A NOVEL WAVELET-BASED SCHEME FOR FAULT DETECTION

J.A.M. Felippe de Souza

Universidade da Beira Interior
Portugal

R.K.H. Galvão and T. Yoneyama

Inst. Tecnológico de Aeronáutica
Brazil

ABSTRACT

This paper presents a novel method for fault detection based on the Wavelet Transform. The proposed technique is similar to the classic method of band-limiting filters, however it is more flexible and less heuristic. Simulation results for fault detection in a servomechanism are presented, the wavelet approach being compared favourably to a standard observer-based scheme.

KEYWORDS

Fault Detection, Analytical Redundancy, Wavelet Transform.

INTRODUCTION

In a wide sense, a fault can be understood as an unexpected change in the behavior of a system. Prompt detection of such occurrence is essential to prevent further deterioration, which could lead to dangerous operating conditions and even physical breakdown.

Early fault detection (FD) schemes resorted mainly to *physical redundancy*, i.e., cross-comparison of measurements from redundant sensors. However, in the early 70's, the increase in computational power made possible the paradigm of *analytical redundancy* [2]. In this case, signals generated by a mathematical model of the monitored system are compared with actual measurements, the differences (residues) being used to detect faults (fig.1).

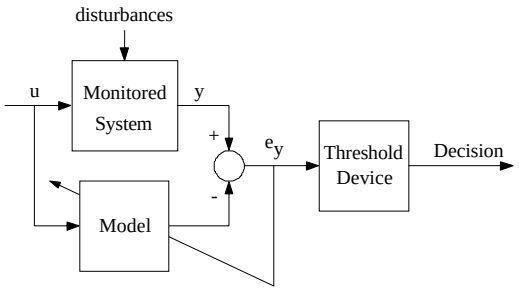

Figure 1: Model-based fault detection scheme. Signal e_y is called residue.

In recent years, FD methods have also benefited from exploiting the decomposition of signals by joint time-frequency analysis [4]. Spectral analysis is important for detecting the periodicity of features in a signal. However, faults are usually non-repetitive at their earlier stages, which means that faults detectable by spectrum analysis alone may be already much advanced. Also, analysis in the frequency domain does not cope well with high frequency bursts associated with some types of faults. Thus, time information, as well as spectral information, should be employed by efficient FD algorithms. In this context, the Wavelet Transform (WT) has shown to be a useful tool, since it provides efficient ways to perform non-stationary analysis and also filter measurement noise [1],[10],[4].

The technique proposed in this paper employs the WT to revitalize the classic method of band-limiting filters [5], which is made more flexible and less heuristic. A preliminary "tuning" phase is required, but the computational workload involved is modest.

Simulation results of fault detection in a servomechanism are presented, the wavelet approach being compared favourably to a standard observer-based one.

THE WAVELET DECOMPOSITION OF A SIGNAL

The WT is a time-frequency or, strictly speaking, time-scale analysis tool. It is obtained as the inner product of a signal $f(t)$ with analyzing functions $\psi_{a,b}(t) = \frac{1}{|a|^{1/2}}\psi(\frac{t-b}{a})$, whose width and center position vary according to parameters a and b. The WT is thus a function of two variables, as shown in eq. 1.

$$(T^{wav}f)(a,b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t)dt \qquad (1)$$

If function $\psi(t)$ satisfies the condition:

$$\int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|^2}d\omega < \infty \qquad (2)$$

where the hat denotes the Fourier transform, then the transformation given by eq. 1 is invertible. The set $\{\psi_{a,b},\ a \in \mathbb{R}^*,\ b \in \mathbb{R}\}$ is then called a family of wavelets, derived from the so-called mother wavelet $\psi(t)$ through dilation and time-shifting operations. It is interesting to note that eq. 2 implies $\hat{\psi}(0) = 0$, i.e., the mother wavelet should be an oscillation of zero mean. Besides that, $\psi(t)$ is usually chosen to

have good localization in time (i.e., compact support or rapid decay), so that the WT can reflect not only the frequency content of the analyzed signal, but also its evolution in time.

