

Disturbance Classification Using Hidden Markov Models and Vector Quantization

T. K. Abdel-Galil, *Member, IEEE*, E. F. El-Saadany, *Member, IEEE*, A. M. Youssef, and M. M. A. Salama, *Fellow, IEEE*

Abstract—This paper presents a novel approach to the classification of power quality disturbances by the employment of Hidden Markov Models. In these models, power quality disturbances are represented by a sequence of consecutive frames. Both the Fourier and Wavelet Transforms are utilized to produce sequence of spectral vectors that can accurately capture the salient characteristics of each disturbance. Vector Quantization is used to assign chain of labels for power quality disturbances utilizing their spectral vectors. From these labels, a separate Hidden Markov Model is developed for each class of the power quality disturbances in the training phase. During the testing stage, the unrecognized disturbance sequence is matched against all the developed Hidden Markov Models. The best-matched model pinpoints the class of the unknown disturbance. Simulation results prove the competence of the proposed algorithm.

Index Terms—Classification, hidden Markov models, monitoring techniques, power quality, vector quantization.

I. INTRODUCTION

POWER quality monitors are increasingly being used to assist electric utilities in the evaluation of the present attributes of electric power so that quality improvement actions can be taken. Since these monitors capture immense amounts of power quality data, the analyzes prohibitively expensive due to time and expertise. Therefore, the automated classification of power quality disturbances has become a significant issue, especially in a deregulated era where power quality is becoming one of the critical differentiating factors for choosing different suppliers [1], [2].

Recently, many automated systems for the classification of power quality have been designed. Typically, They are based on Artificial Neural Networks (ANN's) [3], [4], template matching [5], dynamic time warping (DTW) [6], and rule-based systems [7]–[10]. The main disadvantage of ANN based classifiers is the retraining requirement subsequent to the addition of a new phenomenon to the set of classifier outputs. On the other hand, building a rule-based classifier is tedious, and results in a complicated inference mechanism when a large number of disturbance classes are present. The DTW approach [6] can deal with the time alignment issue between the test and the stored power quality templates, but it requires too much computation time.

Manuscript received May 20, 2003; revised July 14, 2004. Paper no. TPWRD-00235-2003.

T. K. Abdel-Galil is with King Fahd University Petroleum and Minerals, Dhahran 31261, Saudi Arabia.

E. F. El-Saadany, and M. M. A. Salama are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: tkabdelg@hivolt.uwaterloo.ca).

A. M. Youssef is with Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC H3G 1M8, Canada.

Digital Object Identifier 10.1109/TPWRD.2004.843399

Therefore, Hidden Markov Models (HMMs) are introduced in [10] to classify between fast transient phenomena by using wavelet packet coefficients. Since the authors [10] used wavelet packet coefficients as the HMMs observation sequences, HMMs have not been able to classify slow phenomena such as sag because a very long observation sequence is required for slow disturbances.

To overcome this limitation, power quality disturbance classification utilizing Discrete Density Hidden Markov Models combined with Vector Quantization (VQ) is proposed in this paper. The proposed method can classify between fast and slow phenomena and it is computationally efficient. To achieve this goal, a sliding window of either Wavelet Transform (WT) or Fourier transforms (FT) is applied to the recorded voltage signal. Then, the VQ is utilized to convert the continuous observation (FFT or DWT vectors) into a discrete observation sequence. The labels produced by VQ step are utilized as an input observation to HMMs. This results in a short observation sequence, which can be used to distinguish between slow and fast phenomena efficiently.

In the training stage, an HMM is constructed for each phenomenon under consideration. In the classification (testing) stage, these models are used to produce a ranked list of the most probable power quality disturbances associated with the signal under consideration. Moreover, with the inherent scalability of the HMMs, models representing new phenomena can be constructed without the need to retrain the already existing models.

The architecture and methodology of the proposed system is discussed in Section II. The WT and FT that are used in feature extraction are presented in Section III. Section IV details the implementation of the proposed HMM. The VQ algorithm is described in Section V. The supporting results are given in Section VI, and Section VII concludes the paper.

II. SYSTEM ARCHITECTURE

Designing a classifier to fit a specific application entails two fundamental steps. The first is to attain a good extraction of the features that ensure highly discriminatory attributes between different disturbances, and the second is to select an appropriate classifier paradigm. Fig. 1 illustrates the architecture of the proposed classification system. In this paper, the required features are extracted by a sliding window of either the FT or the WT. VQ allows each waveform under consideration to be represented as a time sequence of coded spectral vectors. The development of a proper HMM classifier requires two stages: training and testing.

