

Input	File location, file_path Template matrix of all sound types, *X_allclass* Normalized energy matrix, X_{eN}
Output	Decision matrix, $\boldsymbol{D}$
1:	Calculate the proximity between X_{eN} and *X_allclass*
2:	*X_allclass* = load([file_path,'X_allclass.mat'])
3:	n_c =size(*X_allclass*,2); n_e =size(X_{eN},2); d_{pro} = zeros(n_e,n_c)
4:	**for** i=1:n_e
5:	**for** j=1:n_c
6:	r= X_{eN} (:,i); t=*X_allclass*(:,j); d_{pro}(i,j)=pdist2(r',t','cosine')
7:	**end**
8:	**end**
9:	Figure out the decision matrix $\boldsymbol{D}$
10:	**for** i=1:n_e
11:	min1=min(d_{pro}(i,:)); $\boldsymbol{D}$(i,1)=find(min1 == d_{pro}(i,:)); d_{pro}(i, $\boldsymbol{D}$(i,1))=inf
12:	**end**
13:	**for** i=1:n_e
14:	min2=min(d_{pro}(i,:)); $\boldsymbol{D}$(i,2)=find(min2 == d_{pro}(i,:)); d_{pro}(i, $\boldsymbol{D}$(i,2))=inf
15:	**end**
16:	**for** i=1:n_e
17:	min3=min(d_{pro}(i,:)); $\boldsymbol{D}$(i,3)=find(min3 == d_{pro}(i,:))
18:	**end**

–Integrate the nearest categories of all vectors in one period into a decision matrix $\boldsymbol{D}$, where an element d_{ij} represents the i^{th} nearest sound type of the j^{th} energy vector.

–Tabulate the decision matrix, and find top three categories for all vectors according to the occurrence rate.

Taking one acoustic data of train whistling (type 10) as an example, the decision matrix of template matching is indicated in Table 3. The length of this acoustic data is 1 s, and the corresponding energy matrix includes 15 column vectors. Top three nearest templates of every vector has been figured out. Obviously, different energy vector corresponds to different templates. However, the most three frequent templates of d_{ij} are type 10, type 11 and type 3, where type 10 has the top rate in N_1. The three nearest templates will be output as the basis of SVM classification.

D. SVM CLASSIFICATION

As a binary classification algorithm [42], SVM algorithms usually adopt two approaches for multi classification, namely the one-against-one strategy and the one-against-many strategy [43]. Assuming there are m data categories, the one-against-one strategy needs to train C_m^2 classifiers, and the one-against-many strategy needs to train $m-1$ classifiers. It is obvious that the one-against-one outperforms the one-against-many in the reliability, while the latter outperforms

TABLE 3. The decision matrix *D* of template matching for given data of type 10.

Column	N_1	N_2	N_3
1	5	6	3
2	3	5	6
3	10	11	3
4	10	3	9
5	10	3	11
6	10	11	1
7	10	11	6
8	10	11	5
9	10	11	3
10	10	11	3
11	10	11	5
12	10	11	6
13	10	11	1
14	10	11	6
15	11	10	5

N_1, N_2 and N_3: the first nearest template, the second nearest template and the third nearest template.

the former in the efficiency. In the field of audio surveillance, the one-against-many and one-class SVM classification are frequently adopted due to the high efficiency [44]. However, specific case needs specific analysis.

According to Table 1, there are 12 categories of acoustic data in this study. If we adopt the one-against-many strategy, every energy vector will activate 11 classifiers without template matching. Through template matching, only 3 classifiers will be activated, but it may cause difficulty in identifying the category of abnormal data. If we adopt the one-against-one strategy, the sound category can be identified by probability, but it needs 66 classifiers theoretically, which will cause some computation burden to SVM training.

Considering the safety requirement for recognition accuracy, finally 45 classifiers have been determined based on the one-against-one strategy. The 12 sound categories are split into set **A** and set **B**, where set **A** includes the 5 abnormal sounds and set **B** includes the 7 background sounds. On one hand, every type in set **A** needs to be identified with each other, therefore there should be 10 classifiers. On the other, every type in set **A** also need comparisons with every type in set **B**, therefore there should be another 35 classifiers. Because the purpose is to recognize the abnormal sound of unreleased braking, it is not necessary to develop classifiers between types in set **B**. All 45 classifiers are stored in the database. The binary classification result is marked as '1' or '-1'. The trained SVM classifiers will be activated after the template matching, where the top three nearest templates are denoted by t_1, t_2 and t_3 respectively. The activation and recognition algorithm should obey the following rules under different cases.

1) CASE 1

If none of the three templates belongs to set **A**, then output the safe value.

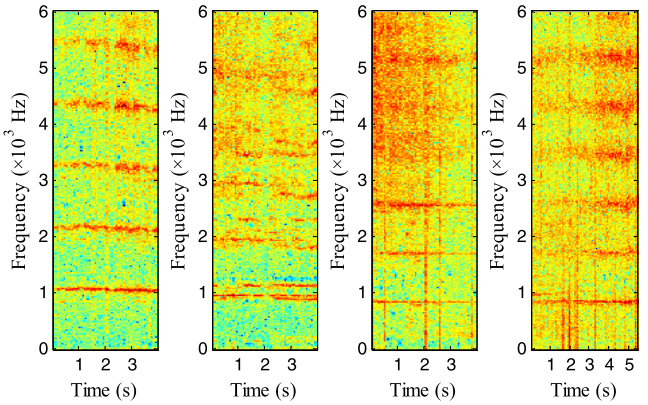


FIGURE 8. The energy spectrum of selected samples (tank car).

