Real-Time Recognition Algorithm of Vehicle External Specific Sounds Based on Spectrum Characteristics

Hongbin Lei and Botao Wang
Faculty of Information Technology, Beijing University of Technology, Beijing, China
Email: 839381149@qq.com; wangbt@bjut.edu.cn

Abstract. As the development of vehicle sound insulation technology, drivers ignore some external sound information regularly. It is particularly important for the recognition of external sounds when the vehicle is driving. The real-time detection of the fundamental frequency of the external sound is not accurate due to the complexity of the external environment and the noise interference. In this paper, a recognition algorithm of vehicle external specific sounds is presented. The spectrum information of the external sound is converted into an image form, and the analysis result is obtained after processing the spectrum image. We choose five different types of sounds and the total number of 1000 samples for testing. The experiment data shows that the overall accuracy of the algorithm's processing data can reach more than 90%, and the recognition rate can reach more than 90%. This algorithm has the characteristics of high accuracy and strong robustness, and it can easily be realized in real time.

1. Introduction
With the development of science and technology, people’s requirement of vehicles is gradually increasing, and the requirement of driving comfort are gradually increasing [1]. Vehicle design focuses on sound insulation. But the better the sound insulation effect, the less external sound can be heard from the inside, which can affect the driver’s decision. The external sound reflected in car horns, pets, and the siren sounds of some special vehicles [1]. Therefore, we should design an algorithm that can make accurate and real-time decisions and reflect it to the car itself. This algorithm is also an important function for intelligent assisted driving.

The decisive role in sound information is the fundamental frequency and the multi-frequency generated by the sound itself [2]. Therefore, this paper focus on the fundamental frequency information of the external sound during driving, realizes the requirement for recognition of different sounds and real-time recognition.

2. Requirement Analysis
We need to identify valid sound information that can make decisions during driving [3]. First of all, we need to filter the sound. The effective information are some sounds with high intensity that satisfy the recognition characteristics [4]. The fundamental frequency of the sound represents its own information during sound recognition. Therefore, it is very important to confirm the fundamental frequency information of sounds. In the frequency spectrum, the fundamental frequency of sounds generally has the largest frequency response at the same time and has multi-frequency information [5]. When confirming the fundamental frequency information, the multi-frequency information is supposed to be considered, and the relationship between the fundamental frequency and the multi-frequency is used to verify whether the analysis result is correct.
As shown in figure 1, the blue arrow is the fundamental frequency information and the green arrow is the double frequency [6]. It shows that the energy information of double frequency is much attenuated than the energy of the fundamental frequency.

2.1. Influencing Factors
In some cases, the sound is composed of composite frequencies rather than a single frequency, we should identify the information of the composite frequency and the harmonic frequency [7]. The following points need to be considered:

2.1.1. Existence the Dissonance. The n times harmonic frequency of an ideal sound should be n times the fundamental frequency. Due to the situation of adding dissonance, the relationship between fundamental frequency and harmonic frequency can be:

\[ f_n = n \times f_1 \sqrt{1 + (n^2 - 1)I} \]  

where \( f_1 \) is the fundamental frequency, \( f_n \) is the harmonic frequency, \( n \) is the harmonic order, and \( I \) is the coefficient of disharmony.

2.1.2. Frequency Overlap. Due to overlapping frequencies, and the multi-frequency energy value is less than the fundamental frequency, other multi-frequency doped in a fundamental frequency cannot be distinguished in the spectrum.

2.2. Specific Sound
Through the analysis of the external sounds during driving, we can use the following specific sounds for judgment and recognition [8]:

2.2.1. Pet Sound. Since it is difficult to observe low objects while driving, it is particularly important to distinguish pet sounds effectively.

\[ f(x) = a + bx + cx^2 + dx^3 + o(x^4) \]

where \( f(x) \) is the frequency value, \( x \) is time, \( a, b, c, d \) are constants. Since the fundamental frequency information is more obvious and representative, the multi-frequency information can be ignored [8].

![Figure 1. Partial spectrum image of guitar single string.](image1)

![Figure 2. Spectrum image of dog barking.](image2)

2.2.2. Car Whistle. During driving, the most frequent sound is the whistle of other cars. We select the whistle of a car and a truck as a reference for research.

