

A Decision Tree and S-Transform Based Approach for Power Quality Disturbances Classification

A. Rodriguez¹, E. Merino, J. Aguado, J.J. Lopez, F. Muñoz, F.I. Martin, J. Muñoz

Electrical Engineering Department, University of Malaga
Dr. Ortiz Ramos s/n, 29071-Malaga, Spain

¹ ale@uma.es

Abstract— In this paper, it is presented an automated classification based on S-transform as feature extraction tool and Decision Tree as algorithm classifier. The signals generated according to mathematical models, including complex disturbances, have been used to design and test this approach, where noise is added to the signals from 40dB to 20dB.

Finally, several disturbances, simple and complex, have been considered to test the implemented system. Evaluation results verifying the accuracy of the proposed method are presented.

I. INTRODUCTION

Power system perturbations can lead to failures or malfunctions of the loads connected to the power system and thus incurring in economical costs for consumers. One alternative of improving the power quality (PQ) parameters consists of analyzing these disturbances efficiently and understanding them deeply [1]. Many recent approaches in PQ monitoring try to achieve it through the automated classification of different disturbances.

Two major aspects have to be taken into account in an automated classification system, Fig. 1. The first focusing in obtaining a suitable pattern that allow distinguish clearly each disturbance, by the use of time-frequency transforms for feature extraction, as Wavelet transform (WT) [2] and S-transform (ST) [3]. The second is oriented to use a classifier able to assign each disturbance correctly in its class, so the most of the artificial intelligent techniques have been combined with WT or ST, as Artificial Neural Networks (ANN) [4],[5], Decision Tree [6],[7], Fuzzy Logic [8], Hidden Markov Model [9] or Support Vector Machines [10],[11].

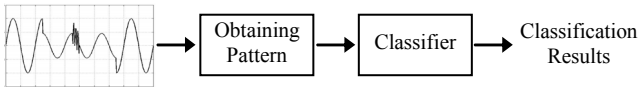


Fig. 1. Automated classification scheme.

Many works consider power quality disturbances as signals that contain a unique given event in a time interval. In electrical systems several disturbances are usually presented consecutively in a very short interval of time, but may do so simultaneously.

In this work complex disturbances have been considered as combinations of two simple ones, simultaneously or consecutively in the same interval. The ST has been chosen for feature extraction, and the pattern is obtained from ST

contours looking for distinctive features of each disturbances. DT has been elected as classifier, using C5.0 algorithm.

Simple and complex disturbance signals have been considered and noise is added to the signals so that the resulting signal to noise ratio (SNR) ranges from 40dB to 20dB.

This paper is organized as follows. Section II presents a simplified outline of ST. Section III is devoted to the pattern design, describing the pattern obtained using the ST in the feature extraction. Section IV briefly defines the DT used. In fifth section the classification results with simple and complex disturbances generated by mathematical model are presented. In section VI different disturbances are generated by simulation software to test the behaviour of the system with real signals. The last section draws conclusions from the results.

II. S-TRANSFORM

The S-Transform [12] is a powerful mathematic tool for time-frequency analysis generated by the combination of WT and STFT. The S-transform of the signal $x(t)$ is defined as:

$$S(\tau, f) = \int_{-\infty}^{\infty} x(t) g_f(\tau-t) \exp(-j2\pi ft) dt \quad (1)$$

Where

$$g_f(\tau-t) = \frac{|f|}{\sqrt{2\pi}} \exp\left(-\frac{f^2(\tau-t)^2}{2}\right) \quad (2)$$

is the Gaussian window function, τ is a shift parameter for adjusting the position in the time axis and f is the scale parameter. $X(f)$ is the Fourier transform of $x(t)$. The relationship between the S-transform and the Fourier transform is:

$$X(f) = \int_{-\infty}^{\infty} S(\tau, f) d\tau \quad (3)$$

The discrete ST is defined as follows:

$$S\left(\frac{n}{NT}, jT\right) = \sum_{m=0}^{N-1} X\left(\frac{m+n}{NT}\right) \cdot \exp\left(-\frac{2\pi^2 m^2 k^2}{n^2}\right) \exp\left(\frac{i2\pi nj}{N}\right) \quad (4)$$

The FFT of the original signal is computed for obtaining the ST [12].

