

TRABALHO DE GRADUAÇÃO

**MINIMIZAÇÃO DE RUÍDO GAUSSIANO  
EM SINAIS DE VOZ UTILIZANDO  
TRANSFORMADA DE WAVELETS**

**Rodrigo Batista Balthazar**

**Brasília, dezembro de 2006**

**UNIVERSIDADE DE BRASÍLIA**

**FACULDADE DE TECNOLOGIA**

UNIVERSIDADE DE BRASÍLIA  
Faculdade de Tecnologia

TRABALHO DE GRADUAÇÃO

**MINIMIZAÇÃO DE RUÍDO GAUSSIANO  
EM SINAIS DE VOZ UTILIZANDO  
TRANSFORMADA DE WAVELETS**

**Rodrigo Batista Balthazar**

Relatório submetido ao Departamento de Engenharia Elétrica  
da Faculdade de Tecnologia da Universidade de Brasília como  
requisito parcial para obtenção do grau de Engenheiro Eletricista.

Banca Examinadora

Prof. Francisco Assis de Oliveira Nascimento, \_\_\_\_\_  
UnB/ENE (Orientador)

Prof. Pedro de Azevedo Berger, UnB/CIC (Co-  
orientador)

## **Dedicatória**

*A meus pais, que nunca deixaram de acreditar em mim, mesmo quando eu não conseguia mais.  
A Bárbara, aquela que tem me inspirado, alegrado, suportado e acompanhado em todos esses  
últimos momentos da minha vida, fáceis ou difíceis, alegres ou tristes.*

*Rodrigo Batista Balthazar*

## Agradecimentos

*A todos os mestres que me infundiram da experiência e segurança necessárias para me deparar com as encruzilhadas da vida, em especial àqueles que me permitiram desfrutar de seu convívio e da sabedoria de suas lições nestes últimos e determinantes momentos; a Cláudia Sigilião, Pedro Berger e Francisco Assis e a todos aqueles que a ingratidão da minha memória não permite recordar. Ao grande Mestre que iluminou a mente de todos estes mestres e me permitiu trilhar meus caminhos com segurança e honestidade, ao Senhor dos Exércitos e das Virtudes. A minha família, colegas, amigos, todos aqueles que, seja pela tolerância com meus comportamentos egoístas, pela compreensão das minhas necessidades ou pela inata e despretenciosa benevolência que me dedicaram, permitiram que esse trabalho se consolidasse.*

*Rodrigo Batista Balthazar*

---

## RESUMO

Este trabalho utiliza a Transformada de *Wavelets* para a minimização de ruído em sinais de voz. A técnica empregada implementa o refinamento da invariância ao deslocamento para aumentar a eficiência no *de-noising*. Com o desenvolvimento atual da Transformada de *Wavelets* Discreta (DWT) já é possível a análise com algoritmos rápidos, viabilizando um processamento massivo. Para a eficiente utilização da Transformada de *Wavelets*, é necessária uma escolha adequada das funções-base a serem empregadas na aplicação pretendida, juntamente aos parâmetros de decomposição associados à mesma. Foram realizadas, portanto, diferentes análises de sinais de voz falada gravados por diferentes locutores na busca de uma generalização de parâmetros que permitam um melhor aproveitamento da transformada na implementação de etapas de *de-noising* para sistemas sujeitos a ruído.

---

## ABSTRACT

This work uses the Wavelet Transform applied to the de-noising of voice signals. The chosen technique implements the refinement of the translation invariance in order to increase the de-noising efficiency. With the recent developments of the DWT, it has become possible to deal with massive processing, due to the fast algorithms available. The efficient transform application needs to be associated with an adequate choice of the base-functions that will model the intended signals together with the best decomposition parameters associated with it. Hence, different analysis were done for voice signals recorded by different speakers seeking a parameter generalization to allow better exploration of the transform potential in the implementation of de-noising stages to potential subject-to-noise systems.

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>1</b>
1.1	CONTEXTUALIZAÇÃO .....	1
1.2	DEFINIÇÃO DO PROBLEMA .....	1
1.3	OBJETIVOS DO PROJETO.....	1
1.4	APRESENTAÇÃO DO TRABALHO .....	2
<b>2</b>	<b>O SOM.....</b>	<b>3</b>
2.1	INTRODUÇÃO .....	3
2.2	PERCEPÇÃO DO SOM.....	5
2.3	SISTEMA AUDITIVO .....	6
2.4	A VOZ .....	8
2.5	O RUÍDO .....	8
<b>3</b>	<b>FOURIER.....</b>	<b>10</b>
3.1	TRANSFORMADA DE FOURIER.....	10
3.2	TRANSFORMADA DE FOURIER DE CURTO TERMO.....	12
3.3	PRINCÍPIO DA INCERTEZA DE HEISENBERG .....	13
3.4	TRANSFORMADA DISCRETA DE FOURIER .....	15
3.4.1	JANELAMENTO .....	15
<b>4</b>	<b>WAVELETS.....</b>	<b>16</b>
4.1	ANÁLISE EM MULTI-RESOLUÇÃO.....	16
4.2	TRANSFORMADA DE WAVELETS.....	17
4.2.1	HISTÓRICO .....	17
4.2.2	TRANSFORMADA DE WAVELETS CONTÍNUA.....	18
4.2.3	CÁLCULO DA CWT .....	19
4.2.4	PROPRIEDADES DAS WAVELETS .....	21
4.2.5	TRANSFORMADA DE WAVELETS DISCRETA .....	22
4.3	FUNÇÕES-BASE .....	27
<b>5</b>	<b>DE-NOISING .....</b>	<b>29</b>
5.1	INTRODUÇÃO .....	29
5.2	THRESHOLD .....	30
5.3	SELEÇÃO DO NÍVEL DO THRESHOLD .....	31
5.4	FIGURAS DE MÉRITO .....	32
5.4.1	RELAÇÃO SINAL-RUÍDO (SNR) .....	32
5.4.2	MSE .....	33
5.4.3	PESQ .....	33
<b>6</b>	<b>METODOLOGIA.....</b>	<b>35</b>
6.1	INTRODUÇÃO .....	35
6.2	DADOS .....	35
6.3	FERRAMENTAS COMPUTACIONAIS .....	36
6.4	TÉCNICAS UTILIZADAS .....	36
6.4.1	ADIÇÃO DE RUÍDO BRANCO .....	36
6.4.2	DE-NOISING WAVELET INVARIANTE AO DESLOCAMENTO .....	37
6.4.3	RECONSTRUÇÃO WAVELET .....	39

6.4.4	MANIPULAÇÕES DA TÉCNICA TIWD .....	40
<b>7</b>	<b>RESULTADOS EXPERIMENTAIS .....</b>	<b>42</b>
7.1	INTRODUÇÃO .....	42
7.2	ANÁLISE DO BANCO DE 22KHz .....	43
7.2.1	ANÁLISE DAS AMOSTRAS .....	44
7.2.2	ANÁLISE DOS SINAIS INTEIROS.....	48
7.2.3	ANÁLISE PARA DIFERENTES NÍVEIS DE RUÍDO .....	49
7.2.4	UTILIZAÇÃO DOS PARÂMETROS OBTIDOS .....	50
7.3	ANÁLISE DO BANCO DE 8KHz.....	50
7.3.1	ANÁLISE DAS AMOSTRAS .....	51
7.3.2	ANÁLISE DOS SINAIS INTEIROS.....	52
<b>8</b>	<b>CONCLUSÕES.....</b>	<b>60</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>62</b>
	<b>ANEXOS .....</b>	<b>64</b>
<b>I</b>	<b>DESCRIÇÃO DO CONTEÚDO DO CD.....</b>	<b>65</b>
<b>II</b>	<b>TEXTOS.....</b>	<b>67</b>
<b>III</b>	<b>TABELAS.....</b>	<b>69</b>

# LISTA DE FIGURAS

2.1	Características da Onda Sonora .....	4
2.2	A escrita musical .....	6
2.3	O Ouvido Humano .....	7
3.1	FT dos sinais $x_1(t)$ e $x_2(t)$ .....	11
3.2	Quadro de resolução da STFT .....	13
3.3	Espectrograma da função $x_3(t)$ .....	14
4.1	Escalonamento da <i>Wavelet</i> -mãe .....	19
4.2	Cálculo da CWT .....	20
4.3	CWT da função $x_3(t)$ .....	20
4.4	Resolução tempo x escala da WT .....	21
4.5	Configuração de Banco de Filtros .....	25
4.6	A função escalonante .....	25
4.7	Decomposição <i>Wavelets</i> implementada com o banco de filtros iterativos .....	26
4.8	Estágio do filtro digital utilizado na decomposição .....	27
4.9	<i>Wavelets</i> -mãe utilizadas neste trabalho .....	28
5.1	Etapas envolvidas no processo de <i>de-noising</i> .....	30
5.2	Sinal $s = x$ submetido aos esquemas <i>hard</i> e <i>soft threshold</i> .....	31
5.3	Diagrama algoritmo do PESQ .....	34
6.1	Função TIWD .....	37
6.2	Processo de Escolha de Parâmetros Ótimos para o <i>De-Noising</i> .....	40
7.1	Curvas de MSE para a escolha de <i>wavelets</i> -mãe .....	45
7.2	<i>De-noising</i> do sinal <i>lm101ls11.wav</i> .....	49



# LISTA DE TABELAS

5.1	Escala de qualidade de escuta ACR usada no desenvolvimento do PESQ .....	34
7.1	Listagem dos arquivos utilizados na análise das vozes em 22kHz .....	44
7.2	Escolha da melhor <i>Wavelet</i> -mãe dentre as amostras analisadas .....	46
7.3	Resumo das análises de quantidade de deslocamentos e tamanho da janela de corte .....	47
7.4	Tempos de execução em segundos para o teste de deslocamentos e janelas .....	47
7.5	Resultados do <i>de-noising</i> dos sinais inteiros com a extrapolação dos parâmetros .....	54
7.6	Quantificação dos erros da extrapolação para o grupo feminino .....	55
7.7	Quantificação dos erros da extrapolação para o grupo masculino .....	55
7.8	Análises dos grupos de sinais .....	56
7.9	Análise de amostras com vários níveis de ruído .....	56
7.10	Análise de amostras com vários níveis de ruído .....	56
7.11	Melhores funções-base para cada nível de ruído .....	57
7.12	Arquivos utilizados do banco de 8kHz .....	57
7.13	Melhores resultados para todos os locutores .....	57
7.14	Melhores resultados para locutoras femininas .....	58
7.15	Melhores resultados para os locutores masculinos .....	58
7.16	Análise dos sinais inteiros do banco de 8kHz .....	59
III.1	Escolha wavelet-mãe para amostra do arquivo lf501ls11.wav .....	69
III.2	Escolha wavelet-mãe para amostra do arquivo lf502ls11.wav .....	69
III.3	Escolha wavelet-mãe para amostra do arquivo lf503ls11.wav .....	70
III.4	Escolha wavelet-mãe para amostra do arquivo lf504ls11.wav .....	70
III.5	Escolha wavelet-mãe para amostra do arquivo lf505ls11.wav .....	71
III.6	Escolha wavelet-mãe para amostra do arquivo lm101ls11.wav .....	71
III.7	Escolha wavelet-mãe para amostra do arquivo lm102ls11.wav .....	72
III.8	Escolha wavelet-mãe para amostra do arquivo lm103ls11.wav .....	72
III.9	Escolha wavelet-mãe para amostra do arquivo lm104ls11.wav .....	73
III.10	Escolha wavelet-mãe para amostra do arquivo lm105ls11.wav .....	73
III.11	Deslocamentos para amostra do arquivo lf501ls11.wav .....	74
III.12	Deslocamentos para amostra do arquivo lf502ls11.wav .....	75
III.13	Deslocamentos para amostra do arquivo lf503ls11.wav .....	76
III.14	Deslocamentos para amostra do arquivo lf504ls11.wav .....	77
III.15	Deslocamentos para amostra do arquivo lf505ls11.wav .....	78
III.16	Deslocamentos para amostra do arquivo lm101ls11.wav .....	79
III.17	Deslocamentos para amostra do arquivo lm102ls11.wav .....	80
III.18	Deslocamentos para amostra do arquivo lm103ls11.wav .....	81
III.19	Deslocamentos para amostra do arquivo lm104ls11.wav .....	82
III.20	Deslocamentos para amostra do arquivo lm105ls11.wav .....	83
III.21	Escolha da wavelet-mãe para arquivo fernandaAfogamento.wav .....	84
III.22	Escolha wavelet-mãe para o arquivo flaviaDivas.wav .....	84
III.23	Escolha da wavelet-mãe para arquivo izumiPoesia.wav .....	85
III.24	Escolha da wavelet-mãe para arquivo fredMagoa.wav .....	85
III.25	Escolha da wavelet-mãe para arquivo ortis-grua.wav .....	86
III.26	Escolha da wavelet-mãe para arquivo tiagoPalocci.wav .....	86
III.27	Deslocamentos para amostra do arquivo fernandaAfogamento.wav .....	87
III.28	Deslocamentos para amostra do arquivo fredMagoa.wav .....	88
III.29	Análise de vários sinais de um locutor masculino .....	89

# LISTA DE SIMBOLOS

## Símbolos Latinos

$f$	Frequência	[Hz]
$T$	Período	[s]
$v$	Velocidade	[m/s]
$s$	Escala	[1/Hz]

