Research on Power Equipment Fault Diagnosis Technology Based on Acoustic Signal

Hongsen Zou\textsuperscript{1,*}, Haoyang Chen\textsuperscript{1}, Zhiyuan Liu\textsuperscript{2} and Guohua Wang\textsuperscript{3}

\textsuperscript{1}State Grid Ningxia Maintenance Company, Ningxia, China
\textsuperscript{2}State Grid Ningxia Electric Power Company, Ningxia, China
\textsuperscript{3}School of Telecommunications Engineering, Xidian University, Xi’an, China

*Corresponding author: zouhongsen@jx.nx.sgcc.com.cn

Abstract. In recent years, with the continuous growth of China's electricity load, the power industry has developed rapidly. Power transformer is the most important and expensive in transmission and distribution system of large-scale power equipment, which undertakes the important task of power transmission. With the development of power system towards ultra-high voltage, large power grid and intelligence, it is particularly important to improve the safe operation level of transformers. Once the power transformer accident occurs, the repair time is longer and the influence is more serious. To solve this problem, a new power equipment fault diagnosis technology based on acoustic signals is studied in this paper, which is used to accurately diagnose and analyse the running state of the transformer. The simulation results show that the fault diagnosis based on acoustic signal is more accurate and can effectively diagnose the fault of power equipment.

1. Introduction
In recent years, with the continuous growth of China's electricity load, the power industry has developed rapidly. As of September 2016, a total of 21 UHV projects have been put into operation or started construction in China, including 7 UHV AC projects and 14 UHV DC projects. Power transformers are the most important and most expensive in transmission and distribution systems of large-scale power equipment to undertake power transmission's critical task. With the development of power systems towards ultra-high voltage, large power grid, and intelligence, it is essential to improve transformers' safe operation level. Once the power transformer accident occurs, the repair time is long, and the influence is serious. In general, the larger the capacity, the higher the voltage level, the greater the loss caused by transformer failure.

For a long time, people have mostly classified the transformer's sound like noise and ignored its research value. However, only experienced master workers who have been engaged in substation operation for many years can judge the transformer's running state according to the sound emitted by the transformer with the human ear close to the box when doing the transformer inspection. This method, which mainly relies on human ear auscultation, is very restrictive. Generally, it can only be distinguished when a fault occurs, and the fault is relatively apparent. The auscultation method is mainly suitable for workers with rich working experience and a good understanding of transformer structure. Due to large human factors, it is not widely applicable to the application of online
monitoring technology [1-3]. Besides, in high pressure, ultra-high pressure and thunderstorm, and other harsh weather conditions, the manual inspection operation method has a greater safety risk. With the continuous development of transformer acoustic diagnosis technology, noise analysis methods and the ultrasonic analysis method are widely used.

In the middle of the last century, acoustic-based fault diagnosis equipment for electrical equipment has increased attention. The standard method is to determine the outdoor high voltage insulator corona with directional portable ultrasonic sensor, using an acoustic sensor to detect partial discharge in the insulated device of compressed air, and using acoustic detection technology to detect defects in dielectric materials. With the development of computer technology, acoustic detection technology is developing rapidly in the direction of digitization, automation, and real-time. Acoustic detection technology has gradually developed from the initial sound distinguishing by human ears to audio frequency measurement [4-5]. Modern acoustic detection overcomes the disadvantage that traditional acoustic detection is only limited to obtaining the practical value of sound pressure and ignoring its phase position and pays attention to and utilizes the phase position information of sound pressure.

Since the 1960s, the development of array signal processing technology has promoted the new development of noise recognition technology. The noise identification technology of array signal processing technology has been widely used in aircraft noise source distribution, train noise distribution, noise measurement, and fault diagnosis of large equipment. The basic principle of array signal processing noise identification (array noise identification) technology is that a group of microphones is distributed in the sound field space to be measured in a certain way. The microphones array receives the spatial sound field signals. After relevant processing, the signal and signal source and other relevant information can be extracted. Array noise recognition mainly includes two categories of methods: one is called acoustic holography, the other one is called beamforming. Acoustic holography and beamforming are complementary.

1) Acoustic holographic: Acoustic holographic is a method that uses the diffraction and refraction principle of the acoustic wave to reversely derive the sound source information from a specific plane of the measured sound field. In terms of research direction, it can be divided into acoustic imaging and acoustic field analysis. According to the shape of the acoustic holographic surface, it can be divided into planar acoustic holography, cylindrical acoustic holography, and spherical acoustic holography. According to the reconstruction principle, acoustic holography can be divided into far-field acoustic holography, and near field acoustic holography (NAH), far-field and near field are defined according to the ratio of the distance and wavelength between the signal receiving plane and the reconstruction plane.