During the training stage, a separate HMM for each power quality disturbance class under consideration is constructed.

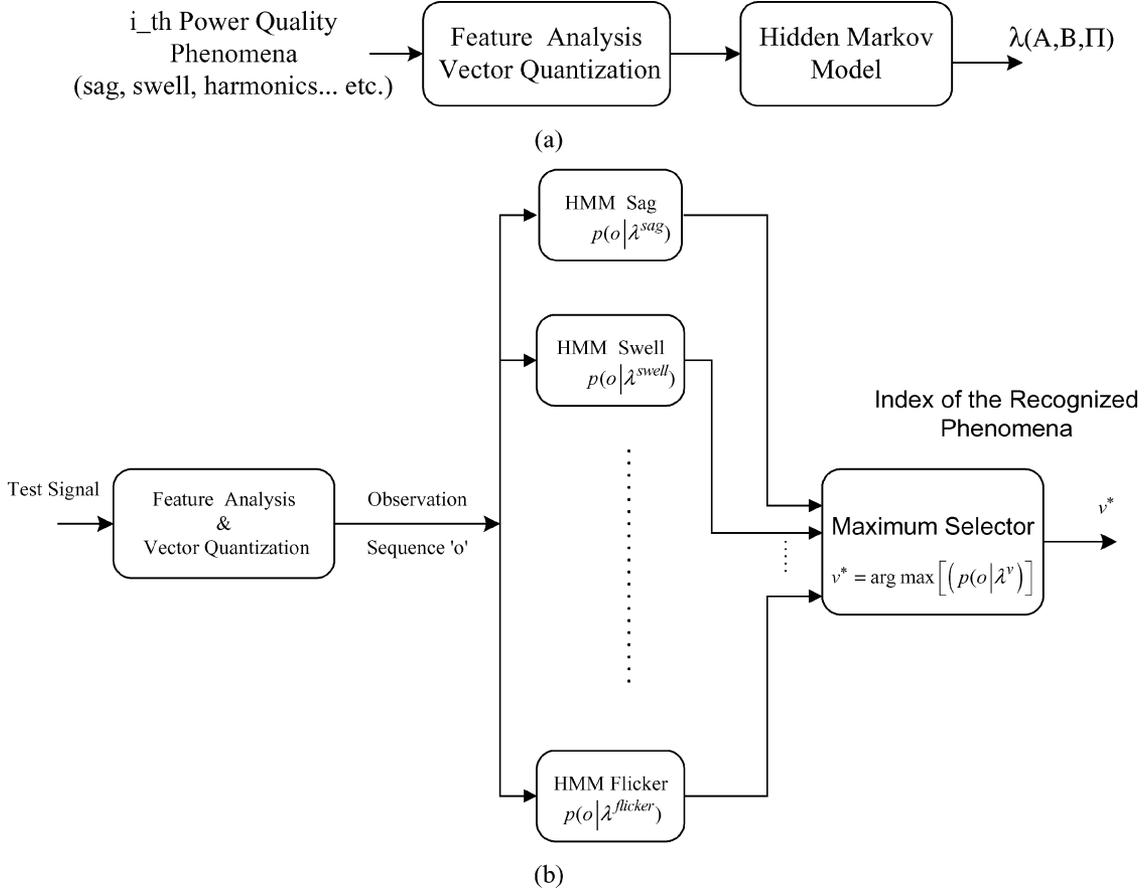


Fig. 1. Hidden Markov Models (a) training stage. (b) Testing stage.

The model parameters are optimized by using observation sequences resulting from the VQ stage. Therefore, the training stage is necessary to obtain models that best characterize each class of power quality disturbances under consideration.

In the testing stage, the observed sequence of the phenomenon under consideration is matched against all the HMMs previously obtained in the training stage. The HMM which attains the highest score in the matching process identifies the class of the test phenomenon.