## Símbolos Gregos

$\lambda$	Comprimento de Onda	[m]
$\phi$	Fase	[rad]
$\omega$	Frequência Angular	[rad/s]

## Siglas

<i>ACR</i>	Avaliação de Categoria Absoluta ( <i>Absolute Category Rating</i> )
<i>CWT</i>	Transformada de Wavelets Contínua ( <i>Continuous Wavelet Transform</i> )
<i>DCT</i>	Transformada de Cosseno Discreta ( <i>Discrete Co-sine Transform</i> )
<i>DFT</i>	Transformada Discreta de Fourier ( <i>Discrete Fourier Transform</i> )
<i>DWT</i>	Transformada de Wavelets Contínua ( <i>Continuous Wavelet Transform</i> )
<i>FFT</i>	Transformada Rápida de Fourier ( <i>Fast Fourier Transform</i> )
<i>FT</i>	Transformada de Fourier ( <i>Fourier Transform</i> )
<i>MOS</i>	Pontuação Média de Opinião ( <i>Mean Opinion Score</i> )
<i>MSE</i>	Erro Médio Quadrático ( <i>Mean Square Error</i> )
<i>PESQ</i>	Avaliação Perceptual da Qualidade de Voz ( <i>Perceptual Evaluation of Speech Quality</i> )
<i>SNR</i>	Relação Sinal-Ruído ( <i>Signal-to-Noise Ratio</i> )
<i>SPL</i>	Nível de Pressão Sonora ( <i>Sound Pressure Level</i> )
<i>SSNR</i>	Relação Sinal-Ruído Segmentar ( <i>Segmental Signal-to-Noise Ratio</i> )
<i>STFT</i>	Transformada de Fourier de Curto Termo ( <i>Short Term Fourier Transform</i> )
<i>TIWD</i>	<i>De-Noising Wavelet</i> Invariável ao Deslocamento ( <i>Translation-Invariant Wavelet De-Noising</i> )
<i>WS</i>	Série de Wavelets ( <i>Wavelet Series</i> )
<i>WT</i>	Transformada de Wavelets ( <i>Wavelet Transform</i> )

# 1 INTRODUÇÃO

*Este capítulo apresenta a principal motivação do projeto, a definição do problema e a apresentação dos objetivos. Por fim, o trabalho é apresentado.*

## 1.1 CONTEXTUALIZAÇÃO

O ruído muitas vezes se apresenta como um enorme empecilho na implementação eficiente de sistemas reais, se mostrando como um elemento aleatório de corrupção da informação que se pretende transmitir e/ou processar [1]. Diante deste problema, muito estudo se investiu no sentido de obter uma ferramenta que permita eliminar ou, ao menos, reduzir significativamente os efeitos indesejáveis produzidos pelas limitações físicas dos sistemas reais. Uma das ferramentas que vêm se consolidando na teoria de processamento de sinais para este tipo de aplicação é a Transformada de *Wavelets*, que incorpora evoluções nos modelos de representação de sinais como a análise em multi-resolução e a codificação de sub-banda.

## 1.2 DEFINIÇÃO DO PROBLEMA

Este projeto surgiu da necessidade de maior imunização a ruído de um sistema de reconhecimento de locutor, implementado como projeto de doutorado pelo Perito Criminal Federal Frederico Quadros D’Almeida. Para o sistema em questão, a inserção de ruído nos sinais de voz utilizados no treinamento da rede neural reduz sensivelmente a eficiência do sistema. Portanto, foi sugerido que se procedesse com um pré-processamento dos sinais ruidosos na tentativa de eliminar o ruído numa etapa de *de-noising* anterior à inserção dos mesmos no sistema. A escolha da transformada de *Wavelets* é motivada pela consolidada eficiência de técnicas de *de-noising* utilizando-a, permitindo um acesso a extensa e recente pesquisa.

## 1.3 OBJETIVOS DO PROJETO

Este projeto objetiva o estudo da aplicação da técnica de *de-noising* utilizando transformada de *Wavelets* a sinais de vozes masculinas e femininas impregnados de ruído com característica de distribuição gaussiana. A aplicação da técnica necessita de um conjunto de parâmetros para a obtenção de melhores resultados. Este trabalho visa a generalização de tais parâmetros para locutores de ambos os sexos, de forma que se possa utilizar tais parâmetros para a aplicação em sistemas que utilizem a técnica.

## 1.4 APRESENTAÇÃO DO TRABALHO

A presente dissertação foi estruturada em oito capítulos. No capítulo 2 são expostos aspectos relacionados ao som, sua percepção e geração pelos órgãos auditivo e fonador, constituindo-se introdução necessária para um estudo relacionado ao tema. Em seguida, o capítulo 3 faz referência à transformada pioneira no sentido de análise espectral, a Transformada de Fourier, e cujo desenvolvimento e limitações levaram ao desenvolvimento da transformada de *Wavelets*. O capítulo 4 apresenta a teoria relacionada à Transformada de *Wavelets*, abordando desde o aspecto evolutivo até uma introdução ao desenvolvimento dos algoritmos utilizados para seu processamento. Também é feita uma breve alusão ao tema da análise em multi-resolução/codificação de sub-banda. Finalizando a parte de revisão bibliográfica, o capítulo 5 apresenta uma sucinta introdução à teoria de *de-noising* utilizando a Transformada de *Wavelets*, abordando, de forma geral, as técnicas que se destacam na teoria atual. O capítulo 6 descreve a metodologia empregada no desenvolvimento do projeto e os algoritmos utilizados no mesmo. Resultados experimentais são discutidos no capítulo 7, seguidos das conclusões no capítulo 8. Os anexos contém material adicional, como tabelas de dados que complementam os resultados apresentados no capítulo 7, os textos gravados pelos locutores e o conteúdo do CD que acompanha este trabalho.

## 2 O SOM

*Este capítulo apresenta aspectos relacionados ao som, sua geração e percepção, além de uma discussão inicial a respeito do ruído.*

### 2.1 INTRODUÇÃO

O som é uma perturbação ondulatória de energia mecânica que se propaga através da matéria. Esta propagação se dá através de ondas de pressão alternada em um meio deformável, causando compressões e rarefações localizadas e gerando um deslocamento das partículas componentes desse meio [2]. De certa forma, a definição acima pode ser tomada como uma definição em sentido amplo, porém, o som é muito comumente definido dentro do subconjunto das ondas mecânicas que podem ser captadas pelo sentido da audição (humana). A onda sonora é caracterizada fisicamente por sua amplitude, frequência, fase, período, comprimento de onda e velocidade de propagação. Seja uma onda sonora dada pela equação:

$$x(t) = A \sin(\omega t + \varphi) \quad (2.1)$$

Frequência é definida como a quantidade de ocorrências de um evento num intervalo de tempo fixo. Para uma onda sonora, costuma ser registrada em ciclos por segundo, ou Hertz [Hz]. Também, devido à facilidade de representação com funções senoidais, costuma-se utilizar uma medição em radianos por segundo [rad/s] (nesse caso, utiliza-se a letra grega “ $\omega$ ” (ômega) e costuma-se chamá-la de frequência angular). Há outras unidades comuns, como rotações por minuto [rpm], batidas por segundo [bps]. Pode-se relacionar a frequência “ $f$ ” [Hz] à frequência “ $\omega$ ” [rad/s] através da seguinte expressão:

$$\omega = 2\pi f \quad (2.2)$$

Período é o intervalo de tempo fixo que se utiliza para registrar a quantidade de ocorrências do evento, ou seja, é o tempo de duração de um ciclo. Frequência e período não podem ser definidos isoladamente, e ambos se relacionam como grandezas inversas:

$$T = \frac{1}{f} \quad (2.3)$$

O período costuma ser registrado em segundos ou minutos, até mesmo horas, a depender do fenômeno estudado. Além disso, muitas vezes a unidade definida para o período limita a escolha da unidade para a frequência (e vice-versa).

Comprimento de onda é a distância entre unidades repetidas de um padrão de onda. No caso de uma onda senoidal, se refere à distância entre os picos ou vales da onda. É uma grandeza dependente da

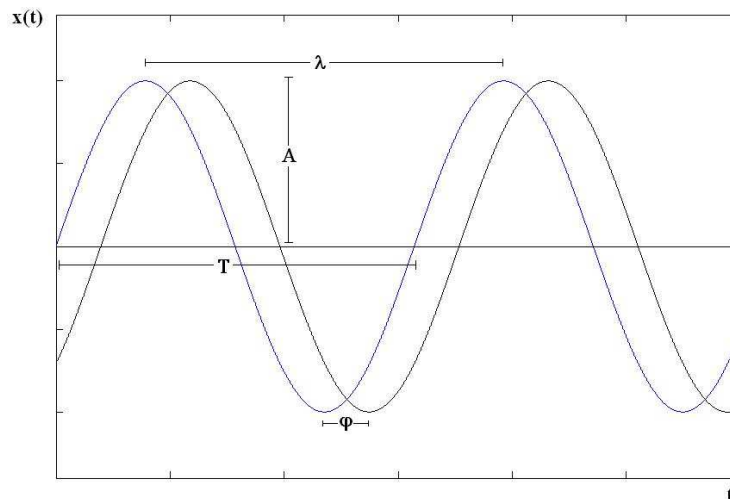


Figura 2.1: Características da Onda Sonora

freqüência e da velocidade de propagação do meio para sua caracterização. Costuma ser medida em metros [m] e representada pela letra grega “ $\lambda$ ” (lâmbda), ou na unidade espacial utilizada para a velocidade e é muito utilizada para caracterizar ondas eletromagnéticas.

A velocidade de propagação caracteriza o meio que será deformado pela onda sonora, sendo afetada pela temperatura e umidade desse mesmo meio. A velocidade de propagação do som no ar é tomada, em média, como aproximadamente 344 metros por segundo [m/s]. A velocidade se relaciona à freqüência e ao comprimento de onda pela seguinte expressão:

$$v = \lambda f \quad (2.4)$$

A fase é definida como o deslocamento da onda no eixo horizontal (no caso, ao deslocamento temporal), sendo definida como o atraso ou defasagem em relação a um padrão de fase nula. A diferença de fase é geralmente caracterizada pela letra grega “ $\phi$ ” (phi) e se refere a um ângulo em rad/s. Pode-se interpretá-la como uma fração do período em que a onda se encontra adiantada ou atrasada em relação à onda de fase nula.

Amplitude é a medida da magnitude de oscilação de uma onda, dada pela metade da distância entre um pico e um vale da onda sonora, e é proporcional à energia transmitida nessa mesma onda. Uma vez que se caracteriza como um valor no eixo vertical, é tomada na mesma unidade que se utiliza para o sinal em si, como decibéis [dB], Volts [V]. A Fig. 2.1 permite a identificação das grandezas anteriormente explicadas.

## 2.2 PERCEPÇÃO DO SOM

O som é percebido através do sentido da audição, por intermédio de órgãos chamados ouvidos, presentes nos seres humanos e em muitos animais (apesar disso, por se tratar de uma onda mecânica, o som também pode ser percebido pelo sentido do tato, a depender de sua intensidade). Os sons são utilizados principalmente para a comunicação, mas possuem diversas outras utilizações, principalmente para a localização [3]. A audição humana é usualmente apresentada limitada à faixa de 16Hz a 20kHz<sup>1</sup>, havendo degradação progressiva ao longo da vida, e sendo os sons abaixo e acima dessa faixa classificados como infra-sons e ultra-sons, respectivamente. Além disso, os níveis de intensidade sonora que podem ser percebidos encontram-se a partir de 20 Pa, ou um nível de pressão sonora (SPL) de 0dB. Intensidades superiores a 85 dB são lesivas ao aparelho auditivo. Uma vez que a audição humana pode detectar sons em uma larga faixa de amplitudes, a pressão sonora é geralmente medida como um nível em uma escala logarítmica em decibel [dB]. O nível de pressão sonora (SPL) é definido como:

$$SPL = 20 \log \left( \frac{p}{p_0} \right), \quad (2.5)$$

onde  $p_0$  é a pressão sonora de referência, geralmente utilizada como 1 ou 20 Pa.

