IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 May 2023, accepted 12 June 2023, date of publication 15 June 2023, date of current version 7 July 2023.

*Digital Object Identifier 10.1109/ACCESS.2023.3286539*

APPLIED RESEARCH

# Loudspeaker Abnormal Sound Classification Using Auditory Perception Weighted by Energy Entropy

**HAITAO SU**[1,2,3], **JIALUN LI**[1,3], **JIEPENG GU**[1,3], **HONGZHI HU**[1,2,3], **AND CUIFENG XU**[1,2,3]
[1]School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China
[2]Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, Guilin University of Electronic Technology, Guilin 541004, China
[3]Nanning Research Institute, Guilin University of Electronic Technology, Nanning 530033, China

Corresponding author: Cuifeng Xu (xcf4100@guet.edu.cn)

This work was supported in part by the Guangxi Key Laboratory of Automatic Detecting Technology and Instruments under Grant YQ22105 and Grant YQ23106, and in part by the Graduate Education Innovation Program of Guilin University of Electronic Science and Technology under Grant 202210595048.

**ABSTRACT** In order to improve the classification accuracy of loudspeaker abnormal sounds, this paper proposes a method based on time-varying specific loudness weighted by energy entropy and principal component analysis. This method simulates human auditory perception mechanism to process loudspeaker sound response signal to obtain more effective features. The human hearing system can be divided into many parallel and functionally independent conduction pathways according to frequency. The energy of the acoustic signal in each pathway can be discerned. Therefore, the time-varying specific loudness is calculated firstly to build the quantitative correlation of loudspeaker's acoustic response signals and human hearing sensations. Each sub band loudness is weighted by energy entropy to highlight the acoustic strength variation with time and frequency. Then, important features are extracted by two dimensional-principal component analysis (2D-PCA). Finally, the whale optimization algorithm-least squares support vector machine (WOA-LSSVM) is adopted for classification. Visual analysis of the extracted features shows that this method can extract loudspeaker response signal features with better discriminability. Classification experimental results show that the average accuracy of this method reached 98.4%, which is higher than the classification method based on traditional time-frequency domain statistical features. The loudspeaker abnormal sound classification method in this paper simulates human auditory perception to extract features and is able to improve classification accuracy and automation effectively.

**INDEX TERMS** Loudspeaker abnormal sound classification, time-varying specific loudness, energy entropy, two dimensional-principal component analysis (2D-PCA), whale optimization algorithm-least squares support vector machine (WOA-LSSVM).

## I. INTRODUCTION

The loudspeaker, as an important electro-acoustic converter, is widely used in human-computer interaction technologies such as augmented reality (AR), virtual reality (VR), and mixed reality (MR). At the same time, the quality of loudspeakers has also received attention. However, due to mistakes in the production process, the loudspeaker will produce abnormal sound in the process of using. It is important to carry out quality inspection and fault classification in the production process to guarantee loudspeaker quality. Traditional loudspeaker abnormal sound classification methods rely on professional workers to listen to the sound on the production line. However, the results of workers' listening are affected by many subjective personal factors; it is difficult to further improve the accuracy and stability of loudspeaker detection. Compared with artificial inspection, reliable classification algorithms are a

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

better choice in industrial production to improve loudspeaker quality.

Currently, researchers in the industry often use loudspeakers' electrical response signals, vibration signals, and acoustic response signals for loudspeaker fault detection. Illana analyzed the current signal flowing through the loudspeaker by means of Zhao-Atlas-Marks distribution (ZAMD) and proposed a failure extractor based on relevant ZAMD frequency regions segmentation and Mahalanobis distance to detect the rub defect [1]. Serban et al. designed an experimental system used for fault detection of the direct radiator loudspeaker based on phase characterization of consumed current and the applied voltage [2]. The system is only used to distinguish between qualified and faulty products. Izzo et al. used radar micro Doppler approach to analyze the vibration signal of a loudspeaker in order to detect irregular defects affecting the motion of the voice coil [3]. Peulraja et al. proposed a method for detecting loudspeaker abnormal sound based on harmonic distortion, which uses an excitation signal from 20 Hz to 20 kHz to make a sound from a loudspeaker. The spectrum was obtained by performing a fast Fourier transform on the response signal, and then the top six frequency bands and their corresponding total energy were computed and chosen for further analysis. Finally, they were used as features for training the neural network for automatic detection of loudspeaker faults. However, the neural network was not able to give an efficiency of more than 90% [4]. Klippel and Werner proposed a method based on asynchronous demodulation and envelope averaging between the higher-order acoustic harmonics to improve the sensitivity of dust cap leaking fault detection [5]. Wenhua and Yuming converted the response acoustic signal of frequency sweep into two-dimension time-frequency image signal and designed a feature extraction method for images. The fault identification accuracy rate can reach 95% based on the image features [6]. However, the image processing method increases the feature extraction complexity of acoustic response signal.

Although the above methods can have a good effect on the recognition and classification of loudspeaker abnormal sound, they don't pay attention to the human auditory characteristics and only continue to follow the conventional signals processing method. They generally extract features from the time domain, frequency domain and time-frequency domain of the loudspeaker response signal. There are errors between test results and the results of manual listening.

Psychoacoustics describes the connection between sound stimulation and human perception. At present, psychoacoustics is widely used in the design and manufacturing process of products to improve the acoustic performance and competitive advantage of products [7], [8], [9], [10], [11]. Under the same excitation signal, the acoustic response signals of the qualified loudspeaker and the faulty loudspeaker have different energy values in different frequency bands. Human listeners are able to distinguish this difference based on previous experience. Temme proposed a simplified perceptual model based on the masking threshold of human auditory and a cepstrum loudness enhancement algorithm for the detection of Rub & Buzz distortion [12], [13]. Finally, additional psychoacoustic variables are added and an artificial neural network approach is used to measure harmonic distortion audibility [14]. Those methods are based on the Perceptual Evaluation of Audio Quality (PEAQ) standard. The PEAQ requires no time lag between the test signal and the reference signal, which is difficult to guarantee in loudspeaker field tests.