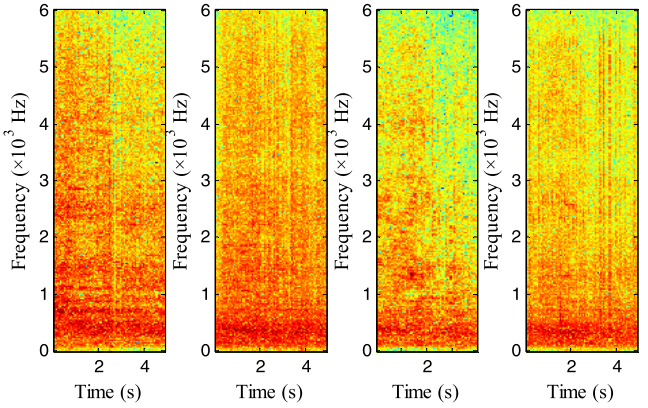


FIGURE 9. The energy spectrum of selected samples (Adjacent uncoupling trains).

2) CASE 2

If there is only one template belongs to set **A**, assuming that this template is t_1, then activate classifier t_1-t_2 and classifier t_1-t_3. Meanwhile, give a safe value to the virtual classifier t_2-t_3, because t_2 and t_3 both belong to set **B**. The final output is a warning value only when the classification results are 1 and 1, and t_1 is recognized as the abnormal sound type.

3) CASE 3

If there are two templates belong to set **A**, assuming that the two templates are t_1 and t_2, then activate classifier t_1-t_2, classifier t_1-t_3 and classifier t_2-t_3. The output value depends on the results of classifiers. E.g. if the classification results are 1, 1 and -1, which correspond to t_1 (abnormal), t_1 (abnormal) and t_3 (normal), then output a warning value and determine t_1 as the sound type. When outputting a warning value, result of classifier t_1-t_2 plays a decisive role in the recognition of sound category.

4) CASE 4

If all the three templates belong to set **A**, then output a warning value without question, and activate classifier t_1-t_2, classifier t_1-t_3 and classifier t_2-t_3 for recognition. Usually, the classification results correspond to two templates, and the template appearing twice will be identified. E.g. if the

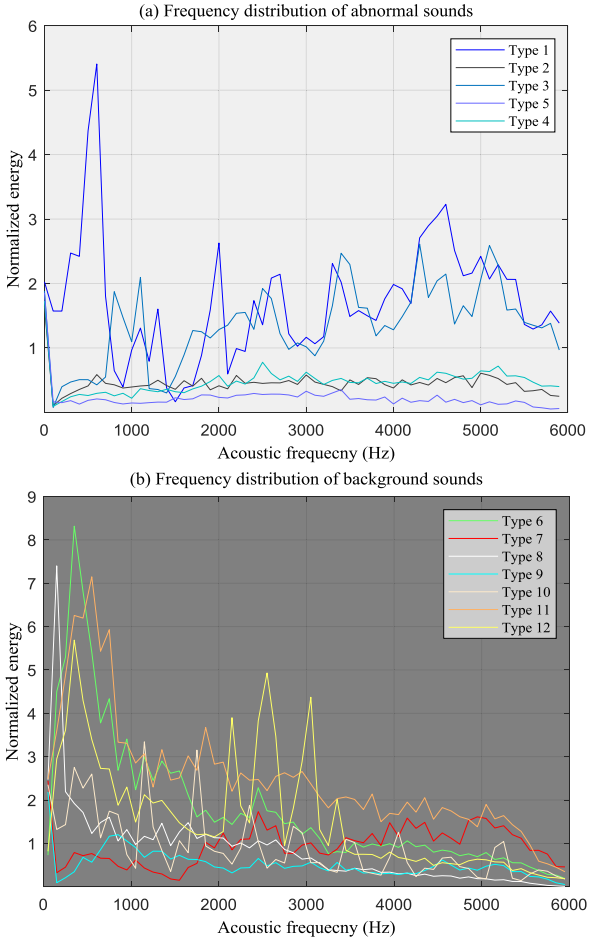


FIGURE 10. The frequency distribution of 12 acoustic templates.

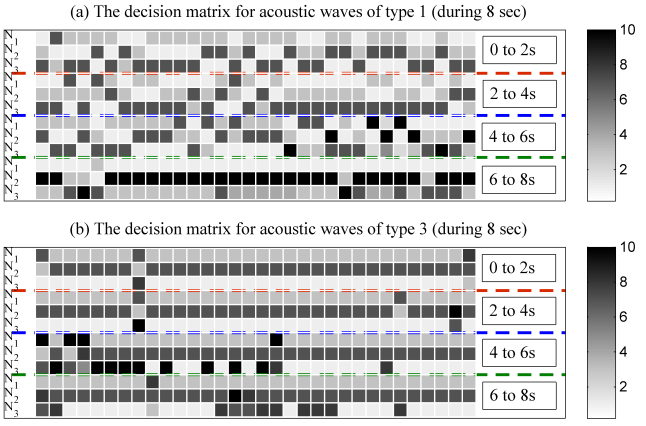


FIGURE 11. Template matching results for abnormal braking sounds.

classification results are t_2, t_3, t_2 respectively, then t_2 is determined as the sound type. However, if the results correspond to three templates, namely every template appears once, then the template with the top rate in N_1 will be identified.

