

Pattern Recognition of Monitored Waveforms from Power Supplies Feeding High-Speed Rail Systems

Wei Gu[†], Shuai Zhang*, Xiaodong Yuan**, Bing Chen** and Jingjing Bai*

Abstract – The development of high-speed rail (HSR) has had a major impact on the power supply grid. Based on the monitored waveforms of HSR, a pattern recognition approach is proposed for the first time in this paper to identify the operating conditions. To reduce the data dimensions for monitored waveforms, the principal component analysis (PCA) algorithm was used to extract the characteristics and their waveforms from the monitored waveforms data. The dynamic time wrapping (DTW) algorithm was then used to identify the operating conditions of the HSR. Cases studies show that the proposed approach is effective and feasible, and that it is possible to identify the real-time operating conditions based on the monitored waveforms.

Keywords: High-speed rail (HSR), Power quality, PCA, DTW, Pattern recognition

1. Introduction

Because high-speed rail (HSR) has the advantages of convenience and speed, it has been developed rapidly in the world since the inception of the first HSR, the Tokaido Shinkansen. In particular, because many countries are building HSRs, the proportion of the HSR power load has increased each year. Because HSR has a high capacity, impact and random power load, it has a tremendous power quality (PQ) impact on the power supply grid, such as voltage fluctuations, harmonics and negative-sequence currents [1-3]. However, given current technology, it is difficult to analyze HSR power quality (PQ) characteristics by monitoring waveforms because of its complicated and variable operating conditions, especially when it switches between different operating conditions, because the switching process is so short. Therefore, it is urgent to provide an in-depth study of the PQ problems caused by the HSR operating conditions, and research on how to identify operating conditions based on monitored waveforms. Then, corresponding relation between PQ anomalies and HSR operating conditions can be established, and targeted PQ control measures can be taken to the reliability, security and economy of the power suppliers and the railway authorities [4].

The traction load of the HSR is a PQ disturbance source for the power grid, and the changes in the operating conditions of HSR cause disturbances that affect the quality of the grid power supply. The few research studies on the PQ problems of HSR mainly concentrate on the

upgrading and transformation of the related technical and PQ issues it caused [5-7]; while the research on PQ disturbances is relatively mature [9-13]. Mishra et al. [9] presented an S-Transform based probabilistic neural network (PNN) classifier for the recognition of PQ disturbances, which can reduce the features of the disturbance signal and needs less memory space and PNN learning time. To eliminate the high-order harmonics in the grid currents under grid voltage disturbances, Nguyen et al. [10] proposed a novel scheme for the current controller of the grid-side converter of a permanent-magnet synchronous generator. Huang et al. [11] proposed a chaos synchronization (CS)-based detector for PQ disturbance classification, and realized a CS-based detector to track the dynamic errors from the fundamental and distorted signals. Artificial neural networks were designed to classify different types of disturbances, including multilayer neural networks and wavelet neural networks [12, 13].

Based on the above survey, although there is still no research on how to identify the operating conditions, the development of a PQ data acquisition system [14-15] and previous studies provided the motivation to carry out the research study reported in this paper. In this study, we established a pattern recognition approach and according to the different characteristics of the monitored waveforms, identified the operating conditions of a HSR. The identification of HSR operating conditions which cause the PQ anomalies has at least three aspects of the application: 1) for the railway authorities, the most concern is HSR security operation, and it does not care about PQ so much. Through the research in this paper, it can provide basis for explicating PQ severity of different HSR operating conditions. Then, it is beneficial for optimizing the operation of HSRs at the rush hour, and mobile PQ improvement of HSR to reduce adverse impact of this type of PQ disturbance source; 2) for the power supply company,

[†] Corresponding Author: School of Electrical Engineering, Southeast University, China. (wgu@seu.edu.cn)

* School of Electrical Engineering, Southeast University, China. (765099360@qq.com, jingjing_bai@163.com)

** Jiangsu Electrical Power Company Research Institute, China. (lannyuan@hotmail.com, cbsure@163.com)

Received: February 6, 2014; Accepted: October 1, 2015

traction substation is a link with power suppliers and users. Therefore, based on probabilistic operating conditions of the accessed HSR, targeted PQ control measures can be taken to traction substation, in order to reduce PQ pollutions and improve the operation and management of the power grid; 3) from the perspective of the long-term planning and evaluation, this study can provide technical references for secure and stable operation of the existed traction substations, the designing and planning of a new traction substation and transforming HSR to adjust its own PQ characteristics.

This paper is organized as follows. The characteristics of the high-speed electrified railway load are presented in Section 2. The conditions recognition approach model and approach flow are presented in Section 3. Numerical results and discussions are given in Section 4. In Section 5, conclusions are drawn.

2. Characteristics of HSR Load

Electric railway traction loads are characterized by changing (in time and location) large amplitude variations, which are typical of the daily fluctuating load and have heavy duty, nonlinear short-term impact characteristics [1, 8, 16]. The reasons for these daily fluctuations include the conditions of the transportation routes, locomotive types, operation modes and speeds of locomotives, traction and weight, all of which are random influencing factors. Thus, an electric railway traction load is not typical of a continuous power load. Low power factor, large negative-sequence power and abundant harmonics are the typical PQ problems of HSR [6, 7, 17]. With the increase in HSR, these problems will become more complex with increasing influences on the grid. In addition, with the development of high-tech industries and the improvement of automation technology in equipment, users are demanding higher PQ requirements [18-19]. Therefore, it is important to understand the PQ problems caused by HSR.