Parameter a is also called *scale*, and is responsible for changing the central frequency of the wavelet. Note that, by the "Uncertainty Principle" [8], there is always a compromise between time and frequency resolutions. Decreasing a compresses the wavelet in time (increasing time resolution), but spreads it in frequency (losing frequency resolution).

The representation obtained from eq. 1 is highly redundant, since parameters a and b vary continuously in $\mathbb{R}$. Hence, a subset of scales and translations is usually used, leading to the Discrete Wavelet Transform (DWT). It can be shown [1] that, sampling a and b in a dyadic grid, that is,

$$a = 2^m \qquad b = n2^m \tag{3}$$

there is no loss of information. Substituting Eq. (3) in eq. 1, one can write:

$$\begin{aligned} (DWTf)(m,n) &= (T^{wav}f)(2^m, n2^m) \\ &= d_{m,n} \quad (m,n \in \mathbb{Z}) \end{aligned} \tag{4}$$

Connections between the Wavelet Transform and Subband Coding [9] allowed the development of fast algorithms for the DWT, which use banks of digital filters in a tree structure, as shown in fig. 2. In this figure, ($\downarrow$ 2) denotes the downsampling operation, which consists of removing each other sample from a sequence. It is the digital counterpart of the time compression and also the key to avoid redundancy and increase computational speed. Note that the downsampling operator is linear, but *it is not shift-invariant.*

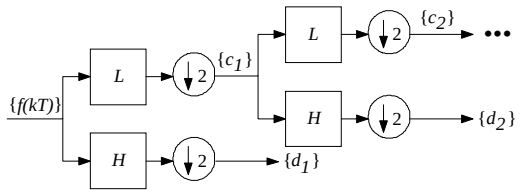

Figure 2: Wavelet decomposition tree. H and L are, respectively, highpass and lowpass filters. T is the sampling period.

Coefficients $\{c\}$ and $\{d\}$ are related, respectively, to approximations and details of the signal at different resolution levels. Since wavelets have a bandpass nature, the details can be regarded as frequency slices of the analyzed signal, each situated approximately between $2^{-m}f_s$ and $2^{-m-1}f_s$ ($m = 1, 2, ...$), where $f_s = 1/T$ is the sampling rate [10].

THE PROPOSED TECHNIQUE

A classic approach to fault detection consists of comparing, over an adequate frequency band, signals measured at different points of the system. The basic premise is that, within a narrow frequency band, there may be simple dynamic relationships between signals that are normally very different when viewed over a wide bandwidth. So, if conveniently selected passband filters are used to process input and/or sensor data, the difference between the outputs of the filters will form a residue. The residue magnitude exceeding some threshold will indicate modifications in the information transmission path, i.e., faults. Fig. 3 illustrates a typical configuration.

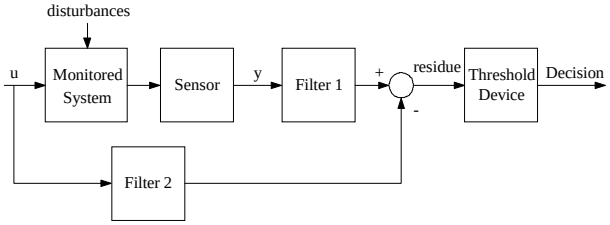

Figure 3: Input-output consistency check employing band-limiting filters.

As reported in [5], this method deals satisfactorily with transient inputs (either commands or unmeasured disturbances), which are known to be false alarm sources.

The theoretical basis of the band-limiting filters approach lies on Gabor's minimum-uncertainty cells [3]. In fact, the output of filters 1 and 2 in fig. 3 should be matched, not only in frequency, but also in time, that is, their time-frequency cells should be coincident, whatever the inputs to the system. This remark allows the establishment of a bridge to Wavelet Theory.

Due to the passband nature of the wavelet transform, it can replace advantageously the conventional filters employed here, as shown in fig. 4 (henceforth, only stable, minimum-phase linear-time-invariant systems will be considered).