V. CASE ANALYSIS

A. FEATURE EXTRACTION OF SELECTED SAMPLES

Figure 8 shows the representative acoustic spectrograms of the unreleased braking from tank cars (type 3), and Figure 9 shows the spectrograms of normal running from adjacent uncoupling trains (type 6). The frequency distribution of type 3 has a strong regularity, where 5 to 6 frequency peaks are uniformly distributed in the energy spectrum. In contrast, the frequency distribution of type 6 concentrates on the low frequency band, where the spectrum data shows that over 50 % acoustic energy is distributed below 1300 Hz.

After running the code of feature extraction, representative frequency characteristics of 12 sound categories can be figured out, as listed in Table 4.

B. TEMPLATE MATCHING ANALYSIS

Based on the acoustic data of selected sound waves, the templates of 12 sound types can be obtained from their normalized energy matrices. The frequency distribution of abnormal unreleased braking sounds and background sounds are displayed in Figures 10(a) and 10(b). It is reflected that sounds of abnormal braking have energy peaks in the middle-high frequency band, while the background sounds focus in the low-middle frequency band. Meanwhile, in Figure 10(a), type 1 and type 3 have more obvious energy peaks than other abnormal types. It is also found that the acoustic characteristics of type 2, type 4 and type 5 are much weaker than type 1 and type 3 due to different brake blocks. In the experimental train, the flat car, open-top car and box car are equipped with composite brake shoes, while he JSQ car and tank car are equipped with traditional iron brakes.

Taking the samples from type 1 and type 3 as examples, both acoustic data sets are during 8 seconds, with 128 continuous column vectors. The results of corresponding template matching are indicated in Figure 11, where the values of d_{ij} are displayed in a greyscale grid.

As shown in Figure 11(a), the original acoustic sample belongs to the unreleased braking of JSQ car, while the nearest templates appear most frequently are type 1 (33.1%), type 3 (29.2%) and type 7 (25.3%). Despite type 1 has the highest rate, it doesn't have a dominate proportion in N_1 during 0 to 2s. Similarly, in Figure 11(b), the original sample data belongs to type 3, the nearest templates appear most frequently are type 7 (33.1%), type 3 (32.8%) and type 1 (25.5%). It should be noted that type 3 has the top rate in N_1, which means that type 3 is the first nearest template. Therefore, it is hard to identify the sound category only by the proportion of nearest templates, and that's why we need the SVM classification.

C. CLASSIFICATION RESULTS

Through template matching, top three nearest templates will be activated for a real-time collected sample within 0.5 s. To better illustrate the classification process, test samples of JSQ car's abnormal braking (type 1), tank car's abnormal braking (type 3) and train whistling (type 10) are taken as examples. Theoretically, different acoustic samples will

TABLE 4. Extracted frequency features of 12 sound types.

ID	1	2	3	4	5	6	7	8	9	10	11	12
f_{ave}	2867	2878	3066	3028	2715	1671	3081	1537	2625	2170	2379	2081
f_{sum}	103.5	27.2	80.8	28.4	13.4	107.3	51.2	57.1	30.9	55.2	149.6	95.2
f_{1p}	7.1%	4.2%	6.8%	4.6%	6.8%	7.8%	5.0%	19.7%	4.8%	12.7%	5.8%	9.9%
f_{2p}	4.5%	2.9%	4.8%	3.2%	3.0%	5.4%	3.9%	4.2%	3.8%	8.4%	4.6%	7.0%
f_{3p}	3.7%	2.7%	3.9%	2.9%	2.7%	4.3%	3.3%	3.4%	3.1%	6.0%	3.8%	5.2%
f_{4p}	3.2%	2.5%	3.4%	2.7%	2.5%	3.5%	3.0%	2.9%	2.8%	4.9%	3.0%	4.0%
f_{5p}	2.8%	2.4%	3.1%	2.5%	2.4%	3.0%	2.8%	2.6%	2.6%	4.3%	2.7%	3.4%
f_{6p}	2.6%	2.3%	2.7%	2.4%	2.3%	2.7%	2.6%	2.4%	2.4%	3.6%	2.5%	2.9%
f_{1a}	600	100	4300	100	100	300	100	200	100	1100	500	300
f_{2a}	4600	5000	5100	2500	3400	700	2400	700	900	1700	700	2500
f_{3a}	2000	600	3400	5200	3000	900	4900	1600	1200	300	1800	3000
f_{4a}	4300	3000	1100	3000	2400	1100	4100	1300	1500	2300	1100	2100
f_{5a}	3300	2200	4600	5000	4500	1400	2600	900	2400	700	2800	900
f_{6a}	2600	4800	2500	4500	1800	2400	5200	1800	4500	2800	4300	3300

TABLE 5. Representative classification results of acoustic wave data (type 1).

