

Real-Time Monitoring of Unmanned Substation Based on Audio Recognition

Wenming Cao ^a, Erqian Tang ^{b, *}, Guanzheng Tan ^c

College of Information Science and Engineering, Central South University, China

^acaowenming9@163.com, ^btangerqian@163.com, ^ctgz@csu.edu.cn

Keywords: Audio recognition; real-time monitoring; failure recognition; unmanned substation.

Abstract. To make the unmanned power substation perform independently and efficiently, the theory and structure of a real-time monitoring system based on audio recognition is introduced in this paper. It focuses on the monitoring and identification process of audio data. Firstly the audio signal of substation is acquired by system using sensors. Then multi-audio data is transmitted to monitoring host. After spectrum analysis and process, MFCC feature parameters are extracted as signal features. Feature parameters are inputted into sinusoidal neural network for identification which can identify the operating status and type of failure efficiently and precisely. Numerical results illustrate that this method improves the traditional manual monitoring method and significantly increases the efficiency of device detection.

1. Introduction

With the development of power grid and the increase of requested security, transformer construction is developing towards unattended or less people on duty. At present, unmanned transformer substations mainly tends to develop towards “five remotes” (remote measurement, remote communication, remote control, remote adjusting and remote viewing) ^[1,2], transformer substation has gradually made remote viewing using video monitoring system realized. However, monitoring on inner sound of the operating equipment is not paid enough attention and developed suit facility is of a small quantity ^[3]. Smart audio identification system used on operating conditions of power equipment is much less.

Electrical devices can make different kinds of noise during operation. We can identify the operating status of devices according to changes of the intensity of sound, even failure category [4]. For example, when a 10KV distribution transformer is operating normally, there is a slight and uniform “buzzing” sound, which is a normal phenomenon of the core vibrations. If abnormal sound of distribution transformer occurs, the reason can be complex. Hence we may consider in the following aspects:

If the sound is heavier than usual, it shows that the distribution transformer is overloaded.

If the sound is pitched than usual, it shows that the power supply voltage is too high.

There will be chaotic and noisy sound when the inner core structure of distribution transformer is loose.[5]

When crackle sound exists in distribution transformer, it shows that there is a coil or core insulation breakdown.

When the drop-out fuse or the switch tap has bad contact, or other outer circuit has a failure, it can also cause the sound variation of distribution transformer.

In order to judge the specific type of distribution transformer failure, the people on duty in the transformer substation usually put one side of insulation rod on the oil tank of distribution transformer and on the other side close ears for careful listening [6]. Although this method operates easily, the remote online detection can't be realized which is not matched with the developing thread of unmanned transformer substation [7]. Also it requires testing personnel with strong practical experience which brings unstable factors to the accuracy of the judgment.

In order to master the real time operating status in a comprehensive way, detect failure devices in time and make scientific decision and process, it is necessary to study a new method of sound monitoring on unmanned devices which can monitor the operating status online via monitoring the sound signal during operation [8]. Also we should pay attention to match the transformer substation with existing monitoring system, master the actual operating status of devices in an all-around way, and optimize the management of transformer substation which is of practical significance to extend the equipment life [9].

2. Voice frequency monitoring principle and system structure

2.1 Voice frequency monitoring principle and system structure

Unmanned transformer substation voice frequency monitoring system is mainly used on detection of sound signal from transformer, capacity, reactance, GIS and other primary devices during operating.

As shown in Figure 1, multi-channel audio signal is transmitted to the microcontroller processor after sensor data acquisition, filtering, amplification and AD conversion. After procession, audio data is then transmitted to industrial personal computer of transformer substation^[10], extract audio feature parameters after the audio spectrum analysis, then recognize the audio features using neural network, output the state of the device and the fault type according to the audio feature vector^[11], and the monitoring host give alarm or take other measures to tackle with the failures based on the judgment results^[12].