2) Beamforming: Using an array of microphones distributed at a fixed position in space to measure the spatial sound field, detailed source information can be obtained by special processing of the sound pressure measured by the microphone at each fixed position. This microphone array signal processing technology is called "Beamforming". Array signal processing technology is widely used in radar, sonar, antenna array, medical imaging, geological exploration, radio astronomy, and other fields. With the increasing progress of science and technology and the continuous development of information processing technology, array signal beamforming technology has also made significant progress and has been widely used.

Array signal processing has become one of the effective methods for noise source location identification. The testing method has developed from the traditional single microphone to the array testing system composed of multiple microphones and has developed to real-time testing and processing, and visualization of the sound field.

Sound waves interfere when they travel through the air, which results in different measuring points on a noise receiving surface give different measurement results. Thus, a single microphone cannot make it in the noise measurement of large electrical equipment and noise extraction when receiving point of transmission line noise measurement is far from the sound source. Therefore, we need to design a multiple microphone array, which uses the beamforming algorithm to sum the collected data by delayed stacking. Moreover, we try the equipment by low-frequency noise extraction and test.
systems suitable for power transmission and transformation projects. Doing so can make it more accurate to extract the noise source of electrical equipment and transmission line and provide technical support for subsequent research on noise control, noise abatement, and other tasks.

With the continuous development of signal processing technology and the constant application maturity, speech recognition has been applied more and more in fault diagnosis [6-8]. In any speech recognition system, the first step is to extract features -- that is, to identify the audio signal components that help us recognize the semantic content while discarding other irrelevant information, such as background noise. For the sound signal processing, the more advanced method is based on the MFCC (MEL frequency cepstrum parameter) feature extraction method. In the MFCC extraction process, the traditional signal processing and Fourier transform methods are used to achieve frequency division. MFCC feature extraction method combines the principle of human auditory bionics and the MEL frequency cepstrum coefficient of cepstrum correlation characteristics and can also compensate for the distortion of the convolutional channel to optimize feature extraction. MFCC usually goes through the following steps: pre-weighting, windowing, Fast Fourier Transform (FFT), Meier Filter Bank, Discrete Cosine Transform (DCT). The purpose of pre-weighting is to enhance the high-frequency part so that the signal's frequency spectrum becomes flat. Therefore, the spectrum can be calculated with the same signal-to-noise ratio in the whole frequency band from low frequency to high frequency. Each frame is substituted into the window function, the value outside the window is set to 0, the purpose of which is to eliminate the signal discontinuity caused at both ends of each frame. Transformation because of the sound signal in the time domain are generally hard to see how the signal characteristics, so it is usually converted into frequency domain on the energy distribution to observe, multiplied by the hammering window after each frame must also be through the fast Fourier transform to get on the frequency spectrum energy distribution, so the frame and window after each frame by frame is obtained to carry on the fast Fourier transform spectrum. As there is overlap between filters, the energy values obtained above are correlated. DCT can also reduce dimension compression and abstraction of data to obtain the final characteristic parameters. MEL frequency is proposed based on the acoustic characteristics of human ears. Wavelet transform is used to divide the frequency band. Wavelet function is used as an impulse response, and an appropriate wavelet basis is selected for extraction of the audio feature vector. MEL frequency is put forward based on the acoustic characteristics of human ears, with very high accuracy and efficiency. The accuracy is inversely proportional to the frequency, and it is suitable for use under low-frequency conditions. Therefore, MFCC is very suitable for the noise processing of large electrical equipment.

2. System design

2.1. System design
The hardware of this system is shown in Figure 1, which is mainly composed of a sound signal acquisition module, sound signal processing module, online monitoring module, and power supply
module. The sound signal acquisition module includes a sound sensor and sound collector, the sound processing module includes an emergency storage unit, pre-processing unit, and communication transmission unit, online monitoring module includes host, memory, and fault alarm unit.

**Figure 1.** Hardware block diagram of the system.

The sound sensor collects the on-site noise of the substation in real-time. The sound collector collects the substation site's sound, including the sound of the equipment's regular operation, the sound of various faults, and the noise of the external environment. The sound signal processing module mainly preprocesses and emergency stores the audio signals collected by the sound sensor and the sound collector and transmits the preprocessed audio data to the online monitoring module's host through the communication transmission unit. The online monitoring module host carries out feature extraction, analysis, processing, and fault early warning diagnosis of the audio signal and displays the noise spectrum analysis results through the human-computer interaction interface. Audio data are stored in the memory of the online monitoring module. The fault alarm unit gives an early warning to the fault when the equipment runs abnormal and can make an accurate judgment on the fault situation.