The output of the ST is a complex matrix whose rows and columns values are frequency and time values, respectively. Each column represents the local spectrum in time. Frequency-time contours with the same amplitude in the time-frequency spectrum are also obtained. The information contained in the time-frequency spectrum is used in this paper to detect and characterize power disturbance events.

The graphical output of the ST yields frequency-time, amplitude-time and frequency-amplitude plots. An example of multi-resolution ST analysis for an electric signal containing an oscillatory transients and fifth harmonic is presented in Fig. 2. In the figure it is depicted the disturbance signal (a), the time-frequency contours (b), and 3-D mesh giving amplitude, frequency and time plots (c).

III. OBTAINING PATTERN

The observation of ST contours allows to obtain some efficient characteristic features for simplifying recognition of disturbances. Some examples of signals with different disturbances are analyzed below in order to illustrate the proposed pattern.

A sinusoidal signal with an oscillatory transient and a fifth harmonic is depicted in Fig. 2, with the time-frequency contours and the three dimensional plot time-frequency-amplitude. One can observe the energy in different zones in the time-frequency plot.

In Fig. 3 it is depicted a sinusoidal signal with a voltage sag and several frequency contours. The contour corresponding to 50 Hz is noted to decrease its value during the voltage sag. In the case of an interruption (not shown), a similar behaviour can be observed, but with a deeper diminution. The frequency contours corresponding to other frequencies present a low energy value at the beginning and end of the sag.

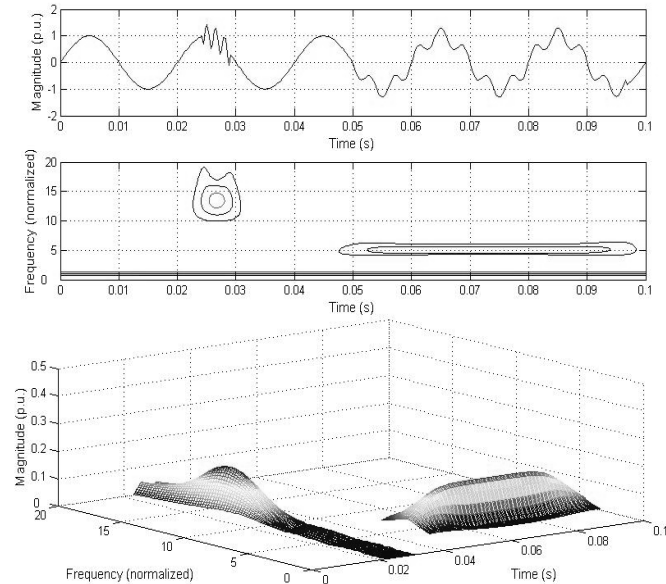


Fig. 2. Signal with an oscillatory transient and 5th harmonic content (a). S-transform Time-Frequency contours (b). 3D mesh Time-Frequency-Amplitude (c).

A signal with an oscillatory transient and 5th harmonics is shown in Fig. 4 (a) and the 150 Hz, 250 Hz, 350 Hz and 700 Hz contours in (b) to (e), respectively. A big amount of energy can be noted in the contours corresponding to the fifth harmonic and the frequency present in the transient signal, 700 Hz in this particular case.

A signal with an oscillatory transient and 5th harmonics is depicted in Fig. 5 (a) and the 150 Hz, 250 Hz, 350 Hz and 700 Hz contours in (b) to (e), respectively. A big amount of energy can be noted in the contours corresponding to the fifth harmonic and the frequency present in the transient signal, 700 Hz in this particular case.

The extraction of characteristic features of the different disturbances can be done taking advantage of the ST contours. In the examples shown above one can see the way of searching for specific frequency disturbances such as any order harmonics or particular oscillatory transients. The results shown above have been taken into account in the election of the characteristic features and in the elaboration of the strategy of classification.

The 50 Hz contour has been proven to contain valuable information about sags, swells and interruptions. Therefore, the mean value of the 50 Hz contour has been taken as a distinctive feature. But this value does not clearly discriminate among sags and interruptions, and therefore the minimum value of the 50 Hz contour, which gives an idea of the severity of the disturbance, has been taken as the second feature.

