

Article

# A Novel Driving Noise Analysis Method for On-Road Traffic Detection

Qinglu Ma <sup>1</sup>, Lian Ma <sup>1,\*</sup>, Fengjie Liu <sup>1</sup> and Daniel (Jian) Sun <sup>2</sup> 

<sup>1</sup> School of Traffic and Transportation, Chongqing Jiaotong University, Chongqing 400074, China; mql@cqu.edu.cn (Q.M.); lfj@mails.cqjtu.edu.cn (F.L.)

<sup>2</sup> School of Future Transportation, University of Chang'an, Xi'an 710064, China; danielsun@sjtu.edu.cn

\* Correspondence: ml@mails.cqjtu.edu.cn

**Abstract:** Effective noise reduction and abnormal feature extraction are important for abnormal sound detection occurring in urban traffic operations. However, to improve the detection accuracy of continuous traffic flow and even overlapping vehicle bodies, effective methods capable to achieve accurate signal-to-noise ratio and appropriate characteristic parameters should be explored. In view of the disadvantages of traditional traffic detection methods, such as Short-Time Energy (STE) and Mel Frequency Cepstral Coefficients (MFCC), this study adopts an improved spectral subtraction method to analyze traffic noise. Through the feature fusion of STE and MFCC coefficients, an innovative feature parameter, E-MFCC, is obtained, assisting to propose a traffic noise detection solution based on Triangular Wave Analysis (TWA). APP Designer in MATLAB was used to establish a traffic detection simulation platform. The experimental results showed that compared with the accuracies of traffic detection using the traditional STE and MFCC methods as 67.77% and 76.01%, respectively, the detection accuracy of the proposed TWA is significantly improved, attaining 91%. The results demonstrated the effectiveness of the traffic detection method proposed in solving the overlapping problem, thus achieving accurate detection of road traffic volume and improving the efficiency of road operation.



**Citation:** Ma, Q.; Ma, L.; Liu, F.; Sun, D. A Novel Driving Noise Analysis Method for On-Road Traffic Detection. *Sensors* **2022**, *22*, 4230. <https://doi.org/10.3390/s22114230>

Academic Editor: Felipe Jiménez

Received: 4 April 2022

Accepted: 30 May 2022

Published: 1 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** traffic detection; traffic noise; feature extraction; Triangular Wave Analysis

## 1. Introduction

In many cities throughout of the world, the existing road network infrastructure needs to be maintained and improved, such as widening existing roads and making use of intelligent transportation system technologies. The number of on-road sensors were usually huge, and could not capture all traffic flow conditions for short-term traffic flow predictions. At the same time, traffic noise emissions may be related to the following attributes: e.g., power unit component; the interaction between speed, traffic flow type and road slope; rolling noise component; and the relation function between traffic speed and road surface [1]. As one of the major input data sources, a traffic noise spectral profile was used for the real-time traffic detection to obtain road traffic flow estimation in real time, with an average correct classification rate of about 96% [2]. Taking audio signals from two nearby sensors, the generalized cross correlation (GCC) function was combined with a particle filter to jointly estimate speed and wheelbase length. Using voice for road traffic monitoring, a design method for the microphone array was put forward based on the correlation function of acoustic observation vehicle trajectory [3]. Probabilistic noise models with more explicit road surface maker (RSM) features were developed to analyze the results of the RSM feature detection under various driving conditions [4]. Based on the probabilistic sensor model, an RSM model designed by Jo et al. a utilized particle filter to update the measurements, thereby improving the localization performance [5]. By extracting the peak power envelope in the traffic noise signal, Torija used a microphone array to record the sound of vehicles and estimate the number of vehicles, in which lanes were divided

according to the signal propagation time [6]. Audio sensors were used for traffic volume collection based on existing traffic detection sensors (coil and video detectors), and Support Vector Regression was introduced for audio data process [7]. Based on the traditional traffic noise prediction models and the improvement, it was stipulated that an efficient strategy processed a dual-channel audio record with a portable system. Analyzing the mechanism of the vehicle signal collected in detail, Steele et al. proposed a blind separation method to improve the classification of traffic noise [8,9]. To this end, a stereo signal recording system was proposed for detecting vehicles passing by and vehicle types (motorcycle, ordinary car and large truck) [10]. In order to develop an artificial neural network (ANN) model for predicting highway traffic noise, a method using the Levenberg-Marquardt (L-M) algorithm was proposed to train the multi-layer feed-forward back-propagation (BP) neural network. The results show that the ANN, a powerful traffic noise modeling technique, has a much smaller percentage difference than regression analysis [11]. Aiming at combining the genetic algorithm with machine learning, focusing on the features of traffic noise selection and vehicle classification, the correct classification rate was finally largely improved [12]. Based on MFCC representation to classify traffic flows of three different density levels, Borkar et al. developed a neuro-fuzzy classifier that achieved a classification of over 95% [13]. Linear SVM was used to extract vehicles from all approaching vehicles detected by radar sensors; meanwhile, hierarchical clustering was used to classify different traffic patterns according to the time series [14]. Emphasizing that traffic detection based on traffic sound is gradually becoming an important issue in the field of traffic flow parameter acquisition [15], when utilizing the noise differences perceived by passing road vehicles (approach, pass and departure), the major challenge is to improve the recognition accuracy of road noise events. By preprocessing the vehicle signal and extracting the features, Nooralahiyan et al. used the directional microphone to collect four vehicles: a small saloon (1300 VW), a medium saloon (2L Ford), a 250 cc motorcycle, and a Ford Diesel. Without controlling vehicle speed or background noise, the same equipment and 700 roadside recordings were made at road sites in the city, in which parameter transformation based on autocorrelation analysis was used to preprocess vehicle signals and extract features [16]. On the grounds of acoustics principle, Schclar et al. proposed a real-time vehicle automatic detection algorithm, which extracted spatiotemporal features from records using wavelet packet transform, and these features constituted the unique acoustic features of each recording [17]. Traffic noise signals were used to measure vehicle speed, and then the cruise speed is detected according to the magnitude relationship between the vehicle speed and the Doppler frequency shift. The traffic flow volume and vehicle types on different road levels were obtained to determine the weight of various noise influencing factors [18,19]. Due to the significant impact of driving signal feature selection on the performance of vehicle detection and identification, some essential features that can effectively represent vehicle signals need be extracted. Dinesh and Naveen introduce two different types of characteristics of traffic noise, Meir Frequency Cepstrum Coefficient and Wavelet Packet Transform, respectively. For binary traffic scene classification, MFCC features are up to 100% [20]. To solve the problem of poor recognition rate of overlapping noise signals, Dennis et al. extracted the local spectral features of noise signals, thus separating background noise and identify overlapping acoustic events, so as to provide a reference approach for overlapping signal recognition studies [21]. In the traffic detection method based on spectrum analysis, vehicle noise detection signals along certain corridor passages was obtained to perform spectrum detection and feature separation of multiple endpoints at the same time, thereby realizing the separation of overlapping vehicle bodies in congested road sections [22]. A smart phone-based sound measurement calibration protocol was tested under real conditions using a group of eight smart phones. Comparison with class 1 sound level meters at six control points showed an average error of  $-0.6 \pm 1.2$  dB for all handsets [23]. In the test area, 24 monitoring sensors were installed and six dynamic noise maps were obtained. By associating the prediction error of traffic noise with the error generated by the traffic flow model, Dynamap prediction was successfully improved, and the overall error was limited within about 3 dB [24]. Using

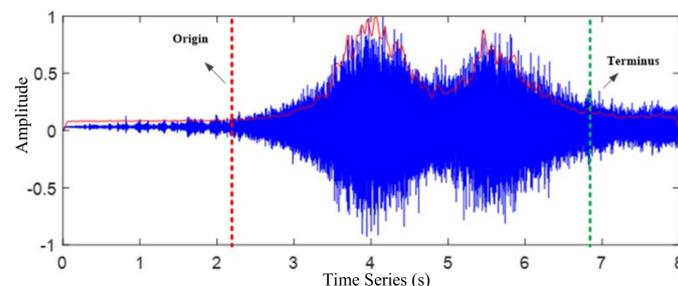
the local spectral features of sound to identify overlapping sound events and separate the background noise has a certain effect on the recognition of overlapping signals, and can be applied to the situation of overlapping traffic noise.

This paper proposed an improved spectral subtraction method to deal with traffic noise. A new feature parameter E-MFCC was defined through the feature fusion of STE and MFCC coefficient, so that a new traffic detection method based on traffic noise is proposed in the frame of TWA, to improve traffic detection accuracy. The remainder of the paper is structured as follows: Section 2.1 introduces the principle of traffic detection based on traffic noise, including pretreatments and analyses on the collected traffic noise data. Section 2.2 analyzes the characteristic parameters. With the field collected experimental data, the detection performance of the three traffic detection methods was assessed based on the experimental results in Section 2.3, so as to verify the effectiveness of the detection method. Finally, conclusions and future research directions are provided in Section 3.