Figures 3-4 show that whether the whistle of a truck or a car has distinctive characteristics and it can be fitted into a multivalued function or find the relationship between time and frequency:

\[ f_1(x_n) = N_1 a_1 \]

\[ N_1 \in [1,7], \ a_1 \in (300,350) \]
\[ f_2(x_n) = N_2 a_2 \quad N_2 \in [1,4], \quad a_2 \in (350,400) \cup (450,500) \]  

where \( f_1(x_n) \) is the frequency of the truck whistle, \( f_2(x_n) \) is the frequency of the car whistle, \( x_n \) is time, \( a_1, a_2 \) is the fundamental frequency, and \( N_1, N_2 \) are constants. It shows that the car frequency is within 1500 Hz, and the truck frequency is within 3000 Hz [9-10].

\[ f_1(x_n) \]

\[ f_2(x_n) \]

\[ f(x) = a \sin(bx + c) + d \]  

where \( f(x) \) is the fundamental frequency, \( x \) is time, \( a, b, c, d \) are constants.

2.2.3. The Sound of Emergency Vehicles. It is very important for the identification of emergency vehicles. During the driving, if you encounter an emergency vehicle passing by, you need to slow down or take measures to make way for it.

The frequency spectrum image of police car sounds shows a very regular sine wave shape in figure 5, which can be fitted into the form of a sine function:

\[ f(x) = Na \]  

\[ f(x+T) = Na \]  

where \( f(x_n) \) is frequency and \( T \) is period. \( a \) is the fundamental frequency, and \( N \) is a constant.

Figure 6 shows the spectrum image of the ambulance is in the form of a segmented horizontal line. And there is a clear correspondence between fundamental frequency and frequency multiplication. From this we can find the relationship between time and frequency:

\[ f(x) = \begin{cases} 
ax + b & x \in [0, \frac{T}{2}] \\
2a(T - x) + 2b & x \in \left[\frac{T}{2}, T\right]
\end{cases} \]  

where \( f(x) \) is frequency, \( a \) and \( b \) are constants and \( T \) is period. This spectrum has the characteristics of high odd frequency multiplication energy, and the highest frequency can reach more than 5000Hz.

Figure 5. Spectrum image of the siren of police car. Figure 6. Spectrum image of the siren of ambulance. Figure 7. Spectrum image of the siren of the fire truck.
3. Fundamental Frequency Detection Algorithm Based on Spectrum Characteristics

This paper designs a fundamental frequency detection algorithm based on spectrum characteristics of image morphology, and realizes the fundamental frequency detection and recognition of sound files by performing image morphology processing on the spectrogram. The main implementation process is divided into the following aspects:

3.1. Grey Scale Processing

In order to improve the analysis of the image, it is necessary to grey scale the spectrogram file. After grey scale processing, the frequency and time of sound data in the original spectrum image are unchanged, and the energy value is converted into grey scale value [11]. The grey scale processing of image is shown in figure 8. For the convenience of calculation, the original energy value of 0~1dB is extended to 0~255, and the zero point of the image is adjusted from the lower left corner of the image to the upper left corner.

![Figure 8. Grey scale processing of spectrum images (5-20 seconds, 0-500Hz).](image)

3.2. Morphological Closed Operation

Through the analysis of the spectrogram, the original continuous image area has break points due to noise interference and other factors. Morphological closed operation can filter the image by filling the concave angle of the image. It can fill the small holes between the images and fill in the small cracks, while ensuring that the overall position and shape remain unchanged. The morphological closed operation on the original image can effectively solve the phenomenon of “burr” in the subsequent process.

3.3. Binary Threshold Processing and Image Thinning

Binary threshold processing is performed on the spectrum image that has completed the morphological closed operation. The multi-frequency information is common in the image. By setting the threshold, it can effectively filter out high frequency and noise. After testing, the threshold is generally set between 150 and 170.

Using skeletonization to process binary images. Since the value of binary images are all 0 and 1, it is easy to find the front-end point information of the frequency by refine large-scale frequency information into one or more line segments.

3.4. Front-End Point Detection

By finding the starting point of the data through the judgment mechanism, the front-end point of each line segment can be obtained, and then it is judged whether it meets the standard of the fundamental frequency (frequency and time information), so the front-end point information corresponding to each fundamental frequency data point can be determined. The information at the front-end point can be used to distinguish different types of sounds and realize the discrimination of the detected sounds. The results of front-end point detection are shown in figure 9. The circle mark is the position of the front-end point, which contains the frequency, time, intensity and other information.
3.5. Real-Time Identification of Fundamental Frequency

In order to ensure the real-time of identification, the real-time recognition algorithm for fundamental frequency is divided into the following steps:

- Step 1: Read the file frame by frame and obtain the characteristic value of the front-end point in the audio file through the algorithm in Section 3.
- Step 2: Determine whether the feature value belongs to the fundamental frequency feature of several types of sound in the section 2. If not, go to step 1.
- Step 3: Determine whether the data after the front-end point conform the correspondence between the fundamental frequency of the sound and time. The duration is the period of the characteristic sound in step 2. If not, go to step 5.
- Step 4: Count the fundamental frequency and multi-frequency during this period of time to determine whether they conform the correspondence between the fundamental frequency and the multi-frequency of the sound.
- Step 5: Output the judgment result and repeat step 1.