A percepção dos sons pelo aparelho auditivo também possui uma particularidade da qual evoluiu uma representação muito interessante na música ocidental consolidada desde o advento do “Cravo bem temperado” de Johann Sebastian Bach. A voz cantada difere da falada por variar de acordo com uma sucessão de notas, a melodia, podendo se estender em toda a extensão vocal do cantor. Mais do que isso, a imensa diversidade de instrumentos musicais que evoluiu junto à história da humanidade permite que se gerem sons nas mais diversas frequências, ocupando praticamente todo o espectro auditivo humano, e nos mais variados timbres, gerando uma diversidade fascinante de sons influenciados pelos mais diversos fatores (culturais, tecnológicos). A evolução no estudo das sensações geradas no sistema auditivo pela música e uma tentativa de sistematização de um método de registro para os sons que se manipulava culminaram no sistema chamado de música ocidental [4]. A música ocidental utiliza um sistema de representação que divide o espectro auditivo em faixas de frequência chamadas de oitavas ( $8^{as}$ ). Cada oitava é subdividida em 12 unidades de frequência, das quais 7 (sete) são predominantes. É o que se chama de escala musical, as notas Dó(C), Ré(D), Mi(E), Fá(F), Sol(G), Lá(A) e Si(B).<sup>2</sup> A escala é continuada para cada oitava, sendo utilizados nomes para as notas devido à similaridade perceptiva que se obtém entre as notas de diferentes oitavas. Isso se deve à forma como o ouvido humano percebe as frequências de forma logarítmica, em base 2 dois, o que significa que cada vez que a frequência dobra, um som de mesma característica é percebido (a mesma nota é percebida, uma oitava acima). A música utiliza o que se chama de Lá<sub>3</sub> como referência de

<sup>1</sup>apesar de haver estudos que mostrem a possibilidade de audição de sons abaixo de 16Hz, como batimentos

<sup>2</sup>C,C#/Db,D,D#/Eb,E,F,F#/Gb,G,G#/Ab,A,A#/Bb,B



Figura 2.2: A escrita musical

afinação dos instrumentos na frequência de 440Hz. Dessa forma, os sons de 220Hz ou 880Hz também correspondem ao som da nota Lá. Seguindo essa representação por oitavas, dividiu-se o espectro atribuindo números às oitavas desde o -1 até o 8, ou seja, do  $D\acute{o}_{-1}$  ao  $Si_8$  (na música só se utiliza comumente do  $A_0$  ao  $A_8$ , aproximadamente, observando que essas são as frequências das fundamentais) [5]. A figura 2.2 ilustra a notação musical.

## 2.3 SISTEMA AUDITIVO

O aparelho auditivo é responsável pela captação, percepção e interpretação dos estímulos sonoros submetidos aos seres humanos. Para isso, esse aparelho executa uma sequência de transformações de energia durante as etapas de processamento desse som captado. O ouvido é dividido em três partes, o ouvido externo, o médio e o interno. O ouvido externo consiste da pinna (a orelha), o ouvido médio inclui o tímpano e os ossículos; finalmente, o ouvido interno é composto pela cóclea, membrana basilar, escala vestibular, órgão de Corti e nervo auditivo [6]. A energia sonora é inicialmente captada pelo pavilhão auditivo (orelha) e penetra o conduto auditivo até sua extremidade, que coincide com uma delicada membrana - o tímpano. Este serve como transdutor das vibrações sonoras captadas para vibrações mecânicas a serem comunicadas aos ossículos (martelo, bigorna e estribo). O tímpano se desloca em ressonância com a pressão e descompressão alternadas do ar. O centro da membrana timpânica se conecta ao martelo, que, por sua vez, está conectado à bigorna e esta ao estribo. Tais estruturas encontram-se suspensas através de ligamentos, de modo que podem oscilar em um eixo amplificando sua energia e reduzindo sua amplitude, procedimento que será necessário para impor movimento ao líquido coclear. Os ossículos também permitem a seletividade do som que se ouve. Essas ondas que penetram a cóclea forçam o movimento da janela oval para



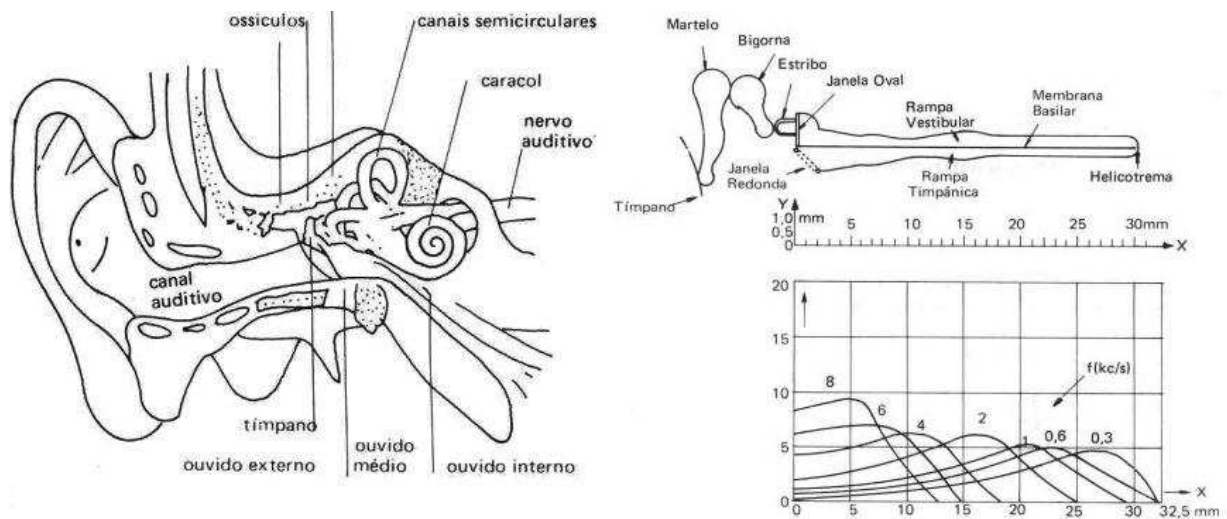


Figura 2.3: O Ouvido Humano

dentro, lançando o líquido da escala vestibular numa profundidade maior dentro da cóclea. A pressão aumentada na escala vestibular desloca a membrana basilar para dentro da escala timpânica e isso faz com que o líquido dessa câmara seja empurrado na direção da janela oval, arqueando-a para fora. Assim, quando as vibrações sonoras provocam o recuo do estribo, o processo é invertido e o líquido se move na direção oposta, pelo mesmo caminho, e a membrana basilar se desloca para dentro da escala vestibular. A vibração da membrana basilar faz com que as células ciliares do órgão de Corti se agitem para frente e para trás, o que excita as células sensoriais e gera impulsos nas pequenas terminações nervosas filamentosas da cóclea, que enlaçam essas células. Tais impulsos são transmitidos pelo nervo coclear até os centros auditivos do tronco encefálico e córtex cerebral. Os sinais nervosos conduzidos já possuem informação sobre as frequências que compõem o som, tal análise é realizada pela membrana basilar do caracol, na qual estão dispostas milhares de células ciliadas. A identificação de frequência sonora se dá através de pontos de ressonância ao longo da membrana basilar. Quando sons de alta frequência penetram a janela oval, a propagação se faz em apenas um trecho da membrana até que se alcance um ponto de ressonância, assim, a membrana se move forçosamente nesse ponto, enquanto o movimento de vibração é mínimo pela membrana. Para uma frequência média, um trecho maior da membrana é percorrido antes de atingir o ponto de ressonância. Os sons graves percorrem praticamente toda a membrana basilar. Dessa forma, as células ciliadas mais próximas à janela oval (início das ondas hidráulicas) transmitem a informação de alta frequência, funcionando como um filtro passa alta não linear. À medida que se distancia da janela oval, as células terão uma melhor resposta em frequência para frequências mais baixas. A intensidade do som é transmitida pela intensidade no movimento das fibras basilares, quanto maior o deslocamento, maior a intensidade do estímulo. A Fig. 2.3 mostra a anatomia do ouvido e o espectro auditivo.

## 2.4 A VOZ

O aparelho fonador é o responsável pela geração do som, geralmente na forma falada ou cantada. Ele é composto por praticamente todos os elementos do sistema respiratório e pelo trato vocal [7]. A voz humana produz sons da seguinte maneira: Primeiro a pressão de ar oriunda dos pulmões gera um fluxo de ar pela traquéia, laringe e faringe. As pregas vocais na laringe vibram, criando flutuações na pressão do ar, que são as ondas sonoras. Essas ondas sonoras são modificadas por ressonâncias impostas pelo trato vocal de acordo com uma série de fatores, como posição dos lábios, da mandíbula, da língua, do palato mole, entre outros aspectos particulares, criando regiões de molde e, portanto, diferentes tipos e qualidades de sons. Estes são, então, expelidos pela boca e nariz em direção ao ambiente. O nível de tensão muscular impresso às pregas vocais é o que vai determinar a frequência fundamental do som expelido. As componentes harmônicas são acumuladas e amplificadas de acordo com as regiões de ressonância presentes em diversas partes do corpo (como seios faciais, peito, parede faríngea). Há também uma caracterização dessas componentes de acordo com o formato do aparelho fonador, ou seja, largura do canal faríngeo, extensão do mesmo, tamanho da boca, tônus habitual das pregas vocais e do diafragma. Isso permite que se identifique uma pessoa através de sua voz ao ouvi-la falada, ou mesmo cantada.

## 2.5 O RUÍDO

Ruído é qualquer som indesejável impregnado a um sinal. O ruído só é um problema quando seu nível é superior ao *threshold* do sentido de audição, situação em que pode ser percebido. O ruído difere da distorção por ser um elemento agregado ao sinal e não parte do mesmo. De acordo com Godsill e Rayner [8], há cerca de cinco classes de degradação de sinal que podem ser agrupadas em duas categorias: globais e locais. A categoria global engloba as degradações que afetam todas as amostras do sinal, entre elas se encontram as classes:

1. ruído de banda-larga ou contínuo - comumente conhecido como *hiss* (chiado);
2. *wow* e *flutter* - resultante de variação de frequência (afinação, *pitch*), o que resulta em modulação de todas as componentes de frequência;
3. distorção - se refere aos demais efeitos não-lineares que influenciam todo o sinal.

Pode-se classificar a degradação local em duas classes:

1. *click* - estalo curto de interferência aleatória no tempo e na frequência;

## 2. transientes ruidosos de baixa frequência.

Pode-se, portanto, destacar seis formas básicas de ruído que podem ser combinadas para influenciar o sinal com diferentes efeitos.

Para este projeto, utilizou-se exclusivamente o ruído gaussiano, ou branco com valor médio nulo. Esse tipo de ruído é importante no estudo teórico e na modelagem de sistemas, pois se estende uniformemente por todo o espectro de frequência [9].

## 3 FOURIER

*Este capítulo relaciona a evolução da transformada de Fourier e a histórica identificação de suas limitações à análise de sinais, convergindo na STFT.*

### 3.1 TRANSFORMADA DE FOURIER

No início do século XIX, o físico-matemático francês Jean-Baptiste Joseph Fourier publicou sua *Théorie Analytique de la Chaleur* (Teoria Analítica do Calor), concluída em 1822. Através do estudo da solução de equações diferenciais parciais na propagação de calor, Fourier conseguiu conceber um método para representação de funções periódicas como uma soma (combinação) de funções exponenciais complexas (senos e cossenos). Muitos anos após o amadurecimento de sua descoberta dessas propriedades das funções periódicas, suas idéias foram generalizadas inicialmente para funções não periódicas, e, posteriormente, para sinais discretos periódicos ou não. Foi após esta última generalização que a manipulação de Fourier se tornou uma ferramenta adequada para cálculos computacionais. Em 1965, um novo algoritmo chamado de *Fast Fourier Transform* (FFT) foi desenvolvido e a Transformada de Fourier se tornou ainda mais popular [10], [11].

A transformada de Fourier é definida da seguinte forma:

$$X(\omega) = F[x(t)] = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt \quad (3.1)$$

e, ainda,

$$x(t) = F^{-1}[X(\omega)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)e^{j\omega t} d\omega \quad (3.2)$$

onde  $x(t)$  representa o sinal no domínio do tempo e  $X(\omega)$  o mesmo sinal no domínio da transformada de Fourier, ou, simplesmente, no domínio da frequência. A equação 3.1 é chamada de transformada de Fourier de  $x(t)$ , e a equação 3.2 é chamada de transformada inversa de Fourier de  $X(\omega)$ . Aprofundando a análise das equações em questão e observando que  $\delta(t) = F^{-1}[1]$ , pode-se concluir que a transformada de Fourier de uma função é a soma de todos os pontos desta função multiplicados por funções impulso em cada intervalo de tempo. A informação provida pela integral corresponde a todos os intervalos de tempo, uma vez que é em  $(-\infty, \infty)$ . Isso mostra que não importa onde no tempo a componente de frequência “f” aparece, ela vai afetar o resultado da integração da mesma forma. A transformada de Fourier mostra a existência ou não de determinadas componentes de frequência. Tal informação é independente da posição no tempo dessa componente. A transformada de Fourier, portanto, converte um sinal de um domínio temporal inteiramente para um domínio frequencial. Neste processo, é obtida com a maior precisão possível a definição do sinal no domínio da frequência, mas toda a informação temporal é descartada. Esta questão

