unesp

UNIVERSIDADE ESTADUAL PAULISTA

Pós-Graduação em Engenharia Mecânica

M.E.S.T.R.A.D.O

1210001256

"Utilização da Transformada Wavelet e Redes Neurais Artificiais para Reconhecimento de Padrões Aplicados à Manutenção Preditiva"

EDMILSON RICARDO DE OLIVEIRA SANTOS

Ilha Solteira - SP

cm 1 2 3 4 5 6 7 8 unesp 12 13 14 15 16 17 18 19

UNESP
UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE ENGENHARIA DE ILHA SOLTEIRA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA

UTILIZAÇÃO DA TRANSFORMADA WAVELET E REDES NEURAIS ARTIFICIAIS PARA RECONHECIMENTO DE PADRÕES APLICADOS A MANUTENÇÃO PREDITIVA

Edmilson Ricardo de Oliveira Santos

1210001256

Dissertação apresentada à Faculdade de Engenharia de Ilha Solteira da Universidade Estadual Paulista "Júlio de Mesquita Filho", como parte dos requisitos exigidos para a obtenção do título de **Mestre em Engenharia Mecânica**

Proc. 071/2002 - NED 16
UNESP - "CAMPUS DE ILHA SOLTEIRA"
SERVIÇO TÉC. DE BIBLIOTECA E DOCUMENTAÇÃO
DATA DE CHEGADA 06/05/02
DATA DE TOMBO 30/05/02
REGISTRO Oferta
TOMBO Te. 1256
AQUISIÇÃO Doação Autor R$ 10,00
CLASSIFICAÇÃO S237u

Orientador: **Prof. Dr. Vicente Lopes Jr.**
Co-orientador: **Prof. Dr. João Antônio Pereira.**

Ilha Solteira, Agosto de 2000

30500001 ENGENHARIA MECÂNICA
sys - 186296
sys - 54802
BCpIS - FEIS - UNESP

"Utilização da Transformada Wavelet e Redes Neurais Artificiais para Reconhecimento de Padrões Aplicados à Manutenção Preditiva"

EDMILSON RICARDO DE OLIVEIRA SANTOS

ESTA DISSERTAÇÃO FOI JULGADA ADEQUADA PARA OBTENÇÃO DO TÍTULO DE

MESTRE EM ENGENHARIA MECÂNICA

NA ÁREA DE CONCENTRAÇÃO EM MECÂNICA DOS SÓLIDOS E APROVADA EM SUA FORMA FINAL PELO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA

Prof. Dr. André Luiz Seixlack - Coordenador

COMISSÃO EXAMINADORA:

Prof. Dr. Vicente Lopes Júnior - Orientador

Prof. Dr. José Maria Campos dos Santos

Prof. Dr. Adyles Arato Júnior

Ilha Solteira – SP, agosto de 2000

BCpIS - FEIS - UNESP

Dedico este trabalho ao meu

pai Juraci e a minha mãe Elza.

BCpIS - FEIS - UNESP

Agradecimentos

Ao Prof. Dr. Vicente Lopes Jr e Prof. Dr. João Antônio Pereira pela orientação;

Ao Prof. Edison Righeto pela amizade e auxílio no estudo de Wavelet;

Ao Prof. Adyles Arato Jr pelas sugestões e auxílio na montagem experimental;

Ao meus amigos Marcos Rogério Barbazelli e Adriano Silva Borges pela amizade e incentivo;

Ao técnico Carlos Santana pela colaboração na montagem experimental;

A todos os professores do Departamento de Engenharia Mecânica de Ilha Solteira.

À CAPES pelo apoio financeiro.

BCpIS - FEIS - UNESP

Sumário

1 INTRODUÇÃO 1

2 REVISÃO BIBLIOGRÁFICA 5

3 FUNDAMENTOS TEÓRICOS 11

3.1 ANÁLISE DE SINAIS ESTACIONÁRIOS 12

3.1.1 SÉRIE DE FOURIER E TRANSFORMADA DE FOURIER 12

3.1.2 TRANSFORMADA DE FOURIER DISCRETA 16

3.2 ANÁLISE DE SINAIS NÃO-ESTACIONÁRIO 19

3.2.1 TRANSFORMADA DE FOURIER COM JANELA 20

3.2.2 TRANSFORMADA WAVELET 21

3.2.3 FAMÍLIA WAVELETS 23

3.2.4 O CONCEITO DE ESCALA 23

3.2.5 RELAÇÃO DE INCERTEZA DE HEISENBERG 25

3.3 WAVELETS ORTOGONAIS 27

3.3.1 EQUAÇÕES DE DILATAÇÃO 27

3.3.2 FUNÇÃO WAVELET 31

3.3.3 PROPRIEDADES DOS COEFICIENTES 35

3.4 TRANSFORMADA WAVELET CIRCULAR 38

3.5 ALGORITMO DE MALLAT (TRANSFORMADA WAVELET DISCRETA) 42

3.6 TRANSFORMADA WAVELET CONTÍNUA 54

4. REDES NEURAIS ARTIFICIAIS 61

unesp

4.1 FUNDAMENTOS DAS REDES NEURAIS ARTIFICIAIS 63

4.2 NEURÔNIO ARTIFICIAL 64

4.3 REDES NEURAIS DE SIMPLES CAMADAS 66

4.4 REDES NEURAIS DE MULTI-CAMADAS 67

4.5 TREINAMENTO DAS REDES NEURAIS 68

4.6 REGRA DELTA 69

4.7 ALGORITMO *BACK-PROPAGATION* 71

5 TÉCNICAS DE REDUÇÃO DE DADOS E APLICAÇÕES 79

5.1 SINAIS DE VIBRAÇÃO SIMULADOS 84

5.2 RESULTADOS 87

5.3 CONSIDERAÇÕES FINAIS 88

6 ANÁLISE EXPERIMENTAL E RESULTADOS 89

6.1 MONTAGEM EXPERIMENTAL 89

6.2 PROGRAMA E DESCRIÇÃO DO FLUXOGRAMA DE AQUISIÇÃO DE DADOS 92

6.3 DESCRIÇÃO DO EXPERIMENTO 94

6.4 SINAIS DE VIBRAÇÃO EXPERIMENTAIS 94

6.5 DISCUSSÃO DOS RESULTADOS 99

7 CONCLUSÃO 103

7.1 SUGESTÕES PARA TRABALHOS FUTUROS 104

REFERÊNCIAS BIBLIOGRÁFICAS 105

Lista de Figuras

Figura 3.1	Funções periódicas.	13
Figura 3.2	Representação gráfica dos coeficientes de Fourier.	14
Figura 3.3	Transformada de Fourier.	15
Figura 3.4	Amostragem de uma função contínua de intervalos regulares no tempo.	16
Figura 3.5	Aproximação envolvida no cálculo dos coeficientes de Fourier de uma série.	18
Figura 3.6	Transformada de Fourier com janela "Short time Fourier Transform".	21
Figura 3.7	Domínio tempo-freqüência (STFT).	21
Figura 3.8	Relação entre freqüência e escala.	22
Figura 3.9	Família Wavelets.	23
Figura 3.10	A escala da TW é implementada pelo parâmetro de escala *a*.	24
Figura 3.11	Elipse de Heisenberg para STFT.	26
Figura 3.12	Elipse de Heisenberg para TW.	26
Figura 3.13	As primeiras sete iterações da construção da função escala, *N*=4.	28
Figura 3.14	Função escala.	29
Figura 3.15	As primeiras oito iterações do desenvolvimento da função Wavelet db4.	34

unesp

Figura 3.16	Wavelet db4 construída a partir da função escala.	34
Figura 3.17	(a) Função escala de Haar ; (b) Função Wavelet de Haar.	38
Figura 3.18	Expansão da função Wavelet de Haar com amplitudes de unitárias.	39
Figura 3.19	Processo em decomposição em vários níveis.	54
Figura 3.20	Sinal temporal.	55
Figura 3.21	FFT do sinal temporal.	55
Figura 3.22	TWC do sinal temporal.	56
Figura 3.23	Diagrama tridimensional da TWC para o sinal temporal.	56
Figura 3.24	TWD do sinal temporal, Fig. 3.20.	57
Figura 3.25	Sinal temporal.	58
Figura 3.26	Mapa tempo-escala (TWC).	58
Figura 3.27	Diagrama tridimensional da TWC para o sinal temporal.	59
Figura 3.28	TWC do sinal temporal.	59
Figura 3.29	TWD do sinal temporal.	60
Figura 4.1	Neurônio biológico.	62
Figura 4.2	Esboço de uma rede neural multi-camadas.	63
Figura 4.3	Neurônio artificial projetado por McCulloch.	64

unesp

Figura 4.4	Funções de Transferência.	66
Figura 4.5	Rede Neural com uma camada.	67
Figura 5.1	Etapas principais do processo de classificação.	80
Figura 5.2	(a) Valores não-correlacionados e (b) Valores correlacionados.	81
Figura 5.3	Linha de regressão para diferentes valores do coeficiente de correlação (ρ).	82
Figura 5.4	Método de redução de dados.	83
Figura 5.5	Sinal Normal.	84
Figura 5.6	Sinal Defeituoso (1).	85
Figura 5.7	Sinal Defeituoso (2).	85
Figura 5.8	Decomposição de um sinal normal pela TWD.	86
Figura 6.1	Montagem experimental do sistema sem falha.	91
Figura 6.2	Montagem experimental do sistema com falha.	92
Figura 6.3	Fluxograma para aquisição dos dados no tempo.	93
Figura 6.4	Sinal sem falha.	96
Figura 6.5	Sinal sem falha.	96
Figura 6.6	Sinal sem falha.	96
Figura 6.7	Sinal sem falha.	96

unesp

Figura 6.8	Sinal sem falha.	96
Figura 6.9	Sinal com falha.	97
Figura 6.10	Sinal com falha.	97
Figura 6.11	Sinal com falha.	97
Figura 6.12	Sinal com falha.	97
Figura 6.13	Sinal com falha.	97
Figura 6.14	Sinal de referência (sem falha) decomposto via TWD.	98
Figura 6.15	Sinal com falha decomposto via TWD .	99
Figura 6.16	Componente (detalhe - d1) da decomposição de um sinal com falha.	100

unesp

Lista de Tabelas

Tabela 5.1	Características geométricas do redutor.	84
Tabela 5.2	Arquivos utilizados para testar a RNA.	87
Tabela 6.1	Arquivos utilizados para treinar e testar a RNA.	101

Nomenclatura

A	Amplitude da função harmônica.	
a_0, a_k	Coeficientes de Fourier	
a	Dilatação na Transformada Wavelet	
a_0, a_2^{j+k}	Coeficientes Wavelet	
ap	Componentes de baixa freqüência da TWD	
b_k	Coeficientes de Fourier	
b	Translação na transformada de Fourier com Janela.	
b	Translação na Transformada Wavelet.	
c_k	Construtores	
$c_{j.k}$	Coeficientes Wavelet	
C_W	Condição de Admissibilidade	
$c_{\phi.k}$	Coeficientes Wavelet das escalas negativas	
dt	Componentes de alta freqüência da TWD	
d_i	Saída desejada do i-ésimo elemento da última camada da RNA	
E	Energia da função janela	
f	Freqüência	(Hz)
f_0	Freqüência central	(Hz)
$f(x)$	Sinal decomposto em uma série infinita de Wavelets de diferentes escalas e posições	
f^{ϕ}	Contribuição das escalas negativas para a amostra do sinal	
$f^{(j)}$	Contribuição das escalas j para a amostra do sinal	
$f^{(j.k)}$	Contribuição da Wavelet $W(2^j x+\Delta x)$ para a amostra do sinal	
$g(t)$	Função janela no tempo	
$G(f)$	Função janela em freqüência	
G_r	Matriz operadoras	
H	Matriz operadoras	
I	Matriz Identidade	
i	Constante imaginária	

k	Localização de uma função	
L	Filtro passa-baixa	
m_x	Valores médios de x	
m_y	Valores médios de y	
M_r	Matriz de coeficientes c_k	
N	Número de pontos de uma amostra	
N	Número par de coeficientes	
n	Índice de iteração	
ns	Número de neurônios da última camada da RNA	
S_b	Transformada de Fourier com Janela	
s_i	Constante da função de ativação	
$\Re$	Conjunto dos números reais	
t	Tempo	(s)
T	Duração do sinal	(s)
$X(\omega)$	Transformada de Fourier	
X_k	Transformada de Fourier Discreta	
x	Ordenada x	
$x(t)$	Sinal temporal	
x_r	Série discreta	
x_i	Grupo de dados de entradas da rede	
X_i	Vetor padrão de entradas	
y_i	Saída do i-ésimo elemento da última camada da RNA	
y	Função de ativação	
y	Ordenada y	
$W_{(a,b)}$	Família de Funções geradas pela variação dos parâmetros a e b	
W	Wavelet mãe	
w_i	Valores de ponderação (peso)	
W	Vetor contendo os pesos	
$\tilde{W}$	Transformada de Fourier de uma função Wavelet	

unesp

Letras Gregas

ω	Freqüência (rad/s)	
ω_k	Freqüência dos harmônicos que compõe um sinal	(rad/s)
$\Delta\omega$	Espaçamentos entre as freqüências dos harmônicos	
τ	Período	
Δ	Intervalo de amostragem	
Δt_{RMS}	Duração RMS da Wavelet	
Δf_{RMS}	Largura de banda RMS do filtro Wavelet	
$\phi(t)$	Função escala	
$\phi_0(t)$	Função caixa unitária	
$\phi_j(t)$	Função escala após j interações	
α	Fator escalar	
δ	Erro	
γ	Parâmetro de controle da estabilidade	
ε^2	Erro quadrático	
Δ_i	Correção associada com a i-ésima entrada x_i	
$\nabla_i(n)$	Gradiente do erro quadrático	
λ	Constante que determina a inclinação da função de ativação	
σ_x	Desvio Padrão em relação a x	
σ_y	Desvio Padrão em relação a y	
$\rho_{x,y}$	Coeficiente de correlação	
Σ	Somatório	

Superescristros

r	Índice referente a ordem da matriz
t	Transposto
j	Níveis da expansão da função Wavelet
*	Conjugado complexo

Subscritos

k	Índice de amplitude dos coeficientes de Fourier
x,y	Índices auxiliares

Siglas

WT	*Wavelet Transform* (Transformada Wavelet)
TW	Transformada Wavelet
TWD	Transformada Wavelet Discreta
STFT	*Short Time Fourier Transform* (Transformada de Fourier com Janela)
TWC	Transformada Wavelet Contínua
ANN	*Artificial Neural Network* (Rede Neural Artificial)
AR	Método Paramétrico Auto-Regressivo
RNA	Redes Neurais Artificiais
EEG	Eletroencefalograma
FFT	*Fast Fourier Transform* (Transformada de Rápida Fourier)
RMS	*Root Mean Square* (Valor médio quadrático)
TWDI	Transformada Wavelet Discreta Inversa
TF	Transformada de Fourier
TFD	Transformada de Fourier Discreta
TFDI	Transformada de Fourier Discreta Inversa
ADELINE	*Adaptive Linear Element*
MADELINE	*Multi Adaptive Linear Element*

Resumo

SANTOS, Edmilson Ricardo de Oliveira, *Utilização da Transformada Wavelet e Redes Neurais Artificiais para Reconhecimento de Padrões Aplicados a Manutenção Preditiva*, Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista, 2000, 117 p. Dissertação de Mestrado.

Uma das ferramentas que tem sido bastante investigadas para localização e detecção de falhas em sistemas mecânicos são as Redes Neurais Artificiais - RNA. No entanto esta ferramenta depende enormemente do padrão de entrada utilizado durante o processo de treinamento da rede. Se o padrão de entrada não for adequado, as vantagens da aplicação de RNA poderão ser perdidas. A Transformada Wavelet (TW) é uma técnica que possui excelentes características de localização tempo–freqüência e, proporciona maiores informações sobre as características do sinal. Portanto, em algumas aplicações, a utilização da TW na escolha do padrão de entradas do sinal poderá melhorar a eficiência da rede neural. A proposta deste trabalho é introduzir a Transformada Wavelet como um meio alternativo de extrair as características mais representativas dos dados de vibração e, utiliza-las como padrões de entrada no treinamento de Rede Neural Artificial. Neste caso, as informações a respeito das condições de funcionamento do sistema serão obtidas a partir dos coeficientes da TW afim de fornecer uma representação mais compacta dos sinais, quando comparado com sistemas convencionais como análise de Fourier.

Palavras chave: Transformada Wavelet, Redes Neurais Artificiais, Monitoramento, Daubechies, Pré-processamento.

unesp

Abstract

SANTOS, Edmilson Ricardo de Oliveira, *Utilização da Transformada Wavelet e Redes Neurais Artificiais para Reconhecimento de Padrões Aplicados a Manutenção Pred*itiva, Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista, 2000, 117 p. Dissertação de Mestrado.

Artificial Neural Network (ANN) has been widely investigated as damage detection tools in mechanical systems. However, this tool depends strongly of the contents of the pattern recognition used during the training process of the network. So, the advantages of the application of ANN can be better exploited if the input pattern considers specific effects caused by damages. The wavelet transform (WT) is a technique that possesses excellent characteristics of location time-frequency; hence, it provides larger amount of information of the signal characteristics. Therefore, in some application, the use of WT in the definition of the signal input pattern could characterize its main features, improving the efficiency of the ANN. The proposal of this work is to introduce WT as an alternative way of extracting the main features of vibration data in order to define a set of pattern recognition to be used in the training of ANN for damage detection purpose. In this case, the information regarding the conditions of operation of the system will be obtained starting from the coefficients of the WT that provides a more compact representation of the signals if compared with conventional tools, as for example analysis of Fourier.

Keywords: Wavelet transform, Artificial neural network, Monitoring, Daubechies, Preprocessing.

cm 1 2 3 4 5 6 7 8 unesp 12 13 14 15 16 17 18

CAPÍTULO 1

1. INTRODUÇÃO

O monitoramento e diagnose de sistemas mecânicos formam uma área emergente em engenharia. Os dados obtidos dessa área informam as condições de operação do sistema. Os dados adquiridos podem ser: aceleração, deslocamento, pressão e outros. Para o monitoramento de sistemas mecânicos, habitualmente são utilizados os sinais de vibração. O controle da variação dos níveis do sinal é uma boa indicação do estado do sistema, visto que o perfil de vibração pode mudar devido ao aparecimento de falhas.

O avanço tecnológico atual associado às exigências de mercado faz com que as indústrias almejem uma maior produtividade. Portanto, altas velocidades de operação juntamente com as condições severas de trabalho decorrentes desta situação levam a uma deterioração mais rápida dos sistemas mecânicos. Então, a manutenção preditiva surge como um ponto estratégico do processo produtivo, cujo papel principal é tornar a manutenção mais rápida e eficiente, mantendo a confiabilidade e evitando paradas precoces.

No monitoramento de sistemas mecânicos a manutenção preditiva tem sido um dos assuntos de grande interesse dos pesquisadores nos últimos anos. Ela utiliza o monitoramento regular do comportamento dos sistemas mecânicos, tentando prever o possível momento da falha, aumentando a vida útil e também as condições de segurança.

cm 1 2 3 4 5 6 7 8 unesp 12 13 14 15 16 17 18

Um número expressivo de técnicas já demonstrou um variado grau de sucesso em monitoramento de sistemas mecânicos. Entretanto, existe uma demanda por técnicas que possuam um melhor desempenho com relação à detecção de falhas, principalmente quando estas causam efeitos transientes. Uma das mais recentes técnicas de processamento de sinais, Transformada Wavelet (TW), tem-se mostrado adequada para o tratamento deste tipo de problema. Embora, uma inspeção visual de certas características da análise espectral seja adequada para identificar falhas, um método menos subjetivo e confiável é necessário para a diagnose do sinal de vibração. A questão de detecção de falhas neste caso é essencialmente um problema de reconhecimento de padrão; dados do sistema sem defeito devem ser distinguíveis dos dados de um sistema com falha. As Redes Neurais Artificiais (RNAs) são técnicas computacionais baseadas em modelos matemáticos inspirados nos neurônios biológicos dos organismos inteligentes, que têm se mostrado como uma excelente ferramenta para o reconhecimento de padrões. Isto tem despertado um grande interesse na área de monitoramento e detecção de falhas e a sua utilização nesta área tem-se mostrado bastante promissora. As RNAs são bem conhecidas por suas capacidades de identificar e classificar, e nos últimos anos tem atraído a atenção de muitos pesquisadores. No entanto, o sucesso da capacidade de reconhecimento está fortemente relacionado com a qualidade dos dados utilizados no treinamento da rede. Este trabalho propõe introduzir a Transformada Wavelet como um meio alternativo de extrair as informações das condições de funcionamento de um sistema, a partir de sinais de vibração ocasionados por falhas, e utiliza-las como um padrão de entrada para o treinamento de uma RNA.

No capítulo 2, em uma primeira etapa, é realizada uma revisão bibliográfica sobre diferentes métodos de processamento de sinais, em particular será dada ênfase à TW. Em uma segunda etapa, será discutida a utilização da TW juntamente com RNAs, para detecção de falhas com base em sinais de vibração. O objetivo é mostrar que a união destas técnicas tende a melhorar o diagnóstico de problemas que ocorrem em um sistema.

O capítulo 3 relata sobre o estudo de identificação de falhas em sistemas mecânicos, com base em medidas de vibração, para isso o processamento de sinais é um tópico fundamental para a avaliação e interpretação correta do problema.

Na natureza existem diferentes classes de sinais, e o processamento adequado de cada tipo depende de algumas características específicas verificadas nestes sinais. Alguns métodos são mais adequados para uma determinada classe de sinal do que outros. Neste contexto, os sinais temporais podem ser classificados, inicialmente, em estacionários e não-estacionários. Portanto,

cm 1 2 3 4 5 6 7 8 unesp 12 13 14 15 16 17 18

serão apresentadas técnicas adequadas para cada classe de sinal e a exatidão com que se pode obter as informações através delas. Foi comentado anteriormente que seria dada ênfase à TW, então será apresentado um método de construção de Wavelet, algoritmo de Mallat. Esse algoritmo pode ser visto como um método de decomposição (ou recuperação) de um sinal utilizando-se uma seqüência de multiplicação de matrizes.

No capítulo 4 são apresentados os fundamentos básicos de RNA e quais as suas principais aplicações.

No capítulo 5 é descrita a técnica de redução de dados que é aplicada com o objetivo de compactar a grande quantidade de informações geradas na fase de aquisição de dados. Ainda neste capítulo, é apresentada uma aplicação que envolve a técnica de pré-processamento utilizando a Transformada Wavelet Discreta (TWD) em conjunto com RNA para sinais simulados.

O capítulo 6 tem como objetivo validar a técnica utilizada para sinais simulados aplicando-a em sinais experimentais.

Finalmente, o capítulo 7 trata das conclusões e propostas de trabalhos futuros.

cm 1 2 3 4 5 6 7 8 unesp 12 13 14 15 16 17 18

CAPÍTULO 2

2. REVISÃO BIBLIOGRÁFICA

A área de processamento e identificação de sinais tem um conjunto expressivo de métodos e técnicas de análise, que tem por objetivo extrair características peculiares dos dados coletados em testes experimentais. A utilização de algumas destas técnicas, de forma complementar, tem permitido o desenvolvimento de várias metodologias de avaliação e detecção de falhas em sistemas mecânicos. As ferramentas de avaliação não destrutiva que utilizam medidas de vibração, estão conquistando cada vez mais importância no contexto de monitoramento das condições de operação e detecção de falhas de máquinas e equipamentos. Em um sentido mais convencional, elas monitoram e comparam o nível atual de vibração do sistema com um nível padrão definido para o sistema sem falhas, objetivando identificar mudanças causadas por eventuais falhas do sistema.

Recentes literaturas correlacionam alguns tipos de sinais de vibração com tipos característicos de falhas e procuram identificar eventuais falhas ainda em um estágio inicial. A análise de tendência é uma técnica bem conhecida nesta área, a qual compara o espectro de

vibração de uma máquina ou equipamento com um espectro de referência e avalia a intensidade das mudanças de vibração de cada espectro em um intervalo temporal especificado. Entretanto, nesta clássica técnica de reconhecimento de padrão, uma significante parte da análise consiste no pré-processamento e separação dos dados medidos em um conjunto relacionado com cada tipo de falha. O estabelecimento de uma relação confiável entre estes conjuntos de dados e os diferentes tipos de falhas é considerado o principal problema desta técnica. Habitualmente, isto é desempenhado pela manutenção particular, monitoramento e registro de sinais e também por analistas que examinam a condição de monitoramento da máquina para detectar falhas incipientes. A tendência atual é projetar sistemas robustos, que automaticamente desempenhe este tipo de análise em tempo real. Isso inclui ainda, o desenvolvimento de ferramentas que sejam capazes de identificar a ocorrência de alguns tipos específicos de eventos, principalmente os transitórios, os quais podem dificultar a utilização de alguns métodos clássicos de análise para avaliação das condições de vibração e diagnóstico de máquinas e equipamentos.

Mori *et al.* (1996) utilizaram os coeficientes da Transformada Wavelet Discreta (TWD) para produzir informações detalhadas de fenômenos transitórios, os quais ocorrem quando uma esfera passa sobre um risco na pista interna do rolamento. A predição destes fenômenos é possível devido à inspeção feita na tendência dos coeficientes da TWD, ou seja, o estudo é realizado sobre uma taxa definida pela razão média dos coeficientes máximos em cada nível de decomposição do sinal para diferentes tamanhos de blocos de amostragem. O numerador é a média dos coeficientes Wavelets máximos em cada bloco de divisão, cujo tamanho é definido pelo tempo total de aquisição pelo número de blocos. Da mesma forma o denominador é outra média dos coeficientes máximos, agora, dividindo-se o tempo total em um número maior de blocos. A relação utilizada entre o numerador e o denominador foi de 10 a 100. A média dos coeficientes máximos em relação a 10 pontos é mais sensível a mudanças locais do que a média dos coeficientes máximos em relação a 100 pontos.

Moraes (1996) utilizou uma outra forma de representação, chamada de mapas médios quadráticos. Nessa representação, o domínio é tempo-escala e, cada escala j é dividida em 2^j intervalos, sendo que cada um desses é centrado exatamente no instante temporal correspondente ao centro de uma Wavelet de mesma escala. Assim, a cada um desses intervalos é associado o valor do quadrado da amplitude do coeficiente Wavelet correspondente, formando uma superfície que representa a distribuição do valor quadrático médio do sinal entre diferentes escalas e também entre diferentes Wavelets de mesma escala. O autor aplicou o mapa médio

unesp

quadrático para um sinal simulado composto de três freqüências (100 Hz, 60 Hz e 25 Hz), sendo a amplitude da terceira modulada por uma freqüência de 12 Hz. As freqüências são ponderadas respectivamente por 1.0, 0.4 e 0.6. Em seguida, foram somados ao sinal 8 pulsos transitórios e o mapa também foi aplicado para este sinal. Uma comparação foi feita entre as análises e detectou-se alteração dos coeficientes em uma faixa de escala. Aplicou-se a transformada Wavelet discreta inversa, recuperando o sinal no tempo para as faixas de escalas correspondentes as perturbações.

