

Classification of Power Quality Events Using Optimal Time-Frequency Representations

– Part 1: Theory

Min Wang, *Student Member, IEEE*, Alexander V. Mamishev, *Member, IEEE*

Abstract—Better software and hardware for automatic classification of power quality (PQ) disturbances are desired for both utilities and commercial customers. Existing automatic recognition methods need improvement in terms of their capability, reliability, and accuracy. This paper presents the theoretical foundation of a new method for classifying voltage and current waveform events that are related to a variety of PQ problems. The method is composed of two sequential processes: feature extraction and classification. The proposed feature extraction tool, time-frequency ambiguity plane with kernel techniques, is new to the power engineering field. The essence of the feature extraction is to project a PQ signal onto a low-dimension time-frequency representation (TFR), which is deliberately designed for maximizing the separability between classes. The technique of designing an optimized TFR from time-frequency ambiguity plane is for the first time applied to the PQ classification problem. A distinct TFR is designed for each class. The classifiers include a Heaviside-function linear classifier and neural networks with feedforward structures. The flexibility of this method allows classification of a very broad range of power quality events. The performance validation and hardware implementation of the proposed method are presented in the second part of this two-paper series [1].

Index Terms—Power Quality, Classification-Optimal TFR, Time-Frequency Ambiguity Plane, Fisher’s Discriminant Function, Linear Classifier, Neural Networks.

I. INTRODUCTION

THE proliferation of highly sensitive computerized equipment places increasingly more stringent demands on the quality of electric power supplied to the customer [2]. Today, power quality (PQ) has become a very interesting cross-disciplinary topic, coupling power engineering and power electronics with digital signal processing, software engineering, networking, and VLSI.

Voltage disturbances are the most frequent cause of a broad

range of disruption in industrial and commercial power supply systems. These disturbances, often referred to as power quality problems, significantly affect many industries. Major causes of PQ-related revenue losses are interrupted manufacturing processes and computer network downtime. The examples abound in semiconductor industry, chemical industry, automobile industry, paper manufacturing, and e-commerce. A report by CEIDS (Consortium for Electric Infrastructure to Support a Digital Society) shows that the U.S. economy is losing between \$104 billion and \$164 billion a year due to outages and another \$15 billion to \$24 billion due to PQ phenomena [3].

The conventional methods currently used by utilities for power quality monitoring are primarily based on visual inspection of voltage and current waveforms. Highly automated monitoring software and hardware is needed in order to provide adequate coverage of the entire system, understand the causes of these disturbances, resolve existing problems, and predicting future problems. Several research institutions and companies are conducting PQ research and development. Software and hardware products, such as PQSoft by Electrotek Concepts and PQNode by Dranetz-BMI have demonstrated benefits of the PQ monitoring methodology to utility companies. Signature System, a web-based power information system recently developed by Dranetz-BMI, Electrotek Concepts, EPRI, and TVA, integrates a network of monitoring software and hardware. It incorporates many internet-related features and motivates multiple future directions in PQ monitoring.

Built-in event diagnosis and assessment modules are the key to these monitoring systems. One way to improve the performance of a monitoring system is to incorporate a more reliable and accurate waveform recognition algorithm, which recognizes a broader range of PQ events. The traditional method is based on RMS measurements and constrained by its accuracy. Approaches for automated detection and classification of PQ disturbances proposed recently are based on wavelet analysis, artificial neural networks, hidden Markov models, and bispectra [4-8]. These techniques have been successfully employed in other pattern recognition and signal processing applications, such as speech recognition, audio processing, communications, and radar and sonar applications. While PQ waveform recognition processes voltage and current signals and speech recognition processes audio

This work is supported by the Advanced Power Technologies (APT) Center at the University of Washington. The APT Center is supported by ALSTOM ESCA, LG Industrial Systems, RTE France, and Mitsubishi Electric Corp. This work is also partially supported by the National Science Foundation Career Award Grant #0093716 and the American Public Power Association DEED program.

Both authors are with the SEAL (Sensors, Energy, and Automation Laboratory), Department of Electrical Engineering, Box 352500, University of Washington, Seattle, WA98195. (E-mails: mwang@ee.washington.edu; mamishev@ee.washington.edu)

signals, these two problems are very similar in nature. Recent development in the speech processing area has shown us the potential of designing more accurate and more intelligent PQ monitoring algorithms, with advanced signal processing techniques. The potential improvement on monitoring capability implies significant economic values to utilities and PQ-sensitive energy customers.

In this paper, the technique of designing an optimized time-frequency representation (TFR) from a time-frequency ambiguity plane is applied to the PQ classification problem for the first time. Because the proposed method is entirely new to the power engineering field and many new concepts are introduced, we present the material with a two-paper series: Part 1 -- Theory (i.e., this paper) and Part 2 -- Application [1]. A significant classification application with this algorithm and its DSP-based real-time hardware implementation are presented in Part 2, to validate the presented method.

II. OVERVIEW OF PQ CLASSIFICATION ALGORITHMS

Although recently proposed PQ classification algorithms provide improved recognition rates, they are not yet sufficient for supporting a robust PQ monitoring system. The described algorithm offers an additional step towards this goal.