In this paper, we propose a method for loudspeaker abnormal sound classification based on auditory perception, drawing on the judgment of good listeners on loudspeaker abnormal sounds. To establish a quantitative correlation between loudspeaker acoustic response signals and human hearing sensations, we first calculate the time-varying specific loudness. Next, we use entropy to enhance the characteristics of the time-varying specific loudness. Among the different kinds of entropy, energy entropy, fractional order fuzzy dispersion entropy [15], and simplified coded dispersion entropy [16] are considered. Energy entropy is chosen because the energy of different frequency bands is one of the important factors for humans to identify the sound character. Then, we use two-dimensional principal component analysis (2D-PCA) to perform feature dimensionality reduction, which can extract the most important features and improve the classification processing speed. Finally, the whale optimization algorithm-least squares support vector machine (WOA-LSSVM) is adopted for classification. The proposed method can reach high accuracy and meet the requirements of loudspeakers abnormal sound classification online.

We analyze 200 loudspeakers; 50 of each kind (normal, slight sound, voice coil rubbing, and dust cap leaking). In total, 1000 response acoustic signals are obtained with 5 measurements per loudspeaker. Slight sound refers to the low loudness of the loudspeaker due to the high rated impedance; voice coil rubbing refers to the collision between the voice coil and the permanent magnet; dust cap leaking refers to the air leakage of the dust cap. Section II presents the proposed psychoacoustic approach to signal processing and describes the classification algorithm in detail. In Section III, the identification results are depicted. We discuss our findings in section IV and provide conclusions in section V.

## II. ABNORMAL SOUND CLASSIFICATION BASED ON ENTROPY-WEIGHTED TIME-VARYING SPECIFIC LOUDNESS AND 2D-PCA

The experimental platform shown in Figure 1 is set up to acquire loudspeaker's acoustic response signal. The platform mainly includes a computer, a power amplifier module, a B&K 4096 microphone, and the B&K input module (Type 3057-B-030). The excitation signal generated by the computer will be amplified by the power amplifier module to drive the loudspeaker. At the same time, the acoustic response signal will be collected by the microphone and acquired by the B&K input module.

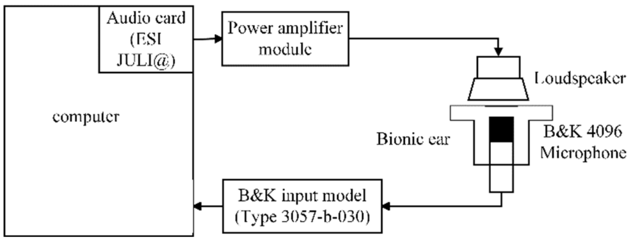



**FIGURE 1.** Test system composition diagram.

The loudspeaker under test is used in the headphone, and its diameter is 40 mm, the rated power is 0.5 W, the impedance is 8 Ω. In the experiment, the continuous logarithmic sweep signal is used for the excitation signal, its duration is 1s, the frequency range is 20 Hz∼20 kHz and the root mean square value is 0.9 V. The acquired acoustic response signal is saved in computer for the subsequent analysis.

To implement abnormal sound classification, we first extract features of the acoustic response based on auditory perception. The time-varying specific loudness is calculated to describe the acoustic strength variation with the time and frequency. The time-varying specific loudness result is weighted by energy entropy to enhance this variation. Then 2D-PCA is used to remove redundant features and enhance the effect of classification. Finally, the WOA-LSSVM model is employing as a classifier. The proposed abnormal sound classification using auditory perception weighted by energy entropy is shown in Figure 2.

### A. TIME-VARYING SPECIFIC LOUDNESS MODEL WEIGHTED BY ENERGY ENTROPY

Loudness is a psychoacoustic parameter that reflects the human subjective perception of sound intensity and is related to both the frequency and amplitude of the sound signal. The computational procedure for calculating loudness considers both auditory frequency decomposition and auditory masking effects. The loudness is assigned to 24 critical bands and specific loudness is calculated for each band. In this paper, we use time-varying specific loudness, which represents changes in specific loudness over time, to represent differences in the loudspeaker's acoustic response signal in different states.

To get a better human hearing system's sensation, the method of time-varying specific loudness weighted by energy entropy is proposed in the paper. Firstly, the time-varying specific loudness of the loudspeaker's acoustic response signal is calculated by the DIN 45631/A1 standard. Then the weight coefficients are obtained through the energy entropy of each critical band. Finally, multiply the weight coefficient and corresponding critical band's specific loudness together and the weighted time-varying specific loudness is gotten. Thus, the quantitative correlation of loudspeaker's acoustic response signals and human hearing sensations is built by the weighted time-varying specific loudness in this paper. And the detailed steps are as follows in Figure 3:

1) Calculate the time-varying specific loudness $v(n, z_i)$ of the acquired acoustic response signal by the standard DIN 45631/A1, $n$ is the number of points obtained by 2ms step for the duration of 1s, $n = 0, 1, 2, \cdots, 500$; $z_i$ is the frequency scale of the critical bands; the 24 critical bands are divided into 240 sub-bands by 0.1 step, $i = 1, 2, \cdots, 240$.

2) Calculate each sub-band energy $E(z_i)$ of $v(n, z_i)$ according to the following formula:

$$E(z_i) = \sum_{n=0}^{\left\lfloor \frac{T}{0.002} \right\rfloor} |v(n, z_i)|^2 \tag{1}$$

3) Calculate the energy entropy weighting factor $g(z_i)$ for each sub-band as follows.

$$P(z_i) = \frac{E(z_i)}{\sum_{i=1}^{240} E(z_i)} \tag{2}$$

$$g(z_i) = -P(z_i)logP(z_i) \tag{3}$$

4) The $s(n, z_i)$ is obtained by multiplying each sub-band specific loudness with its corresponding weighting factor, as shown in the following equation.

$$s(n, z_i) = v(n, z_i)g^T(z_i) \tag{4}$$

By weighting the time-varying specific loudness, it can make its features more obvious. However, redundant information still exists. This redundant information may cause the classifier to become more complex. Principal component analysis (PCA) can be considered to remove the redundant information. Traditional principal component analysis suffers from the problems of complex covariance matrix calculation and time-consuming determination of the corresponding feature vectors, and 2D-PCA methods are proposed to overcome these problems [17]. In this paper, the 2D-PCA method is used to extract the main information of the weighted time-varying specific loudness and obtain the reduced-dimensional feature matrix of the loudspeaker acoustic response signal. Four kinds of the loudspeakers were tested, and the corresponding feature set $Y_i$ was obtained for each acoustic response signal by the above mentioned processing.