Tai e Sung (1997) aplicaram a STFT "*Short Time Fourier transform*" e a TWC, Transformada Wavelet Contínua, para analisar sinais de vibração que representam o movimento de oscilação de um vagão, o qual é acoplado a um motor por sistema de correias. As análises foram feitas utilizando-se os espectrogramas (tempo-freqüência) gerados pela STFT, e também pelos escalogramas (tempo-escala) gerados pela TWC. Para auxiliar a análise da TWC utilizou-se um índice que representa a porcentagem de energia (PE) encontrada através dos coeficientes extraídos da TWC para cada escala. A porcentagem de energia é dada pelo quociente entre a soma do quadrado dos coeficientes de cada escala e a soma total dos coeficientes ao quadrado. A PE é um índice útil para representar a distribuição de energia, assim como a variação de um sinal. Portanto, foi possível detectar mais facilmente os eventos transitórios no escalograma (tempo-escala) do que no espectrograma (tempo-freqüência), concluindo-se que os escalogramas (tempo-escala) são mais eficazes na análise e diagnose de sinais não-estacionários.

Najar e Hashemi (1999) apresentaram a aplicabilidade da TWC e TWD para análise de sinais acústicos provenientes do carregamento unidirecional de um fio (teste de fadiga). O objetivo da utilização das transformadas era visualizar com maior precisão as emissões dos sinais acústicos, ou seja, fazer uma comparação direta entre a TWC e a TWD. As análises resultaram na seguinte conclusão: as emissões podem ser melhor visualizadas no escalograma (tempo-escala) gerado pela TWC, pois sua resolução (escala) é muito refinada e consequentemente sua precisão será melhor do que a TWD, que possui uma resolução em potência de dois.

Autores como, Staszewski e Worden (1993) buscam associar as características da TW com as da RNA para a condição de monitoramento e detecção de falhas. Neste artigo eles discutem a detecção de falhas em um redutor acoplado a um motor elétrico utilizando um

cm 1 2 3 4 5 6 7 8 unesp 12 13 14 15 16 17 18

processo de dois estágios, ou seja, primeiro os sinais são processados pela TWC, em seguida classificados por uma RNA. Várias situações são discutidas, um dente quebrado foi simulado através da remoção progressiva (25%, 50%, 70% e 100%) até uma profundidade de 2 mm da face do dente do pinhão em uma posição pré-estabelecida. Os dados foram registrados para a condição normal e para cada condição de falha. Embora a TWC fosse suficiente para detectar a falha, ela não conseguiu distinguir os vários tipos de falhas. Neste caso utilizou-se uma RNA visando classificar os vários tipos de falhas. Os resultados da classificação da RNA foram considerados bons, pois nenhuma resposta da RNA desviou além de 12% do resultado final.

Petrilli *et al.* (1995) aplicaram as seguintes técnicas de processamento: Cepstrum, Transformada de Fourier e Transformada Wavelet em conjunto com RNAs para monitoramento de um sistema constituído de um redutor acoplado a um motor. Uma falha foi introduzida intencionalmente em um dente da engrenagem. Os sinais sem defeito e com defeito foram pré-processados visando uma redução dos dados e em seguida, os 5 mais importantes valores da energia espectral obtidos de cada técnica foram utilizados para treinar e testar as suas respectivas RNAs. Os resultados utilizando a Transformada de Fourier não foram tão adequadas como os apresentados pelas técnicas Cepstrum e Wavelet devido aos harmônicos que geram muitas informações redundantes para rede. As técnicas Cepstrum e Wavelet demonstraram ser mais eficientes, ou seja, apresentaram uma grande capacidade em compactar informações, mantendo a fidelidade das principais informações. As informações características a respeito do estado de funcionamento do sistema foram mantidas no processo de redução dos dados favorecendo o bom desempenho das suas respectivas RNAs.

Paya *et al.* (1997) também propuseram a utilização de Wavelet e redes neurais para o monitoramento de um sistema constituído de um redutor, dois mancais e um motor elétrico. Este sistema foi monitorado em condições normais, com falhas individuais, e por último com falhas múltiplas. Os sinais adquiridos para cada caso foram pré-processados pela Transformada Wavelet, sendo considerando somente os 10 maiores coeficientes. Estes coeficientes foram utilizados para treinar e testar uma RNA do tipo perceptron multi-camada. Uma taxa de 96% de classificação sobre os dados testes foram obtidos, levando à uma indicação clara dos tipos exato de falhas no sistema monitorado.

Hazarika *et al.* (1997) utilizaram técnicas de processamento de sinais em conjunto com RNAs para classificação de anomalias no cérebro humano. Para isso, registros da variação

unesp

temporal da atividade elétrica cerebral foram empregados. Nesses registros, os quais são denominados de Eletroencefalograma (EEG), estão presente informações sobre doenças psiquiátricas, as quais são muito sutis de serem observadas. Os sinais EEG foram divididos em três classes: normal, esquizofrenia e doença compulsiva-obsessiva. Então, em uma primeira etapa as atividades de processamento de sinais foram desenvolvidas pela TW e também pelo método paramétrico Auto-Regressivo (AR). Em uma segunda etapa, os coeficientes originários do processamento de sinais através das duas técnicas foram utilizados como padrão de entrada para treinar e testar as suas respectivas RNA. Nessa etapa o tipo de RNA utilizada para classificação foi o perceptron multi-camadas. Os resultados obtidos pela RNA treinada com coeficientes Wavelets, para as classes de doenças esquizofrenia e compulssiva-obsessiva, foram superiores aos resultados obtidos pela RNA treinada com coeficientes do método AR, somente para classe normal os resultados foram iguais. A substituição da RNA perceptron multi-camadas por uma RNA recorrente apresentou melhores resultados. Após o treinamento, a RNA utilizando coeficientes da TW obteve resultados superiores em todas as classes de doenças em relação a RNA treinada com coeficientes do método Auto-Regressivo.

unesp

CAPÍTULO 3

3. FUNDAMENTOS TEÓRICOS

No estudo de identificação de falhas em sistemas mecânicos, com base em medidas de vibração, a análise e processamento de sinais são tópicos fundamentais para a avaliação e interpretação correta do problema. As propriedades de vibração de um sistema estão associadas, diretamente, com as forças atuantes no mesmo e às suas respectivas propriedades físicas, sendo que qualquer mudança nas propriedades físicas do sistema (massa, rigidez e amortecimento) irá refletir imediatamente nos sinais de vibração. Isto pode ser adequadamente explorado para avaliação das condições de funcionamento do equipamento.

Na natureza existem diferentes classes de sinais e o processamento adequado de cada tipo depende das características específicas verificadas no sinal. Alguns métodos são mais adequados para uma determinada classe de sinal do que outros. Os sinais temporais podem ser classificados, inicialmente, em estacionários e não-estacionários. No primeiro caso, o sinal não sofre alterações ao longo do tempo, ou seja, as suas propriedades estatísticas não mudam no tempo. Na segunda classe, as propriedades estatísticas mudam significativamente num dado intervalo de tempo, isto é, um sinal (ou parte dele) é considerado não-estacionário quando ele não apresenta um padrão de repetição distinto.

cm 1 2 3 4 5 6 7 8 unesp 12 13 14 15 16 17 18

3.1 ANÁLISE DE SINAIS ESTACIONÁRIOS

Para muitos sinais, a análise espectral é extremamente útil, pois o conteúdo de freqüência é suficientemente rico em informações a respeito do comportamento do sistema. Portanto, os sinais podem ser caracterizados por suas freqüências. Isso pode ser constatado por uma freqüência de um sinal de áudio que é responsável pelo que se ouve de sons graves e agudos. Logo, a análise de uma função pode ser feita pelo estudo de suas freqüências, (Gomes *et al.*, 1997). O ponto de partida para o entendimento das características de freqüência dos sinais é dado pela análise de funções periódicas, devido à facilidade de entendimento das propriedades destas funções. Para isso pode ser utilizada a série de Fourier, que representa adequadamente esta classe de sinais estacionários.

3.1.1 SÉRIE DE FOURIER E TRANSFORMADA DE FOURIER

Seja uma função harmônica $x(t)=A\sin(\omega t)$, para $\omega=2\pi/\tau$, $\tau>0$, definido no intervalo [-1,1]. O parâmetro A mede a amplitude da função, Fig. 3.1. O parâmetro ω indica quantos ciclos completos de períodos τ existem no intervalo de tempo T. Esse número está diretamente relacionado com o número de oscilações da função na unidade de tempo, ou seja, a freqüência da função. As figuras 3.1b e 3.1c mostram o gráfico de $x(t)$ para diferentes valores de ω.

Uma função periódica $x(t)=f(t+\tau)$, pode ser decomposta em suas componentes harmônicas, sendo expressa como série infinita da forma:

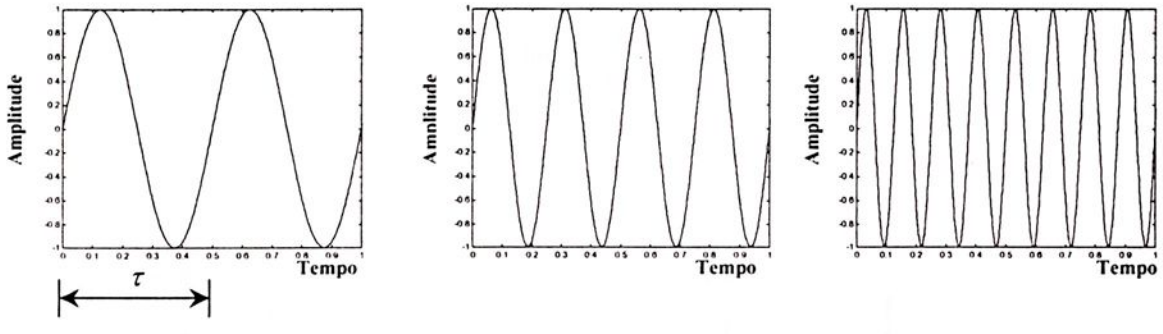
$$x(t) = a_0 + \sum_{k=1}^{r}\left(a_k \cos\frac{2\pi kt}{\tau} + b_k \operatorname{sen}\frac{2\pi kt}{\tau}\right) \tag{3.1}$$

sendo a_0, a_k e b_k constantes denominadas de coeficientes de Fourier, dadas por:

$$a_0 = \frac{1}{\tau}\int_{-\tau/2}^{\tau/2} x(t)dt \tag{3.2}$$

$$a_k = \frac{2}{\tau}\int_{-\tau/2}^{\tau/2} x(t)\cos\frac{2\pi kt}{\tau}\,dt\,;\, k = 1, 2, \ldots\ldots, n, \tag{3.3}$$

$$b_k = \frac{2}{\tau}\int_{\tau/2}^{\tau/2} x(t)\,\mathrm{sen}\,\frac{2\pi kt}{\tau}\,dt\,;\, k = 1, 2, \ldots\ldots, n, \tag{3.4}$$



(a) (b) (c)

Figura 3.1 - Funções periódicas

A condição matemática para convergência da equação (3.1) é representada pela equação (3.5). Somente uma importante restrição deve ser feita quando a função *x(t)* é descontínua: a série gera um valor médio de *x(t)* na descontinuidade. Supõe-se que a posição do eixo do tempo na, figura 3.1 é ajustada e por isso o valor médio de *x(t)* é zero. Então, de acordo com a equação (3.2), o coeficiente a_0 será nulo. Para sinais periódicos, os coeficientes remanescentes a_k e b_k são calculados e os seus valores podem ser ilustrados graficamente como mostra a figura 3.2. O eixo horizontal representa as freqüências e a localização ω_k dos respectivos coeficientes,

$$\int_{-\tau/2}^{\tau/2} |x(t)|^2 \, dt < \infty \tag{3.5}$$

$$\omega_k = \frac{2\pi k}{\tau}, \tag{3.6}$$

ω_k é a freqüência dos harmônicos que compõe o sinal. Os espaçamentos entre as freqüências dos harmônicos são representados por:

$$\Delta\omega = \frac{2\pi}{\tau}. \tag{3.7}$$

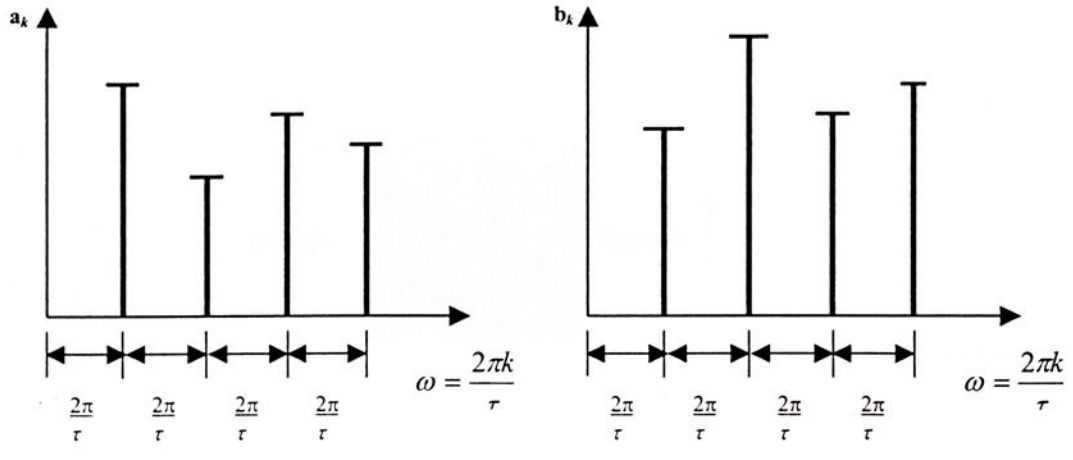


Figura 3.2 - Representação gráfica dos coeficientes de Fourier.

Uma outra classe de sinal de suma importância é a dos sinais transientes. Observando a equação (3.7) nota-se que quando o período τ torna-se grande, o espaçamento entre os

coeficientes, $\Delta\omega$, torna-se pequeno e vice-versa. Portanto, para $\tau \rightarrow \infty$, os espaçamentos entre as raias de fato estarão unidos. Neste caso *x(t)* não mais representa um fenômeno periódico e portanto não pode ser analisado em componentes de freqüências discretas. Assim, as equações (3.2) a (3.4) tornam-se funções contínuas de freqüência e a análise das componentes de freqüência do sinal é feita pela chamada transformada de Fourier, Newland (1993).

As equações (3.8) e (3.9) formam o par de transformada de Fourier na forma complexa, que permitem a transformação do sinal temporal *x(t)* para o domínio da freqüência e a transformação do sinal no domínio da freqüência para o domínio do tempo, Fig. 3.3. Uma análise da Fig.3.3 permite identificar claramente quais são as componentes de freqüência do sinal.

$$X(\omega) = \int_{-\infty}^{+\infty} x(t) e^{-i2\pi\omega t} dt \tag{3.8}$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(\omega) e^{i2\pi\omega t} d\omega \tag{3.9}$$

Figura 3.3 - Transformada de Fourier.

As equações (3.8) e (3.9) são denominadas, respectivamente, Transformada de Fourier e Transformada de Fourier Inversa e levam a uma avaliação do sinal no domínio do tempo ou da freqüência. Entretanto, estas equações não são adequadas para a utilização em testes práticos

porque na maioria dos casos, os sinais a serem analisados são discretos e as mesmas devem ser redefinidas como tal.

3.1.2 TRANSFORMADA DE FOURIER DISCRETA

Medidas experimentais de um processo geralmente envolvem a digitalização do sinal através de um conversor digital. Neste caso, a função $x(t)$ é amostrada em uma série temporal regularmente espaçada, Fig.3.4. Se o intervalo de amostragem é Δ (constante) então o valor discreto da função no tempo $t = r\Delta$, é escrito como $x_r=(r\Delta)$, r=. . . . .,-1, 0, 1 , 2, 3. . . . . ,. Desde que, as séries discretas têm sido derivadas de uma função contínua no tempo, é usual referir-se a mesma como uma série temporal contínua. As palavras "séries temporais" são portanto utilizadas para referirem-se a uma seqüência de números discretos x_r ordenados no tempo, ou para referirem-se a uma amostra temporal contínua original, da qual a série discreta foi obtida.

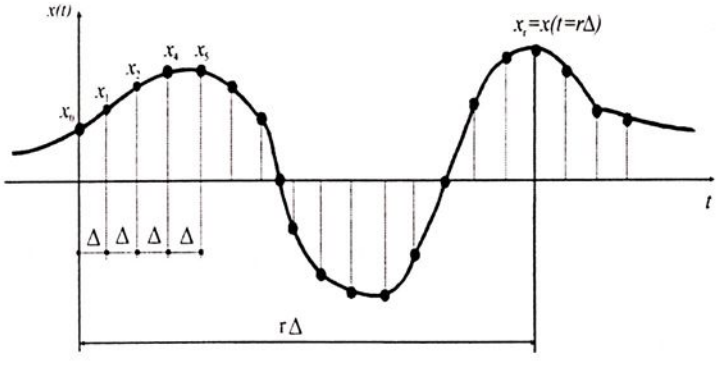


Figura 3.4 - Amostragem de uma função contínua de intervalos regulares no tempo.

Na série de Fourier, se uma função $x(t)$ é periódica, esta pode ser escrita na forma dada pelas equações (3.1) a (3.4). As equações (3.3) e (3.4) podem ser escrita no intervalo de 0 a τ, ao

invés de $-\tau/2$ a $\tau/2$, desde que $x(t)$ tenha período τ. Utilizando notação complexa, as equações (3.3) e (3.4) podem ser combinadas em uma única equação, então, define-se:

$$X_k = \mathrm{a}_k - i\mathrm{b}_k, \tag{3.10}$$

e utilizando a equação (3.11), tem-se:

$$e^{-i(2\pi kt/\tau)} = \cos\frac{2\pi kt}{\tau} + i\,\mathrm{sen}\frac{2\pi kt}{\tau} \tag{3.11}$$

assim, tem se:

$$X_k = \frac{1}{\tau}\int_0^{\tau} x(t)e^{-i(2\pi kt/\tau)}dt\,;\, k = 1, 2, \ldots\ldots, n. \tag{3.12}$$

Considerando que a série temporal contínua $x(t)$ não é conhecida e que somente as amostras igualmente espaçadas são disponíveis, portanto, pode-se redefinir a equação (3.12) para um sinal discreto, representado pela série discreta $\{x_r\}$, r = 1, 2, 3. . . . . ., (N-1), Newland, (1993). Para um tempo $t = r\Delta$, e $\Delta = \tau/N$, a integral na equação (3.12) pode ser considerada como uma soma:

$$X_k = \frac{1}{\tau}\sum_{r=0}^{N-1} x_r e^{-i(2\pi k/\tau)(r\Delta)}\Delta \tag{3.13}$$

A equação (3.13) assume que a área total sobre a curva, na figura 3.5, é dada pela soma de todas as faixas sombreadas. Substituindo $\tau = N\Delta$ na equação (3.13), tem-se:

$$X_k = \frac{1}{N}\sum_{r=0}^{N-1} x_r e^{-i(2\pi kr/N)} \tag{3.14}$$

a qual pode ser considerado como uma fórmula aproximada para calcular os coeficientes da série representada pela equação (3.1).

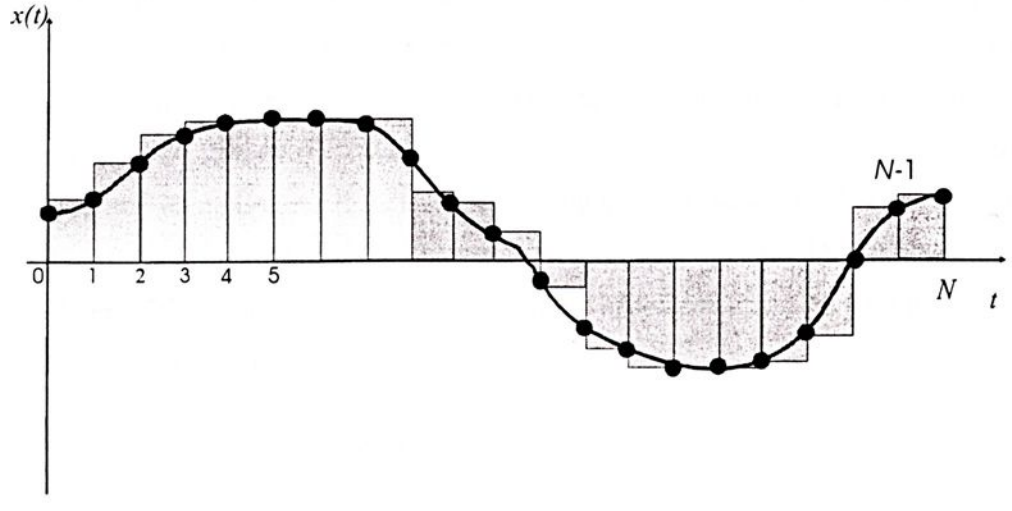


Figura 3.5 - Aproximação envolvida no cálculo dos coeficientes de Fourier de uma série.

No entanto, a equação (3.14) não fornece informação suficiente para permitir que a série temporal contínua seja reconstituída e que os valores discretos $\{x_r\}$ sejam recuperados exatamente. Portanto, qualquer valor da série $\{x_r\}$ é dado pela fórmula inversa:

$$x_r = \sum_{k=0}^{N-1} X_k e^{j(2\pi kr/N)} \tag{3.15}$$

que representa a definição formal da Transformada de Fourier Discreta da série x_r; r = 1, 2, 3, ..., (N-1).

O intervalo das componentes de Fourier X_k é limitado para k = 0 à (N –1), correspondendo às harmônicas de freqüência $\omega_k = \frac{2\pi k}{\tau} = \frac{2\pi k}{N\Delta}$, a fim de manter a simetria do par de transformada representado pelas equações (3.14) e (3.15).

A TFD é utilizada em grande variedade de aplicações em processamento de sinais e, portanto, é de grande interesse a implementação de um algoritmo eficiente para o cálculo da Transformada discreta, Eq. (3.14), e transformada inversa Eq. (3.15). Assim, para o cálculo da TFD ou TFDI de uma amostra com N pontos, é necessária a ordem de N^2 operações matemáticas (multiplicações e adições). Esse número pode ser reduzido utilizando a Transformada Rápida de Fourier (FFT), que é um algoritmo baseado na TFD com N pontos e consiste em calculá-la através de duas TFDs com N/2 pontos ou quatro TFDs com N/4 pontos, e assim por diante. Dessa forma, o número de operações necessárias para transformação (direta ou inversa) é reduzido à ordem de Nlog N, Newland, (1993).

3.2 ANÁLISE DE SINAIS NÃO-ESTACIONÁRIOS

As técnicas de análise espectral tradicionais, com base na transformada de Fourier ou em filtros digitais, fornecem uma descrição melhor de sinais estacionários e pseudo-estacionários. Infelizmente, estas técnicas enfrentam algumas limitações quando os sinais a serem analisados são não-estacionários, isto é, sinais com propriedades espectrais variando no tempo. Em tais casos, a solução seria calcular um espectro instantâneo para cada índice temporal do sinal. As ferramentas que trabalham dessa forma são chamadas de técnicas de análise tempo-freqüência, as quais serão discutidas a seguir.

unesp

3.2.1 TRANSFORMADA DE FOURIER COM JANELA

A transformada de Fourier discutida anteriormente não se mostra adequada para a classe de sinal não-estacionária, ou seja, existem informações temporais (eventos transitórios) que não são observadas por ela. Assim, Gabor, (1946) tentando corrigir esta limitação, adaptou-a para analisar somente uma pequena seção do sinal no tempo (técnica chamada de janelamento do sinal). Essa adaptação é conhecida como Transformada de Fourier com janela, "*Short time Fourier transform*" (STFT). A idéia é dividir o sinal não-estacionário em frações, nas quais aplicam-se hipóteses estacionárias e realiza-se a transformada de Fourier de cada uma destas frações. A divisão é feita por meio de uma janela $g(t\text{-b})$, na qual b representa a localização da janela e, portanto, a localização temporal do correspondente espectro, Fig. 3.6. A série de espectros, cada um relacionado a um índice temporal, forma uma representação tempo-freqüência do sinal. Note, que o comprimento, a forma da janela e também as suas translações estão fixadas. Estes valores devem ser escolhidos antes de se iniciar a análise. Desta forma, a STFT oferece uma resolução constante no tempo e na freqüência, e é definida como a transformada de Fourier de um sinal janelado para as várias posições b da janela,

$$\begin{aligned} S_{\mathrm{b}} &= \int_{-\infty}^{+\infty} x(t) g^{*}(t-\mathrm{b}) e^{-j2\pi(t-\mathrm{b})} dt \\ S_{\mathrm{b}} &= \left\langle x, g_{\mathrm{b},f} \right\rangle \end{aligned} \tag{3.16}$$

$g_{\mathrm{b},f}(t) = g^{*}(t-\mathrm{b}) e^{-j2\pi f(t-\mathrm{b})}$. A figura 3.7 mostra que a análise de um sinal através da STFT fornece um reticulado uniforme, ou seja, as células de informações que formam o domínio tempo-freqüência são uniformes.

unesp

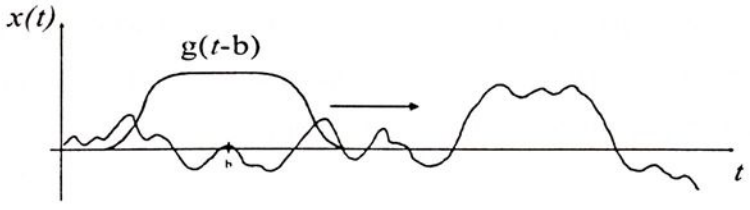


Figura 3.6 - Transformada de Fourier com janela "*Short time Fourier transform*".

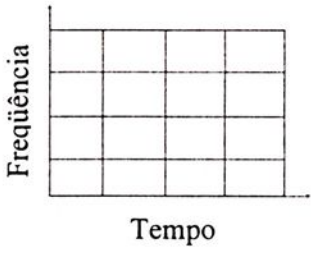


Figura 3.7 - Domínio de tempo-freqüência (STFT).

Com a STFT pode-se obter informações com exatidão limitada, visto que esta é determinada pelo tamanho da janela. A vantagem é que uma vez escolhido o tamanho de uma janela em particular, esta é a mesma para todas as freqüências. Muitos sinais requerem um enfoque mais flexível, ou seja, onde se possa variar o tamanho da janela para determinar uma característica particular mais adequada em tempo ou em freqüência. A análise com base na Transformada Wavelet possui esta característica, como será discutido no próximo item.

3.2.2 TRANSFORMADA WAVELET

A STFT introduz uma escala e analisa o sinal do ponto de vista dela (a escala é dada pela largura da janela). Se o sinal possui detalhes importantes fora dessa escala, haverá problemas na análise. Portanto, para a STFT visualizar melhor esses detalhes, o tamanho da janela terá que ser

alterado, e uma vez fixo o tamanho será o mesmo para todo o período T do sinal em análise. As conseqüências dessa escolha serão uma melhor visualização no tempo e um empobrecimento na freqüência ou vice-versa. Isso acontece porque o domínio tempo-freqüência é governado pelo princípio da incerteza.

Para resolver esse problema da STFT, será definido uma transformada que não utiliza uma escala fixa na análise. A escala é definida pela largura da função de modulação, portanto, será utilizada uma função que não tenha escala fixa e que seja localizada no tempo. Partindo de uma função W "Wavelet mãe" como função de modulação, e tomando como base uma família de funções geradas a partir da variação da escala e posição da função W, tem-se:

$$W_{(a,b)}(t) = \frac{1}{\sqrt{a}} W\left(\frac{t-b}{a}\right), \; a \text{ e } b \in \Re, \; a \neq 0 \tag{3.17}$$

A modulação da função W pelo fator $\frac{1}{\sqrt{a}}$, faz com que sua amplitude aumente quando a escala a diminui e vice-versa. Em termos de freqüência, para a pequena $W_{(a.b)}$ possui altas freqüências e à medida que a aumenta, a freqüência detectada diminui, Fig. 3.8.