Test samples	Classification results			Output type	Decision type
	1-3 SVC	1-7 SVC	3-7 SVC		
R_1	1	1	-1	1	1
	1	1	-1	1	
	-1	-1	1	3	
	1	-1	1	1	
	1	-1	1	1	
	1	-1	1	1	
	1	-1	1	1	
	1	1	1	1	
R_2	1	-1	-1	7	7
	1	-1	-1	7	
	1	-1	1	1	
	1	-1	1	1	
	1	-1	-1	7	
	1	1	-1	1	
	1	-1	-1	7	
	1	-1	-1	7	

SVC represents the classifiers trained by SVM.

activate different combinations of nearest templates. After template matching, all test samples of type 1 and type 3 activate the same template set of 1-3-7, while the test samples of type 10 activate different template sets including 7-10-11, 3-7-10 and 3-10-11.

Classification results of two representative samples of type 1 are indicated in Table 5. The decision type is determined by the output types during 0.5 s, because every output type is an instantaneous value. Among the output types of R_1, type 1 accounts for 87.5%; among the output types of R_2, type 1 only accounts for 37.5%. R_1 and R_2 both belong to the sample sets of type 1, while R_1 is recognized as type 1 and R_2 is recognized as type 7. Considering the quality of samples and the complexity of classifiers, it is inevitable to encounter the above recognition errors. Similarly, test samples of type 3 and type 10 will also encounter these errors.

The distribution of output types for test samples of type 1, type 3 and type 10 are shown in Figures 12(a)-(c). According to the test results, 3% of type 1 samples are mistaken as type 3

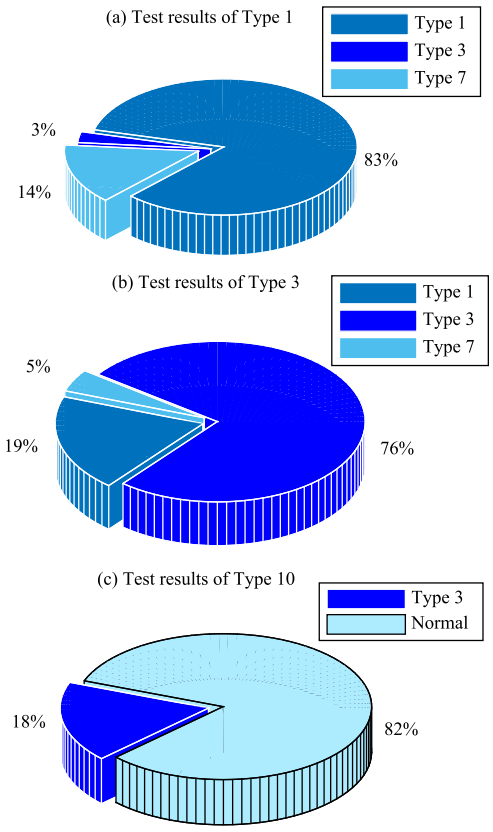


FIGURE 12. Distribution of instantaneous output types for test samples of type 1, 3 and 10.

and 19% of type 3 samples are mistaken as type 1. Meanwhile, 14% of type 1 samples are recognized as type 7 and 5% of type 3 samples are recognized as type 7. Despite these false rates, it should be noted that these output types are under instantaneous recognition, and 8 continuous output types during 0.5 seconds will determine a decision type. Therefore, the final recognition accuracy of decision types should be higher, which will be validated later.

As indicated in Figure 11(c), 18% of type 10 samples are mistaken as type 3, which is acceptable and safety-oriented.

TABLE 6. Recognition results of test samples for 12 sound types.

Sound type		Test sample size	Instantaneous outputs composition	Warning rate
Abnormal	1	20 (513 × 154)	83% (type 1), 3% (type 3), 14% (type 7)	95%
	2	10 (513 × 79)	100% (type 2)	100%
	3	20 (513 × 154)	19% (type 1), 76% (type 3), 5% (type 7)	100%
	4	8 (513 × 62)	6% (type 2), 94% (type 4)	100%
	5	9 (513 × 71)	100% (type 5)	100%
Normal	6	20 (513 × 158)	100% (normal types)	0%
	7	12 (513 × 95)	11% (type 1), 1% (type 3), 88% (normal types)	8%
	8	15 (513 × 118)	100% (normal types)	0%
	9	20 (513 × 155)	4% (type 2), 96% (normal types)	0%
	10	10 (513 × 78)	18% (type3), 82% (normal types)	10%
	11	13 (513 × 102)	100% (normal types)	0%
	12	10 (513 × 79)	100% (normal types)	0%

Besides, the normal sound types need no recognition because we have not designed SVM classifiers between normal sound types to reduce calculation load. Therefore, output types of 7, 10 and 11 are merged into the normal types.

D. ACCURACY VALIDATION

Based on the proposed algorithm, the final classification results of different acoustic samples can be obtained. Table 6 lists the recognition results of 12 sound types. For abnormal sound types, the warning rate is consistent with the recognition accuracy. It is obvious that the warning rate is higher than the proportion of corresponding abnormal type. During the accuracy testing, the average warning rate of 5 abnormal types is 99%, while the average warning rate of 7 normal sound types is 2.6%. Considering the warning rate of normal sound samples, the global decision accuracy is about 98.2%, which is calculated by a weighted equation as

$$A_g = \left[\sum_{i=1}^{5} s_i \times \alpha_i + \sum_{i=6}^{12} s_i \times (1-\alpha_i)\right] \Big/ \sum_{i=1}^{12} s_i \qquad (5)$$

where A_g represents the global decision accuracy, s_i is the sample size of the i^{th} sound type, α_i is the warning rate of the i^{th} sound type. Note that for normal sound types ($i = 6, 7, \ldots, 12$), the i^{th} decision accuracy should be 1 minus α_i.

