Article

Power Quality Disturbances Recognition Based on a Multiresolution Generalized S-Transform and a PSO-Improved Decision Tree

Nantian Huang [1],*, Shuxin Zhang [1,†], Guowei Cai [1,†] and Dianguo Xu [2]

[1] School of Electrical Engineering, Northeast Dianli University, Jilin 132012, China;
E-Mails: zhang_shu_xin@126.com (S.Z.); caigw@mail.nedu.edu.cn (G.C.)

[2] Department of Electrical Engineering, Harbin Institute of Technology, Harbin 150001, China;
E-Mail: xudiang@hit.edu.cn

[†] These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: huangnantian@nedu.edu.cn;
Tel.: +86-135-7855-5676; Fax: +86-432-6480-6691.

Academic Editor: Josep M. Guerrero

Abstract: In a microgrid, the distributed generators (DG) can power the user loads directly. As a result, power quality (PQ) events are more likely to affect the users. This paper proposes a Multiresolution Generalized S-transform (MGST) approach to improve the ability of analyzing and monitoring the power quality in a microgrid. Firstly, the time-frequency distribution characteristics of different types of disturbances are analyzed. Based on the characteristics, the frequency domain is segmented into three frequency areas. After that, the width factor of the window function in the S-transform is set in different frequency areas. MGST has different time-frequency resolution in each frequency area to satisfy the recognition requirements of different disturbances in each frequency area. Then, a rule-based decision tree classifier is designed. In addition, particle swarm optimization (PSO) is applied to extract the applicable features. Finally, the proposed method is compared with some others. The simulation experiments show that the new approach has better accuracy and noise immunity.

Keywords: power quality disturbances; S-transform; multiresolution; particle swarm optimization; decision tree

1. Introduction

Recently, the fast growing microgrid technology provides a good solution to solve the problem of large scale distributed generator coupling. However, compared with the traditional grid, power quality in microgrids has become the subject of interest for utilities as well as energy producers or prosumers [1]. In a traditional grid, power quality can be improved by centralized treatments in substations and it has little influence on the end users, but in microgrids, especially islanded ones, the distributed generators directly power the user loads. Furthermore, the capability of microgrids is relatively small and there exist high proportions of nonlinear and unbalanced loads, so microgrids are more likely to be troubled by PQ (power quality) events, such as harmonics, transient disturbances and so on [2]. These disturbances can lead to serious impacts on power equipment and users' important production processes, which may cause economic damage [3], so monitoring and detailed analysis for power quality are necessary to manage power quality and monitor the status of the devices in microgrids.

Power quality disturbance classification is a basis of analysis and control of power quality, which plays an important role in transient analysis and monitoring of power electronic devices. What's more, through the analysis of PQ disturbances, a large amount of information can be mined to contribute to locating faults, detecting the operation state of systems and troubleshooting [4–6]. Along with electric power technology advances, the monitoring and the analysis of the power quality status will be developed gradually from generation and transmission system to distribution system. As a result, the accuracy and instantaneity of disturbance recognition is much more important now. Additionally, the research now focuses on complex power quality disturbances recognition instead of simplex disturbances [7].

The power quality disturbance classification procedure is always composed of two parts. The first part is signal processing and the second part is pattern recognition [4,5]. In an actual operating process, more than one type of disturbance may take place at the same time. That is what we call a complex disturbance. This results in a higher need for signal processing. Common approaches include short-time Fourier transform (STFT) [8], wavelet transform (WT) [8], S-transform (ST) [9], Generalized S-transform (GST) [10] and Hilbert-Huang Transform (HHT) [11]. WT is a well-developed signal processing method. However, its effect in practical application relies on the choice of base functions and decomposition scale. ST and GST, which are conditioned by the Heisenberg uncertainty principle, have nevertheless been widely used in power quality disturbance classification. When disposing complex disturbances, time resolution and frequency resolution can't be satisfied at the same time. HHT has recently become a popular method because of its full self-adaptability. There exist many problems to solve in HHT, although some progresses have been reached in power quality disturbance classification. For example, end effects, mode mixing problem and bad real-time feature caused by large amount of computation. Pattern recognition techniques include Support Vector Machine (SVM) [12], Artificial Neural Network (ANN) [12], k-Nearest Neighbor (KNN) [13] and decision tree (DT) [14], *etc.* This article chooses DT to design the classifier. Compared with the other approaches, DT can be carried out more easily. Meanwhile, it also has high efficiency that can satisfy the demands of situations when real-time performance is highly needed, although it's worth mentioning that DT's classification effect depends on the feature selected. In the considered situation the classification ability of the feature is great. As a result, DT can be applied. This paper extracts six features from the original signal and S-matrix. In order to strengthen the classification ability of the features, particle swarm optimization

(PSO) is presented to optimize the parameter. PSO is a swarm intelligence method which originates from birds flock's behavior when looking for food [15,16]. The problem discussed in this paper is a one-dimensional continuous optimization problem. Since the traditional methods, analysis methods and methods of exhaustion are not suitable here because the calculation process is complicated, PSO can effectively decrease the computation time, and achieve a solution with a high fitness value.

2. Multiresolution Generalized S-Transform

2.1. Generalized S-Transform

Stockwell proposed the S-transform in 1996 [17]. As an extension to short-time Fourier transform and continuous wavelet transform, it can be expressed as follows:

The input signal is $h(t)$, which after the S-transform will become $S(\tau, f)$:

$$S(\tau, f) = \int_{-\infty}^{\infty} h(t) w(\tau - t, f) e^{-i2\pi ft} dt \tag{1}$$

$$w(t, f) = \frac{|f|}{\sqrt{2\pi}} e^{-t^2 f^2 / 2} \tag{2}$$

where $w(t, f)$ is the Gaussian window function, and $\sigma(f) = \dfrac{1}{|f|}$ is the width of window.

The inverse S-transform can be obtained by utilizing the inverse Fourier transform, which is expressed as:

$$h(t) = \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} S(\tau, f) d\tau \right\} e^{i2\pi ft} df \tag{3}$$

As shown above, considering a discrete signal, T is the sampling interval, N is the total number of sample points. As $f \to n/NT, \tau \to kT$, the discrete form of S-transform can be expressed as:

$$S[k, n] = \sum_{m=0}^{N=1} H\left(\frac{m+n}{NT}\right) e^{-\frac{2\pi^2 m^2}{n^2}} e^{\frac{i2\pi mk}{N}} \tag{4}$$

$$(K = 0, 1, \ldots, N-1; n = 1, \ldots, N-1)$$

With Equation (1), we can see that the S-transform is different from STFT, as the height and the width of gauss window vary with changing frequency, thus it overcomes the defect that the window height and width of the short-time Fourier transform are constant. The S-transform result is a two-dimension matrix, which is defined as the S-matrix. The S modulus matrix can be obtained by modulus arithmetic, whose columns reflect amplitude frequency characteristics of the signal in some certain time, and rows describe time-domain distribution of the signal in the particular frequency.

Different frequency components of the non-stationary signal with distortion will produce different time-frequency distribution characteristics, whereby the high frequency area of a signal changes dramatically, and the low frequency area has a relatively stable change [18]. It is necessary to adjust the window width of the Gaussian window according to the signal analysis requirement, that is, a wider time window in the high frequency area while are more narrow time window can be used in the low frequency area. The width of the S-transform window will be invariant after confirmation of the window function, thus its application is limited and some scholars have thus put forward a generalized S-transform [10],

with the width factor of window function λ, that is, making $\sigma(f) = \dfrac{1}{\lambda|f|}$. With the adjustment of the value λ, the variation rate of width of window varies and the Gaussian window function is expressed as:

$$w(t,f) = \frac{\lambda|f|}{\sqrt{2\pi}} e^{-t^2 f^2/2} \tag{5}$$

Then generalized S-transform is obtained as:

$$S(t,f) = \int_{-\infty}^{+\infty} h(t) \frac{\lambda|f|}{\sqrt{2\pi}} e^{-(\tau-t)^2 \lambda^2 f^2/2} e^{-i2\pi ft} dt \tag{6}$$

According to the Equation (8), the discrete form of the generalized S-transform is expressed as: $\left(f \to n/NT, \tau \to jT\right)$:

$$\begin{cases} S\left[jT, \dfrac{n}{NT}\right] = \displaystyle\sum_{m=0}^{N-1} H\left[\dfrac{m+n}{NT}\right] G(m,n) e^{i2\pi mj/N}, n \neq 0 \\ S[jT, 0] = \dfrac{1}{N} \displaystyle\sum_{m=0}^{N-1} h\left[\dfrac{m}{NT}\right], \qquad\qquad n = 0 \end{cases} \tag{7}$$

where:

$$G(m,n) = e^{-2\pi^2 m^2/\lambda^2 n^2} \tag{8}$$

$G(m,n)$ is obtained from the Gaussian window function by FFT. λ is a constant. When $\lambda=1$ we have a standard form of the S-transform, in other words, the S-transform is a special case of the generalized S-transform. In the generalized S-transform, the time (or frequency) resolution of the time-frequency spectrum can be improved by changing the value of λ, whereby small values of λ correspond to high frequency resolution and large values of λ corresponds to high time resolution. Considering the contrariety between the time resolution and the frequency resolution, in the application of the generalized S-transform, we should choose an appropriate λ value according to the actual situation.

The existing methods to recognize PQ disturbance by using the generalized S-transform generally adjust whether the harmonic components exist. When analyzing the signal with harmonic components, a smaller λ is chosen to obtain a higher frequency resolution. Otherwise, when analyzing the signal without harmonic components, a larger λ is chosen to obtain a higher time resolution, but, when we utilize GST to analyze complex disturbances with harmonic components as well as fundamental frequency disturbances (such as harmonics with voltage sag, harmonics with voltage swell, *etc.*), the GST adopts a smaller λ value. For sag or swell in the complex disturbance signals, GST performs poorly, which affects the classification accuracy.

2.2. Multiresolution Generalized S-Transform

General power quality disturbance signals include voltage sag, voltage swell, voltage interruption, flicker, impulse, notch, harmonics and oscillatory transients, *etc*. According to the time-frequency distribution characteristics of different disturbance signals by S-transform, voltage sag, voltage swell, voltage interruption and flicker are concentrated in the low frequency area, which can be analyzed by the

change of the fundamental frequency amplitude, utilizing higher time resolution. The energy of harmonic signals are distributed in the harmonic frequency area (generally considering odd harmonics under the 13th), which should adopt higher frequency resolution to overcome sidelobe effects. The disturbance energy of the transient oscillation is distributed in the frequency area higher than the harmonic frequency. An adaptively adjusted λ is needed to avoid different kinds of influence including the high frequency energy of other types of disturbance signal and noise. The impulse and notch disturbances occur near the fundamental frequency as well as in the harmonic frequency area, which can be separated from other types by the features of the original signal.

In order to improve the complex disturbance classification ability of the generalized S-transform, the paper proposes MGST. The GST modulus matrix is divided into three areas of low, middle and high frequency. According to the need of disturbance classification in different frequency areas, different λ values are used. The low frequency area ranges from 1 to 100 Hz, which is used to analyze disturbances including sag, swell, interruption, flicker, impulse and notch. The middle frequency area ranges from 101 to 700 Hz, and is used to classify the harmonic components of the disturbance. The part above 700 Hz is the high frequency area, which is used to classify transient components. In order to meet the requirement of the complex disturbance analysis, we need to set different λ values in different frequency areas.