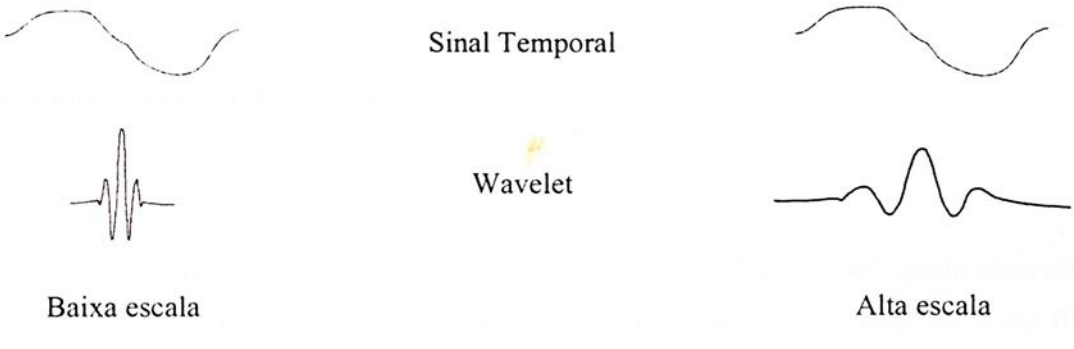


Figura 3.8 - Relação entre freqüência e escala.

3.2.3 FAMÍLIAS WAVELETS

As famílias de Wavelets mais conhecidas são: Daubechies, Morlet e Meyer, Fig. 3.9. As Wavelets da família Daubechies têm sido bastante utilizadas por terem um suporte compacto e também devido à sua assimetria, a qual favorece a visualização de detalhes tais como transientes, Najar e Hashemi (1999). Neste trabalho serão discutidas as Wavelets da família Daubechies devido às suas propriedades mencionadas anteriormente, bem como pela sua aceitação e utilização por vários pesquisadores, como: Paya *et al.* (1997), Newland (1994), Petrilli *et al.* (1995) e outros.

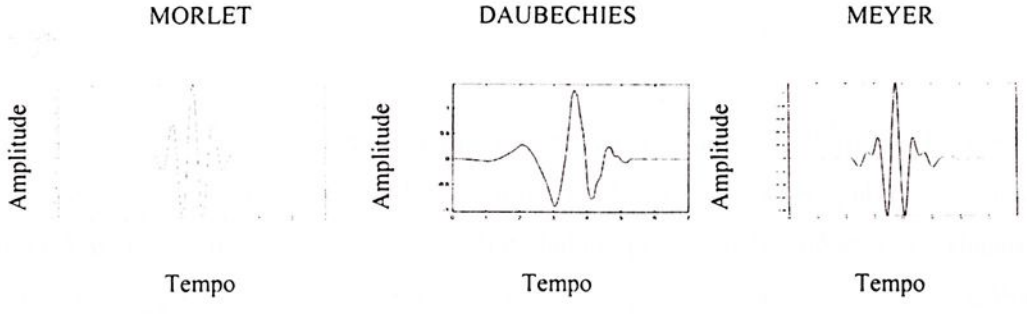


Figura 3.9 - Família Wavelets.

3.2.4 O CONCEITO DE ESCALA

Para estabelecer uma conexão clara entre "escala" e "freqüência", é adequado observar as Wavelets no domínio da freqüência, Fig. 3.10. O espectro da Wavelet corresponde a um filtro passa banda centrado ao redor da freqüência f_0, sendo esta freqüência central o recíproco do período da Wavelet, e a largura de banda depende de quantos períodos (oscilações) estão inclusos na Wavelet, ou seja, o comprimento da Wavelet.

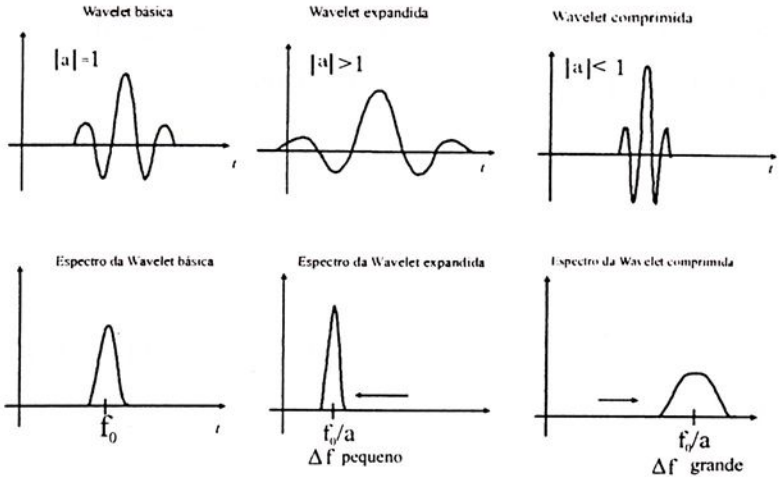


Figura 3.10 - A escala da TW é implementada pelo parâmetro de escala *a*.

A escala no domínio do tempo corresponde a uma translação no domínio da freqüência. O espectro da Wavelet dilatada (expandida) é transladado para baixas freqüências, enquanto o espectro da Wavelet contraída (comprimida) é transladado para altas freqüências. A relação entre "escala" e " freqüência" se torna mais evidente. Outra característica importante é que a Wavelet expandida é mais ampla no tempo, mas exibe um espectro mais concentrado em torno da freqüência central. O inverso aplica-se para a Wavelet comprimida, seu espectro é amplo ao redor da sua freqüência central, e concentrado no tempo. Este fato é conseqüência direta do princípio da incerteza, que será discutido brevemente com mais detalhes. A duração Δt da Wavelet no domínio do tempo é proporcional ao fator de escalar a, enquanto a largura de banda do "filtro Wavelet", Δf, no domínio da freqüência é inversamente proporcional ao fator de escala a. Como conseqüência, o produto entre a duração temporal e a largura de banda do filtro é uma constante ($\Delta t \Delta f = constante$), e ela pode assumir qualquer valor dependendo da definição de Δt e Δf. Definindo-se Δt como a duração RMS da Wavelet e Δf como largura de banda RMS do filtro Wavelet, então o produto é sempre $\geq 1/4\pi$. Isto é chamado de relação de incerteza de Heisenberg, a qual também se aplica para a STFT.

A conclusão é que a TW favorece a resolução temporal quando são analisados componentes de alta freqüência, e oferece vantagens com a resolução de freqüência quando feita

com baixas freqüências comparada a STFT, a qual oferece resolução constante no domínio do tempo e em freqüência. Assim, a TW conduz a uma análise com porcentagem constante (ou relativa) de largura de banda, enquanto STFT fornece análise de largura de banda constante, Gram-Hansen *et al.* (1997).

3.2.5 A RELAÇÃO DE INCERTEZA DE HEISENBERG

Um ponto freqüentemente abordado na análise de sinais é a exatidão que se pode obter para localização das informações, em relação à f, no domínio tempo-freqüência. Esta exatidão está relacionada com as funções janelas g, utilizadas na transformada. O que rege a exatidão da transformada no domínio tempo-freqüência é a relação de incerteza de Heisenberg. De maneira sucinta, esse princípio afirma que não se podem obter localizações precisas simultaneamente no domínio do tempo e da freqüência. A idéia desse princípio é simples: para medir freqüências deve-se observar o sinal por um período de tempo definido e quanto mais exatidão deseja-se na medida da freqüência, maior o período que se deve observar.

A relação de incerteza é conhecida como o produto BT (largura de banda X tempo), que deve ser maior que a unidade para os resultados de análise serem válidos. Para a análise da FFT "*Fast Fourier Transform*", uma versão muito prática e geral do produto BT é ΔfT=1, sendo Δf o espaçamento da FFT e T a duração do registro. Esta relação é independente da escolha da função de ponderação.

A versão fundamental da relação de incerteza utiliza o desvio padrão, que é dado pelas equações:

$$\Delta t_{RMS} = \left(\frac{1}{E_g} \int_{-\infty}^{+\infty} (t - t_0)^2 |g(t)|^2 \, dt \right)^{\frac{1}{2}} \tag{3.18}$$

$$\Delta f_{RMS} = \left(\frac{1}{E_g} \int_{-\infty}^{+\infty} (f - fo)^2 \left| G(f) \right|^2 df \right)^{\frac{1}{2}} \tag{3.19}$$

As equações (3.18) e (3.19) determinam a localização no domínio (f,t) e é representada por elipses de dimensões $\Delta t \Delta f$. Essas elipses são chamadas de células de informação. Pelo princípio da incerteza, a área dessa célula é $\geq 1/4\pi$. O fator de normalização E é a energia da função janela. A função de ponderação escolhida é uma forma Gaussiana para ambas as transformadas: TW e STFT. Neste caso alcança-se o produto de incerteza, o qual é igual a $1/4\pi$. Isto pode ser visualizado pela elipse de Heisenberg. Para STFT, a elipse tem uma forma constante, enquanto que para TW o comprimento da largura de freqüência é proporcional à freqüência na qual ela está localizada. Assim, a STFT oferece uma resolução constante no tempo e em freqüência, Fig. 3.11, enquanto que a TW oferece uma melhor resolução no tempo e em freqüência, Fig. 3.12, (Gram-Hansen *et al.*, 1997)

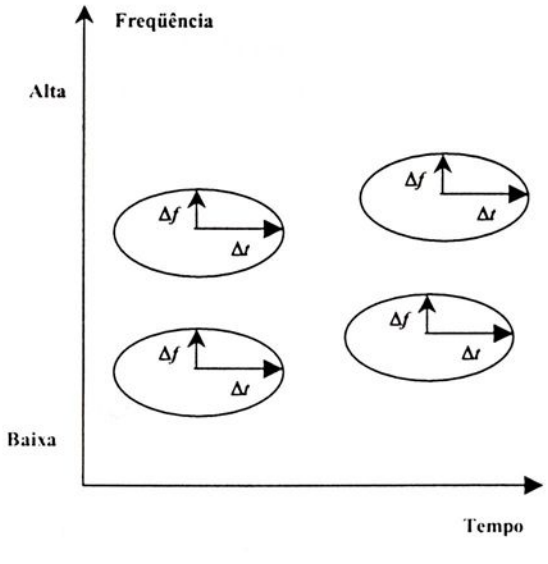


Figura 3.11 - Elipse de Heisenberg para STFT.

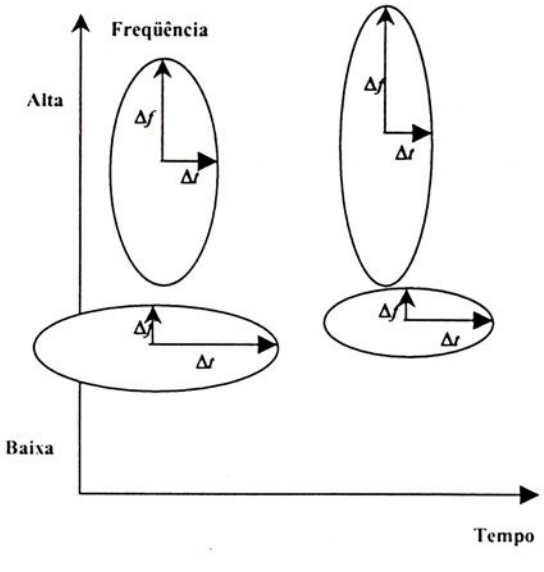


Figura 3.12 - Elipse de Heisenberg para TW.

3.3 WAVELETS ORTOGONAIS

Este item discute a construção de Wavelets ortogonais, de suporte compacto. Nesse método, uma função escala é calculada através de equações de dilatação, a qual é utilizada para gerar a função Wavelet básica. Essas duas funções são a base do algoritmo de Mallat, que pode ser visto como um método para decompor (ou recuperar) um sinal utilizando-se uma seqüência de multiplicação de matrizes.

3.3.1 EQUAÇÕES DE DILATAÇÃO

As funções Wavelets ortogonais são geradas através da utilização das equações de dilatação, chamadas de funções escala. Uma função escala $\phi(x)$ é uma versão dilatada (horizontalmente) de $\phi(2x)$. Ela tem a mesma altura, mas é esticada duas vezes sobre o eixo horizontal x, o qual é uma variável independente que pode representar tempo ou espaço. A equação de dilatação $\phi(x)$ é expressa como uma soma finita de termos da forma $\phi(2x)$. Cada um destes termos é posicionado em um lugar diferente sobre o eixo horizontal tomando o argumento $(2x-k)$ no lugar de $2x$, sendo k um inteiro. A equação de dilatação básica tem a forma:

$$\phi(x) = \sum_{k=0}^{N-1} c_k \phi(2x - k) \tag{3.20}$$

sendo N um número par de coeficientes Wavelet c_k, que são constantes numéricas (geralmente algumas positivas e outras negativas). O passo inicial para a construção da Wavelet mãe é a obtenção de funções escala $\phi(x)$ através de um processo iterativo, cujo passo zero-ésimo pode ser (em nosso contexto será) uma função *caixa* $\phi(x)=1$ definida no intervalo [0,1), e uma função $\phi(2x)=0$, fora do intervalo. A equação (3.21) mostra a conexão entre um passo e o seguinte. Desde o primeiro passo da construção da função escala e também durante a construção da função

Wavelet será utilizado um conjunto particular de coeficientes, o qual faz parte dos resultados publicados por Daubechies, (1988).

$$c_0 = (1+\sqrt{3})/4 \qquad c_1 = (3+\sqrt{3})/4$$
$$c_2 = (3-\sqrt{3})/4 \qquad c_3 = -(\sqrt{3}-1)/4$$

Partindo da citada função *caixa*, a primeira iteração produz uma função escada no intervalo [0,2) e o processo continua até que $\phi_j(x)$ torne-se indistinguível de $\phi_{j-1}(x)$, como mostrado na figura 3.13.

$$\phi_j(x) = \sum_{k=0}^{N-1} c_k \phi_{j-1}(2x-k) \tag{3.21}$$

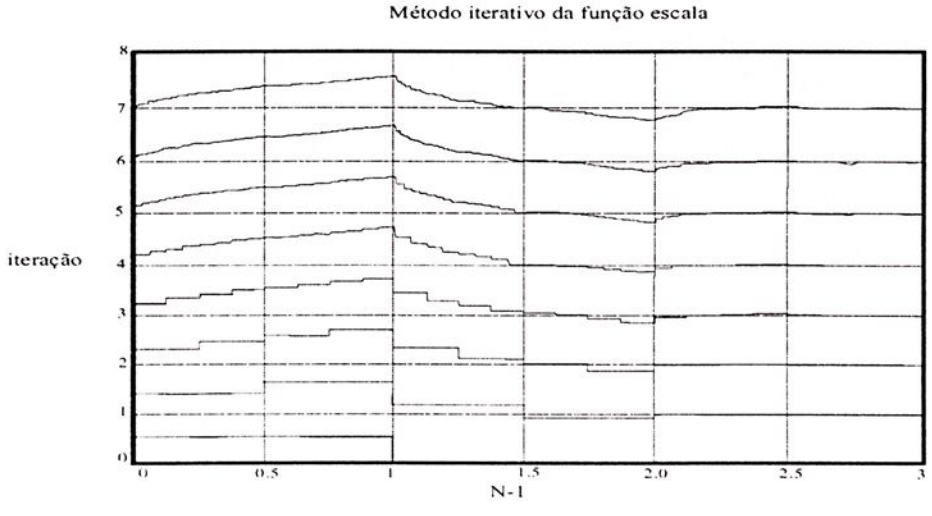
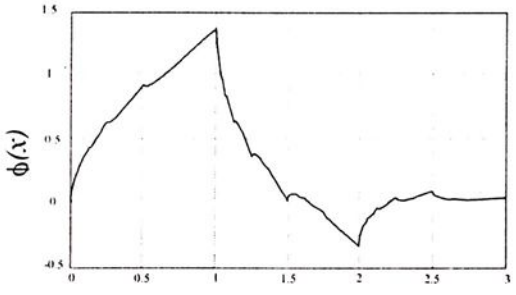


Figura 3.13 - Primeiras sete iterações na construção da função escala, N=4.

Uma característica transparente na construção da função escala é a natureza descontínua do processo. A mesma função é mostrada em escalas maiores na figura 3.14. Para obter uma função mais suave, é necessário incluir mais termos na equação (3.21).



N-1

Figura 3.14 - Função escala

A função $\phi(x)$ gerada a partir de uma *caixa* unitária (unitário em altura e comprimento) é chamada *função escala* e uma correspondente *função Wavelet* será construída a partir dela. Entretanto, veremos em mais detalhes como o método iterativo funciona. Representando a função *caixa* inicial pela ordenada 1 em $x = 0$, a primeira iteração, Fig. 3.13, fornece quatro novas ordenadas c_0, c_1, c_2 e c_3, correspondendo, respectivamente, aos intervalos [0, 0.5), [0.5, 1), [1, 1.5) e [1.5, 2). Na segunda iteração, a ordenada c_0 contribuirá para quatro novas ordenadas $c_0^2, c_0c_1, c_0c_2, c_0c_3$, referentes aos intervalos [0, 0.25), [0.25, 0.5), [0.5, 0.75), [0.75, 1), e a ordenada c_1 também contribui para quatro novas ordenadas $c_1c_0, c_1^2, c_1c_2, c_1c_3$, cujos intervalos referentes são: [0.5, 0.75); [0.75, 1), [1, 1.25), [1.25, 1.5), e assim por diante. Depois da segunda iteração completa, as ordenadas resultantes são:

$$c_0^2, c_0c_1, c_0c_2 + c_1c_0, c_0c_3 + c_1^2; c_1c_2 + c_2c_0; c_1c_3 + c_2c_1; c_2^2 + c_3c_0; c_2c_3 + c_3c_1; c_3c_2; c_3^2,$$

posicionadas em x = 0, 0.25, 0.5, 0.75, ...,2.25. Esse processo pode ser implementado na seguinte forma matricial:

$$[\phi]_2 = \begin{bmatrix} c_0 & 0 & 0 & 0 \\ c_1 & 0 & 0 & 0 \\ c_2 & c_0 & 0 & 0 \\ c_3 & c_1 & 0 & 0 \\ 0 & c_2 & c_0 & 0 \\ 0 & c_3 & c_1 & 0 \\ 0 & 0 & c_2 & c_0 \\ 0 & 0 & c_3 & c_1 \\ 0 & 0 & 0 & c_2 \\ 0 & 0 & 0 & c_3 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} [1] = M_2 M_1 [1]. \tag{3.22}$$

Generalizando-se,

$$[\phi]_r = M_r M_{r-1} \ldots\ldots M_1 [1], \tag{3.23}$$

sendo que M_r denota a matriz de ordem $(2^{r+1} + 2^r - 2) \times (2^r + 2^{r-1} - 2)$, na qual cada coluna é formada por coeficientes c_0, c_1, c_2, c_3, dispostos duas posições abaixo da coluna à sua esquerda. O número de pontos no gráfico cresce na seqüência: 1, 2, 4, 10, 22, 46, ..., $(2^{r+1} + 2^r - 2)$, ou seja, após oito iterações, atingimos $2^9 + 2^8 - 2 = 766$, com cada ponto espaçado de $\frac{1}{2^8} = \frac{1}{256}$ unidades ao longo do eixo horizontal.

3.3.2 FUNÇÃO WAVELET

Será definida agora a função Wavelet mãe, *W(x)*, a qual é construída a partir da função escala $\phi(x)$. Assim, tem-se:

$$W(x) = \sum_{k=0}^{N-1} (-1)^k c_k \phi(2x + k - N + 1) \tag{3.24}$$

Os mesmos coeficientes c_0, c_1, c_2 e c_3, utilizados na definição da função escala $\phi(x)$ serão aplicados para função Wavelet *W(x)*. Para isso, admitem-se as seguintes hipóteses: com apenas uma iteração concluem-se os estágios de construção da função escala, e que a função tenha apenas quatro ordenadas c_0, c_1, c_2 e c_3 no intervalo [0, 1.5). Assim, de acordo com a equação (3.24) estas quatro ordenadas geram dez novas espaçadas de 0.25. O termo $-c_3\phi(2x)$ gera as seguintes ordenadas: $-c_3c_0, c_3c_1, -c_3c_2 - c_3^2$; e o termo $c_2\phi(2x-1)$ gera: $c_2c_0, c_2c_1, c_2^2, c_2c_3$; deslocados dois lugares à direita, e assim por diante para os outros termos. Portanto, as dez novas ordenadas que formam a *W(x)* são:

$$-c_3c_0,\ -c_3c_1,\ -c_3c_2+c_2c_0,\ -c_3^2+c_2c_1,\ c_2^2-c_1c_0,\ c_2c_3-c_1^2,\ -c_1c_2+c_0^2,\ -c_1c_3+c_0c_1,\ c_0c_2,\ c_0c_3.$$

Matricialmente pode-se escrever:

$$[W]_2 = \begin{bmatrix} -c_3 & 0 & 0 & 0 \\ 0 & -c_3 & 0 & 0 \\ c_2 & 0 & -c_3 & 0 \\ 0 & c_2 & 0 & -c_3 \\ -c_1 & 0 & c_2 & 0 \\ 0 & -c_1 & 0 & c_2 \\ c_0 & 0 & -c_1 & 0 \\ 0 & c_0 & 0 & -c_1 \\ 0 & 0 & c_0 & 0 \\ 0 & 0 & 0 & c_0 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} [1], \tag{3.25}$$

ou alternativamente:

$$[W]_2 = \begin{bmatrix} c_0 & 0 & 0 & 0 \\ c_1 & 0 & 0 & 0 \\ c_2 & c_0 & 0 & 0 \\ c_3 & c_1 & 0 & 0 \\ 0 & c_2 & c_0 & 0 \\ 0 & c_3 & c_1 & 0 \\ 0 & 0 & c_2 & c_0 \\ 0 & 0 & c_3 & c_1 \\ 0 & 0 & 0 & c_2 \\ 0 & 0 & 0 & c_3 \end{bmatrix} \begin{bmatrix} -c_3 \\ c_2 \\ -c_1 \\ c_0 \end{bmatrix} [1]. \tag{3.26}$$

Observe os zeros entre as linhas diagonais na equação (3.25). Se as Wavelets fossem calculadas depois de primeiro completar duas iterações da função escala, haveria três zeros entre cada linha diagonal. Isto porque as ordenadas depois de duas iterações da função escala são espaçadas de x=0.25. Como um resultado do $2x$ nos termos do lado direito da equação (3.24), as ordenadas da função Wavelet serão separadas de 0.25/2=0.125. Entretanto, as translações requeridas pela equação (3.25) são x=0.5 (a função $\phi(2x-1)$) transladada de 0.5) com respeito a

$\phi(2x)$, e assim por diante. O resultado é que as linhas diagonais na equação (3.27) são separadas cada uma por 0.5/0.125=4 lugares:

$$[W]_3 = \begin{bmatrix} -c_3 & 0 & 0 & 0 \\ 0 & -c_3 & 0 & 0 \\ 0 & 0 & -c_3 & 0 \\ 0 & 0 & 0 & -c_3 \\ c_2 & 0 & 0 & 0 \\ 0 & c_2 & 0 & 0 \\ 0 & 0 & c_2 & 0 \\ 0 & 0 & 0 & c_2 \\ -c_1 & 0 & 0 & 0 \\ 0 & -c_1 & 0 & 0 \\ 0 & 0 & -c_1 & 0 \\ 0 & 0 & 0 & -c_1 \\ c_0 & 0 & 0 & 0 \\ 0 & c_0 & 0 & 0 \\ 0 & 0 & c_0 & 0 \\ 0 & 0 & 0 & c_0 \end{bmatrix} M_2 M_1 [1]. \tag{3.27}$$

Como pode ser verificado, as equações (3.25) e (3.26) são formas equivalentes, portanto, a equação (3.27) também tem a mesma forma:

$$W_3 = M_3 M_2 \begin{bmatrix} -c_3 \\ c_2 \\ -c_1 \\ c_0 \end{bmatrix} [1], \tag{3.28}$$

sendo que M_3 é uma matriz de ordem 22×10 com submatrizes $[c_0\ c_1\ c_2\ c_3]^t$, cada uma adaptada dois lugares abaixo de sua vizinha do lado esquerdo.

As equações (3.23) e (3.28) são importantes, pois formam a estrutura básica do algoritmo da transformada Wavelet discreta (TWD). Antes da apresentação deste algoritmo, valem alguns comentários sobre os coeficientes c_k. O cálculo da Eq. (3.28) é semelhante ao utilizado para construir a figura 3.13. O que difere um do outro é o primeiro passo $[-c_3\ c_2\ c_1\ -c_0]^t\,[1]$ no lugar de $[c_0\ c_1\ c_2\ c_3]^t\,[1]$. Todos os passos após a primeira iteração usam as matrizes M_r, que consistem de submatrizes $[c_0\ c_1\ c_2\ c_3]^t$ cada uma adaptada dois lugares abaixo de sua vizinha do lado esquerdo.

Os cálculos da equação (3.28) podem ser visualizados na figura 3.15, e em escalas maiores na figura 3.16.

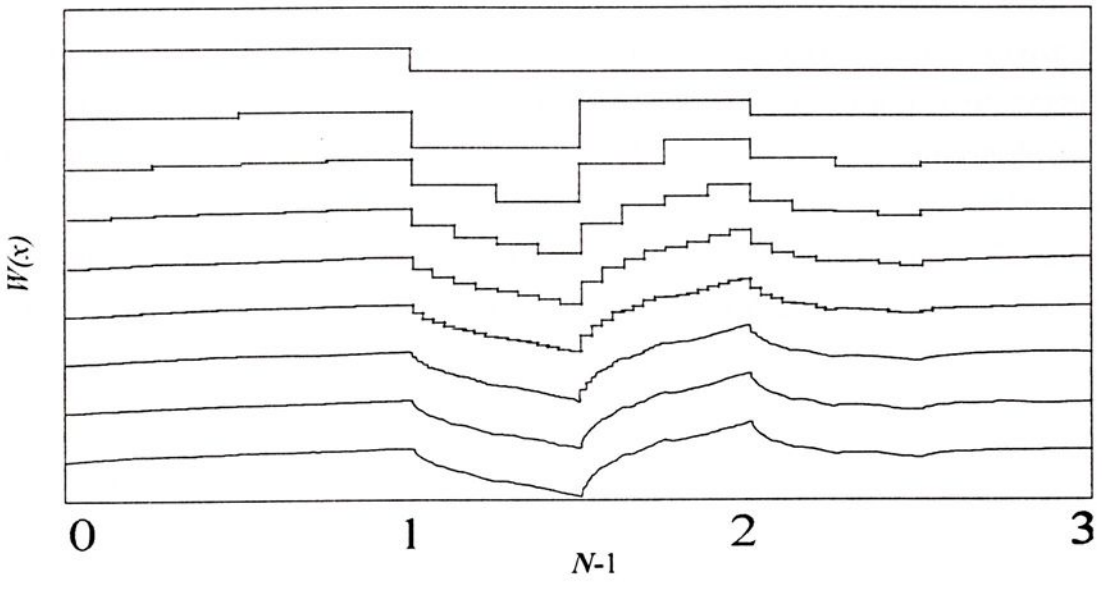


Figura 3.15 - As primeiras oito iterações do desenvolvimento da Wavelet db4.