III. FEATURE REPRESENTATION

A. Feature Extraction Using Discrete Wavelet Multi-Resolution Analysis (MRA)

In this section, the Discrete Wavelet Transform (DWT) is introduced as the basic tool for the decision tree feature extraction. DWT is the discrete counterpart of the Continuous Wavelet Transform (CWT). The CWT of a continuous time signal x_t is defined as

$$CWT_{\psi}x(a, b) = \int_{-\infty}^{+\infty} x(t)\psi_{a,b}^*(t)dt, \quad a, b \in R, \quad a \neq 0 \quad (1)$$

where

$$\psi_{a,b}^*(t) = \frac{1}{\sqrt{a}}\psi^*\left(\frac{t-b}{a}\right). \quad (2)$$

The function $\psi(t)$ is the mother wavelet, and the asterisk denotes a complex conjugate. a and b are the scaling and translating parameters, respectively. In practical applications, the DWT (3) of the sampled signal x_k is used to replace the CWT of x_t such that

$$DWT_{\psi}x(m, n) = \sum_k x_k \psi_{m,n}^*(k) \quad (3)$$

where

$$\psi_{m,n}^*(k) = \frac{1}{\sqrt{a_0^m}}\psi^*\left(\frac{k-nb_0a_0^m}{a_0^m}\right). \quad (4)$$

Both the scaling factor a_0^m and the shifting factor $nb_0a_0^m$ are functions of the integer parameter m , where m and n are scaling and sampling numbers, respectively, and $m = 0, 1, 2, \dots$. By selecting $a_0 = 2$ and $b_0 = 1$, a representation of any signal x_k at various resolution levels can be developed by using the MRA. It is implemented by a set of successive filter banks [11], [12] where the mother wavelet is a low pass filter $h(n)$ and its dual as a high pass filter $g(n)$.

Consider a general stage of the filter bank; then the relation between the input sequence $c_{j-1}(n)$ to this stage and its output, which is the detail coefficients $d_j(n)$ and approximated coefficients $c_j(n)$, is given as [12], [13]

$$c_j(n) = \sum_k h(k-2n)c_{j-1}(k) \quad (5)$$

and

$$d_j(n) = \sum_k g(k - 2n)c_{j-1}(k) \quad (6)$$

where c_j represents the coefficients of the approximate signal at level j , and d_j represents the detailed coefficients of the signal at level j . The same procedure can be applied recursively; thus, the signal is mapped by the following set of coefficients: $[c_0 \ d_0 \ d_1 \ \dots \ d_{j-1}]$.

B. Feature Extraction Using Fourier Transform

The FT is utilized to convert time domain waveforms into their frequency components [13]. For a discrete time signal f of L points, the Discrete Fourier Transform (DFT) is given by

$$X(k) = \sum_{n=0}^{L-1} x(n)e^{-\frac{j2\pi kn}{L}}, \quad 0 \leq k \leq L-1 \quad (7)$$

where $X(k)$ denotes the k^{th} spectral sample, and $x(n)$ denotes the n^{th} input sample. For $L = 2^l$, l is an integer, Fast Fourier Transform (FFT) algorithms allow the computation of the DFT to be performed in $O(L \log L)$ rather than in $O(L^2)$. Different windowing types (such as the rectangular, Hanning, Triangular, and Hamming windows) can be applied in the preprocessing step. The experiments demonstrate that a rectangular window with a width equal to one power cycle and zero overlap between the consecutive windows is sufficient to achieve the desired goal.

IV. HIDDEN MARKOV MODELS (HMMS)

Once the disturbance feature vectors have been extracted, the next step is to determine the similarities between the test and reference sequences. An efficient recognition strategy that makes use of the statistical nature of these power quality disturbances is the use of HMMs. These models have been found to offer a performance comparable to other classification approaches such as DTW [14], [15] at a fraction of the computational cost of other approaches.

An HMM is defined as a doubly stochastic process, comprised of an underlying stochastic process that is not directly observable, but can only be visualized through another set of stochastic processes that produce a sequence of observations [14], [15]. In simpler terms and according to the current application, it is a collection of hidden states connected by transitions with an output emitted during each transition; the output from these hidden states is directly derived from the disturbance sequence. The model is prefixed by the term ‘‘hidden’’, since the states of the model at any given time are not directly observable. A discrete HMM is defined as follows.

- N is the number of states in the model that are hidden and not observable. Each state of the model can emit a set of observation symbols O , which are obtained from the power quality disturbance measurements. The individual states are denoted by $S = \{s_1, s_2, \dots, s_N\}$.
- M is the number of distinct observation symbols for each state. Vector quantization is conducted to convert all power quality disturbance sequences into a sequence of observation symbols.

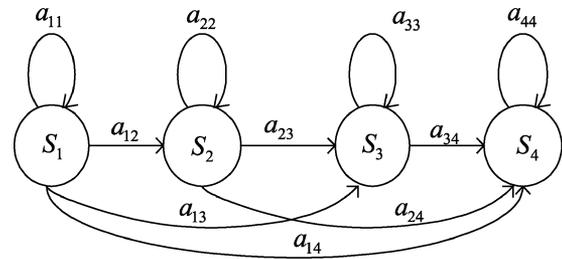


Fig. 2. The four states (left to right) HMM.