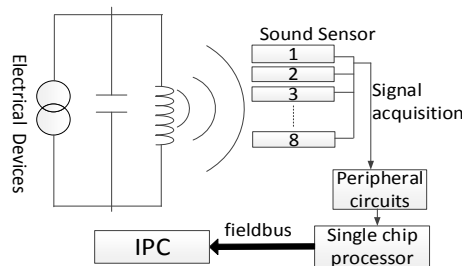


Figure 1. Audio signal monitoring principle

2.2 Voice frequency monitoring system structure

The structure of unmanned devices voice monitoring system is shown in Figure 2, the hardware of system is consisted of sound signal acquisition module (including sound sensor, field bus and data collector), fault type identification module (substation host) and human-computer interaction module (control station host) and power supply module^[13].

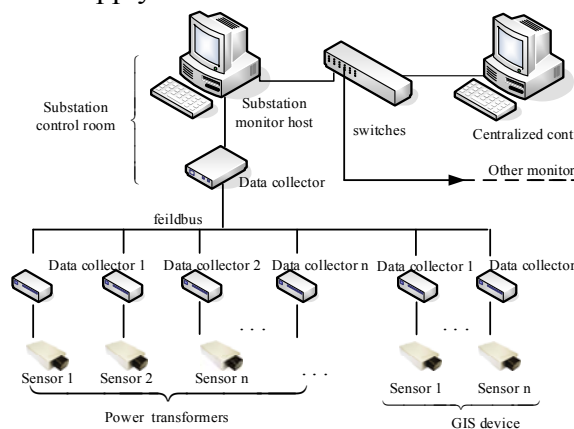


Figure 2. Structure of audio monitoring system in Power Substation

Firstly, many sound sensors are installed in the electrical devices which need monitoring in transformer substations (as shown in Figure 1). Data collector is responsible for multi-channel audio data collection. Data collector is responsible for collecting audio data of multi-collector from a specific substation and it will upload them to substation monitoring host^[14]. Substation monitoring host then shares process multi-channel audio data based on device failure type and device priority and

makes conversion and spectrum analysis to audio data, distracts eigenvectors from audio data and then input eigenvectors to trained neural network for identification, recognize the operating status and failure type of devices which give this sound. Eventually give alarm or other control measures according to the judgement results.

3. System key technologies

3.1 Audio data collector

Data collector is an important component to realize unmanned audio monitoring in substations. It is responsible for audio data processing of electrical devices and primary identification, its inner structure is shown in Figure 3.

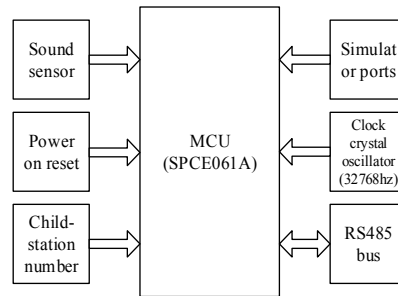


Figure 3. Hardware construction of audio data collection terminal

Audio signal from audio sensor is processed by data collector in the following procedures, signal amplification, low-pass filtering, AD conversion, audio compression, signal primary judgement, serial communication process, and etc. Its core processor is LINGYANG SPCE061A audio processor which integrates components of signal amplification, AD converter and serial communication connector. It has powerful audio process ability which can satisfy audio signal processing of this project. The IOB7 and IOB10 ports of SPCE061A single-chip can be set as UART from general IO Ports, which provides hardware conditions for RS422 communication. The baud rate of UART ranges from 1500bps to 51200bps (51.2Kbps), the communication speed can meet the requirement of status monitoring. In order to realize multi-machine communication, each audio data collector has its own address. Collectors' address coding is realized by single-chip reading seven IO binary codes, according to this address, data collectors make multi-machine communication of each substation possible by polling each data collector.

In order to reduce the upper layer server's calculating and processing to audio data, data collector from down layer has to make simple local pre-judgement to the collected audio data to find whether the collected data belongs to failure audio via pre-judgement^[16]. Its basic principle is: firstly, make feature statistics for typical failure audio and storage the statistic results in local database. Then compare the feature of the current audio data with the feature of failure audio data rapidly. For example, when the amplitude of audio signal appears to be over-high, over-low significantly or noise signal overdue, we can judge that it belongs to failure audio rapidly. For these abnormal situations, data collector immediately send the judgement results to data receiver via RS485 bus, and then inform the monitor host via data collector.