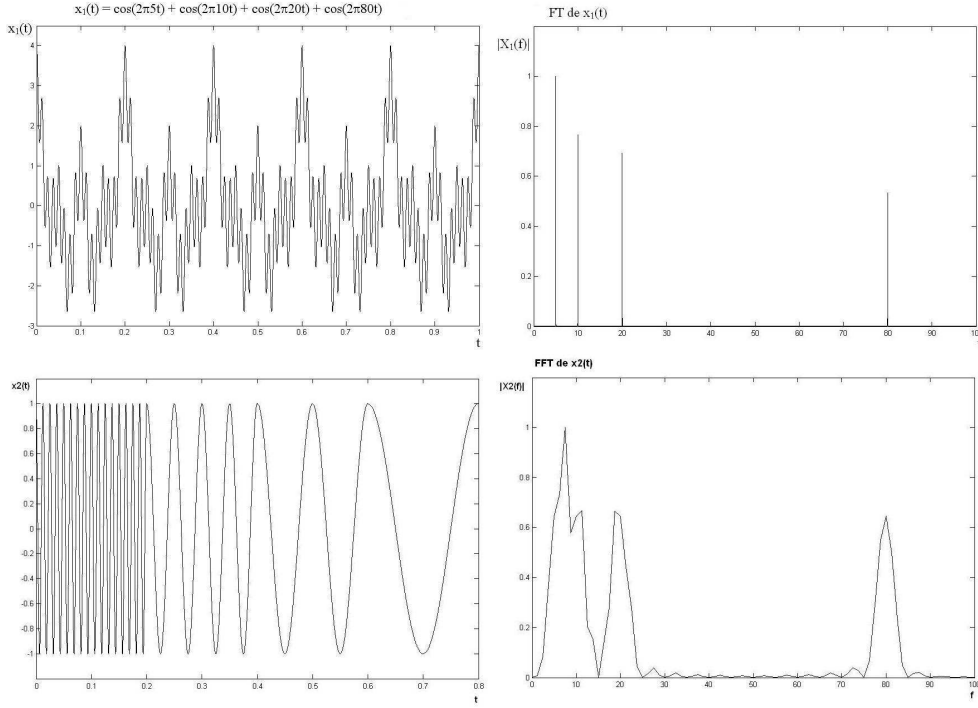


Figura 3.1: FT dos sinais  $x_1(t)$  e  $x_2(t)$

não se apresenta como um problema para sinais cujas componentes de frequência não variam no tempo (sinais estacionários), uma vez que, neste caso, sabe-se que tais componentes estão presentes todo o tempo (ou que não estão). Para situações como esta, a transformada de Fourier realiza análises com perfeição. O grande revés que se apresenta na maioria dos sinais reais é que estes não podem ser considerados absolutamente estacionários. E, sendo assim, a análise através da transformada de Fourier pode trazer alguns inconvenientes [12]. Sejam os sinais  $x_1(t)$  e  $x_2(t)$  definidos como a seguir:

$$x_1(t) = \cos(2\pi 5t) + \cos(2\pi 10t) + \cos(2\pi 20t) + \cos(2\pi 80t) \quad (3.3)$$

$$x_2(t) = \begin{cases} \cos(2\pi 80t), & t = (0; 0, 2) \\ \cos(2\pi 20t), & (0, 2; 0, 4) \\ \cos(2\pi 10t), & (0, 4; 0, 6) \\ \cos(2\pi 5t), & (0, 6; 0, 8) \end{cases} \quad (3.4)$$

Pode-se notar que o sinal  $x_2(t)$ , ao contrário de  $x_1(t)$ , não é um sinal estacionário, uma vez que apresenta composição frequencial variável para diferentes intervalos de tempo, o que não ocorre com  $x_1(t)$ , que é um sinal periódico composto de diversas componentes que ocorrem em todos os instantes. Calculando-se as transformadas de Fourier para ambos os sinais e esboçando o espectro de frequência pelo módulo das componentes, obtêm-se os espectros da Fig. 3.1.

Observa-se que, apesar das distorções causadas pelas descontinuidades no sinal  $x_2(t)$ , ambos os sinais possuem a mesma característica no domínio da frequência, ou, no mesmo sentido, a transformada de

Fourier de ambas fornece a mesma informação. O que se pode observar são quatro impulsos centrados nas frequências das cossenóides que compõem o sinal. Uma análise dos espectros de Fourier obtidos não seria suficiente para diferenciar os dois sinais. Tal análise careceria das informações temporais, que são completamente descartadas quando se passa para o domínio da frequência após a utilização do operador da transformada de Fourier. O problema é que a própria frequência se apresenta “em função do tempo” no sinal  $x_2(t)$ , e, a não ser que se obtenha a informação temporal, não será possível uma caracterização eficiente de ambos. Isso exemplifica uma das limitações da análise pela transformada clássica de Fourier [12].

### 3.2 TRANSFORMADA DE FOURIER DE CURTO TERMO

Uma questão que surge diante do problema deixado pela seção anterior é relacionada à possibilidade de considerar o sinal  $x_2(t)$  (ou qualquer sinal não-estacionário) estacionário para determinados intervalos de tempo e proceder à mesma análise para cada um desses intervalos. Esta é a solução proposta pela chamada Transformada de Fourier de Curto Termo, ou, simplesmente, STFT (*Short Term Fourier Transform*). Há uma sutil diferença entre a STFT e a FT clássica. Na STFT, o sinal é segmentado em intervalos de tempo em que possa ser considerado estacionário. Para isso, utiliza-se uma função de janela “w” de largura igual à do segmento que se pretende utilizar. A função  $w(t)$  (de largura  $T$  segundos) é inicialmente centralizada no ponto  $t = 0$  s, de modo a compreender o segmento  $t = [0, T/2]$  e multiplicada pelo sinal, de modo que se obtenha apenas um segmento resultante. Calcula-se então a FT para o segmento resultante, e isto corresponderá à STFT do primeiro segmento do sinal. A partir de então, a janela é deslocada para cada um dos próximos segmentos do sinal, e a FT é calculada para todos estes segmentos. As FTs de todos os segmentos compõem a STFT do sinal analisado. A STFT é definida da seguinte forma:

$$STFT[x(t)](\tau, \omega) = \int_{-\infty}^{\infty} [x(t)w^*(t - \tau)]e^{-j\omega t} dt \quad (3.5)$$

É interessante observar que a STFT é uma função de duas variáveis, tempo e frequência (na verdade deslocamento e frequência). O módulo da STFT é chamado de espectrograma da função.

$$espectrograma \{x(t)\} \triangleq |X(\tau, \omega)| \quad (3.6)$$

Observa-se, então, que através da utilização da STFT é possível obter uma representação tempo-frequência de um sinal, portanto, um sinal unidimensional (ou bidimensional, se for considerada a amplitude como uma das dimensões, tempo e amplitude) como um sinal de áudio é convertido em um sinal bidimensional (tridimensional) no domínio da STFT. Desta forma, a STFT soluciona parcialmente o problema da representação de sinais não-estacionários. Isso poderia levar à conclusão equivocada que a STFT é

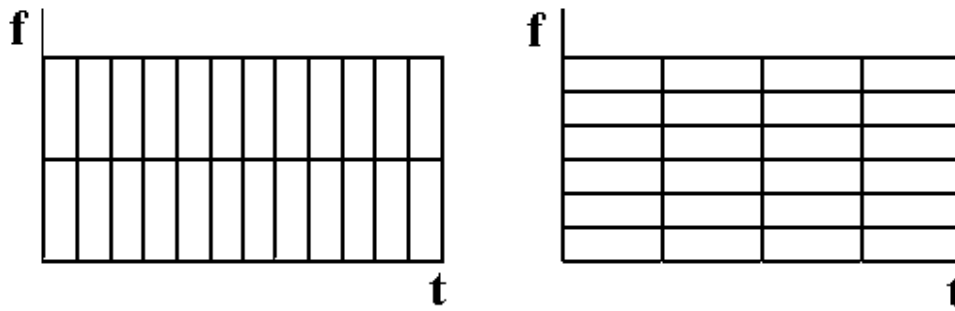


Figura 3.2: Quadro de resolução da STFT

a solução definitiva para a representação de sinais não-estacionários, uma vez que basta escolher janelas de tamanhos infinitesimais (ou funções de Dirac) e assim se obterá a representação de forma excelente [12]. A próxima seção discutirá o problema envolvido na escolha da janela de corte.

### 3.3 PRINCÍPIO DA INCERTEZA DE HEISENBERG

Originalmente enunciado como um dos postulados da mecânica quântica, o princípio da incerteza de Heisenberg afirma que não é possível determinar, ao mesmo tempo, o momentum e a posição de um elétron. Da mesma forma, tal princípio encontra uma aplicação na teoria de processamento de sinais. Não é possível determinar em que instante de tempo está localizada uma componente freqüencial do sinal, ou seja, não é possível representar um sinal como pontos nos domínios tempo-freqüência. O máximo que se pode obter são bandas de freqüência associadas a intervalos de tempo de ocorrência, ou seja, funções de densidade de probabilidade. Uma definição se faz necessária neste ponto; fala-se em resolução da STFT, o que significa a precisão com que um sinal é representado em um dos semi-domínios, tempo ou freqüência. No caso da STFT, tem-se uma resolução fixa, determinada pela função-janela utilizada, que determinará o nível de resolução no tempo ou na freqüência. Assim, se for utilizada uma janela longa, será obtida uma alta resolução de freqüência (no caso extremo, se obterá a FT do sinal, enxergando-se a função janela como uma função  $w(t) = 1$ ) em prejuízo da resolução temporal, que será mínima. Da mesma forma, se a janela de corte for mínima (como uma função impulso), haverá uma boa resolução temporal em detrimento da resolução freqüencial (como a própria função no tempo, no caso extremo). Como se pode notar, a escolha da janela de corte é um compromisso do engenheiro, a depender da necessidade da aplicação. A Fig 3.2 ilustra situações de melhor resolução (a) temporal ou melhor resolução freqüencial (b). É importante ressaltar que, apesar de se poder variar o tamanho da janela de modo a melhorar ou piorar a resolução em um dos dois semi-domínios, tal variação afeta ambos os domínios proporcionalmente. Isso equivale a dizer que a área de cada um dos blocos da figura acima deverá permanecer constante para qualquer

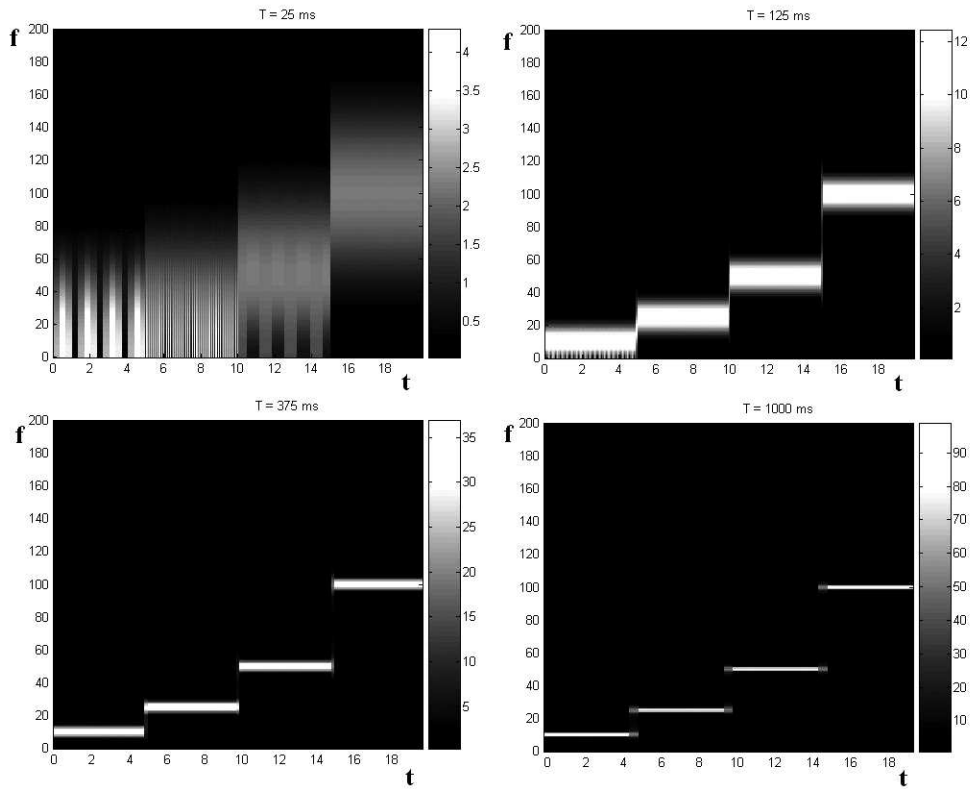


Figura 3.3: Espectrograma da função  $x_3(t)$

tamanho de janela e, no caso da STFT, o mesmo formato de janela é aplicado para todos os segmentos do sinal analisado. Esta limitação desta transformada, é a questão que motivou o início do estudo da análise em multi-resolução[10],[11],[12].

Seja o seguinte sinal,

$$x_3(t) = \begin{cases} \cos(2\pi 10t), & t = (0; 0, 1) \\ \cos(2\pi 25t), & (0, 1; 0, 2) \\ \cos(2\pi 50t), & (0, 2; 0, 3) \\ \cos(2\pi 100t), & (0, 3; 0, 4) \end{cases} \quad (3.7)$$

para este sinal, foram esboçados os espectrogramas para 25, 125, 375 e 1000 ms, de modo a tornar perceptível a variação de resolução nos dois semi-domínios da STFT. Na Fig 3.3, observa-se que, a exemplo do que havia sido discutido, a janela de 25 ms apresenta melhor resolução temporal, e à medida que o tamanho dessa janela é aumentado, até o extremo de 1000 ms, a resolução temporal é deteriorada para a melhora da resolução freqüencial. Fazendo também uma referência ao caso discreto, a área de cada quadro no espectrograma também é limitada pela taxa de amostragem utilizada para o sinal, sendo possível reduzir esta área até o caso contínuo, à medida que a taxa de amostragem tende a valores muito altos.