Wavelet transform on a PQ signal produces a multi-resolution decomposition (MRD) matrix, which contains time domain information of the signal at different scales. This property has made wavelets a promising tool for detecting and extracting disturbance features of various types of PQ events [5,7,8]. However, there are still some issues to be resolved in wavelet-based methods. First, while PQ disturbances cover a wide frequency range, a very small subset of the MRD matrix (e.g., five scales in [8]) may not be a sufficient or optimized selection for capturing features of all different types of PQ events. This feature selection scheme may filter some important information for classification and potentially degrade the recognition rates. Second, the wavelet-based methods require more training examples than the algorithm proposed in this paper. It results in more efforts or difficulties when adapting the algorithm onto a new system.

The method using a wavelet packet-based hidden Markov model provides improved recognition rates [4], however, it is the post-processing procedures that make a significant contribution to the final results (overall classification rate on 507 testing waveforms: 83.2% without post-processing; 95.5% with post-processing). In [4], the post-processing steps are not very exactly defined in terms of implementation. The ambiguity makes it hard to judge whether this method can achieve comparable performance on other power quality data. The method proposed in this paper does not require a post-processing procedure.

The wavelet transform (WT) or short-time Fourier based methods are good in terms of time domain localization of an event or relatively accurate representation of the time-scale (time-frequency) structure of a waveform. On the other hand, accurate representations often include redundant information that confuses the classification process. For example, two

capacitor high-frequency switching waveforms have wavelet coefficient peaks or Fourier coefficient peaks centered at *different* time locations, if taken WT or STFT. Smoothing procedures are usually applied in the wavelet domain or the spectrogram to remove redundancies. However, it has not been proved that the smoothing procedures adopted can significantly remove redundant information for classification. It has not been proved either that a wavelet or Fourier spectrum is an optimal starting point for generating classification features. In this paper, the proposed statistical smoothing and explicit TFR design techniques has an optimization goal of classification.

To build an ideal PQ monitoring system, the following issues have to be addressed in the classification algorithm. First, new algorithms need to be able to classify voltage or current *transient* events due to different power system events, such as motor starting, transformer inrush currents, capacitor switching, and high impedance faults. Transient identification problems, which require higher frequency waveform recording devices and more robust classification methods, are rarely addressed in the literature, with the exception of [9]. Second, new methods need to be capable of implementing cause-based classifications, in addition to phenomenon-based classifications. For example, for a voltage sag event, it is desirable to know not only that it is a voltage sag, but also whether it is caused by the switching of a large load, a line-to-ground fault, or other reasons. Third, new methods should be able to localize a PQ event, which indirectly leads to a solution for the detected PQ problem. Decisions may be made according to the measurements from multiple monitoring nodes. The number of PQ monitors and their layout in a power system need to be optimized. Fourth, the new methods should also have the reduced computational complexity mode to support real-time on-line monitoring.

None of these issues has been resolved perfectly in existing classification methods. We are not even close to an ideal PQ monitoring system. Sustained efforts in exploring better solutions become crucial for improving the overall quality of services of energy providers and rescuing energy customers from expensive PQ-related losses.

The new approach presented in this paper is composed of two sequential processes: feature extraction and classification. The main contribution of this paper is in the area of feature extraction. The essence of the feature extraction is to project a PQ signal onto a low-dimension time-frequency representation (TFR), which is deliberately designed for maximizing the separability between classes. A distinct TFR is designed for each class. Standard classifiers are adapted in this method, including a Heaviside-function linear classifier and neural networks with feedforward structures. The proposed algorithm has been successfully tested with 860 real world power quality events. A recognition rate of 98.0% has been achieved in the experiment that covers five classes; a DSP-based hardware system, which is capable of processing a five-cycle (83.3 ms) PQ waveform within 11.2 ms, has verified the real-time computing capability of this algorithm

[1].

III. POWER QUALITY EVENTS

Power quality problems are caused by many different system events. PQ disturbances cover a broad frequency range and significantly different magnitude variations, which makes the identification of PQ events often difficult. This section presents five major power quality related waveform events in our study: harmonics, voltage sags, capacitor high frequency switching, capacitor low frequency switching, and normal voltage variations.

Fig. 1 shows sample waveforms of these events recorded from real world power systems. Fig. 1(a) represents the waveform affected by harmonics. Harmonics distortion is the most frequently encountered power quality problem [10]. Because of the increasing popularity of electronic and other non-linear loads, such as adjustable-speed drives, arc furnaces, and induction furnaces, perfect sinusoid waveforms often become distorted by both even and odd harmonics. Transformer core saturation is another cause of harmonic distortion. Major consequences of harmonics include equipment overheating, excessive currents and voltage, and tripping of protective relays.