We pre-judge data which does not belong to failure audio via data collector, and then upload it to monitor host for further analysis. When data collector receives the request for monitoring a particular device from monitor host, compress and code the audio data immediately, then transmit it to data receiver. Data receiver transfer data of RS485 proposal to TCP/IP proposal network data using serial ports server and then communicate with monitor host. Finally it will transmit the audio data to monitor host for further judgement and process.

3.2 Device operating status identification module

Device operating status identification module is the core module of transformer substation audio monitoring system, its basic identification process is shown in Figure 4. Monitor host make frequency domain transform and spectral analysis for the uploaded audio data, and then extract the feature parameters of audio signal's MFCC (Mel Frequency Cepstrum Coefficient)^[6], next input

feature vector to trained Sine-based neural network for calculation, identify device operating status characterized by this audio data. At the same time, monitor host make appropriate control and process according to the judgement results, making sure the devices are operating safely.

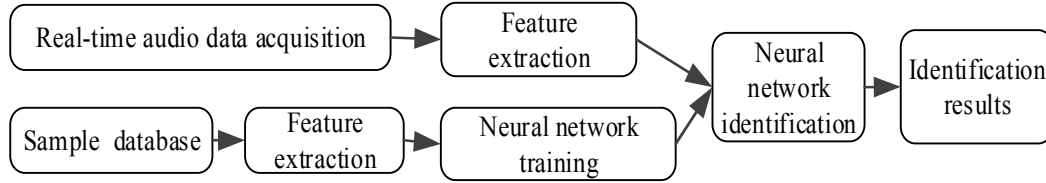


Figure 4. Recognition process of audio data

Audio signal feature extraction

Extracting audio feature effectively is the key for identifying audio. Human's inner ear base membrane can make regulation function to incoming signal. Actually, it acts as filter group which can identify varieties of voice normally even in a noisy environment and under variation circumstances. It has good identifying functions even when SNR (Signal to Noise Ratio) reduces. Mel frequency is raised based on the feature of human's inner auditory sense^[17], it has a nonlinear corresponding relationship with HZ frequency. What's more, the MFCC algorithm simulates human ear filter's function, and designs a model of auditory frontier filter group^[18,19]. This project extracts Mel frequency Cepstral feature parameter as feature parameter for judgement and identification from electrical devices. The steps of extracting Mel frequency Cepstral feature parameter are as follows.

Step1: Pre-emphasis

Make the digital audio signal $s(n)$ after sampling through a high-pass filter: $H(z) = 1 - a \times z^{-1}$ $0.9 < a < 1.0$ (usually takes approximately 0.95). The signal after pre-emphasis is $H(n) = s(n) - a \times s(n-1)$

Step 2: Framing

Usually we take 10~20ms as one frame, in order to avoid the window border spectrum signal leakage, when doing shifting to frames, we need frame selecting (the overlap parts between frames). Usually we take half a frame as a frame shift, in other words, we shift half a frame and then take the next frame, by doing this we can avoid too much feature transformation between different frames.

Step 3: Calculating short-time energy

Short-time energy resembles the amplitude of volume, in other words the size of the sound amplitude. We can filter out the minor noisy of voice signal according to this energy value. When the energy value of a frame is below the set threshold, then this frame can be set as silence segment.

Step 4: Windowed

Voice is constantly changing over a long range, it can't be processed without fixed characteristics. So in order to eliminate the signal inconsistency caused by two sides of each frame, we substitute each frame into window function and set the window value to 0. The most usual window functions in use are square window, Hamming window, Hanning window, etc. According to the frequency specialty of the window functions, we usually take Hamming window. The formula is among the range of window $w(n) = 0.54 - 0.46 \times \cos(2 \times \pi \times n / (n-1))$

Step 5: Fast Fourier transform

As the change of voice signal in the time domain is fast and unstable, it is usually converted into the frequency domain for observation. In this moment, its frequency spectrum will change slowly over time. So the frames after windowed are usually calculated for the frequency parameters via FFT transformation.

Step 6: Triangular band pass filter

The spectrum parameters of each frame are passed through Mel scale filter composed by a set of N triangular band pass filter (N is generally 20 to 30). The cover scope of each triangular window assembles the critical bandwidth of human ear in order to simulate human ear's masking effect. Take the logarithmic of each output band and calculate the log energy of each output, $k = 1, 2, \dots, N$. Then

make cosine transformation for these N parameters to get L-order Mel Frequency Cepstral Coefficients.