### 3.4 TRANSFORMADA DISCRETA DE FOURIER

Na verdade, a transformada original de Fourier não pode ser aplicada diretamente a sinais de áudio digitais, uma vez que estes são discretos [10]. É necessária, portanto, uma versão discreta da transformada de Fourier, a DFT (*Discrete Fourier Transform*), que transforma um sinal discreto em um espectro de frequências discretizado. Ao contrário da transformada contínua, ela pode ser aplicada a um número limitado de amostras com a restrição de que a sequência analisada é admitida como um período de um sinal periódico. Para a análise computacional através da DFT foram desenvolvidos algoritmos velozes, chamados de FFT (Fast Fourier Transform). A saída do operador DFT pode ser interpretada como amplitudes de bandas de frequência. Cada banda tem uma largura fixa e uma frequência central (a frequência principal que é analisada).

#### 3.4.1 Janelamento

A discretização da transformada de Fourier apresenta um problema, há uma distorção causada por frequências que não são múltiplas da fundamental em uma banda, que causam um espalhamento da energia até mesmo em outras bandas. Uma técnica utilizada para lidar com esse problema é chamada de janelamento (do inglês *windowing*). No domínio do tempo o sinal é associado com uma janela que suaviza os extremos reduzindo o espalhamento. A utilização de janelamento causa alterações no espectro da função original. Algumas janelas comumente utilizadas são a Hanning, a Blackman e a Hamming.

A STFT, tanto pelo desenvolvimento histórico da FT, que gerou uma grande quantidade de pesquisas na sua aplicação e desenvolvimento de algoritmos eficientes para o seu processamento, quanto pelo pioneirismo na apresentação de uma solução para análise em tempo e frequência, se tornou uma ferramenta muito poderosa para o processamento de sinais. Algumas de suas aplicações incluem alguns formatos de compactação de áudio, vídeo e dados, como o mp3 (que utiliza a DCT, derivada da STFT).

## 4 WAVELETS

*Este capítulo apresenta a evolução histórica da transformada de Wavelets, contendo aspectos de definição geral que incluem a CWT, as séries de Wavelets e a DWT.*

### 4.1 ANÁLISE EM MULTI-RESOLUÇÃO

Na seção anterior foi apresentada a questão da limitação da resolução no tempo e frequência na análise de sinais e como essa limitação é decorrente de um fenômeno físico, existindo independentemente da transformada utilizada. Apesar disso, o amadurecimento da teoria de processamento de sinais permitiu o desenvolvimento de formas elegantes de contornar o problema da incerteza. Uma dessas abordagens é chamada de análise em multi-resolução (MRA), que consiste, basicamente, como o próprio nome já diz, em analisar o sinal em diferentes frequências com diferentes resoluções. Esse aspecto difere da análise de Fourier em que cada componente espectral (faixa espectral) é analisada da mesma maneira. Uma das possibilidades da MRA é a imposição de boa resolução no tempo (e baixa resolução em frequência) para altas frequências e baixa resolução no tempo (e conseqüente boa resolução em frequência) para baixas frequências. Esse modo de análise é especialmente útil para sinais em que as componentes de alta frequência ocorram em curtos intervalos de tempo e as componentes de mais baixa frequência durem mais tempo, e este tipo de sinal é muito comum de se encontrar em situações reais. A MRA permitiu a inserção de um novo paradigma no processamento de sinais introduzindo essa noção de análise por escala.

Um aspecto interessante pode ser obtido observando-se a análise pela DFT [10]. Uma vez que o sinal analisado pela DFT é discreto, o espectro resultante pode ser interpretado como amplitudes de bandas de frequência. Cada banda tem uma largura específica em torno de uma frequência central (a frequência analisada). Como exemplo, suponha-se um sinal de 1s amostrado a 44100 Hz (2x a largura de banda do sinal analisado). Se forem utilizadas 512 faixas para a análise, as bandas de frequência terão um espaçamento de aproximadamente 43 Hz, sendo a primeira banda com frequência central em 0Hz, a segunda em 43Hz e as demais seguindo a mesma lógica. 43Hz é a frequência fundamental, as outras frequências centrais analisadas são múltiplas dela. Essa distribuição uniforme das bandas de frequência apresenta um problema para a análise de sinais de áudio, já que, como foi discutido anteriormente, a audição humana segue uma escala logarítmica na percepção de frequências. Um intervalo de uma oitava (mais precisamente uma 8ª justa, de um  $Dó_1$  a um  $Dó_2$ ) soa como o mesmo intervalo sendo ele deslocado duas notas para cima ou baixo, mantido o mesmo intervalo entre elas. Dito de outra forma, o intervalo de 100Hz a 200Hz soa como o intervalo de 5kHz a 10kHz. Isso é observado em aplicações da DFT de detecção de altura (frequência

ou “pitch”) de um som; há um problema relacionado às baixas frequências, uma vez que estas possuem baixa resolução logarítmica comparadas às altas frequências. Aplicando-se o exemplo utilizado acima de 512 faixas, há aproximadamente apenas uma banda para análise da oitava que vai de 30 Hz a 60 Hz e, em contrapartida, a oitava de 3kHz a 6kHz é representada por 100 faixas. Isso impede a detecção da altura da nota em 50Hz, uma vez que a única coisa que se sabe é que a mesma se encontra na segunda faixa, próxima aos 43Hz. Para uma nota muito aguda (de frequência muito alta) a altura pode ser determinada com muita precisão, já que há 100 bandas para representar 12 semitons. Daí o fato da DFT gerar poucos detalhes para baixas frequências e muitos detalhes para altas frequências. E é neste ponto que aparece a proposta da transformada de *Wavelets*.

## 4.2 TRANSFORMADA DE WAVELETS

### 4.2.1 Histórico

A Transformada de *Wavelets* é uma das ferramentas recentes criadas na tentativa de se superar as limitações já expostas encontradas na análise através da transformada de Fourier. Apesar do desenvolvimento mais recente, as primeiras menções à teoria de *wavelets* alcançam quase um século, como um apêndice da tese de doutorado do matemático húngaro Alfréd Haar em 1909, que introduziu uma família de funções de suporte compacto à qual foi atribuído seu nome. Muito tempo se passou desde então até que, na década de 30, o físico francês Paul Levy utilizou as funções-base de Haar em sua pesquisa sobre movimento Browniano, e concluiu que estas funções-base de escala variável eram superiores às funções-base de Fourier na representação de pequenos detalhes de uma função. Uma pesquisa paralela por Littlewood, Paley e Stein envolvendo o cálculo da energia de uma função  $f(x)$ , levou a:

$$E = \frac{1}{2} \int_0^{2\pi} |f(x)|^2 dx \quad (4.1)$$

O cálculo produziu diferentes resultados se a energia estava concentrada em alguns pontos ou distribuída em um intervalo maior. Esse resultado perturbou os cientistas porque indicava que a energia podia não ser conservada. Eles descobriram uma função que pode variar em escala e conservar a energia ao calcular-se sua energia funcional. Entre 1960-80, os matemáticos Guido Weiss e Ronald Coifman estudaram os elementos mais simples de um espaço de funções, chamados átomos, com o objetivo de encontrar átomos para uma função comum e descobrir as “regras de estruturação” que permitem a reconstrução de todos os elementos de um espaço de funções utilizando estes átomos. Apesar de se observarem alguns avanços na teoria de *wavelets* desde 1930, foi Jean Morlet que introduziu os desenvolvimentos mais marcantes por volta do ano 1975. Morlet foi o primeiro a utilizar o nome *wavelet* (inicialmente ondelette). Ele trabalhava em uma companhia exploradora de petróleo em um projeto utilizando a FT e a STFT para análise de sinais

sísmicos. Porém, ele se deparou com a limitação de resolução da STFT e percebeu que a análise não era eficiente com uma janela fixa. Ele então inverteu o raciocínio, mantendo fixa a frequência da função e variando a janela. Ele descobriu que alargar a janela aumentava a função e encolher a janela a comprimia. Este conceito de Morlet foi abordado anteriormente com as *wavelets* de Haar. A 0-ésima *wavelet* filha era simplesmente uma versão comprimida da mãe. Começando-se pela *wavelet*-mãe no intervalo  $[0,1]$ , pode-se comprimir a janela na metade e obter a 0-ésima *wavelet*-filha, que possui a mesma forma mas é definida em um intervalo menor,  $[0,1/2]$ . Em 1981, Morlet se juntou ao físico croata Alex Grossman, de modo a trabalhar no desenvolvimento de sua descoberta, a idéia de que um sinal podia ser transformado em um formato de *wavelet* e depois transformado de volta ao original sem perda de informação. Os dois físicos foram bem sucedidos em seus esforços, formalizando a teoria para a Transformada de *Wavelets* Contínua. Os próximos protagonistas na história das *wavelets* foram o matemático francês Yves Meyer e Stephane Mallat, por volta de 1986, introduzindo a idéia da análise multi-resolução para as *wavelets* através da função de escalonamento e a utilização das mesmas no processamento digital de sinais. Há muitos outros nomes de importância no desenvolvimento da teoria atual de *wavelets* na década de 80-90, entre eles Jan-Olov Strömberg, com seu trabalho em *wavelets* discretas e Ingrid Daubechies, que descobriu uma série de funções-base para utilização com *wavelets* (entre elas as Daubechies e as Coiflets).

#### 4.2.2 Transformada de Wavelets Contínua

A CWT, definida por Morlet e Grossman, é formalmente escrita como [13]:

$$\gamma(s, \tau) = \int f(t) \psi_{s,\tau}^* dt \quad (4.2)$$

Esta equação descreve a decomposição de uma função  $f(t)$  em um conjunto de funções-base  $\psi_{s,\tau}(t)$ . A função  $\gamma(s, \tau)$  é a correspondente da função  $f(t)$  no domínio da CWT. As variáveis “s” e “ $\tau$ ” definem as dimensões escala e deslocamento, respectivamente. A transformada inversa é dada por:

$$f(t) = \int \int \gamma(s, \tau) \psi_{s,\tau}(t) d\tau ds \quad (4.3)$$

O termo *wavelet* significa uma pequena onda (ondelette), nomenclatura essa que alude às condições de que esta função (janela) tem largura finita (suporte compacto, que significa a concentração da energia em um intervalo finito de tempo) e é oscilatória (por isso wave). As funções-base  $\psi_{s,\tau}(t)$  são derivadas de funções chamadas de *wavelets*-mãe, uma vez que tais funções são versões escalonadas e deslocadas de uma função modelo. Esta relação é definida na seguinte equação:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi \left( \frac{t - \tau}{s} \right) \quad (4.4)$$

Na equação 4.4, o termo  $s^{-1/2}$  é usado para normalização da energia ao longo de diferentes escalas. Pode-se notar, em comparação com a STFT, que a variável de deslocamento ( $\tau$ ) utilizada nas duas transformadas

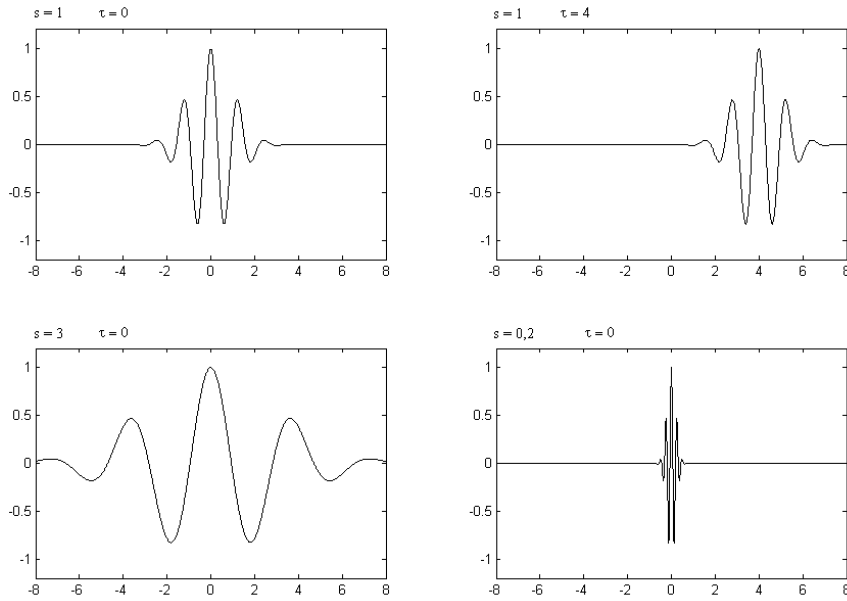


Figura 4.1: Escalonamento da *Wavelet*-mãe

é a mesma, havendo uma compatibilidade na resolução temporal. Porém, a CWT utiliza o termo de escala ( $s$ ) em vez da frequência ( $\omega$ ), utilizada na STFT; isso, aliado ao fato de se utilizar o termo no denominador, gera uma resolução inversa em relação à frequência (e à STFT), ou seja, quanto maior a escala, menor a frequência. A Fig 4.1 ilustra as derivações da *wavelet*-mãe.