The traction power supply system for HSR typically includes external power supplies, traction substations, traction networks and high-speed EMUs [20, 21]. A sample model of it is shown in Fig. 1. As a special high-power

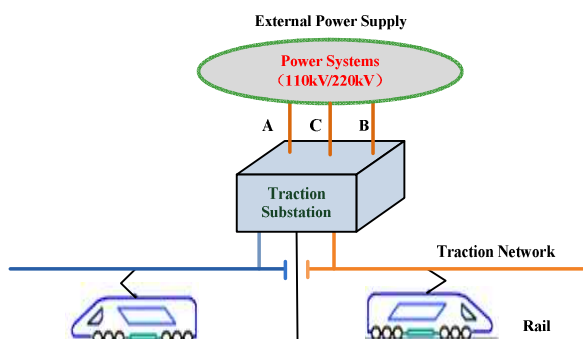


Fig. 1. Sample model of HSR traction power supply system.

and single-phase load, HSR has acceleration, uphill and regenerative braking conditions and the characteristics of a HSR are as follows.

2.1 Frequent fluctuation

The operating conditions of a high-speed train during running are vulnerable to random factors, such as track gradients, bends and climatic conditions. In addition, the rail passenger traffic flows are distributed in different areas and at different times, and the number of high-speed trains in the traction substation contact-network and the load conditions of each train are always changing. Therefore, the traction load has the features of large and frequent fluctuations.

2.2 Large traction power and high reliability requirements

We all know that air resistance increases as a geometric series with speed, and a train has to overcome air resistance to run, so because of the high speed, HSR needs larger traction power, which is about 3 to 5 times the traction load power of conventional rail. With the development of HSR and transportation needs, the speeds and traction power of trains will be further improved. Furthermore, the reliability requirements of the HSR operation will be greater.

2.3 Nonlinearity and asymmetry

Most HSR is from the CRH series and uses an 'AC-DC-AC' type transmission system, which can transform the industrial-frequency alternating current into power suitable for a traction motor. This nonlinear process produces large quantities of harmonic currents. In addition, HSR uses a single-phase power supply. The loads of the two power supply arms are not fully consistent, and for the three-phase grid, it is a type of asymmetric load, which will produce negative-sequence current and results in three-phase unbalance occurring in the power systems.

3. Pattern Recognition Approach for Identifying the Operation Conditions of HSR

From the above analysis, it can be seen that the traction load of HSR has heavy duty, nonlinear and short-term impact characteristics, which will adversely affect the PQ of the grid, bringing hidden dangers to its safe and stable operation. Therefore, to achieve optimal operation and PQ control of HSR, it is necessary to understand the characteristics of the operating conditions, particularly those that cause PQ anomalies during actual operation. Based on a PQ monitoring platform, this paper presents a pattern recognition approach for HSR. First, it uses PQ monitoring equipment to acquire the PQ waveform data

from typical operating conditions and establish a typical condition waveform library (TCWL) for HSR, based on the principal component analysis (PCA) algorithm. And based on the PCA algorithm and a dynamic time wrapping (DTW) algorithm, it identifies the actual operating conditions of HSR through comparative analysis of the monitored characteristic waveforms and reference characteristic waveforms. Fundamentally, to identify the operating conditions of HSR, the pattern recognition approach contains three parts: the TCWL for HSR, PCA algorithm and DTW algorithm. The TCWL for HSR is the Reference Waveform Database of its typical operating conditions, which is the basis of the whole approach. The PCA algorithm is used to reduce the dimensions of monitored waveforms data and acquire the characteristic waveforms. To reduce the redundancy and correlation between the monitored waveforms data, the PCA algorithm is applied to convert three-phase PQ data to two-phase or single-phase data. The DTW algorithm is used to analyze the similarity between the different characteristic waveforms, which in turn identifies the real-time operating conditions of the HSR.

3.1 The establishment of a typical condition waveform library

The pattern recognition approach for the HSR relies on the TCWL, which provides the data supporting environment for this approach. As a reference waveform database of the typical operating conditions of the HSR, the TCWL includes the monitored waveform data of the typical operating conditions, the PQ data (negative sequence, harmonics, unbalance factor, etc.), the reference characteristic waveforms after data dimensionality reduction, and their standardized waveforms. Establishing the typical condition waveform library requires the following steps and a corresponding flowchart is shown in Fig. 2.

3.1.1 Acquisition of data on typical operating conditions

Under certain conditions, the PQ monitor is used to

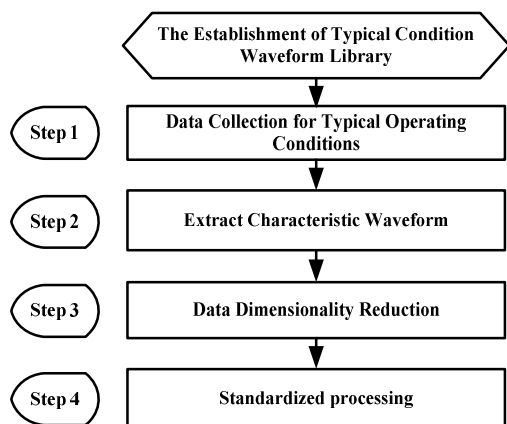


Fig. 2. Flowchart of the establishment of TCWL

acquire the data bands (also called the recorded data) of the HSR typical operating conditions from the traction substation side. This includes the regenerative braking, neutral-section passing, uphill and acceleration conditions.

3.1.2 Extract reference characteristic waveform

- a) Select a PQ indicator which can represent the data bands of the HSR typical operating conditions as the characteristic parameter, such as the fundamental harmonic, harmonics (2~50th), negative sequence current or RMS current;
- b) Extract the data waveforms of the characteristic parameter as the reference characteristic waveforms.