2.2.1. The Setting of the Width Factor λ in the Low Frequency Area

Because signals are processed with different width factors in different frequency areas, there is no need to consider the conflict between time and frequency resolution in different frequency areas. Thus the width of a low frequency window can be narrower than GST. Generally, the amplitude factor of fundamental frequency is used to describe the change of the fundamental [19]. Figure 1 compares the situations of different width factors acting on the amplitude factor of fundamental frequency between sag (black star) and notch (red dot). As shown in Figure 1, when the width factor is increased appropriately, the number of the cross samples declines significantly. Through statistical experiments, the width factor in low frequency area is set as $\lambda_{LF} = \sqrt{5}$ (λ_{LF} represents λ of low frequency).
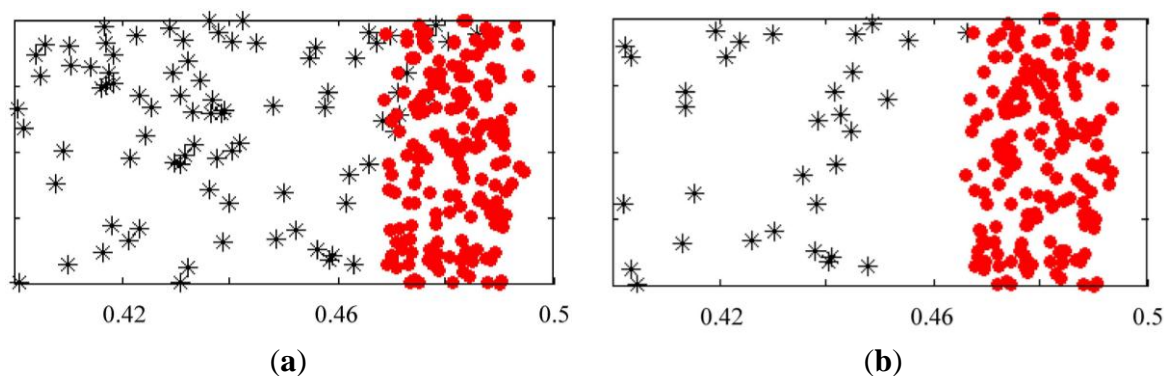

Figure 1. The width factor's impact on the feature in the low frequency area: (**a**) $\lambda = 1$; (**b**) $\lambda = \sqrt{5}$.

2.2.2. The Setting of the Width Factor λ in the Middle Frequency Area

The window width in middle frequency area can be wider than GST so that the classification ability of the harmonic component can be improved to overcome the sidelobe effect. Using the same harmonic

disturbance signal, Figure 2 shows the maximum spectrum amplitude curves with different width factors. As shown in Figure 2, the 5th harmonic, which can't be recognized by traditional ST, can be recognized when the width factor is reduced suitably. Through the statistical experiments, the width factor is selected as $\lambda_{MF} = 1/3$ (λ_{MF} represents λ of middle frequency).
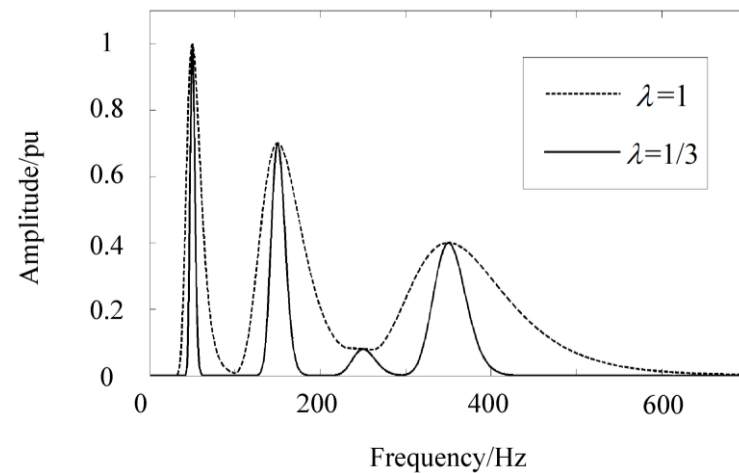

Figure 2. The width factor's impact on the feature in the middle frequency area.

2.2.3. The Setting of the Width Factor λ in the High Frequency Area

The setting of the width factor in the high frequency area aims to reflect the transient component feature. Interruption and sag distribute in the whole frequency range, whose energy changes dramatically near the edge of distortion, so they may be mistakenly identified as the signal containing a transient disturbance. At the same time, a signal with noise may be identified as the signal containing a transient component. However, the requirements in these two cases are different obviously. On the one hand, we should maintain the high frequency energy of the transient to distinguish between sag, interruption and the transient disturbances. In this case, MGST need a higher time resolution. On the other hand, higher frequency resolution is needed to suppress any noise. Therefore, for the width factor in the high frequency area, MGST gives an adaptive setting. Based on the analysis of the fundamental frequency through the Fourier spectrum (A_F), we can judge whether the signal contains a fundamental frequency disturbance (sag or interruption). In this paper, when $0.997\text{pu} \leq A_F \leq 1.003\text{pu}$ (pu represents per unit), we decide that there is no fundamental frequency disturbance and make $\lambda_{HF} = 1/\sqrt{3}$ (λ_{HF} represents λ of high frequency); otherwise, there exists a fundamental frequency disturbance and $\lambda_{HF} = \sqrt{2}$. In this way, by adaptively adjusting the width factor in the high frequency area we can meet different requirements of the window width in different situations.

Figure 3 is the time-frequency contour in the high frequency area, when the same transient signal without disturbance in fundamental frequency is analyzed with different width factors. As is shown, when the window width increases appropriately, the anti-noise performance is improved significantly.
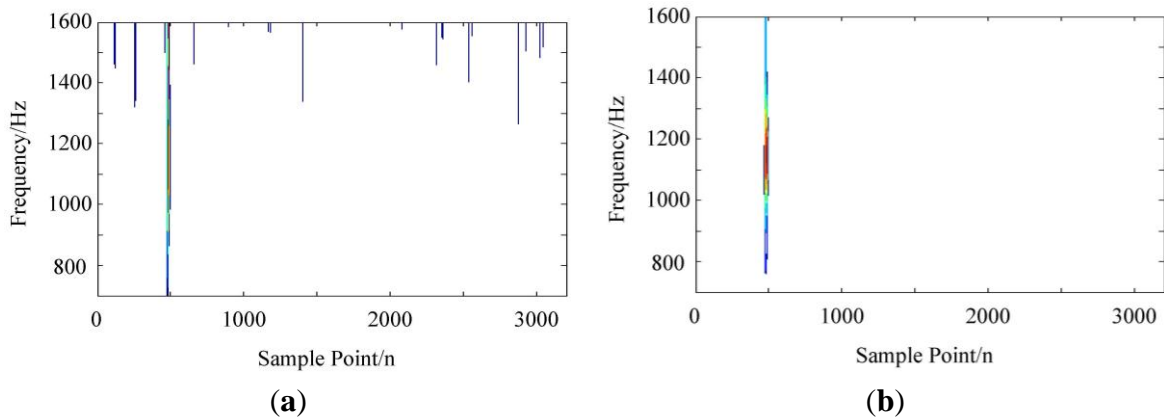
Figure 3. The width factor's impact on the feature in the high frequency area: (**a**) $\lambda=1$; (**b**) $\lambda=1/\sqrt{3}$.

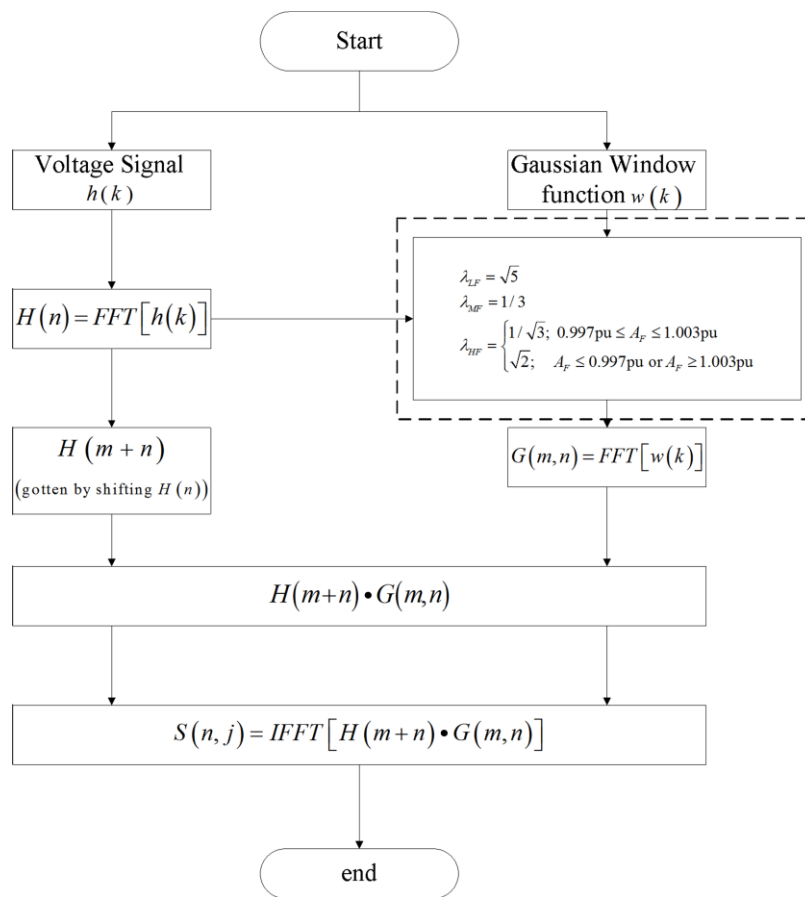The Figure 4 shows the flow chart for MGST.

Figure 4. The flow chart for MGST.

3. PQ Disturbance Classifier Based on Decision Tree

DT recognizes different PQ disturbances by turning a complex classification problem into some binary classification problems. It can be carried out easily and has a high classification efficiency [14]. As the disturbance samples concerned in this paper at each decision node can be differentiated by merely one feature, DT is undoubtedly the best choice.

As a matter of fact, the noise level in power system is not fixed. In [20,21], waveforms are generated through power system simulation which is significant. However, taking various circumstances into consideration, this paper analyzes the PQ disturbance signals simulated by MATLAB 7.0 software with SNR (signal to noise ratio) evenly distributed from 30 to 50 dB. The model references [13,22,23]. The models are given in Table 1. In real power systems, the frequency may change with the fluctuating active power. However, the frequency variation is generally restricted no more than 0.2 Hz (when the capacity is small, the value is 0.5 Hz). The S-transform in this paper divides the frequency domain by 1 Hz so the method is reasonable. What's more, the power quality disturbances researched in this paper usually refer to voltage disturbances occurring in a short time, so there is a negligible effect on the experiment especially after the S-transform method and the assumption is acceptable. The sampling rate is 3.2 kHz. This paper analyzes 13 kinds of disturbances. The types of disturbances and the corresponding class labels considered are denoted in Table 2. For each kind of disturbance 2000 samples are generated. Half of them are used to train the DT, the others are used to test the DT's accuracy.

Table 1. Equations and parameter variations for PQ signals.