- A is a set of transition probabilities where a_{ij} is the probability that the model is making a transition from state i to state j .
- B is the output probability matrix, that is, the probability of emitting distinct observation symbol, which is related to the power quality sequence measurements, when being in state j .
- π is the initial state probabilities.

To sum up, a complete specification of the HMM includes two constant size parameters, N and M , and three matrices of probability measures A , B , π , the HMM is defined by the following notation $\lambda = (A, B, \pi)$.

HMMs in which every state in the model can be reached (in a single step) from any other state of the model are given the appellation ‘‘ergodic’’, meaning that they are fully connected. However, there are other types of HMMs that are known to account for properties of the signal being observed better than the standard ergodic model. One such model is the left-to-right model. In this model, as the time increases, the states proceed from left to right. The fundamental property of the left-to-right model is that $a_{ij} = 0$, when $j < i$. There are many possible permutations and combinations for the shape of the HMM. Unfortunately, there is no rigorous theoretical approach to determine the optimum topology of the model. For the task of power quality disturbance classification, the initial topology for an HMM may be determined by estimating the number of different states that are involved in specifying a given disturbance behavior. The fine-tuning of this topology can be performed empirically. Although different topologies may be specified for each phenomenon, a four-state left to right HMM is found to be sufficient for the power quality classification purpose. The four states of the HMM is depicted in Fig. 2.

HMMs can be classified, according to the observation density, as discrete observation density models DD-HMMs or continuous observation density HMMs [15]. In this research, DD-HMMs are employed to perform the task of power quality disturbance classification, because these models require less computation and less training data without affecting the classification accuracy.

A. HMM Implementation

Essentially, the implementation of HMM first involves a training phase during which a model for each disturbance is built. The training step attempts to optimize the disturbance model parameters $\lambda(A, B, \pi)$ for a given set of disturbance training sequences. Then there is the recognition phase, which involves classification of new unseen sequences.

The calculation of $P(O|\lambda)$, the probability that the observation sequence $O = [O_1, O_2, \dots, O_T]$ is produced by the model λ , is calculated for all available disturbances HMMs in the recognition step. The most intuitive way to calculate $P(O|\lambda)$ is to sum up the probabilities of all possible state sequences

$$P(O, \lambda) = \sum_{\text{all } S} P(S|\lambda)P(O|S, \lambda) \quad (8)$$

Consequently, to compute $P(O|\lambda)$, all possible state sequences S that can generate the observation sequence O are enumerated, and then all the probabilities are added. The probability of each path S is the product of the state sequence probability (the first factor) and the joint output probability (the second factor) along the path.

Although the calculation of (8) is straightforward, its computation requires enumeration over N^T possible state sequences that result in an exponential computation complexity. Alternatively, an efficient computational algorithm [15] called forward algorithm is reported to have a computational complexity N^2T rather than an exponential complexity. The forward algorithm has three principal equations which are the initialization of what is called forward variable α , an induction step, and the termination step.

The forward variable $\alpha_t(i)$ is defined as

$$\alpha_t(i) = P(O_1, O_2, \dots, O_t, \quad s_t = i|\lambda). \quad (9)$$

As (9) implies, $\alpha_t(i)$ is the probability that the HMM is in state $S_t = i$, having generated partial observation sequence (O_1, O_2, \dots, O_t) ; the forward variable is initialized using

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N. \quad (10)$$

For $2 \leq t \leq T$, the forward variable is calculated by the induction step

$$\alpha_t(i) = \left[\sum_{j=1}^N \alpha_{t-1}(j) a_{ij} \right] b_j(O_t), \quad 2 \leq t \leq T, \quad 1 \leq j \leq N. \quad (11)$$

$P(O|\lambda)$ is calculated in the termination step using the termination step

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i). \quad (12)$$

Regarding the training step, there is no analytical procedure to re-estimate the model parameters A, B and π which maximizes $P(O|\lambda)$, given any finite observation sequences as the training data. Therefore an iterative algorithm, the ‘‘Baum-Welsh’’, is employed [15] to choose $\lambda = (A, B, \pi)$ such that $P(O|\lambda)$ is only locally maximized. In the ‘‘Baum-Welsh’’ algorithm the estimated model $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$ is better than or equal to the previous model estimate in the sense that $P(O|\bar{\lambda}) \geq P(O|\lambda)$. Thus, the algorithm iteratively uses $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$ in place of $\lambda = (A, B, \pi)$ to repeat the re-estimation process. The probability of O being observed from the model improves until some limiting point is reached. The final estimate is called maximum likelihood of the HMM.