## 2. Materials and Methods

### 2.1. Pretreatment and Characteristic Analysis of Driving Noise

Voice endpoint detection (VED) technology has been widely used in voice detection [25], with the main purpose being to distinguish voice and non-voice segments from the input signal. The core of using endpoint detection technology to detect traffic volume is to set thresholds and confirm vehicle signal frames. The existence of a vehicle is determined by the threshold value, as shown in Figure 1. According to the corresponding abscissa of the waveform diagram, the oversaturated road segment can be roughly determined. The red dot line within the figure represents the starting point of the vehicle road, and the green line is the ending point, both of which are roughly judged by the waveform of the driving noise signal. In this study, the low signal-to-noise ratio is not recommended to be judged out according to the waveform. For example, two vehicles were included in Figure 1, but unfortunately, only one was detected at this time. Therefore, it is necessary to set up a threshold, as shown in Figure 1, by which two sections of the waveform may be obtained.

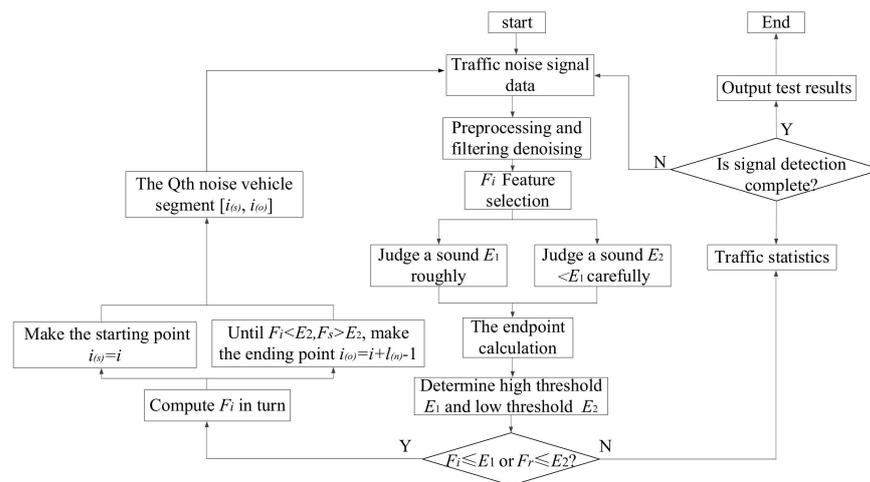


**Figure 1.** Traffic noise signal analysis.

#### 2.1.1. Pretreatment

Borkar et al. extracted the driving noise MFCC, and based on the Neuro-Fuzzy Classifier characterized by MFCC, SVM was used to classify of low traffic (40 km/h), medium traffic (20–40 km/h) and high traffic (0–20 km/h) at three different traffic density levels, showing that the classification accuracy was over 95% [26]. Kaur et al. collected traffic noise data in a ‘busy street’ and a ‘quiet street’, respectively, and extracted various time- and frequency-based features such as short-term zero crossing rate (ZCR), short-term energy (STE), root mean Square (RMS) and MFCC, yielding results with a better classification accuracy of 91.8% with Neural Network and 93% with SVM [27].

Vehicles driving on the road are often accompanied by the influence of abnormal short-term signals, such as vehicle whistle, emergency brake, etc., which may also be erroneously detected as vehicles passing. According to the high and low thresholds, the vehicle noise signal detection and analysis algorithm was proposed in Figure 2.



**Figure 2.** Flow chart of traffic detection and analysis algorithm.

Six steps are included in the algorithm as follows:

**Step 1:** The original traffic noise signal collected in real time on the road is imported into the processing system;

**Step 2:** The collected traffic noise signal is preprocessed and filtered to obtain smooth traffic noise signal data;

**Step 3:** Feature vectors are extracted as the configuration parameters and the eigenvalue  $F_i$  of each frame signal of driving noise is calculated successively;

**Step 4:** The high threshold is calculated as  $E_1 = (a + 0.2) * \max_{1 \leq j \leq n} E_j$ , the low threshold is calculated as  $E_2 = a * \max_{1 \leq j \leq n} E_j$ ,  $0 \leq a \leq 10$ , and increased by 0.1 per time so as to set up the minimum car signal frame;

**Step 5:** If  $F_i \leq E_1$  and  $F_r \leq E_2$ , ( $r = i + 1, i + 2, \dots, i + l_{(v)} - 1$ ), then  $i_{(s)} = i$ , where  $l_{(v)}$  is the minimum length of the vehicle segment and  $i_{(s)}$  is the initial endpoint of the  $Q^{\text{th}}$  driving noise segment; then,  $F_i$  is successively calculated from  $i = i + l$  until  $F_i < E_2$  and  $F_s < E_2$ , ( $s = i + 1, i + 2, \dots, i + l_{(n)} - 1$ ); then,  $i_{(o)} = i + l_{(n)} - 1$ , where  $l_{(n)}$  is the minimum length of the environmental noise segment, and  $i_{(s)}$  is the end point of the  $Q^{\text{th}}$  driving noise segment; and the interval  $[i_s, i_o]$  represents the  $Q^{\text{th}}$  driving noise segment. Repeat the preceding steps;

**Step 6:** The entire algorithm iterates between the detection state and vehicle signal state. To complete a cycle is to detect a vehicle passing, and the number of cycles equals the number of vehicles. By exploring all data sequences, the objective of road traffic flow detection can be achieved.

The preprocessing process mainly uses the following operation modes: pre-emphasis, windowing, framing, normalization and noise reduction [27]. After sampling the noise signal, a FIR high-pass filter called pre-emphasis of audio samples, is inserted to facilitate the analysis of audio samples. The purpose is to increase the high frequency resolution of the audio signal. Pre-emphasis processing has a certain inhibitory effect on the low frequency signal. After pre-emphasis, the high frequency component of the traffic noise signal is significantly increased, and the overall amplitude of the signal becomes smaller. The waveform of the traffic noise signal after pre-emphasis becomes smoother when there are no cars, which is beneficial to the subsequent signal processing and feature extraction.

### 2.1.2. Framing and Windowing

The traffic noise signal is a random signal with non-stationarity, which can be regarded as a quasi-steady state process within a rather short time range. During the traffic noise signal processing, the entire signal processing needs to be framed, generally in the range of

10–30 ms, to ensure the stability of the input signal. To this end, a traffic noise signal with length  $L$  is framed according to Equation (1).

$$\begin{cases} f = \sum_{i=1}^n \frac{L-X_i}{T_i} \\ N_i = X_i - T_i \end{cases} \quad (1)$$

where  $f$  is the total number of frames after splitting,  $L$  is the signal length,  $T_i$  is the displacement of  $i + 1^{\text{th}}$  frame to  $i^{\text{th}}$  frame, namely the frame shift,  $X_i$  is the overlapping part between the two frames,  $N_i$  is the frame length and  $X_i = N_i - T_i$ , and the final data is divided into  $f$  frames.

### 2.1.3. Normalized Processing

During driving noise signal collection, even if the same vehicle is affected by factors such as speed and location of acquisition equipment, the signal amplitude collected is different. To eliminate the influence of various factors on traffic detection, when the position of the weak traffic signal in the strong signal is obtained, a selection judgment is required. Euclidean distance is used as the judgment function. Because of the output of continuous signals in the audio, the collected signal amplitude is used to eliminate the influence and weight of the location of the collection device by using the distance discrimination index, and the expression is shown in Equation (2):

$$D(i, j) = \sum_{m=1}^M \sum_{n=1}^N [X^{ij}(m, n)]^2 - 2 \sum_{m=1}^M \sum_{n=1}^N X^{ij}(m, n)Y(m, n) + \sum_{m=1}^M \sum_{n=1}^N [Y(m, n)]^2 \quad (2)$$

$X^{ij}(m, n)$ ,  $Y(m, n)$  is the driving signal at any point in the selected area. After selecting the template,  $\sum_{m=1}^M \sum_{n=1}^N [X(m, n)]^2$ ,  $\sum_{m=1}^M \sum_{n=1}^N [Y(m, n)]^2$  is fixed; if  $D=0$ , the signal is equal; If  $D > 0$ , the normalization is shown in Equation (3):

$$R(i, j) = \frac{\sum_{m=1}^M \sum_{n=1}^N X^{ij}(m, n)Y(m, n)}{\sqrt{\sum_{m=1}^M \sum_{n=1}^N [X^{ij}(m, n)]^2} \sqrt{\sum_{m=1}^M \sum_{n=1}^N [Y(m, n)]^2}} \quad (3)$$

The normalization process eliminates the number level difference between the data of each dimension and avoids large calculation errors caused by the large difference of the data level of the feature data. To analyze the collected road vehicle data accurately, noise reduction is needed.