Disturbance Class	Modeling Equations	Equations' Parameters
Pure Signal	$h(t) = A\cos(\omega t)$	$A = 1(\text{pu}), f = 50\text{Hz}, \omega = 2\pi f; u(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases}$
Sag	$h(t) = A\{1 - k[u(t_2) - u(t_1)]\}\cos(\omega t)$	$0.1 < k < 0.9; \quad 0.5T \leq t_2 - t_1 \leq 9T$
Swell	$h(t) = A\{1 + k[u(t_2) - u(t_1)]\}\cos(\omega t)$	$0.1 < k < 0.8; \quad 0.5T \leq t_2 - t_1 \leq 9T$
Interruption	$h(t) = A\{1 - k[u(t_2) - u(t_1)]\}\cos(\omega t)$	$0.9 < k < 1; \quad 0.5T \leq t_2 - t_1 \leq 9T$
Flicker	$h(t) = A[1 + \alpha\cos(\beta\omega t)]\cos(\omega t)$	$0.1 \leq \alpha \leq 0.2; \quad 0.1 \leq \beta \leq 0.4$
Transient	$h(t) = A\{\cos(\omega t) + k\exp[-(t - t_1)/\tau]\cos[\omega_n(t - t_1)]\}$	$0.1 < k < 0.8; \quad 150 < 1/\tau < 1000;$ $\omega_n = 2\pi f_n; 700\text{Hz} \leq f_n \leq 1600\text{Hz}$
Harmonics	$h(t) = A\cos(\omega t) + \alpha_3\cos(3\omega t) + \alpha_5\cos(5\omega t) + \alpha_7\cos(7\omega t)$	$0.02 < \alpha_3 < 1, \quad 0.02 < \alpha_5 < 1, \quad 0.02 < \alpha_7 < 1$
Notch	$h(t) = A\cos(\omega t) - \text{sgn}[\sin(\omega t)]\sum_{i=0}^{k}\alpha[u(t_2 + i \cdot 0.02) - u(t_1 + i \cdot 0.02)]$	$1 \leq k \leq 8; \quad 0.1 \leq \alpha \leq 0.4; \quad 0.05T \leq t_2 - t_1 \leq 0.2T$
Impulse	$h(t) = A\cos(\omega t) + \text{sgn}[\sin(\omega t)]\sum_{i=0}^{k}\alpha[u(t_2 + i \cdot 0.02) - u(t_1 + i \cdot 0.02)]$	$1 \leq k \leq 8; \quad 0.1 \leq \alpha \leq 0.4; \quad 0.05T \leq t_2 - t_1 \leq 0.2T$

Table 2. Types of PQ disturbances.

Type of PQ Disturbance	Class Labels
Sag	D1
Swell	D2
Interruption	D3
Flicker	D4
Transient	D5
Harmonic	D6
Notch	D7
Impulse	D8
Harmonic with Sag	D9
Harmonic with Swell	D10
Harmonic with Flicker	D11
Harmonic with Transient	D12
Sag with Transient	D13

3.1. The Analysis of Disturbance Signal via MGST

The time-frequency contour plots of different disturbance signals via MGST are shown in Figure 5. The first part of each disturbance shows the plots of the disturbance signal. The second part is the time-frequency contour that presents frequency values *versus* time values for the S-matrix. The third part, called fundamental frequency amplitude plot, presents amplitude of fundamental frequency *versus* time values. The fourth part, called frequency-mean amplitude plot, presents mean amplitudes *versus* normalized frequency values, and the values in these plots are obtained by calculating the mean value of each row of the S-matrix. The fifth part, called frequency-maximum amplitude plot, presents maximum amplitudes *versus* normalized frequency values, and the values in these plots are obtained by searching the maximum value of the rows of the S-matrix at every frequency. The sixth part, called frequency-standard deviation plot, shows standard deviations (Std) *versus* normalized frequency values, and the values in these plots are obtained by searching the rows of the S-matrix at every frequency. As far as simplex disturbances concerned, features of voltage sag, swell, interruption, flicker, notch and impulse are mainly in the low frequency area. The distortion characteristics of the fundamental frequency are different from the others. Harmonics are mainly in the middle frequency area. Transients are mainly in the high frequency area. Complex disturbances have different features in all three frequency areas. Considering the characteristics of MGST, features are extracted from different frequency areas to recognize disturbances, and in this way more targeted features can be acquired.
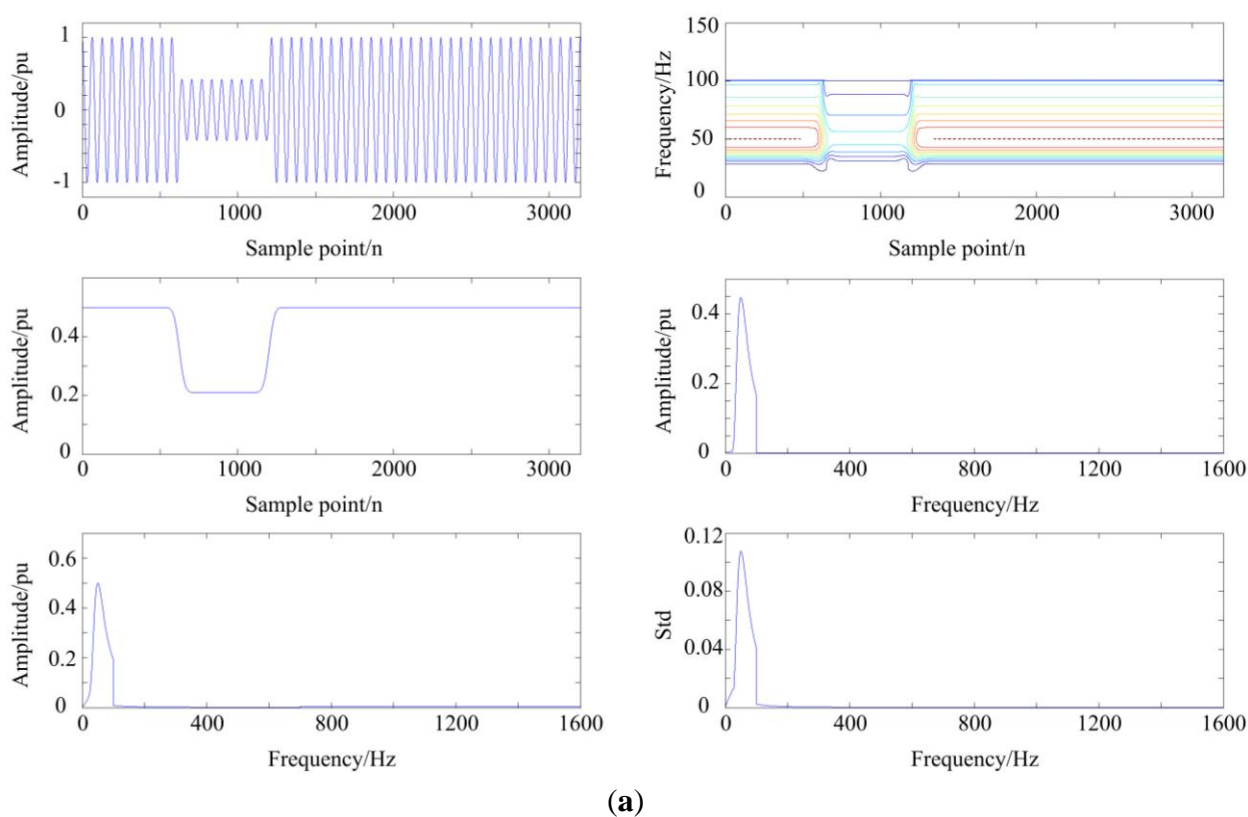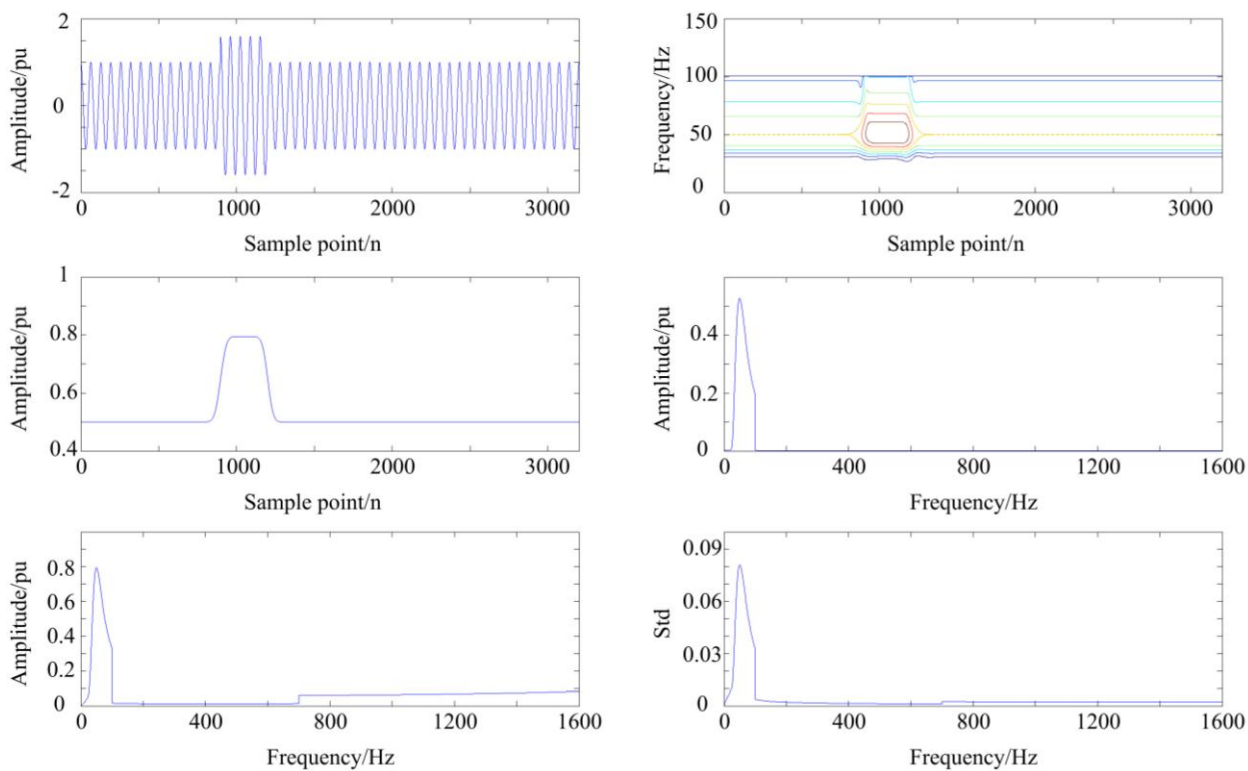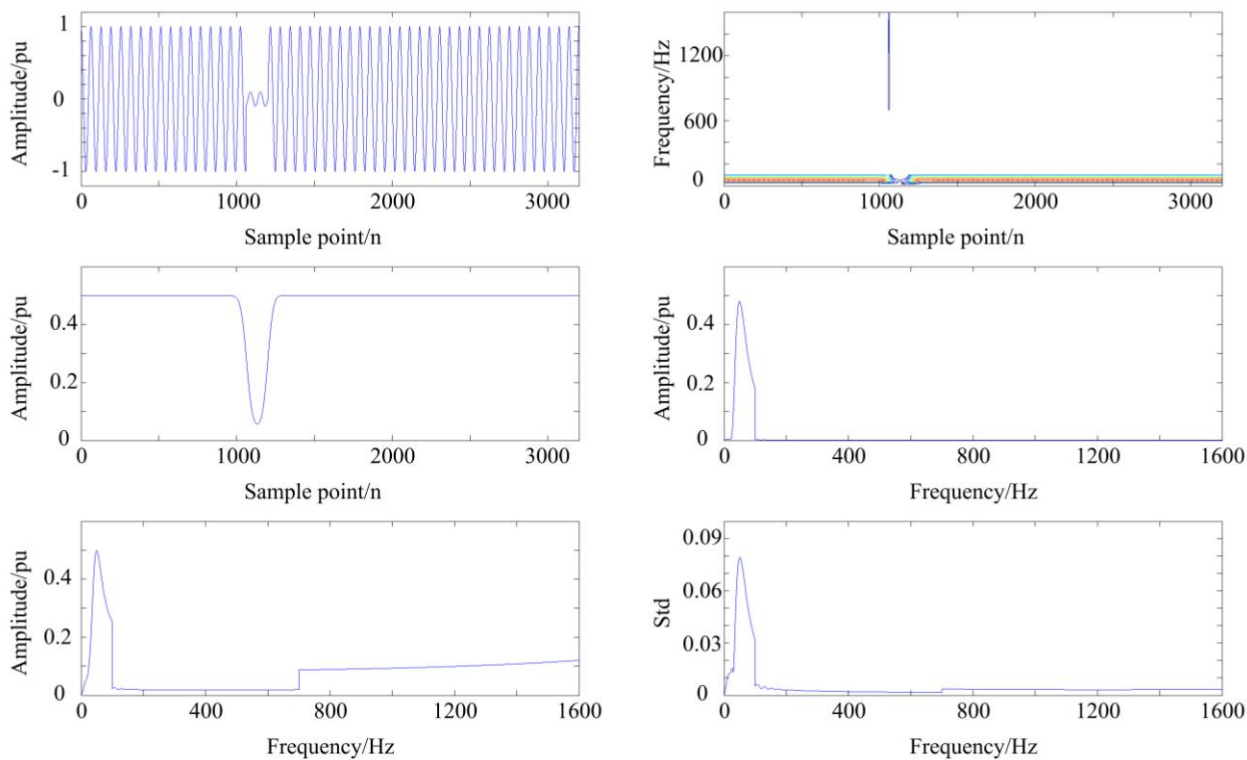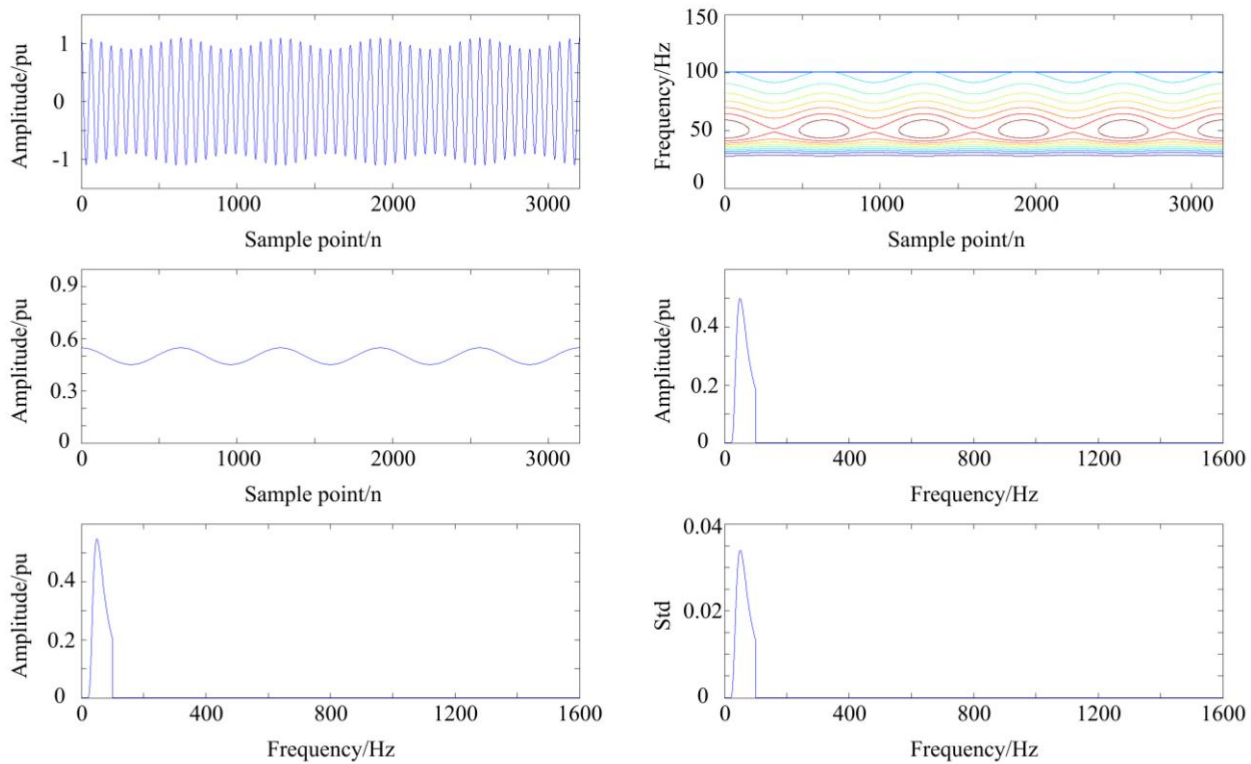