**Figure 2.** Circuit diagram of the sound collector.

**Figure 3.** Software interface diagram of the system.
2.2. Identification and classification of equipment noise

Judging the equipment's running state by monitoring substation equipment operation noise can be summed up in classification problems at the algorithm level. Studies related to noise identification and classification have been extensively developed. The current noise identification and classification is mainly doing pattern recognition based on the equipment noise feature extraction. The traditional signal processing technology can solve the problem of equipment noise feature extraction but show less effect in pattern recognition—it cannot fit the noise model of different types of equipment accurately with changeful background noise.

With the promotion of machine learning and deep learning in pattern recognition, the recognition and classification device based on neural network modeling has been successfully applied to the field of voice recognition. These methods can always be applied to noise pattern recognition operated by the above equipment. At the software level, we use a neural network to classify and recognize equipment noise.

To solve the problem of the limited sound processing capacity of standard model and singleness of extracted sound feature and enable the model to adapt to a broader range of environment, modeling process of recognition and classification device related to neural networks on software level needs to collect a large number of samples to train and adjust the model parameters. After collecting train test data, we still need a lot of voice segmentation, voice tag processing, and other repetitive ways to process data. Sound signal acquisition and training data preprocessing

To ensure that the original sound signal does not lose its feature due to too low sampling rate, a pickup is used to sample the five types of sound related to the transformer sound for 60 minutes at a sampling rate of 44.1 kHz. The data set required by the neural network mainly includes training, validation and testing data.

The training data set is often used to construct the designed neural network model so that the weight parameters of each connection can reach the optimal state. In the construction process, validation data sets can be added to evaluate the network model's performance. The test data set is used to test whether the modeling capability of the established network is perfect. The test data set should not be duplicated with the two data sets mentioned above. Otherwise, it will cause over-fitting. The specific construction process of the two kinds of data sets needs to consider the sample length and sample number.

Since the original sound signals are affected by environment and rhythm changes, if the original sound signals are continuously intercepted as training or test samples, it is easy to lead to insufficient training samples and affect the test results. For the specific segmentation problem of the original data, we consider randomly selecting the starting point. Firstly, the starting position points of random sound segmentation are generated according to the length of the original recording data. The selection of the starting position points shall cover the entire length range of sound data and be evenly distributed. Therefore, select the uniformly distributed random number generation function that takes the length of the recording data as a reference as the random position generation function. To prevent the superposition, repeat problem between generated sound sample data, a generated starting position point need to be compared with all generated starting position points. The gap between the two points is higher than the sample length means that the starting position point has been generated successfully and should be saved to the starting position point array. Otherwise, continue to generate starting position points until they meet the requirements. After a starting position point of sound sample has been determined, we select the data of this position point and the sample data after this position to form another data sample and save it in the corresponding format.
2.2.1. Feature extraction of training data. We choose MFCC, STF, SC, and RMS as feature extraction quantity mainly for the following reasons: 1. MFCC feature comes from the difference in the critical bandwidth of the human ear. Low-frequency linear interval frequency filter and high-frequency logarithmic interval frequency filter are used to retain important speech characteristics of the speech signal. 2. Fourier transform is used to convert signals from the time domain to frequency domain to make it possible to analyze the signal in the frequency domain. However, the disadvantage is that the Fourier transform is a kind of signal of integral transform and unable to provide the frequency spectrum of the signal change over time. Therefore, when analyzing the time-varying signal, using Fourier transform cannot get the ideal effect. The short-time Fourier transform (STFT) is proposed to solve the Fourier transform defects, which can be used to analyze the signal changes in a very narrow time band and a very narrow frequency band. 3. RMS value, also known as an effective value, is the square root of the signal and is used to represent the energy in the signal. 4. Spectral centroid (SC) is an essential feature for signal frequency domain analysis. It is defined as the weighted average frequency within a given frequency band. The weight is the energy of each frequency component, which is the "center of the spectrum" and can detect the approximate position of the spectral peak in the frequency band. As long as the background noise does not contain a strong spectral peak, the signal spectral peak position is almost not affected by the background noise, so the spectral centroid has good robustness to the noise. Specifically, the MFCC, STF, SC, and RMS features are shown from Figure 6 to Figure 12.
2.2.2. Implementation of noise identification network based on LSTM. We use Keras as the neural network framework and then determine the specific parameters of the neural network. These parameters involve the structure of the neural network, the determination of training test data, the activation function of the neurons, the loss function of neural networks, and the optimization function of the neural network.
First, we determined the number of neurons in the input layer, the number of neurons in the output layer, the number of hidden layers, and the number of neurons in each hidden layer. For each training sample, the extracted data dimension is 768. The neuron data in the input layer is initially set to 768 and then adjusted according to the training situation. Since the sound needs to be divided into five categories, the number of neurons in the output layer is set to 5.