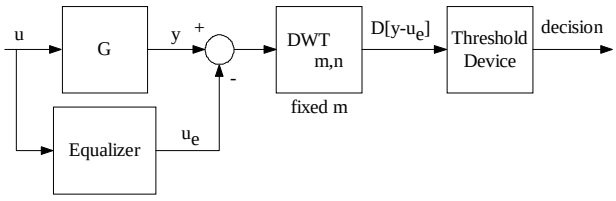

Figure 4: Wavelet-based input-output consistency check.

In this figure, G is a transfer function incorporating the plant and the sensor, and DWT represents the Discrete Wavelet Transform, implemented by a filter bank.

For ease of notation, $\{Df[n],\ n \in \mathbb{Z}\}$ will henceforth represent the coefficients of f, *obtained via the filter bank*, at the fixed level of analysis

m ($m = 1, 2, ...$). Also, $D[f+g]$ will denote the coefficients of $(f+g)$. Finally, G_B will designate the transfer function G after band-limiting, that is:

$$G_B(j\omega) = G(j\omega)\beta_m(\omega) \tag{5}$$

where β_m is a function of effective support $[2^{-m-1}(2\pi f_s), 2^{-m}(2\pi f_s)]$, whose characteristics depend on the mother wavelet employed.

It should also be noted that monitoring can be carried out in the wavelet coefficient domain because reconstruction from the wavelet coefficients is numerically stable [1]. So, denoting by $||\cdot||$ the ℓ^2 norm of a sequence, then $||D[y-u_e]||$ is large if and only if $||y-u_e||$ is large at the frequency band of analysis[1].

The Equalizer

The equalizer block in fig. 4 is a digital lowpass or highpass filter of the form:

$$\begin{aligned} Eq(z) &= \left(\frac{1}{z-1}\right)^s (\alpha + \beta z^{-1}) \\ s &\in \mathbb{Z};\ \alpha, \beta \in \mathbb{R} \end{aligned} \tag{6}$$

Parameters s, α, β should be chosen in order to minimize

$$|G_B(j\omega) - Eq(j\omega)| \tag{7}$$

for the frequencies ω in the analysis band. This can be achieved if the frequency response of the equalizer matches the magnitude and slope of $G_B(j\omega)$. Positive slopes require highpass equalization and negative slopes, lowpass equalization.

Parameter s reflects the estimated slope $\frac{d}{d\omega}|G_B(j\omega)|$: for a slope of λ $db/decade$, one should take $s = round\,(-\lambda/20)$. Parameters α and β are responsible for magnitude matching and also for fine adjustments in the slope.

Fig. 5 depicts the process of adjusting the equalizer parameters[2].

To see how α and β can be determined, suppose initially $s = 0$, i.e., $|G_B(j\omega)|$ approximately flat within the analysis band. In this case, $u_s = u$. Now, the problem consists of minimizing the error e_D, defined as:

$$e_D \triangleq D[y - u_e] \tag{8}$$

To minimize e_D in a least-squares sense, let the cost function be:

$$J = \frac{1}{2}||e_D|| \tag{9}$$

[1] If $\{f[kT], k \in \mathbb{Z}\}$ belongs to $\ell^2(\mathbb{Z})$, then Df also belongs to $\ell^2(\mathbb{Z})$, which allows the use of the ℓ^2 norm. This finite-energy requirement is satisfied automatically by the use of finite-time observation windows.

[2] From this point on, by an abuse of notation, z^{-1} will denote the operator unit time delay. Also, the context will make it clear whether f denotes the sequence $\{f(kT), k \in \mathbb{Z}\}$ or its $\mathcal{Z}$-transform.