3.1.3 Data dimensionality reduction

Based on Step 2), use the PCA algorithm to convert the three-phase reference characteristic waveform data to single-phase data, called the Reference Characteristic Time Sequence (RCTS) in this study.

3.1.4 Standardized processing

Use Formula (9) to standardize the Reference Characteristic Time Sequence and acquire the standardized waveforms for different operating conditions.

3.2 Principal component analysis algorithm

The PCA algorithm is used as a linear mapping method in this study, that is, each new characteristic variable is a linear function of the original characteristic variables after mapping [22]. After linear conversion, the number of new characteristic variables is less than the original number of characteristic variables, as well as retaining the main information of the original characteristic variables.

Initially, the characteristic variables are based on the original sample data: $X = (x_1, x_2, \dots, x_n)^T$, where x is called the original characteristic variable. The PCA algorithm is then used to construct the new sample data: $Y = (y_1, y_2, \dots, y_n)^T$, where y_i is called the new characteristic variable. The steps are as follows:

- 1) Based on the PCA algorithm, calculate the covariance matrix S_X of the original sample data X :

$$S_x = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1n} \\ S_{21} & S_{22} & \dots & S_{2n} \\ \dots & \dots & \dots & \dots \\ S_{n1} & S_{n2} & \dots & S_{nn} \end{bmatrix}. \quad (1)$$

- 2) Calculate the characteristic variables of S_X : $\lambda_1, \lambda_2, \dots, \lambda_n$ and their corresponding eigenvectors: u_1, u_2, \dots, u_n . Place the characteristic variables in descending order:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n. \tag{2}$$

- 3) Define the Cumulative Variance Contribution Rate (CVCR) of the new characteristic variable y_i (also referred to as the i -th principal component) :

$$\lambda_i / (\lambda_1 + \lambda_2 + \dots + \lambda_n). \tag{3}$$

- 4) Calculate the CVCR of the first m principal component y_1, y_2, \dots, y_m :

$$(\lambda_1 + \lambda_2 + \dots + \lambda_m) / (\lambda_1 + \lambda_2 + \dots + \lambda_n). \tag{4}$$

If the CVCR is large enough, such as 80% or larger, it takes the first m principal components as the new main characteristic variables, and discards the remaining components. The relationship between the new main characteristic variables and the original characteristic variables is shown in Formula (5):

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_m \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & u_{22} & \dots & u_{2n} \\ \dots & \dots & \dots & \dots \\ u_{m1} & u_{m2} & \dots & u_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_m \end{bmatrix}, \tag{5}$$

where Y should satisfy the following three conditions:

- a) The new characteristic variables are derived from the original characteristic variables through a linear combination, that is:

$$y_i = u_{i1}x_1 + u_{i2}x_2 + \dots + u_{in}x_n, \tag{6}$$

where, $i = 1, 2, \dots, n$ and u_{ij} is a constant.

- b) The characteristic variables of Y are unrelated, that is:

$$r(y_i, y_j) = 0, \quad i \neq j, \tag{7}$$

where, $i, j = 1, 2, \dots, n$; $r(y_i, y_j)$ is the coefficient of correlation between the characteristic variables: y_i and y_j .

- c) The constant coefficient u_1 maximizes the variance of y_1 and the constant coefficient u_2 makes the variance of y_2 sub-maximum, and so on.

To further reduce the dimensions of the new sample data, for the calculated new characteristic variables, this study made a further amendment. When the CVCR is large enough and meets the requirement, if the number of the new characteristic variable is only 1 after data dimensionality reduction, $Y^1 = y_i$ is the only main characteristic variable. If the number of the new characteristic variable is more than 1, we use the Formula (8) to obtain the main characteristic variable:

$$Y^1 = y_1\lambda'_1 + y_2\lambda'_2 + \dots + y_m\lambda'_m, \tag{8}$$

where, $\lambda'_i = \lambda_i / (\lambda_1 + \lambda_2 + \dots + \lambda_m)$.

3.3 Dynamic time wrapping algorithm

Because the sequence distribution of the operating conditions of the HSR is not consistent over the timeline, even when they are in the same operating conditions, the feature information of the operating condition may be different because of the difference in the operating environment of the traction load. To reduce the difference in the feature information, it is necessary to standardize the characteristic waveform data.

The so-called standardizing treatment is usually known as normal distribution processing, meaning that it should transform the sample data into standard sample data whose mathematical expectation is zero and variance is one. In this study, we used Formula (9) to standardize the characteristic waveform data of the HSR operating conditions.

$$Y^{1*} = \frac{Y^1 - M}{S}, \tag{9}$$

where, Y^{1*} are the standardized sample data; Y^1 is the original sample data; M is the mean value, and S is the standard deviation of the original sample data.

After standardization, the sample data of the characteristic waveform itself provides the distribution on the timeline. Therefore, it can be used as a time sequence for the DTW algorithm.

The DTW algorithm uses dynamic programming principles to search for the minimum bending path between two time sequences [23, 24]. In the DTW algorithm, the points on the time sequences no longer satisfy a one-to-one relationship. To find the minimum bending path between two time sequences, it needs to meet the path W [25], which is shown in Fig. 3, where T and R represent the two different time sequences.

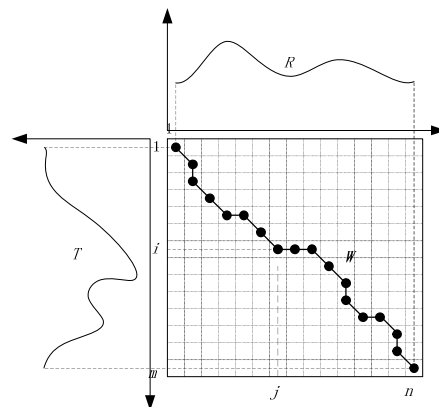


Fig. 3. Schematic diagram of the dynamic bending path.