### 4.2.3 Cálculo da CWT

A CWT é calculada através do corte do sinal a ser transformado ( $f(t)$ ) por uma janela de grande escala (inicialmente) definida pela *wavelet*-mãe que é deslocada por todo o sinal até que este tenha sido inteiramente coberto. Em seguida, o valor da escala diminui, diminuindo também a largura da janela e tornando-se a deslocá-la para cobrir todo o sinal. Tal procedimento encontra-se ilustrado na Fig. 4.2, para a qual se utilizou um sinal com a função de Morlet e a *wavelet*-mãe Mexican Hat como função-base [12]. Para comparar a análise através da CWT com a análise equivalente através da STFT, foram esboçados os coeficientes CWT da função  $x_3(t)$  (equação 3.7) do capítulo anterior, que é uma composição de senóides em vários tempos diferentes. Como foi dito na ocasião, para uma análise eficiente pela STFT, deve-se escolher uma janela de corte de tamanho adequado. Com a CWT não é necessário dimensionar a janela. A CWT da Fig. 4.3 utilizou as funções Mexican Hat como funções-base. Como se pode observar na Fig. 4.3, os eixos utilizados são a escala e a translação(deslocamento). A translação pode ser entendida como o tempo acumulado até o instante marcado desde  $t = 0$ . A escala não permite uma analogia tão direta. Pode-se notar na figura que as frequências mais altas ocupam os menores intervalos de escala, além de terem, na mesma proporção, melhor especificidade no eixo de translação. As frequências mais baixas, porém, têm

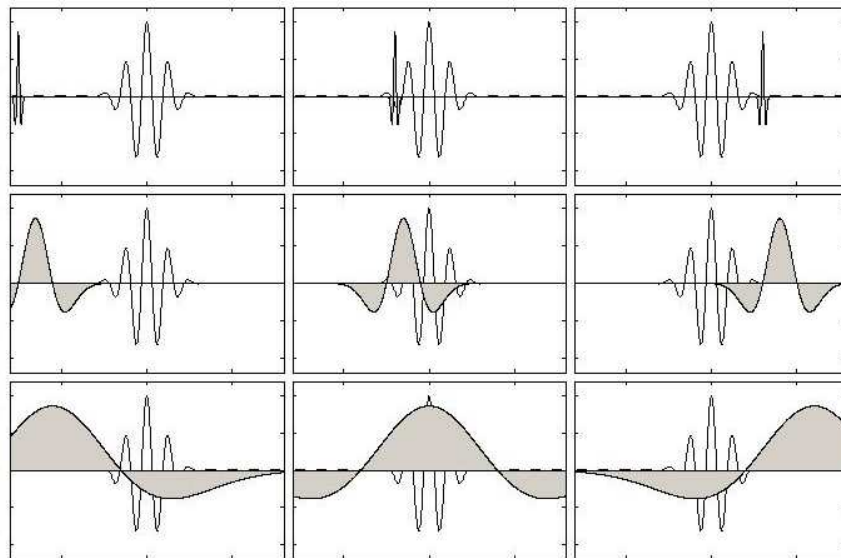


Figura 4.2: Cálculo da CWT

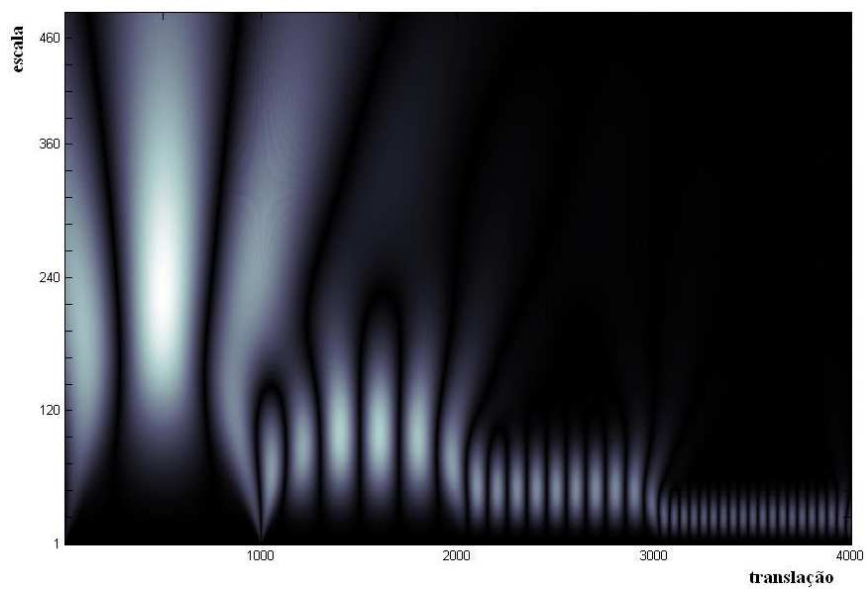


Figura 4.3: CWT da função  $x_3(t)$

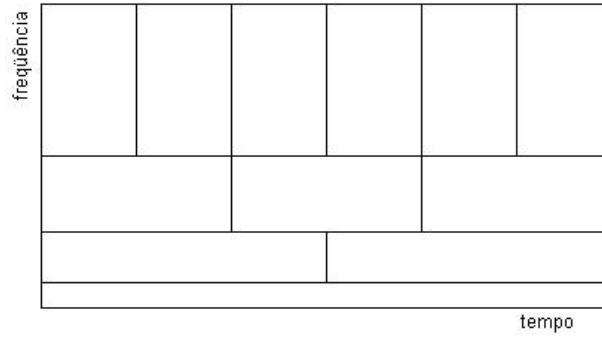


Figura 4.4: Resolução tempo x escala da WT

pior definição no tempo e ocupam maiores intervalos de escala. A figura permite localizar com razoável precisão os intervalos de ocorrência das quatro frequências que compõem a função analisada. Em relação à resolução, ao contrário da STFT, que possui resolução constante, a CWT possui melhor resolução temporal para as altas frequências (e conseqüente pior resolução frequencial) e melhor resolução frequencial para as baixas frequências (pior resolução temporal). Na figura acima, observa-se que as frequências mais altas possuem melhor resolução no eixo das escalas (são mais facilmente localizadas), mas por ser a escala o inverso da frequência, isso indica uma fraca resolução em frequência. Do mesmo modo ocorre com as frequências mais altas que têm maior espalhamento no eixo da escala. Por isso possuem melhor resolução frequencial. Em analogia ao modelo apresentado para a organização das janelas da STFT, a Fig. 4.4 ilustra um modelo para a WT. Como pode ser observado na Fig. 4.4, os comprimentos das bases dos retângulos se alteram à medida que a escala aumenta. Apesar dessa mudança, as áreas dos mesmos se mantêm constantes. A área, como já foi dito, é uma proporção do princípio da incerteza de Heisenberg. Não se pode determinar dentro de cada retângulo a relação pontual entre a frequência e o tempo. Esta área também é igual às áreas dos retângulos da STFT. Um ponto interessante é que, apesar de não ser possível ultrapassar esse limite de área, é possível utilizar diferentes formatos de “lacunas”, através das representações por diferentes *wavelets*-mãe.

#### 4.2.4 Propriedades das Wavelets

As principais propriedades das *wavelets* são as condições de admissibilidade e regularidade, e são essas propriedades que as caracterizam como pequenas ondas (*wavelets*). A primeira condição de admissibilidade é dada por [14]:

$$\int \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty \quad (4.5)$$

Funções  $\psi(t)$  que satisfaçam essa condição podem ser usadas para analisar e reconstruir um sinal sem perda de informação. A equação 4.5 indica que a FT de  $\psi(t)$  desaparece na frequência zero,  $\Psi(\omega)|_{\omega=0} = 0$ .

Também é necessário que o valor médio da *wavelet* no tempo seja nulo e que ela seja oscilante, ou seja,

$$\int \psi(t)dt = 0 \quad (4.6)$$

Como se pode observar pela equação 4.2, uma função unidimensional se torna bidimensional através de sua representação por *wavelets*. Isso significa maior quantidade de dados nesta representação (apesar de que isso não significa maior quantidade de informação), e esta não é uma característica desejável em aplicações práticas. Por este motivo, impõem-se as condições de regularidade às *wavelets*, que requerem que haja alguma suavidade e concentração em ambos os domínios (tempo e frequência). Expandindo a transformada *wavelet* em séries de Taylor com a simplificação de  $\tau=0$  [14],

$$\gamma(s, 0) = \frac{1}{\sqrt{s}} \left[ \sum_{p=0}^n f^{(p)}(0) \int \frac{t^p}{p!} \psi\left(\frac{t}{s}\right) dt + O(n+1) \right], \quad (4.7)$$

onde  $f^{(p)}$  é a p-ésima derivada de  $f$  e  $O(n+1)$  é o resto da expansão. Seguindo a análise, pode-se definir os momentos da *wavelet* como:

$$M_p = \int t^p \psi(t) dt \quad (4.8)$$

Substituindo 4.8 em 4.7, obtém-se a especificação dos momentos:

$$\gamma(s, 0) = \frac{1}{\sqrt{s}} \left[ f(0)M_0s + \frac{f^{(1)}(0)}{1!}M_1s + \frac{f^{(2)}(0)}{2!}M_2s^3 + \dots + \frac{f^{(n)}(0)}{n!}M_ns^{n+1} + O(s^{n+2}) \right] \quad (4.9)$$

Na equação 4.9 acima,  $M_0$  é nulo, por causa da condição de oscilação 4.6. É interessante que mais desses momentos se anulem, de modo a se obter uma redução nos coeficientes  $\gamma(s, \tau)$ . A ordem de aproximação de uma Transformada de *Wavelets* é dada pelo número “N” de momentos desprezados [15].

#### 4.2.5 Transformada de Wavelets Discreta

A CWT possui três peculiaridades na sua aplicação prática [11]. Uma vez que a CWT é calculada através do deslocamento contínuo de uma janela contínua por toda a extensão do sinal contínuo, (calculando-se a correlação entre ambos) e essa mesma janela é escalonada continuamente, refazendo-se os deslocamentos contínuos, isso gera uma enorme quantidade de coeficientes que armazenam a informação do sinal com um alto grau de redundância, tornando inviável a análise utilizando-se a CWT, já que o produto tempo x largura de banda do sinal decomposto é o quadrado do do sinal original. Um outro problema que se apresenta mesmo após a solução da redundância (com a discretização do deslocamento e do escalonamento) é que a decomposição ainda apresenta uma quantidade infinita de coeficientes na representação da função. O terceiro problema é a falta de soluções analíticas, tornando necessário um grande esforço computacional para a decomposição, sendo necessário utilizar-se cálculo numérico ou computação analógica.



#### 4.2.5.1 Redundância

Para o problema da redundância, uma solução imediata é a discretização da função de escalonamento. Isso é possível escolhendo-se passos de dilatação e translação para  $s$  e  $\tau$ . A escolha do parâmetro “ $s$ ” geralmente segue um modelo logarítmico -  $s = s_0^j$  (normalmente de base 2 para corresponder à percepção auditiva humana). E o parâmetro de translação, “ $\tau$ ”, é discretizado usando-se a regra de amostragem em qualquer escala.  $\tau = k\tau_0 s_0^j$ . Desse modo a equação 4.4 se torna:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s_0^j}} \psi\left(\frac{t - k\tau_0 s_0^j}{s_0^j}\right) \quad (4.10)$$

Esse desenvolvimento é chamado de decomposição em Séries de *Wavelets* (WS). Mesmo com a redução da redundância, (e, portanto, discretização dos semi-domínios translação e escala) a reconstrução de um sinal através dos coeficientes de sua WS ainda é possível. Isso é demonstrado em [16] e a condição para que seja possível tal reconstrução é que a energia dos coeficientes seja limitada, ou seja,

$$A \|f\|^2 \leq \sum_{j,k} |\langle f, \psi_{j,k} \rangle|^2 \leq B \|f\|^2 \quad (4.11)$$

Caso se satisfaça essa condição, a família de funções base  $\psi_{j,k}$  é chamada de quadro com fronteiras A e B. Se  $A = B$ , o quadro se torna uma base ortonormal e a reconstrução pode ser conseguida com a mesma *wavelet* da decomposição, o que não ocorre para o caso  $A \neq B$ , quando a reconstrução só é possível com um quadro duplo e uma *wavelet* diferente. Nesse caso as duas funções são ditas ortogonais. Formalmente, a condição para ortonormalidade é:

$$\begin{aligned} \gamma_{j,k}(t) &= \int_{-\infty}^{\infty} x(t) \psi_{j,k}(t) dt \\ x(t) &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \gamma_{j,k}(t) \psi_{j,k}(t) \end{aligned} \quad (4.12)$$

$$\langle \psi_{j,k}(t), \psi_{l,m}(t) \rangle = \begin{cases} 1, & \text{para } j = l \text{ e } k = m \\ 0, & \text{para } j \neq l \text{ e } k \neq m \end{cases}$$

E a condição de ortogonalidade,

$$\begin{aligned} \gamma_{j,k}(t) &= \int_{-\infty}^{\infty} x(t) \psi_{j,k}^I(t) dt \\ x(t) &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \gamma_{j,k}(t) \psi_{j,k}^{II}(t) \end{aligned} \quad (4.13)$$

$$\langle \psi_{j,k}^I(t), \psi_{l,m}^{II}(t) \rangle = \begin{cases} 1, & \text{para } j = l \text{ e } k = m \\ 0, & \text{para } j \neq l \text{ e } k \neq m \end{cases}$$

Apesar de solucionar o problema da redundância, a ortogonalidade não é essencial na representação de sinais [14]. Em alguns casos, uma representação redundante é vantajosa por tornar o sinal mais robusto, aumentando a imunidade a ruído ou, ainda, a invariabilidade a deslocamento [15].