Since signals used in the tuning of the equalizer have a finite number of samples, e_D can be regarded as a row vector and $||\cdot||$ as the euclidean norm. As a result, eq. 9 can be rewritten as:

$$J = \frac{1}{2} e_D e_D^T \tag{10}$$

The linearity of the wavelet filter bank yields:

$$e_D = D[y - u_e] = Dy - Du_e \tag{11}$$

and also:

$$Du_e = D[\alpha u + \beta z^{-1} u] = \alpha Du + \beta D[z^{-1}u] \tag{12}$$

Since the wavelet filter bank is not shift-invariant (due to the downsampling operation), $D[z^{-1}u]$ cannot be written in terms of Du. So, $\{z^{-1}u\}$ must be treated as a separate sequence, which will be denoted by $\{u_d\}$ for short. Now, to find α and β that minimize the cost in eq. 10, it suffices to write:

$$\frac{\partial J}{\partial \alpha} = 0 \Rightarrow -(Dy - \alpha Du - \beta Du_d)(Du)^T = 0 \tag{13}$$

$$\frac{\partial J}{\partial \beta} = 0 \Rightarrow -(Dy - \alpha Du - \beta Du_d)(Du_d)^T = 0 \tag{14}$$

or, in matrix form:

$$M \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} Dy(Du)^T \\ Dy(Du_d)^T \end{bmatrix} \tag{15}$$

where:

$$M = \begin{bmatrix} Du(Du)^T & Du_d(Du)^T \\ Du(Du_d)^T & Du_d(Du_d)^T \end{bmatrix} \tag{16}$$

M is non-singular, provided that the input u is sufficiently rich in a sense that will be made precise below. In fact, for ease of notation let $Du = x$ and $Du_d = y$. Then:

$$M = \begin{bmatrix} ||x||^2 & \langle x, y\rangle \\ \langle x, y\rangle & ||y||^2 \end{bmatrix} \tag{17}$$

and

$$\begin{aligned} \det(M) &= ||x||^2 ||y||^2 - \langle x, y\rangle^2 \\ &= ||x||^2 ||y||^2 - [||x||\,||y|| \cos \measuredangle(x,y)]^2 \\ &= ||x||^2 ||y||^2 \left[1 - \cos^2 \measuredangle(x,y)\right] \end{aligned} \tag{18}$$

Thus, M will only be singular if either $||x||$ or $||y||$ are equal to zero (the input signal has no components in the analysis band) or else if $\cos \measuredangle(x,y) = 1$ (x and y are colinear). Colinearity implies that there is a constant $R \in \mathbb{R}$ such that $Du = RDu_d$. Now, denoting by u^B the input signal on the frequency band delimited by operator D, then $Du = RDu_d$ implies $u^B(k) = Ru^B(k-1)$, or $u^B(k) = u^B(0)R^k$.

The computational effort required when solving for α and β is modest, since the number of wavelet

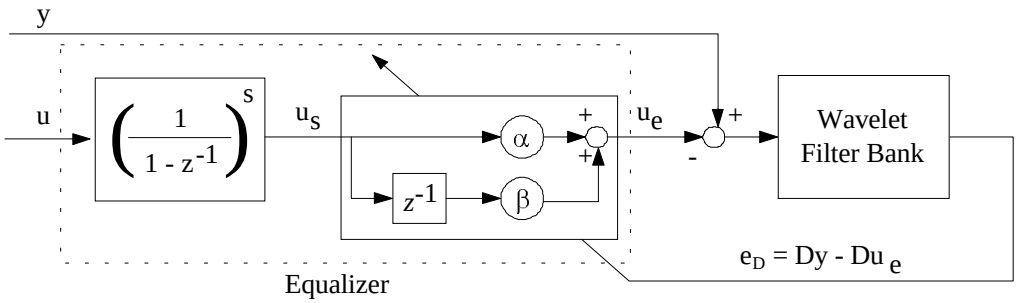

Figure 5: Tuning the equalizer.

coefficients is considerably smaller than the number of samples in the filter bank input.

In the general case ($s \neq 0$), it suffices to employ u_s instead of u. The optimum value for s can be easily obtained by using a search algorithm in $\mathbb{Z}$.

Note that an output (y) equalization scheme might as well be adopted. However, since physical systems usually have a lowpass characteristic, that would lead to a highpass equalizer, undesirably amplifying measurement noise.