Audio data identification method

After extracting MFCC audio features, sinusoidal basis neural network which can respond quickly is taken to make status identification and classification for feature vectors ^[20]. The model of sinusoidal neural network is shown in Figure 5:

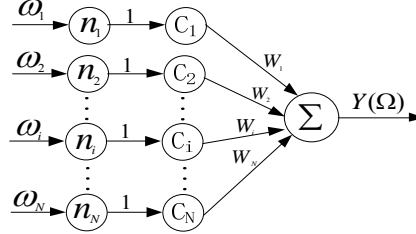


Figure 5. Neural networks model based on sine basis function

Among them, $C_1(\omega_1)$, $C_2(\omega_2)$, ..., $C_N(\omega_N)$ are sinusoidal functions, namely $C_1(\omega_1) = \sin(\frac{1}{2}\omega_1)$, $C_2(\omega_2) = \sin(\frac{3}{2}\omega_2)$... $C_n(\omega_n) = \sin[(n - \frac{1}{2})\omega_n]$, the neural input is N parameters $[\omega_1, \omega_2, \dots, \omega_N]$ extracted by MFCC, namely vector $\Omega = [\omega_1, \omega_2, \dots, \omega_N]$, $\omega_i \in [0, 1]$, among them, ω_i are N parameters extracted by normalized characteristic interval. As the input vector has 16 elements, the number of the selected input layer and hidden layer neural is $N=16$.

As the evaluation result for the transmitter operating status is a determined value, the number of the output neuron is chosen as 1. Assuming that the transformer operation status level is roughly divided into 10 grades, so the expected neural output is $D(\Omega) = [1, 0.9, 0.8, \dots, 0.1]$, neural network output is

$$Y(\Omega) = \sum_{n=1}^N w_n \sin[(n - \frac{1}{2})\omega_n] \quad (1)$$

Assuming that $W = w_n = [\omega_1, \omega_2, \dots, \omega_N]$, $C(\Omega) = [C_1(\omega_1), C_2(\omega_2), \dots, C_n(\omega_n)]$, formula(1) can be rewrite as:

$$Y(\Omega) = \sum_{n=1}^N w_n \sin[(n - \frac{1}{2})\omega_n] = W^T C(\Omega) \quad (2)$$

Assuming that the error function is $e(t) = D(\Omega_t) - Y(\Omega_t)$, among them $t=1, 2, \dots, m$ are trained sample numbers. $D(\Omega)$ is the expected output of neural output, $Y(\Omega)$ is the actual output of neural output. Performance indicator:

$$J = \frac{1}{2} \sum_{t=1}^m e^2(t) \quad (3)$$

Weight adjustment: From (2), (3), there is:

$$\begin{cases} \Delta W = -\eta \frac{\partial J}{\partial W} = -\eta \frac{\partial J}{\partial e(t)} \frac{\partial e(t)}{\partial Y(\Omega_t)} \frac{\partial Y(\Omega_t)}{\partial \Omega_t} \\ = \eta e(t) C(\Omega_t) \\ W(t+1) = W(t) + \eta e(t) C(\Omega_t) \end{cases} \quad (4)$$

Among, η is learning rate and $0 < \eta < 1$. The experiment shows ^[21]: when learning rate is between $0 \leq \eta \leq 2/N$ (N is the number of hidden neuron), this sin-based neural network algorithm is convergent.

The training of audio feature recognizer

Using neural network to classify the audio data automatically, firstly we need to collect audio data of varieties of electrical devices operating under good and error status and train them. M audio signal samples are collected from substation devices, then the feature parameters of m samples are extracted via MFCC algorithm to be trained sample of this status. Each feature parameter is a vector consisting

of N elements. A training method is designed as follows according to the model of neural network and its weight correction method:

Step 1: Initialize neural network weight value ω_i using random number. Make sure the value of learning rate η and arbitrary small positive real number ε .

Step 2: Take the MFCC audio feature value of the first status from substation main devices as the input vector of neural network $\Omega_1 = [\omega_1, \omega_2, \dots, \omega_n]^T$.