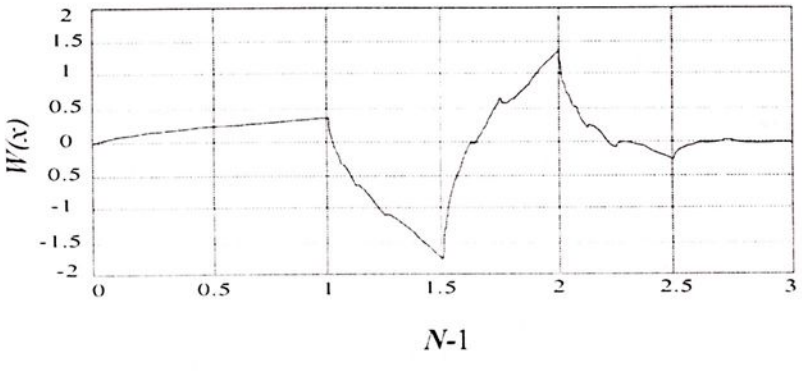


Figura 3.16 - Wavelet db4 construída a partir da função escala.

3.3.3 PROPRIEDADES DOS COEFICIENTES

Na construção de Wavelets, os coeficientes devem ser escolhidos cuidadosamente. Serão apresentadas condições que um bom conjunto de coeficientes de Wavelets devem satisfazer. É de interesse somente o caso em que o número de coeficientes (N) não nulos seja par, isto porque além de todas as condições terem como base esses coeficientes, será apresentado um algoritmo (TWD) que também utiliza esses coeficientes. Estas condições estão descritas em forma de equações, e os resultados são, tanto equações independentes quanto coeficientes. Existem três categorias de condições: conservação da área, ortogonalidade e exatidão.

Conservação da área

Será demonstrado como a área é conservada através das equações (3.20) e (3.24):

$$\phi(x) = \sum_{k=0}^{N-1} c_k \phi(2x-k), \tag{3.20}$$

$$W(x) = \sum_{k=0}^{N-1} (-1)^k c_k \phi(2x+k-N+1). \tag{3.24}$$

Integrando ambos os lados da equação (3.20),

$$\int_{-\infty}^{+\infty}\phi(x)dx = \sum_k c_k \int_{-\infty}^{+\infty}\phi(2x-k)dx\,, \tag{3.29}$$

mudando a variável de integração, $y = 2x - k$,

$$\int_{-\infty}^{+\infty}\phi(x)dx = \sum_k c_k \frac{1}{2}\int_{-\infty}^{+\infty}\phi(y)dy\,, \tag{3.30}$$

como a soma de todos coeficientes de Wavelet deve ser sempre igual a 2, tem-se:

$$\sum_{k=0}^{N-1} c_k = 2\,. \tag{3.31}$$

Aplicando a equação (3.31) para o método iterativo da equação (3.21), observa-se que $\int_{-x}^{x}\phi_j(x)dx$ permanece igual a $\int_{-x}^{+x}\phi_{j-1}(x)dx$, por isso que a área é conservada durante as iterações. Desde que a *caixa* de área unitária é usada para iniciar a iteração, segue que a área sob a função escala permanece igual à unidade:

$$\int_{-\infty}^{+\infty}\phi(x)dx = 1\,. \tag{3.32}$$

BCpIS - FEIS - UNESP

Exatidão

Um sinal pode ser decomposto via TWD em vários níveis de Wavelets, e este mesmo sinal pode ser sintetizado através da TWDI. Então, surge a preocupação com a exatidão, a qual depende da quantidade de coeficientes utilizados pela Wavelet. Para adquirir uma fiel representação do sinal os pesquisadores Strang e Fix (1973) e Strang (1989) apresentaram um conjunto de equações que garante a robustez na representação:

$$\sum_{k=0}^{N-1}(-1)^k k^m c_k = 0 \text{ para } m=0,1,\ldots, n\text{-}1. \tag{3.33}$$

A equação (3.33), mais as outras condições sobre os coeficientes de Wavelet, oferecem uma maior precisão na representação de um sinal.

Ortogonalidade

Para que o algoritmo da TWD funcione com eficiência na análise de um sinal, é imperativo que o processo iterativo seja reversível e esta necessidade impõe a ortogonalidade das matrizes M_r, isto é,

$$\alpha M_r^t M_r = \mathrm{I}, \tag{3.34}$$

sendo que α é um fator escalar. Assim, para tal condição deve-se satisfazer um conjunto de equações:

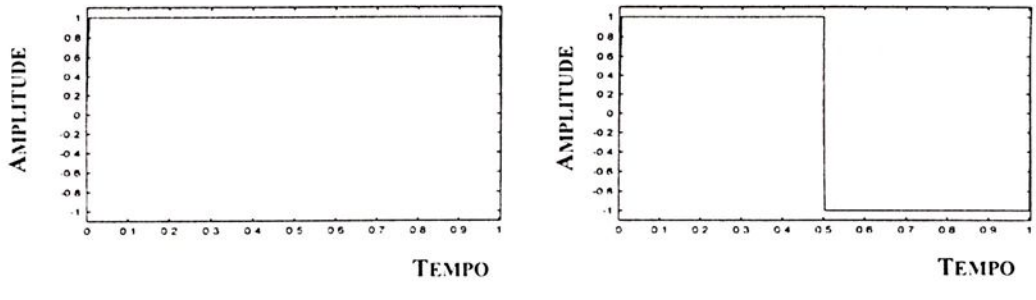
$$\sum_{k=0}^{N-1} c_k c_{k+2m} = 0\,, \qquad \text{para todos } m \text{ exceto para } m=0\,. \quad (3.35)$$

3.4 TRANSFORMADA WAVELET CIRCULAR

A meta é decompor qualquer sinal *f(x)* em uma série infinita de Wavelets de diferentes escalas e posições de acordo com a expansão:

$$f(x) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} c_{j,k} W(2^j x - k)\,. \quad (3.36)$$

Para interpretar esta expansão será utilizada a função escala constante no intervalo unitário, isto é, $\phi(x)$ para o intervalo [0, 1), e a função Wavelet de Haar, com N=2 coeficientes. Sabendo que todas as amplitudes $c_{j,k}$ da Wavelet são unitárias, Fig. 3.17. Esta expansão pode ser vista na figura 3.18.



(a) (b)

Figura 3.17 - (a) Função escala e (b) Função Wavelet de Haar.

Os níveis –3, -2 ,-1, 0, 1, 2 são mostrados na figura 3.18 e correspondem a j= -3, -2, -1, 0, 1, 2. No nível 0 existe $2^0 = 1$ Wavelet de Haar em cada intervalo unitário; no nível 1 existem $2^1=2$ Wavelets de Haar por intervalo unitário; no nível 2 existem $2^2 = 4$ Wavelets de Haar por intervalo unitário, e assim por diante para todos os níveis acima. Para o nível –1 existe uma Wavelet em dois intervalos; no nível -2 existe uma Wavelet em quatro intervalos, e assim segue para todos os níveis abaixo de -1. Para $j \leq -1$, a contribuição de cada nível é constante sobre todo o intervalo unitário, e a soma também é constante.

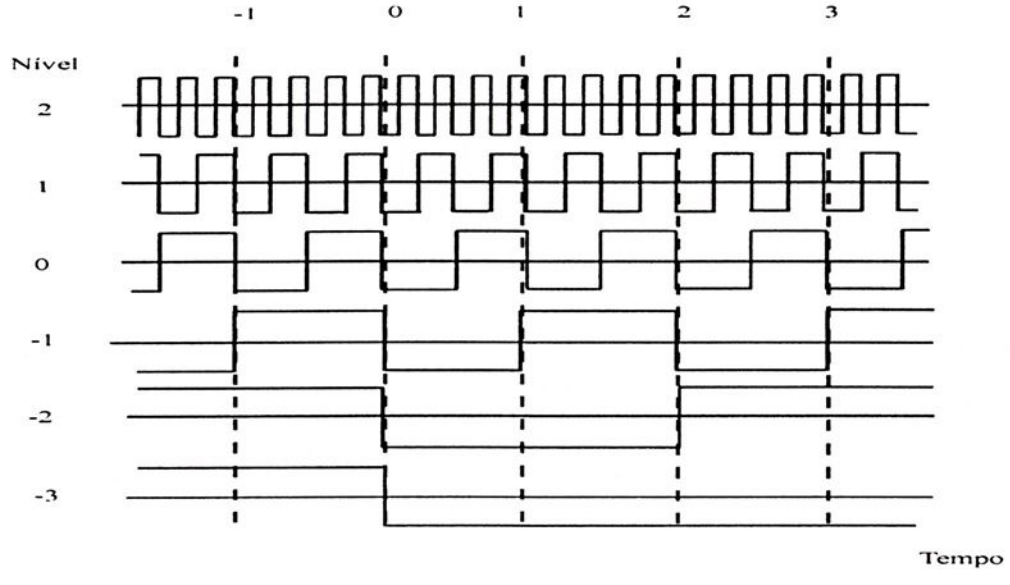


Figura 3.18 - Expansão da função Wavelet de Haar com amplitudes unitárias

Neste exemplo, representa-se a contribuição das escalas negativas por:

$$\sum_{j=-\infty}^{-1}\sum_{k=-\infty}^{+\infty} c_{j,k} W(2^j x - k) = \sum_{k=-\infty}^{+\infty} c_{\phi,k}\phi(x-k)\,, \tag{3.37}$$

sendo que, os $c_{\phi,k}$, são um novo conjunto de coeficientes. Lembrando que:

$$\phi(x) = \sum_{k=0}^{N-1} c_k \phi(2x-k), \tag{3.20}$$

e

$$W(x) = \sum_{k=0}^{N-1} (-1)^k c_k \phi(2x+k-N+1). \tag{3.24}$$

Quando j (escala) é negativo, $W(2^j x-k)$ pode ser sempre expresso como uma soma de termos semelhante à $\phi(x - k)$. Assim, a equação (3.37) é uma forma geral e por isso a equação (3.36) pode ser escrita na forma:

$$f(x) = \sum_{k=-\infty}^{+\infty} c_{\phi,k} \phi(x-k) + \sum_{j=0}^{+\infty} \sum_{k=-\infty}^{+\infty} c_{j,k} W(2^j x-k). \tag{3.38}$$

Pode ser mostrado que, sob condições gerais sobre $f(x)$ e $W(x)$, as séries representadas pelas equações (3.36) e (3.38) convergem, por isso proporcionam uma base prática para análise de sinais, Chui, (1992).

Com o objetivo de estabelecer um algoritmo para a transformada Wavelet discreta, é conveniente limitar a extensão da variável x a um intervalo unitário. Por exemplo, se a variável independente for o tempo t e o interesse está em trabalhar com um sinal de duração T, então

$x = t/T$ proporciona uma variável adimensional, assim, a função *f(x)* estará definida apenas no intervalo [0, 1).

Será considerada agora uma expansão da série de Wavelet no intervalo [0, 1). Entretanto, uma complicação ocorre nas bordas deste intervalo, porque algumas das Wavelets $W(2^j-k)$ podem ultrapassar o intervalo unitário, ou seja, sobrepõe-se nas extremidades. Para equações (3.20) e (3.24), elas sobrepõem-se em x=1, e algumas Wavelets podem sobrepor em ambas as bordas. Para evitar esse tipo de problema, é conveniente assumir que $f(x)$, $0 \le x < 1$, é um período de um sinal e por isso *f(x)* é repetida nos intervalos adjacentes, Portanto:

$$f(x)=\sum_{k=0}^{N-1} f(x-k) \tag{3.39}$$

Supõe-se que será utilizada a Wavelet db4 como exemplo de *W(x)*, Fig. 3.18, a qual cobre o intervalo [0, 3). No intervalo [0, 1), *f(x)* receberá contribuições do primeiro terço de *W(x)*; do terço médio *W(x+1)*; e do último terço *W(x+2)*. Isto é o mesmo que dizer que *W(x)* é uma função enrolada "*wrapped around*". Quando alguma Wavelet definida originalmente no intervalo [0, 1) vai além do extremo direito x=1 ela pode ser assumida dobrada sobre o intervalo (várias vezes se necessário), pois podem existir muitos coeficientes de maneira que a Wavelet eventualmente cobre muitos intervalos. Com essa hipótese, a partir da equação (3.38), a expansão de Wavelet de *f(x)* em [0, 1) pode ser escrita como:

$$f(x) = a_0\phi(x) + a_1W(x) + [a_2a_3]\begin{bmatrix} W(2x) \\ W(2x-1) \end{bmatrix} + [a_4a_5a_6a_7]\begin{bmatrix} W(4x) \\ W(4x-1) \\ W(4x-2) \\ W(4x-3) \end{bmatrix} + \ldots\ldots + a_{2^j+k}W(2^jx-k) + \ldots$$

(3.40)

sendo,

$$a_{2^j+k} = \int_0^1 f(x)W(2^j x - k)dx\,, \tag{3.41}$$

$$a_0 = \int_0^1 f(x)\phi(x)dx\,, \tag{3.42}$$

sendo o coeficiente a_0 a amplitude da função escala e os demais $a_1, a_2, a_3, a_4, ...$são as amplitudes da função Wavelet para um ciclo da função periódica representada pela equação (3.39) no intervalo [0, 1). Será visto mais tarde que a função escala $\phi(x)$ torna-se sempre uma constante. O segundo termo $a_1W(x)$ é uma Wavelet de nível zero; o terceiro e quarto termos, $a_2W(2x)$ e $a_3W(2x-1)$ são Wavelets de nível 1; a segunda inicia transladada de $\Delta x=\frac{1}{2}$ com respeito a primeira, os próximos quatro termos são Wavelets de nível 2, e assim por diante para Wavelets de níveis mais altos. Nos níveis mais altos, detectam-se os mais finos detalhes e no nível j existem 2^j Wavelets cada uma espaçada de $\Delta x=\frac{1}{2^j}$ ao longo do eixo x.

3.5 ALGORITMO DE MALLAT (TRANSFORMADA WAVELET DISCRETA)

A TWD é um algoritmo utilizado para calcular os coeficientes, equações (3.41) e (3.42), quando $f(x)$ é amostrado sempre com o mesmo espaçamento por todo intervalo [0, 1). O algoritmo da TWD foi proposto por Mallat, (1989).

Este algoritmo de decomposição utiliza a mesma estrutura do método de geração das Wavelets e, para melhor compreensão, inicialmente será apresentado o algoritmo inverso de recuperação de uma amostra de um sinal a partir de seus coeficientes Wavelet. Para tanto,

unesp

considera-se que o grupo de coeficientes abaixo é o resultado da decomposição de uma amostra discreta do sinal *f(x)*, Eq. (3.40), com oito pontos, e por Wavelets geradas por quatro coeficientes (*N*=4).

$$a = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 \end{bmatrix} \tag{3.43}$$

Observa-se que na Eq. (3.40) o elemento a_0 corresponde a contribuição de $\phi(x)$ (escala -1) na reconstrução do sinal. Sendo $\phi(x)$ gerada iterativamente a partir de uma função caixa unitária definida em [0, 1), pode-se gerar $a_0\phi(x)$ de forma análoga, a partir de uma função caixa com amplitude a_0, definida nesse intervalo. Então, o primeiro passo de iteração na forma matricial é:

$$\phi_1 = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} \begin{bmatrix} a_0 \end{bmatrix}. \tag{3.44}$$

Sabe-se que a função *caixa* ocupa o intervalo [0, 1), mas na figura 3.13 observa-se que a primeira iteração estende-se até o intervalo [0, 2). Se a parte que ultrapassa o intervalo é enrolada de volta, então obtém-se:

$$\phi_1 = \begin{bmatrix} c_0 + c_2 \\ c_1 + c_3 \end{bmatrix} \begin{bmatrix} a_0 \end{bmatrix}. \tag{3.45}$$

O segundo passo da iteração, antes de considerar o enrolamento, tem-se da equação (3.22):

$$[\phi]_2 = \begin{bmatrix} c_0 & 0 & 0 & 0 \\ c_1 & 0 & 0 & 0 \\ c_2 & c_0 & 0 & 0 \\ c_3 & c_1 & 0 & 0 \\ 0 & c_2 & c_0 & 0 \\ 0 & c_3 & c_1 & 0 \\ 0 & 0 & c_2 & c_0 \\ 0 & 0 & c_3 & c_1 \\ 0 & 0 & 0 & c_2 \\ 0 & 0 & 0 & c_3 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} [a_0], \tag{3.22}$$

A terceira linha da figura 3.13 mostra uma função com 10 ordenadas espalhadas pelo intervalo [0, 2.5). Estes 10 termos são:

$$\begin{aligned} [\phi]_2 = [\ & c_0^2 & c_0c_1 & c_0c_2 + c_1c_0 & c_0c_3 + c_1^2 \ldots\ldots \\ & c_1c_2 + c_2c_0 & c_1c_3 + c_2c_1 & c_2^2 + c_3c_0 & c_2c_3 + c_3c_1 \ldots \\ & c_3c_2 & c_3^2]^{t} [a_0]. \end{aligned} \tag{3.46}$$

Considerando o enrolamento e adicionando termos na mesma posição, no intervalo [0, 1), obtém-se os seguintes termos:

$$[\phi]_2 = \begin{bmatrix} c_o^2 + c_1c_2 + c_2c_0 + c_3c_2 \\ c_oc_1 + c_1c_3 + c_2c_1 + c_3^2 \\ c_0c_2 + c_1c_0 + c_2^2 + c_3c_0 \\ c_0c_3 + c_1^2 + c_2c_3 + c_3c_1 \end{bmatrix} [a_0],$$

unesp

$$[\phi]_2 = \begin{bmatrix} c_0 & c_3 \\ c_1 & c_2 \\ c_2 & c_0 \\ c_3 & c_1 \end{bmatrix} \begin{bmatrix} c_0 + c_2 \\ c_1 + c_3 \end{bmatrix} [a_0]. \tag{3.47}$$

Para o terceiro passo iterativo, a forma matricial de cálculo é a seguinte:

$$[\phi]_3 = \begin{bmatrix} c_0 & & & c_2 \\ c_1 & & & c_3 \\ c_2 & c_0 & & \\ c_3 & c_1 & & \\ & c_2 & c_0 & \\ & c_3 & c_1 & \\ & & c_2 & c_0 \\ & & c_3 & c_1 \end{bmatrix} \begin{bmatrix} c_0 & c_2 \\ c_1 & c_3 \\ c_2 & c_0 \\ c_3 & c_1 \end{bmatrix} \begin{bmatrix} c_0 + c_2 \\ c_1 + c_3 \end{bmatrix} [a_0] = M_3 M_2 M_1 [a_0]. \tag{3.48}$$

Como resultado, obtém-se um vetor com oito ordenadas no intervalo [0, 1) (ver a quarta linha da figura 3.13), representando a função escala enrolada e ponderada pelo coeficiente a_0.

Como pode-se notar, as matrizes possuem ordem $2^r \times 2^{r-1}$, sendo que cada coluna possui uma sub matriz de coeficientes posicionados duas linhas abaixo em relação a coluna à esquerda. Quando os elementos ultrapassam as dimensões de M_r, os elementos excedentes devem ser recolocados no topo da coluna, seguindo a ordem original. Assim, o algoritmo para gerar a contribuição de $a_0\phi(x)$ para $f(x)$ é:

$$f^{\phi}(x) = M_3 M_2 M_1 a_0, \tag{3.49}$$

unesp

ou na forma de diagrama:

$$a_0 = f^{\phi}(1) \xrightarrow{M_1} f^{\phi}(1:2) \xrightarrow{M_2} f^{\phi}(1:4) \xrightarrow{M_3} f(1:8), \qquad (3.50)$$

sendo que $f^{\phi}(1:8)$ é um vetor com oito elementos representando a contribuição de $a_0\phi(x)$ para a amostra do sinal *f(x)* em $[0,\frac{1}{8}),[\frac{1}{8},\frac{1}{4}),.....,[\frac{7}{8},1)$, Newland, (1993).

Retornando à equação (3.43) e considerando-se o segundo termo que é a amplitude da função Wavelet $W(x)$ (escala 0) na reconstrução do sinal. Novamente utilizou-se o processo iterativo, substituindo a função *caixa* unitária por uma amplitude a_1. A diferença em relação à construção da função escala, do método de produzir Wavelets, está na ordem e no sinal dos coeficientes na matriz M_1. A matriz resultante, após o enrolamento, é definida por G_1, então:

$$\begin{bmatrix} -c_3 \\ c_2 \\ -c_1 \\ c_0 \end{bmatrix} \text{torna - se } G_1 = \begin{bmatrix} -c_3 & -c_1 \\ c_2 & +c_0 \end{bmatrix}. \qquad (3.51)$$

Assim, analogamente à equação (3.49), os oito elementos que representam a contribuição de $a_1W(x)$ na reconstrução da amostra de sinal são obtidos através das seguintes operações:

$$f^{(0)}(1:8) = M_3M_2G_1a_1, \qquad (3.52)$$

ou na forma de diagrama:

unesp

$$a_1 = f^{(0)}(1) \xrightarrow{G_1} f^{(0)}(1:2) \xrightarrow{G_2} f^{(0)}(1:4) \xrightarrow{G_3} f^{(0)}(1:8). \quad (3.53)$$

No nível 1, o coeficiente a_2 corresponde à contribuição da Wavelet *W(2x)* na reconstrução do sinal. Essa Wavelet pode ser gerada pelo mesmo processo iterativo descrito no início do capítulo mas, por ser uma versão contraída à metade em relação à *W(x)*, deve-se partir de uma função unitária definida em [0,0.5). Assim, a contribuição de $a_2 W(2x)$ para a amostra de sinal é obtida partindo-se de uma função de amplitude a_2 definida neste intervalo. Na forma matricial, essa contribuição (após o enrolamento) é obtida por:

$$f^{(1,1)}(1:8) = M_3 \begin{bmatrix} -c_3 \\ c_2 \\ -c_1 \\ c_0 \end{bmatrix} [a_2]. \quad (3.54)$$

Ainda nesse nível também existe a $W(2x-1)$, versão deslocada de Δx=1/2 à direita em relação à *W(2x)*. Então, quando da geração de *W(2x-1)*, iniciando-se de uma função unitária definida em [0.5, 1) e, como no caso anterior, obtém-se $a_3 W(2x-1)$ a partir de uma função caixa de amplitude a_3 definida nesse intervalo. Na forma matricial, representa-se essa contribuição (após o enrolamento) da seguinte maneira:

$$f^{(1,2)}(1:8) = M_3 \begin{bmatrix} -c_1 \\ c_0 \\ -c_3 \\ c_2 \end{bmatrix} [a_3]. \quad (3.55)$$

unesp

Portanto, combinando as equações (3.54) e (3.55), obtém-se a contribuição total da escala 1 para a amostra de sinal. Definindo-se:

$$G_2 = \begin{bmatrix} -c_3 & -c_1 \\ c_2 & c_0 \\ -c_1 & -c_3 \\ c_0 & c_2 \end{bmatrix}, \tag{3.56}$$

então, a contribuição total será:

$$f^{(1)}(1:8) = M_3 \begin{bmatrix} -c_3 & -c_1 \\ c_2 & c_0 \\ -c_1 & -c_3 \\ c_0 & c_2 \end{bmatrix} \begin{bmatrix} a_2 \\ a_3 \end{bmatrix} = M_3 G_2 \begin{bmatrix} a_2 \\ a_3 \end{bmatrix}, \tag{3.57}$$

ou na forma de diagrama:

$$\begin{bmatrix} a_2 \\ a_3 \end{bmatrix} = f^{(1)}(1:2) \xrightarrow{G_2} f^{(1)}(1:4) \xrightarrow{M_3} f^{(1)}(1:8). \tag{3.58}$$

Os quatro elementos que restaram da equação (3.43), a_4, a_5, a_6, a_7 são as amplitudes das Wavelets *W(4x)*, *W(4x-1)*, *W(4x-2)*, *W(4x-3)* as quais são transladadas de Δx=1/4 com respeito à outra. Neste caso os dois primeiros passos são omitidos e, para os oito termos da seqüência, existe somente um passo a efetuar os cálculos. O intervalo [0, 1) é fracionado em oito

cm 1 2 3 4 5 6 7 8 unesp 12 13 14 15 16 17 18

subintervalos, por isso uma Wavelet é transladada de duas posições com respeito à sua vizinha. Cada Wavelet inicia-se com os elementos $[-c_3, c_2, -c_1, c_0]^t$ e num único passo de cálculo tem-se:

$$f^{(2)}(1:8) = \begin{bmatrix} -c_3 & & & -c_1 \\ c_2 & & & c_0 \\ -c_1 & -c_3 & & \\ c_0 & c_2 & & \\ & -c_1 & -c_3 & \\ & c_0 & c_2 & \\ & & -c_1 & -c_3 \\ & & c_0 & c_2 \end{bmatrix} \begin{bmatrix} a_4 \\ a_5 \\ a_6 \\ a_7 \end{bmatrix} = G_3 \begin{bmatrix} a_4 \\ a_5 \\ a_6 \\ a_7 \end{bmatrix}. \tag{3.59}$$

Como no caso das matrizes M_r, as matrizes G_r possuem ordem $2^r \times 2^{r-1}$, sendo que cada coluna possui uma submatriz de coeficientes posicionada duas linhas abaixo em relação à coluna à esquerda. A diferença está na ordem e no sinal dos coeficientes (igual ao método de construção das Wavelets) e, novamente, quando as submatrizes ultrapassam as dimensões de G_r, os elementos excedentes devem ser recolocados no topo da coluna, seguindo na mesma ordem.