The implementation of the DTW algorithm is as follows: Take the two different time sequences T and R whose data dimension are m and n respectively: $T = (t_1, t_2, \dots, t_m)^T$, $R = (r_1, r_2, \dots, r_n)^T$.

- 1) Where, T is the test time sequence, and $T = Y^{1*}$; R is the reference time sequence, which are the known characteristic waveform data.
- 2) Sort T and R in accordance with the corresponding time sequence position, and then, construct the distance matrix $A_{m \times n}$ of T and R :

$$A_{m \times n} = \begin{bmatrix} d(t_1, r_1) & d(t_1, r_2) & \dots & d(t_1, r_n) \\ d(t_2, r_1) & d(t_2, r_2) & \dots & d(t_2, r_{n-1}) \\ \vdots & \vdots & \ddots & \vdots \\ d(t_m, r_1) & d(t_m, r_2) & \dots & d(t_m, r_n) \end{bmatrix}, \quad (10)$$

where the element $a_{ij} = d(t_i, r_j) = \sqrt{(t_i - r_j)^2}$ in matrix A represents the Euclidean distance between the time sequence points t_i and r_j .

- 3) For the distance matrix $A_{m \times n}$, the set consisting of its adjacent elements is called the DTW bending path, which is defined as $W = (w_1, w_2, \dots, w_m)$, where, $w_k = (a_{ij})_k$, $k=1, 2, \dots, m$, and W satisfy the following conditions:
 - a) Boundedness: $\max\{m, n\} \leq k \leq m+n-1$;
 - b) Critical Condition: $w_1 = a_{11}$, $w_m = a_{mn}$;
 - c) Continuity and Monotonicity: for $w_k = a_{ij}$, $w_{k-1} = a_{i',j'}$, they should satisfy $0 \leq i-i' \leq 1$, $0 \leq j-j' \leq 1$.
- 4) Calculate the minimum bending path between T and R , which is shown as Formula (11):

$$\mathcal{D}_{DTW}(T, R) = \min\left(\frac{1}{K} \sqrt{\sum_{k=1}^K w_k}\right)$$

$$= \begin{cases} D(<, >, <) = 0, \\ D(i, <) = D(<, j) = \infty, \\ D(1, 1) = a_{11}, \\ D(i, j) = a_{ij} + \min\{D(i-1, j-1), D(i, j-1), D(i-1, j)\} \end{cases} \quad (11)$$

where, $i = 2, 3, \dots, m$; $j = 2, 3, \dots, n$; and $D(m, n)$ is the accumulated value of the minimum bending path.

In fact, the matrix that the DTW algorithm constructed is the ‘‘cumulative distance matrix’’ of the time sequences T and R . At the same time, the smaller $D(m, n)$, the greater the similarity between the two time sequences. To meet the objectives of this study, $D(m, n)$ is expressed as a similarity, which is given by Formula (12):

$$s(T, R) = \frac{10}{10 + D(m, n)} \times 100\%, \quad (12)$$

where, $s(T, R)$ represents the similarity between the time sequences T and R .

Formula (12) shows that $s(T, R)$ is in the range [0, 100 %], and the larger $s(T, R)$, the smaller $D(m, n)$.

3.4 Implementation of the proposed approach

This study uses the HSR as an example combined with the typical HSR operating conditions. The flowchart of the proposed approach is shown as Fig. 4.

The implementing process of the proposed approach is given as follows:

- 1) Use the PQ monitoring equipment to acquire the waveform data from the traction substation side (also called system side in this study) during HSR operation. Extract the abnormal band data of the real-time monitored waveforms which are called the test waveform data in this study.
- 2) Extract the test characteristic waveform
 - a) Select a PQ indicator, which can represent the abnormal bands, as the characteristic parameter, such as the fundamental harmonic, harmonics (2~50th), negative sequence current, or current RMS.
 - b) Extract the data waveforms characteristic parameter as the test characteristic waveforms.
- 3) Use the PCA algorithm to convert the three-phase PQ data to single-phase data. The specific steps are as follows:
 - a) Take the characteristic waveform data of the

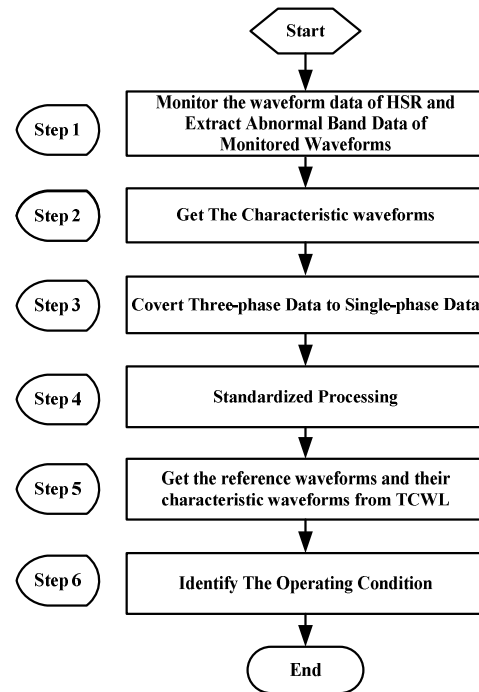


Fig. 4. Flowchart of the proposed approach.

A/B/C phases as the test waveform data, and take the value at each time as a characteristic variable, the test waveform data can then be regarded as a set of several characteristic variables.