(**a**)

Figure 5. *Cont.*

(**b**)

(**c**)

Figure 5. *Cont.*

(**d**)
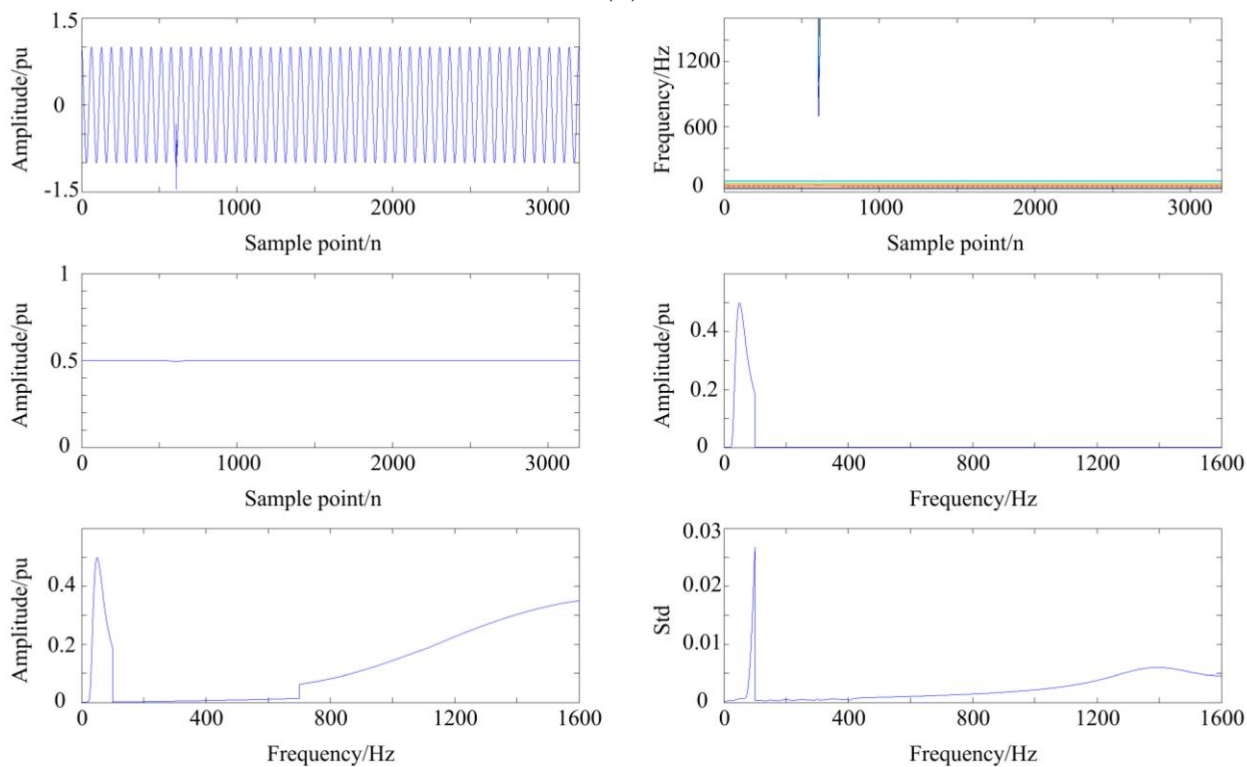
(**e**)

Figure 5. *Cont.*

(**f**)

(**g**)

Figure 5. *Cont.*

(**h**)