Step 3: Calculate the output vector of sinusoidal neuron $C(\Omega) = [C_1(\omega_1), C_2(\omega_2), \dots, C_n(\omega_n)]^T$, then calculate the output of neural network $Y(\Omega_1)$.

Step 4: Compare the actual output with the expected output $D_1(\Omega_1)$ to get error $e(t) = D(\Omega_1) - Y(\Omega_1)$.

Step 5: Calculate $\eta e(t)C(\omega_i)$ according to $e(t)$, adjust output neural weight value $W(t+1) = W(t) + \eta e(t)C(\omega_i)$.

Step 6: Take the second status MFCC audio feature value $\Omega_2 = [\omega_1, \omega_2, \dots, \omega_n]^T$ of substation main device operating status as input. Calculate step 3, 4, 5. Complete m sample data calculation and weight correction of P grades, complete one learning of m trained samples.

Step 7: Calculate overall error $J_{all} = \frac{1}{2} \sum_{i=1}^m e_i(t)$. When $J_{all} > \varepsilon$, turn to step 1 for further training.

When $J_{all} < \varepsilon$, neural network training is completed.

In this way, m samples are taken as feature to train neural network, they convergence to P expected outputs after repeated neural network learning. At last, neural network training are completed to make it have ability of identifying multi operating status of substation devices.

4. Experiments and simulation

4.1 Introduction of experiment devices

The audio monitoring system of unmanned substation was tested in Xinyang 220KV Shagang centralized control center and 110KV industry city substation in Henan province, China. Shagang centralized control center is more than 50km from industry city substation, system devices include: 1 centralized station host, 2 substation hosts, 8 data collectors, terminal box, distribution box, switch machine and communication cables. Centralized station host is installed in Shagang 220KV control center in Xinyang Power Company. The rest main devices are installed in 110KV industry substation in Xinyang province. So far, the performance of audio monitoring system is stable and operating well, its installation site is shown in the following picture.



Figure 6. Audio monitoring host in power substation

In order to ensure the accuracy of signal acquisition, 4 data collectors are installed around NO.1 main transformer in industry city substation and monitor the sound of transformer from different locations. 4 audio data collector are installed in main GIS switch to monitor the operating status of GIS devices (as shown in Figure 7)

The monitor host of substation receive audio signal of electrical devices collected by data collector, analyze the spectrum analysis synchronously, identify the anomalies, and then transmit the spectrum diagram and waveform amplitude diagram to the workers on duty in the centralized host via Ethernet.

Figure 8 shows the waveform spectrum diagram and amplitude spectrum of the audio signal from main transformer. From the figure we can notice that the audio signal is periodic sine wave with

voltage amplitude 120mV, frequency 300Hz. In addition, monitor host extract the audio feature of sound signal using MFCC algorithm. Feature parameters are input into neural network for identifying the operating status of devices. If belongs to error audio, auto alarm are given to inform technicians for handing in time.



Figure 7.Audio collector of GIS switchgear

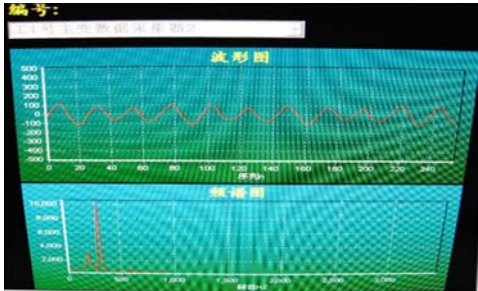


Figure 8.Oscillogram and Spectrum amplitude diagram of voice data for transformer

4.2 calculation and simulation

The feature of audio signals in time domain and frequency domain

First the audio data collectors installed in the substation of Xinyang industry city are used for recording, the audio are stored as 50_good.wav file. Use MATLAB to read audio file for calculating the feature in time domain, the operating results are shown in Figure 9.