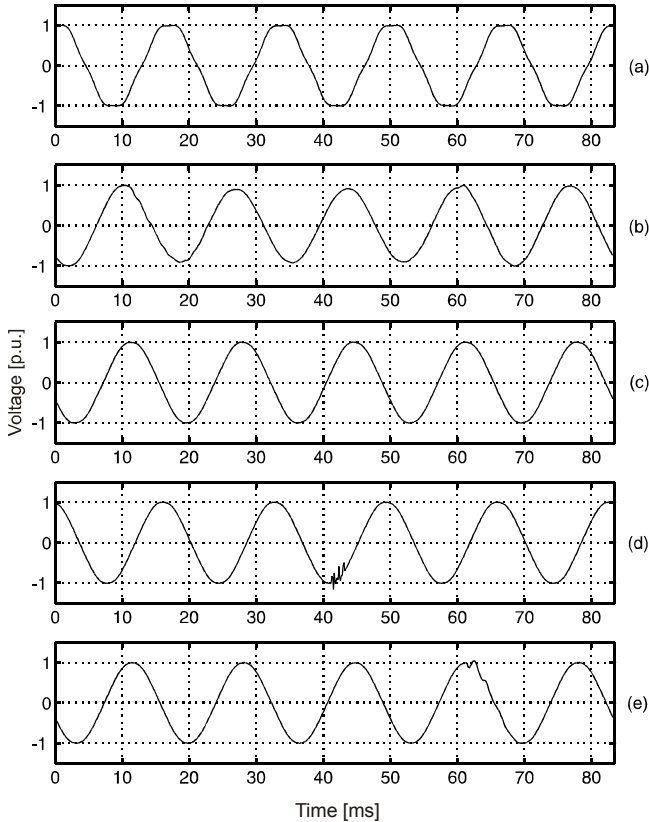


Fig. 1. Five classes of PQ signals for classification.

(a) harmonics; (b) voltage sag; (c) normal voltage variations; (d) capacitor high frequency switching; (e) capacitor low frequency switching.

A momentary voltage dip that lasts for less than one minute and more than eight milliseconds is classified as voltage sag [11]. It is caused by faults, the start of large loads (e.g. motors), and brown-out recovery [10]. Fig. 1(b) shows an example of voltage sag. Major consequences of voltage sag

include short circuits, process interruptions, and loss of data. Process industry equipment is particularly sensitive to voltage sag problems, because conveyor-like machinery is interconnected and temporary failure of any component in the process can cause system downtime.

Fig. 1(c) shows a waveform of normal voltage variations. A normal variation signal is a sinusoidal voltage waveform with electrical system noise on a normal level.

Transients caused by capacitor switching are also among common sources of power quality degradation in utility systems [10]. Two sample signals that correspond to high and low frequency capacitor switching are shown on Fig. 1(d) and Fig. 1(e), respectively. The low frequency range is usually defined as 500 Hz to 2 kHz, and the high frequency range is usually defined as greater than 2 kHz. The frequency of a transient is determined mostly by the capacitance and inductance of system components. Capacitor banks are commonly used in power systems for power factor correction. Major consequences of capacitor switching events include insulation breakdown or sparkover, semiconductor device damage, short circuits, accelerated aging, and loss of data or stability.

These five types of PQ events were used to validate the proposed new classification algorithm, in the Application Part of this paper [1]. The case study tested a total of 860 real-world waveforms.

IV. THEORETIC BACKGROUND

A. Classification-optimal TFR

The solution for a pattern recognition problem usually includes two sequential processes: feature extraction and classification. For waveform recognition problems, given a time series, features are often extracted from some forms of TFRs. To gain more understanding of a given signal, it is often advantageous to study the signal in a different time-frequency representation, based on a different weighting function. From a mathematical point of view, there is an infinite number of ways this can be done. In other words, there is an infinite number of possible TFRs corresponding to a signal [12].

For example, for the high-frequency capacitor switching signal in Fig. 1(d), two TFRs, spectrogram and Rihaczek TFR, are shown in Fig. 2(a) and Fig. 2(b), respectively. Although the high frequency components due to the switching event around 43 ms can be roughly located on both TFRs, the matrices corresponding to these two TFRs show significantly different patterns. This is caused by their different underlying weighting functions. In real applications, one TFR may show its advantages over certain applications, but expose its disadvantages over other applications.

To achieve good classification results, features need to be selected from a TFR that is optimal for the classification task. The spectrogram is traditionally used for PQ event classification. Recently, wavelet multi-resolution decomposition matrix (MSD) becomes dominant for this

purpose. However, a spectrogram or a wavelet MSD is defined to accurately describe the component frequencies (scales) of a signal over the time, while the classification problem requires a TFR that maximizes the separability of signals in different classes. There are not enough evidences showing that existing spectrogram or wavelet-based methods have generated such TFRs. Therefore, it is desirable to design a classification-optimal representation TFR_i that specifically emphasizes the differences between classes [13,14]. It is not necessary for the representation TFR_i to accurately describe the time-frequency information of the signal. Given the desired classification-optimal representation TFR_i , features can be generated by mapping signals to TFR_i , as illustrated in Fig. 3. The question here is how to find the TFR_i .

Time-frequency ambiguity planes have been an important tool in the radar field, in analyzing and constructing radar signals, formulating the performance characteristics of a waveform, and relating range and velocity resolution [15]. It has also been used extensively in the fields of sonar, radio astronomy, communications, and optics [16]. Gillespie and Atlas have recently proposed feature extraction methods based on designing class-dependant TFRs from time-frequency ambiguity plane. This class of new techniques has been successfully applied for tool-wear monitoring and radar transmitter identification [14,17,18].