(**i**)
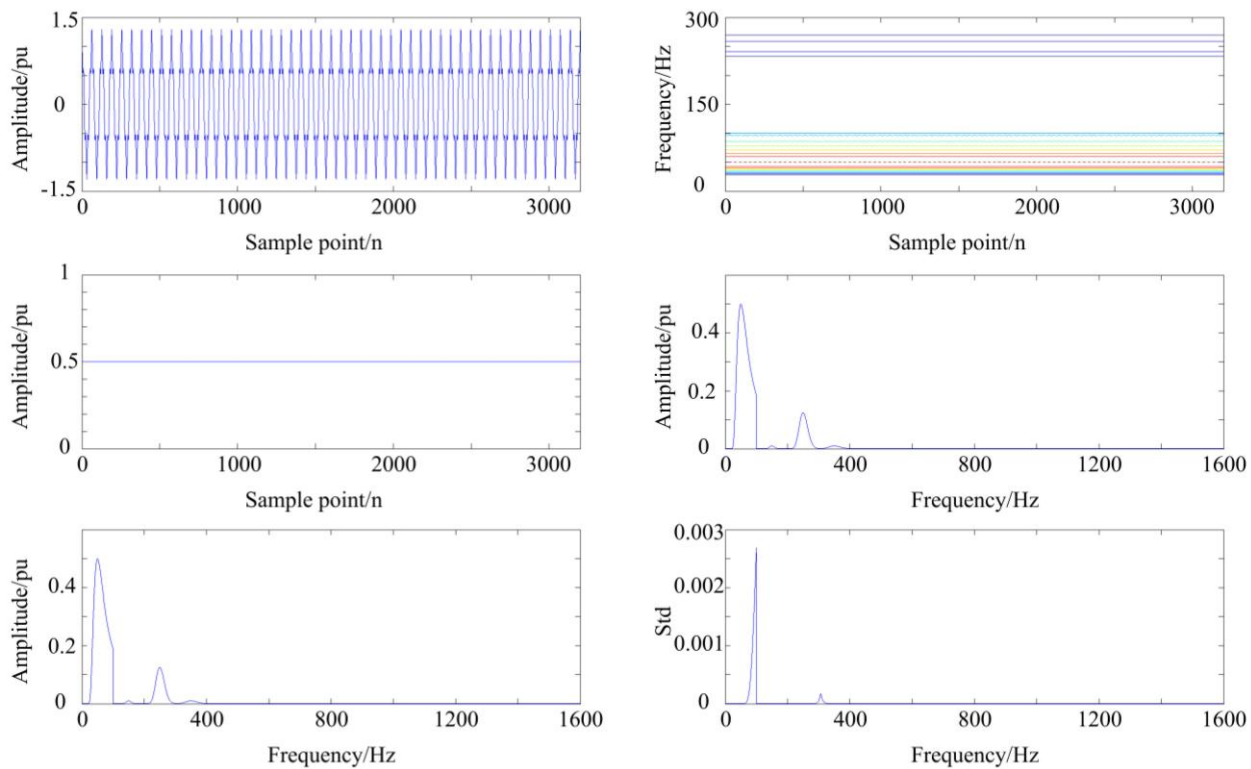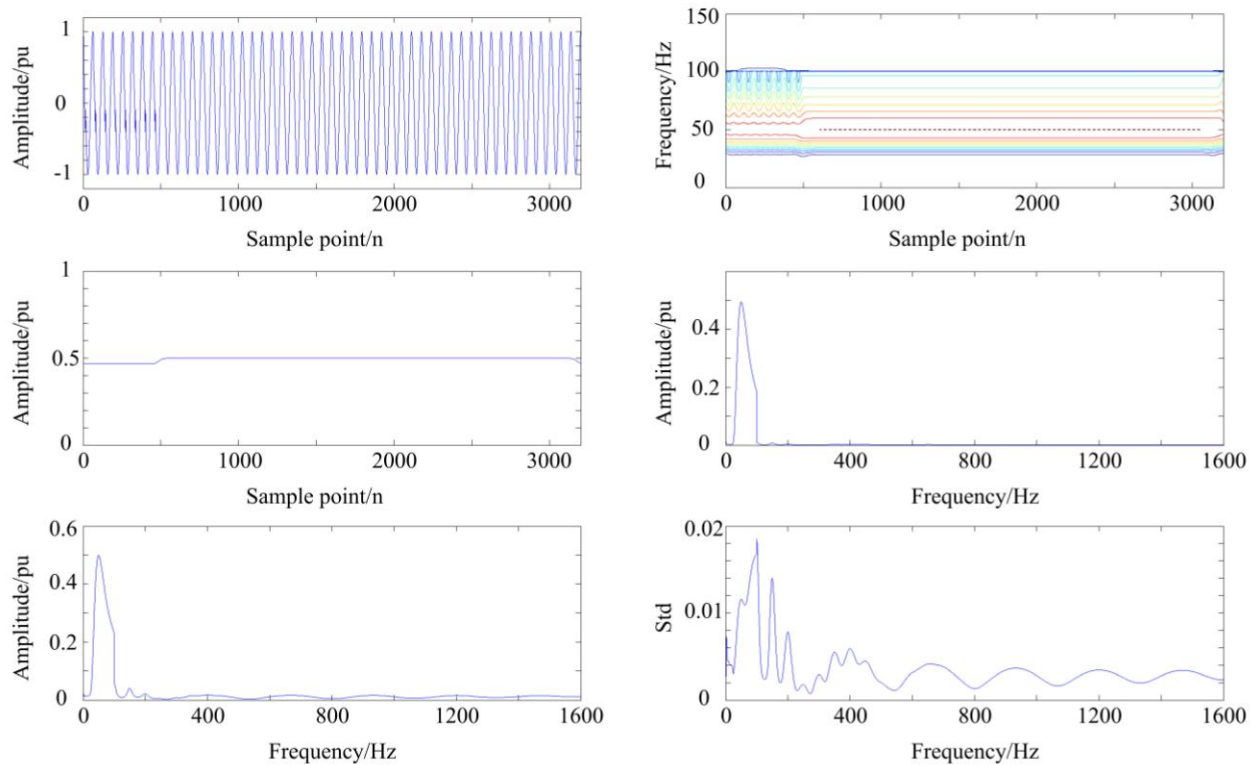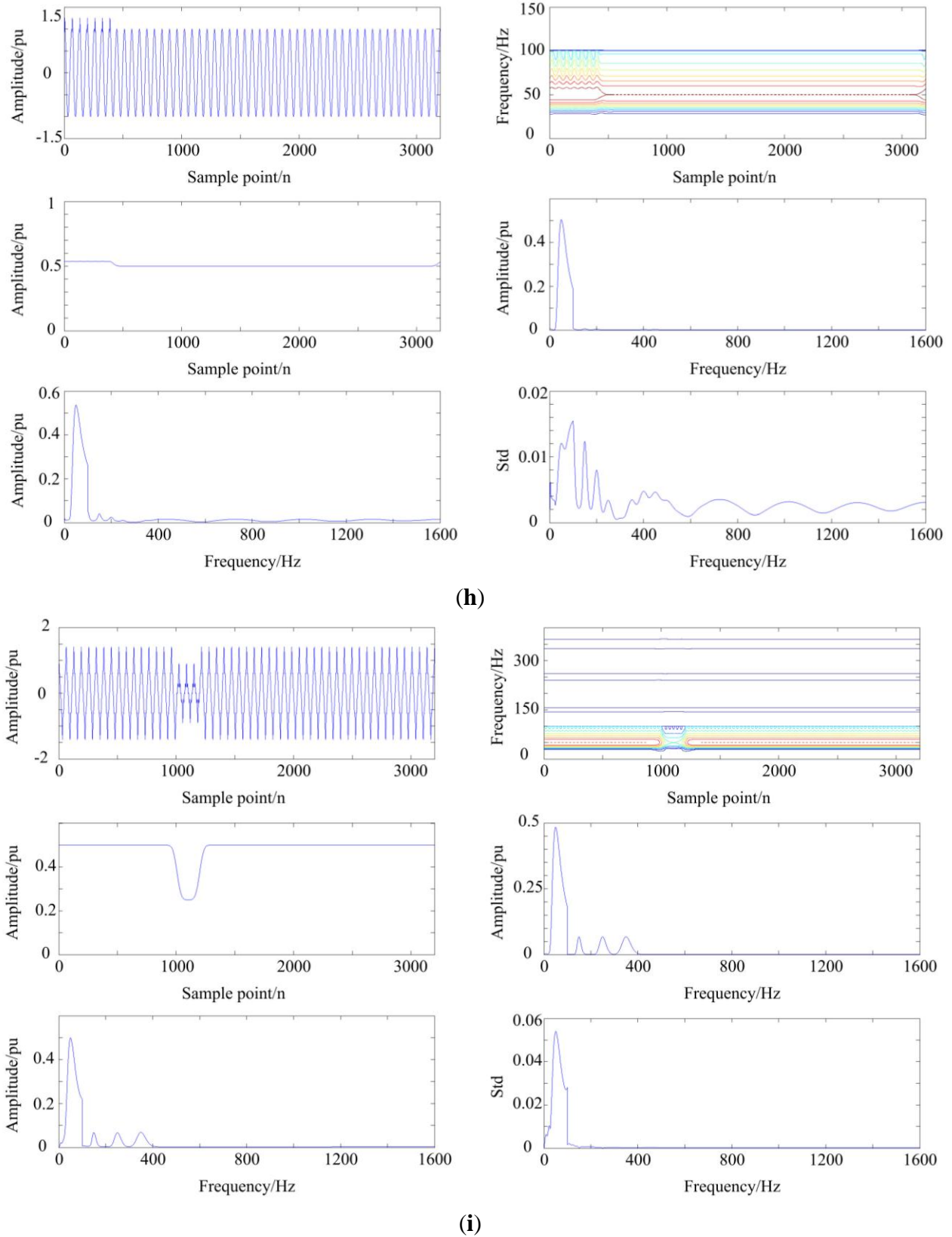
Figure 5. *Cont.*

(**j**)

(**k**)

Figure 5. *Cont.*

(**l**)