This article identifies several specific sounds in section 2. We choose 200 samples for each sound, and the total number of samples is 1000 for testing. The results are shown in table 1.

Table 1. The result of different types of sound recognition.

| Types of sound | Number of samples identified | The correct number of samples identified | Accuracy % | Precision % | Recall % |
|----------------|------------------------------|----------------------------------------|-----------|-------------|---------|
| Dog            | 153                          | 130                                    | 90.7      | 84.97       | 65      |
| Car            | 186                          | 171                                    | 95.6      | 91.94       | 85.5    |
| Police car     | 187                          | 179                                    | 97.1      | 95.72       | 89.5    |
| Fire truck     | 162                          | 151                                    | 94        | 93.21       | 75.5    |
| Ambulance      | 175                          | 158                                    | 94.1      | 90.29       | 79      |

Table 1 shows that the overall prediction reaches more than 90%, the accuracy is more than 90%, and the recall is about 80%. The recognition effect of pet sounds is poor, and the recognition of various vehicle sounds is generally better, which has a certain relationship with the stability of the sound’s own frequency.

4. Conclusion

In this paper, we design a real-time recognition algorithm of vehicle external specific sounds based on spectrum characteristics. The accuracy and precision of this algorithm for specific sounds is above 90%, and it can satisfy the need of real-time recognition. This algorithm can be applied on different platforms and it has strong applicability. It can be a part of the vehicle’s intelligent assisted driving system, and it can make decisions by real-time recognition of the external sound during driving, which can effectively improve the safety and comfort during driving. Next, we will try to solve the accurate
rate of low voice recognition of pets in the article, and research the spectrum characteristics of more types of sound.

References
[1] Lu L, Jiang H and Zhang H J 2001 A robust audio classification and segmentation method Ninth ACM International Conference pp 203-211.
[2] Zhu Y W, Kankanhalli M and Gao S 2005 Music Key Detection for Musical Audio The 11th International Multimedia Modeling Conference pp 30-37.
[3] Bahja F, Martino J, Ibn-Elhaj E and Aboutajdine D 2016 A corroborative study on improving pitch determination by time–frequency cepstrum decomposition using wavelets SpringerPlus (Germany: Springer Science and Business Media Deutschland GmbH) pp 564-580.
[4] Ali S M and Karule P 2016 MFCC, LPCC, formants and pitch proven to be best features in diagnosis of speech disorder using neural networks and SVM International Journal of Applied Engineering Research 897-903.
[5] Ryu J and Song H 2019 Effect of building façade on indoor transportation noise annoyance in terms of frequency spectrum and expectation for sound insulation Applied Acoustics 152 21.
[6] Yang W J, Cai M and Luo P 2020 The calculation of road traffic noise spectrum based on the noise spectral characteristics of single vehicles Applied Acoustics 160 21.
[7] Sato K, Tomita Y, Kageyama R, Takane Y, Kayano S and Saito H 2019 Method to calculate frequency characteristics of reconstruction filter kernel in X-ray computed tomography Australasian Physical & Engineering Sciences in Medicine 43 69-78.
[8] Virtanen T, Plumbley M and Ellis D 2018 Introduction to sound scene and event analysis Computational Analysis of Sound Scenes and Events (Springer International Publishing) pp 3-12.
[9] Wieczorkowska A, Kubera E, Słowik T and Skrzypiec K 2017 Spectral features for audio based vehicle and engine classification Journal of Intelligent Information Systems 50 265-290.
[10] Fukui M, Watanabe T and Kanazawa M 2018 Sound source separation for plural passenger speech recognition in smart mobility system IEEE Transactions on Consumer Electronics 66 125-133.
[11] Chandrakala S and Jayalakshmi S 2019 Environmental audio scene and sound event recognition for autonomous surveillance: A survey and comparative studies ACM Computing Surveys 52 1-34.