The important connection between the ambiguity plane and time-frequency representations has been recognized for a long time. Any bilinear (Cohen class) TFR $P(t, f)$ can be expressed as the two-dimensional Fourier transform of the product of the ambiguity plane $A(\eta, \tau)$ of the signal and a kernel function $\varphi(\eta, \tau)$ [12]:

$$P(t, f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(\eta, \tau) \varphi(\eta, \tau) e^{j2\pi\eta t} e^{-2\pi jf\tau} d\eta d\tau \quad (1)$$

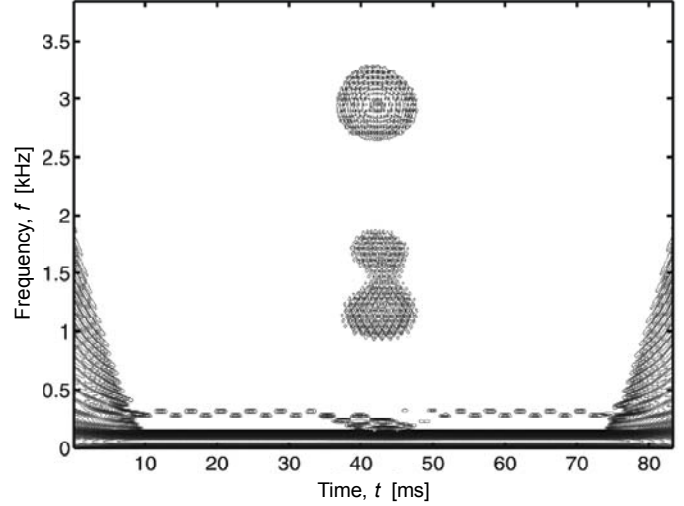
where t represents time, f represents frequency, η represents continuous frequency shift, and τ represents continuous time lag. The ambiguity plane $A(\eta, \tau)$ for a given signal $s(t)$ is defined as:

$$A(\eta, \tau) = \int_{-\infty}^{\infty} s(t) s^*(t + \tau) e^{j2\pi\eta t} dt \quad (2)$$

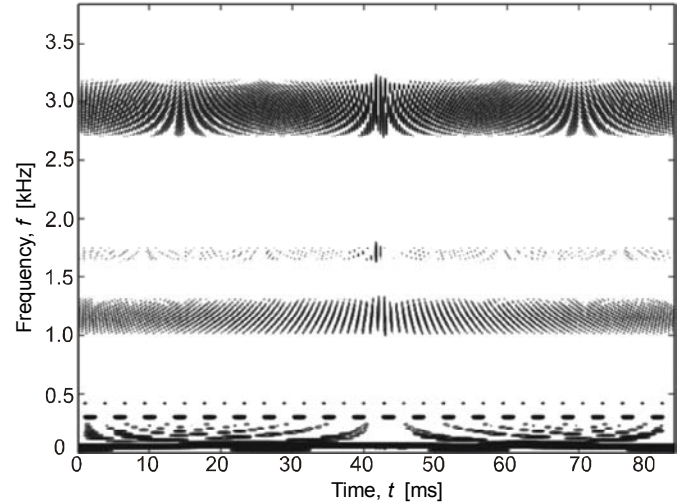
Here $s(t)$ represents the signal at time t , and $s(t + \tau)$ represents the signal at a future time $t + \tau$, and the $s^*(t + \tau)$ means the complex conjugate of $s(t + \tau)$.

The ambiguity plane is utilized in TFR design because of its desirable properties for classification. It can be interpreted as a joint time-frequency correlation function. Each location in this plane contains global time-frequency information about the signal. When evaluated along two axes, i.e. $\eta = 0$ and $\tau = 0$, it simplifies to either the time-domain or the frequency domain correlation function. If only the locations on the axis $\eta = 0$ are considered and other locations are ignored, then the designed TFR only contains steady-state frequency

information and all temporal information is smoothed. If only the locations on the axis $\tau = 0$ are considered and other locations are ignored, then the designed TFR only contains temporal information and all spectral information is smoothed. Locations not on either axis correspond to a sloped representation of temporal and spectral information. Using the appropriate combination of weighted points, any TFR can be generated [14,16].



(a) The log scale spectrogram of the PQ signal in Fig. 1-(d).



(b) The Rihaczek TFR of the PQ signal in Fig. 1-(d).

Fig. 2. A PQ signal and its two TFRs.

The infinite number of possible quadratic TFRs for a signal results from the infinite number of possible choices for the kernel. The kernel determines the representations and its properties. A kernel function is a generating function that operates upon the signal to produce the TFR. The characteristic function for each TFR $P(t, f)$ is $A(\eta, \tau) \varphi(\eta, \tau)$. In other words, given a signal, a TFR can be uniquely mapped from a kernel.

The relationships above provide two important implications. First, just like any other TFR, the classification-optimal representation TFR_i can be obtained through

smoothing the ambiguity plane with an appropriate kernel φ_i , which is a classification-optimal kernel. The problem of designing the TFR_i becomes equivalent to designing the classification-optimal kernel $\varphi_i(\eta, \tau)$. Second, because of the characteristic (uniquely mapping) relationship between a TFR $P(t, f)$ and a kernel-smoothed ambiguity plane $A(\eta, \tau)\varphi_i(\eta, \tau)$, features can also be extracted directly from the $A(\eta, \tau)\varphi_i(\eta, \tau)$, instead of the classification-optimal TFR_i . This shortcut simplifies the computation complexity of feature extraction, by reducing the calculations described in Equation (1).