(**m**)
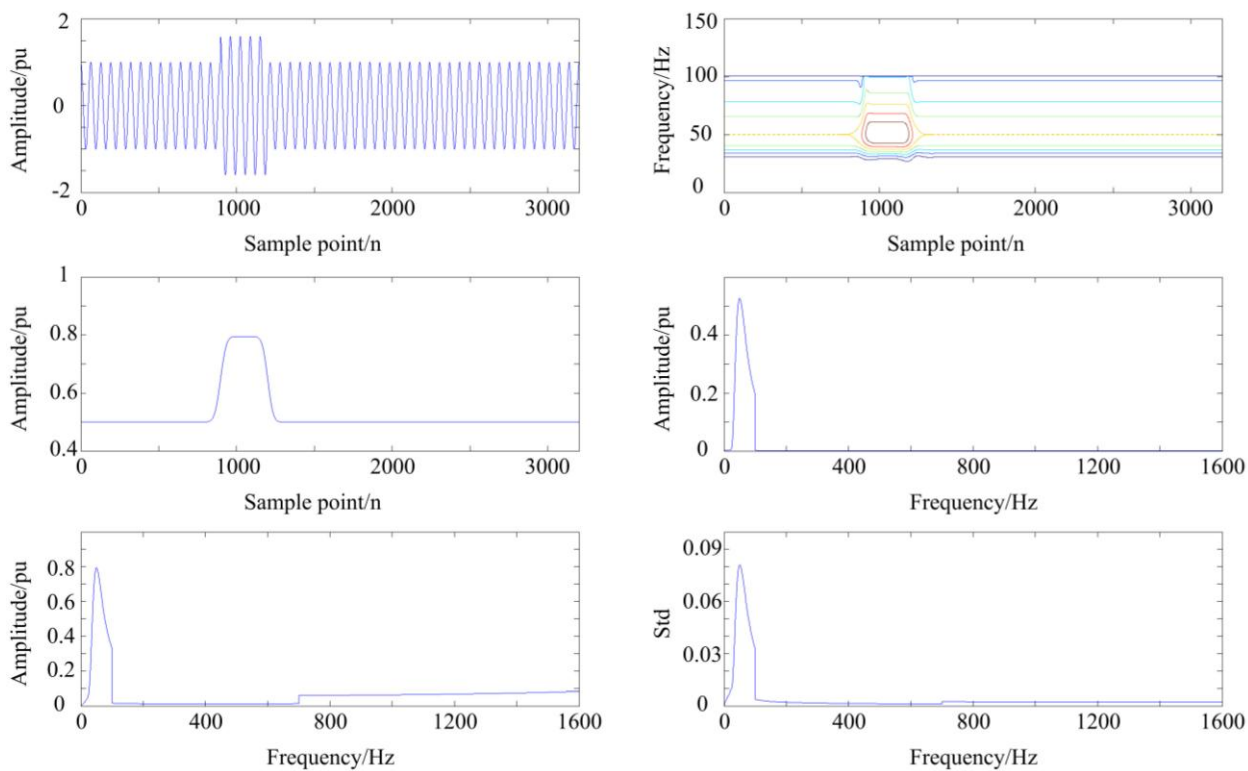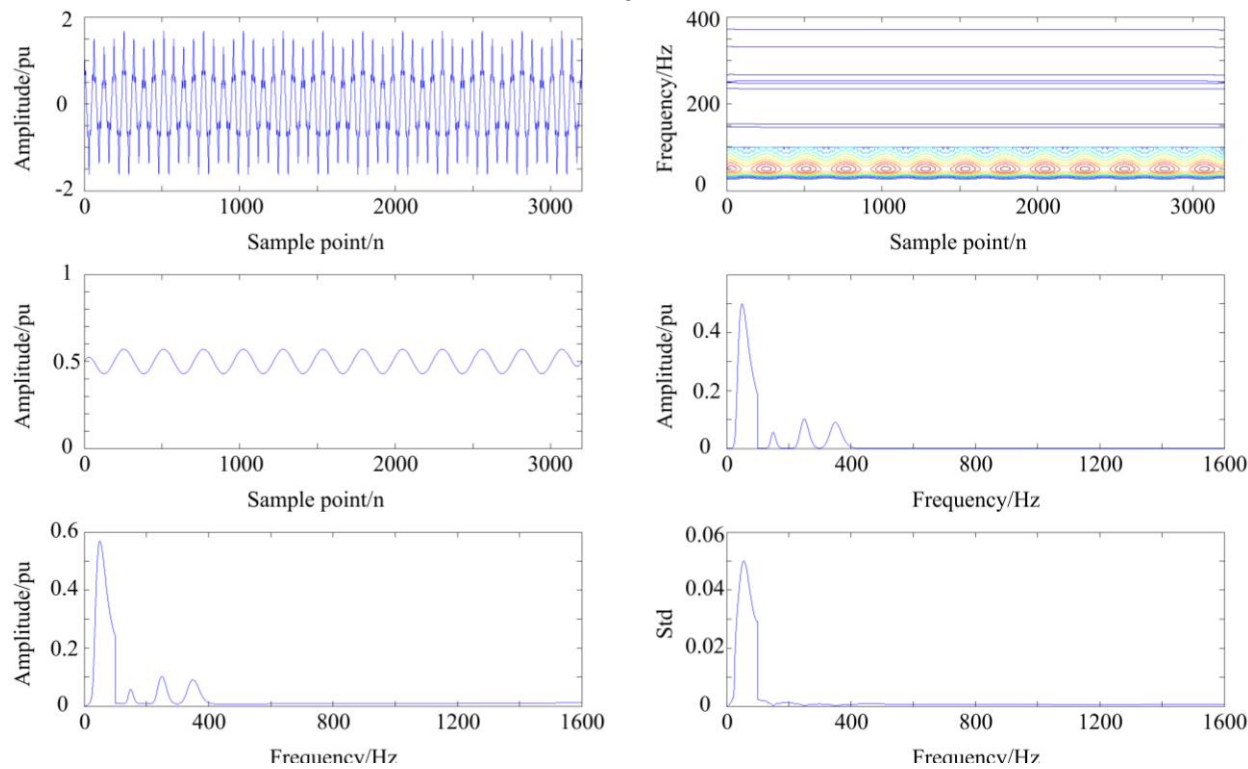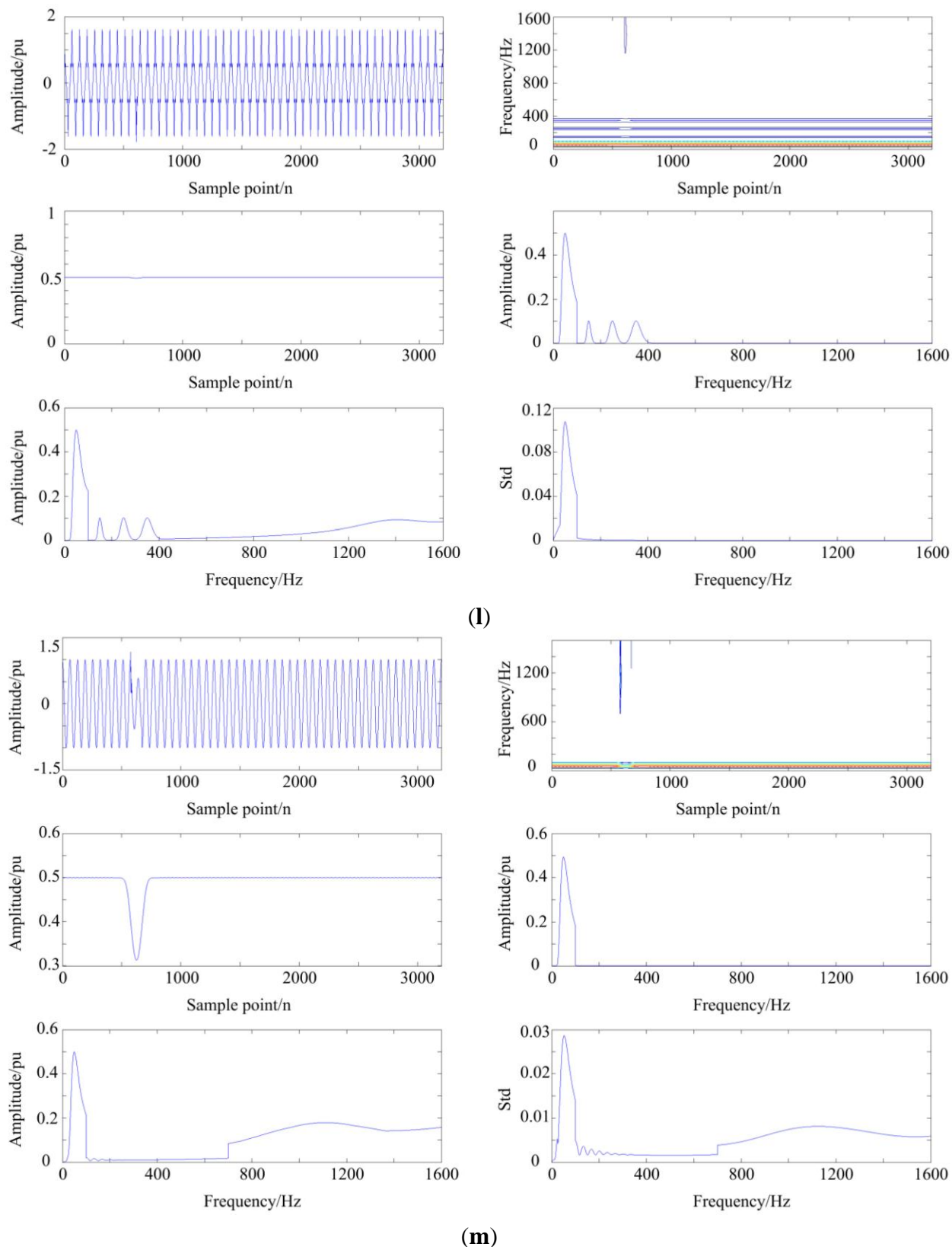
Figure 5. The MGST analysis of 13 disturbances: (**a**) sag; (**b**) swell; (**c**) interruption; (**d**) flicker; (**e**) transient; (**f**) harmonic; (**g**) notch; (**h**) spike; (**i**) harmonic with sag; (**j**) harmonic with swell; (**k**) harmonic with flicker; (**l**) harmonic with transient; (**m**) sag with transient.

3.2. The Structure of DT

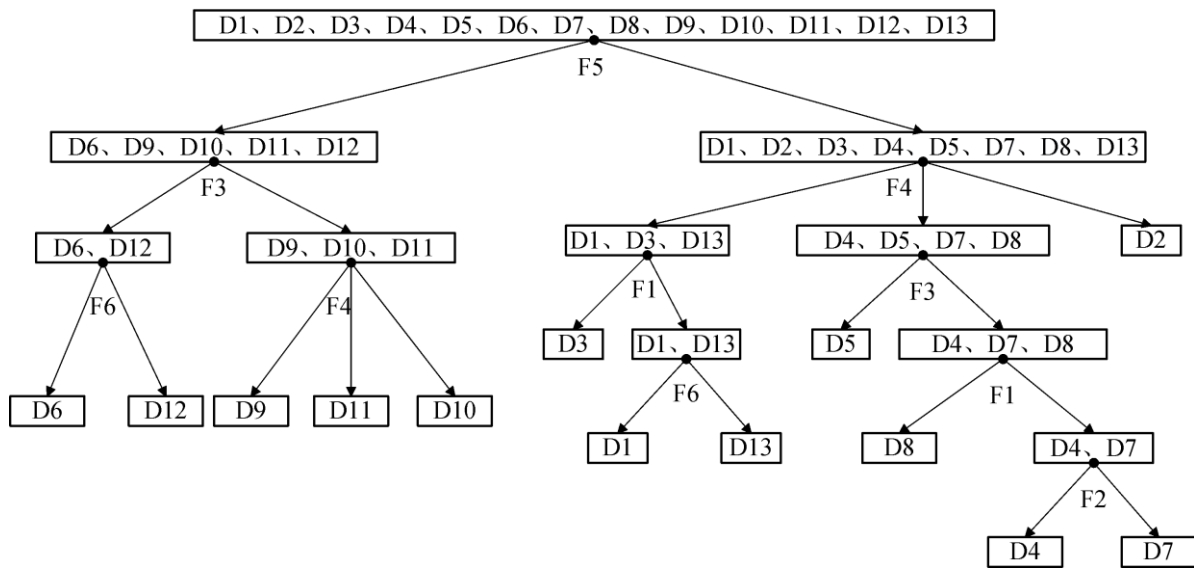The structure of DT classifier is shown in Figure 6.

Figure 6. The structure of DT.

As shown in Figure 6, the features at each decision node are as follows:

Feature 1: The extent of energy falling in 1/4 cycle, we call it *D*.

$$D = \frac{\min\left[R(m)\right]}{R_0} \tag{9}$$

Thereinto, $R(m)$ is root mean square (RMS) of original signal in 1/4 cycle, and R_0 is RMS of standard signal in 1/4 cycle:

$$R(m) = \sqrt{\frac{1}{16}\sum_{k=16m-15}^{k=16m} h^2(k)} \tag{10}$$

Feature 2: The extent of energy rising in 1/4 cycle, we call it *R*:

$$R = \frac{\max\left[R(m)\right]}{R_0} \tag{11}$$

Features 1 and 2 are both selected from original signal, by which we can distinguish whether a sample signal's energy increases or decreases.

Feature 3: Standard deviation of amplitude of fundamental frequency, we call it σ_{Fstd}:

$$\sigma_{\text{Fstd}} = \left\{\frac{1}{N}\sum_j\left[S(n_0,\ j) - \frac{1}{N}\sum_j S(n_0,\ j)\right]^2\right\}^{1/2} \tag{12}$$

where n_0 represents fundamental frequency that in this paper equals 50 Hz. Feature 3 is used to judge whether there exists a disturbance component in fundamental frequency in a sample.

Feature 4: Normalizing factor of amplitude of fundamental frequency, we call it A_f;

$$A_f = \frac{A_{\mathrm{Max}} + A_{\mathrm{Min}} - 1}{2} \tag{13}$$

Thereinto, A_{Max} is the max amplitude value of fundamental frequency, and A_{Min} is the min amplitude value of fundamental frequency. Feature 4 is used to judge whether the amplitude of fundamental frequency of sample signal is increasing, decreasing or stable.

Feature 5: The max value of average amplitude of every frequency in middle frequency area, we call it A_{Mmax};

$$A_{\mathrm{Mmax}} = \max\left[\frac{\sum_j S(n_{\mathrm{H}}, j)}{N}\right] \tag{14}$$

n_{H} represent frequencies in middle frequency area. Feature 5 is used to judge whether there exists harmonic component in a sample signal.

Feature 6: Modified energy of high frequency area, we call it E_{HF}:

$$E_{\mathrm{HF}} = \sum_{n=701}^{1600} \sum_{j=1}^{3200} S_{\mathrm{Th}}^{2}(n, j) \tag{15}$$

Feature 6 is used to judge whether there exists transient component in high frequency area in a sample signal.