### B. CLASSIFICATION METHODS

The support vector machine (SVM) is a supervised learning algorithm that can be used for classification, clustering, prediction, and regression analysis. SVM model has gained popularity due to several attractive features such as stability, robustness [18]. Least squares support vector machine (LSSVM) is selected as the classifier in this paper. LSSVM is an improvement of SVM, it not only reduces the classification complexity but also improves the operation speed and convergence accuracy of the classification process, at the same time it has good noise immunity [18]. In order to improve the classification performance of loudspeaker abnormal sound, this paper uses whale optimization algorithm (WOA) to optimize the parameters of LSSVM, where the number of search agents is set to 20, and the maximum number of iterations is set to 50. In this paper, accuracy is used as fitness function, and

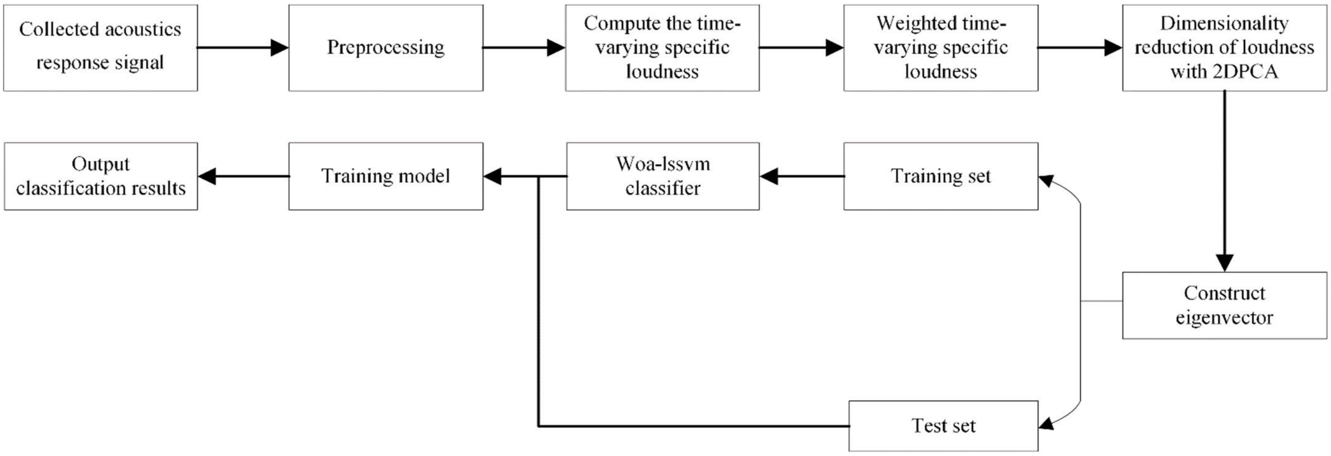



**FIGURE 2.** A psychoacoustical based system for classifying loudspeaker abnormal sounds.

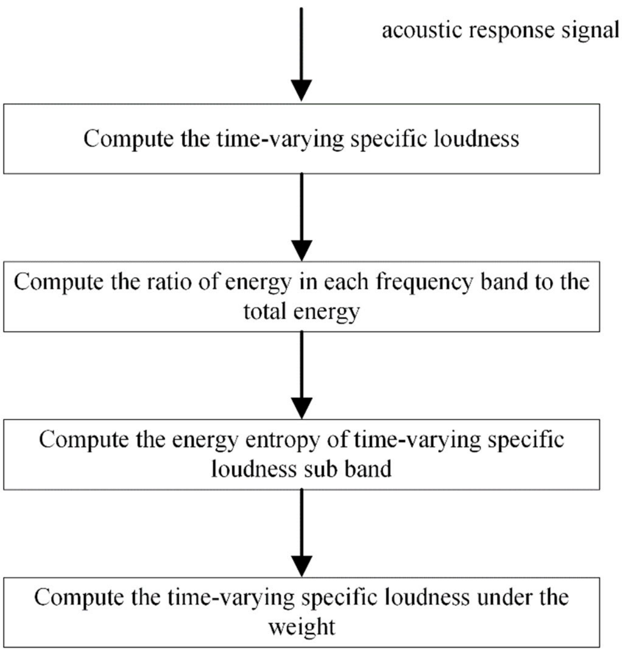



**FIGURE 3.** Algorithm diagram.

the fitness curve of a training is randomly selected, as shown in Figure 4. It can be seen from Figure 4 that the parameter settings have met the optimization requirements.

A detailed description of the WOA algorithm for LSSVM optimization is given in the literature [19]. The recognition accuracy of WOA-LSSVM is:

$$Acc = \frac{(N_{\text{Pass}})}{(N_{\text{Total test samples}})} \tag{5}$$

where $Acc$ is the acoustic signal recognition accuracy for the selected category, $N_{\text{Pass}}$ is the number of correct sound signal files tested and $N_{\text{Total test samples}}$ is the number of all sound signal files tested for the selected category.

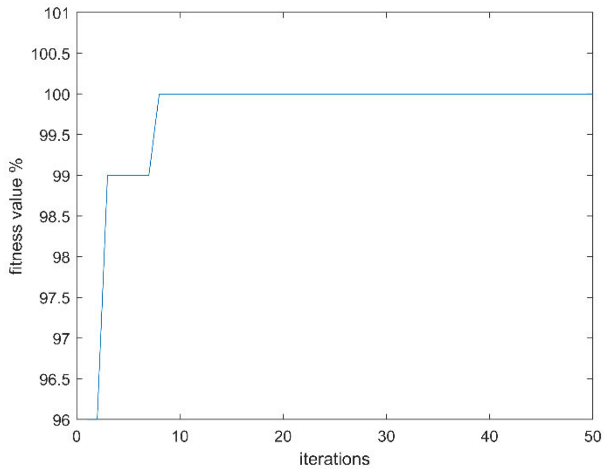



**FIGURE 4.** Fitness function curve.

The overall recognition accuracy is defined as follows

$$TAcc = \frac{Acc_a + Acc_b + Acc_c + Acc_d}{4} \tag{6}$$

where $TAcc$ is the total loudspeaker recognition accuracy, $Acc_a$ is the accuracy of recognizing normal loudspeakers, $Acc_b$ is the accuracy of recognizing voice coil rubbing loudspeakers, $Acc_c$ is the accuracy of recognizing dust cap leaking loudspeakers and $Acc_d$ is the accuracy of recognizing slight sounding loudspeakers.