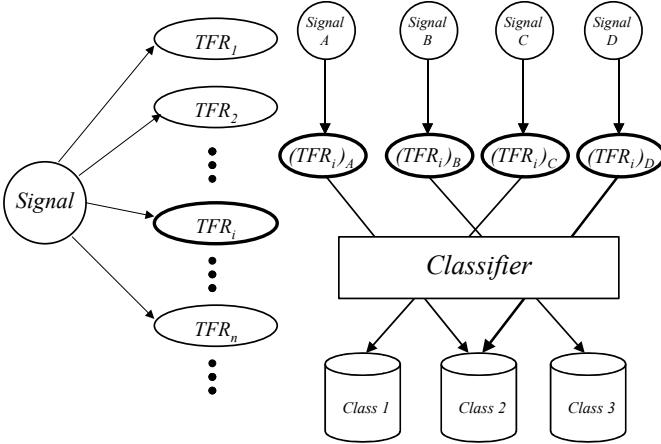


Fig. 3. Extract features by mapping signals to the classification-optimized representation TFR_i .

B. Fisher's discriminant function

Fisher's discriminant function (FDF), which was developed by R. A. Fisher in 1930s, is a method that projects high-dimensional data onto low-dimensional space for classification. The projection maximizes the distances between the means of the different classes while minimizing the variances within each class. Equation (3) defines the Fisher's discriminant function:

$$J_F(\omega) = \frac{(m_1 - m_2)^2}{D_1^2 + D_2^2} \quad (3)$$

where m represents a mean, D^2 represents a variance, and the subscripts denote the two classes. The Fisher's criterion defines that the best ω 's are the ones that maximize $J_F(\omega)$. The criterion is also known as signal-to-interface ratio.

In this paper, with the dimension reduction capability, the Fisher's discriminant function is adopted for the design of the classification-optimal kernel functions.

V. CLASSIFICATION ALGORITHM AND TRAINING TECHNIQUES

In this study, all 860 real world voltage signals from five event classes were collected from industrial databases for the training and the testing of the algorithm. Each voltage signal to be identified consists of five cycles of voltage waveform sampled 128 times per cycle, and has a length of 640 sampling

points. A window size of 83.3 ms (five sinusoidal waveform cycles in a 60 Hz system) is adopted because of the following two reasons. First, a five-cycle window is long enough for capturing the characteristics of all types of PQ events under our study. Second, a five-cycle window is short enough for generating real-time monitoring outputs. This window size is used in this paper to demonstrate the algorithm, and can be adjusted appropriately upon specific applications. For example, when this method is dedicated for discriminating different types of high frequency power system transients, the window size can be reduced to a fraction of a cycle.

The classification algorithm is composed of two parts: the training algorithm and the implementation algorithm. In the training stage, four classification-optimal kernels are designed for separating four classes sequentially. The kernel design process selects nine locations from the time-frequency ambiguity plane. In the training stage, four different classifiers (linear or neural net classifier) are optimally selected and trained. In the implementation stage, nine feature points are generated directly from the signal and sent to four different classifiers.

A. Design of classification kernels

According to the Fisher's discriminant function, four classification-optimal kernels are designed for four classes: harmonics, voltage sags, normal variations, and capacitor high-frequency switching, respectively.

The discrete versions of equations (1) and (2) are [14]

$$TFR[n, k] = \mathcal{F}_{\eta \rightarrow n}^{-1} \{ \mathcal{F}_{\tau \rightarrow k} \{ A[\eta, \tau] \varphi[\eta, \tau] \} \} \\ = \frac{1}{N} \sum_{\eta=0}^{N-1} \sum_{\tau=0}^{N-1} A[\eta, \tau] \varphi[\eta, \tau] e^{-j(2\pi/N)\tau k} e^{j(2\pi/N)\eta n} \quad (4)$$

and

$$A[\eta, \tau] = \mathcal{F}_{n \rightarrow \eta} \{ R[n, \tau] \} \\ = \sum_{n=0}^{N-1} R[n, \tau] e^{-j(2\pi/N)\eta n} \quad (5)$$

where the function $\mathcal{F}()$ represents the Fourier transform, the function $\mathcal{F}^{-1}()$ represents the inverse Fourier transform, n represents sample, k represents discrete frequency, η represents discrete frequency shift, and τ represents discrete time lag. The instantaneous autocorrelation function $R[n, \tau]$ is defined as

$$R[n, \tau] = x^*[n]x[\text{mod}(n + \tau, N)] \quad (6)$$

where the function $\text{mod}(p_1, p_2)$ represents modulus after dividing the first parameter p_1 by the second parameter p_2 . As stated in Section IV-A, features can be extracted directly from $A[\eta, \tau]\varphi_i[\eta, \tau]$. The kernel $\varphi_i[\eta, \tau]$ has the same dimensions as the ambiguity plane $A[\eta, \tau]$. In this application, the kernel $\varphi_i[\eta, \tau]$ is defined as a binary matrix (each matrix element is either 0 or 1), so,

$$A[\eta, \tau] \varphi_i[\eta, \tau] = \begin{cases} A[\eta, \tau], & \text{if } \varphi_i[\eta, \tau] = 1 \\ 0, & \text{if } \varphi_i[\eta, \tau] = 0 \end{cases} \quad (7)$$

Feature points are ambiguity plane points of locations (η, τ) where $\varphi_i[\eta, \tau] = 1$. Therefore, the process of feature extraction is to select points that are optimal for the classification task from the ambiguity plane.