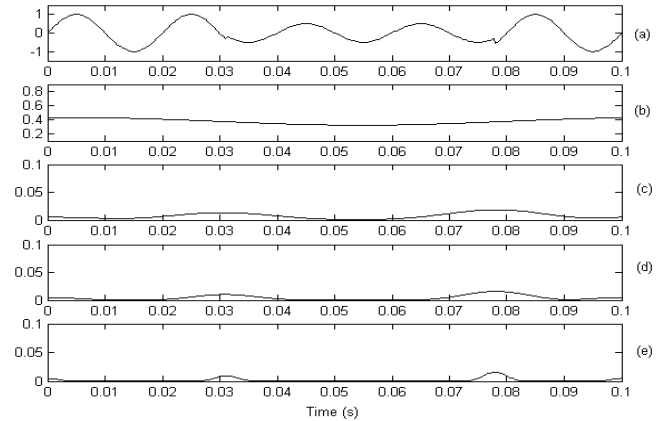


Fig. 4. Voltage sag (a). 50 Hz contour (b). 150 Hz contour (c). 250 Hz contour (d). 700 Hz contour (e).

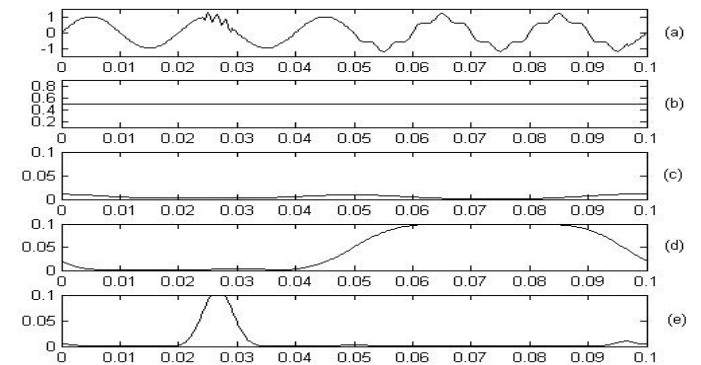


Fig. 5. Signal with an oscillatory transient and 5th harmonic (a). 50 Hz contour (b). 150 Hz contour (c). 250 Hz contour (d). 700 Hz contour (e).

The 50 Hz contour has been proven to contain valuable information about sags, swells and interruptions. Therefore, the mean value of the 50 Hz contour has been taken as a distinctive feature. But this value does not clearly discriminate among sags and interruptions, and therefore the minimum value of the 50 Hz contour, which gives an idea of the severity of the disturbance, has been taken as the second feature.

For detecting disturbances with presence of the third, fifth and seventh harmonics, the energy of these harmonics (150 Hz, 250 Hz and 350 Hz contours) are used as distinctive features. This approach could be extended to other harmonics within the Nyquist condition.

The sum of the energies from 600 Hz to the Nyquist frequency (1600 Hz) has been taken as another characteristic feature. A high value of this energy gives information related to high frequency transient events.

The distinctive features used in this automated classification system is listed below:

- F1: Mean of the fundamental frequency contour (50 Hz)
- F2: Minimum of the fundamental frequency contour
- F3: Energy of the 3rd harmonic contour (150 Hz)
- F4: Energy of the 5th harmonic contour (250 Hz)
- F5: Energy of the 7th harmonic contour (350 Hz)
- F6: Sum of energy from 600 to 1600 Hz contours

IV. DECISION TREE

Decision tree is one of the most widely used classifications methods. There are many decision tree algorithms, such as CART, ID3, C4.5, etc. The C5.0 algorithm as successor of ID3, uses the concept of information gain ratio to build the decision tree. It consists in two distinct phases, build and pruning phases.