Admittedly, some types may be confused with other types, like type 1 and type 7, type 3 and type 7, type 4 and type 2. This phenomenon of confusion may arise from their similar frequency distribution, and it may also arise from their disturbance with each other. Nevertheless, the recognition error is acceptable and the overall algorithm accuracy is reliable.

As to the recognition efficiency, the combined algorithms were coded in MATLAB R2018a and were performed on an Intel Core 6700k @4.18 GHz, with 16 Gb of RAM. The computational time for an acoustic data during 0.5 s is about 0.11 s, which satisfies the demand of real-time recognition.

VI. DISCUSSION

Compared with current studies related to train braking faults identification, the biggest difference lies in the monitoring method. This paper presented an acoustic-based recognition algorithm for wagons' unreleased braking, while the prior studies usually adopt onboard sensors to monitor the status of brake system. Based on the time-varying monitoring data of brake valve status and cylinder air pressure, Liu *et al.* [45] achieved a classification accuracy of 94.3% by using the method of Gradient Boosting Decision Tree (GBDT). The accuracy of our acoustic-based recognition algorithm is about 98.2%. It is hard to judge between the two methods due to the differences in measuring approach and original data set. Another difference lies in the train running environment. Current researches focus on trains running on railway lines, while this study targets at the uncoupling trains in the marshaling yard, the running speed, infrastructure condition and surrounding environment are different. Meanwhile, in the marshaling yard, it is feasible for us to install the way-side acoustic monitoring system, which is easier to maintain and upgrade.

TABLE 7. Comparisons with other acoustic recognition applications.

Application context	Sound types	Algorithm	Accuracy
Home sound classification [46]	Male speech, female speech, cough, laughing, screaming, dog barking, etc.	SVM KNN	85.1%
Bird sound classification [47]	Flight calls of 43 different birds.	CNN	86.31%
Emergency vehicle detection [48]	Siren sounds of ambulances, police cars and fire engines, car horns, urban noise.	CNN	98.24%
Animal sound classification [49]	Sounds from anurans, birds, and insects.	CNN SVM	97.18%
Public scene classification [50]	Sounds of bus, busy street, office, park, supermarket, etc.	SVM	73%

Since the proposed method is an acoustic-based recognition algorithm, comparisons with prior studies in acoustic recognition and classification have been performed from the perspectives of application context, sound types, recognition algorithm and recognition accuracy, as indicated in Table 7.

SVM and convolutional neural networks (CNN) are most frequently used algorithm for recognition, while the k nearest neighbour (KNN) is also applicable. The recognition algorithm presented in this paper is an improvement of SVM, where the classifiers and decision rules are newly designed considering the recognition requirements. Meanwhile, the extracted features are determined upon the spectrum analysis of 12 sound types, which makes the results of data training more pertinent. As shown in Table 7, the accuracy of acoustic recognition usually ranges from 70% to 99%. The global recognition accuracy of our algorithm is in the high level of current studies. However, it should be noted that the acquisition quality and the number of sound types will affect the accuracy to some extent.

VII. CONCLUSION

This work researches the detection of wagons unreleased braking at railway marshaling yard, and proposes an acoustic-based recognition algorithm considering the application context. Repeated experiments have been conducted to collect acoustic data through a self-installed wayside acquisition system, where the acoustic waves are classified into 5 abnormal types and 7 background types.

The hybrid algorithm consists of data pre-processing, feature extraction, template matching, SVM training and classification. Upon a detailed spectrum analysis through FFT, 14 frequency features have been selected during extraction, where the discreetness of extracted frequency peaks is guaranteed under a minimum interval of 200 Hz. A matching algorithm is then carefully designed to figure out a decision matrix before the activation of SVM classifiers. The matching proximity between real-time vectors and template vectors is calculated by the cosine distance, which outperforms the Euclidean distance for the high-dimensional acoustic data. To increase the efficiency, top three nearest templates in the decision matrix are selected for SVM classification. The mechanism of SVM classification has been redesigned, considering the selected templates and trained classifiers. Meanwhile, algorithms of off-line training and real-time computation have been integrated in the algorithm procedure organically. Testing results of 12 sound types have validated the reliability of proposed algorithm, with an overall decision accuracy over 98%.

The limitation of proposed method lies in the acoustic samples and the disturbance of background noise. On one hand, some sound types may be confused with each other, and the average recognition error for all sound types is about 6.75%. Given more acoustic samples before feature extraction and SVM training, the false rate will decrease to some extent. On the other hand, there exist some unexpected background noises, which will affect the recognition accuracy of unreleased braking. Meanwhile, when faced with adverse weather conditions, the warning system needs an alternative database including different original samples and classifiers.

Future work will be focused on improving the pre-processing method and recognition algorithm for a higher accuracy during instantaneous recognition. Algorithms like wavelet transform and decision tree are worth applying to enhance the recognition performance. In summary, the proposed recognition method can effectively extract the acoustic features and recognize the sound type for the real-time identification of wagons unreleased braking.