### 2.1.4. Improved Spectral Subtraction Noise Reduction

In the process of spectral subtraction, it is necessary to determine the length of the leading noise segment and the value of the sum of the parameters. The acquisition of an audio signal is completely random background noise; thus, when using spectral subtraction, it is likely the line value will be greater than this one. In this case, if spectral subtraction is used to reduce noise, the background noise environment cannot be removed and many burr peak points will be retained, which greatly reduces the effect of noise reduction. The flow chart of this method is shown in Figure 3. The improved spectral subtraction of multi-window spectral estimation is mainly based on the basic spectral subtraction technology. An orthogonal data window is improved to multiple orthogonal data windows, and the red dot line within Figure 3 represents the improved part.

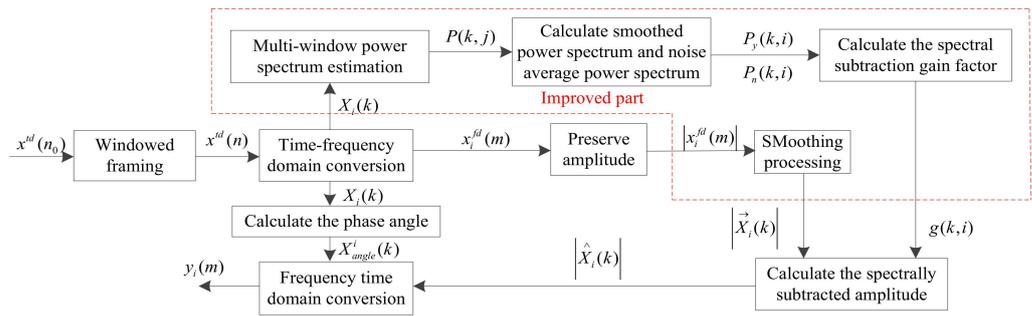


Figure 3. Improved spectral subtraction.

The original traffic noise signal time domain sequence is  $x^{td}(n_0)$ , the signal sequence after pretreatment is  $x^{td}(n)$  and  $x^{td}(n)$  is expressed as the  $i^{th}$  frame traffic noise signal. The general spectral subtraction noise reduction steps include time-frequency domain conversion, noise estimation, phase angle calculation, spectral subtraction and the five steps of frequency-time domain conversion. The three steps of time-frequency domain conversion, noise estimation and phase angle calculation are important to the pre-work of spectral subtraction, as shown in Equation (4) below:

$$\begin{cases} x_i^{fd}(m) = \sum_{n=0}^{N-1} x_i^{td}(n) \exp(-j\frac{2\pi nm}{N}) \\ D(m) = \frac{1}{NIS} \sum_{i=1}^{NIS} |x_i^{fd}(m)|^2 \\ x_i^{fd(angle)}(m) = \arctan \left[ \frac{\text{Im}(x_i^{fd}(m))}{\text{Re}(x_i^{fd}(m))} \right] \end{cases} \quad (4)$$

where  $x_i^{fd}(m)$  is the frequency domain signal of  $x_i^{td}(n)$  after fast Fourier transform,  $m = 0, 1, \dots, N-1$ ,  $D(m)$  is the average energy of the leading non-vehicle signal segment in the collected traffic noise signal,  $IS$  is the time length of the leading noise segment (composed of the leading non-vehicle segment) and  $NIS$  is the corresponding frame number. The noise signal amplitude is  $|x_i^{fd}(m)|$ .  $x_i^{fd_1}(m)$  represents the phase angle of frame  $i$  of  $x_i^{fd}(m)$ ,  $\text{Im}(x_i^{fd}(m))$  is the imaginary part of  $x_i^{fd}(m)$  and  $\text{Re}(x_i^{fd}(m))$  is the real part.

The spectral subtraction process reduces the energy of the frequency domain signal  $x_i^{fd}(m)$  and the average noise energy of each frame, and the amplitude value  $|x_i^{ss}(m)|$  after spectral subtraction is realized as follows in Equation (5):

$$|x_i^{ss}(m)|^2 = \begin{cases} |x_i^{fd}(m)|^2 - a|D(m)|^2 & , |x_i^{fd}(m)|^2 - |D(m)|^2 > 0 \\ b|D(m)|^2 & , \text{else} \end{cases} \quad (5)$$

where  $a$  is a minus factor and  $b$  is a gain compensation factor. Then, the signal sequence  $y_i(k)$  after spectral subtraction is obtained by combining them. The basic spectral subtraction only uses one data window in the process of noise reduction, and it is improved through multi-window spectral estimation. The flow chart of this method is shown in Figure 3.

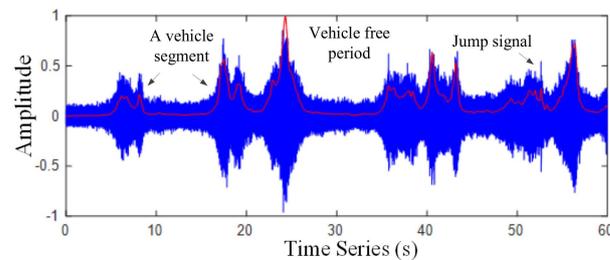
$$\begin{cases} P(k, i) = \text{PMTM} [x^{td}(n)] \\ P_y(k, i) = \frac{1}{2M+1} \sum_{j=-M}^M P(k, i+j) \\ |\vec{X}_i(k)| = \frac{1}{2M+1} \sum_{j=-M}^M |X_{i+j}(k)| \end{cases} \quad (6)$$

where the multi-window spectral power spectral density  $P(k, i)$  and  $P_y(k, i)$  are the multi-window spectral power density and the smoothed power spectral density of the  $k$ th spectral line of the  $i$ th frame. PMTM means to estimate the multi-window spectral power spectral density and the first and last  $M$  frames of the  $i$ th frame as the center. Then,  $2M + 1$  frames are averaged.  $|\vec{X}_i(k)|$  is the average magnitude spectrum,  $g(k, i)$  is the gain factor and the magnitude spectrum after spectral subtraction is  $|\hat{X}_i(k)|$ .

The short time energy of the traffic noise signal  $y_i(n)$  in  $i$ th frame is calculated as shown in Equation (7):

$$E(i) = \sum_{n=0}^{L-1} y_i^2(n) \quad (7)$$

where  $y_i(n)$  is the value of a frame and  $n = 1, 2, \dots, L$ ,  $1 \leq i \leq fn$ .  $L$  is the frame length and the square sum of the amplitude of the  $i$ th frame signal is the STE value of the corresponding time point of the signal. Figure 4 represents the original waveform of a section of traffic noise and the corresponding STE. In Figure 4, the STE of the traffic noise signal changes with time, and the energy difference between the congested segments with and without car is significant.

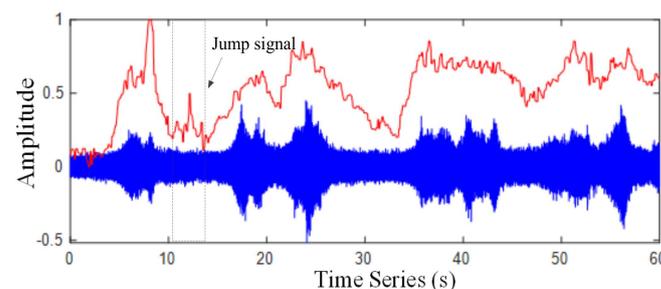


**Figure 4.** Waveform of traffic noise signal and its STE.

In the case of great environmental noise, the existence of congested segments should be identified, so that the STE feature extraction algorithm becomes relatively easy to implement, which saves the processing time in the case of large amounts of data. Zero Crossing Rate (ZCR) indicates the number of times a voice signal waveform passes through the horizontal axis (zero level) in a frame of speech [28]. The definition of short-time average zero crossing rate is shown in Equation (8):

$$Z(i) = \frac{1}{2N} \sum_{n=0}^{L-1} |\text{sgn}[y_i(n)] - \text{sgn}[y_i(n-1)]| \quad (8)$$

where  $\text{sgn}[\cdot]$  is a sign function and  $\text{sgn}[x] = 1$ , when  $x \geq 0$ ; otherwise,  $\text{sgn}[x] = -1$ .  $N$  is the length of the selected window. The original waveform of a section of traffic noise and the corresponding Zero Crossing Rate are shown in Figure 5.