A total number of $N-1$ kernels need to be designed for a N -class PQ classification system (N is the number of classes, here $N=5$). A kernel K_i works as either a single-class separator or a group-class separator. In the single-class separator case, kernel K_i is dedicated to discriminate class i from all remaining classes $\{i+1, \dots, N\}$. In the group-class separator case, kernel K_i is dedicated to discriminate a class group $\{i, i+1, \dots, i+m\}$ from all remaining classes $\{i+m+1, i+m+2, \dots, N\}$. In the second case, additional kernels are needed in order to uniquely identify class i from the class group $\{i, i+1, \dots, i+m\}$, and the total number of kernels required for a N -class classification is still $N-1$.

In the algorithm presented in this paper, four kernels designed include: harmonic kernel, sag kernel, capacitor switching kernel, and capacitor high-frequency switching kernel. The harmonics kernel is to separate the class *harmonics* from all remaining classes $\{sag, capacitor\ high\ frequency\ switching, capacitor\ low\ frequency\ switching, normal\ variations\}$. The sag kernel is to separate the class *sag* from all remaining classes $\{capacitor\ high\ frequency\ switching, capacitor\ low\ frequency\ switching, normal\ variations\}$. The capacitor switching kernel is to separate the class group $\{capacitor\ high\ frequency\ switching, capacitor\ low\ frequency\ switching\}$ from the remaining class *normal variations*. The capacitor high-frequency switching kernel is separate the class *capacitor high-frequency switching* from the remaining class *capacitor low-frequency switching*.

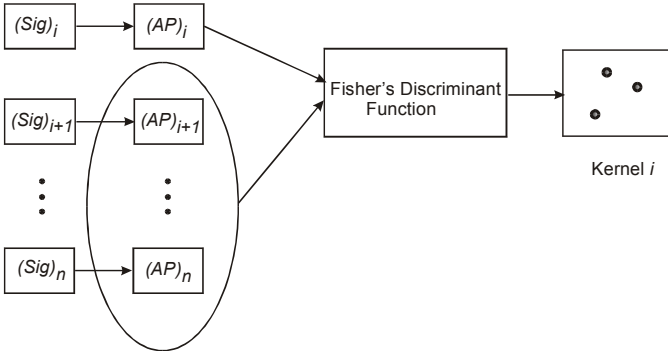


Fig. 4. Kernel design.

As shown in Fig. 4, after calculating ambiguity planes for all training signals, the Fisher's discriminant function is applied for the kernel design. Assume there are n classes and totally N_i training examples for class i . The notation $A_j[\eta, \tau]$ represents the ambiguity plane of the j th training example in the i th class.

With the Fisher's criterion, locations on the ambiguity

plane are ranked according to their importance for classification. A small amount of training data from each class is needed feature ranking in this statistical method. For example, when designing kernel i , a Fisher's discriminant score is calculated for each location (η, τ) on the ambiguity plane,

$$J_{F_i}(\eta, \tau) = \frac{(m_i[\eta, \tau] - m_{i-remain}[\eta, \tau])^2}{D_i^2[\eta, \tau] + D_{i-remain}^2[\eta, \tau]} \quad (8)$$

where $m_i[\eta, \tau]$ and $m_{i-remain}[\eta, \tau]$ represent two means of location (η, τ) ,

$$m_i[\eta, \tau] = \frac{1}{N_i} \sum_{j=1}^{N_i} A_{ij}[\eta, \tau] \quad (9)$$

$$m_{i-remain}[\eta, \tau] = \frac{\sum_{k=i+1}^5 \sum_{j=1}^{N_k} A_{kj}[\eta, \tau]}{\sum_{k=i+1}^5 N_k} \quad (10)$$

and $D_i^2[\eta, \tau]$ and $D_{i-remain}^2[\eta, \tau]$ represent two variances of location (η, τ) ,

$$D_i^2[\eta, \tau] = \frac{1}{N_i} \sum_{j=1}^{N_i} (A_{ij}[\eta, \tau] - m_i[\eta, \tau])^2 \quad (11)$$

$$D_{i-remain}^2[\eta, \tau] = \frac{\sum_{k=i+1}^5 \sum_{j=1}^{N_k} (A_{kj}[\eta, \tau] - m_{i-remain}[\eta, \tau])^2}{\sum_{k=i+1}^5 N_k} \quad (12)$$

Locations $[\eta, \tau]$ that receive the highest discriminant score $J_{F_i}(\eta, \tau)$ are selected as feature locations. The factor of correlation is also considered. In this case, only one point is selected as a feature among a group of points that are highly correlated. Because the ambiguity plane of a signal is symmetric according to two dimensions, only points on a quarter plane are considered.

Specifically, one feature point is selected for the harmonic kernel, two for sag kernel, three for normal variation kernel, and three for the capacitor high-frequency switching kernel. Therefore, only nine points on the ambiguity plane need to be calculated and sent to classifiers. The feature extraction process reduces the dimension by the factor of 71.1 (from 640 to 9). The number of feature locations needed for each kernel is optimized by training experiments.