- b) Based on the principle of the PCA algorithm, calculate the covariance matrix of the test waveform data.
- c) Calculate all the characteristic variables and the corresponding eigenvectors of the test waveform data, and arrange the characteristic variables in descending order.
- d) Calculate the CVCR and analyze the number of the new characteristic variables after the data dimensionality reduction.

If the number of new characteristic variables is greater than one, it should use the Formula (8) to obtain the main characteristic variable. We can then obtain the “test characteristic time sequence (TCTS)” and its data dimension is 1.

- 4) Standardized Processing: Use Formula (9) to standardize the TCTS, and obtain their standardized characteristic time sequences.
- 5) Obtain the reference waveforms and their characteristic waveforms which are called the “reference characteristic time sequence (RCTS)” from the TCWL of the HSR;
- 6) Identify the operating conditions of the real-time monitored waveforms. Using the DTW algorithm to calculate the dynamic bending distance between TCTS and RCTS, and then convert the dynamic bending distance into a similarity value, and the conditions of the RCTS corresponding to the maximum similarity are the operating conditions of the real-time monitored waveforms. The specific steps are as follows:
 - a) Calculate the Euclidean Distance between TCTS and RCTS, and then build their corresponding cumulative distance matrix;
 - b) According to Formula (11), use dynamic programming principles to calculate the minimum bend distance between TCTS and RCTS.

4. Case Studies

To verify the effectiveness of the above proposed approach, this study takes three segments of monitored waveform data from the HSR traction substation of the Beijing-to-Shanghai railway as test samples, called Test Sample 1, Test Sample 2 and Test Sample 3. The transformer of this traction substation uses three-phase V/V

wiring and the short-circuit capacity is 18000 MVA. The case studies for the HSR operating conditions are all based on a single condition, and they do not consider factors such as more than one condition occurring at the same time, e.g. a tunnel and a bend.

Because of the dramatic changes, strong randomness and short duration of frequent switching conditions of the HSR traction load, we need to use high-sampling-frequency PQ monitoring equipment to acquire its PQ data, typically at least 10 kHz to capture the information of waveform variation during the switching process conditions.

4.1 Reference waveforms in the TCWL of the HSR

Before obtaining the pattern recognition for the HSR, it is necessary to provide the reference monitored waveforms in the TCWL of the HSR, in addition to their characteristic waveforms that the pattern recognition process needs. All the reference data are monitored from the traction substation of Beijing-to-Shanghai railway as test samples.

Fig. 5 shows the reference monitored waveforms for typical operating conditions of the HSR which are called reference waveforms in this study. The reference characteristic waveforms for different operating conditions, which have been standardized, are shown in Fig. 6, where the horizontal coordinates represent the data after standardization. The CVCR for different operating conditions are shown in Table 1. The conditions are regenerative braking, neutral-section passing, uphill and acceleration conditions. We observe from Table 1 that, for the reference waveforms of all the operating conditions in TCWL, the CVCR is at least 99 %, that is, a single new characteristic variable retained the most information of the original characteristic variables.

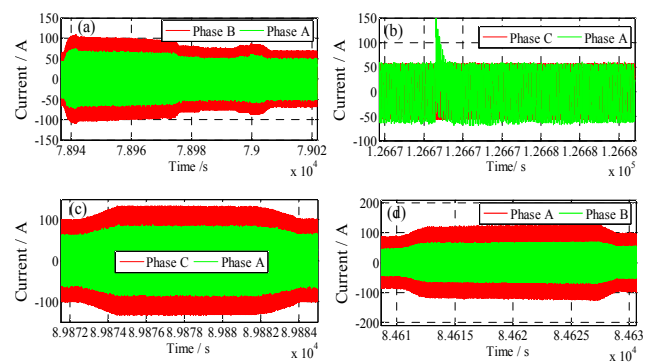


Fig. 5. Reference monitored waveforms for typical operating conditions in the TCWL of the HSR: (a) Regenerative braking; (b) Neutral-section passing; (c) Uphill; (d) Acceleration.

Table 1. Cvr of characteristic variable

	Acceleration condition	Uphill condition	Regenerative braking condition	Neutral-section passing condition	Test Sample 1	Test Sample 2	Test Sample 3
CVCR	99.95%	99.92%	99%	99.42%	99.52%	99.51%	99.63%

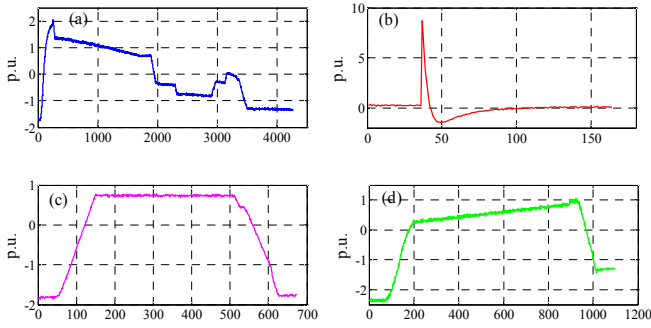


Fig. 6. The characteristic waveforms for the reference sample under conditions of (a) Regenerative braking; (b) Neutral-section passing; (c) Uphill; (d) Acceleration.

4.2 Data dimensionality reduction of the test samples

Of the electrical quantities of a HSR traction load, the most dramatic change is evident in the three-phase current on the system side, where its waveform morphology is obviously different for different conditions. In addition, it can reflect the distribution of harmonics and unbalance of negative-sequence voltages/currents. Therefore, the RMS current is selected as the characteristic quantity to identify the operating conditions of the HSR.