**Figure 5.** Waveform of traffic noise signal and the corresponding zero crossing rate.

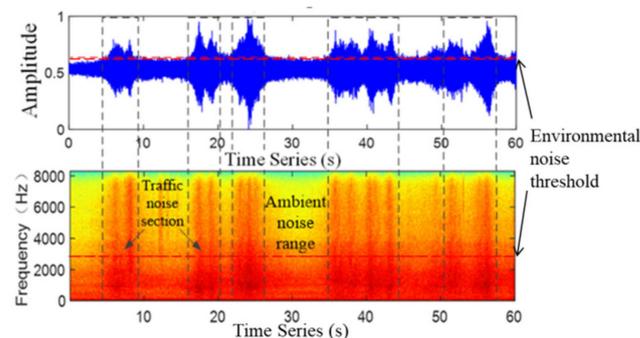
The sonogram reflects the dynamic spectrum characteristics of the sound signal, which plays an important role in signal processing [29]. In the sonogram, the abscissa represents the time and the ordinate represents the amplitude. Because the three-dimensional information is mapped to the two-dimensional plane, the amplitude is expressed by the depth of the color. The deeper color indicates large amplitude at this time. The result of FFT of traffic noise signal is shown in Equation (9) below:

$$Y_i(k) = \sum_{n=1}^N y_i(n-1) \exp\left(-j \frac{2\pi(n-1)(k-1)}{N}\right) \quad (9)$$

The spectrum characteristics of traffic noise signal calculated can be expressed by the matrix shown in Equation (10).

$$A_{sv} = [a_i]_{m \times L} = \begin{bmatrix} |Y_1(1)| & |Y_2(2)| & \dots & |Y_L(1)| \\ |Y_1(2)| & |Y_2(2)| & \dots & |Y_L(2)| \\ \dots & \dots & \dots & \dots \\ |Y_1(m)| & |Y_2(m)| & \dots & |Y_L(m)| \end{bmatrix} \quad (10)$$

where  $A_{sv}$  represents the spectrum feature and  $a_i$  represents the amplitude value of the traffic noise signal after FFT transformation.  $m = 1, 2, \dots, [N/2]_{\text{int}} + 1$ , and when the  $k$  indicates 1,  $|Y_i(k)|$  and  $|Y_i(N-k+2)|$  is equal, only  $([N/2]_{\text{int}} + 1)$  sample points are needed. According to Equations (8) and (9), the spectrum feature of the traffic noise signal is extracted, as shown in Figure 6.



**Figure 6.** Waveform of traffic noise signal and its sonogram.

The Figure 6 represents a section of traffic noise signal waveform and its corresponding spectrogram extraction results, respectively. The frequency of the signal collected in the road environment is found below 8000 Hz, and the color of the traffic noise section is deeper than that of the ambient noise section. The traffic noise information is continuously distributed from the low frequency region to the medium and high frequency region, whereas the environmental noise information is mainly concentrated in the low and medium frequency region. It can be figured out from Figure 6 that above the intermediate frequency region of about 3000 Hz, the frequency distribution of the environmental noise signal is rather small, and the frequency relative to the traffic noise signal can be ignored. Therefore, the existence of congested segments can be easily judged by setting the frequency threshold.

## 2.2. Traffic Detection Based on Triangle Wave Analysis of Traffic Noise

### 2.2.1. Feature Extraction and Fusion

This study proposes the Triangular Wave Analysis (TWA) technology for traffic volume processing. Through the analyses of STE and MFCC characteristic traffic detection methods in [30], both methods have certain limitations in detecting the performance of overlapping congested segments. The indicator of STE ( $E_i$ ) of the traffic noise signal is calculated by taking  $d_m^0(i, n)$  from  $d_m(i, n)$  in MFCC, and the first and last two frames of the short-time

energy  $E_i$  of the traffic noise signal are discarded because the frames were not included in calculating  $d_m(i, n)$ . To match the length of  $d_m^0(i, n)$ , the short-time energy  $E_i$  is also discarded. The  $E_i$  is then multiplied by the index, that is, the characteristics of the frame  $i$  signal of  $E$ -MFCC.

$E$ -MFCC does not need to set the front NIS frame audio by the MFCC describing the background noise, so as to calculate the average. The  $E$ -MFCC feature is feasible for traffic detection, which is superior to Short Term Energy and MFCC.

### 2.2.2. Extremum Extraction

For the extracted new feature envelope, the extremum extraction method searches the local extremum points of the numerical sequence to extract the extremum of the original sequence. The new feature  $E$ -MFCC is taken as the original data of the traffic noise signal after digitization. Assuming that the original data matrix is  $X$ , Equation (11) can be expressed as follows:

$$X = \begin{bmatrix} T_1, T_2, \dots, T_i, \dots, T_m \\ P_1, P_2, \dots, P_i, \dots, P_m \end{bmatrix}^T \quad (11)$$

where  $X$  is the original data matrix,  $i = 1, 2, \dots, m$ ,  $T_i, P_i$  are the original time and amplitude, respectively,  $i$  is the amount of original data and  $m$  is the end. The specific process is as follows:

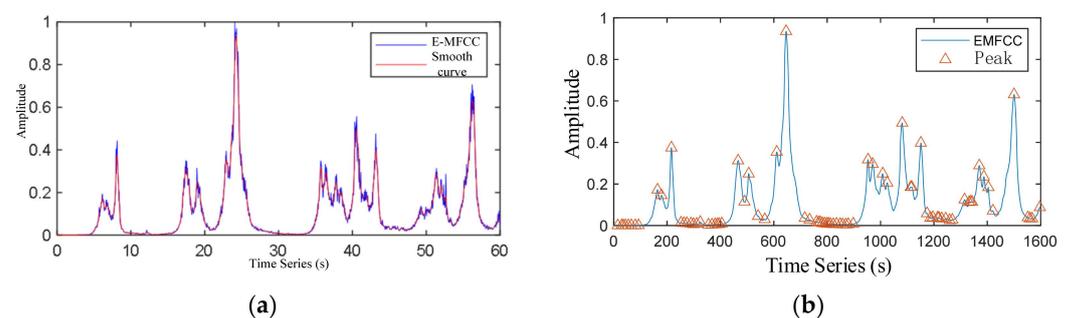
First, the beginning and end of initialization are presented as Equation (12):

$$\begin{cases} E(1, 1) = T_1; E(1, 2) = P_1 \\ E(d, 1) = T_m; E(d, 2) = P_m \end{cases} \quad (12)$$

where  $E(\cdot)$  is the extremum matrix and  $d$  is the number of extremum points. Then, judging from the given conditions: when  $P_{i-1} > P_i > P_{i+1}$  is satisfied, it is the minimum point; when  $P_{i-1} < P_i < P_{i+1}$  is satisfied, it is the maximum point. The corresponding extremum point  $T_i$  and  $P_i$  are stored in  $E(\cdot)$ .

The above two steps are the basic principle and implementation process of extremum extraction.

The extracted  $E$ -MFCC features and the smoothed  $E$ -MFCC features are shown in Figure 7, where the peak represents the extreme, and the waveform after extremum extraction of smoothed  $E$ -MFCC characteristic curve is shown in Figure 7b.



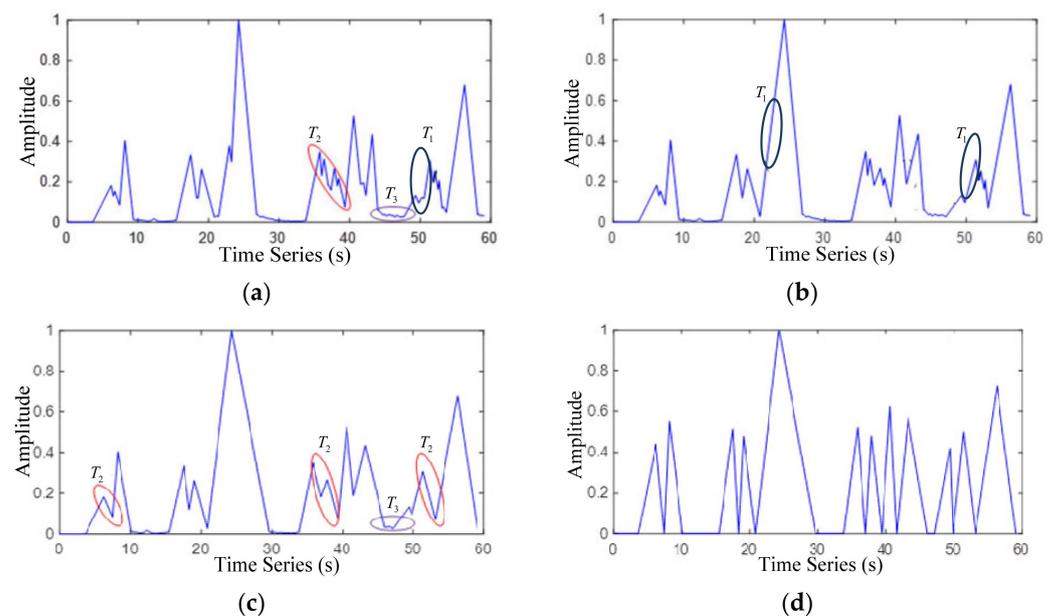
**Figure 7.**  $E$ -MFCC characteristic curve and extremum extraction. (a)  $E$ -MFCC characteristic curve and its smoothed characteristic curve and (b)  $E$ -MFCC extremum extraction.