S_{Th} present elements that are larger than a certain threshold "T_{s}" in the high frequency area. T_{s} is utilized to reduce the impact of noise, and in this paper we call it "cut-off threshold". By setting the cut-off threshold "T_{s}", the noise resistant ability of classifier is improved without increasing the system complexity, however, it is difficult to determine a reasonable value. In next part, the paper proposes a modified PSO method to solve this problem.

4. Feature Optimization via PSO

4.1. Background

In order to improve the noise resistance ability of the feature, this paper proposes a modified high frequency area energy as Feature 6 to judge whether there exists a transient component in the high frequency area in a sample signal. The main principle is that a cut-off threshold is presented to easily filter the noise in the high frequency area. It is proved that this method effectively improves the accuracy of recognition. Meanwhile, it can be carried out easily without influencing the reliability.

Nevertheless, there is certain difficulty to determine the cut-off threshold. On one hand, choosing too large a value may give rise to excessive energy reduction in some slight transient samples that results in failure. On the other hand, too small a value can't achieve the expected targets of resisting noise. What's more, large amount of samples are needed to enumerate kinds of disturbance signals with different parameters. When a certain "T_{s}" is selected, every sample in sample set should be processed by MGST. As a result, the amount of computations is very huge. Optimization problem about the cut-off threshold in this paper is a one-dimensional continuous optimization problem whose objective function can be expressed as $y = f(T_{\mathrm{s}})$. Thereinto, y represents false recognition rate that is a function depending

on independent variable "T_s". The function can be realized through programming. So the problem can be expressed as follows:

$$\begin{cases} \min f(T_s) \\ s.t. \quad T_s \in S \end{cases} \tag{16}$$

S is the value range of cut-off threshold "T_s". $y = f(T_s)$ can be achieved through programming basing on MGST and feature calculation. On account of the complex computation process, an untraditional computational intelligence technology must be chosen to solve the problem.

4.2. The Search of Optimized Threshold via Modified PSO

Particle swarm optimization (PSO) was proposed in 1995 by Kennedy and Eberhart [15,16]. It is a bionic algorithm which imitates birds flock's looking for food. PSO shows a strong charm in various kinds of problems because of easy realization, less parameters and simple concept. The problem concerned in this paper can be solved effectively by PSO without wasting too much time.

4.2.1. Basic Fundamentals of PSO

The basic fundamentals of PSO can be described as follows:

A swarm constituted by m particles is flown through the D-dimensional search space of the problem. Each particle is regarded as a unit without volume whose velocity is adjusted dynamically based on its own experience and its companions' experience. On that basis, its position (that is, the value of cut-off threshold "T_s") changes. Thereinto, the position of the ith particle is expressed as $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iD})$ and the velocity of the ith particle is expressed as $\boldsymbol{v}_i = (v_{i1}, v_{i2}, \cdots, v_{iD}), 1 \leq i \leq m$. The best position of the ith particle so far is represented as $\boldsymbol{p}_i = (p_{i1}, p_{i2}, \cdots, p_{iD})$, also known as p_{best}. The best particle among all p_{best} is represented as $\boldsymbol{p}_g = (p_{g1}, p_{g2}, \cdots, p_{gD})$, also known as g_{best}. The fitness function is used to evaluate if a certain position is better and in this paper the fitness function is $y = f(T_s)$ in which y represents false recognition rate.

In the kth iteration, the position and velocity in the dth dimension ($1 \leq d \leq D$) of the particles are updated according to the formulas as follows:

$$v_{id}^k = w v_{id}^{k-1} + c_1 rand_1 (p_{id} - x_{id}^{k-1}) + c_2 rand_2 (p_{gd} - x_{id}^{k-1}) \tag{17}$$

$$v_{id}^k = \begin{cases} v_{\max} v_{id}^k & > v_{\max} \\ -v_{\max} & v_{id}^k < -v_{\max} \end{cases} \tag{18}$$

$$x_{id}^k = x_{id}^{k-1} + v_{id}^k \tag{19}$$

The parameter w represents inertia weight, whose value is always set to 1 in the basic PSO. The parameters c_1 and c_2 represent acceleration coefficient. In the basic PSO, it's often the case that c_1 and c_2 are all equal to 2. In the formula, $rand_1$ and $rand_2$ are two pseudo-random numbers between 0 and 1. The velocity of particles are limited to $v_{\max}$. After several iterations, the value of "T_s" is refreshed again and again with the moving of the particles. Along with the changing of "T_s", the false recognition rate is reduced at the same time.

4.2.2. The Improvement of PSO

The above-mentioned PSO is a basic PSO. Many scholars have done a lot of work on different parts of PSO that include population size [24], inertia weight [25], neighborhood topology [26] acceleration coefficient [27,28] and so on. According to the real conditions in this paper, we make improvements of PSO as follows:

Improvement of the inertia weight w

Inertia weight is a very important parameter in PSO that directly affects the balance of global and local searching ability. A large inertia weight promotes global searching ability while a small inertia weight promotes local searching ability.

According to the requirements of the problem concerned in this paper, the authors designed a fuzzy control system to adaptively adjust the inertia weight. The best fitness value at present, known as $f(\hat{y})$, and the inertia weight at present are selected as inputs. The percentage change of the inertia weight is selected as output. Because different problems have different ranges of fitness value, Equation (20) is presented to normalize $f(\hat{y})$:

$$f_{\text{norm}}(\hat{y}) = \frac{f(\hat{y}) - f_{\min}}{f_{\max} - f_{\min}} \tag{20}$$

In Equation (20), $f_{\max}$ and $f_{\min}$ are decided by the specific problem. It is necessary to confirm or estimate their values. Three membership functions (left, middle and right) are presented as Equations (21)–(23) to provide the inputs corresponding to the three fuzzy sets: LOW, MEDIUM and HIGH where a and b are parameters determined by the practical problems. The curves of the membership functions are shown in Figure 7:

$$f_{\text{left}} = \begin{cases} 1, & x < a \\ \dfrac{1}{2} - \dfrac{1}{2}\sin\dfrac{\pi}{b-a}(x - \dfrac{a+b}{2}), & a \leq x \leq b \\ 0, & x > b \end{cases} \tag{21}$$

$$f_{\text{middle}} = \begin{cases} 0, & x < a \\ \dfrac{1}{2} + \dfrac{1}{2}\sin\dfrac{2\pi}{b-a}(x - \dfrac{3a+b}{4}), & a \leq x \leq \dfrac{a+b}{2} \\ \dfrac{1}{2} - \dfrac{1}{2}\sin\dfrac{2\pi}{b-a}(x - \dfrac{3b+a}{4}), & \dfrac{a+b}{2} < x \leq b \\ 0, & x > b \end{cases} \tag{22}$$

$$f_{\text{right}} = \begin{cases} 0, & x < a \\ \dfrac{1}{2} + \dfrac{1}{2}\sin\dfrac{\pi}{b-a}(x - \dfrac{a+b}{2}), & a \leq x \leq b \\ 1, & x > b \end{cases} \tag{23}$$
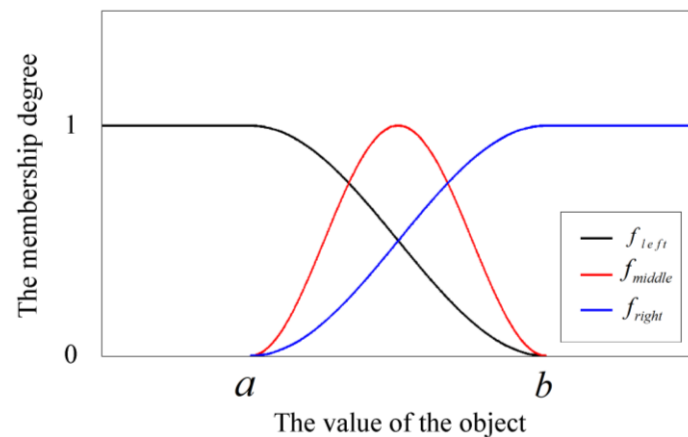
Figure 7. The curves of three membership functions.

In this way, the inertia weight is adjusted adaptively on the basis of the current situation during each iteration. The adjusting strategy is shown in Table 3 below.

Table 3. The adjusting strategy of inertia weight.

Inputs &	Inputs Fuzzy Sets		Output Fuzzy Set
Output	Best Fitness Value at Present	Inertia Weight at Present	Percentage Change of the Inertia Weight
RULE1	LOW	LOW	MEDIUM
RULE2	LOW	MEDIUM	LOW
RULE3	LOW	HIGH	LOW
RULE4	MEDIUM	LOW	HIGH
RULE5	MEDIUM	MEDIUM	MEDIUM
RULE6	MEDIUM	HIGH	LOW
RULE7	HIGH	LOW	HIGH
RULE8	HIGH	MEDIUM	MEDIUM
RULE9	HIGH	HIGH	LOW

Take RULE1 as an example. It expresses that percentage change of the inertia weight should be attached to the MEDIUM fuzzy set (that is, there is no need to change it too much) if the best fitness value and inertia weight at present relatively belong to the LOW fuzzy set. In engineering, there is always too much noise in PQ disturbance signals. Through this method, the local best solution caused by noise is avoided by adjusting the inertia weight. The anti-noise effects and robust performance are enhanced and that is good for improving accuracy.

Improvement of the acceleration coefficient

The acceleration coefficients c_1 and c_2 reflect the communication between the particles. A large c_1 leads to the particles relying on their own experience. As a result, the particles hover in the area near themselves. However, a large c_2 leads to all particles rapidly moving towards the best unit at present. In this case, the particles may converge on local best solution at the beginning of the algorithm. To balance the contradiction, c_1 and c_2 are usually set to be the same value. Nevertheless, this can't satisfy the actual needs sometimes. For the purpose of having a larger c_1 and smaller c_2 at the beginning of the algorithm, decreasing c_1 and increasing c_2, adjustment as follows is utilized:

$$c_1 = (c_{1f} - c_{1i}) \frac{I}{I\_\max} + c_{1i} \qquad (24)$$

$$c_2 = (c_{2f} - c_{2i}) \frac{I}{I\_\max} + c_{2i} \qquad (25)$$

Hence, at the beginning of the algorithm, the particles tend to search for a better T_s in the whole search space. At the end of the algorithm, the particles tend to move towards the best unit of the swarm. By this means, computation time is reduced, but no solutions would be left out.

In the formula, c_{1i} and c_{1f} are the initial and final value of c_1. Analogously, c_{2i} and c_{2f} are similar to c_{1i} and c_{1f}. I is the current iteration time and $I\_\max$ is the maximum number of iteration. In [27], they change symmetrically, that is c_1 decreases linearly from 2.5 to 0.5 and c_2 increases linearly from 0.5 to 2.5. This method behaves well concerning single hump functions but not ideally concerning multi-peak functions, and it may lead to premature convergence. To solve the problem, [28] proposes asymmetric changing. It's found that when c_1 decreases linearly from 2.75 to 1.25 and c_2 increases linearly from 0.5 to 2.25, most of the functions can be solved well, so this paper selects this method to adjust the acceleration coefficient.

In this paper, the false recognition rate of different cut-off thresholds T_s is chosen as the fitness function. The simulation disturbances are generated by MATLAB and then white Gaussian noise is added to make the SNR distribute from 30 to 50 dB. When a certain cut-off threshold T_s is input, the values of the features are calculated by MGST, so the false recognition rate corresponding to the certain T_s is acquired. The fitness function is realized through programming on the basis of the abovementioned way and then the modified PSO is utilized to optimize the cut-off threshold T_s. After multiple tests, 0.0192 is taken for the optimal T_s.