Taking Test Sample 1 as an example, Fig. 7 shows the three-phase waveform of the RMS current and its corresponding characteristic waveform. In this example, when extracting the characteristic variable, its CVCR is 99.52 %. In other words, after data dimensionality reduction, the extracted characteristic waveform basically retains

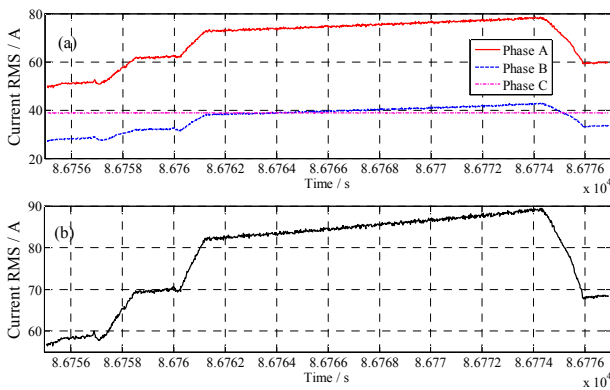


Fig. 7. Data dimensionality reduction for Test Sample 1: (a) Waveforms of three-phase current RMS of Test Sample 1; (b) The characteristic waveform after data dimensionality reduction of Test Sample 1.

Table 2. Similarity Results of Different Test Samples

Condition mode	Acceleration condition	Uphill condition	Regenerative braking condition	Neutral-section passing condition
Test sample				
Test Sample 1	78.0940%	61.4360%	39.3920%	26.4460%
Test Sample 2	24.3152%	28.1313%	27.0862%	80.7912%
Test Sample 3	60.9751%	84.3858%	30.8192%	25.1879%

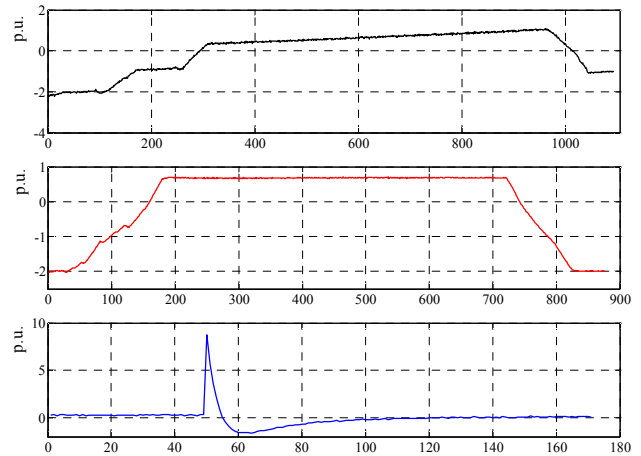


Fig. 8. The characteristic waveforms of the test samples: (a) Characteristic waveform of Test Sample 1; (b) Characteristic waveform of Test Sample 2; (c) Characteristic waveform of Test Sample 3.

the information contained in the original characteristic waveform.

The CVCRs of the test samples are shown in Table 1. and their CVCRs are at least 99 %. In addition, because all the extracted characteristic variables of the test samples are three-phase RMS currents that have the same dimension, they do not need dimensionless processing before PCA analysis.

We observe from Fig. 7 that the characteristic waveform after data dimensionality reduction is very similar to the original waveforms of the A/B phase, and the characteristic waveform not only retains most of the information of the original characteristic waveforms, but also strengthens their turning points. Therefore, we conclude that the proposed PCA algorithm is feasible and effective in handling data dimensionality reduction and the extraction of the characteristic variables. Fig. 8 shows the characteristic waveforms of the test sample, which have been standardized for different conditions.

4.3 Pattern recognition of test samples

After standardization, the characteristic variables of the reference and test samples are at the same standard, so we can use the DTW algorithm to calculate the similarity between the test samples and the reference samples.

The results are shown in Table 2.

We observe from Table 2 that Test Sample 1 has the maximum similarity with the reference sample for the acceleration condition, Test Sample 2 has the maximum

similarity with the reference sample for the neutral-section passing condition, and Test Sample 3 has the maximum similarity with the reference sample for the uphill condition. The operating condition for Test Sample 1 is similar to the acceleration condition, the operating condition for Test Sample 2 is similar to the uphill condition, and the operating condition for Test Sample 3 is similar to the neutral-section passing condition. Therefore, we conclude that the operating condition for Test Sample 1 is the acceleration condition, the operating condition for Test Sample 2 is the neutral-section passing condition, and the operating condition for Test Sample 2 is the uphill condition.

Figs. 9, 10 and 11 show the monitored waveforms of the test sample and the corresponding monitored waveforms of the reference sample. We observe from Figs. 8, 9 and 10 that the waveform trends of the three test samples are very similar to their corresponding reference sample waveforms, which further demonstrated and verified the accuracy and reliability of the pattern recognition algorithm proposed in this study.

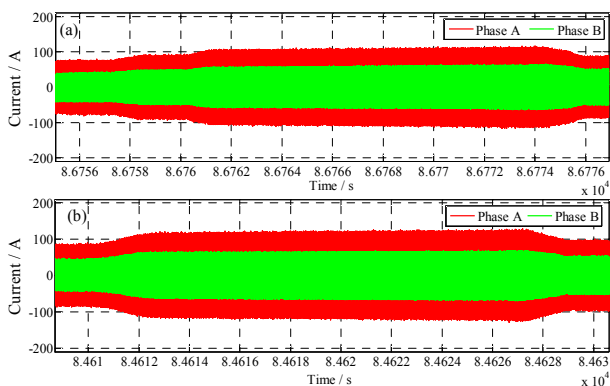


Fig. 9. Waveforms for Test Sample 1 and the corresponding reference sample: (a) Waveform for Test Sample 1; (b) Reference waveform for the acceleration condition.