Dessa maneira pode-se recuperar exatamente a amostra de um sinal com a somatória das contribuições de todas as escalas, ou seja:

$$f(1:8) = f^{\phi}(1:8) + f^{(0)}(1:8) + f^{(1)}(1:8) + f^{(2)}(1:8). \tag{3.60}$$

Na forma de diagrama:

$$\begin{array}{ccccccc} f^i(1) & \xleftarrow{M_1} & f^i(1:2) & \xleftarrow{M_2} & f^i(1:4) & \xleftarrow{M_3} & f^i(1:8) = f \\ \uparrow & & G_1 \uparrow & & G_2 \uparrow & & G_3 \uparrow \\ a = [a(1) & & a(2) & & a(3:4) & & a(5:8)] \end{array} \tag{3.61}$$

Considera-se agora o problema inverso, ou seja, obter os coeficientes Wavelet $a(1:2^j)$ a partir de uma amostra de sinal $f(1:2^j)$. Devido à ortogonalidade, esta inversão ocorre de maneira trivial. Como conseqüência da condição de ortogonalidade, tem-se:

$$\begin{aligned} \frac{1}{2}M_r^t M_r &= I & M_r^t G_r &= 0 \\ G_r^t M_r &= 0 & \frac{1}{2}G_r^t G_r &= I \end{aligned} \quad . \tag{3.62}$$

Definindo-se novos símbolos para as transpostas de M e G,

$$L = M^t \qquad H = G^t \tag{3.63}$$

então,

$$\begin{aligned} \frac{1}{2}LM &= I & LG &= 0 \\ HM &= 0 & \frac{1}{2}HG &= I \end{aligned} \quad , \tag{3.64}$$

e as dimensões das matrizes são:

$$\begin{aligned} G_r &= G(2^r \times 2^{r-1}) & H_r &= H(2^{r-1} \times 2^r) \\ L_r &= L(2^{r-1} \times 2^r) & M_r &= M(2^r \times 2^{r-1}). \end{aligned} \tag{3.65}$$

unesp

Assim, devido à equação (3.64), cada um dos passos que faz o ajuste do algoritmo de Mallat pode ser revertido, por exemplo:

$$a_0 = \frac{1}{2} L_1 \frac{1}{2} L_2 \frac{1}{2} L_3 f^{\phi} . \tag{3.66}$$

Considerando-se a troca d

e f^{ϕ} por:

$$f = f^{\phi} + f^{(0)} + f^{(1)} + f^{(2)}, \tag{3.67}$$

calcula-se,

$$\frac{1}{2} L_1 \frac{1}{2} L_2 \frac{1}{2} L_3 f . \tag{3.68}$$

Substituindo-se as contribuições $f^{(0)}$, $f^{(1)}$ e $f^{(2)}$, respectivamente pelas equações (3.52), (3.57) e (3.59), resultando em:

$$f = f^{\phi} + M_3 M_2 G_1 [a_1] + M_3 G_2 \begin{bmatrix} a_2 \\ a_3 \end{bmatrix} + G_3 \begin{bmatrix} a_4 \\ a_5 \\ a_6 \\ a_7 \end{bmatrix} . \tag{3.69}$$

unesp

Então,

$$\frac{1}{2}L_1\frac{1}{2}L_2\frac{1}{2}L_3 f = \frac{1}{2}L_1\frac{1}{2}L_2\frac{1}{2}L_3 f^{\phi} + 0 + 0 + 0\,, \tag{3.70}$$

de acordo com a equação (3.64), conclui-se que f^{ϕ} na equação (3.66) pode ser trocado por f, então:

$$a_0 = \frac{1}{2}L_1\frac{1}{2}L_2\frac{1}{2}L_3 f\,, \tag{3.71}$$

e conclusão análoga é sustentada para outros passos de forma que, no lugar da equação (3.52), tem-se:

$$a_1 = \frac{1}{2}H_1\frac{1}{2}L_2\frac{1}{2}L_3 f\,, \tag{3.72}$$

em vez da equação (3.57), tem-se:

$$\begin{bmatrix} a_2 \\ a_3 \end{bmatrix} = \frac{1}{2}H_2\frac{1}{2}L_3 f\,, \tag{3.73}$$

unesp

e em vez da equação (3.59) tem-se:

$$\begin{bmatrix} a_4 \\ a_5 \\ a_6 \\ a_7 \end{bmatrix} = \frac{1}{2} H_3 f \,. \tag{3.74}$$

Portanto, todo algoritmo de Mallat pode ser revertido, de forma que a equação (3.61) torna-se:

$$\begin{array}{ccccccc} f^i(1) & \xleftarrow{\frac{1}{2}L_3} & f^i(1:2) & \xleftarrow{\frac{1}{2}L_3} & f^i(1:4) & \xleftarrow{\frac{1}{2}L_3} & f^i(1:8)=f \\ \downarrow & & \frac{1}{2}H_1\downarrow & & \frac{1}{2}H_2\downarrow & & \frac{1}{2}H_3\downarrow \\ a=[a(1) & & a(2) & & a(3:4) & & a(5:8)] \end{array} \,. \tag{3.75}$$

Este é o algoritmo de Mallat para a TWD, o qual permite que o sinal $f(1:2^j)$ seja transformado em seus coeficientes $a(1:2^j)$ de uma maneira muito eficiente, Fig. 3.19. As matrizes L e H atuam como filtros passa baixa (a-aproximações) e passa alta (d-detalhes), respectivamente. Então, L filtra os detalhes mais finos, consequentemente será obtida uma representação mais grosseira em cada escala. Por outro lado H oferece uma visão mais fina em cada escala.

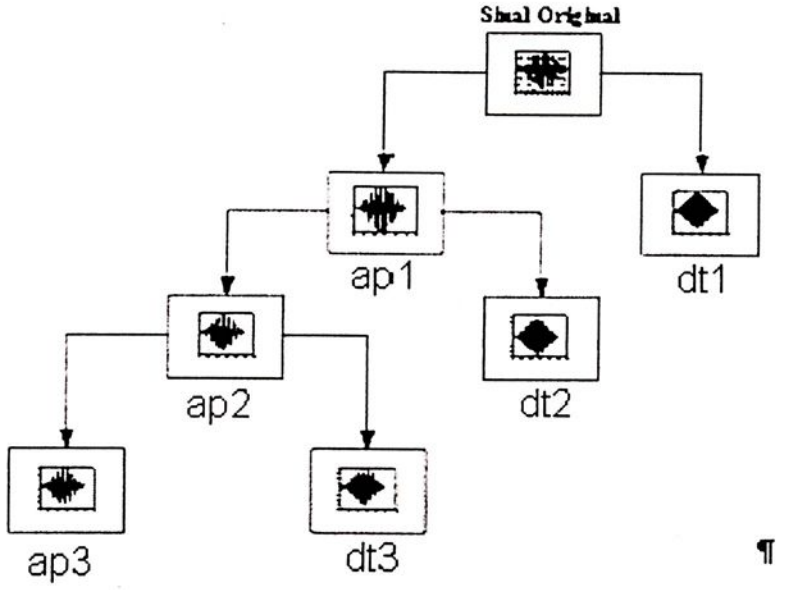


Figura 3.19 - Processo de decomposição em vários níveis.

3.6 TRANSFORMADA WAVELET CONTÍNUA

A Transformada Wavelet Contínua (TWC), de um modo similar à transformada de Fourier com janela (STFT), é definida utilizando o conjunto de funções base $W_{(a.b)}$ como função moduladora de f, então, tem-se:

$$TW_{(a,b)} = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) W\left(\frac{t-b}{a} dt\right), \tag{3.76}$$

e sua transformada inversa, Eq. (3.77):

$$f(t) = \frac{1}{C_W} \iint_{\Re} TW_{(a,b)} W_{(a,b)} dadb. \tag{3.77}$$

Para que existam as condições de operações da transformada direta/inversa, a "Wavelet mãe" deve satisfazer a condição de admissibilidade dada pela equação (3.78):

$$C_W = \int_{-\infty}^{+\infty} \frac{\left|\hat{W}(\omega)\right|^2}{|\omega|} d\omega < \infty, \tag{3.78}$$

sendo que $\hat{W}(\omega)$ denota a Transformada de Fourier de *W(t)*.

Uma aplicação da TF, TWC e TWD para um sinal será apresentada com o objetivo de confrontar as três técnicas. O sinal é formado por um seno com freqüência de 20 Hz e um impulso na posição 600 (0.5 s), Fig. 3.20. Os coeficientes gerados pelo cálculo da TF e TWC são representados pelo gráfico de freqüência, Fig. 3.21 e pelo mapa tempo-escala, Fig. 3.22, respectivamente. Esse mapa é um diagrama de contorno de uma superfície tridimensional, Fig. 3.23, e descreve a distribuição de energia do sinal (Newland, 1997). O sinal proposto está definido para um tempo de aquisição de 0.85 s e amostrado com 1024 pontos.

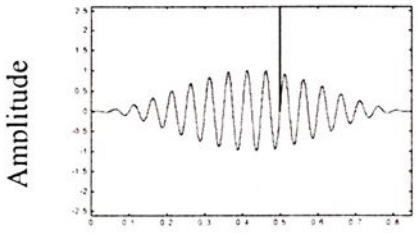


Figura 3.20 - Sinal temporal

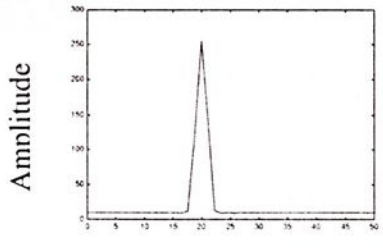


Figura 3.21 - FFT do sinal temporal

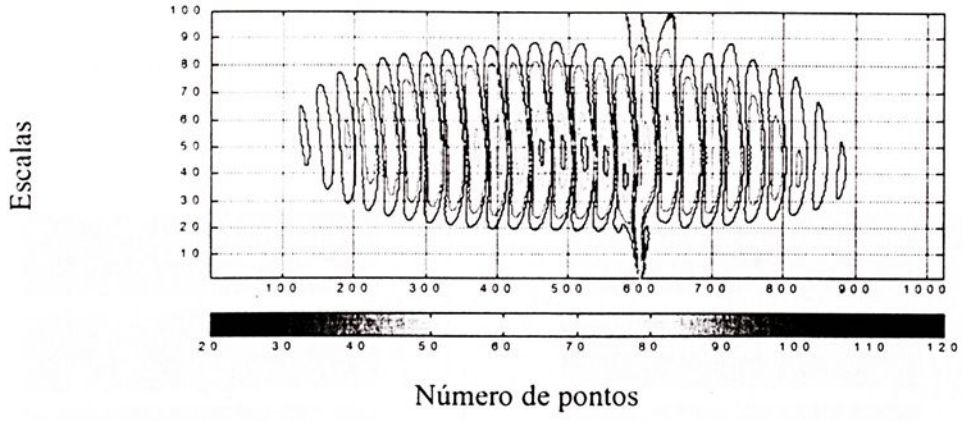


Figura 3.22 - TWC do sinal temporal.

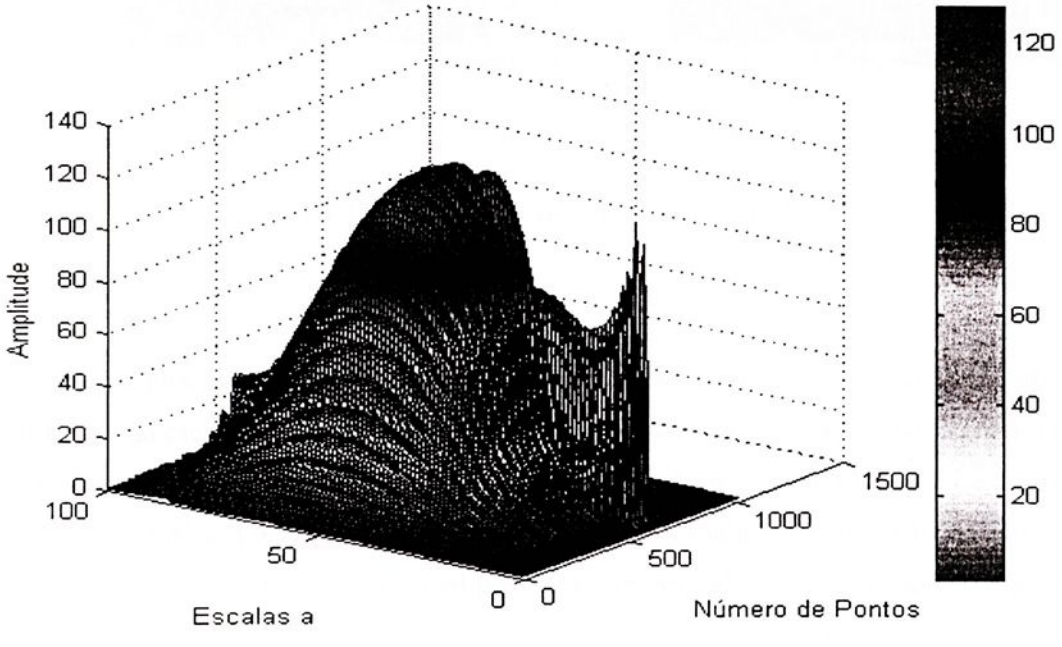


Figura 3.23 - Diagrama tridimensional da TWC para o sinal temporal.

O gráfico da FFT "*Fast Fourier Tranform*", Fig. 3.21, do sinal mostra somente um pico de amplitude representando a freqüência do sinal. Entretanto, a TWC do sinal temporal, Fig. 3.22, mostra o exato momento em que ocorre o impulso.

O mesmo sinal temporal da figura 3.20 foi utilizado para análise através da TWD e o exato momento do impulso também pode ser visto, Fig. 3.24.

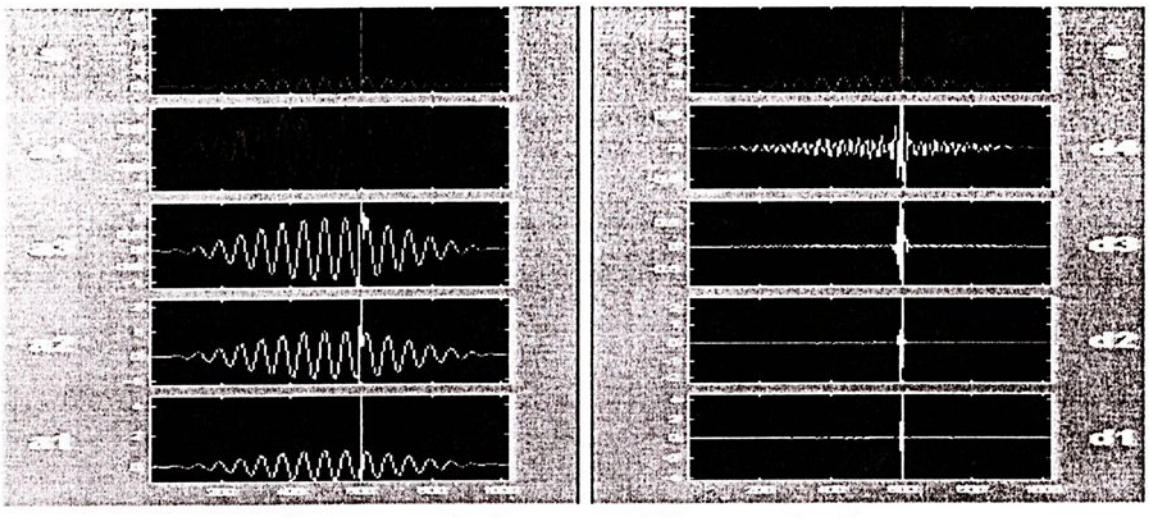


Figura 3.24 - TWD do sinal temporal, Fig. 3.20.

O sinal processado anteriormente pela TWC e TWD é simples e portanto não precisa de uma variação da escala para que se possa capturar o transitório e a senóide, assim, em um mesmo gráfico pode-se visualizar os dois eventos. Neste contexto, será considerada novamente a aplicação da TWC e TWD para um sinal temporal que simula uma condição de falha de um redutor de engrenagens, Fig.3.25. O sinal proposto está definido para um tempo de aquisição de 0.128 s e amostrados com 1024 pontos. Este sinal é composto de freqüência de engrenamento, 5% de ruído branco, e adicionalmente, um impulso a cada 256 pontos (ou 0.032 s) representando o efeito de um dente quebrado. Os coeficientes gerados pelo cálculo da TWC são representados pelo mapa tempo-escala, Fig. 3.26. Esse mapa é um diagrama de contorno de uma superfície tridimensional, Fig.3.27, que proporciona uma visão global do sinal. No mapa tempo-escala da figura 3.26 pode-se verificar facilmente o impulso devido à falha, mas não se consegue visualizar a faixa de escala referente à freqüência de engrenamento, portanto, deve-se variar a escala para se ter uma visão mais clara do evento, Fig. 3.28. A TWC é muito refinada, ou seja,

ela pode garantir mais exatidão sobre os eventos que se deseja visualizar. Mas esta melhor visualização tem um preço, que é um tempo de processamento elevado.

A detecção da falha ainda é feita pela TWD, que decompõe o sinal temporal em componentes de baixa freqüência (a-aproximações) e componentes de alta freqüência (d-detalhes). A TWD é uma transformada semelhante à TWC, e difere somente por operar em uma escala de potência de dois. A conseqüência disso é uma menor exatidão em relação à TWC, mas o tempo computacional é drasticamente reduzido. Para o presente estudo optou-se pela TWD. Apesar de uma menor exatidão em relação à TWC, ela ofereceu bons resultados nas análises dos eventos estudados. Isso pode ser visualizado na figura 3.29, onde as componentes de altas d3 e d4 mostram claramente os impulsos a cada 256 pontos (ou 0.032 s).

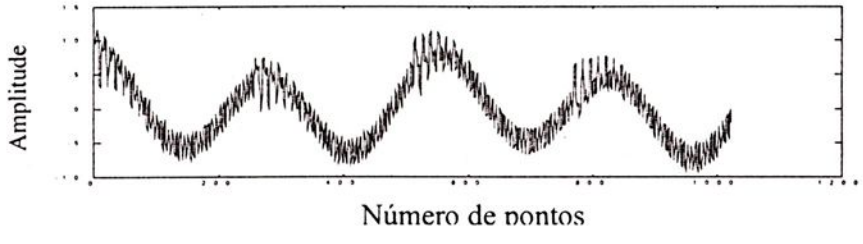


Figura 3.25 - Sinal temporal

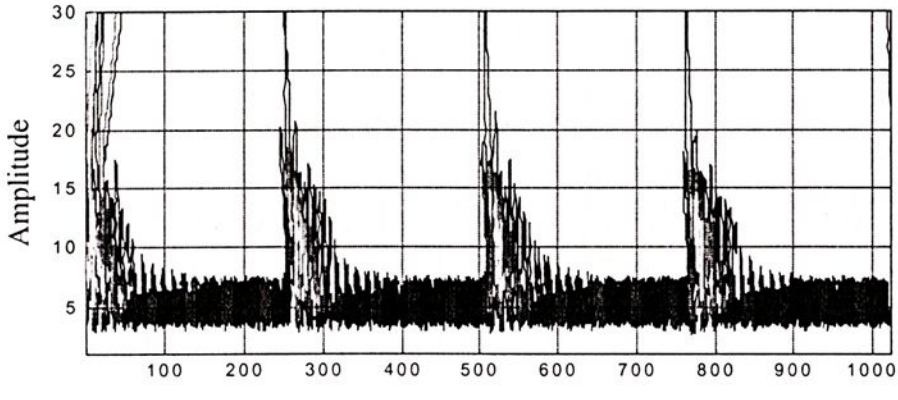


Figura 3.26 - Mapa tempo-escala (TWC)

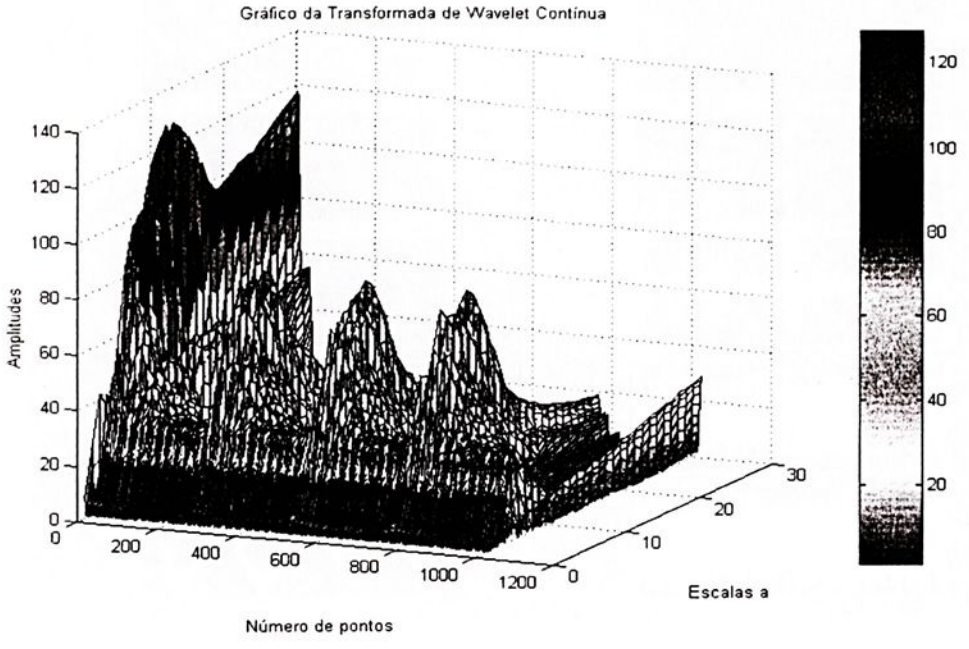


Figura 3.27 - Diagrama tridimensional da TWC para sinal temporal.

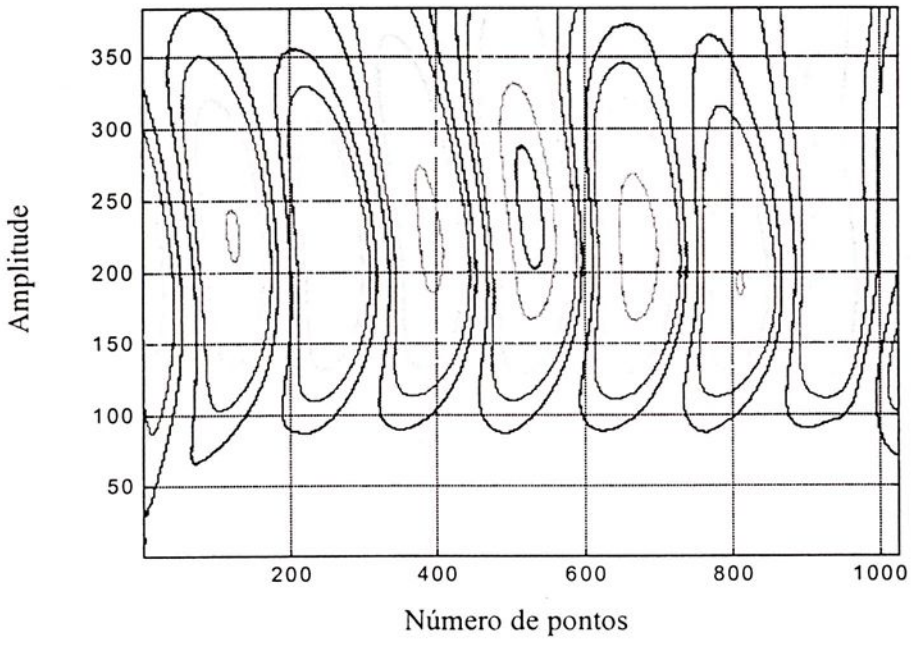


Figura 3.28 - TWC do sinal temporal.

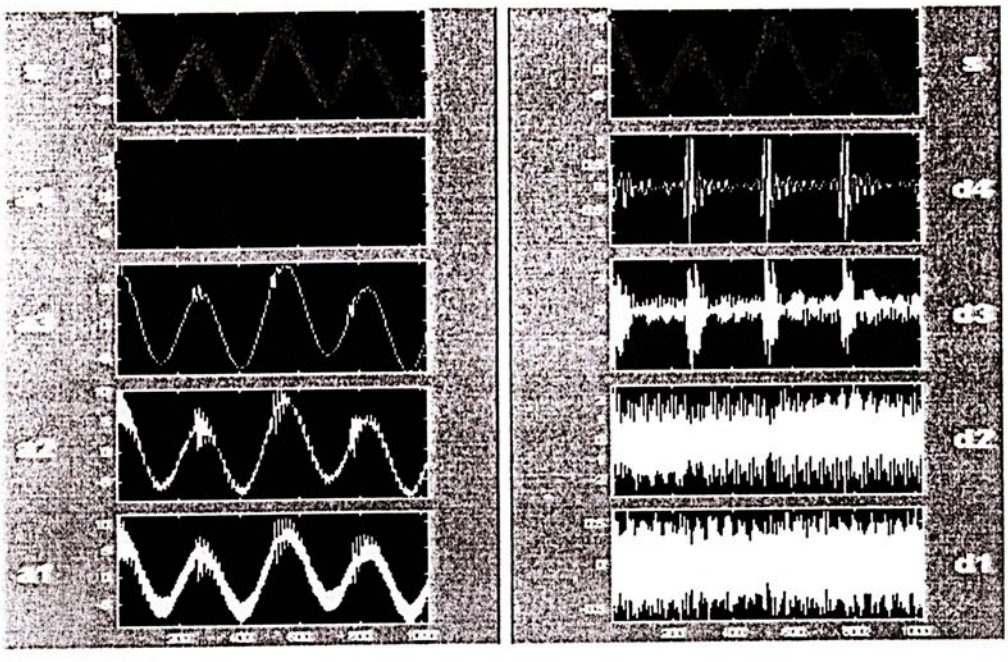


Figura 3.29 - TWD do sinal temporal.

CAPÍTULO 4

4 REDES NEURAIS ARTIFICIAIS

Modelos de redes neurais artificiais ou simplesmente "redes neurais" recebem muitos nomes tais como modelos de conexão, modelos de processamento, modelos de processamento distribuído em paralelo, e sistemas de neuromorfos. Qualquer que seja o nome, todos estes modelos tentam obter um bom desempenho via interconexão densa de elementos computacionais simples em paralelo. A estrutura da rede neural artificial é inspirada em pesquisas biológicas sobre o funcionamento do cérebro humano. O cérebro humano consiste aproximadamente de 100 bilhões de células nervosas, chamadas de neurônios. Uma célula neural, Fig. 4.1, é composta de dendritos, núcleo e axônio. As ligações entre os neurônios são chamadas de sinapses. Os dendritos recolhem os sinais dos outros neurônios, via sinapses, e conduzem estes sinais para o núcleo do neurônio. O núcleo é responsável pelo processamento que ocorre no neurônio. Os sinais provenientes dos dendritos são somados, e caso o total ultrapasse um determinado limiar "*Threshold*", a célula nervosa dispara e envia um sinal pelo axônio em direção às outras células.

unesp

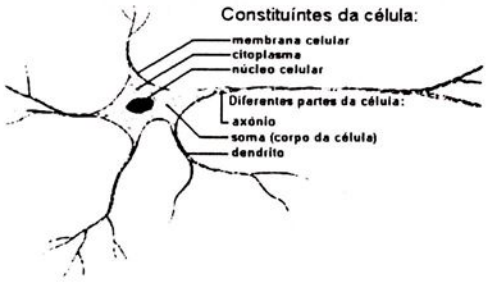


Figura 4.1 - Neurônio biológico.

As Redes Neurais Artificiais (RNA) são compostas de elementos (unidades de processamentos) que realizam, de maneira similar, a maior parte das funções elementares do neurônio biológico.

Os trabalhos sobre redes neurais artificiais têm uma longa história. O desenvolvimento de modelos matemáticos detalhados começaram há mais de 50 anos com o trabalho de McCulloch e Pitts (1943), Rosenblatt (1959), Widrow (1960).

As redes neurais são normalmente utilizadas em problemas de reconhecimento de padrões e, em muitos casos, têm mostrado sua superioridade sobre os métodos tradicionais. Modelos de redes neurais tem um maior potencial em áreas como reconhecimento de voz e imagem, onde muitas hipóteses são colocadas em paralelo e para casos onde altas velocidades computacionais são requeridas. Alguns estudos têm sidos elaborados no campo de identificação de sistemas e uma possível aplicação, no caso de identificação não-linear, é a detecção e localização de falhas em sistemas mecânicos. Worden e Tomlinson (1994) mostraram que uma RNA multi-camadas pode ser utilizada para modelar com acurácia um sistema dinâmico não-linear e detectar a influência de não-linearidades em sinais temporais de estruturas. Lopes *et al.*, (1997) discutem a utilização de RNAs para monitoramento e detecção de falhas em que técnicas de ajuste de modelos são utilizadas para geração dos dados de treinamento da rede. A técnica de ajuste de modelos fornece um modelo confiável que permite simular qualquer condição de dano, o qual estabelece uma correlação direta entre o desvio da resposta e a correspondente falha. O aprendizado da RNA é desenvolvido utilizando um sinal de espectro comprimido criado para cada tipo de falha.

unesp

4.1 FUNDAMENTOS DAS REDES NEURAIS ARTIFICIAIS

As RNAs são inspiradas em pesquisas desenvolvidas sobre a organização do cérebro em que se considera as configurações e os algoritmos utilizados para a transmissão de informações da rede. Entretanto, o conhecimento sobre a operação global do cérebro é muito limitado e há pouco conhecimento de seu processamento. Consequentemente, a configuração da rede deve procurar ter, além do conhecimento biológico corrente, estruturas com as funções mais usuais. A obtenção de uma RNA demanda o desenvolvimento de um modelo matemático que é inspirado no funcionamento do sistema biológico.