V. VECTOR QUANTIZATION

Since the discrete observation density is used in implementing HMMs, a vector Quantization (VQ) step is required to map the continuous observation (FFT or DWT vectors into a discrete observation sequence). The VQ process entails two major steps [16]: designing the appropriate codebook vector for quantization, and assigning a label to each vector of the input signal from the codebook.

A. Codebook Design

The problem of designing the codebook is formulated as follows:

Given

Training set $\tau = \{x_1, x_2, \dots, x_N\} \in R^K$, where x is the feature vector, and the size of codebook M .

Find

The codebook vector, $C = \{c_1, c_2, \dots, c_M\}$, and the partitions of space, $P = \{S_1, S_2, \dots, S_M\}$, where S_m is the encoding region associated with code vector c_m , which minimize the average distortion index.

$$D_{ave} = \frac{1}{N \cdot K} \sum_{n=1}^N \sum_{m=1}^M g_{n,m} \|x_n - c_m\| \quad (13)$$

where

$$g_{n,m} = \begin{cases} 1, & \text{if } x_n \in c_m \\ 0, & \text{if } x_n \notin c_m \end{cases}$$

such that the following two conditions are fulfilled

$$S_m = \left\{ x : \begin{aligned} &\|x_n - c_m\|^2 \leq \|x_n - c_{m'}\|^2, \\ &\forall m' = 1, 2, \dots, M, \forall n = 1, 2, \dots, N \end{aligned} \right\}. \quad (14)$$

The foregoing condition implies that, the encoding region S_m should consist of all vectors that are closer to c_m than to any of the other codebook vectors, where

$$c_m = \frac{1}{\eta} \sum_{x_n \in S_m} x_n \quad (15)$$

η is the number of the training sequences, which belongs to class c_m , a condition which infers that the code vector c_m should be the average of all the training vectors that are in the encoding region S_m . In this study, the Linde-Buzo-Gray vector quantization (LBG-VQ) technique has been adopted to determine the required codebook size. The details of this algorithm are discussed in [16]. Furthermore, the training sequence is obtained by recording the feature vectors extracted from several different power quality disturbance patterns obtained from both EMTDC/PSACD3.08 simulation and MATLAB6.0 software.

B. Data Labeling

Given

Feature vector set: $\tau = \{x_1, x_2, \dots, x_N\} \in R^K$, codebook consisting of M codeword such that, $\{c_i, 1 \leq i \leq M\}$ obtained from the codebook design stage, and the distance function $d(x, c_i)$ defined between feature vector $x \in R^K$ and codeword $c_i \in R^K$.

Find

TABLE I
 DISTURBANCE SIGNAL MODELING

Disturbance		Controlling Parameters	Equations
Pure Sinusoidal	M1		$v(t) = A \sin(\omega t)$
Sudden Swell	M2	$0.1 \leq \alpha \leq 0.8,$ $T \leq t_2 - t_1 \leq 9T$	$v(t) = A(1 + \alpha(u(t_2) - u(t_1))) \sin(\omega t)$
Sudden Sag	M3	$0.1 \leq \alpha \leq 0.9,$ $T \leq t_2 - t_1 \leq 9T$	$v(t) = A(1 - \alpha(u(t_2) - u(t_1))) \sin(\omega t)$
Harmonics	M4	$0.05 \leq \alpha_3 \leq 0.15,$ $0.02 \leq \alpha_5 \leq 0.1,$ $0.02 \leq \alpha_7 \leq 0.1,$ $\sum \alpha_i^2 = 1$	$v(t) = A(\alpha_1 \sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 \sin(5\omega t) + \alpha_7 \sin(7\omega t))$
Flicker	M5	$0.1 \leq \alpha \leq 0.2$ $0.1 \leq \beta \leq 0.5$	$v(t) = A(1 + \alpha \sin(\beta \omega t)) \sin(\omega t)$
Oscillatory Transient	M6	$0.1 \leq \alpha \leq 0.8$ $0.1ms \leq \tau \leq 0.2ms$ $0.5T \leq t_2 - t_1 \leq 3T$	$v(t) = A[\sin(\omega t) + \alpha e^{-(t-t_1)/\tau} \sin(\omega_n(t-t_1))(u(t_2) - u(t_1))]$

The particular codeword index i^* , which is associated with the codeword, $c_i \in R^k$ that possesses the minimum distance from

$$x : i^* = \arg \min [d(x, c_i)]. \quad (16)$$

Therefore, instead of dealing with the feature vector x , the emphasis is on the index i^* for achieving the optimum matching, which greatly reduces the calculation time.