DATA AVAILABILITY

The original experimental data used during the study are available from the corresponding author by request.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] K. Sato and N. Fukumura, "Real-time freight locomotive rescheduling and uncovered train detection during disruption," *Eur. J. Oper. Res.*, vol. 221, no. 3, pp. 636–648, Sep. 2012, doi: 10.1016/j.ejor.2012.04.025.

[2] F. Corman, A. D'Ariano, D. Pacciarelli, and M. Pranzo, "Bi-objective conflict detection and resolution in railway traffic management," *Transp. Res. C, Emerg. Technol.*, vol. 20, no. 1, pp. 79–94, Feb. 2012, doi: 10.1016/j.trc.2010.09.009.

[3] H. Henao, S. H. Kia, and G.-A. Capolino, "Torsional-vibration assessment and gear-fault diagnosis in railway traction system," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 1707–1717, May 2011, doi: 10.1109/TIE.2011.2106094.

[4] M. McClanachan and B. Payne, "Improving brake propagation in long freight trains," in *Proc. Conf. Railw. Eng. (CORE)*, Sep. 2008, pp. 537–548.

[5] S. Zhang, Z. Huang, Y. Yang, K. Gao, and C. Zhu, "A safe and reliable anti-lock wheel control with enhanced forgotten factor for brake operation of heavy train," in *Proc. IEEE Int. Symp. Parallel Distrib. Process. with Appl. IEEE Int. Conf. Ubiquitous Comput. Commun. (ISPA/IUCC)*, Dec. 2017, pp. 822–826.

[6] L. S. Ye, "Identification of the continuous brake seizure troubles due to brake slack adjuster locking on freight cars and the disposition," *Rolling Stock*, vol. 48, no. 2, pp. 43–44, Feb. 2010, doi: 10.3969/j.issn.1002-7602.2010.02.014.

[7] S.-W. Nam and H.-J. Kim, "A study on the improvement of release application characteristics of pneumatic brakes for freight train," *KSME Int. J.*, vol. 16, no. 6, pp. 776–784, Jun. 2002, doi: 10.1007/BF02939337.

[8] H. Zhang, "Research on the control of wheel-rail noise by wheel sound-proof skirt for high-speed train," in *Proc. IEEE Int. Conf. Mechatron. Autom. (ICMA)*, Aug. 2019, pp. 2064–2069, doi: 10.1109/ICMA.2019.8816428.

[9] P. J. Bolton and P. Clayton, "Rolling-sliding wear damage in rail and tyre steels," *Wear*, vol. 93, no. 2, pp. 145–165, Jan. 1984, doi: 10.1016/0043-1648(84)90066-8.

[10] S. J. Kwon, D. H. Lee, S. T. Kwon, and B. C. Goo, "Failure analysis of railway wheel tread," *Key Eng. Mater.*, vols. 321–323, pp. 649–653, Oct. 2006, doi: 10.4028/www.scientific.net/kem.321-323.649.

[11] S. Bähler, "Methods and results of field testing of a retrofitted freight train with composite brake blocks," *J. Sound Vibrat.*, vol. 293, nos. 3–5, pp. 1041–1050, Jun. 2006, doi: 10.1016/j.jsv.2005.08.055.

[12] E. H. Jansen, M. G. Dittrich, and E. L. Sikma, "Brake noise measurements on mixed freight trains with composite brake blocks," *J. Acoust. Soc. Amer.*, vol. 123, no. 5, pp. 4521–4526, May 2011, doi: 10.1121/1.2933586.

[13] M. Durali and B. Shadmehri, "Nonlinear analysis of train derailment in severe braking," *J. Dyn. Syst., Meas., Control*, vol. 125, no. 1, pp. 48–53, Mar. 2003, doi: 10.1115/1.1541669.

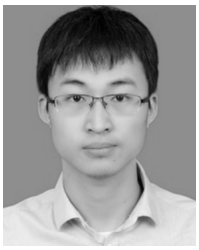
[14] E. Bernal, M. Spiryagin, and C. Cole, "Onboard condition monitoring sensors, systems and techniques for freight railway vehicles: A review," *IEEE Sensors J.*, vol. 19, no. 1, pp. 4–24, Jan. 2019, doi: 10.1109/JSEN.2018.2875160.

[15] C. Lonsdale and B. Wilson, "Use of on-board hand brake monitoring to prevent freight car wheel damage and improve maintenance and safety," in *Proc. ASME Rail Transp. Division Fall Tech. Conf.*, Jan. 2011, pp. 11–16, doi: 10.1115/RTDF2011-67002.