Macro-F1 is also used to evaluate classification performance in multi classification. The calculation formula for macro-F1 in this paper is as follows

$$Macro\text{-}F1 = \frac{F1\text{-}score_a + F1\text{-}score_b + F1\text{-}score_c + F1\text{-}score_d}{4} \tag{7}$$

where $F1\text{-}score_a$ is the $F1\text{-}score$ of normal loudspeakers, $F1\text{-}score_b$ is the $F1\text{-}score$ of voice coil rubbing loudspeakers,

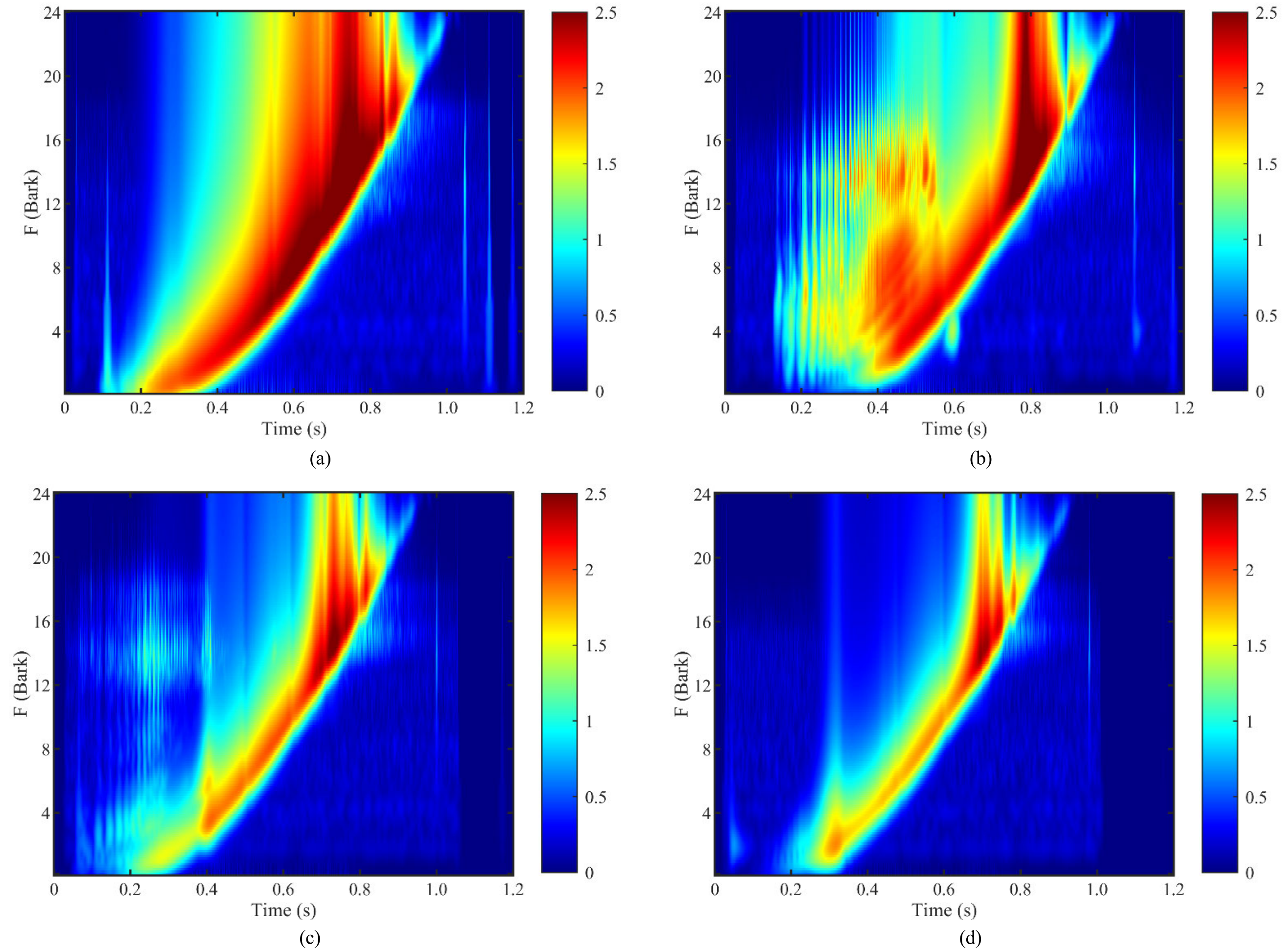



**FIGURE 5.** Time-varying specific loudness of a loudspeaker in different states (a) normal (b) voice coil rubbing (c) dust cap leaking (d) slight sound.

$F1\text{-}score_c$ is the $F1\text{-}score$ of dust cap leaking loudspeakers and $F1\text{-}score_d$ is the $F1\text{-}score$ of slight sound loudspeakers.

The calculation formula for F1-score is as follows

$$F1\text{-}score = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{8}$$

TP refers to True Positive, which means that a positive class is determined as a positive class; FP refers to False Positive, which means that negative classes are judged as positive classes; FN refers to False Negative, which means that a positive class is judged as a negative class.

## III. RESULTS ANALYSIS

### A. THE RESULTS OF TIME-VARYING SPECIFIC LOUDNESS BASED ON ENTROPY WEIGHTING

After pre-processing the loudspeaker response signal data collected from 1000 sets of valid data, the proposed method was used to analyze the loudspeaker response signals. Firstly, the time-varying specific loudness information of every loudspeaker response signal was obtained. Figure 5 depicts the time-varying specific loudness results of the loudspeaker acoustic response signal in four states. In order to collect the complete response signal, we started the collection 0.1s before playing the excitation signal and waited 0.1s after playing to finish the collection. As a result, the acoustic response signal is within the time range of 0.1s to 1s. Figure 5(a) shows the results of a normal loudspeaker and it can be seen that the energy of each critical band changes approximately logarithmically over time. The corresponding energy of the low and high critical band is smaller than the energy of the mid critical band part, which is consistent with the auditory response characteristics to different frequencies. Figure 5(b) shows the time-varying specific loudness of the voice coil rubbing state response signal. Compared with Figure 5(a), there are two differences: the energy from 4 to 12 critical band is smaller than that in Figure 5(a), and many harmonic components are generated between 0.2s and 0.6s. This is due to the collision between the voice coil and the permanent magnet during the loudspeaker's sound production. Figure 5(c) shows the time-varying specific loudness of the dust cap leaking state response signal. It can be observed that the energy of the signal on all bands is