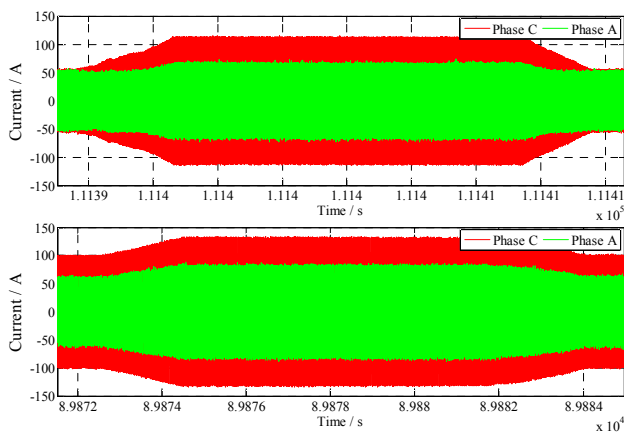


Fig. 10. Waveforms for Test Sample 2 and the corresponding reference sample: (a) Waveform for Test Sample 2; (b) Reference waveform for the uphill condition.

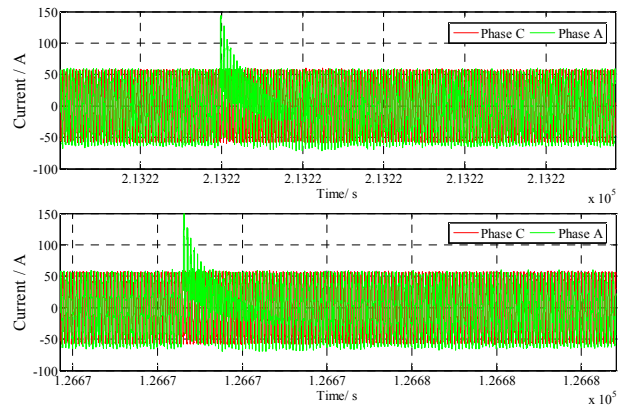


Fig. 11. Waveforms for Test Sample 3 and the corresponding reference sample: (a) Waveform for Test Sample 3; (b) Reference waveform for the neutral-section passing condition.

Based on the above analysis, it can be found that the main advantages of the proposed algorithm are as follows: 1) accuracy: by establishing the TCWL and analyzing pattern matching, the proposed method can accurately identify HSR operating conditions that cause PQ anomalies. The recognition results are reliable and can be applied to PQ control; 2) simplicity: from analysis steps, required inputs and computation process are not complex at all, and computation speed is fast. All in all, time complexity and space complexity of the proposed method are in acceptable ranges and have many advantages over general recognition methods; 3) applicability: the proposed approach and flow can not only be applied for recognition of HSR operating conditions to carry out PQ analysis and control, but also be suitable for other PQ disturbance sources.

5. Conclusions

Based on the monitored waveforms of the HSR, a pattern recognition approach for the HSR operating conditions is first proposed in this study. Based on the analysis of the PQ characteristic waveforms under different HSR operating conditions, we performed Condition Matching Analysis for the real-time monitored waveforms. To verify the accuracy and effectiveness of the proposed algorithm, we compared and analyzed the characteristics and trends of the reference sample and test sample waveforms.

This study has realized new investigation techniques in the field of PQ analysis. The proposed approach and the results of the pattern recognition can be used to better understanding the corresponding relation between PQ anomalies and HSR operating conditions, so as to provide reference information for reducing PQ pollutions, suggestions for optimizing HSR operations, long-term planning of the power suppliers and the railway authorities, and so on. Besides, this paper can lay a foundation for the

analysis and control of other PQ disturbance sources, to make the power grid become more safe and reliable.

Acknowledgements

This work was supported in part by the National Science Foundation of China (Grant No. 51277027) and the State Grid Corporation of China (Voltage sag).