As RNAs geralmente têm um conjunto de camadas de elementos (neurônios), as quais são classificadas como camada de entrada[1], camada intermediária e camada de saída, Fig 4.2. As redes com um conjunto de neurônios arranjados em uma única camada são denominadas RNAs de simples camadas e as RNAs com mais de uma camada são denominadas de multi-camadas.

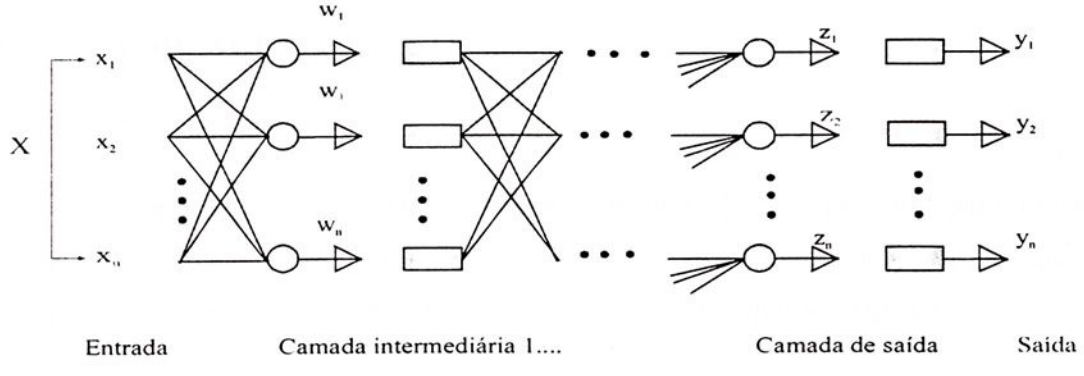


Figura 4.2 - Esboço de uma rede neural multi-camadas.

Cada neurônio, em uma camada, é ligado a todos os neurônios da camada seguinte. Essas ligações são ponderadas e os valores passam de um neurônio para o próximo. Os dados de entrada $\{x\}_i$ são ponderados pelos coeficientes $\{w\}_i$ e passados para a primeira camada

[1] **Camada de entrada:** alguns autores não a classificam como camada, pois não possue neurônios com capacidade de processamento, portanto não é computada no total de camadas de uma RNA.

intermediária, e os neurônios desta camada são ativados e produzem as saídas, as quais têm como base a somatória dos pesos dos valores passados a elas pela função de ativação. Da mesma forma, a primeira camada intermediária passa os valores à segunda camada (se ela existir) e assim, sucessivamente. No final, os valores são passados para a camada de saída que produz as saídas da rede. As saídas da rede são repetidamente comparadas com as saídas desejadas e a cada iteração os pesos de conexão são ajustados na direção das saídas desejadas. Este processo é denominado de "treinamento" ou "aprendizado" da rede. Este aprendizado é feito através dos ajustes dos pesos de interconexão entre as camadas. Este primeiro estágio é o mais importante.

Um dos algoritmos mais utilizados no treinamento de redes neurais de multi-camadas é o "*Back-Propagation*", o qual será brevemente discutido no item 4.7. Se uma rede é bem treinada, isto é, ela tem uma boa capacidade de generalização, ela pode fazer o trabalho de classificação correto, inclusive para novos dados de entrada que não foram utilizados no treinamento. As informações contidas nas entradas e saídas da rede serão melhor detalhadas nos exemplos apresentados.

4.2 O NEURÔNIO ARTIFICIAL

O neurônio artificial (unidades de processamento) foi desenvolvido para imitar as características principais do neurônio biológico. O primeiro modelo de unidade de processamento foi proposto por McCulloch e Pitts no início dos anos 40, Fig. 4.3.

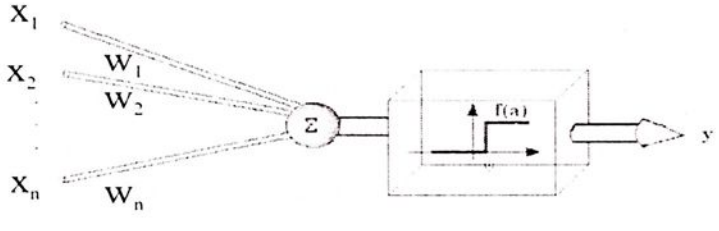


Figura 4.3 - Neurônio artificial projetado por McCulloch

Esta unidade de processamento pode ser dividida em duas partes. A primeira parte engloba as conexões de entrada da unidade. Em essência, um conjunto de entradas é aplicado, e cada entrada representa a saída de um outro neurônio. Cada entrada é multiplicada por um peso correspondente, análogo ao comprimento sináptico, e todos os pesos de entrada são, então, somados em um determinado nível de ação do neurônio. Apesar da diversidade de padrões da rede, todas têm como base essa configuração. Aqui, um grupo de entradas (x_1, x_2,..., x_n) é aplicado ao neurônio artificial. Essas entradas, vetor $\{X\}$, correspondem aos sinais dentro da sinapse de um neurônio biológico. Cada sinal é multiplicado por um peso associado (w_1, w_2,..., w_n) e é aplicado a um bloco, denominado soma ($\sum$). Cada peso corresponde ao comprimento de uma simples conexão sináptica biológica (o grupo de pesos é referido coletivamente como um vetor W). O bloco de soma correspondente, simbolicamente à célula biológica, todos os pesos de entrada são adicionados algebricamente, produzindo uma saída denominada NET. Isto pode ser descrito em notação vetorial por:

$$NET = \{x\}^t \{w\} . \tag{4.1}$$

A segunda parte do neurônio artificial recebe a saída do bloco de soma (NET), que é processado por uma função de ativação F para produzir a saída do neurônio (Wasserman, 1989).

$$y = F(NET), \tag{4.2}$$

F pode ser uma função linear simples, uma função "*threshold*" ou uma função mais complexa capaz de simular características de transferência não-lineares do neurônio biológico. No neurônio de McCulloch-Pitts, a função de transferência consiste de um degrau. No caso de uma RNA com características não-lineares, uma das funções de transferência não-linear mais utilizada é a sigmoidal, que restringe a saída do neurônio entre 0 e 1. O desenvolvimento das redes levou à utilização de vários tipos de função de transferência. A figura 4.4 mostra

cm 1 2 3 4 5 6 7 unesp 12 13 14 15 16 17 18

esquematicamente algumas das principais funções: rampa, tangente hiperbólica, logística, degrau e gaussiana.

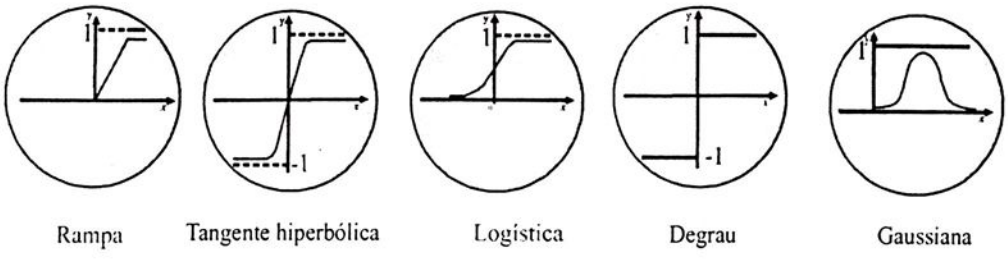


Figura 4.4 - Funções de transferência.

4.3 REDES NEURAIS DE SIMPLES CAMADAS

A rede mais simples é composta por um grupo de neurônios arranjados em uma camada, como mostra o lado direito da figura 4.5. Nota-se que os círculos do lado esquerdo da figura servem somente para distribuir as entradas. Eles não realizam computação e, conseqüentemente, não constituem uma camada. Por essa razão, são mostrados como círculos para distinguí-los dos neurônios de computação, que serão mostrados como quadrados. O conjunto de entradas {X} tem cada um dos seus elementos conectados a cada um dos neurônios artificiais ponderados por w. Cada neurônio emite uma soma ponderada das entradas para rede. As redes artificiais atuais e biológicas têm muitas das conexões apagadas, mas toda conectividade é mostrada neste caso por razões de generalidade. É conveniente considerar os pesos como elementos $w_{i,j}$ de uma matriz W. As dimensões da matriz são m linhas por n colunas, sendo que i= 1, 2, ...,m são as entradas e j=1, 2,.....,n são os neurônios.

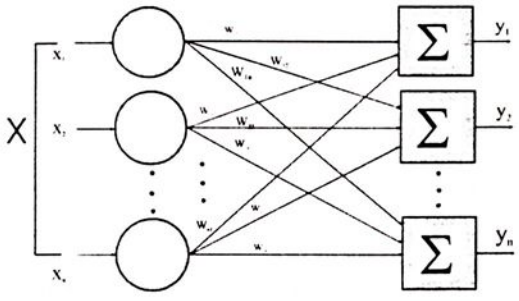


Figura 4.5 - Rede Neural com uma camada.

4.4 REDES NEURAIS ARTIFICIAIS DE MULTI-CAMADAS

Em geral, redes mais complexas oferecem maior capacidade computacional. Embora as redes têm sido construídas nas mais variadas configurações, o arranjo de neurônios em camadas imita a estrutura de certas partes do cérebro. Dessa forma, as redes de multi-camadas são providas de capacidade além daquela de simples camadas e, nos últimos anos, diferentes algoritmos têm sido desenvolvidos para treiná-las.

Desde que se determinaram as limitações, quanto à capacidade de representação das redes neurais com uma camada de neurônios, ficou claro que tais limitações podiam ser superadas pelas redes com mais de uma camada de neurônios. As redes de multi-camadas podem ser formadas em cascata a partir de um grupo de camadas simples. A saída de uma camada provém da entrada da camada subsequente, Fig. 4.3. Entretanto, a utilização de redes multi-camadas não aumenta a capacidade computacional em relação à rede de simples camada a menos que se utilize uma função de transferência não-linear entre as camadas como mostrado em Wasserman, (1989).

A capacidade de representação de uma rede com mais de uma camada de neurônios pode ser entendida como uma combinação das capacidades de representação de cada rede individualmente. Esta combinação dependerá dos pesos associados as conexões entre a primeira e última camada bem como da função de transferência dos neurônios nas camadas.

4.5 TREINAMENTO DAS REDES NEURAIS

Em redes neurais artificiais, considera-se como aprendizado o processo de determinação dos pesos, de maneira que a rede possa implementar a função desejada. Os métodos de treinamento de uma rede neural podem ser divididos em duas categorias: aprendizado supervisionado e aprendizado não supervisionado.

O aprendizado supervisionado consiste em apresentar à rede neural uma série de dados de entrada e as respectivas saídas esperadas. Durante o treinamento, um estímulo de entrada é aplicado e isso resulta em uma reação de saída. Esta reação é comparada com o sinal de saída desejado. Se a saída atual diferenciar da saída desejada, a cadeia neural gera um sinal de erro, que é então utilizado no cálculo de ajuste dos pesos de ponderação para que a saída atual seja corrigida, de forma alcançar a saída desejada ou meta. Em outras palavras, o erro é minimizado iterativamente em um processo de minimização de erro que requer um circuito especial como um professor ou supervisor, por isso o nome aprendizagem supervisionada (Wasserman, 1989). Se o erro não for suficientemente pequeno, o processo é repetido iterativamente até se obter um erro menor do que aquele pré-fixado e, então, finaliza-se o aprendizado. Na cadeia neural artificial, a quantidade de cálculo requerida para minimizar o erro depende do algoritmo utilizado, e isto é puramente uma ferramenta matemática derivada das técnicas de otimização. O treinamento supervisionado, apesar de muitas aplicações com sucesso, tem sido criticado por não apresentar uma concepção biologicamente plausível. É difícil conceber um mecanismo de treinamento do cérebro que compara saídas desejadas e reais, realimentando o processo de correções da rede. O treinamento não-supervisionado é mais plausível, e neste sentido, ele se assemelha ao modelo de entendimento do sistema biológico.

O treinamento não-supervisionado não requer o estabelecimento de um vetor alvo para as saídas e, consequentemente, também não requer comparações com saídas pré-determinadas, pois utiliza apenas um vetor de entrada. O algoritmo de treinamento modifica os pesos da rede para produzir vetores de saída consistentes. O processo de treinamento, por essa razão, extrai as propriedades estatísticas do grupo de treinamento e agrupa os vetores similares em classes. Aplicando um vetor de uma dada classe, esta entrada produzirá um vetor de saída específico, mas não há como determinar, a priori, qual padrão de saída específico será produzido por uma

cm 1 2 3 4 5 6 7 8 unesp 12 13 14 15 16 17 18

dada classe de vetor de entrada. Conseqüentemente, as saídas da rede devem, geralmente, ser transformadas em forma compreensível para o processo de treinamento.

Em contraste com a aprendizagem supervisionada, a aprendizagem não-supervisionada não requer um professor, isto é, não há alvo de saída. Ainda que a aprendizagem não-supervisionada não requeira um professor, ela necessita de princípios gerais para determinar como formar-se-ão os grupos. O agrupamento pode ser baseado em forma, cor, consistência material ou sob alguma outra propriedade do objeto. Portanto, se os princípios gerais não foram dados, e nem os tipos de características que devem ser usadas para o agrupamento dos objetos, a aprendizagem não-supervisionada pode ou não ser bem sucedida, (Wasserman, 1989).

4.6 REGRA DELTA

A regra delta foi a principal inovação introduzida por Widrow (1959) para o treinamento de redes do tipo ADELINE, que são compostas por uma unidade de processamento com função de transferência linear, por isso o nome ADELINE (ADAPTIVE LINEAR ELEMENT). Este modelo foi posteriormente estendido para o caso de redes multi-camadas denominadas MADELINE (MULTI ADAPTIVE LINEAR ELEMENT). Em seguida, a regra delta foi adaptada para outros tipos de redes neurais, formando a base de um dos tipos de redes mais empregado atualmente, as redes multi-camadas com retro-propagação do erro, (*Back-Propagation*). A diferença básica do algoritmo de treinamento da rede ADALINE em relação ao perceptron, que utiliza valores de saída binários, é que esta utiliza uma saída contínua, (Wasserman, 1989).

A regra delta para uma rede neural com uma camada de neurônios e com função de transferência linear é de aplicação bastante restrita, devido a uma série de limitações. Entretanto, desenvolvimentos subsequentes levaram à aplicação da regra delta a redes com múltiplas camadas de neurônios e função de transferência não linear. Diferentemente do perceptron, que verifica se a saída da rede é correta ou incorreta, na regra delta tem um erro em que a saída é comparada com a saída desejada, d, o que permite a introdução do termo δ,

unesp

$$\delta = (d\text{-}y), \tag{4.3}$$

Para o caso $\delta = 0$, a saída é correta e não são necessários cálculos adicionais, e para $\delta > 0$ e $\delta < 0$ a saída é incorreta. No primeiro caso, adiciona-se cada valor de entrada nos correspondentes pesos, e no segundo, os valores de entrada são subtraídos dos pesos. Em qualquer caso, o algoritmo é satisfeito se δ é multiplicado pelo valor de cada entrada x_i e este produto é adicionado ao peso correspondente. Generalizando, define-se uma "taxa de aprendizagem" representada pelo coeficiente γ, que multiplica o produto δx_i. Isso permite o controle do tamanho médio da mudança dos pesos,

$$\Delta = \gamma \delta x_i \tag{4.4}$$

$$w_i(n+1) = w_i(n) + \Delta_i \tag{4.5}$$

sendo:

Δ_i: correção associada com a *i-ésima* entrada x_i;
$w_i(n+1)$: valor de ponderação *i* após o ajuste;
$w_i(n)$: o valor de ponderação *i* antes do ajuste.

A regra delta modifica os pesos aproximadamente a partir das entradas desejadas e calculadas.

4.7 ALGORITMO DE RETROPROPAGAÇÃO ("*BACKPROPAGATION*")

O algoritmo de retropropagação "*Backpropagation*", foi criado para generalizar o controle instruído Widrow e Hoff para redes de trabalho MADALINE (múltiplas camadas), ou seja, ele é utilizado para o ajuste dos pesos e dos "*bias*" (utilizado para equiparar a origem da função de ativação produzindo um efeito que é similar a ajustar o "*threshold*" do neurônio, permitindo uma convergência mais rápida do processo de treinamento) da rede, com o objetivo de minimizar a soma dos erros quadráticos. A idéia do *Back-propagation* é determinar os erros das camadas ocultas pelo erro propagado na ordem inversa das unidades de saída. É um procedimento sistemático de adaptação dos pesos de RNA com base no método do gradiente descendente.

O *Back-propagation* treinado pode conduzir a um mínimo local., que pode ser satisfatório. Caso contrário, uma rede com mais camadas e neurônios pode fazer uma melhor tarefa. De qualquer modo, o número de neurônios ou camadas a aumentar pode não ser óbvio.

Derivadas do erro (chamadas vetores delta) são calculadas para as camadas de saída da rede e, estas são depois utilizadas para completar a rede e serem avaliadas em cada camada oculta.

Para treinar uma rede, vetores são apresentados à rede e os seus vetores de saída são calculados. A soma dos erros quadrados é então calculada. Se a soma dos erros quadrados de todos os vetores for menor que o erro meta, o treinamento termina. Se os erros da rede forem pequenos o bastante, os vetores podem ser apresentados para a rede e ela responde com um vetor de saída correto.

O algoritmo de retropropagação consiste na adaptação de pesos, tal que, o erro quadrático da rede é minimizado. A soma do erro quadrático instantâneo de cada neurônio alocado na última camada (saída da rede) é dada por:

$$\varepsilon^2 = \sum_{i=1}^{ns} \varepsilon_i^2 \qquad (4.6)$$

unesp

sendo:

ε_i: d_i - y_i ;
d_i: saída desejada do *i-ésimo* elemento da última camada da rede;
y_i: saída do *i-ésimo* elemento da última camada da rede;
ns: número de neurônios da última camada da rede.

Considerando-se o neurônio de índice *i* da rede, e utilizando-se o método do gradiente descendente, o ajuste de pesos pode ser formulado como:

$$W_i(n+1) = W_i(n) - \theta_i(n), \tag{4.7}$$

sendo:

$\theta_i(n)$: $\gamma\,[\nabla_i(n)]$;
γ: parâmetro de controle da estabilidade ou taxa de treinamento;
n: representa o índice de iteração;
$\nabla_i(n)$: gradiente do erro quadrático com relação aos pesos do neurônio *i* avaliado em *n*
W_i: vetor contendo os pesos do neurônio *i*
$[\, w_{0i} \;\; w_{1i} \;\; w_{2i} \;\; \ldots \;\; w_{ni} \,]^T$

Na equação (4.7) a direção adotada para se minimizar a função objetivo do erro quadrático corresponde à direção contrária ao gradiente. O gradiente $\nabla_i(n)$ pode ser expresso por:

$$\nabla_i(n) = \frac{\partial \varepsilon_i^2}{\partial W_i(n)} = 2\,\varepsilon_i\,\frac{\partial \varepsilon_i}{\partial W_i(n)}\,. \tag{4.8}$$

A função não-linear adotada é a função sigmóide definida por:

$$y = \frac{1 - e^{(-\lambda S)}}{1 + e^{(-\lambda s)}}, \tag{4.9}$$

sendo:

λ: constante que determina a inclinação da curva y_i.

Diferenciando a equação (4.6) com relação ao vetor w_k, obtém-se:

$$\frac{\partial \varepsilon_i}{\partial W_i} = -\frac{\partial y_i}{\partial W_i} = -\frac{\partial y_i}{\partial s_i}\,\frac{\partial s_i}{\partial W_i}. \tag{4.10}$$

assim, tem-se:

$$\frac{\partial \varepsilon_i}{\partial W_i} = -\,sgm'(s_i)\,\frac{\partial s_i}{\partial W_i}, \tag{4.11}$$

em que:

$sgm'(s_i) = \dfrac{\partial y_i}{\partial s_i}$, derivada parcial de y_i com relação a s_i.

Observa-se que:

$$\frac{\partial s_i}{\partial W_i} = X_i \tag{4.12}$$

sendo:

X_i: vetor padrão;
$[\, x_{0i} \;\; x_{1i} \;\; x_{2i} \;\; \ldots \;\; x_{ni} \,]^T$.

Substituindo na equação (4.11), tem-se:

$$\frac{\partial \varepsilon_i}{\partial W_i} = - sgm'(s_i) \;\; X_i \tag{4.13}$$

Inserindo-se a equação (4.13) na equação (4.8), resulta em:

$$\nabla_i(n) = - 2 \, \varepsilon_i \;\; sgm'(s_i) \;\; X_i \tag{4.14}$$

Utilizando o gradiente estimado no método do gradiente descendente, obtém-se o seguinte algoritmo:

$$W_i(n+1) = W_i(n) + 2\,\gamma\,\varepsilon_i\; sgm'(s_i)\; X_i \tag{4.15}$$

O treinamento através do algoritmo de retropropagação, é iniciado pela apresentação de um padrão $\{X\}$ à rede, o qual produzirá uma saída $\{Y\}$. Em seguida, calcula-se o erro de cada saída (diferença entre o valor desejado e saída). O próximo passo consiste na determinação do erro propagado no sentido inverso, através da rede associada à derivada parcial do erro quadrático de cada elemento com relação aos pesos e, finalmente, ajusta-se os pesos de cada elemento. Um novo padrão é apresentado e assim, o processo é repetido, para todos os padrões, até a convergência (erro $\leq$ tolerância pré-estabelecida). Os pesos iniciais são normalmente adotados como números randômicos, visto que a inicialização nula quase sempre proporciona baixa convergência.

O processo de treinamento pode ser visto contendo cinco fases:

- Inicialização dos pesos: os pesos de cada conexão são inicializados aleatoriamente, visto que valores iniciais nulos em redes multi-camadas quase sempre proporcionam baixa convergência;

- Apresentação das entradas e saídas desejadas: submete-se a RNA a um conjunto de treinamento;

- Cálculo das saídas da RNA: as entradas são propagadas através da RNA segundo as funções de rede e de transferência dos neurônios. Estas saídas são comparadas com aquelas esperadas no próximo passo;

- Atualização dos pesos: é feita a minimização do erro quadrático, definida entre a saída desejada e a calculada pela RNA. Não é possível calcular esse erro para as camadas ocultas (intermediárias). Portanto, a retropropagação do erro é feita

no sentido das camadas ocultas, atualizando os pesos de cada uma delas. Esta minimização é obtida através do método dos gradientes conjugados;

- Repete-se o processo desde o segundo passo: parar somente se o erro cometido pela rede estiver abaixo de um valor pré-determinado ou o número de iterações ultrapassar um limite.

Convergência por Padrão: consiste em ajustar os pesos da rede (considerando-se todas as camadas), fazendo-se com que haja convergência para cada padrão, até que se complete o conjunto de padrões de entrada. O processo deverá ser repetido até a total convergência, isto é, o erro quadrático seja inferior a uma tolerância pré-estabelecida para os padrões considerados.

Uma Iteração por Padrão: este procedimento é idêntico ao primeiro, porém, fazendo-se somente uma iteração (ajuste de pesos) por padrão.

Uma das vantagens da RNA é que esta pode capturar muitos tipos de informações que permitem o usuário explicar com maior rapidez e facilidade os problemas que, por outro lado poderiam ser difíceis ou impossíveis de serem resolvidos. Dentre outras utilizações das RNAs, pode-se citar a inferência de múltiplas variáveis, a grande tolerância a falhas, o modelamento direto do problema e o paralelismo inerente.

Apesar do sucesso aparente do algoritmo de aprendizado *Back-Propagation*, existem alguns problemas que afetam o seu desempenho. A maior parte dos problemas é que o processo de treinamento é longo, e pode causar um resultado de momento e uma taxa de treinamento não ótimos. Os fracassos do treinamento geralmente surgem de duas fontes: paralisia da rede e mínimo local.

- **Paralisia da Rede:** ocorre quando esta opera com valores elevados para os pesos, fazendo com que as saídas calculadas pelos neurônios sejam muito altas. Isto provocará uma parada do reajuste dos pesos nas camadas próximas à de saída, visto que os neurônios passam a operar numa faixa onde o ganho, representado pela derivada da função de transferência dos mesmos, é próximo de zero, portanto o treinamento fica paralisado, (Wasserman 1989). A função de transferência utilizada foi a do tipo tangente hiperbólica.

- **Mínimos Locais:** um problema que ocorre quando se utiliza o algoritmo *Back-Propagation* é a possibilidade do treinamento convergir para um mínimo local da função erro quadrático, e

cm 1 2 3 4 5 6 7 8 unesp 12 13 14 15 16 17 18

não para o mínimo global, conforme o esperado. Para a ADALINE há sempre um mínimo global, garantido pelo fato das funções de transferência serem lineares. Já no caso dos conjuntos de neurônios que possuem função de transferência não-linear, a função erro quadrático pode apresentar um ou mais mínimos locais, (Wasserman, 1989).

cm 1 2 3 4 5 6 7 8 unesp 12 13 14 15 16 17 18

CAPÍTULO 5

TÉCNICA DE REDUÇÃO DE DADOS E APLICAÇÃO

A maioria das máquinas e equipamentos utilizados na atualidade operam por meio de motores e outras partes rotativas que podem desenvolver falhas, e estas podem causar a parada imediata do sistema de operação. Para prever esse tipo de problema, as máquinas e equipamentos devem ser monitorados, e os sinais provenientes da vibração tratados por técnicas de análise de sinais. Este tipo de análise pode ser efetuada por diferentes técnicas, algumas mais adequadas ou não a determinadas classes de sinais (estacionário e não-estacionário).

No caso de sinais não-estacionários provenientes de máquinas e equipamentos, muitas destas técnicas não são adequadas, visto que não são sensíveis a eventos transitórios, e mesmo as que são sensíveis, por exemplo TW, são incapazes de classificar eventos transitórios de fontes distintas. Uma alternativa é utilizar estas técnicas conjuntamente com métodos menos subjetivos de interpretação, como as redes neurais artificias (RNAs). Um problema crítico na utilização destas técnicas com RNA é a grande quantidade de dados produzidos no processamento. Embora existam muitas RNAs que possam processar os dados temporais diretamente, a utilização de técnicas de redução de dados, ou seja, reduzir os vetores de entradas (padrão de entrada) da rede

cm 1 2 3 4 5 6 7 8 unesp 12 13 14 15 16 17 18

para um conjunto de dados que contenha apenas as principais informações do sinal é um ponto chave neste processo. A inclusão de um algoritmo de redução de dados pode diminuir o tempo de treinamento e também melhorar a capacidade de generalização (desempenho) da rede treinada (Murray e Penman, 1997).