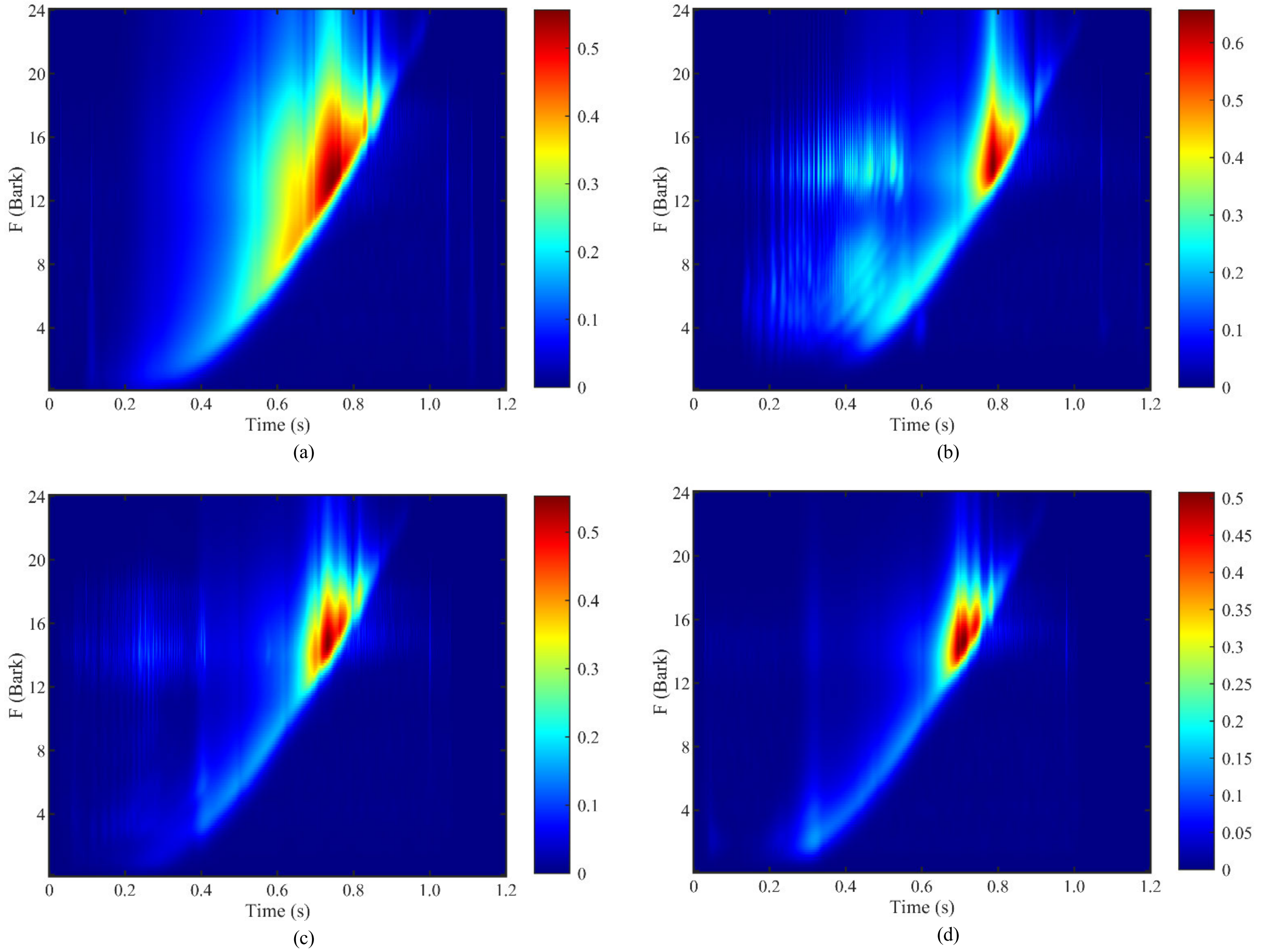



**FIGURE 6.** **Weighted time-varying specific loudness of loudspeakers in different states (a) normal (b) voice coil rubbing (c) dust cap leaking (d) slight sound.**

smaller than that in Figure 5(a), and there are many anomalies between 0.2s and 0.4s, which may be due to the decrease in electroacoustic conversion performance caused by dust cap leaking. Figure 5 (d) shows the time-varying specific loudness of the loudspeaker response in slight sound state. It can be seen that the signal energy on all bands is smaller than normal, and there are no harmonic components or other abnormalities compared to Figure 5(b) and Figure 5(c). Through the comparison in Figure 5, it can be found that there are significant differences in the time-varying specific loudness of the four situations. It proves that auditory perception is very effective to analyze the loudspeaker acoustic response signal.

In order to increase the difference of the time-varying specific loudness results in different states, the time-varying specific loudness is weighted by the energy entropy, results are shown in Figure 6. It is found that in the time-varying specific loudness graph under the energy entropy weighting, the different frequency bands of the loudspeaker in different states are enhanced, the time-varying specific loudness of the loudspeaker response signal in different faulty states has obvious difference characteristics.

However, the dimensionality of the time-varying specific loudness spectrum after the energy entropy weighting is too high. If the classifier's input dimensionality is too high, the classification efficiency will be affected and the classification time can increase significantly. Therefore, the main information quantity of the time-varying specific loudness spectrum is extracted by 2D-PCA method, and the redundant information is eliminated to obtain the reduced dimensional feature matrix as the classifier's input.

### *B. RESULTS BASED ON ENTROPY-WEIGHTED TIME-VARYING SPECIFIC LOUDNESS COMBINED WITH 2D-PCA*

In order to more vividly illustrate the feature extraction effect of the algorithm, the t-distribution stochastic neighbor embedding (t-sne) algorithm is used to visually analyze the original acoustic response signal data and the feature data extracted in this paper, and the results are shown in Figure 7.