### 2.2.3. Formation and Combination of Triangular Waves

The formation and combination of triangular waves mainly includes four parts: triangular wave formation, combination on triangular wave, combination under triangular wave and expansion of frame width.

$T_1$  is the interference triangle wave which appears in monotonic increasing,  $T_2$  is the interference triangle wave which appears in monotonic decreasing and  $T_3$  is the interference triangle wave which includes  $T_1$  and  $T_2$ . In order to achieve satisfactory detection

accuracy, it is necessary to carry out triangle wave up combinations, triangle wave down combinations and frame width expansion for these three cases. Figure 8b represents the combination of triangular waves, which is the first step of the combination rule of triangular waves. It is a solution to the  $T_1$  situation. Comparing Figure 8b with Figure 8a, the monotone increasing interference triangular wave of type  $T_1$  has been eliminated after the combination operation on the triangular wave. Figure 8c is the second step of the combination rule of triangular wave. It is a solution to  $T_2$  case. After the combination of triangular wave up and down processing, the peak value of some triangular waves is still very small; that is, the signal in some environmental noise segments is very weak. Comparing Figure 8c with Figure 8d, it can be figured out that after the operation of triangle waves lowers, the combination algorithm, the monotone decreasing interference triangle wave of type  $T_2$  has been eliminated. At the same time, after the combination of up and down triangle waves,  $T_3$  interference wave is basically eliminated. The waveform formed by frame width expansion is shown in Figure 8d. The frame width expansion algorithm is designed to solve the problem of abnormal noise (whistle, birdsong). The core idea is to expand the frame width of overlapped congested segment. By setting the minimum frame length of the congested segment, the traffic volume is obtained. The triangle wave analysis algorithm solves the problem of separating overlapping vehicle sections and can be applied.



**Figure 8.** Formation and combination of triangular waves. (a) triangular wave formation, (b) combination on triangular wave, (c) combination under triangular wave and (d) frame width extension.

### 2.3. Experiment and Empirical Cases

#### 2.3.1. Experimental Data Acquisition

The Nanbin Road segment between the Caiyuanba Yangtze River Bridge and Chongqing Yangtze River Bridge was selected as the experimental data acquisition section. The entire road lays out from west to east along the Yangtze River (the north side adjacent to the river), with a total length around 1.3 km, and the south side is surrounded by mountains with no redundant branch section. The segment is intentionally selected to avoid pedestrian noise disturbance, as no shops, restaurants etc. are nearby, thus ensuring the quality of traffic noise collection. For urban arterial roads, noise intervention for pedestrians and noise intervention for shops are included and the data is collected from factors such as congested roads, etc. The data collection is seriously disturbed, so the two-lane off-peak hours are selected to collect the traffic noise. The driving noise signal can be converted

from the time domain to the frequency domain by the fast Fourier transform (FFT), and then the spectrogram in the signal is extracted and analyzed.

The data collection points and collection schematic diagram are shown in Figure 9. As shown in Figure 9b, the equipment used in the study included a recording pen, a mobile phone and a computer. The recording pen is used for lossless recording with a 1536 KBPS/48 kHz sampling rate, and the recording format is WAV, supporting line-in recording and built-in microphone recording. The sensitivity of the microphone is  $-58 \pm 3$  dB, the working temperature is  $-25\sim 70$  °C and it has full directivity, electret type capacitor microphone-to-text. The frame length is set to 25 ms (namely 1200 sample points) and the frame shift is set to 12.5 ms. The video data were synchronously collected by the smart phone to compare the actual traffic volume. A video corresponding to the same segment will be recorded using the smartphone, to obtain the statistics of the vehicles on the road, which will be verified as the actual traffic volume of the experiment. Then, save the recorded traffic noise data in WAV format, and save the video data in MP4 format as a reference to obtain accurate traffic volume. The collected data were processed by Premiere Pro CC 2017 to remove the reverse traffic noise data and count the actual traffic volume for 60 min. The final synthetic sample was named ‘traffic noise’, and the actual traffic volume is 667 pcu.



**Figure 9.** Schematic diagram of data acquisition. (a) selection of data acquisition points and (b) data acquisition (left bracket: a recording pen, right bracket: a smartphone).

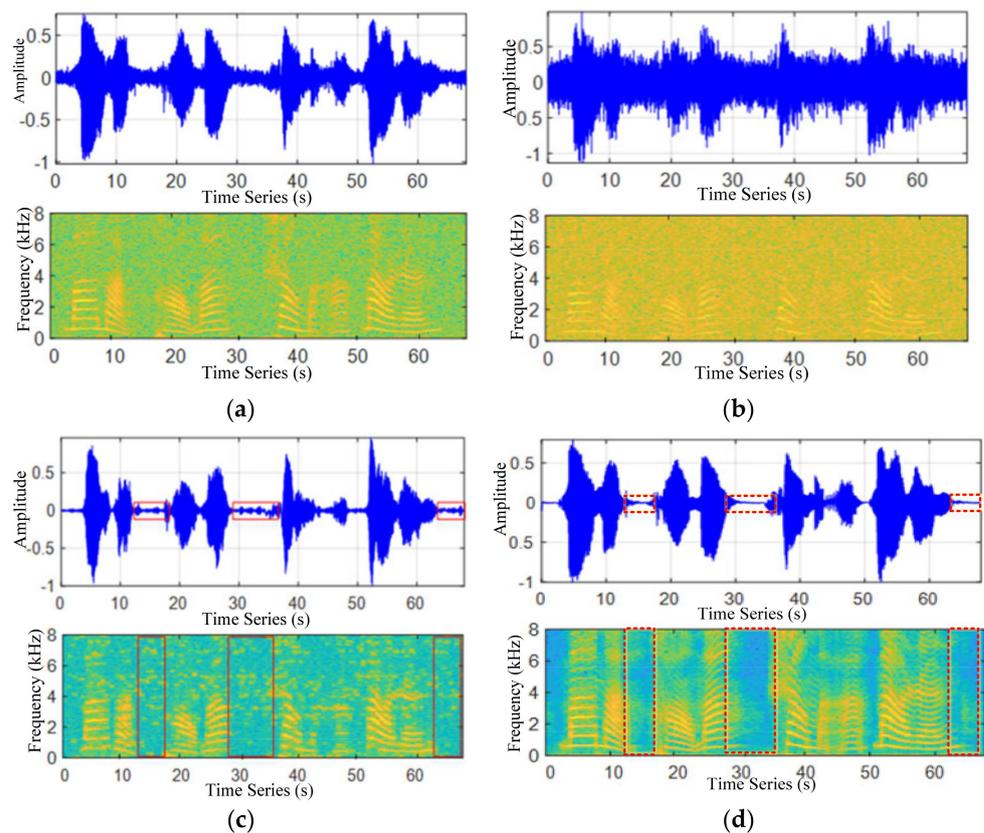
### 2.3.2. Evaluating Indicator

According to Ma et al. [31], the traffic detection evaluation indexes in this study were selected as traffic detection accuracy  $r_c$ , false detection rate  $r_w$  and missed detection rate  $r_m$ , with definitions provided in Equation (13) below:

$$\begin{cases} r_c = \frac{n_{(t,c)}}{n_a} \times 100\% \\ r_w = \frac{n_t - n_{(t,c)}}{n_a} \times 100\% \\ r_m = \frac{n_a - n_t}{n_a} \times 100\% \end{cases} \quad (13)$$

where  $n_a$  is the actual number of vehicles in video synchronous acquisition data,  $n_t$  is the total number of vehicles detected by traffic noise,  $n_{(t,c)}$  is the correct number of vehicles in the total number of detected vehicles,  $n_{(t,w)}$  is the number of vehicles wrongly detected from the noise signal,  $n_{(t,w)} = n_t - n_{(t,c)}$  and  $n_m$  is the number of undetected vehicles. The basic spectral subtraction and improved spectral subtraction are used to filter and denoise a section of original traffic noise signal, as shown in Figure 10.