B. Training of classifiers

Multiple classifiers are adopted in the presented method. As shown in Fig. 5, each classification node consists of a kernel function and a classifier. Heaviside-function linear classifier is used for the task of separating two classes that have a big

distance and are relatively easy to discriminate. Neural networks with small numbers of input nodes are used for all other classification tasks. The feedforward structure is selected for the following reasons: a) the feature extraction method does not emphasize the inter-window correlations; b) the feedforward structure is easier to implemented than recurrent structures. Depending on the nature of the kernel, classification node i is to either discriminate signals that belong to class i from signals that belong to class $\{i+1, \dots, n\}$, as shown in Fig. 5(a), or discriminate signals in class $\{i, \dots, i+m\}$ from signals in class $\{i+m+1, \dots, n\}$, as shown in Fig. 5(b).

Four classifiers are employed in this classification application for five types of PQ events. They are a Heaviside linear classifier (for the class of harmonics) and three feedforward neural network classifiers (for the other three classes) with simple structures. The Heaviside linear classifier is chosen for discriminating harmonics class from all other classes, because the distance between the harmonics class and all other classes is relatively big and the classification task is relatively simple.

For a two-class classification problem and an input f , the Heaviside linear classifier is defined as,

$$H(f-t) = \begin{cases} 1 & (f \text{ belongs to class 1}), \text{ if } f-t \geq 0 \\ 0 & (f \text{ belongs to class 2}), \text{ if } f-t < 0 \end{cases} \quad (13)$$

where t is a threshold value. Training this classifier is to determine the parameter t .

$$t = \frac{f_{1-\min} + f_{2-\max}}{2} \quad (14)$$

where $f_{1-\min}$ represents the minimum input from class 1 and $f_{2-\max}$ represents the maximum input from class 2.

Each of the three feedforward artificial neural network (ANN) classifiers adopted in this algorithm has three layers. Extensive classification experiments were conducted to determine the optimized neural network structures. The structure of the ANN for discriminating sags is 2-12-2 (input layer node number - hidden layer node number - output layer node number); the one for capacitor switching is 3-10-2; the one for capacitor high-frequency switching is 3-10-2. The transfer and training functions adopted for the ANN include: the hyperbolic tangent sigmoid transfer function as the transfer function for the hidden layer, the linear transfer function as the transfer function for the output layer, the Levenberg-Marquardt backpropagation as the network training function, the gradient descent learning function as the weight learning function, and the mean squared error function as the performance evaluation function.

VI. CONCLUSIONS

Identification and classification of voltage and current disturbances in power systems is an important task in power system monitoring and protection. A new classification algorithm for power quality disturbances has been proposed

and successfully validated. The proposed feature extraction tool, a time-frequency ambiguity plane with kernel techniques, is entirely new to the power engineering field. By designing a classification-optimal TFR, features are directly selected from the time-frequency ambiguity plane based on Fisher's discriminant function. Four linear and neural network classifiers with simple structures are used for the final decision-making. This paper (the first part of a two-paper series) presents the background and theory of the proposed PQ classification method. An application of the proposed method to real world power quality data is presented on the second part of this paper. The proposed algorithm has been successfully tested with 860 real world power quality events. A recognition rate of 98.0% has been achieved in a classification experiment that covers five classes; a DSP-based hardware system, which is capable of processing a five-cycle (83.3 ms) PQ waveform within 11.2 ms, has verified the real-time computing capability of this algorithm [1].

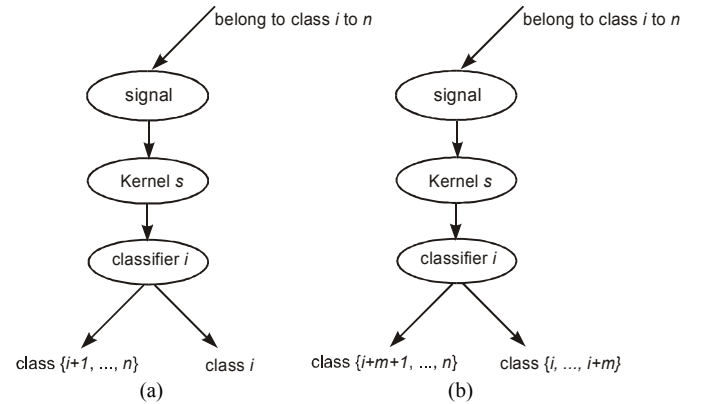


Fig. 5. at each classification node i .

VII. ACKNOWLEDGMENTS

Special thanks go to the UW EE graduate student Piotr Ochenkowski, and undergraduate students Gabriel Rowe, Cheuk-Wai Mak, Jeff Chen, Rejo Jose, My Hua, and Gabriel Rowe for their work on power quality monitoring equipment, event simulation software, and signal processing algorithms, respectively. The authors also would like to acknowledge the help on power quality data collection from Joe Wilson (Eugene Water and Electric Board), James Harris (Bonneville Power Administration), Mark Kirschenbaum (University of Washington Power Plant), and Stephen Sturgill (Salt River Project). The authors would also like to thank scholarship sponsorships from APPA's (America Public Power Association) DEED (Demonstration of Energy-Efficient Developments) program and EEIC (Electrical Energy Industrial Consortium) undergraduate research program.