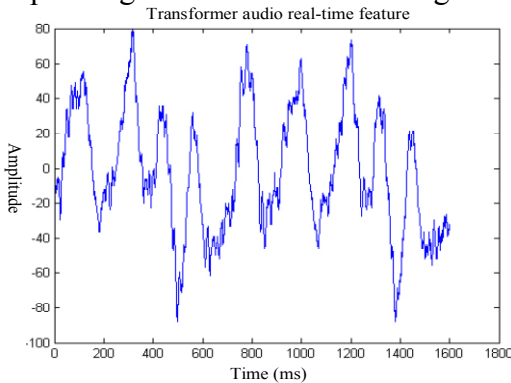


Figure 9.Oscillogram of voice signal for transformers in time domain

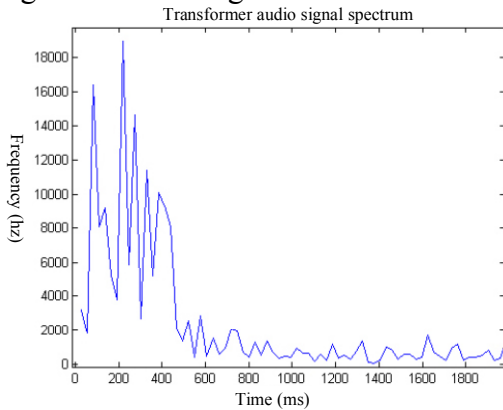


Figure 10.Spectrum amplitude diagram of voice signal for transformers in frequency domain

From the diagram above, the signals are relatively stable in a short time period (20ms) in the time domain which are the most important features, and shows periodism. The feature of other audio signals cannot be characterized significantly, so the audio signals cannot show characteristics completely in the time domain, signals need to be transmitted into frequency domain for analysis with the use of FFT. The operating results are shown in Figure 10.

From the diagram above, the frequency of transformer audio signal mainly focus on 50Hz-500Hz, mainly due to the iron core vibration of the transformer focus on 5 frequency domains. Experiment shows: the peak frequency value and amplitude of audio signal are different significantly under

different operating status. So it is essential to set the initial condition for MFCC calculation according to this feature and extract stable audio signal feature as the basis for identification.

Feature extraction of transformer audio signal

The MFCC feature parameters of main transformer audio data are calculated using MATLAB, the results are shown in Figure 11.

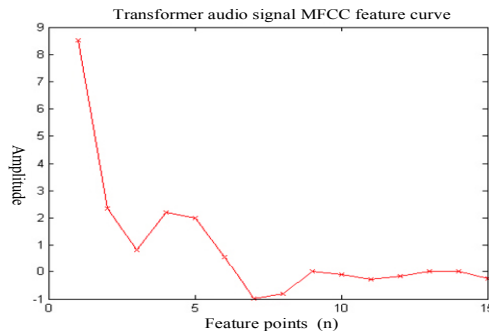


Figure 11. The MFCC characteristic curve of voice signal for transformer

When transformers are under the same operating status, its audio has stable consistency, also the MFCC feature diagram of the same status during different period shows similarity, as shown in Figure 12.

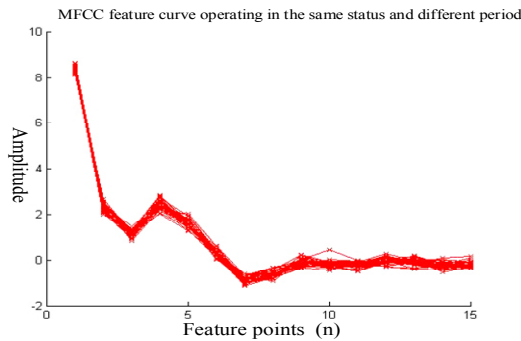


Figure 12. The characteristic curve of transformer voice signal in different periods of the same state

4.2.3 The category of sin-basis neural network identification

Select audio data from 5 different statuses as told before, each 2000 groups with 10000 groups in total. Firstly, select 1500 groups (7500 groups in total) randomly from each status for neural network training and make sin-basis neural network identification for the feature vectors of the rest 2500 groups to get the correct identification rate during different operating status, as shown in Table 1.

Table 1. The recognition rate of transformer voice signal in different states

	Status1	Status2	Status3	Status4	Status5
Accuracy	1.0	0.9861	0.9629	0.9399	0.9206

5. Conclusions

Monitoring the audio signal during electrical devices operating period can detect the error situations effectively. Through this method, effective measures are conducted in time to isolate and eliminate failures which can avoid worsening of the situation. This paper presents a method of identifying audio features precisely and judging the failure type of devices using neural network. It fundamentally improved the tradition manual method and the level of electrical devices failure detection. It also enriches the online device monitoring and maintenance technology which has important application values in future smart grid construction.