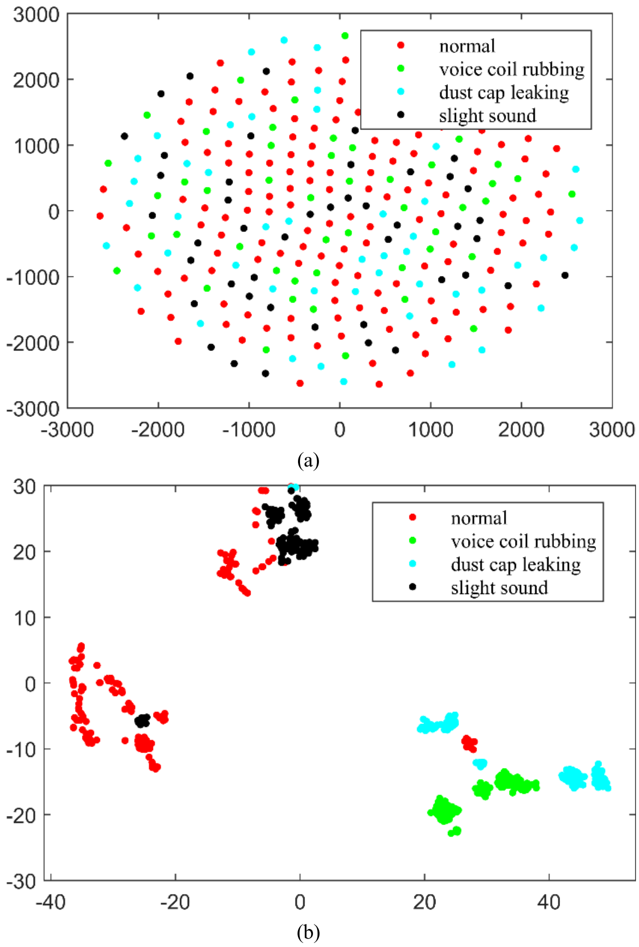



**FIGURE 7. (a) Raw data. (b) Two-dimensional visualization after the output of the proposed method.**

The t-sne plots show the trends and patterns in the data. However, it cannot be used to directly find outliers in the dataset as the data alignment does not directly represent the distance between data clusters. The analysis of Figure 7 shows that most of the fault types in the original acoustic response signal data overlap and it's difficult to distinguish from each other. In contrast, the feature data extracted in this paper can distinguish well between the abnormal sound types. In particular, the distribution of features between normal, voice coil rubbing, and slight sound are almost completely separated and they clustered in the corresponding area.

### C. CLASSIFICATION RESULTS

In the loudspeaker abnormal sound classification experiments, this paper refers to the 5-fold hierarchical cross-validation method for the selection of training and test sets. The method randomly divides the feature set of the loudspeaker response signal into 5 parts according to the proportion of the original data sample, with each part containing the same proportion of each type of data as in the original sample. One part is then used for testing while the remaining four parts are used for training. The average recognition rate from 5 experiments is used as the experimental result. This avoids the situation that random division may produce all one category in one copy, and reduces the error caused by selecting different samples in the test set and training set, and also reduces the influence of the randomness of the WOA-LSSVM model in the process of population initialization and optimization on the test results.

In order to better verify the accuracy of the algorithm in classifying loudspeaker anomalies, the statistical features of the loudspeaker response signal in the time-frequency domain were extracted from the above-mentioned data sets respectively, and then the data sets were normalized and input to the WOA-LSSVM model for loudspeaker anomaly classification discrimination.

The time and frequency domain characteristic parameters include mean ($\overline{x}$), standard deviation ($\sigma$), root mean square ($x_{RMS}$), skewness ($x_{skew}$), kurtosis ($x_{peak}$), mean square frequency, frequency center, frequency variance, etc. are commonly used. In the text, the time domain and frequency domain characteristic parameters are used in combination, and their calculation formulae are shown in Table 1.

**TABLE 1. Partial time domain frequency domain statistical characteristics parameters.**

| Characteristic parameters | Calculation formula |
|---|---|
| mean | $\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ |
| standard deviation | $\sigma=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}}$ |
| root mean square | $x_{RMS}=\sqrt{\frac{\sum_{i=1}^{n}(x_i)^2}{n}}$ |
| skewness | $x_{skew}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^3}{(n-1)std^3}$ |
| kurtosis | $x_{peak}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^4}{(n-1)rms^4}$ |
| mean square frequency | $MSF=\frac{\int_0^{+\infty}f^2P(f)df}{\int_0^{+\infty}P(f)df}$ |
| frequency center | $FC=\frac{\int_0^{+\infty}f\,P(f)df}{\int_0^{+\infty}P(f)df}$ |
| frequency variance | $VF=\frac{\int_0^{+\infty}(f-FC)^2P(f)df}{\int_0^{+\infty}P(f)df}$ |

In this paper, a total of 1000 loudspeaker response signal samples were used as the dataset for the experiments, and the number of each kind (normal, slight sound, voice coil rubbing, and dust cap leaking) is 250. Afterward, the proposed method and the traditional time-frequency domain statistical feature method were used to extract features from the response signals, and then the WOA-LSSVM was used to classify them, and the classification results of the training and test sets were calculated, as shown in Table 2. It can be concluded that the recognition results of the method proposed in this paper are superior to time-frequency domain statistical

**TABLE 2.** Recognition rate of loudspeaker abnormal sound classification by different methods.

| Methods | Recognition rate (%) | |
|---|---|---|
| | Training set | Test set |
| Statistical characteristics in the time-frequency domain | 91.4 | 88.5 |
| Methodology of this paper | 99.8 | 98.4 |

features. Under the test set, the accuracy rate of our method of the method proposed in this paper is 9.9% higher.

The hardware environment of this paper is inter(R) Core(TM) i7-10870H CPU @2.20GHz 2.21GHz and 16GB RAM, the software environment is Windows 10, and the software version is MATLAB R2020b. The total running time of the method in this paper is 133.72s, while the total running time of traditional time-frequency domain methods is 60.94s. For 1000 samples, although this method consumes more time than traditional time-frequency domain methods, the success rate is also 9.9% higher, which is acceptable in industrial production.

Tables 3 and 4 show the detailed classification results of traditional time-frequency domain feature extraction methods and proposed methods. In the tables, the row represents real values and the column represents predicted values.