Background from the environmental noise (distant vehicle) is superimposed. For Figure 10c, it can be seen that the traffic noise after noise reduction by spectral subtraction has an obvious discrimination compared with the original waveform, in which noise is marked in the red rectangular box. For Figure 10d, after filtering and noise reduction, when using the improved spectral subtraction, the signal of the non-congested segment is close to zero within the time domain diagram, which is ignorable, as marked in the red dotted rectangular box. In the spectrogram, there is no frequency distribution in the non-vehicle section, and the spectrum of the vehicle noise signal section is smoother.

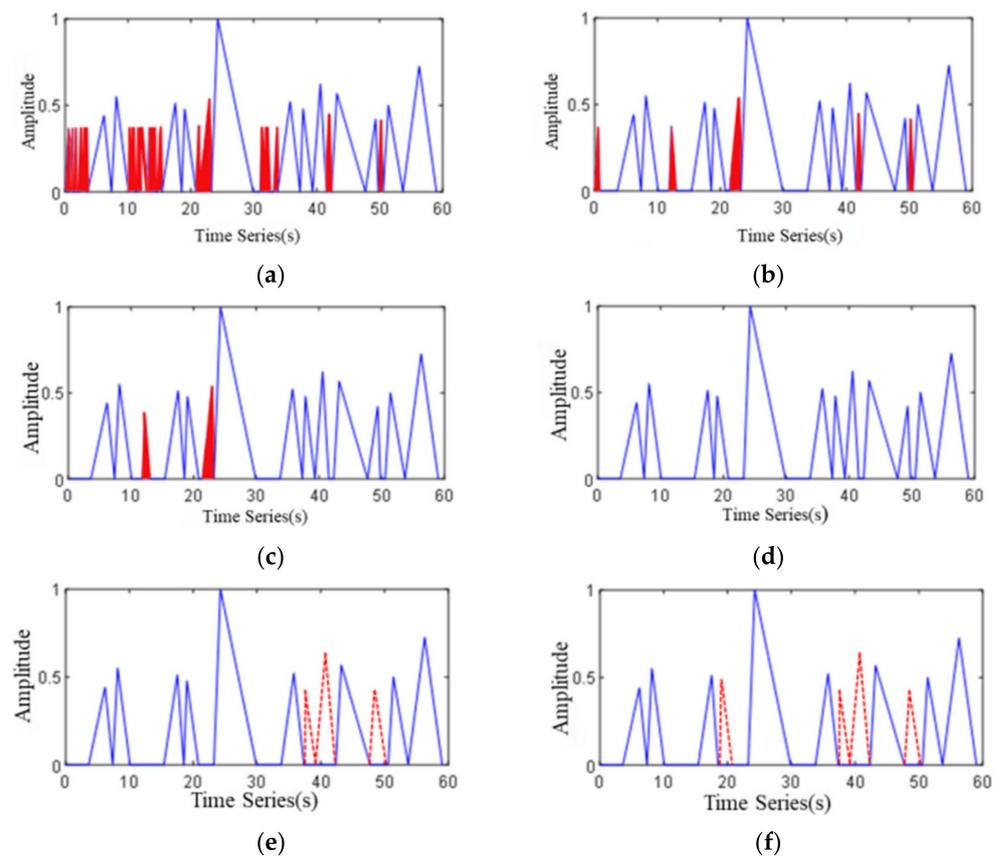


**Figure 10.** Noise reduction algorithm sonogram and waveform comparison. (a) original driving noise signal and its spectrogram, (b) noised-driving noise signal and its spectrogram, (c) spectral subtraction driving noise signal and its spectrogram and (d) improved spectral subtraction for driving noise signal and its spectrogram.

### 2.3.3. Noise Reduction Performance Comparison

A traffic statistics system based on traffic noise signal is established by App Designer in MATLAB R2020a. Different frame lengths directly affect the accuracy of detection. With regard to TWA characteristics, the parameter setting of the minimum frame length of the traffic segment signal is the key to detect the traffic volume.

Taking 10 frames as the initial minimum frame length of the vehicle signal segment, 20 frames are successively increased to verify the traffic noise data of a segment with a traffic flow of 12 vehicles, as presented in Figure 11. In the Figure, the red triangle represents the triangular waves detected more than once, whereas the triangular box with a red dotted line represents the triangular waves missed. As can be figured out from Figure 11a–c, with the increase of frame length, the number of detected triangular waves becomes closer to the real number of vehicles, and the number of detected triangular waves decreases from six to two. In Figure 11d, when attaining 40 frames, the number of detected triangular waves equals the real vehicles. As the transition of frames increases, few triangular waves are detected and three vehicles missed detection at 50 frames and four vehicles missed detection at 60 frames. That is to say, when the minimum frame length of the vehicle signal section is set within certain range, the number of vehicles can be accurately detected. Through multiple experiments and comparisons, the minimum frame length of the vehicle signal segment was set within the range of 35–45 frames.



**Figure 11.** Comparison of the minimum frame length of different vehicle signal segments. (a) 10-frame, (b) 20-frame, (c) 30-frame, (d) 40-frame, (e) 50-frame and (f) 60-frame.

#### 2.3.4. Establishment of Simulation Platform

First, the platform selection area for traffic noise signal acquisition and feature extraction operation plate, acoustic signal acquisition acoustic signal reading are determined. Then, noise signal denoising and preprocessing occurs, followed by the selection of a traffic detection algorithm, which is divided into the STE detection algorithm, MFCC detection algorithm and TWA algorithm.

Among the three traffic detection algorithms, traffic detection based on STE needs to consider the selection of high and low thresholds. Traffic detected by the three detection algorithms is presented in Figure 12. Referring to the threshold parameter selection method of endpoint detection in Zhang and Pan [32], the high and low threshold parameters were set as 0.13 and 0.11, respectively, in order to obtain the number of vehicles closer to the field situations. TWA is used for traffic detection, and the minimum length of vehicle signal segment is set as 40 frames, and other parameters are the same as MFCC detection.

#### 2.3.5. Test Results and Analysis

Since the collected traffic noise data are large, the 60 min traffic noise signal acquisition data are divided into 1 to 12 groups with the length of 300 s in each group. Figure 13 represents the detection results of the first group of noise signals, with the statistical results provided in Table 1.

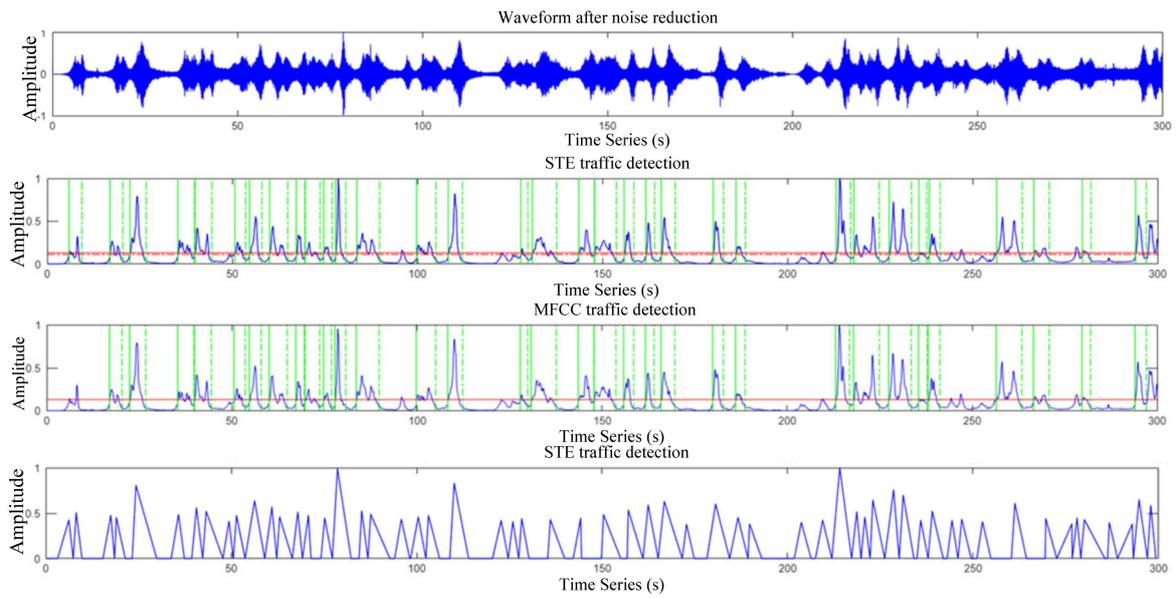


Figure 12. Waveform after noise reduction and three traffic detection characteristic curves.