VI. RESULTS

A. Signal Modeling

To obtain representative signals that mimic the inherent characteristics of the most common power quality disturbances, power quality disturbance signals are initially simulated by MATLAB 6.0. Six disturbance categories are simulated: namely, undisturbed sinusoid, sudden sag, sudden swell, harmonics, oscillatory transient, and voltage flicker. The utilized sampling frequency, 256 samples/cycle, allows for the accurate detection of high frequencies up to approximately 4 KHz. Some unique attributes for each disturbance type are used and allowed to change randomly, within specified limits, in order to create different disturbance cases. The key parameters such as, the disturbance starting time, magnitude, duration, frequency, and damping are implemented. The randomness in the generated signals is intended to test the reliability and accuracy of the utilized HMM, since none of the mentioned attributes is fixed for real distribution system disturbances. Table I provides a detailed summary of all the disturbance types along with the controlling parameters, definitions and equations. As a rule of thumb, the accuracy of the HMMs is increasing proportional to the data size utilized in the training process, that is, proportional to the number of examples/class. However, in order to limit the complexity, only 200 different examples of each disturbance class are generated and utilized. Yet, a large variety of examples are needed to test the proposed algorithm. Therefore, another 200 examples of each disturbance class that are different from the training set are generated randomly for testing the classifier.

B. EMTDC Simulation

The generation of the utilized signals for both the training and testing of the novel HMM using MATLAB6.0 does not take into consideration the noisy environment associated with distribution systems. To carry out a more thorough testing for the proposed HMM under actual signals contaminated with different noise levels, a thirteen-bus industrial distribution system, is employed to generate actual disturbances. The system data is found in [17], where the harmonics are generated by loading the system with various nonlinear loads such as diode bridge rectifiers, phase angle voltage controllers and multiple adjustable speed drives which are placed at different buses in the system. Moreover, fast transients are simulated by switching different capacitors, representing both power factor and voltage profile capacitors that are placed at different voltage levels. Voltage flicker is generated by connecting a simulated arc-furnace. Furthermore, sags and swells are introduced by creating artificial faults at different voltage levels in the distribution system. The measurements from different voltage levels are recorded and then normalized. Two hundred examples/class are stored and used for training and testing the HMMs.

C. Analysis of Results

Computer simulations have been conducted to the clustering of the input data utilizing the vector quantization. Different codebook sizes, namely, 8, 16, 32, and 64 are utilized to test the HMM's classifier. The classification accuracy is found to increase with increasing the codebook size. Moreover, it is observed that saturation is achieved for codebook sizes beyond 64. A further increase in the codebook size does not result in any classification accuracy improvement, and may even worsen it. The results obtained for codebook size 64 are reported, since it yields the best classification accuracy. The results for the six-class classification that employs the HMM algorithm and implements FT and DWT extraction techniques are shown in Table II for the MATLAB-based data, and in Table III for the EMTDC simulation data. The results show a high and comparable accuracy for each extraction tool. The results of

TABLE II
MATLAB-BASED HMM CLASSIFICATION

Class	Fast Fourier Transform (FFT)						Wavelet Transform (WT)					
	M1	M2	M3	M4	M5	M6	M1	M2	M3	M4	M5	M6
Sinusoidal	100 %	0	0	0	0	0	100%	0	0	0	0	0
Sudden Sag	0	97%	0	0	3%	0	0	95%	0	0	5%	0
Sudden Swell	0	0	93%	0	7%	0	0	0	97%	0	3%	0
Harmonics	0	0	0	100%	0	0	0	0	0	100%	0	0
Flicker	0	1%	0	0	99%	0	0	1%	3%	0	96%	0
Transient	0	2%	0	0	0	98%	0	0	0	0	0	100%
Classification Accuracy	97.83%						98%					

TABLE III
EMTDC-BASED HMM CLASSIFICATION

Class	Fast Fourier Transform (FFT)						Wavelet Transform (WT)					
	M1	M2	M3	M4	M5	M6	M1	M2	M3	M4	M5	M6
Sinusoidal	100%	0	0	0	0	0	100%	0	0	0	0	0
Sudden Sag	0	93%	0	0	7%	0	0	94%	0	0	6%	0
Sudden Swell	0	0	96%	0	4%	0	0	0	92%	0	8%	0
Harmonics	1%	0	0	99%	0	0	2%	0	0	98%	0	0
Flicker	0	4%	2%	0	94%	0	0	3%	7%	0	90%	0
Transient	1%	2%	0	0	0	97%	0	0	0	0	1%	99%
Classification Accuracy	96.5 %						95.17 %					

the FFT are comparable to those of WT and may even outperform the latter. Given that the choice of the mother wavelet is optional and may degrade the process of classification, the FFT-based HMM classifier is highly recommended as a base for the scalable power quality disturbance classifier.