AN APPLICATION EXAMPLE

To exemplify the described technique, a velocity servomechanism will be considered. Its transfer function, between the input signal u and the shaft speed ω, is:

$$G(s) = \frac{\Omega(s)}{U(s)} = \frac{2.5}{(s+20)(Is^2 + 55Is + 0.125)}$$

where $I = 5.0 \times 10^{-4} Kgm^2$ is the inertia momentum of the load connected to the servo shaft (the inertia of the servo rotor is supposed much smaller than I). Let us assume that the load consists of two interconnected rigid bodies, each one with an inertia momentum of $0.5I$.

In the simulation, the fault consists of the rupture of the connection between the two bodies, which is modelled by an abrupt 50% reduction in the load inertia.

In normal conditions, $G(s)$ has poles in $-5.0, -20, -50$ and a DC gain of one. The above-mentioned fault changes the poles to $-11, -20, -44$, preserving the DC gain.

Sampling frequency was set to $250Hz$ ($1571rad/s$).

RESULTS

The mother wavelet chosen was $db8$ (from the Daubechies family). The choice of the best wavelet for a given problem is still a subject of much research. As a rule of thumb, wavelets with longer support yield better resolution in frequency (improving the matching process and noise rejection) at the expense of worse time resolution (larger delays in the detection). In this work, the dbN family was chosen *a priori,* the index $N = 8$ yielding a good compromise between noise rejection and time resolution (dbN mother wavelets have a support width of $2N-1$ and associated filters with $2N$ taps).

In the tuning and test phases, measurements were subjected to an additive white gaussian noise with a standard deviation of 10^{-2}.

Results will be evaluated using the following index:

$$\sigma \triangleq \frac{max(abs(residue))_{post-fault}}{max(abs(residue))_{pre-fault}}$$

Equalizer tuning

To adjust the parameters of the equalizer, the servomechanism was excited with a PRBS input during $16.4s$ (4096 sampling intervals), as seen in fig. 6.

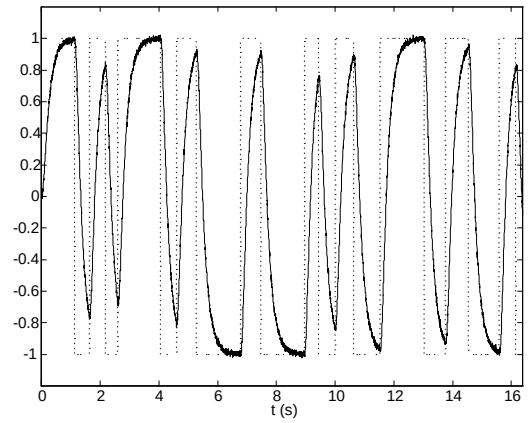

Figure 6: Tuning phase. Dotted line: input u, solid line: output y.

The detail level of the analysis was set to $m = 6$, i.e., the frequency band considered was approximately between $f_s/128$ ($12.3rad/s$) and $f_s/64$ ($24.6rad/s$). Table 1 presents the minimum costs attained for different values of the slope s:

TABLE 1 - Finding the optimum s.

s	0	1	2	3	4
J_{min}	3.9	0.46	1.5	2.6	4.0

As it can be seen, $s = 1$ yields the best result. In this case, the values found for the remaining equalizer parameters are:

$$\alpha = -0.124 \quad \beta = 0.136$$

Fig. 7 illustrates the matching of the wavelet coefficients of the output and equalized input.

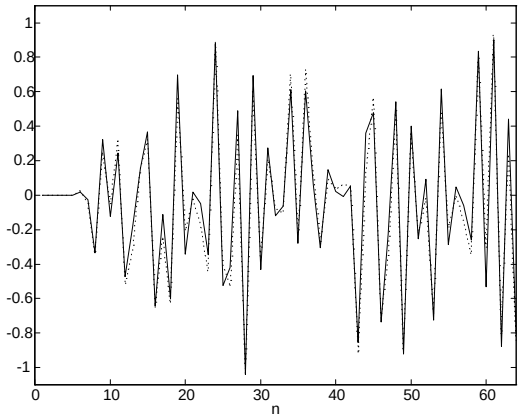

Figure 7: Matching between the wavelet coefficients of the output (solid line) and the equalized input (dotted line).