VIII. REFERENCES

- [1] M. Wang and A. V. Mamishev, "Classification of Power Quality Events Using Optimal Time-Frequency Representations -- Part 2: Application," *IEEE Transactions on Power Delivery*, 2003.
- [2] G. T. Heydt, "Electric Power Quality: a Tutorial Introduction," *IEEE Computer Applications in Power*, vol. 11, pp. 15-19, Jan. 1998.

- [3] CEIDS (Consortium for Electric Infrastructure to Support a Digital Society), "The Cost of Power Disturbances to Industrial and Digital Economy Companies Executive Summary," July 2001.
- [4] J. Chung, E. J. Powers, W. M. Grady, and S. C. Bhatt, "Electric Power Transient Disturbance Classification Using Wavelet-Based Hidden Markov Models," *Proceedings of 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, 2000, pp. 3662-3665.
- [5] A. M. Gaouda, M. A. Salama, M. R. Sultan, and A. Y. Chikhani, "Power Quality Detection and Classification Using Wavelet-Multiresolution Signal Decomposition," *IEEE Transactions on Power Delivery*, vol. 14, no. 4, pp. 1469-1476, Oct. 1999.
- [6] J. S. Lee, C. H. Lee, J. O. Kim, and S. W. Nam, "Classification of Power Quality Disturbances Using Orthogonal Polynomial Approximation and Bispectra," *Electronics Letters*, vol. 33, no. 18, pp. 1522-1524, Aug. 1997.
- [7] B. Perunicic, M. Mallini, Z. Wang, Y. Liu, and G. T. Heydt, "Power Quality Disturbance Detection and Classification Using Wavelets and Artificial Neural Networks," *The 8th International Conference on Harmonics and Quality of Power*, 1998, pp. 77-82.
- [8] S. Santoso, E. J. Powers, W. M. Grady, and A. C. Parsons, "Power Quality Disturbance Waveform Recognition Using Wavelet-Based Neural Classifier. I. Theoretical Foundation," *IEEE Transactions on Power Delivery*, vol. 15, no. 1, pp. 222-228, Jan. 2000.
- [9] A. V. Mamishev, S. X. Short, T. W. Kao, and B. D. Russell, "Analysis of High Impedance Faults Using Fractal Techniques," *IEEE Transactions on Power Systems*, vol. 11, no. 1, pp. 435-440, 1996.
- [10] R. C. Dugan, M. F. McGranaghan, and H. W. Beaty, *Electrical Power Systems Quality*, McGraw-Hill, 1996.
- [11] B. Kennedy, *Power Quality Primer*, McGraw-Hill, 2000.
- [12] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall, 1995.
- [13] M. Davy and C. Doncarli, "Optimal Kernels of Time-Frequency Representations for Signal Classification," *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, 1998, pp. 581-584.
- [14] B. W. Gillespie and L. Atlas, "Optimizing Time-Frequency Kernels for Classification," *IEEE Transactions on Signal Processing*, vol. 49, no. 3, pp. 485-496, Mar. 2001.
- [15] L. Cohen, "Time-Frequency Distributions - A Review," *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941-981, July 1989.
- [16] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and Quadratic Time-Frequency Signal Representations," *IEEE Signal Processing Magazine*, vol. 9, no. 2, pp. 21-67, Apr. 1992.
- [17] B. W. Gillespie and L. Atlas, "Optimization of Time and Frequency Resolution for Radar Transmitter Identification," *Proceedings of 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, 1999, pp. 1341-1344.
- [18] B. W. Gillespie and L. Atlas, "Data-Driven Time-Frequency Classification Techniques Applied to Tool-Wear Monitoring," *Proceedings of 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, 2000, pp. 2045-2048.

IX. BIOGRAPHIES



Min Wang received a Bachelor's degree from Tsinghua University, Beijing, China, in 1999, and a M.S.E.E. degree from the University of Washington, Seattle, in 2001. Currently he is a Ph.D. student and research assistant at the Department of Electrical Engineering, University of Washington. His research interests include digital signal processing, power quality monitoring, RFID, embedded systems, and ubiquitous computing. He is a recipient of the First Prize Award in the Student Paper-Poster Competition in the 2002 IEEE Power Engineering Society Summer Meeting in Chicago.



Alexander V. Mamishev received an equivalent of B.S. degree from the Kiev Polytechnic Institute, Ukraine in 1992, M.S. degree from Texas A&M University in 1994, and a Ph.D. degree from MIT in 1999, all in electrical engineering. Currently he is an assistant professor and director of SEAL (Sensors, Energy, and Automation Laboratory) in the Department of Electrical Engineering, University of Washington, Seattle. Prof. Mamishev is an author of about 60 journal and conference papers, and one book chapter. His research interests include PQ monitoring, sensor design and integration, dielectrometry, and electrical insulation diagnostics. He serves as Associate Editor for the IEEE Transactions on Dielectrics and Electrical Insulation and a reviewer for the IEEE Transactions on Power Delivery. He is a recent recipient of the NSF CAREER Award and the IEEE Outstanding Branch Advisor Award.