[16] G. Y. Lu and H. H. Zhang, "Research and design of the passenger train brake fault alarm," *J. Zhengzhou Railw. Vocat. Tech. Coll.*, vol. 28, no. 2, pp. 25–27, Jul. 2016.
[17] M. Aimar and A. Somá, "Study and results of an onboard brake monitoring system for freight wagons," *Proc. Inst. Mech. Eng., F, J. Rail Rapid Transit*, vol. 232, no. 5, pp. 1277–1294, Jul. 2017, doi: 10.1177/0954409717720348.
[18] W. J. Lee, "Contextual air leakage detection in train braking pipes," in *Proc. Int. Conf. Industrial, Eng. Other Appl. Appl. Intel. Sys. (IEA/AIE)*, Jun. 2017, pp. 191–200, doi: 10.1007/978-3-319-60045-1_22.
[19] Y. Z. Yang, "Research and application on pneumatic pipe leakage diagnosis technology for freight train," in *Proc. Int. Conf. Transp. Mech. Electr. Eng. (TMEE)*, Dec. 2012, pp. 2046–2049, doi: 10.1109/TMEE.2011.6199618.
[20] G. J. Gutierrez, O. Garrido, R. Jimenez, and E. Mena, "Thermal characterization of a gray cast iron vented brake rotor during severe-braking using infrared thermography," in *Proc. 7th Int. Conf. Heat Transf. Fluid Mech. Thermodyn.*, Jul. 2010, pp. 1407–1412.
[21] N. Li, Z. Wei, and Z. Cao, "Automatic fault recognition for brake-shoe-key losing of freight train," *Optik*, vol. 126, no. 23, pp. 4735–4742, Dec. 2015, doi: 10.1016/j.ijleo.2015.07.120.
[22] M. Cerullo, G. Fazio, M. Fabbri, F. Muzi, and G. Sacerdoti, "Acoustic signal processing to diagnose transiting electric trains," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 238–243, Jun. 2005, doi: 10.1109/TITS.2005.848361.
[23] C. Pronello, "The measurement of train noise: A case study in northern Italy," *Transp. Res. D Transp. Environ.*, vol. 8, no. 2, pp. 113–128, Mar. 2003, doi: 10.1016/S1361-9209(02)00036-6.
[24] H. Huang, F. Liu, L. Geng, Y. Liu, Z. Ren, Y. Zhao, X. Lei, and X. Lu, "Fault diagnosis accuracy improvement using wayside rectangular microphone array for health monitoring of railway-vehicle wheel bearing," *IEEE Access*, vol. 7, pp. 87410–87424, 2019, doi: 10.1109/ACCESS.2019.2924832.
[25] C.-H. Jiang, "A fault diagnosis system of railway vehicles axle based on translation invariant wavelet," in *Proc. Int. Conf. Mach. Learn. Cybern.*, 2007, pp. 1045–1050, doi: 10.1109/ICMLC.2007.4370297.
[26] R. R. K. Jones and A. J. Packham, "Acoustic monitoring of rail-head roughness for targeted grinding and noise modelling," in *Proc. IET Int. Conf. Railway Condition Monitor.*, 2006, pp. 61–62.
[27] R. Janeliukstis and S. Kaewunruen, "A novel separation technique of flexural loading-induced acoustic emission sources in railway prestressed concrete sleepers," *IEEE Access*, vol. 7, pp. 51426–51440, 2019, doi: 10.1109/ACCESS.2019.2912011.
[28] J. R. Aguilar, R. A. Salinas, and H. E. Latorre, "Automatic recognition of working states of a rock breaker machine from acoustic signal processing," *J. Acoust. Soc. Amer.*, vol. 110, no. 5, p. 2766, Oct. 2001, doi: 10.1121/1.4777678.
[29] S. Zhao, "Intelligence expert system of transformer running state diagnosis based on acoustic signal analyzing," in *Proc. 2nd Int. Sym. Knowl. Acquis. Modeling*, Dec. 2009, pp. 52–55, doi: 10.1109/KAM.2009.153.
[30] J. Cao, W. Wang, J. Wang, and R. Wang, "Excavation equipment recognition based on novel acoustic statistical features," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4392–4404, Dec. 2017, doi: 10.1109/TCYB.2016.2609999.
[31] T. A. Stout, K. L. Gee, T. B. Neilsen, A. T. Wall, D. W. Krueger, and M. M. James, "Preliminary analysis of acoustic intensity in a military jet noise field," *J. Acoust. Soc. Amer.*, vol. 133, no. 5, p. 3421, May 2013, doi: 10.1121/1.4806000.
[32] J. Altmann, "Acoustic and seismic signals of heavy military vehicles for co-operative verification," *J. Sound Vib.*, vol. 273, nos. 4–5, pp. 713–740, Jun. 2004, doi: 10.1016/j.jsv.2003.05.002.
[33] L. B. Jackson, "Discrete Fourier Transform," in *Digital Filters and Signal Processing*. Boston, MA, USA: Springer, 1996, pp. 189–248.
[34] E. P. Serrano and M. A. Fabio, "Application of the wavelet transform to acoustic emission signals processing," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1270–1275, May 1996, doi: 10.1109/78.502340.
[35] Y.-S. Jeong, M. K. Jeong, and O. A. Omitaomu, "Weighted dynamic time warping for time series classification," *Pattern Recognit.*, vol. 44, no. 9, pp. 2231–2240, Sep. 2011, doi: 10.1016/j.patcog.2010.09.022.
[36] R. Scharrer and J. Fels, "Fuzzy sound field classification in devices with multiple acoustic sensors," in *Proc. Int. Workshop Acoust. Signal Enhanc. (IWAENC)*, Sep. 2012, pp. 1–4.
[37] A. Temko and C. Nadeu, "Classification of acoustic events using SVM-based clustering schemes," *Pattern Recognit.*, vol. 39, no. 4, pp. 682–694, Apr. 2006, doi: 10.1016/j.patcog.2005.11.005.
[38] J. W. Cooley, P. A. W. Lewis, and P. D. Welch, "The fast Fourier transform and its applications," *IEEE Trans. Educ.*, vol. 12, no. 1, pp. 27–34, Mar. 1969, doi: 10.1109/TE.1969.4320436.
[39] A. B. Hamida, M. Samet, N. Lakhoua, M. Drira, and J. Mouine, "Sound spectral processing based on fast Fourier transform applied to cochlear implant for the conception of a graphical spectrogram and for the generation of stimulating pulses," in *Proc. 24th Annu. Conf. IEEE Ind. Electron. Soc.*, 1998, pp. 1388–1393, doi: 10.1109/IECON.1998.722854.
[40] L. J. Winderbaum, I. Koch, O. J. R. Gustafsson, S. Meding, and P. Hoffmann, "Feature extraction for proteomics imaging mass spectrometry data," *Ann. Appl. Statist.*, vol. 9, no. 4, pp. 1973–1996, Dec. 2015, doi: 10.1214/15-AOAS870.
[41] L. Egghe and L. Leydesdorff, "The relation between Pearson's correlation coefficientrand Salton's cosine measure," *J. Amer. Soc. for Inf. Sci. Technol.*, vol. 60, no. 5, pp. 1027–1036, May 2009, doi: 10.1002/asi.21009.
[42] A. Mathur and G. M. Foody, "Multiclass and binary SVM classification: Implications for training and classification users," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 241–245, Apr. 2008, doi: 10.1109/LGRS.2008.915597.
[43] W. Chmielnicki and K. Stapor, "Combining one-versus-one and one-versus-all strategies to improve multiclass SVM classifier," in *Proc. 9th Int. Conf. Comput. Recognit. Sys. (CORES)*, Mar. 2016, pp. 37–45, doi: 10.1007/978-3-319-26227-7_4.
[44] S. Lecomte, R. Lengelle, C. Richard, F. Capman, and B. Ravera, "Abnormal events detection using unsupervised one-class SVM–Application to audio surveillance and evaluation -," in *Proc. 8th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2011, pp. 124–129, doi: 10.1109/AVSS.2011.6027306.
[45] Z. Liu, M. Zhang, F. Liu, and B. Zhang, "Fault diagnosis on the braking system of heavy-haul train based on multi-dimensional feature fusion and GBDT enhanced classification," *IEEE Transac. Ind. Inform.*, Mar. 2020, p. 9, doi: 10.1109/TII.2020.2979467.
[46] J.-C. Wang, J.-F. Wang, K. Wai He, and C.-S. Hsu, "Environmental sound classification using hybrid SVM/KNN classifier and MPEG-7 audio low-level descriptor," in *Proc. IEEE Int. Joint Conf. Neural Netw. Proc.*, Mar. 2006, pp. 1731–1735, doi: 10.1109/IJCNN.2006.246644.
[47] J. Xie, K. Hu, M. Zhu, J. Yu, and Q. Zhu, "Investigation of different CNN-based models for improved bird sound classification," *IEEE Access*, vol. 7, pp. 175353–175361, 2019, doi: 10.1109/ACCESS.2019.2957572.
[48] V.-T. Tran and W.-H. Tsai, "Acoustic-based emergency vehicle detection using convolutional neural networks," *IEEE Access*, vol. 8, pp. 75702–75713, 2020, doi: 10.1109/ACCESS.2020.2988986.
[49] K. Ko, S. Park, and H. Ko, "Convolutional feature vectors and support vector machine for animal sound classification," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 376–379, doi: 10.1109/EMBC.2018.8512408.
[50] J. T. Geiger, B. Schuller, and G. Rigoll, "Large-scale audio feature extraction and SVM for acoustic scene classification," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, Oct. 2013, pp. 1–4, doi: 10.1109/WASPAA.2013.6701857.