The information gain can be described as effective decrease in entropy and can be calculate [13]. So for a data set S of instances from m classes C1, C2 ... Cm, the information (entropy) is describes in (5), where pk is the probability of Ck in S [14].

$$information(S) = - \sum_k^m p_k \cdot \log_2(p_k) \quad (5)$$

In a classification problem of m classes, given a set of training instances S, for a test y with n outcomes, the entropy is given in (6).

$$information_y(S) = \sum_{j=1}^n \frac{|T_j|}{|T|} \cdot info(T_j) \quad (6)$$

Information gain is calculated by (7).

$$gain(Y) = information(S) - information_y(S) \quad (7)$$

Split information is calculated by (8)

$$split_information(Y) = - \sum_{j=1}^n \frac{|T_j|}{|T|} \cdot \log_2 \left(\frac{|T_j|}{|T|} \right) \quad (8)$$

Information gain ratio is determined by (9), the split with maximum gain ratio selected.

$$gain_ratio(x) = \frac{gain(Y)}{split_information(Y)} \quad (9)$$

Power quality disturbances can be classified using a deterministic decision tree by sorting them from the root to some leaf node; this node identifies a class value (disturbance). A decision tree is form, after calculating the gain ratio, the tree can be represented by If...then rules, with these rules, classifications can be implemented easily.

For each branch node, the left child node corresponds to the points that satisfy the conditions, and the right child node corresponds to the points that do not satisfy the condition. The attributes or features used for train the decision tree are energy at different resolution level for different power quality disturbances.

In this case 90% of inputs data are considered for tree decision training and 10% rest are using for testing. Training generates a model with error; this model is applied to test data for further analysis and observes if error produced can be neglected.

The automated classification system proposed using DT as classifier is shown in Fig. 5.

V. CLASSIFICATION RESULTS

The algorithm selected has to be trained. The training stage consists of adjusting the function that assigns each data to its correct class with the minimum possible error. Once the classifier has been adjusted, the results are validated with a different data set, in order to test the effectiveness of the whole system. For this purpose, a dataset with the correct mapping to the corresponding class is used for this process, usually 90% for the training stage and 10% for the test one. In this work the data set has 11000 signals.

More usual power quality disturbances have been considered such as sags, interruptions (Int), swells, oscillatory transients (OT) and harmonics (Harm) moreover signals considered as normal voltage (Normal), besides different complex power quality disturbances. In real power systems the disturbances often occurs consecutively in a very short interval of time but may do so simultaneously.

These signals has been generated following mathematical models, simulated in Matlab [15] with a fundamental frequency of 50 Hz and a virtual voltage of 1V. The signals have a length of 100 ms and the sampling frequency is 3.2 kHz. The classification results are presented in Table 1. As can be observed the classification results are quite robust against the noise.

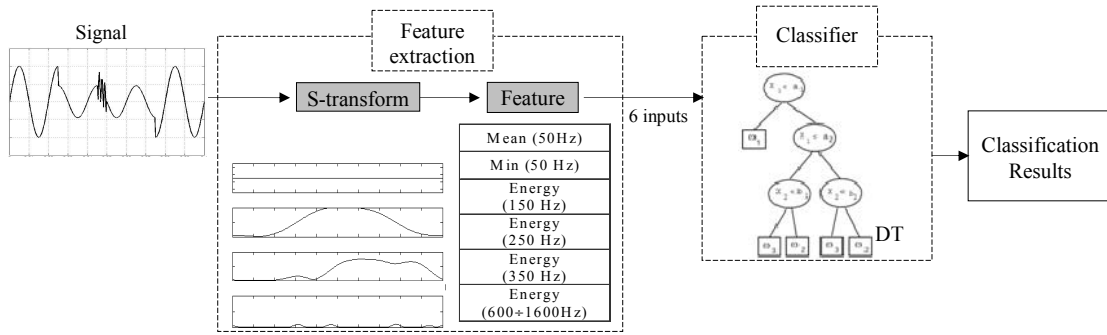


Fig. 5. Automated classification system.

VI. CHECKING THE CLASSIFICATION SYSTEM

In this section a simulation software, PSCAD [16], has been used to generate a data set of distorted signals. So a completely different power quality disturbances has been used to check the implemented system.

Diverse electrical systems have been simulated to obtain power quality disturbances and performing multiple simulations can be obtained several signals with different characteristics, e.g. the initial phase of disturbance, from the same system.

One of the case examples consists of a power system where a single phase fault occurs in the middle of one of the transmission lines as shown in Fig. 7.

In Fig. 8 (a) can be seen one of the signals generated for the previous system, where it can be observed an oscillatory transient produced by the fault, approximately at 0.014 s., a voltage sag during two cycles, and other oscillatory transient produced at the clear of the fault.