References

- [1] A. Bueno, J. M. Aller, J. A. Restrepo, et al, "Harmonic and Unbalance Compensation Based on Direct Power Control for Electric Railway Systems," *IEEE Trans on Power Electron*, Vol. 28, No. 12, pp. 5823-5831, Dec. 2013.
- [2] C. Wu, A. Luo, J. Shen, F. J. Ma, et al, "A negative sequence compensation method based on a two-phase three-wire converter for a HSR way traction power supply system," *IEEE Trans on Power Electron*, Vol. 27, No. 2, pp. 706-717, Feb. 2012.
- [3] G. W. Chang, H. W. Lin, S. K. Chen, "Modeling characteristics of harmonic currents generated by HSR way traction drive converters," *IEEE Trans on Power Del*, Vol. 19, No. 2, pp. 766-733, Apr. 2004.
- [4] H. Q. Wang, Y. J. Tian, Q. C. Gui, "Evaluation of negative-sequence current injecting into the public grid from different traction substation in electrical railways," in *Proceeding of 20th Int. Conf. Exhib. Electr. Distrib*, pp. 1-4, 2009.
- [5] M. Brenna, F. Foiadelli, D. Zaninelli, "Electromagnetic Model of High Speed Railway Lines for Power Quality Studies," *IEEE Trans on Power Syst*, Vol. 25, No. 3, pp. 1301-1308, Aug. 2010.
- [6] A. Capasso, G. Ghilardi, G. G. Buffarini, "Bologna-Florence High Speed Railway Line: MV Emergency Traction Power Supply, Operating Conditions and PQ Issues," in *Proceeding of IEEE_ESARS*, pp.1-5, 2010.
- [7] F. Ciccarelli, M. Fantauzzi, D. Lauria, "Special Transformers Arrangement for AC Railway Systems," in *Proceeding of IEEE_ESARS*, pp.1-6, 2012.
- [8] E. Cinieri, A. Fumi, V. Salvatori, et al, "A New High-Speed Digital Relay Protection of the 3-kVdc Electric Railway Lines," *IEEE Trans on Power Del*, Vol. 22, No. 4, pp. 2262-2270, Oct. 2007.
- [9] S. Mishra, C. N. Bhende, B. K. Panigrahi, "Detection and Classification of Power Quality Disturbances Using S-Transform and Probabilistic Neural Network," *IEEE Trans on Power Del*, Vol. 23, No. 1, pp. 280-287, Jan 2008.
- [10] N. H. Nguyen, D. C. Lee, "Control Strategy for Three-Phase Grid-Connected Converters under Unbalanced and Distorted Grid Voltages Using Composite Observers," *Journal of Power Electronics*, Vol. 13, No. 3, pp. 469-478, May. 2013.
- [11] C. H. Huang, C. H. Lin, C. L. Kuo, "Chaos Synchronization-Based Detector for Power-Quality Disturbances Classification in a Power System," *IEEE Trans on Power Del*, Vol. 26, No. 2, pp.944-953, April. 2011.
- [12] C. H. Lin, C. H. Wang, "Adaptive wavelet networks for power quality detection and discrimination in a power system," *IEEE Trans on Power Del.*, Vol. 21, No. 3, pp. 1106-1113, Jul. 2006.
- [13] W. M. Lin, C. H. Wu, C. H. Lin, et al, "Detection and classification of multiple power quality disturbances with wavelet multiclass SVM," *IEEE Trans on Power Del*, Vol. 23, No. 4, pp. 2575-2582, Oct. 2008.
- [14] M. Simic, D. Denic, D. Zivanovic, et al, "Development of a Data Acquisition System for the Testing and Verification of Electrical Power Quality Meters," *Journal of Power Electronics*, Vol. 12, No. 5, pp. 813-820, Sep. 2012.
- [15] K. Yingkayun, S. Premrudeepreechacharn, "A Power Quality Monitoring System for Real-Time Detection of Power Fluctuations," in *Proc. IEEE 40th North American Power Symposium*, pp. 53-57, 2008.
- [16] S. L. Chen, R. J. Li, P. H. Hsi, "Traction system unbalance problem analysis methodologies," *IEEE Trans on Power Del*, Vol. 19, No. 4, pp. 1877-1883, Oct 2004.
- [17] Z. L. Shu, S. F. Xie, Q. Z. Li, "Single-Phase Back-To-Back Converter for Active Power Balancing, Reactive Power Compensation, and Harmonic Filtering in Traction Power System," *IEEE Trans on Power Elec*, Vol. 26, No. 2, pp. 334-343, Feb. 2011.
- [18] P. Moallem, A. Zargari, A. Kiyoumars, "Improvement in Computation of Delta V-10 Flicker Severity Index Using Intelligent Methods," *Journal of Power Electronics*, Vol. 11, No. 2, pp. 228-236, May. 2011.
- [19] S. A. Yin, C. L. Su, R. F. Chang, "Assessment of power quality cost for high-tech industry," in *Proc. IEEE Power India Conference*, pp. 927-932, 2006.
- [20] S. L. Chen, R. J. Li, P. H. Hsi, "Traction system unbalance problem analysis methodologies," *IEEE Trans. Power Del*, Vol. 19, No. 4, pp. 1877-1883, Oct. 2004.
- [21] Z. W. Zhang, B. Wu, J. S. Kang, et al, "A multi-purpose balanced transformer for railway traction applications," *IEEE Trans on Power Del*, Vol. 24, No. 2, pp. 711-718, Apr. 2009.
- [22] I. T. Jolliffe, "Principal Component Analysis," Springer, 2002, Chap.3.
- [23] D. J. Berndt, J. Clifford, "Finding patterns in time series: a dynamic programming approach," *AAAI/MIT Press*, 1996, pp. 229-248.
- [24] C. Myers, L. R. Rabiner, A. E. Rosenberg, "Performance tradeoffs in dynamic time warping algorithms

for isolated word recognition,” *IEEE Trans on Acoustics, Speech, and Signal Processing*, Vol. 28, No. 6, pp. 623-635, Dec. 1980.

- [25] L. R. Rabiner, A. E. Rosenberg, S. E. Levinson, “Considerations in dynamic time warping algorithms for discrete word recognition,” *IEEE Trans on Acoustics, Speech, and Signal Processing*, Vol. 26, No. 6, pp. 575-582, Dec. 1978.



Jingjing Bai received the B.Eng. degree in Electrical Engineering from Jiangsu University, China in 2012. He is currently pursuing the M.Eng degree in Electrical Engineering at Southeast University. His research interest is power quality.



Wei Gu received his B.Eng degree and Ph.D. degree in Electrical Engineering from Southeast University, China, in 2001 and 2006. From 2009 to 2010, he was a Visiting Scholar in the Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287, USA. He is now a professor in the School of Electrical Engineering, Southeast University. His research interests are power system stability and control, smart grid, renewable energy technology and power quality.



Shuai Zhang received the B.Eng. degree in Electrical Engineering from Southeast University, China in 2011. He is currently pursuing the M.Eng degree in Electrical Engineering at Southeast University. His research interest is power quality.



Xiaodong Yuan received the B.S. and M.S. degrees from Southeast University, China, in 2002 and 2005, respectively. Currently he is a director with new energy and distribution network laboratory, Jiangsu Electrical Power Company Research Institute, China. He is now taking the responsibility of smart distribution power system, renewable energy technology and power quality.



Bing Chen received the Ph.D. degree from China. Currently he is a senior engineer with new energy and distribution network laboratory, Jiangsu Electrical Power Company Research Institute, China. He is now taking the responsibility of smart distribution power system, renewable energy technology and power quality.