**TABLE 3.** Abnormal sound classification results based on traditional time-frequency domain statistical features.

| | Normal | Slight sound | Voice coil rubbing | dust cap leaking |
|---|---|---|---|---|
| Normal | 89.07% | 4.93% | 0.13% | 5.87% |
| Slight sound | 15.60% | 84.40% | 0 | 0 |
| Voice coil rubbing | 2.00% | 0 | 98.00% | 0.00% |
| dust cap leaking | 16.80% | 0.80% | 0 | 82.40% |

**TABLE 4.** Entropy weighted time-varying specific loudness combined with 2DPCA for abnormal sound classification results.

| | Normal | Slight sound | Voice coil rubbing | dust cap leaking |
|---|---|---|---|---|
| Normal | 98.67% | 1.33% | 0 | 0 |
| Slight sound | 1.60% | 98.40% | 0 | 0 |
| Voice coil rubbing | 0.40% | 0 | 99.20% | 0.40% |
| dust cap leaking | 0 | 2.00% | 0.80% | 97.20% |

TABLE 5 shows the F1 score of four states, from which the macro-F1 can be calculated. By comparing Table 3 and Table 4, it can be seen that the method proposed in this paper has higher recognition rates for various types. At the same time, it can be calculated that the macro-F1 value of the method proposed in this paper is 98.37%, which is higher than 88.74% based on traditional time-frequency domain statistical. This fully demonstrates the advantages of features extracted based on psychoacoustic models in the classification of loudspeaker abnormal sounds.

**TABLE 5.** F1-score of loudspeaker abnormal sound classification by different methods.

| | Normal | Slight sound | Voice coil rubbing | dust cap leaking | Macro-F1 |
|---|---|---|---|---|---|
| Statistical characteristics in the time-frequency domain | 79.71% | 88.78% | 98.92% | 87.53% | 88.74% |
| Methodology of this paper | 98.34% | 97.56% | 99.20% | 98.38% | 98.37% |

C. Zhang et al. has compared many classifiers such as relevance vector machines (RVMs), SVM, etc. in experiments, and also compared some optimization algorithms such as particle swarm optimization (PSO), genetic algorithm (GA), etc. [20], [21], [22]. In order to verify the reliability of WOA-LSSVM, this paper also carried out comparative experiments. In addition, we tried the multiclass multi-kernel relevance vector machines (mRVMs) classifier, PSO and GA. The mRVMs classifier uses the WOA algorithm to optimize the parameters of the RBF kernel function, and their results are shown in the following table. As can be seen from the table, the accuracy of WOA-LSSVM in our experiment was 1.7%, 1.6% and 1.2% higher than that of WOA-mRVMs, PSO-LSSVM and GA-LSSVM, respectively.

**TABLE 6.** Accuracy of several methods.

| Methods | Recognition rate (%) |
|---|---|
| WOA-LSSVM | 98.4 |
| PSO-LSSVM | 96.8 |
| GA-LSSVM | 97.2 |
| WOA-mRVMs | 96.7 |

## IV. DISCUSSION

Loudspeaker abnormal sound classification is an essential task in the production of loudspeakers. The conventional classification methods are mostly based on time-frequency domain feature extraction, without paying attention to the human auditory characteristics. This paper calculates the time-varying specific loudness of loudspeaker abnormal sound and extracts features, which is closer to human perception and judgment mechanism. However, this method has a relatively low recognition rate for loudspeakers in the state of dust cap leaking. This is because the time-varying specific loudness of the response signal in the high frequency part decreases, which is similar to the response of the loudspeaker in the slight sound. The auditory perception model may not be sensitive to these small differences, which may be due to human auditory masking effects. This also requires to improve auditory perception model to better distinguish these differences. Also, LSSVM model used is very mature, and more novel classification models will be designed and observed to achieve more accuracy in the future. But the other effects brought by these models should also be evaluated in more detail.

As the acquisition of loudspeaker acoustic response signals places high demands on the acquisition environment, this study was conducted in a quiet experimental environment. Therefore, when testing in a new environment, it is important to ensure that there is no substantial noise interference before experimenting. Afterwards, we will attempt to collect data from the factory to make it closer to actual usage.

In addition, the loudspeaker anomalous sound classification method proposed in this paper classifies a single fault in a loudspeaker; we do not consider multiple faults in a loudspeaker (e.g., having both a voice coil rubbing ring fault and a slight sound fault), and if multiple faults in a loudspeaker need to be detected, more dimensions of the signal need to be acquired. For example, we can combine the electrical response signal and the acoustic response signal of the loudspeaker for fault detection.

## V. CONCLUSION

This paper presents a psychoacoustics-based method for loudspeaker abnormal sound classification. The main conclusions are as follows.

1. This paper extracts psychoacoustics features which are different from traditional time-frequency domain features for loudspeaker abnormal sound classification. The weighted time-varying specific loudness is proposed to build the quantitative correlation of loudspeaker's acoustic response signals and human hearing sensations. The 2D-PCA is used to extract the main information of the weighted time-varying specific loudness and obtain the reduced-dimensional feature matrix of the loudspeaker acoustic response signal.

2. The experiment result shows that, compared with the time-frequency domain statistical features, the weighted time-varying specific loudness spectrum can more accurately characterize features and has higher average recognition rate. The proposed method in this paper simulates human auditory perception to extract features, and is able to improve the classification accuracy effectively. It has great theoretical significance and practical value for Industrial production of loudspeakers.

## REFERENCES

[1] G. R. Illana. (2011). *Loudspeaker Rub Fault Detection by Means of a New Nonstationary Procedure Test in Audio Engineering Society*. Accessed: Oct. 9, 2022. [Online]. Available: http://www.researchgate.net/publication/286129204_Loudspeaker_rub_fault_detection_by_means_of_a_new_nonstationary_procedure_test

[2] G. Serban, G. V. Iana, and P. Anghelescu, "Fault detection of direct radiator loudspeaker systems by phase characterisation," in *Proc. 8th Int. Conf. Electron., Comput. Artif. Intell. (ECAI)*, Mar. 2023, pp. 1–4. [Online]. Available: http://ieeexplore.ieee.org/document/7861082

[3] A. Izzo, L. Ausiello, C. Clemente, and J. J. Soraghan, "Loudspeaker analysis: A radar based approach," *IEEE Sensors J.*, vol. 20, no. 3, pp. 1223–1237, Feb. 2020, doi: 10.1109/JSEN.2019.2946987.

[4] M. P. Paulraj, S. Yaacob, and M. R. Saad, "Loudspeaker fault detection using artificial neural network," in *Proc. 5th Int. Colloq. Signal Process. Appl.*, New York, NY, USA, Mar. 2009, pp. 362–366, doi: 10.1109/CSPA.2009.5069251.

[5] W. Klippel and R. Werner, "Measurement of turbulent air noise distortion in loudspeaker systems," *J. Audio Eng. Soc.*, vol. 59, nos. 7–8, pp. 475–486, Aug. 2011.