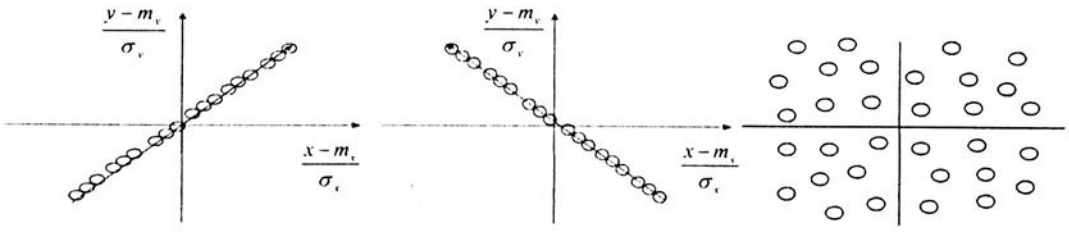
O presente trabalho propõe uma metodologia de classificação e reconhecimento de padrão de falhas, utilizando o processamento de sinais de vibração por Wavelet, associado com redes neurais. A figura 5.1 mostra esquematicamente as várias etapas da metodologia.

Figura 5.1 - Etapas principais do processo de classificação.

Inicialmente, os sinais de vibração são decompostos em várias componentes, aproximações e detalhes. Estas componentes são correlacionadas com padrões de referência e apenas os coeficientes de correlação são utilizados para o treinamento, reduzindo assim a quantidade de dados de entrada da rede. A redução de dados é feita utilizando técnicas de correlação entre as várias componentes dos sinais pré-processados.

A correlação das componentes permite estabelecer padrões de comparação que possibilitam estimar quão próximos os sinais estão uns em relação aos outros. A correlação de dois sinais pode ser representada qualitativamente por um gráfico de dispersão. Os valores dos sinais são os pares de coordenadas (x,y). Se eles mantêm um padrão definido, conseqüentemente, existe uma correlação entre eles. E se não existe uma relação definida entre as coordenadas pode-se dizer que os sinais são não-correlacionados, Fig. 5.2.

ρ_{xy}=0 não haverá correlação, portanto, as retas são paralelas ao eixo x e y, respectivamente, Fig. (5.3c).



(a) (b) (c)

Figura 5.3 - Linha de regressão para diferentes valores do coeficiente de correlação (ρ).

O primeiro passo para o treinamento da rede a ser utilizado na metodologia proposta é separar em classes o conjunto de dados disponíveis.

Os sinais são agrupados em famílias e cada família contém sinais obtidos a partir de condições de operações semelhantes, sendo assim, estas famílias representam a condição operação normal (sem falhas) e as possíveis condições de falha do sistema. Cada família é composta por um conjunto de sinais, e o que difere um sinal do outro em cada família, é o efeito causado por ruído inerente ao processo de medição e pequenas perturbações provenientes do próprio sistema. As famílias de sinais são pré-processadas via Transformada Wavelet Discreta (TWD), ou seja, cada sinal de uma família é decomposto em várias componentes. A decomposição em componentes leva a uma avaliação mais minuciosa das várias informações do sinal, permitindo uma maior extração de informações. A seleção do nível e da família de Wavelet permite decompor os sinais com mais ou menos componentes e também com maior ou menor resolução, como foi discutido na seção referente à TW. Parte dessas componentes são responsáveis pelas baixas freqüências (aproximações) e as outras pelas altas freqüências (detalhes).

Os dados de treinamento da rede são obtidos a partir da redução do sinal pré-processado pela TW. Define-se um sinal de referência a partir da família de sinais normais, e as componentes deste sinal são correlacionadas com as respectivas componentes dos demais sinais normais, gerando os coeficientes de correlação do sistema sem falha. Os coeficientes do sistema

com falha são definidos através da correlação das componentes do sinal de referência com as respectivas componentes dos sinais de cada família de sinais com falha.

A correlação entre as componentes dos pares de sinais decompostos geram vetores com dimensão n × 1, sendo n o número de componentes da TW. Estes vetores são utilizados como entrada da RNA. Os vetores de entrada da rede expressam quanto um sinal assemelha-se com o sinal de referência, permitindo uma avaliação das condições de operação do sistema. Esta correlação poderia ser feita com os próprios sinais de cada família, entretanto, as informações obtidas da correlação entre dois sinais, por exemplo, normal e defeituoso, apresentam um caráter geral e podem não exibir informações específicas sobre o problema. Já a correlação entre as componentes do sinal processado pela TW pode fornecer informações mais específicas, visto que parte das componentes do sinal decomposto representam as baixas freqüências (a-aproximações) e as demais as componentes de alta freqüência (d-detalhes). Portanto, a forma em que é feita a correlação, entre componentes das decomposições, faz com que se possa ter informações gerais e minuciosas ao mesmo tempo. Estas informações são utilizadas para alimentar o treinamento da rede.

A figura 5.4 mostra, esquematicamente, a etapa de redução de dados. Os sinais são decompostos em componentes de altas e baixas freqüências através da TWD e correlacionados utilizando uma função do Matlab (corrcoef), a qual está relacionada com a equação (5.3). Assim, são obtidos os vetores de entrada (padrão de entrada) da rede neural artificial. O procedimento é o mesmo para os vetores que serão utilizados para testar a eficiência da RNA.

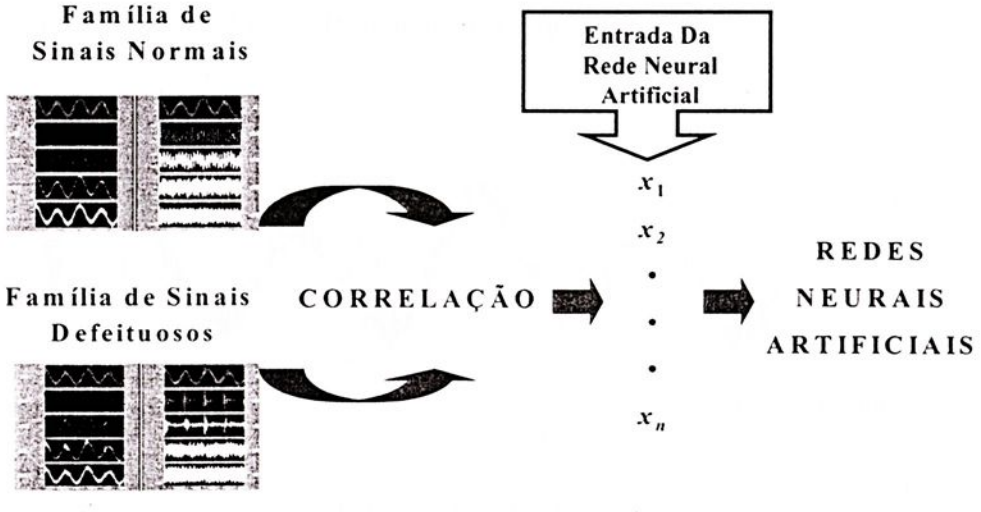


Figura 5.4 - Método de redução de dados.

5.1 SINAIS DE VIBRAÇÃO SIMULADOS

O principal objetivo deste estudo é demonstrar a aplicabilidade da metodologia proposta para o reconhecimento das condições de operação de um sistema utilizando dados reais. Entretanto, com o objetivo de testar a potencialidade e as limitações do método, a mesma será avaliada inicialmente com sinais simulados de um sistema redutor com um par de redução. Os sinais utilizados são classificados por famílias (cada uma delas com 8 sinais), que representam sinais normais e defeituosos. Cada família representa uma condição do redutor e o que difere um sinal do outro, em cada família, é o efeito causado pela introdução de um ruído branco no sinal. Os sinais propostos estão definidos para um tempo de aquisição de 0.128 s e amostrados com 1024 pontos. Os sinais simulados utilizam dados de um redutor de engrenagens helicoidais de um único par de redução, Tab. 5.1.

Tabela 5.1 - Características geométricas do redutor

Número de dentes do pinhão	44
Número de dentes da coroa	95
Freqüência do eixo de entrada	30 (Hz)
Freqüência do eixo de saída	14 (Hz)
Freqüência de engrenamento	1320 (Hz)

Cada sinal da família de sinais normais é constituído da freqüência de engrenamento, freqüência de giro (pinhão) e 5% de ruído branco, Fig. 5.5.

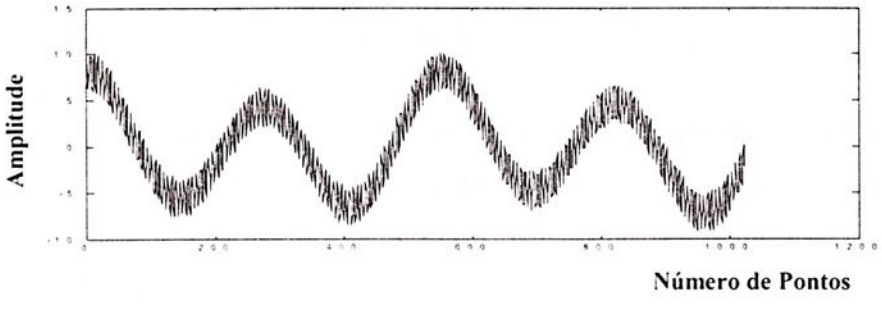


Figura 5.5 - Sinal Normal.

Um sinal da família de sinais defeituosos (1) tem as mesmas características de um sinal normal, freqüência de engrenamento, freqüência de giro (pinhão), 5% de ruído branco, e adicionalmente, um impulso a cada 256 pontos (ou 0.032 s) representando o efeito de um dente quebrado, Fig. 5.6.

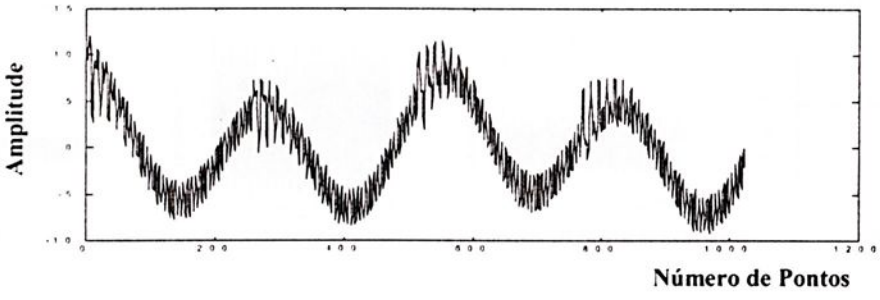


Figura 5.6 - Sinal Defeituoso (1).

Um sinal da família de sinais defeituosos (2) tem as mesmas características de um sinal normal, freqüência de engrenamento, freqüência de giro (pinhão), 5% de ruído branco e, adicionalmente, um impulso a cada 64 pontos (ou 0.008 s) representando um defeito de rolamento, Fig. 5.7,

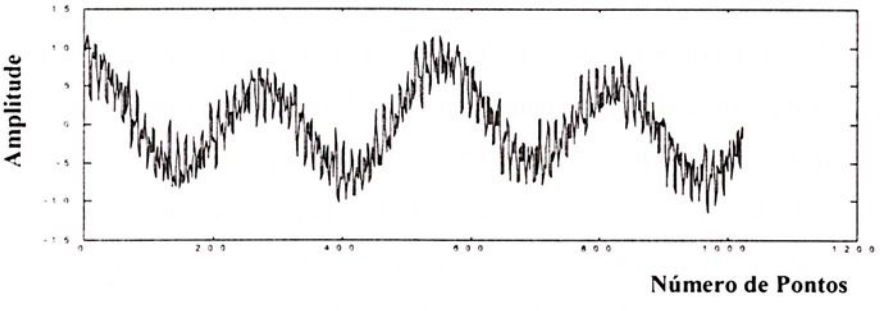


Figura 5.7 - Sinal Defeituoso (2).

As famílias de sinais simulados são pré-processadas via transformada Wavelet discreta (TWD), ou seja, cada sinal de uma família é decomposto em 8 componentes, Fig. 5.8. Parte dessas componentes são responsáveis pelas baixas freqüências (aproximações) e as outras pelas altas freqüências (detalhes).

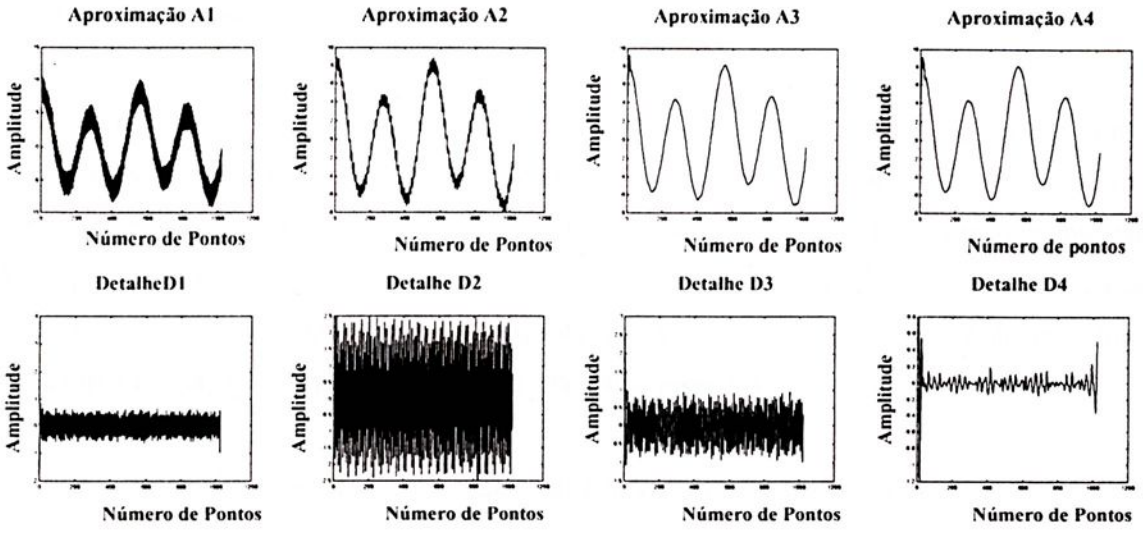


Figura 5.8 - Decomposição de um sinal normal pela TWD.

Mediante a seleção do nível e da família de Wavelet pode-se decompor os sinais com mais ou menos componentes e também com maior ou menor resolução. Um sinal da família de sinais normais decomposto foi escolhido como referência para estabelecer a correlação entre suas componentes e as componentes dos demais, e também em relação às componentes das outras famílias (defeituosos (1) e (2)). Assim, a correlação entre pares de sinais decompostos geram vetores com dimensão 8 × 1, os quais foram utilizados como entrada da RNA. Esses vetores representaram quanto um sinal assemelha-se com o sinal escolhido como referência. Parte desses vetores foram utilizados para o treinamento da RNA (arquivos Aq1, Aq1.3 e Aq2.6) e os demais, como relacionados na tabela 2, foram utilizados para testar a capacidade de reconhecimento da rede. A RNA supervisionada de multi-camadas utilizou o algoritmo de treinamento retropropagação, com função tangente hiperbólica. A camada intermediária foi estabelecida através de tentativas, e assim, pela experiência adquirida, chegou-se a 10 neurônios com o melhor desempenho para o treinamento da RNA. Para simplificar os resultados de classificação ou distinção entre tipos diferentes falhas, as saídas designadas para cada situação são as seguintes:

1 sinal normal: saída 0;
2 sinal defeituoso (1): saída 0,5;
3 sinal defeituoso (2): saída 1.

5.2 RESULTADOS

Na tabela 5.2 estão todas as respostas da RNA para os vários testes de reconhecimento dos vetores não utilizados durante o treinamento. Para o primeiro bloco referente aos sinais normais, pode-se observar que as respostas da RNA foram as esperadas (aproximadamente zero). As respostas da RNA para o segundo e terceiro bloco, também foram muito satisfatórias, ou seja, os arquivos que apresentam as piores resposta ainda mostram uma identificação próxima do esperado, ou seja, apenas uma diferença de 2% e 4%, respectivamente.

Tabela 5.2 - Arquivos utilizados para testar a RNA

Famílias de Sinais	Arquivos	Saída da Rede
NORMAIS	Aq2	0.0007
	Aq3	0.0023
	Aq4	0.0004
	Aq5	0.0007
	Aq6	0.0023
	Aq7	0.0004
	Aq8	0.0005
DEFEITO (1)	Aq1.1	0.5196
	Aq1.2	0.5186
	Aq1.4	0.5077
	Aq1.5	0.5196
	Aq1.6	0.5186
	Aq1.7	0.5077
	Aq1.8	0.5146
DEFEITO (2)	Aq2.1	0.9735
	Aq2.2	0.9783
	Aq2.3	0.9794
	Aq2.4	0.9730
	Aq2.5	0.9684
	Aq2.7	0.9786
	Aq2.8	0.9789

Os resultados apresentados mostram que a TW é uma ferramenta de pré processamento de sinais de vibração muito eficiente, e que estes dados podem ser utilizados como entradas para uma RNA com objetivo de melhorar o desempenho desta na identificação dos tipos de falhas. Embora, não tenha sido feita uma comparação direta com outros tipos de ferramentas, um

unesp

trabalho desenvolvido por Petrilli *et al.* (1995), mostrou que uma RNA que utilizava sinais pré-processados pela TW obteve uma convergência muito mais rápida que os mesmos sinais pré-processados pela FFT. Isto ocorre porque a capacidade da TWD em analisar transientes e guardar informações importantes do sinal permite associar as falhas com as respectivas componentes, assim a RNA pode ter informações mais precisas, e conseqüentemente, um melhor diagnóstico.

5.3 CONSIDERAÇÕES FINAIS

Os resultados apresentados e discutidos neste capítulo mostram que o desempenho da TWD como um pré-processador dos sinais de vibração para detecção de eventos do tipo transiente foi satisfatório. A análise via TWD requer uma familiaridade do usuário com a técnica, ou seja, na interpretação e na escolha do tipo de Wavelet. O tipo de Wavelet utilizada foi a Daubechies, a qual também mostrou-se eficiente no processamento de sinais em outros trabalhos.

A combinação da TWD e RNA é uma técnica eficaz e adequada para monitoramento contínuo de máquinas. O próximo passo será utilizar a técnica para a detecção de falhas utilizando dados experimentais.

cm 1 2 3 4 5 6 7 8 unesp 12 13 14 15 16 17 18

CAPÍTULO 6

6 ANÁLISE EXPERIMENTAL E RESULTADOS

6.1 MONTAGEM EXPERIMENTAL

Para validação da metodologia proposta foi utilizada uma estrutura composta por três mesas (placas de alumínio) e lâminas verticais. As características dinâmicas desta estrutura podem ser encontradas em Melo (1998). Na bancada experimental, foram utilizados os seguintes materiais e equipamentos:

MATERIAIS:

[1] - 3 chapas de alumínio com as seguintes dimensões: (0.350x0.350x0.010)m, (0.255x0.255x0.010)m, (0.158x0.158x0.010)m;

cm 1 2 3 4 5 6 7 8 unesp 12 13 14 15 16 17 18

[2] - parafusos para fixação, placas de aço para suporte lateral e a base;

[3] - 1 haste de ferro;

[4] - 2 suportes de madeira;

[5] - 16 lâminas metálicas de aço inoxidável:

- Placa superior: 6 lâminas (0.135x0.025)m;
- Placa central: 4 lâminas (0.135x0.025)m;
- Placa inferior: 4 lâminas (0.115x0.023)m.

EQUIPAMENTOS:

[1] - transdutor de força - PCB, modelo 208C02, série 15908, sensibilidade: 47,10 mV/lbf e linearidade: 0,25% FS;

[2] - acelerômetro - PCB, modelo 353b33, série 54656, sensibilidade 103,7 mV/g;

[3] - condicionador de sinais - PCB, modelo 480E09;

[4] - amplificador de sinais - MMF - LU 103, série 3997/8;

[5] - sistema de aquisição de dados Daq Book/112, 6 canais, com taxa máxima de aquisição de 40 kHz;

[6] - gerador de funções - PT – 930, All – In – One Instrument, série 000125;

[7] - computador PC Pentium - 133 MHz;

[8] - programa de aquisição - Dasylab 4.

cm 1 2 3 4 5 6 7 8 unesp 12 13 14 15 16 17 18

O sistema mecânico composto de três placas de alumínio e lâminas de aço inoxidável foi montado sobre uma placa de aço, que por sua vez é apoiada sobre dois suportes de madeira. As lâminas são fixas em suas respectivas placas e a placa de aço é fixa na mesa inercial.

Os materiais e equipamentos acima foram utilizados para simular o sistema original, ou seja, sistema sem falha. A figura 6.1 mostra a configuração utilizada nos ensaios experimentais.

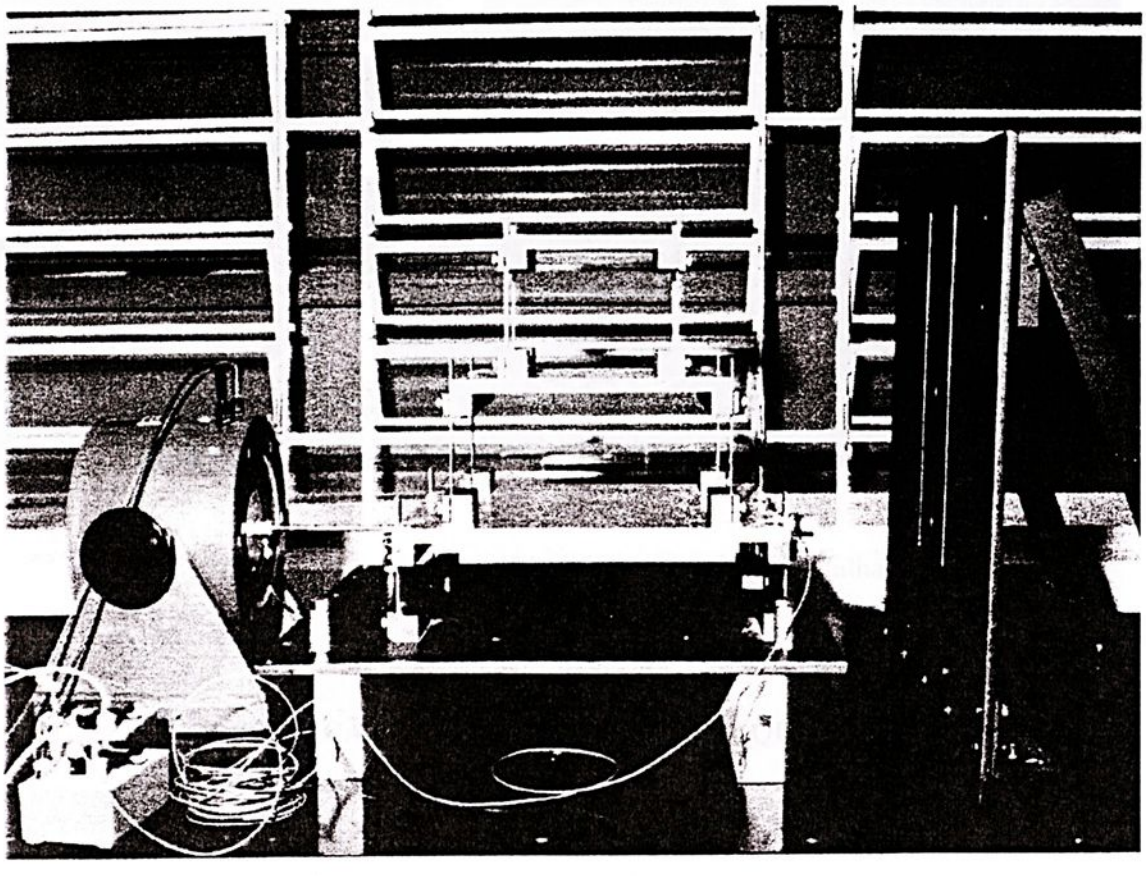


Figura 6.1 - Montagem experimental do sistema sem falha.

O sistema com falha foi simulado fixando-se uma haste em uma placa de aço vertical parafusada na mesa inercial. A falha é considerada a partir do momento em que a placa intermediária encosta na haste, Fig. 6.2.

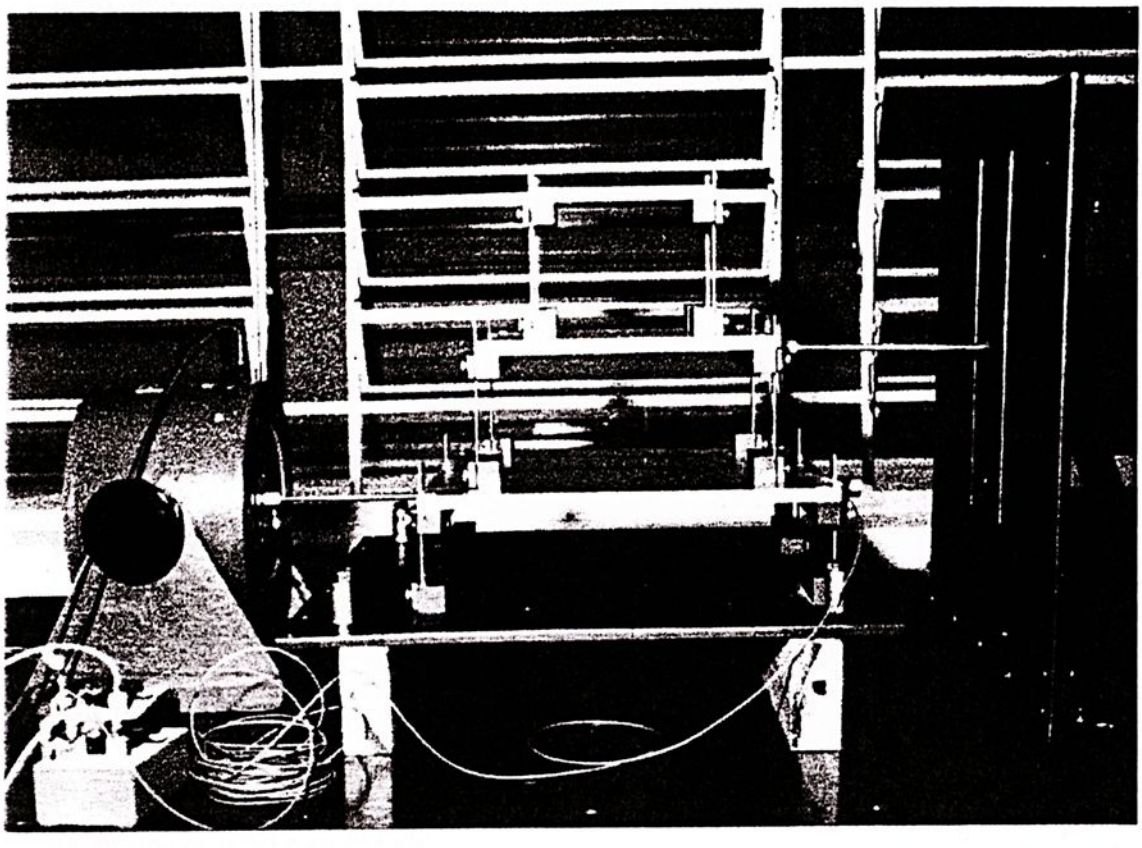


Figura 6.2 - Montagem experimental do sistema com falha.