D. Discussion

Neural Networks (NN's) are widely viewed as having a desirable topology for power quality disturbance recognition [3], [4]. In this paper, a comparison between the performance of NN-based classifiers and that of HMM-based classifiers for power quality disturbance classification is provided; the comparison considers the following criteria: classification accuracy, training, testing, speed, scalability, and programming complexity.

Classification Accuracy: NN-based classifiers [3], [4] and HMM-based classifiers achieve comparable classification accuracy, which is more than 95%.

Training: With HMMs, a separate model must be trained for each class of the input data; the objective of the training step is to adjust the model parameters in order to obtain the maximum likelihood $P(O|\lambda)$ of the HMM $\lambda(A, B, \pi)$ for the class training data. For NN's, whole data sets are used for training, and the net output is a binary number "0" when the input data does not belong to the class, and "1" when it does belong to the class. It is worth mentioning that, as a special case, modular NN's can be trained in a manner similar to HMM training.

Testing: Although NN's produce a binary decision, HMMs produce a maximum likelihood $P(O|\lambda)$ estimate for each class. This represents a significant advantage over the NN's, since a ranked list of the classification decision based on the output maximum likelihood for each class, is thereby provided.

Speed: The complexity of the viterbi decoding is exponentially proportional to the number of the states in the HMM model. Thus the complexity of the overall recognition task is exponentially proportional to the number of HMM states and linear with the number of reference models (in our case we have one model for each power quality phenomenon, i.e., we have 6 models). This does not represent any practical problem because this number is still quite manageable. However, as the size of the problem grows, the difference in speed between the ANN and the HMM approach becomes increasingly apparent. However there are algorithms to speed up the HMMs in such cases to yield a performance comparable to that of NN's. We have tested one of these approaches, which is a fast match based vector procedure that provides encouraging results [5].

Scalability: The principal advantage of the proposed HMM based classification scheme over the standard back-propagation NN's is its scalability. A new disturbance class can be included readily in the classification procedure by adding an HMM, which is trained only for the newly considered disturbance class. In contrast to the standard back-propagation NN's, the addition of a new disturbance class to an already trained NN implies the retraining of the entire NN.

VII. CONCLUSIONS

The identification and classification of power quality is a vital task for the precise monitoring of distribution systems. In this paper, a new classification methodology based on the use of HMM and VQ algorithms is introduced. The computational effort and the storage requirements in the classification process are decreased. This is due to the utilization of the labels of the quantized vectors, instead of the use of spectral coefficients. Moreover, the use of VQ labels allows classification of fast phenomena such as transients and slow phenomena such as sag. Two different feature extraction techniques, namely, the FFT and the WT have been adopted to extract the salient features of the studied signals. The degree of success of the HMM-based classifier depends mainly on the chosen codebook size with success rate reaching 96.5% for a codebook size equal to 64. The test of the proposed classifier has been preliminary obtained through numerical results and needs to be supported by field-measured data. The classification results show a consistency when both signals are derived from MATLAB and EMTDC simulations. The proposed algorithm demonstrates the potential for the development of fully automated monitoring systems with on-line classification abilities.

REFERENCES

[1] J. Arrillaga, M. H. J. Bollen, and N. R. Watson, "Power quality following deregulation," *Proc. IEEE*, vol. 88, pp. 246–260, Feb. 2000.

[2] T. K. Abdel-Galil, E. F. El-Saadany, and M. M. A. Salama, "Power quality assessment in deregulated power systems," in *Proc. IEEE Power Engineering Soc. Winter Meeting*, vol. 2, New York, Jan. 2002, pp. 952–958.

[3] A. K. Ghosh and D. L. Lubkeman, "The classification of power system disturbance waveforms using a neural network approach," *IEEE Trans. Power Del.*, vol. 10, no. 1, pp. 109–115, Jan. 1995.

[4] 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 Trans. Power Del.*, vol. 15, no. 1, pp. 222–228, Jan. 2000.