[6] G. Wenhua and Q. Yuming, "A method of loudspeaker's pure tone fault detection based on time-frequency image fractal," *Recent Patents Signal Process.*, vol. 4, no. 2, pp. 116–123, Sep. 2015, doi: 10.2174/221068630402150922131806.

[7] G. Pietila and T. C. Lim, "Intelligent systems approaches to product sound quality evaluations—A review," *Appl. Acoust.*, vol. 73, no. 10, pp. 987–1002, Oct. 2012, doi: 10.1016/j.apacoust.2012.04.012.

[8] M. Schneider and C. Feldmann. *Psychoacoustic Evaluation of Fan Noise*. Accessed: Mar. 7, 2023. [Online]. Available: http://www.mb.unisiegen.de/iftsm/forschung/veroeffentlichungen_pdf/150_2015.pdf

[9] D. Guoqing, "An improved psychoacoustic annoyance model based on tonal noises," in *Proc. Inter Noise Con Congr. Conf.*, vol. 253, no. 8, 2016, Accessed: Mar. 13, 2023. [Online]. Available: http://schlr.cnki.net/new/Detail/index/GARJ2016/SFJGE8B34DE08E46F4450567738705C344DF

[10] Y. Jing, H. Su, S. Wang, W. Gui, and Q. Guo, "Fault diagnosis of electric impact drills based on time-varying loudness and logistic regression," *Shock Vib.*, vol. 2021, Mar. 2021, Art. no. 6655090, doi: 10.1155/2021/6655090.

[11] T. Novaković, M. Ogris, and J. Prezelj, "Validating impeller geometry optimization for sound quality based on psychoacoustics metrics," *Appl. Acoust.*, vol. 157, Jan. 2020, Art. no. 107013, doi: 10.1016/j.apacoust.2019.107013.

[12] S. Temme, P. Brunet, and D. Keele, "Practical measurement of loudspeaker distortion using a simplified auditory perceptual model," in *Proc. 127th Audio Eng. Soc. Conv.*, vol. 2, Jan. 2009.

[13] P. Brunet, B. Fallon, and S. Temme, "Practical implementation of perceptual rub & buzz distortion and experimental results," presented at the Audio Eng. Soc. Conv., Nov. 2010. Accessed: Oct. 10, 2022. [Online]. Available:https://www.aes.org/e-lib/browse.cfm?elib=15596

[14] S. Temme, P. Brunet, and P. Qarabaqi, "Measurement of harmonic distortion audibility using a simplified psychoacoustic model—Updated," presented at the Audio Eng. Soc. Conf., 51st Int. Conf., Loudspeakers Headphones, Aug. 2013. Accessed: Oct. 10, 2022. [Online]. Available: https://www.aes.org/e-lib/browse.cfm?elib=16885

[15] Y. Li, B. Tang, B. Geng, and S. Jiao, "Fractional order fuzzy dispersion entropy and its application in bearing fault diagnosis," *Fractal Fractional*, vol. 6, no. 10, p. 544, Sep. 2022, doi: 10.3390/fractalfract6100544.

[16] Y. Li, B. Geng, and B. Tang, "Simplified coded dispersion entropy: A nonlinear metric for signal analysis," *Nonlinear Dyn.*, vol. 111, no. 10, pp. 9327–9344, May 2023, doi: 10.1007/s11071-023-08339-4.

[17] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004, doi: 10.1109/TPAMI.2004.1261097.

[18] H. Hamedi, M. Ehteshami, S. A. Mirbagheri, and S. Zendehboudi, "New deterministic tools to systematically investigate fouling occurrence in membrane bioreactors," *Chem. Eng. Res. Design*, vol. 144, pp. 334–353, Apr. 2019, doi: 10.1016/j.cherd.2019.02.003.

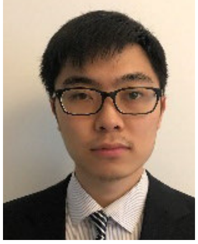
[19] F. Zhao and W. Li, "A combined model based on feature selection and WOA for PM2.5 concentration forecasting," *Atmosphere*, vol. 10, no. 4, p. 223, Apr. 2019, doi: 10.3390/atmos10040223.

[20] C. Zhang, S. Zhao, Z. Yang, and Y. Chen, "A reliable data-driven state-of-health estimation model for lithium-ion batteries in electric vehicles," *Frontiers Energy Res.*, vol. 10, Sep. 2022, Art. no. 1013800, doi: 10.3389/fenrg.2022.1013800.

[21] C. Zhang, Y. He, S. Jiang, T. Wang, L. Yuan, and B. Li, "Transformer fault diagnosis method based on self-powered RFID sensor tag, DBN, and MKSVM," *IEEE Sensors J.*, vol. 19, no. 18, pp. 8202–8214, Sep. 2019, doi: 10.1109/JSEN.2019.2919868.

[22] C. Zhang, Y. He, B. Du, L. Yuan, B. Li, and S. Jiang, "Transformer fault diagnosis method using IoT based monitoring system and ensemble machine learning," *Future Gener. Comput. Syst.*, vol. 108, pp. 533–545, Jul. 2020, doi: 10.1016/j.future.2020.03.008.

**HAITAO SU** received the Ph.D. degree from Xiamen University, in 2015. He is currently an Associate Professor with the Guilin University of Electronic Technology. His research interests include acoustic signal processing and embedded system design.



**JIALUN LI** received the B.S. degree in engineering and the B.S. degree in management from the Beijing University of Technology, in 2020. He is currently pursuing the M.S. degree in engineering with the Guilin University of Electronic Technology. His research interests include machine learning and sound enhancement.

**JIEPENG GU** received the B.S. degree in engineering from the Guilin University of Electronic Technology, in 2021, where he is currently pursuing the M.S. degree in engineering. His research interests include deep learning and auditory perception.

**HONGZHI HU** received the Ph.D. degree from the University of Electronic Science and Technology of China, in 2015. He is currently an Associate Professor with the Guilin University of Electronic Technology. His research interests include fault diagnosis, weak signal processing, and embedded system design.

**CUIFENG XU** received the master's degree from the Guilin University of Electronic Technology, in 2010. She is currently a Senior Laboratory Master with the Guilin University of Electronic Technology. Her research interests include digital signal processing and embedded system design.