Acknowledgements

This work is supported by the Surface project of the State Natural Science Foundation (NO. 2015M572267) and project of the Zhejiang Provincial Natural Science Foundation (NO. LY16F030013).

Reference

- [1] Ding Z.S. Remote Video Monitoring System for Non-attended Substation, J. Popular Utilization of Electricity 2009, 03, 24-25.
- [2] Kong Y.H. A recognition method of abnormal patterns for video surveillance in unmanned substation, J. Journal of North China Electric Power University (Natural Science Edition), 2011,06(38):11-16.
- [3] Xu Lin. Preliminary discussion of Audio Real-Time Monitoring in Transformer Substation J. Mechanical and Electrical Information, 2009(36):155-156.
- [4] Yang X.J. Analysis and Diagnosis of Abnormal Phenomenon Occurring for Distribution Transformer, J. Journal of Beijing Electric Power College, 2011, 12:2-3.
- [5] Cai S.Z. Distributed network of measuring SF6 system for GIS, J. Chinese Journal of Scientific Instrument, 2006, 27(9):1033-1036.
- [6] Chen Y. The extraction and application of phonetic characteristic parameter MFCC, J. Journal of Hunan Agricultural University (Natural Sciences), 2009, 35(1):106-107.
- [7] Zhang X.M. A Noise-robust Speech Recognition Based on Wavelet MFCC, J. Journal of CAEIT, 2008, 3(2):187-189.
- [8] Heittola, T.; Mesaros, A.; Eronen, A.; Virtanen, T. Audio context recognition using audio event histograms, J. 2010. 1272-1276.
- [9] Li Wern Chew; Seng, Kah Phooi; Ang, Li-Minn; Ramakonar, V.;Gnanasegaran, A., J. Audio-Emotion Recognition System Using Parallel Classifiers and Audio Feature Analyzer. 2011,210-215.
- [10] Erdem, E.S.; Sert, M. Efficient Recognition of Human Emotional States from AudioSignals, J. 2014. 139-142.
- [11] Palecek, K.; Chaloupka, J. Audio-visual speech recognition in noisy audioenvironments J. 2013. 484-487.
- [12] Sandhan, T.; Sonowal, S.; Jin Young Choi, J. Audio Bank: A high-level acoustic signal representation for audio event recognition, 2014,82-87.
- [13] Kyuwoong Hwang; Soo-Young Lee, J. Environmental audio scene and activity recognition through mobile-based crowdsourcing, 2012,700-705.
- [14] Weninger, F.; Lehmann, A.; Schuller, B.,J. OpenBliSSART: Design and evaluation of a research toolkit for Blind Source Separation in Audio Recognition Tasks, 2011, 1625 - 1628
- [15] Tamura, S.; Seko, T.; Hayamizu, S.,J. Data collection for mobile audio-visual speech recognitionin various environments. 2014,1-6.
- [16] Grais, E.M.; Topkaya, I.S.; Erdogan, Hakan, J. Audio-visual speech recognition with background music using single-channel source separation. 2012,1-4.
- [17] Alepis, E.; Stathopoulou, I.-O.; Virvou, M.; Tsihrintzis, G.A.;Kabassi, K. J. Audio-lingual and Visual-facial Emotion Recognition: Towards a Bi-modal Interaction System. 2010,274-281.
- [18] Guo C.X. Study of MFCC Speaker Recognition, J. Electronic Science and Technology, 2005(11):53-56.
- [19] Jiang X.H. Audio Data Retrieval Method Based on LPCMCC, J. Computer Engineering, 2009,35(11):246-248.
- [20] ZOU A-jin. Design on the Filter Based on the Sine Basis Function Neural Networks, J. Journal of Changsha University of Electric Power(Natural Science),2001,16(2):16-18.
- [21] Wang X.H. Optimized Design of the Type-four FIR Filter Based on Neural Networks with Sine Basis Functions, J. Journal of Circuits And Systems,2003,8(10):97-100.