6.2 PROGRAMA E DESCRIÇÃO DO FLUXOGRAMA DE AQUISIÇÃO DE DADOS

Com o objetivo de tornar a aquisição de forma automática utilizou-se o programa de aquisição de dados Dasylab 4. A figura 6.3 mostra o fluxograma empregado para aquisição dos dados que foram utilizados no pré-processamento através da TWD.

O fluxograma montado para aquisição dos dados é constituído de:

1 - um conversor analógico/digital com taxa de amostragem 500 Hz, tamanho do bloco de aquisição de 4096 pontos pré-determinados;

2 - filtro digital passa-alta (5Hz) "*cut-off*" utilizado para separar as freqüências que não estão na faixa de trabalho do transdutor (acelerômetro);

3 - controlador do relé;

4 - relé;

5 - sistema responsável por armazenar os dados (arquivo);

6 - sistema de coordenadas utilizado para construir os respectivos gráficos;

7 - janela Hanning para evitar o aparecimento de freqüências inexistentes;

8 - FFT "*Fast Fourier Transform*";

9 - sistema de coordenadas utilizado para construir os respectivos gráficos;

10 - sistema de parada;

11 - trigger responsável por dar início à aquisição do sinal, ou seja, as aquisições são feitas a partir da mesma posição e amplitude.

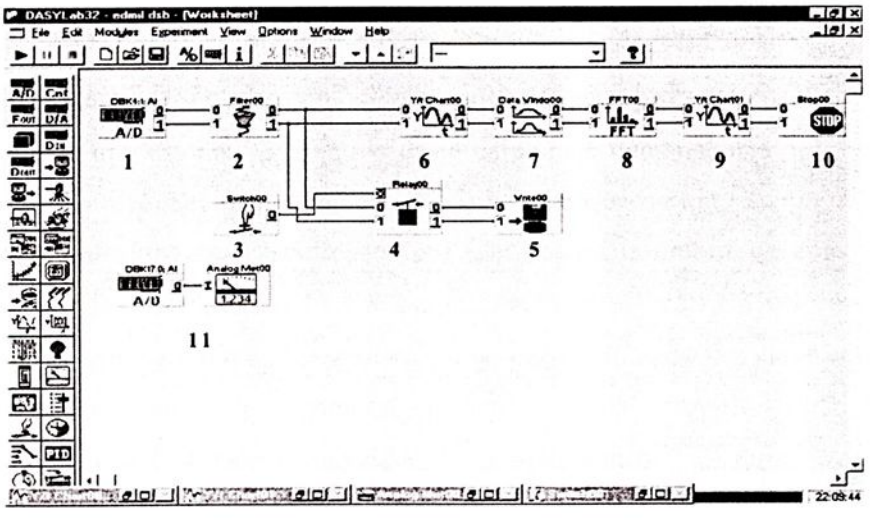


Figura 6.3 - Fluxograma para aquisição dos dados no tempo.

6.3 DESCRIÇÃO DO EXPERIMENTO

O sistema mecânico sem falhas, Fig. 6.1, foi excitado com uma onda senoidal (f_e=40 Hz), utilizando-se um transdutor de força. Esta onda senoidal foi produzida por um gerador de funções PT-930 All – In – One Instrument. A resposta do sistema foi medida através de um acelerômetro fixo na placa de alumínio inferior. O mesmo foi feito para o sistema com falha, Fig. 6.2, que possui uma haste presa a uma placa de aço fixa à mesa inercial. A oscilação da estrutura faz com que a placa de alumínio intermediária toque a haste produzindo impulsos, que representam situações anormais do sistema. Este sistema pode ser considerado bi-linear, pois a partir do instante em que a mesa intermediária toca a haste, a rigidez lateral da estrutura é alterada.

Durante o experimento, os sinais condicionados e filtrados são amostrados pelo conjunto computador-placa de aquisição. No total foram feitas 10 medidas, sendo cinco para sinais normais e cinco para sinais com falha.

6.4 SINAIS DE VIBRAÇÃO EXPERIMENTAIS

No capítulo anterior apresentou-se a metodologia proposta para o processo de monitoramento de um sistema. O objetivo deste capítulo é demonstrar a aplicabilidade desta metodologia para o reconhecimento das condições de operação de um sistema utilizando dados reais. Os sinais utilizados são classificados por famílias (cada uma delas com 5 sinais), que representam sinais sem falha, Fig. 6.4 a 6.8, e com falha, Fig. 6.9 a 6.13. Cada família representa uma condição do sistema e o que difere um sinal do outro, em cada família, é o efeito causado pelas fontes de ruído oriundo do sistema de aquisição. Os sinais propostos estão definidos para um tempo de aquisição de 4,096 s e amostrados com 4096 pontos.

As famílias de sinais sem falha e com falha foram pré-processadas via Transformada Wavelet Discreta (TWD). O número de níveis que um sinal pode ser decomposto é dado com base no seu comprimento N. Quando $N=4096=2^{12}$ pontos o sinal pode ser

unesp

decomposto no máximo em 12 níveis, 12 componentes de aproximações e 12 de detalhes Parte dessas componentes são responsáveis pelas baixas freqüências (aproximações) e as outras pelas altas freqüências (detalhes). A opção em decompor o sinal em mais ou menos componentes depende do que se pretende visualizar, ou seja, deve-se determinar primeiro um número de níveis adequado para decomposição que proporcione informações pertinentes sobre o sinal. Foram feitas várias análises utilizando-se diferentes níveis de decomposição. O sinal estudado foi decomposto em 8 componentes, ou seja, 4 níveis, Fig. 6.14. Quando mais níveis de decomposição foram utilizados, não se observou um ganho significativo nas informações obtidas.

Um sinal da família de sinais normais decomposto, Fig. 6.14, foi escolhido como referência para estabelecer a correlação entre suas componentes e ser utilizado como padrão de entrada para a rede. Da mesma maneira, um sinal da família de sinais com defeito, Fig. 6.15, foi escolhido para ser referência e utilizado no processo de treinamento. O restante dos sinais foram utilizados para verificar a capacidade de generalização da rede. Assim, a correlação entre pares de sinais decompostos gera vetores com dimensão 8×1, os quais foram utilizados como entrada da RNA. Esses vetores representam quanto um sinal assemelha-se com o sinal escolhido como referência. Além da correlação, para o padrão de entradas utilizou-se também um outro tipo de padrão com o objetivo de complementar e aumentar as informações fornecidas para o processo de treinamento da rede. Para isso, utilizou-se a soma da diferença das amplitudes das componentes ao quadrado. As decomposições dos sinais de referência foram utilizadas como base. Parte dos vetores foram utilizados para o treinamento da RNA e os demais, como relacionado na tabela 6.1, foram utilizados para testar a capacidade de reconhecimento da RNA. A RNA supervisionada de multi-camadas utilizou o algoritmo de retropropagação, com função tangente hiperbólica. A camada intermediária foi estabelecida através de tentativas e chegou-se a 30 neurônios como tendo o melhor desempenho para o treinamento da RNA. Para identificar os resultados de classificação ou distinção entre as diferentes situações, as saídas designadas para cada delas foram:

1 sinal normal: saída 0;
2 sinal com falha: saída 1.

SINAIS SEM FALHA

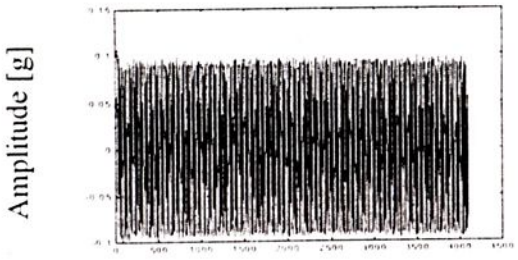


Figura 6.4

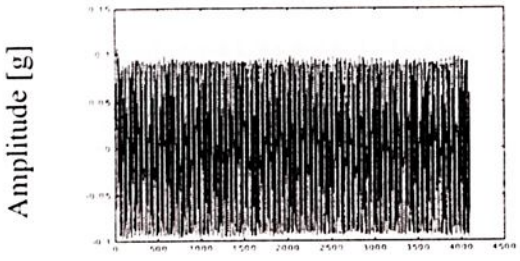


Figura 6.5

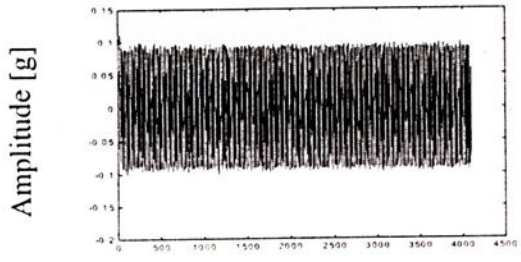


Figura 6.6

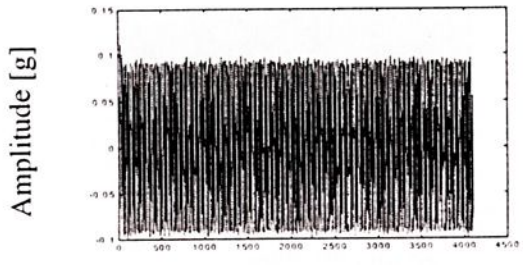


Figura 6.7

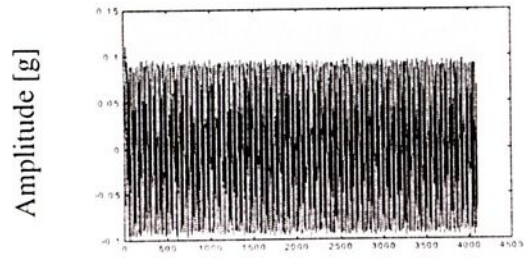


Figura 6.8

SINAIS COM FALHA

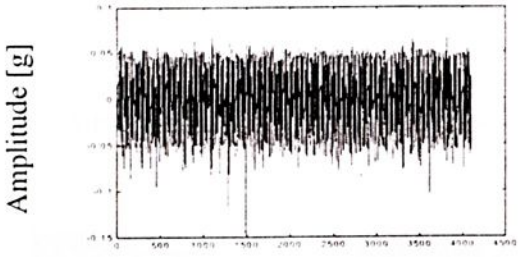


Figura 6.9

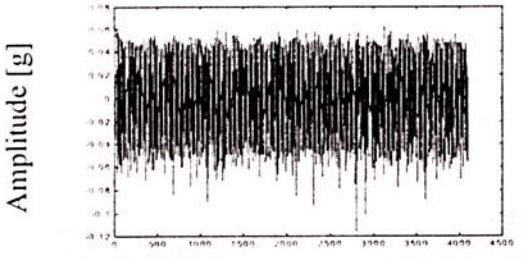


Figura 6.10

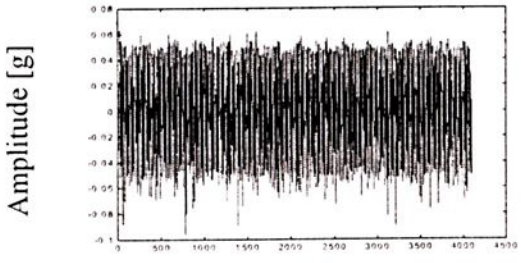


Figura 6.11

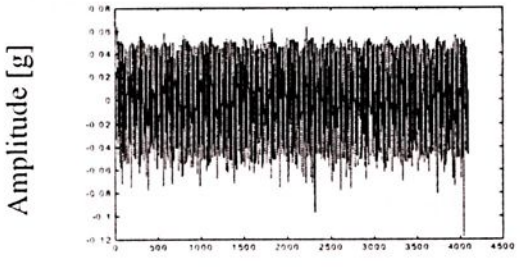


Figura 6.12

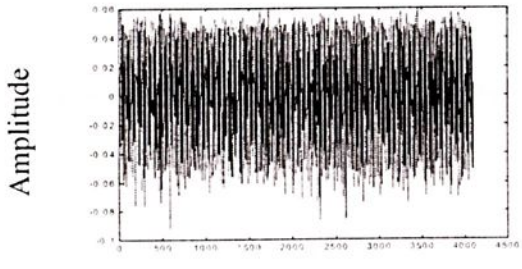


Figura 6.13

SINAL SEM FALHA DECOMPOSTO

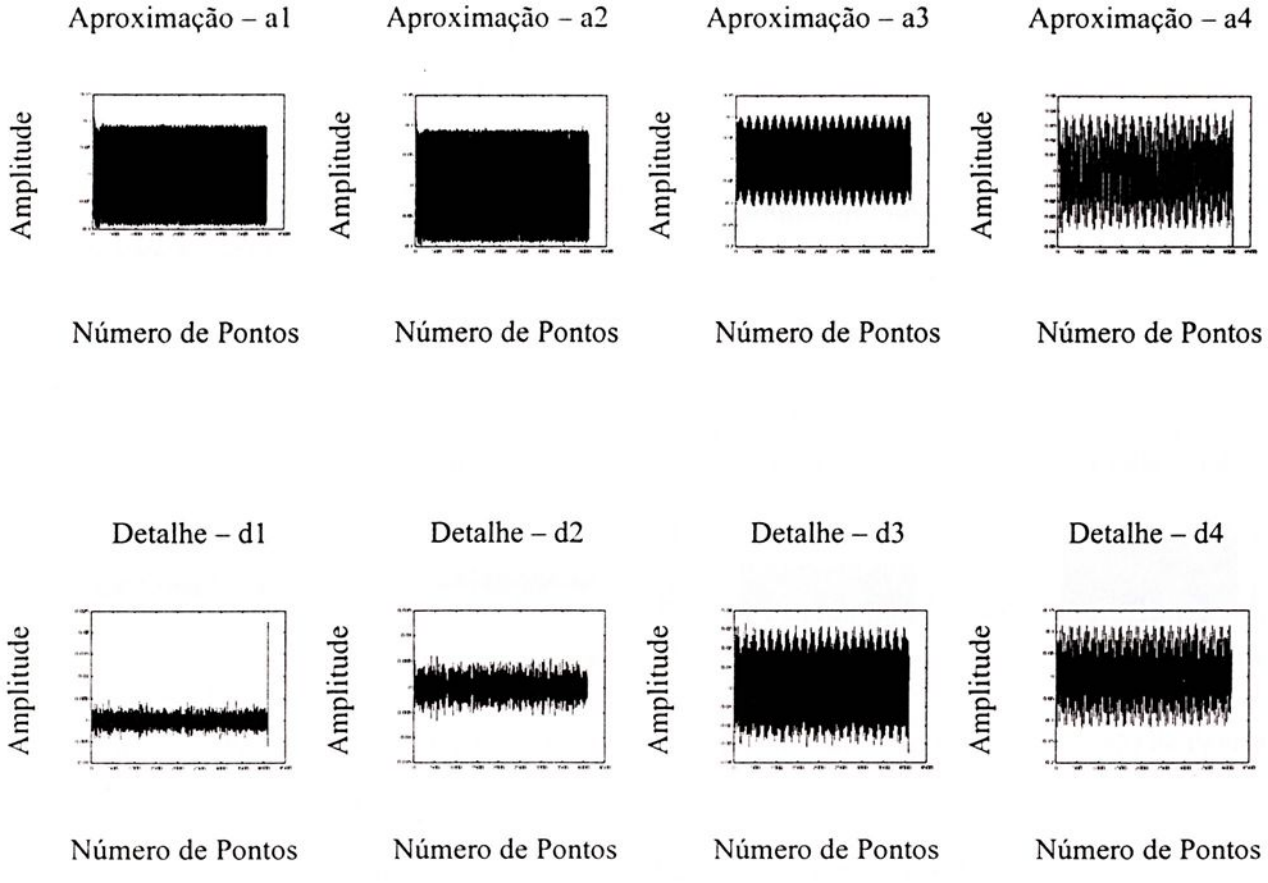


Figura 6.14 - Sinal de referência (sem falha) decomposto via TWD.

SINAL COM FALHA DECOMPOSTO

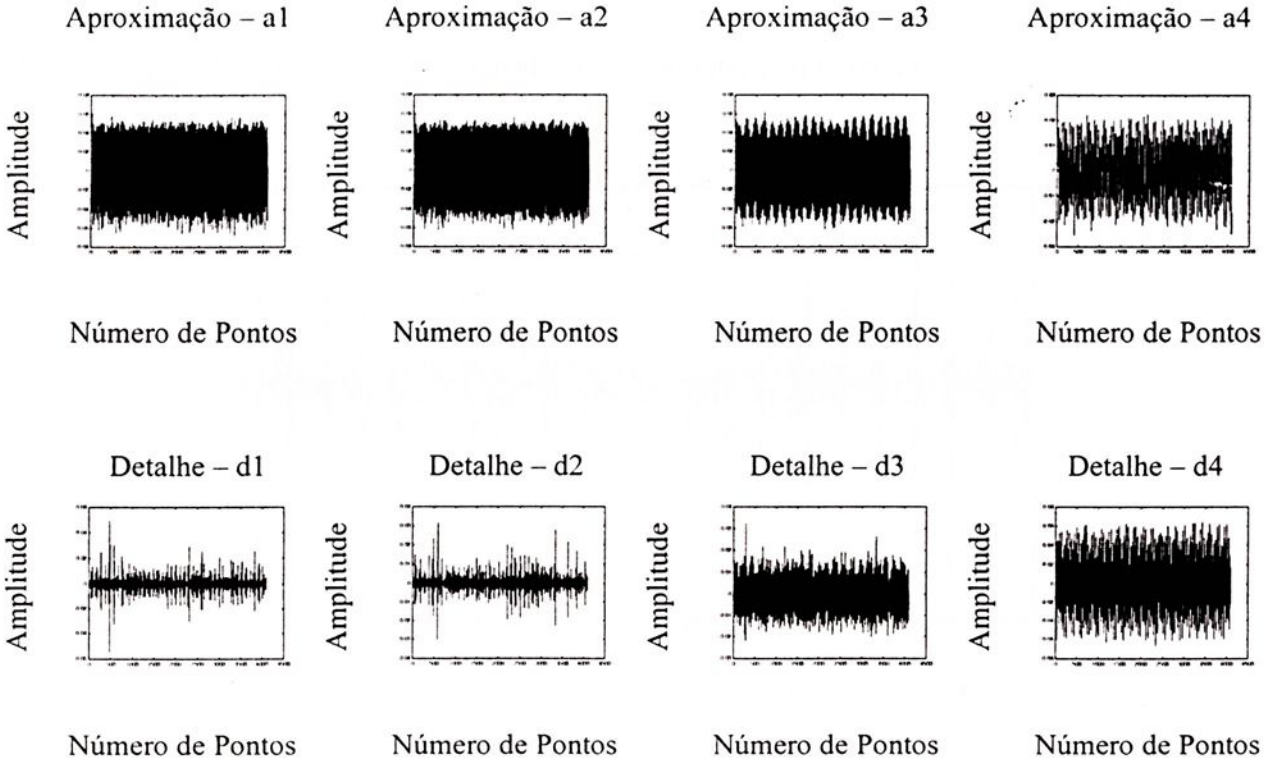


Figura 6.15 - Sinal com falha decomposto via TWD.

6.5 DISCUSSÃO DOS RESULTADOS

O pré-processamento dos sinais temporais é a etapa mais importante quando se utiliza redes neurais para classificação de padrões. A RNA necessita de padrões que sejam bem definidos para que o seu treinamento (ajuste das interligações dos neurônios que compõem a RNA) tenha um bom desempenho. Então, o objetivo de se utilizar a TWD como pré-processador é extrair características dos sinais de forma que se tenha informações específicas deles, e

Nesta tabela os arquivos decorr são vetores de dimensão 8x1 e representam os dados obtidos através da correlação do sinal e do sinal de referência. Os arquivos detot são vetores de dimensão 8x1 e representam as somas do quadrado das diferenças das amplitudes entre os sinais e o sinal de referência.

Tabela 6.1 - Arquivos utilizados para treinar e testar a RNA

Famílias de Sinais	Arquivos	Saída da Rede
NORMAIS	**decorr(:,1) detot(:,1)**	0.0003
	decorr(:,2) detot(:,2)	0.0026
	decorr(:,3) detot(:,3)	0.0025
	decorr(:,4) detot(:,4)	0.0023
	decorr(:,5) detot(:,5)	0.0029
FALHAS	**decorr(:,6) detot(:,6)**	0.9999
	decorr(:,7) detot(:,7)	1.0038
	decorr(:,8) detot(:,8)	1.0017
	decorr(:,9) detot(:,9)	1.0055
	decorr(:,10) detot(:,10)	0.9983

Como pode ser visto na tabela 6.1, a rede neural foi capaz de identificar todos os casos testados com grande precisão.

CAPÍTULO 7

7 CONSIDERAÇÕES FINAIS

Os resultados apresentados e discutidos mostram a importância do pré-processamento na extração de características específicas dos sinais de vibração que representam os estados do sistema mecânico. As informações extraídas e, conseqüentemente, a diferença de uma situação (sinal normal e sinal com defeito) para outra determina o nível de sucesso da aplicação de redes neurais para classificação e identificação de falhas estruturais. A análise via TWD requer uma familiaridade do usuário com a técnica, ou seja, na interpretação e escolha do tipo de Wavelet. Portanto, com base na literatura existente, estabeleceu-se qual o tipo de Wavelet é mais adequada para o processamento. Após esse processamento, a quantidade de dados gerada é extremamente grande, então, a compactação dessas informações torna-se necessária, ressaltando que as características dos sinais originais devem ser mantidas. Essa compactação ou redução dos dados foi feita utilizando-se a técnica de correlação entre as várias componentes da decomposição dos sinais e também a soma do quadrado das diferenças das amplitudes das componentes obtidas das decomposições dos sinais. Estes valores podem ser relacionados à diferença de energia entre as componentes da decomposição de cada estado com as decomposições do sinal de referência. Essa etapa de redução de dados é muito importante para minimizar o tempo de treinamento e

unesp

melhorar a capacidade de generalização da RNA. A combinação da TWD e RNA mostrou ser uma técnica eficaz e adequada para o monitoramento contínuo de sistemas.

7.1 SUGESTÕES PARA TRABALHOS FUTUROS

Para trabalhos futuros, propõe-se um estudo mais detalhado dos vários tipos de Wavelets, desde que diferentes tipos de sinais possam ser melhor representados por outras famílias de Wavelets.

unesp

REFERÊNCIAS

Arato, A. J. , Silva D. G., 2000, "Análise por Demodulação Aplicada ao Monitoramento de Falhas em Engrenagens", CONEM 2000.

Chui, Ch. K., 1992, " An Introduction to Wavelet", Academic Press Inc., Boston. Vol. 2, pp. 158-160.

Coifman, R.R. e Rochberg, R., 1980 "Representation theorems for holomorphic and harmonic functions in L^p", Asterisque, Vol. 77, pp. 11-66.

Daubechies, I., 1992, "Ten Lectures on Wavelets", CBMS-NSF Series in Appl. Math. SIAM publi., 357p.

Gabor, D., "Theory of Comunications", Journal of the Institute of Electrical Engineering, III, V 93, London, pp. 429-457,

Gomes, J.M. , Velho, L. e Goldenstein, S., 1997, "Wavelets: Teoria, Software e Aplicações", 216p.

Gram-Hansen, K., Gram & Juhl, 1997, "The Analysis of Nonstationary Signals", Sound and Vibration, Janeiro, pp. 40-46.

Grossmann, A., Holschneider, M., Martinet, R. K. e Morlet, J., 1987, "Detection of Abrupt Changes in Sound Signal with the Help of Wavelet Transform", Avances in Electronics and Electron Phusics, supp. 19.

Haar, A., 1910, " Zur Theorie der Orthogonalen Funktionensysteme", Math. Am., Vol.69, pp.331-371.

unesp

Hazarika, N., Chen, J.Z., Tsoi, A.C. e Sergejew, A., 1997, "Clasification of EEG Signals Using the Wavelet Transform", Vol. 59, pp. 61-72.

Lopes V., Pereira, A.J, Weber I.H., 1997, "Using Model Updating Technique to Train Neural Network for Fault Detection", Proceedings of DETEC'97, ASME Design Technical Conferences, September 14-17, Sacramento, California.

Mallat, S., 1989, "A Theory for Multeresolution Signal Decomposition: the Wavelet Representation", IEEE Transactions on Pattern Analysis and Machine Intell, vol. 11, pp-674-693.

Melo, G. P., "Detecção e localização de Falhas via Observadores de Estado de Ordem Reduzida", Campinas, Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas, 1998. 125p., Tese (Doutorado)

Meyer, Y., 1992, "Wavelets and Operators", pp. 223.

Mori, K., Kasashima, N., Yoshioka, T. e Ueno, Y., 1996, "Prediction of Spalling on a Ball Bearing by Applying the Discrete Wavelet transform to Vibration of Signals", Elsevier, Vol. 195, pp. 162-168.

McCulloch, W. W., Pitts, W., 1943, A Logical Calculus of the Ideas Imminent in Nervous Activity Bulletin of Mathematical Biophysiscs 5:115-33.

Moraes, F. J. V., Teoria de Wavelet Aplicada à Análise de Vibrações. Campinas, Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas, 1996. 87p, Tese (mestrado).

Murray, A., Penman, J., 1997, "Wavelets as Alternative to the FFT in Condition Monitoring Shemes Using ANN'S", University of Aberdeen, UK.

Najar-Munoz, A., Hashemi, J., 1999, "Continuous Wavelet Transform Analysis of Acoustic Emission Signal", IMAC.

cm 1 2 3 4 5 6 7 unesp 12 13 14 15 16 17 18

Newland, D.E., 1993, "An Introduction to Random Vibrations, Spectral and Wavelet Analysis", Logman Scientific & Technical, Third edition, 472p.

Newland, D.E., 1994, "Wavelet Analysis of Vibration, Part 1: Theory", Journal of vibration and Acoustics, Vol. 116, pp. 409-416.

Newland, D.E., 1994, "Wavelet Analysis of Vibration, Part 2: Wavelet Maps", Journal of Vibration and Acoustics, Vol. 116, pp. 416-425.

Newland, D.E., 1997, "Practical Signal Analysis: Do Wavelets Make any Diference?", ASME Proceedings of DETC 97, pp. 1-12.

Paya, B.A., Esat, I.I. e Badi, M.N.M, 1997, "Artificial Neural Network Based Fault Diagnostics of Rotating Machinery Using Wavelet Transforms as a Preprocessor, Mechanical Systems and Signal Processing, Vol. 11, pp. 751-765.

Petrilli, O., Paya, B.A., Esat, I.I. e Badi, M.N.M., 1995, " Neural Network based Fault Detection Using Different Signal Processing Techniques as Pre-Processor", Structural Dynamics and Vibration Vol. 70, pp. 97-101.

Rosenblatt, R.,1959, Principles of Neurodynamics, New York, Spartan Books.

Staszewski, W. J., Worden, K., 1993, "Classification of Faults in Spur Gears", The World Congress on Neural Networks, Portland Oregon, pp.11–15.

Tai, H-M., Sung, C-K, 1997, " Motion Detection for Precision Positioning Systen by Wavelet Analysis", Proceedings of DETC'97, ASME Design Engineering Technical Conference.

Wasserman, Philip D., 1989, " Neural Computing Theory And Practice".

Widrow, B., Hoff, M. E., 1960, "Adaptive Switching Circuits", IRE WESCON Convention Record, part 4, pp. 96-104 New York, Institute of Radio Engineers.

cm 1 2 3 4 5 6 7 unesp 12 13 14 15 16 17 18

unesp

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de Ilha Solteira
Programa de Pós-Graduação em Engenharia Mecânica
Av. Brasil Centro, 56
15385-000 Ilha Solteira - SP
www.dem.feis.unesp.br