[5] T. K. Abdel-Galil, E. F. El-Saadany, A. M. Youssef, and M. M. A. Salama, "On-line disturbance recognition utilizing vector quantization based fast match," in *Proc. IEEE Power Engineering Soc. Summer Meeting*, Chicago, IL, Jul. 21–25, 2002.

[6] A. M. Youssef, T. K. Abdel-Galil, E. F. El-Saadany, and M. M. A. Salama, "Disturbance classification utilizing dynamic time warping classifier," *IEEE Trans. Power Del.*, vol. 19, no. 1, pp. 272–278, Jan. 2004.

[7] T. K. Abdel-Galil, M. Kamel, A. M. Youssef, E. F. El-Saadany, and M. M. A. Salama, "Power quality disturbance classification using inductive inference approach," *IEEE Trans. Power Del.*, vol. 19, no. 4, pp. 1812–1818, Oct. 2004.

[8] S. Santoso, J. Lamoree, W. M. Grady, E. J. Powers, and S. C. Bhatt, "A scalable PQ event identification system," *IEEE Trans. Power Del.*, vol. 15, no. 2, pp. 738–743, Apr. 2000.

[9] M. Kezunovic and Y. Liao, "A novel software implementation concept for power quality study," *IEEE Trans. Power Del.*, vol. 17, no. 2, pp. 544–549, Apr. 2002.

[10] J. Chung, E. J. Powers, W. M. Grady, and S. C. Bhatt, "Power disturbance classifier using a rule-based method and wavelet packet-based hidden Markov model," *IEEE Trans. Power Del.*, vol. 17, no. 1, pp. 233–241, Jan. 2002.

[11] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[12] D. C. Robertson, O. I. Camps, J. S. Mayer, and W. B. Gish, "Wavelets and electromagnetic power system transient," *IEEE Trans. Power Del.*, vol. 11, no. 2, pp. 1050–1058, Apr. 1996.

[13] E. Brigham, *The Fast Fourier Transform*. Upper Saddle River, NJ: Prentice-Hall, 1974.

[14] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[15] L. Rabiner and S. Levinson, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, pp. 257–286, Feb. 1989.

[16] A. Gersho, R. M. Gray, and R. Gallager, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer, Nov. 1991.

[17] E. F. El-Saadany, T. K. Abdel-Galil, and M. M. A. Salama, "Application of wavelet transform for assessing power quality in medium voltage industrial distribution system," in *Proc. Transmission Distribution Conf. Expo.*, vol. 1, Atlanta, GA, Oct. 2001, pp. 427–432.



T. K. Abdel-Galil (M'04) received the B.Sc. and M.Sc. degrees in electrical engineering from Ain Shams University, Cairo, Egypt, in 1992 and 1998, respectively, and the Ph.D. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 2003.

After his graduation, he was a Postdoctoral Fellow with the University of Waterloo, where he was involved in conducting funded research projects related to power quality and distribution system operation. Currently, he is a Research Engineer (III) with the Research Institute-King Fahd University Petroleum and Minerals, Dhahran, Saudi Arabia. His research interests are in the area of the operation and control of distribution systems, power quality analysis, application of artificial intelligence algorithms in power systems, and high voltage and insulation systems.



E. F. El-Saadany (M'98) received the B.Sc. and M.Sc. degrees from Ain Shams University, Cairo, Egypt, and the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada, all in electrical engineering, in 1986, 1990, and 1998, respectively.

He is an Assistant Professor in the Department of Electrical and Computer Engineering, University of Waterloo. His interests include the operation and control of distribution systems, power electronics application in power systems, power quality analysis, and mechatronics.

A. M. Youssef received the B.Sc. and M.Sc. degrees from Cairo University, Cairo, Egypt, in 1990 and 1993, respectively, and the Ph.D. degree from Queens University, Kingston, ON, Canada, in 1997.

He worked for Nortel Networks from 1997 until mid-1999, and then joined the Combinatorics and Optimization Department (Center for Applied Cryptographic Research) at the University of Waterloo, ON, until the end of 2001. He is currently an Associate Professor with Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada.



M. M. A. Salama (F'02) received the B.Sc. and M.Sc. degrees in electrical engineering from Cairo University, Cairo, Egypt, in 1971 and 1973, respectively, and the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada, in 1977.

Currently, he is a Professor in the Electrical and Computer Engineering Department at the University of Waterloo. His research interests include the operation and control of distribution systems, power-quality analysis, artificial intelligence, and electromagnetic field. He is consulted widely with govern-

ment agencies and electrical industry.
Dr. Salama is a registered professional engineer in the Province of Ontario.