#### 4.2.5.2 Redução da quantidade de coeficientes

O segundo problema identificado na CWT e que também se estende às WS se refere à quantidade infinita de coeficientes que ambas utilizam na decomposição dos sinais [11]. Para lidar com o problema é preciso observar sua origem, seu fator gerador. O sinal analisado sempre terá duração limitada, isso limita também a quantidade de deslocamentos (no caso discreto). Ademais, a maior escala possível de se obter no sinal também é limitada pelo teorema de Nyquist. O problema aparece, então, no limite inferior da escala. Uma forma de mostrar isso é observando a similaridade do procedimento de escalonamento com os deslocamentos temporais. O escalonamento executa deslocamentos de filtros de diferentes larguras de banda no domínio da frequência no intuito de cobrir todo o espectro ocupado pelo sinal. A equação 4.5 permite concluir que as *wavelets* têm um espectro que se comporta como um filtro passa faixa e os diversos escalonamentos deslocam e redimensionam a frequência central e a banda passante desse filtro, respectivamente. Isso pode ser observado na figura 4.5, tornando claro o motivo dos infinitos coeficientes: Se se utiliza um passo de 2 na discretização do escalonamento (por exemplo), isso quer dizer que cada vez que muda o valor de “s” a banda da *wavelet* é reduzida na metade. Isso invoca o paradoxo de Zeno de Elea, pois, para cobrir todo o espectro inferior do sinal, é necessário utilizar uma quantidade infinita de filtros (que, a cada iteração, cobrem metade do espectro restante). Um detalhe que não foi discutido é que é preciso que os escalonamentos das *wavelets* realizem algum tipo de sobreposição entre os espectros cobertos pelos filtros. A solução proposta por Mallat [17] para evitar o uso de infinitos coeficientes é simplesmente não tentar cobrir todo o espectro inferior com as *wavelets* e, em vez disso, utilizar uma outra função chamada função escalonante que possui uma natureza de filtro passa-baixa para cobrir o resto do espectro. Desta forma se obtém o resultado da figura 4.6. A função escalonante utiliza infinitos coeficientes até uma escala “j”, e isso desonera as *wavelets*, sendo possível a reconstrução do sinal utilizando-se ambas. O único revés é que agora houve perda no sentido de representação do sinal através de coeficientes *wavelet*. A largura da função escalonante é um parâmetro importante na informação da WT, uma vez que reflete a quantidade de informação representada pelos coeficientes *wavelets*. A  $\phi(t)$  pode ser decomposta em *wavelets* e expressa como em 4.2:

$$\phi(t) = \sum_{j,k} \gamma_{j,k}(t) \psi_{j,k}(t) \quad (4.14)$$

e ela precisa satisfazer uma condição de admissibilidade de

$$\int \phi(t) dt = 1 \quad (4.15)$$

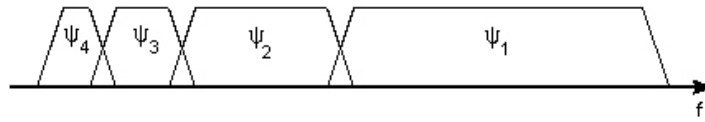


Figura 4.5: Configuração de Banco de Filtros

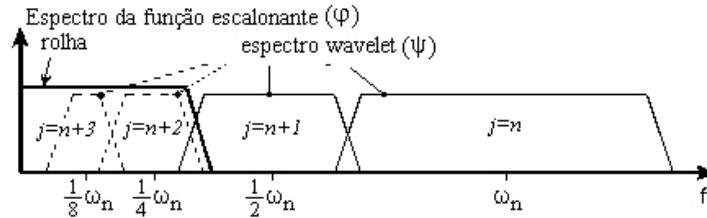


Figura 4.6: A função escalonante

isso significa que o 0-ésimo momento não pode ser nulo.

Portanto, a WT pode ser entendida como um banco de filtros passa faixa de espectro variável. A razão entre a frequência central de um espectro *wavelet* e a largura desse espectro é a mesma para todas as *wavelets* e é referida como o fator de fidelidade  $Q$  de um filtro. A função escalonante é um filtro passa baixa utilizado para cobrir a parte inferior do espectro. O esquema da seção anterior pode ser implementado utilizando-se um banco de filtros, sendo os coeficientes a saída do filtro nos diferentes estágios. A forma usual de se implementar esse banco de filtros é através da utilização de um banco composto por dois filtros: um passa-altas (limitado dentro do espectro do sinal) e um passa baixa, dividindo o espectro em duas faixas, sendo a faixa alta composta pelos pequenos detalhes. A mesma configuração de filtro é aplicada à banda baixa para dividi-la também em duas, prosseguindo desta forma até que se atinja uma quantidade de bandas suficiente. Uma vantagem deste esquema em relação ao projeto isolado de todos os filtros é que só é necessário projetar dois desses. Seguindo a lógica discutida, ao atingir-se o nível pretendido, utiliza-se um filtro passa-baixa para complementar o processo (função escalonante). A figura 4.7 mostra o banco de filtros iterativo. Esse tipo de análise em multi-resolução é referida como codificação de sub-banda utilizando um banco de filtros com  $Q$  constante. O termo “transformada discreta de *wavelets*” (DWT) se refere ao algoritmo de Mallat de 1988. A DWT divide o sinal analisado em componentes de alta escala e baixa escala, sendo as componentes de alta escala chamadas de “aproximações”, e correspondendo ao conteúdo de baixa frequência do sinal. As variações rápidas do sinal são chamadas de “detalhes”. O processo de decomposição da DWT é repetido atuando-se sempre sobre as componentes de aproximações. Sucessivas decomposições geram a chamada “árvore de decomposição *wavelet*”[18].

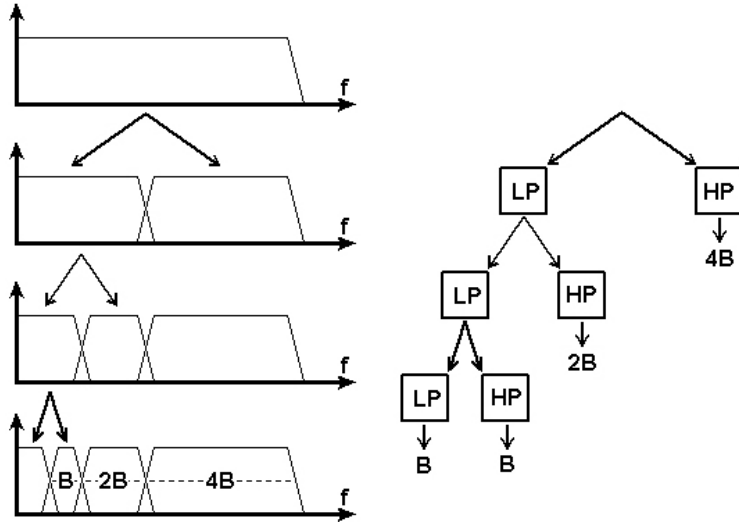


Figura 4.7: Decomposição *Wavelets* implementada com o banco de filtros iterativos

#### 4.2.5.3 A DWT e sua implementação

Apesar de se haver discretizado a translação e escalonamento, as WTs apresentadas até este ponto ainda não são discretas no tempo. Foi visto que a função escalonante pode ser decomposta em *wavelets* até uma escala “j” eq. 4.14. É interessante adicionar um espectro de *wavelets* ao espectro da função escalonante que seja duas vezes mais largo que o primeiro e que contenha a informação da primeira função escalonante. Isso pode ser resumido em uma relação de duas escalas ou formulação em multi-resolução [14] e [15]:

$$\phi(2^j t) = \sum_k h_{j+1}(k) \phi(2^{j+1} t - k) \quad (4.16)$$

Esta expressão significa que a função de escalonamento em uma determinada escala pode ser expressa em termos de funções de escalonamento transladadas na próxima escala reduzida. O conjunto de *wavelets* substituído pela função escalonante também pode ser expresso em termos das funções escalonantes da próxima escala,

$$\psi(2^j t) = \sum_k g_{j+1}(k) \phi(2^{j+1} t - k) \quad (4.17)$$

E a função  $f(t)$  pela escala j,

$$f(t) = \sum_k \lambda_j(k) \psi(2^j t - k) \quad (4.18)$$

Para uma escala j-1, a representação de  $f(t)$  precisa considerar as *wavelets*,

$$f(t) = \sum_k \lambda_{j-1}(k) \phi(2^{j-1} t - k) + \sum_k \gamma_{j-1}(k) \psi(2^{j-1} t - k) \quad (4.19)$$

Se  $\lambda$  e  $\gamma$  são ortogonais ou um quadro curto,

$$\begin{aligned} \lambda_{j-1}(k) &= \langle f(t), \phi_{j,k}(t) \rangle e \\ \gamma_{j-1}(k) &= \langle f(t), \psi_{j,k}(t) \rangle \end{aligned} \quad (4.20)$$

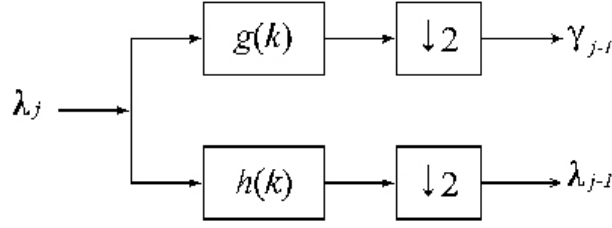


Figura 4.8: Estágio do filtro digital utilizado na decomposição

Desenvolvendo a partir das eqs. 4.16 e 4.17, obtém-se [15]:

$$\begin{aligned}\lambda_{j-1}(k) &= \sum_m h(m-2k)\lambda_j(m) \\ \gamma_{j-1}(k) &= \sum_m g(m-2k)\gamma_j(m)\end{aligned}\tag{4.21}$$

que significa que se pode encontrar os coeficientes de uma certa escala através de uma combinação dos coeficientes da escala anterior. Isso prova que se pode utilizar um banco de filtros digitais (versões discretas dos filtros da seção anterior) para a obtenção dos coeficientes da transformada, sendo que os coeficientes  $\lambda_j(k)$  vêm da parte de baixa frequência e  $\gamma_j(k)$  da parte alta. Os fatores de ponderação  $h(k)$  e  $g(k)$  formam os filtros passa baixa e alta respectivamente. Além disso, pode-se acrescentar que, devido à característica de redução da banda passante dos filtros à medida que se aumenta o nível de decomposição, também a frequência de Nyquist decresce na mesma proporção da banda, e, por isso, as operações de convolução de “h” e “g” só são calculadas para uma em duas amostras. Dessa forma, o sinal é subamostrado a cada escala, sendo, para a DWT, estabelecido um limite inferior na quantidade de escalas, limite este que é atingido quando se chega a uma amostra. Um dos estágios do filtro pode ser observado na figura 4.8.

### 4.3 FUNÇÕES-BASE

Dentre os diversos fatores que impulsionaram a difusão da teoria de *wavelets* nos últimos anos, se destacam os estudos que introduzem funções ou famílias de funções-base apropriadas para serem utilizadas como *wavelets* [19]. As funções-base utilizadas neste trabalho podem ser verificadas na Fig. 4.9. Como característica geral, pode-se verificar que todas as funções decaem rapidamente para zero. Há diferentes tipos de famílias de funções *wavelet* e suas qualidades variam de acordo com critérios como o comprimento (suporte), velocidade de convergência para zero, simetria, regularidade. O número que costuma acompanhar a designação compacta de uma wavelet (como, por exemplo, “db4”) corresponde ao número de momentos nulos (*vanishing moments*) da função.

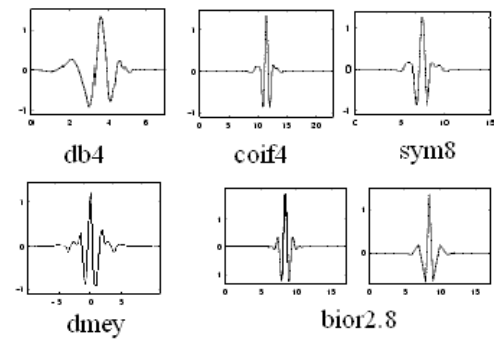


Figura 4.9: *Wavelets*-mãe utilizadas neste trabalho

## 5 DE-NOISING

*Esta seção introduz o problema de restauração de um sinal que foi submetido a uma fonte de ruído (no caso deste trabalho, ruído gaussiano).*

### 5.1 INTRODUÇÃO

O modelo básico para o sinal ruidoso é da forma,

$$r(t) = x(t) + \sigma e(t), \quad (5.1)$$

onde  $e(t)$  é o ruído gaussiano e  $\sigma$  é o nível de ruído. O objetivo do procedimento de *de-noising* é suprimir a parte ruidosa e retornar somente o termo referente ao sinal na equação 5.1. O método a ser utilizado é eficiente para famílias de funções cuja decomposição resulte em uma pequena quantidade de coeficientes *wavelet* significativos. Diz-se que tais funções possuem representação esparsa no domínio das *wavelets*. De um ponto de vista estatístico, este é um modelo de regressão no tempo e o método é uma estimação não-paramétrica da função utilizando bases ortogonais [20]. Existem duas propriedades das transformadas de *wavelets* que tornam possível a supressão de ruído nos sinais. A primeira é que apenas alguns poucos coeficientes de decomposição serão não-nulos se as funções-base forem selecionadas adequadamente de acordo com as características do sinal analisado, daí surge a alta concentração da energia nesses poucos coeficientes. A outra propriedade é que, se o sinal apresenta distribuição gaussiana, os coeficientes *wavelet* também apresentarão tal distribuição. Neste sentido, a transformada de *wavelets* de ruído gaussiano branco é ruído gaussiano branco no domínio de *wavelets* e a energia total está espalhada por todos os coeficientes da transformada do ruído. Esta diferença de amplitude torna possível uma operação de filtragem onde as componentes espectrais do sinal e do ruído podem estar superpostas em tempo e frequência, o que não é possível em métodos baseados na transformada de Fourier [18].