The fundamental frequency ST contour of simulation signal and a normal signal to 0.9 one is showed in (b). The 700 Hz ST-contour of the simulation signal and an oscillatory transient is depicted in (c). It can be observed that the simulation signal contours advise the presence of voltage sag, the 50 Hz ST contour is below normal voltage to 0.9, moreover an oscillatory transient; the energy in 700 Hz ST contour is higher than oscillatory transient used as reference.

TABLE 1
CLASSIFICATION RESULTS USING DT

	No noise	40 dB	30 dB	20 dB
Normal	0.99	0.99	0.98	0.96
Sag	0.97	0.96	0.92	0.84
Int	0.87	0.85	0.93	0.93
Swell	1	1	1	0.95
OT	1	1	1	1
Harm	1	1	1	0.98
Harm + Sag	0.99	0.99	1	1
Harm + Swell	1	1	1	1
OT + Sag	1	1	1	1
OT + Swell	1	1	1	1
Harm + OT	1	1	0.99	1
Total	0.984	0.980	0.983	0.969

Others systems have been simulated using PSCAD and the power quality disturbances obtained are showed in Table 3, which also indicates the number of signals.

These signals generated by a simulation program based in electrical models allow to check the behaviour of the implemented system because this set of signals is completely different of that one obtained by mathematical models.

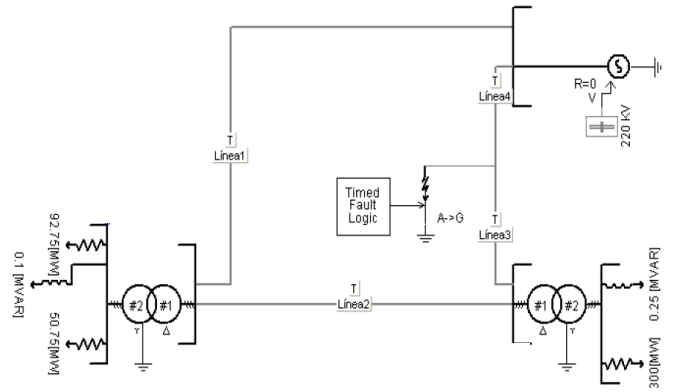


Fig. 7. Electrical system scheme in PSCAD.

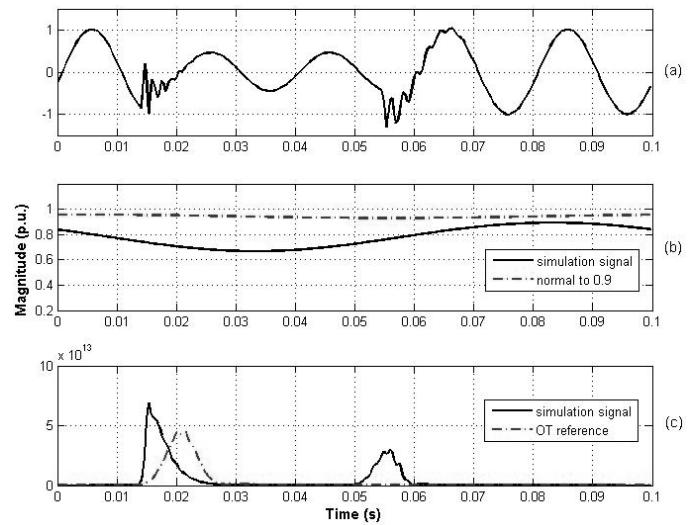


Fig. 8. Comparison of ST contours.

VII. CONCLUSION

An automated classification system has been implemented, based on a reduced and simple set of features extracted from the ST in order to advance accuracy, simplicity and reliability. Even in the presence of complex disturbances with different levels of noise, these features characterize the signals successfully.

Robust results have been achieved even in the presence of complex disturbances. The implemented system has been compared against signals generated by simulation software.