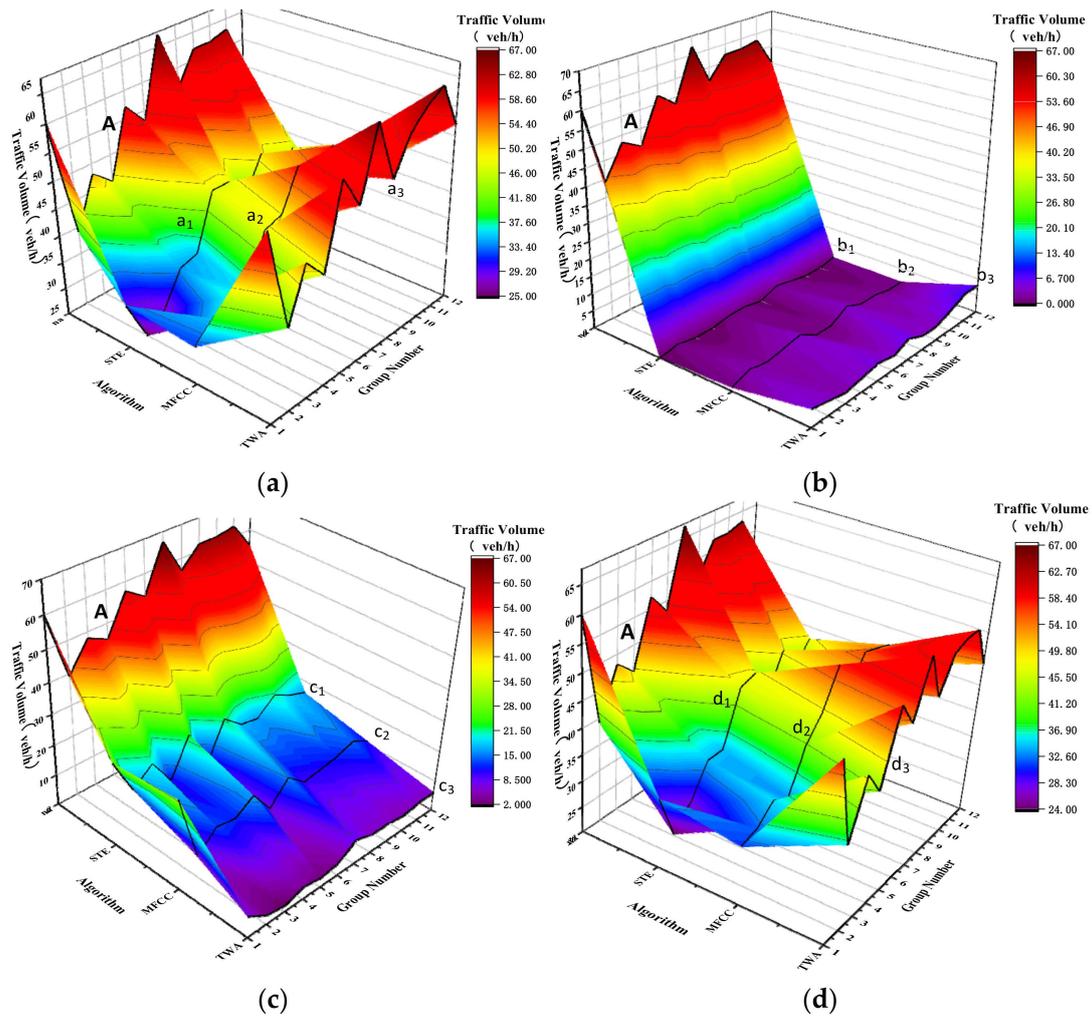


Figure 13. Analysis diagram of detection results of different algorithms. (a)  $n_t$  comparison, (b)  $n_{(t,w)}$  comparison, (c)  $n_m$  comparison and (d)  $n_{(t,c)}$  comparison.

**Table 1.** Results of three traffic detection algorithms.

| Group | $n_a$ | $n_t^s$ | $n_{(t,w)}^s$ | $n_m^s$ | $n_t^m$ | $n_{(t,w)}^m$ | $n_m^m$ | $n_t^t$ | $n_{(t,w)}^t$ | $n_m^t$ |
|-------|-------|---------|---------------|---------|---------|---------------|---------|---------|---------------|---------|
| 1     | 60    | 33      | 0             | 27      | 32      | 1             | 29      | 58      | 5             | 7       |
| 2     | 39    | 25      | 1             | 15      | 35      | 3             | 7       | 40      | 4             | 3       |
| 3     | 48    | 28      | 0             | 20      | 37      | 2             | 13      | 49      | 3             | 2       |
| 4     | 45    | 34      | 0             | 11      | 36      | 3             | 12      | 45      | 4             | 4       |
| 5     | 57    | 35      | 1             | 23      | 45      | 4             | 16      | 59      | 5             | 3       |
| 6     | 53    | 45      | 1             | 9       | 46      | 1             | 8       | 53      | 4             | 4       |
| 7     | 67    | 45      | 0             | 22      | 53      | 1             | 15      | 65      | 6             | 8       |
| 8     | 56    | 41      | 2             | 17      | 50      | 4             | 10      | 54      | 4             | 6       |
| 9     | 62    | 47      | 1             | 16      | 53      | 2             | 11      | 60      | 4             | 6       |
| 10    | 62    | 41      | 0             | 21      | 52      | 2             | 12      | 63      | 6             | 5       |
| 11    | 63    | 45      | 0             | 18      | 50      | 1             | 14      | 65      | 8             | 6       |
| 12    | 55    | 40      | 1             | 16      | 45      | 1             | 11      | 57      | 8             | 6       |

Table 1 presents the statistical results of 12 groups of different detection algorithms, in which  $n_{(\bullet)}^t, n_{(\bullet)}^m$  and  $n_{(\bullet)}^s$  represent the corresponding number of vehicles of STE, MFCC and TWA algorithms, respectively. The corresponding 3D views are provided in Figure 13.

The broken line A within the Figure is the distribution broken line of  $n_a, a_1, a_2$  and  $a_3$  represent the  $n_t$  distribution under the action of STE, MFCC and TWA methods;  $b_1, b_2$  and  $b_3$  represent the  $n_{(t,w)}$  distribution under the action of STE, MFCC and TWA methods;  $c_1, c_2$  and  $c_3$  represent the  $n_m$  distribution under the action of STE, MFCC and TWA methods;  $d_1, d_2$  and  $d_3$  represent the  $n_{(t,c)}$  distribution under the action of STE, MFCC and TWA methods. From the Figure 13, the distributions of  $n_t$  and  $n_{(t,c)}$  were found to be the closest to that of the broken with the lowest  $n_m$  distribution. The statistical results of the three detection algorithms are shown in Table 2.

**Table 2.** Comparison of detection results of different algorithms.

| Methods | Missed Detection Rate | False Detection Rate | Accuracy |
|---------|-----------------------|----------------------|----------|
| STE     | 32.23%                | 1.05%                | 67.77%   |
| MFCC    | 23.69%                | 3.75%                | 76.01%   |
| TWA     | 9%                    | 9.15%                | 91%      |

In the second column of Table 2, the vehicle missing rate of the three methods are provided. The average missing rate of TWA is 9%; that of MFCC is 23.69%; and that of STE is 32.23%,  $R_1^T < R_1^M < R_1^S$ .

In the third column of the table, the vehicle error detection rate of the three methods is provided, in which the average error detection rate of TWA is 9.15%; that of MFCC is 3.75%; and that of STE is 1.05%,  $R_2^S < R_2^M < R_2^T$ . In the fourth column of the table, the vehicle accuracy of the three methods was provided, in which the average accuracy of TWA attains 91%, much higher than that of MFCC (76.01%) and STE (67.77%),  $R_3^S < R_3^M < R_3^T$ . TWA can better achieve the detection of overlapping car bodies and has good stability within the time-varying environment.

### 3. Conclusions and Future Research Directions

This paper proposes a novel traffic noise analysis method for on-road traffic detection. The innovative feature parameter E-MFCC was defined for the fusion of STE and MFCC principal component features. Then, the extremum point was identified by exploring the entire characteristic curve, in which the overlapped segment signals were separated by triangular wave algorithms. The final triangular wave number is the number of vehicles detected in the running noise signal, which has theoretical and practical significance to expand and improve current traffic detection technologies, e.g., by loop detector or video camera sensors.