Fault Detection Test

In the test phase, the servomechanism was excited with a square wave, as seen in fig. 8. A fault occurs exactly in the middle of the test ($t = 8.2s$).

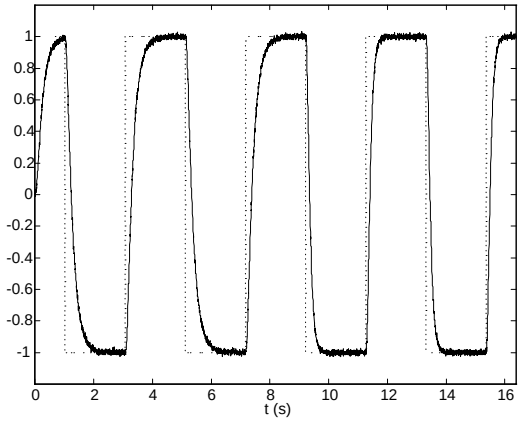

Figure 8: Test phase (a fault occurs at $t = 8.2s$). Dotted line: input u, solid line: output y.

Fig. 9 displays the output of the wavelet fault detector (wavelet coefficients of the difference between the system output and the equalized input). Remark that the first post-fault coefficient corresponds to $n = 33$ (marked by a vertical line). The performance index in this case is $\boldsymbol{\sigma = 5.9}$.

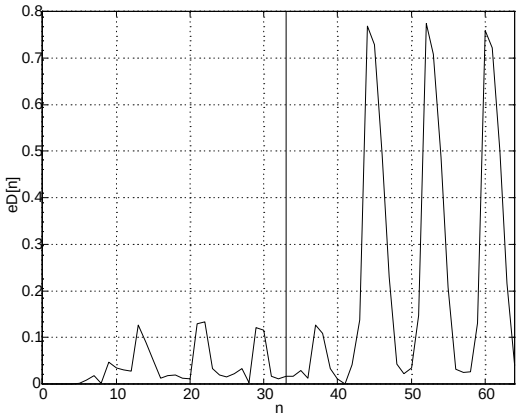

Figure 9: Output of the wavelet fault detector (absolute value).

For means of comparison, a Luenberger observer was used to detect the fault in the configuration seen in fig. 1. Since the fastest pole in the monitored system is -50, the observer poles were all set to -50ρ, $\rho > 1$. The observer employed an exact model of the system and initial errors on state estimation were set to zero.

Table 2 presents the fault detection performance of the observer as a function of the position of its poles. Remark that, due to the adverse effects of measurement noise [7], lower values of ρ yield better results, though still inferior to the one obtained with the wavelet detector.

TABLE 2 - Effect of changing the observer poles.

ρ	1.0	1.1	1.2	1.5	2.0
σ	2.2	1.8	1.6	1.2	1.1

Fig. 10 displays the observer-generated residue for $\rho = 1.0$ (best result for the observer detector). It is important to note that, in a real situation, when only an approximate model is available for the system, the use of larger values for ρ might be required. In this case, the superiority of the wavelet detector would be even greater.

CONCLUDING REMARKS

When compared to conventional Band-Limiting Filters, the wavelet approach is seen to be less heuristic, since parameter tuning can be made automatically, with modest computational effort. That potentially allows an easier on-line reconfiguration of the fault monitor, though this possibility was not exploited in this work. Also, different frequency bands can be easily selected, by changing the level of the

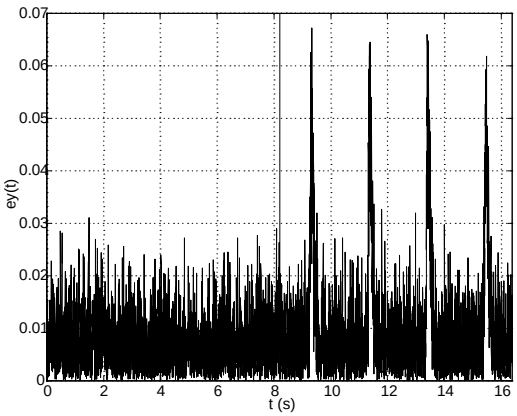

Figure 10: Observer-generated residue (absolute value). Fault onset is marked by a vertical line.

analysis (m). Note that the restriction of constant quality factor (that is, bands situated between $2^{-m-1}f_s$ and $2^{-m}f_s$) can be alleviated by using *wavelet packets*, which allow more generic frequency partitions [9].