TABLE 3

CLASSIFICATION OF DISTURBANCES OBTAINED BY SIMULATION SOFTWARE

Disturbance	Number	Accuracy
Normal	10	1
Sag	42	0.643
Int	4	0.5
Swell	16	1
OT	30	1
Harm	30	1
Harm + Sag	20	1
Harm + Swell	30	1
OT + Sag	28	0.714
OT + Swell	-	-
Harm + OT	-	-
Total	210	0.881

REFERENCES

- [1] A. Rodriguez, J. Aguado, J.J. Lopez, F.I. Martin, F.J. Muñoz, J.E. Ruiz, "Time-frequency comparisons for power quality analysis," in *Power Quality*, Intech, ISBN 978-953-307-180-0.
- [2] T. Radil, V. Matz, F.M. Janeiro, P. Ramos, A. Cruz, "On-line Detection and Classification of Power Quality Disturbances in a Single-phase Power System," *Power Engineering, Energy and Electrical Drives*, pp. 713 – 718, Setubal, Portugal, 2007
- [3] A. Rodriguez, J.E. Ruiz, J.A. Aguado, F.I. Martin, J.J. Lopez, F. Muñoz, "Classification of Power Quality Disturbances using S-transform and Artificial Neural Networks," *Power Engineering, Energy and Electrical Drives*, pp. 1-6, May. 2011.
- [4] D. Borrás, M. Castilla, N. Moreno, J.C. Montano. "Wavelet and neural structure: a new tool for diagnostic of power system disturbances", *IEEE on Industry Applications*, vol. 37, pp. 184-190, Jan. 2001.
- [5] S. Mishra, C. N. Bhende, B. K. Panigrahi, "Detection and Classification of Power Quality Disturbances Using S-Transform and Probabilistic Neural Network", *IEEE Transactions on Power Delivery*, vol. 23, Jan. 2008.
- [6] T.K. Abdel-Galil, M. Kamel, A.M. Youssef, E.F. El-Saadany, M.A. Salama, "Power quality disturbance classification using the inductive inference approach", *IEEE Transactions on Power Delivery*, vol. 19, pp. 1812-1818, Jan. 2004.
- [7] O. Ozgonenel, D.W.P. Thomas, T. Yalcin, Superiority Of Decision Tree Classifier On Complicated Cases For Power System Protection, The 11th International Conference on Developments in Power System Protection, 23-26 April 2012, The ICC, Birmingham, UK.
- [8] M.V. Chilukuri, P.K. Dash, "Multiresolution S-Transform-Based Fuzzy Recognition System for Power Quality Events", *IEEE Transactions on Power Delivery*, vol. 19, pp. 323-330, Jan. 2004.
- [9] T.K. Abdel-Galil, E.F. El-Saadany, A.M. Youssef, and M. M. A. Salama, "Disturbance Classification Using Hidden Markov Models and Vector Quantization" *IEEE Transactions on Power Delivery*, vol. 20, pp. 2129-2135, Jul 2005.
- [10] G.S. Hu, J. Xie, F.F. Zhu "Classification of power quality disturbances using wavelet and fuzzy support vector machines", *Proceedings of International Conference on Machine Learning and Cybernetics*, vol. 7, pp. 3981-3984, Aug. 2005.
- [11] O. Ozgonenel, T. Yalcin, I. Guney, U. Kurt. "A New Classification for Power Quality Events in Distribution System," *Electric Power System Research*, vol. 95, pp. 192-199, 2013.
- [12] R.G. Stockwell, L. Mansinha, R.P. Lowe. "Localization of the Complex Spectrum: The S Transform", *IEEE Transactions on Signal Processing*, vol. 44, pp. 998-1001, 1996
- [13] Ying-Hui Kong, Jin-Sha Yuan, Jing An, Lin-Lin Che. "Online Power Quality Disturbances Detection and Classification Using One-Pass Wavelet Decomposition and Decision Tree," *Proceeding of Sixth International Conference on Machine Learning and Cybernetics*, Vol. 5, pp. 2990-2995, Aug. 2007.
- [14] Sub Wei, Huang Wen-fang, Yan Gui, Dong Li-Fang. "Classification for Power Disturbances based on Cubic B-spline Wavelet and Decision Tree". *International Conference on Computer Science and Software Engineering*, Vol. 1, pp. 823-826, Dec. 2008.
- [15] MATLAB. Natick, MA: Math Works, Inc., 2000.
- [16] PSCAD. Manitoba HVDC Research Centre, Inc., 2005