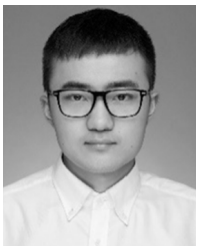
YULING YE received the B.S. degree in transportation engineering from Southwest Jiaotong University, Chengdu, China, in 1993, the M.S. degree in transportation engineering from Shanghai Tiedao University, Shanghai, China, in 1996, and the Ph.D. degree in transportation planning and management from Tongji University, Shanghai, in 2005.

She is currently a Professor with the College of Transportation Engineering, Tongji University. She is the author of four books and more than 60 articles. Her main research interests include the optimization technology of railway management and operation, the planning and design of railway infrastructure, and the economy and policy of railway transportation.



JUN ZHANG received the B.S. and M.S. degrees in transportation engineering from Chang'an University, Xi'an, China, in 2014 and 2017, respectively. He is currently pursuing the Ph.D. degree in transportation engineering with Tongji University, Shanghai, China.

He is the author of two inventions and more than ten articles. His current research interests include the smart rescheduling system of railway operation, the intelligent technology of railway management, and the planning method of rail transit networks.



HENGDA LIANG received the B.S. degree in mechanical and electronic engineering from Shanghai Maritime University, Shanghai, China, in 2014.

He is currently an Engineer with the Department of Logistics Planning and Control, SAIC Volkswagen Automotive Company Ltd. His current research interests include the computer control technology, the automatic detection method and application, and other mechatronic directions.