5. Simulation and Experiment

The simulation signals of different SNR are generated to test the new method. There are 1000 signals in each group whose SNR are 30~50 dB, 30 dB, 40 dB and 50 dB. Decision trees based on ST [9] and GST [10] are also made for comparison with MGST. The result for 30~50 dB is shown in Table 4 and the results of 30 dB, 40 dB and 50 dB are shown in Figure 8.

Table 4. Accuracy in percentage using different techniques in 30 to 50 dB.

Disturbances	SNR = 30–50 dB		
	MGST	GST	ST
D1	99.7	99.3	98.5
D2	99.6	98.9	98.0
D3	99.5	99.5	99.5
D4	100	99.7	99.7
D5	100	100	100
D6	100	100	93.3
D7	99.4	97.0	97.0
D8	99.7	96.6	96.6

Table 4. *Cont.*

Disturbances	MGST	GST	ST
SNR = 30–50 dB			
D9	100	98.8	99.8
D10	100	98.7	99.6
D11	100	45.5	96.5
D12	94.7	93.9	98
D13	86.4	84.9	73.4
%Average accuracy	98.38	93.29	96.15

As shown in Table 4 and Figure 8, MGST does better than ST and GST when recognizing all kinds of disturbances in different noise levels. What's more, consistent with the theory, complex disturbances are recognized much better by MGST.
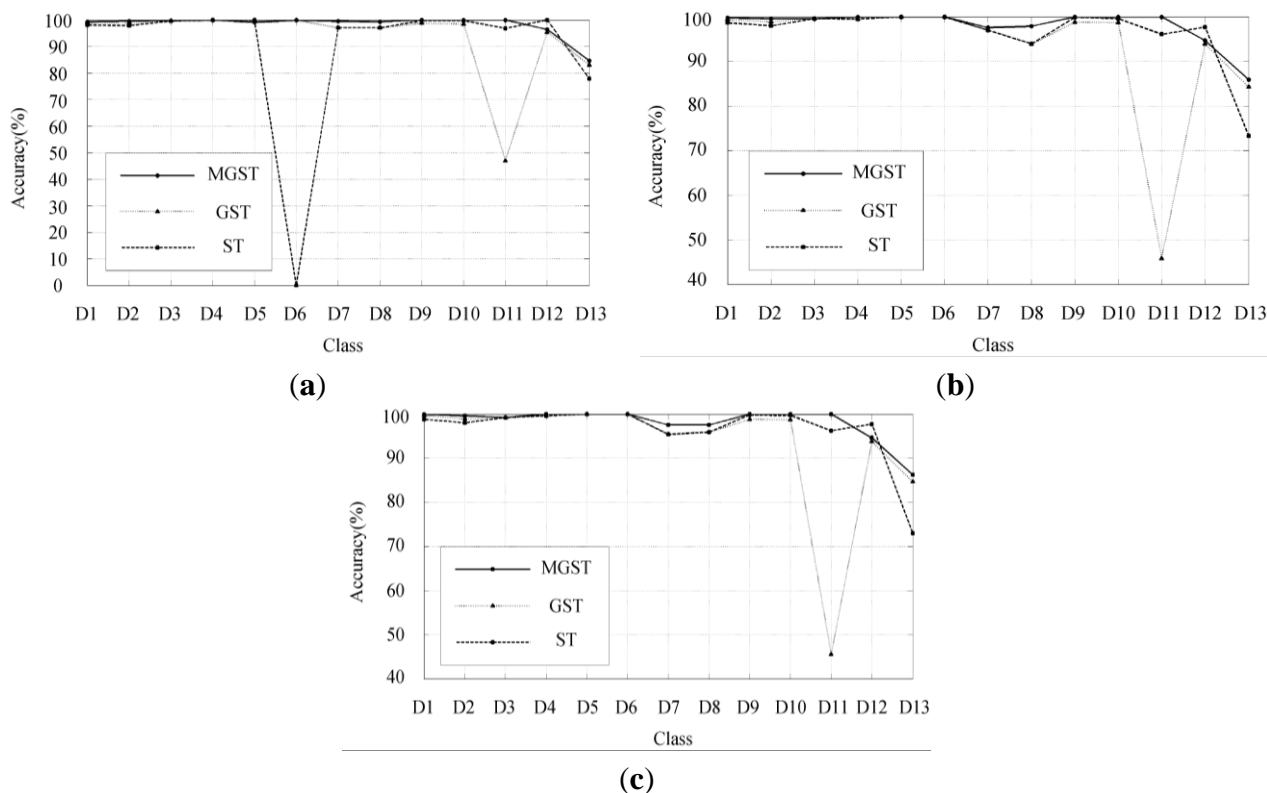

(**a**)

(**b**)

(**c**)

Figure 8. Comparison of classification accuracy. (**a**) 30 dB; (**b**) 40 dB; (**c**) 50 dB.

6. Conclusions

(1) The width factor of the window function in the S-transform is set respectively in different frequency areas, which improves MGST's ability to recognize complex disturbances.

(2) Features are extracted in different frequency areas so as to avoid the disturbances caused by components of different frequency areas. Moreover, the amount of calculation is reduced.

(3) A cut-off threshold T_s is proposed and optimized by modified PSO, which further improves the anti-noise performance.

(4) Simulation experiments show its effectiveness and practicability.

Future works will focus on further optimizing the width factor of the window function in different frequency areas and reducing the computational burden of MGST.

Author Contributions

The paper was a collaborative effort between the authors. The authors contributed collectively to the theoretical analysis and manuscript preparation.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Wasiak, I.; Pawelek, R.; Mienski, R. Energy storage application in low-voltage microgrids for energy management and power quality improvement. *IET Gener. Transm. Distrib.* **2014**, *8*, 463–472.
2. Prodanovic, M.; Green, T.C. High-quality power generation through distributed control of a power park microgrid. *IEEE Trans. Ind. Electron.* **2006**, *53*, 1471–1482.
3. Baggini, A. *Handbook of Power Quality*; Wiley Blackwell: New York, NY, USA, 2008.
4. Saini, M.K.; Kapoor, R. Classification of power quality events—A review. *Int. J. Electr. Power Energy Syst.* **2012**, *43*, 11–19.
5. Granados-Lieberman, D.; Romero-Troncoso, R.J.; Osornio-Rios, R.A.; Garcia-Perez, A.; Cabal-Yepez, E. Techniques and methodologies for power quality analysis and disturbances classification in power systems: A review. *IET Gener. Transm. Distrib.* **2011**, *5*, 519–529.
6. Salles, D.; Wilsun, X. Information extraction from PQ disturbances—An emerging direction of power quality research. In Proceedings of the 2012 IEEE 15th International Conference on Harmonics and Quality of Power (ICHQP), Hong Kong, China, 17–20 June 2012; pp. 649–655.
7. Farzanehrafat, A.; Watson, N.R. Power Quality State Estimator for Smart Distribution Grids. *IEEE Trans. Power Syst.* **2013**, *28*, 2183–2191.
8. Gu, Y.H.; Bollen, M.H.J. Time-frequency and time-scale domain analysis of voltage disturbances. *IEEE Trans. Power Deliv.* **2000**, *15*, 1279–1284.
9. Rodríguez, A.; Aguado, J.A.; Martín, F.; López, J.J.; Muñoz, F.; Ruiz, J.E. Rule-based classification of power quality disturbances using S-transform. *Electr. Power Syst. Res.* **2012**, *86*, 113–121.
10. Xu, F.; Yang, H.; Ye, M.; Liu, Y.; Hui, J. Classification for power quality short duration disturbances based on generalized S-transform. *Proc. Chin. Soc. Electr. Eng.* **2012**, *4*, 77–84.
11. Biswal, B.; Biswal, M.; Mishra, S.; Jalaja, R. Automatic classification of power quality events using balanced neural tree. *IEEE Trans. Ind. Electron.* **2014**, *61*, 521–530.
12. Janik, P.; Lobos, T. Automated classification of power-quality disturbances using SVM and RBF networks. *IEEE Trans. Power Deliv.* **2006**, *21*, 1663–1669.

13. Panigrahi, B.; Pandi, V.R. Optimal feature selection for classification of power quality disturbances using wavelet packet-based fuzzy k-nearest neighbour algorithm. *IET Gener. Transm. Distrib.* **2009**, *3*, 296–306.

14. Fengzhan, Z.; Rengang, Y. Power-quality disturbance recognition using S-transform. *IEEE Trans. Power Deliv.* **2007**, *22*, 944–950.

15. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks, Perth, Australia, 27 November–1 December 1995; Volume 1944, pp. 1942–1948.

16. Eberhart, R.; Kennedy, J. A new optimizer using particle swarm theory. In Proceedings of the Sixth International Symposium on Micro Machine and Human Science (MHS '95), Nagoya, Japan, 4–6 October 1995; pp. 39–43.

17. Stockwell, R.G.; Mansinha, L.; Lowe, R.P. Localization of the complex spectrum: The S transform. *IEEE Trans. Signal Process.* **1996**, *44*, 998–1001.

18. Pinnegar, C.R.; Mansinha, L. The S-transform with windows of arbitrary and varying shape. *Geophysics* **2003**, *68*, 381–385.

19. Uyar, M.; Yildirim, S.; Gencoglu, M.T. An expert system based on S-transform and neural network for automatic classification of power quality disturbances. *Expert Syst. Appl.* **2009**, *36*, 5962–5975.

20. Nguyen, T.; Liao, Y. Power quality disturbance classification utilizing S-transform and binary feature matrix method. *Electr. Power Syst. Res.* **2009**, *79*, 569–575.

21. Nguyen, T.; Liao, Y. Power quality disturbance classification based on adaptive neuro-fuzzy system. *Int. J. Emerg. Electr. Power Syst.* **2009**, *10*, doi:10.2202/1553-779X.2110.

22. Hooshmand, R.; Enshaee, A. Detection and classification of single and combined power quality disturbances using fuzzy systems oriented by particle swarm optimization algorithm. *Electr. Power Syst. Res.* **2010**, *80*, 1552–1561.

23. Huang, N.; Xu, D.; Liu, X.; Lin, L. Power quality disturbances classification based on S-transform and probabilistic neural network. *Neurocomputing* **2012**, *98*, 12–23.

24. Chen, D.; Zhao, C. Particle swarm optimization with adaptive population size and its application. *Appl. Soft Comput.* **2009**, *9*, 39–48.

25. Yuhui, S.; Eberhart, R.C. Fuzzy adaptive particle swarm optimization. In Proceedings of the 2001 Congress on Evolutionary Computation, Seoul, Korea, 27–30 May 2001; Volume 101, pp. 101–106.

26. Kennedy, J.; Mendes, R. Neighborhood topologies in fully informed and best-of-neighborhood particle swarms. *IEEE Trans. Syst. Man Cybern. C Appl. Rev.* **2006**, *36*, 515–519.

27. Ratnaweera, A.; Halgamuge, S.; Watson, H.C. Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. *IEEE Trans. Evol. Comput.* **2004**, *8*, 240–255.

28. Wenzhong, G.; Guolong, C.; Xiang, F. A new strategy of acceleration coefficients for particle swarm optimization. In Proceedings of the 10th International Conference on Computer Supported Cooperative Work in Design (CSCWD '06), Nanjing, China, 3–5 May 2006; pp. 1–5.