The problem of detection delay was not considered in this paper. However, it could be argued that the use of wavelets with better time resolution would result in smaller delays.

Improvements in the wavelet fault monitor might be achieved by using a more complex structure for the equalizer or by performing an adaptation on the wavelet filters themselves [6].

Generalizing this technique to nonlinear systems is an interesting and challenging task. In fact, due to harmonics generation, outputs within a frequency band may be influenced by inputs at several different bands. Thus, to perform input-output matching, it would be necessary to decompose the input signals at several resolution levels, which should then be nonlinearly combined (by means of fuzzy inference, or an artificial neural network, for example).

Future works could attempt to extend the wavelet approach to the Fault Isolation problem. In this case, simultaneous analysis of residues in different frequency bands would be probably required (fig. 11). However, if the effects of different faults on the output frequency spectrum are similar, more information transmission paths may have to be monitored (i.e., more sensors will be needed).

ACKNOWLEDGMENTS

The authors would like to thank FAPESP for funding this research under grant 96/11821-8, and also Prof. Tânia Nunes Rabello for providing suggestions and material related to wavelet theory.

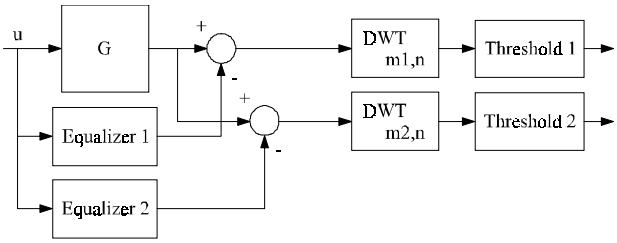

Figure 11: Monitoring two frequency bands.

REFERENCES

[1] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, Philadelphia, 1992.

[2] P. M. Frank. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy - a survey and some new results. *Automatica*, 26(3):459–474, 1990.

[3] D. Gabor. Theory of communication. *Journal of the IEE*, 93:429–457, 1946.

[4] Roberto Kawakami Harrop Galvão. *Wavelet-Based Techniques for Adaptive Feature Extraction and Pattern Recognition.* PhD thesis, ITA, São José dos Campos, 1999.

[5] J. G. Jones and M. J. Corbin. Band-limiting filter approach to fault detection. In R. Patton, P. Frank, and R. Clark, editors, *Fault Diagnosis in Dynamic Systems - Theory and Application*, chapter 6, pages 189–251. Prentice Hall, New York, 1989.

[6] Y. Mallet, D. Coomans, J. Kautsky, and O. De Vel. Classification using adaptive wavelets for feature extraction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(10):1058–1066, Oct. 1997.

[7] R. Patton, P. Frank, and R. Clark, editors. *Fault Diagnosis in Dynamic Systems - Theory and Applications.* Prentice-Hall International, Cambridge, UK, 1989.

[8] S. Qian and D. Chen. *Joint Time-Frequency Analysis - Methods and Applications.* Prentice Hall PTR, Upper Saddle River, 1996.

[9] G. Strang and T. Nguyen. *Wavelets and Filter Banks.* Wellesley-Cambridge Press, Wellesley, 1996.

[10] X. Yibin, D. C. T. Wai, and W. W. L. Keerthipala. A new technique using wavelet analysis for fault location. In *Proc. 6th Int. Conf. On Developments in Power Systems Protection, Nottinghnam, 25-27 Mar 1997*, pages 231–234, London, 1997. IEE.