A hidden layer was initially set, and the number of neurons in the hidden layer was set as 256. After training, it was found that the cross-validation result of the neural network with a hidden layer was only about 50%, indicating that the structure of the neural network was too simple and the depth of the neural network needed to be increased. Based on this, two hidden layers were set, and the number of neurons in each hidden layer was 128 and 32, respectively. After training, it is found that the cross-validation result of the neural network with two hidden layers is about 80%, which indicates that the neural network has preliminary learning ability, and the specific parameters need to be adjusted to optimize the neural network.

The final network structure is shown in Fig. 14 by constantly adjusting network parameters and combining them with the actual system operation measurement. The final neural network consists of an input layer, an output layer, and two hidden layers. The input layer is fully connected to the hidden layer, and the network input layer is the extracted sound features. The output layer is fully connected to the hidden layer. There is no dropout layer in the whole network.

![Neural network structure](image)

**Figure 14.** Neural network structure.

We tested the performance of the system, and the results met the requirements. The system was installed in the substation, and the sound collector of the system was installed 5 meters away from the transformer to monitor the audio signals of the equipment running in real-time. The monitored audio signal is transmitted to the sound signal processing module through the audio line for emergency storage, preprocessing, and communication transmission. Then the audio signal is transmitted to the host through the 4G private network of the substation for frequency spectrum monitoring, analysis and display, fault early warning and diagnosis, and audio data storage. The results show that the function
of the system is useful. The effective monitoring distance of the system is 30m, the frequency response range is 20Hz~20kHz, the sensitivity is -68dB, and the signal-to-noise ratio is greater than 80dB.

3. Conclusions
Because the regular operation of the transformer has its inherent acoustic characteristics, and each fault has its inherent acoustic characteristics, the fault diagnosis system based on acoustic characteristics proposed by us can monitor various sounds in the operation of the transformer. Through data processing, analysis, and comparison with the database data, it can effectively realize the on-line intelligent fault diagnosis for the transformer. The advantages of this system are obvious: First, the measurement signal is completely free from the influence of the electromagnetic interference of the power system. Second, the monitoring and diagnosis system does not affect the regular operation of the system at all. Thirdly, because different kinds of sounds have their inherent acoustic characteristics, the accuracy of the monitoring and diagnosis system identification is high, and it is entirely credible.

Through this system, the automatic judgment of the equipment running state is realized. Moreover, remote use of the system is convenient for real-time monitoring of the running condition of equipment in remote areas and unmanned power facilities. It can find problems in advance, avoid the damage of power equipment, and reduce all kinds of unnecessary losses caused by equipment damage and power failure.

References
[1] Shanmugam Chellamuthu,E. Chandira Sekaran. Fault detection in electrical equipment’s images by using optimal features with deep learning classifier [J]. Multimedia Tools and Applications, 2019, 78 (19).
[2] Camila Paes Salomon, Claudio Ferreira, Wilson Cesar Sant’Ana, Germano Lambert-Torres, Luiz Eduardo Borges da Silva, Erik Leandro Bonaldi, Levy Ely de Lacerda de Oliveira, Bruno Silva Torres. A Study of Fault Diagnosis Based on Electrical Signature Analysis for Synchronous Generators Predictive Maintenance in Bulk Electric Systems [J]. Energies, 2019, 12(8).
[3] Haikun Shang, Feng Li, Yingjie Wu. Partial Discharge Fault Diagnosis Based on Multi-Scale Dispersion Entropy and a Hypersphere Multiclass Support Vector Machine [J]. Entropy, 2019, 21(1).
[4] Ali Reza Abbasi, Mohammad Reza Mahmoudi, Zakieh Avazzadeh. Diagnosis and clustering of power transformer winding fault types by cross-correlation and clustering analysis of FRA results [J]. IET Generation, Transmission & Distribution, 2018, 12(19).
[5] He Xiuzhi, Liu Qiang, Yu Wennian, Mechefské Chris K., Zhou Xiaoin. A new autocorrelation-based strategy for multiple fault feature extraction from gearbox vibration signals [J]. Measurement, 2021, 171.
[6] Fawcett Timothy J, Cooper Chad S., Longenecker Ryan J., Walton Joseph P.. Machine learning, waveform preprocessing and feature extraction methods for classification of acoustic startle waveforms [J]. MethodsX, 2021, 8.
[7] Liu Peng Yi, Li Zhi Ming. A Feature Extraction Method based on Local Binary Pattern Preprocessing and Wavelet Transform [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2020, 34(13).
[8] Bibin Sam Paul S, Antony Xavier Glittas, Lakshminarayanan Gopalakrishnan. A low latency modular-level deeply integrated MFCC feature extraction architecture for speech recognition [J]. Integration, 2020.