The average accuracy of STE, MFCC and TWA were calculated as 67.77%, 76.01% and 91%, respectively, indicating that TWA has rather good accuracy and is effective to detect traffic volume. Although the results are promising, this study indeed has limitations with regards to the data and the approach. First, this study only focuses on one-way lane traffic detection, without considering the two-way lane, which may lead to underutilization of information. In the follow-up work, the applicability of TWA in multi-lane processes will be investigated. Furthermore, during traffic noise pretreatment and feature extraction, window function, frame length, frame shift and feature dimension were mostly selected by personal experiences, without a set of mature theories or methods to guide the parameters' configuration. Different parameter settings may help to shed lights on investigating the result of the study. Driving sound is not the end of traffic detection, because different models need to be converted into Passenger Car Units to get more accurate traffic statistics; therefore, the next part of the study will involve vehicle type recognition research based on the driving voice, and theoretical research efforts are required to improve result quality in the field.

**Author Contributions:** The authors confirm contribution to the paper as follows: study conception and design: Q.M. and D.S.; data collection: L.M. and F.L.; analysis and interpretation of results: Q.M., draft manuscript preparation: L.M., F.L. and D.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Social Science Fund National Emergency Management System Construction Research special project: Research on the path of improving the quality and upgrading of urban comprehensive emergency management capability under major emergencies, grant number 20VYJ023 and this research was funded by National Key Research and Development Program of China: Intelligent perception theory and method for road infrastructure, grant number 2018YFB1600200.

**Institutional Review Board Statement:** Not available.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## References

1. Hamet, J.F.; Besnard, F.; Doisy, S.; Lelong, J.; Le Duc, E. New vehicle noise emission for French traffic noise prediction. *Appl. Acoust.* **2010**, *71*, 861–869. [[CrossRef](#)]
2. Torija, J.A.; Ruiz, P.D. Using recorded sound spectra profile as input data for real-time short-term urban road-traffic-flow estimation. *Sci. Total Environ.* **2012**, *435*, 270–279. [[CrossRef](#)] [[PubMed](#)]
3. Marmaroli, P.; Carmona, M.; Odobez, J.M.; Falourd, X.; Lissek, H. Observation of Vehicle Axles Through Pass-by Noise: A Strategy of Microphone Array Design. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 1654–1664. [[CrossRef](#)]
4. Feng, C.; Chen, S. Reliability-based assessment of vehicle safety in adverse driving conditions. *Transp. Res. Part C Emerg. Technol.* **2011**, *19*, 156–168.
5. Jo, K.; Jo, Y.; Suhr, J.K.; Jung, H.G.; Sunwoo, M. Precise Localization of an Autonomous Car Based on Probabilistic Noise Models of Road Surface Marker Features Using Multiple Cameras. *IEEE Trans. Intell. Transp. Syst.* **2016**, *6*, 3377–3392. [[CrossRef](#)]
6. Torija, J.A.; Ruiz, D.P. Automated classification of urban locations for environmental noise impact assessment on the basis of road-traffic content. *Expert Syst. Appl.* **2016**, *53*, 1–13. [[CrossRef](#)]
7. Lefebvre, N.; Chen, X.; Beuseroy, P.; Zhu, M. Traffic flow estimation using acoustic signal. *Eng. Appl. Artif. Intell.* **2017**, *64*, 164–171. [[CrossRef](#)]
8. Steele, C. A critical review of some traffic noise prediction models. *Appl. Acoust.* **2001**, *62*, 271–287. [[CrossRef](#)]
9. Mato-Méndez, F.J.; Sobreira-Seoane, M.A. Blind separation to improve classification of traffic noise. *Appl. Acoust.* **2011**, *72*, 590–598. [[CrossRef](#)]
10. Chakraborty, P.; Hegde, C.; Sharma, A. Data-driven parallelizable traffic incident detection using spatio-temporally denoised robust thresholds. *Transp. Res. Part C Emerg. Technol.* **2019**, *105*, 81–99. [[CrossRef](#)]
11. Kumar, P.; Nigam, S.P.; Kumar, N. Vehicular traffic noise modeling using artificial neural network approach. *Transp. Res. Part C Emerg. Technol.* **2014**, *40*, 111–122. [[CrossRef](#)]
12. Alexandre, E.; Cuadra, L.; Salcedo-Sanz, S.; Pastor-Sánchez, A.; Casanova-Mateo, C. Hybridizing Extreme Learning Machines and Genetic Algorithms to select acoustic features in vehicle classification applications. *Neurocomputing* **2015**, *152*, 59–68. [[CrossRef](#)]

13. Borkar, P.; Sarode, M.V.; Malik, L.G. Employing speeded scaled conjugate gradient algorithm for multiple contiguous feature vector frames: An approach for traffic density state estimation. *Procedia Comput. Sci.* **2016**, *78*, 740–747. [[CrossRef](#)]
14. Rahman, M.; Kang, M.W.; Biswas, P. Predicting time-varying, speed-varying dilemma zones using machine learning and continuous vehicle tracking. *IEEE Trans. Intell. Transp. Syst.* **2021**, *130*, 103310. [[CrossRef](#)]
15. Kim, K. Lightweight filter architecture for energy efficient mobile vehicle localization based on a distributed acoustic sensor network. *Sensors* **2013**, *13*, 11314–11335. [[CrossRef](#)]
16. Nooralahiyani, Y.; Kirby, H.R.; Mckeown, D. Vehicle classification by acoustic signature. *Math. Comput. Model.* **1998**, *27*, 205–214. [[CrossRef](#)]
17. Schlar, A.; Averbuch, A.; Rabin, N.; Zheludev, V. A diffusion framework for detection of moving vehicles. *Digit. Signal Processing* **2010**, *20*, 111–122. [[CrossRef](#)]
18. Yuan, Y.K.; Fan, L.; Wang, Z.S. A novel method for speed measurement based on vehicle noise. *J. Dalian Polytech. Univ.* **2014**, *33*, 300–304.
19. Ma, A.Y.; Wang, J.L. Statistical analysis of source noise and speed of different vehicles on Expressway. *Highway* **2005**, *7*, 199–202.
20. Dinesh, V.; Naveen, A. Smartphone based traffic state detection using acoustic analysis and crowdsourcing. *Appl. Acoust.* **2018**, *138*, 80–91.
21. Dennis, J.; Tran, H.D.; Chng, E.S. Overlapping sound event recognition using local spectrogram features and the generalised Hough transform. *Pattern Recognit. Lett.* **2013**, *34*, 1085–1093. [[CrossRef](#)]
22. Ma, Q.L.; Zou, Z. Traffic detection algorithm for identify overlapping noise. *Appl. Res. Comput.* **2020**, *37*, 1069–1072 + 1080.
23. Aumond, P.; Can, A.; Gozalo, G.R.; Fortin, N.; Suarez, E. Method for in situ acoustic calibration of smartphone-based sound measurement applications. *Appl. Acoust.* **2020**, *166*, 107337. [[CrossRef](#)]
24. Benocci, R.; Molteni, A.; Cambiaghi, M.; Angelini, F.; Roman, H.E.; Zambon, G. Reliability of Dynamap traffic noise prediction. *Appl. Acoust.* **2019**, *156*, 142–150. [[CrossRef](#)]
25. Yang, G.; Mo, H.; Li, W.; Lian, L.; Zheng, Z. The Endpoint detection of cough signal in continuous speech. *J. Biomed. Eng.* **2010**, *27*, 544–547, 555.
26. Borkar, P.; Malik, L.G. Acoustic Signal based Traffic Density State Estimation using SVM. *Int. J. Image Graph. Signal Processing* **2013**, *8*, 37–44. [[CrossRef](#)]
27. Ma, Q.L.; Zou, Z. Traffic State Evaluation Using Traffic Noise. *IEEE Access* **2020**, *8*, 120627–120646. [[CrossRef](#)]
28. Kaur, A.; Sood, N.; Aggarwal, N.; Vij, D.; Sachdeva, B. Traffic state detection using smartphone based acoustic sensing. *J. Intell. Fuzzy Syst.* **2017**, *32*, 3159–3166. [[CrossRef](#)]
29. Cai, P. A fast algorithm of speech end-point detection combined with short-time zero crossing rate. *J. Xiamen Univ. Technol.* **2013**, *21*, 48–51.
30. Jodra, J.L.; Gurrutxaga, I.; Muguerza, J.; Yera, A. Solving Poisson's equation using FFT in a GPU cluster. *J. Parallel Distrib. Comput.* **2017**, *102*, 28–36. [[CrossRef](#)]
31. Ma, Q.L.; Zou, Z.; Liu, F.J. Traffic statistics based on the endpoint detection of driving acoustic signals. *Sci. Technol. Eng.* **2020**, *20*, 1676–1683.
32. Zhang, H.; Pan, Z.M. Parameters selection of stationary wavelet denoising algorithm. *J. Natl. Univ. Def. Technol.* **2019**, *41*, 165–170.