As etapas envolvidas no *de-noising* são:

1. Decomposição - Escolha da *wavelet*-mãe e do nível do filtro passa-baixa(e dos coeficientes) a serem utilizados na representação do sinal que será processado. O sinal  $r(t)$  é então decomposto em  $N$  níveis utilizando a *wavelet*-mãe como função-base.
2. *Threshold* - É a etapa mais importante do *de-noising*, consistindo na determinação e aplicação das funções de eliminação de coeficientes ao sinal decomposto.
3. Reconstrução - Reconstrução do sinal após a aplicação do *threshold* utilizando os coeficientes de aproximação resultantes daquela operação.

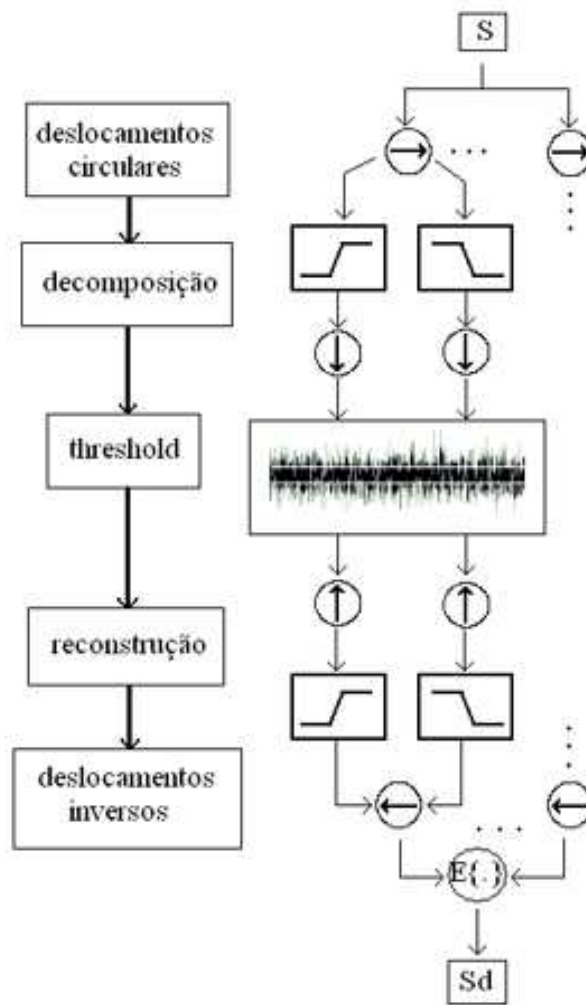


Figura 5.1: Etapas envolvidas no processo de *de-noising*

4. Deslocamentos circulares - Rotação do sinal em blocos e repetição dos procedimentos 1, 2 e 3 descritos para o sinal rotacionado sendo, ao final, invertidos os deslocamentos e calculada a média dos sinais resultantes (WT invariante ao deslocamento).

## 5.2 THRESHOLD

O *threshold* é o elemento essencial do *de-noising*, consistindo na eliminação ou suavização dos coeficientes abaixo de um certo nível a ser determinado. Esse procedimento se justifica pela característica exigida de representação esparsa de modo que a informação do sinal se encontra representada em uma pequena quantidade desses coeficientes, assim, o restante dos coeficientes corresponde a uma representação predominantemente de ruído. Como a energia do sinal é superior à energia do ruído, é de se esperar que o sinal esteja caracterizado pelos coeficientes de maior nível, podendo-se eliminar os outros. Dois esquemas



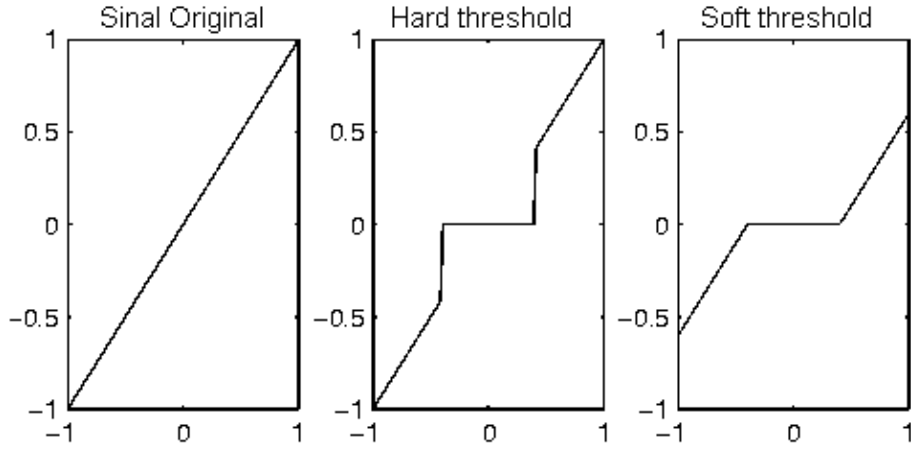


Figura 5.2: Sinal  $s = x$  submetido aos esquemas *hard* e *soft threshold*

de *threshold* se destacam no trabalho de Donoho e Johnstone [21],[22], o *hard threshold* (limiar rígido) e o *soft threshold* (limiar suave). O *hard threshold* é o esquema cuja função anula todos os coeficientes de valor inferior ao nível do *threshold* escolhido. No *soft threshold*, os coeficientes de valor inferior também são eliminados, mas os coeficientes maiores que o nível são suavizados linearmente, evitando o aparecimento de descontinuidades no sinal resultante. As funções de definição para os esquemas de *threshold* são

$$hardthreshold : x_h(t) = \begin{cases} x, & se |x| > t \\ 0, & se |x| \leq t \end{cases} \quad (5.2)$$

$$softthreshold : x_s(t) = \begin{cases} sign(x)(|x| - t), & se |x| > t \\ 0, & se |x| \leq t \end{cases}$$

A figura 5.2 ilustra a aplicação dos dois esquemas de *threshold*.

### 5.3 SELEÇÃO DO NÍVEL DO THRESHOLD

Para aplicações *off-line*, o nível do *threshold* pode ser determinado de forma empírica utilizando-se um percentual do maior coeficiente que controla progressivamente a quantidade desses coeficientes. Dessa forma, utilizaria-se um critério de avaliação da qualidade do *de-noising*, registrando o nível de *threshold* ideal. Este procedimento é bastante preciso e é aquele que permite os melhores *de-noisings*. Porém, é inviável para aplicações que necessitam do *de-noising* em tempo real.<sup>1</sup> Para este tipo de aplicação existem

<sup>1</sup>É possível, entretanto, determinar um intervalo de níveis de *threshold* para uma determinada aplicação (como no caso deste trabalho, em que este intervalo será estipulado para o *de-noising* de sinais de voz). A utilização deste intervalo de *thresholds* em conjunto com um algoritmo de seleção do nível dentro desse intervalo poderia apresentar resultados interessantes.

regras de seleção do nível do *threshold*. Entre estes, se destacam 2[20]:

- minimax - estimador que realiza o mínimo do máximo MSE (Erro Médio Quadrático) sobre um dado conjunto de funções.
- SURE - utiliza para o *soft threshold* uma regra baseada na *Stein's Unbiased Estimate of Risk* (função de perda quadrática).

O *threshold* independente do nível é calculado da observação dos coeficientes no nível de maior resolução da transformada e é geralmente usado para remover ruído branco, devido à forma de distribuição deste no domínio de *wavelets*. Se a base da transformada for apropriadamente escolhida, se observará uma maior concentração da energia do sinal nos coeficientes dos primeiros níveis, podendo-se esperar que os de maior resolução representem principalmente o ruído, motivando o cálculo do limiar no nível de maior resolução e mantendo-o constante para os outros [18].

## 5.4 FIGURAS DE MÉRITO

Os critérios utilizados na avaliação do processo de *de-noising* são essenciais na determinação dos parâmetros ótimos para a aplicação pretendida, ou mesmo para a extrapolação de parâmetros a serem utilizados em aplicações análogas. Há uma grande quantidade de medidas consolidadas na doutrina do processamento de sinais que buscam expressar a qualidade de um sinal através de um valor. Três dessas medidas serão utilizadas nesse trabalho.

### 5.4.1 Relação Sinal-Ruído (SNR)

A relação sinal-ruído, ou, simplesmente, SNR, é um conceito amplamente utilizado na teoria de comunicações e no processamento de sinais para medir a razão entre um sinal e o ruído de fundo associado ao meio. Esta medida calcula a razão logarítmica entre as energias do sinal e do ruído, permitindo uma caracterização do processo de *de-noising* através de um valor que avaliará a eficiência do processo. A relação sinal-ruído é calculada pela seguinte expressão:

$$SNR(x, x_d) = 10 \log_{10} \frac{P_x}{P_{x_d-x}}, \quad (5.3)$$

onde “ $x$ ” e “ $x_d$ ” são o sinal base e o sinal recuperado após o *de-noising*, respectivamente.  $P_x$  e  $P_{x_d-x}$  se referem à energia desses dois sinais (o base e o recuperado). Observe-se que é necessário subtrair do sinal distorcido a energia referente ao sinal original, uma vez que a relação leva em consideração apenas a energia do ruído.

#### 5.4.1.1 SSNR

Uma variação possível no cálculo da relação sinal-ruído quando se procede com a análise segmentada de sinais é a Relação Sinal-Ruído Segmentar (*Segmental Signal-to-Noise Ratio*), na qual se calcula a relação sinal ruído para cada segmento do sinal original analisado e, ao final, calcula-se a média do grupo. A equação 5.4 expressa o cálculo da SSNR:

$$SSNR(x, x_d) = \sum_i SNR(x_i, x_{d_i}) \quad (5.4)$$

Uma discussão importante se faz necessária na utilização dessa medida. Uma vez que o logaritmo é calculado antes do cálculo da média, essa medida expressa uma “média de produtos”, que só deve ser comparada com relações que utilizem a mesma forma de cálculo. A medida é útil na verificação da variação da qualidade do *de-noising* para os diferentes segmentos analisados no sinal.

#### 5.4.2 MSE

Na estatística, o erro médio quadrático ou MSE de um estimador é o valor esperado do quadrado do erro. O erro é a quantidade pela qual o estimador difere da quantidade a ser estimada. A diferença ocorre devido à aleatoriedade ou porque o estimador não contabiliza informação que poderia produzir uma estimativa mais acurada [23]. Neste trabalho o MSE para um sinal é definido através da seguinte expressão:

$$MSE(x(t)) = \frac{\sum_{t=0}^T x(t)^2}{x(0) * x(T)} \quad (5.5)$$

Calculado desta forma, o MSE permite a associação de um valor ao sinal representado, de tal modo que é possível a utilização do mesmo como figura de mérito na comparação com outros sinais. Para o *de-noising*, o valor em que se está interessado se refere à quantidade de erro remanescente no sinal submetido ao *de-noising* “ $x_d(t)$ ”. Além disso, este valor deve ser normalizado utilizando-se o MSE do próprio sinal, com o intuito de permitir a comparação entre diferentes processos. Tudo isso pode ser resumido na seguinte expressão:

$$MSE_N = \frac{MSE(x_d(t) - x(t))}{MSE(x(t))} \quad (5.6)$$

#### 5.4.3 PESQ

Para muitas aplicações, a medição da qualidade de voz pelas técnicas clássicas, que utilizam conceitos como SNR ou funções de resposta em frequência, não são suficientes para obter um resultado que possa

assegurar uma boa percepção subjetiva. Um modelo baseado na percepção humana vem se consolidando na última década. Nesse sentido evolui-se até o atual algoritmo PESQ (*Perceptual Evaluation of Speech Quality*) - Avaliação Perceptiva de Qualidade da Voz. Esse algoritmo encontrou uma excelente aceitabilidade na doutrina do processamento de sinais para a avaliação de qualidade de voz desde sua padronização pelo ITU-T (*International Telecommunication Union*) como recomendação P.862. O algoritmo basicamente mapeia os sinais base e degradado em uma representação interna usando um modelo perceptivo e a diferença entre essas representações é usada num modelo cognitivo para prever a qualidade do sinal degradado em termos da pontuação MOS (*Mean Opinion Score*) - Pontuação Média de Opinião - (a maioria dos experimentos subjetivos usa a escala ACR de opinião). A escala ACR (*Absolute Category Rating*) - Avaliação de Categoria Absoluta - usa valores de pontuação de 1 a 5 conforme a tabela 5.1. As representações internas usadas com o modelo cognitivo de predição são calculadas com base em representações de sinal que usam equivalentes psico-acústicos de frequência e intensidade [24]. A figura 5.3 ilustra, em um diagrama de blocos, a implementação básica do algoritmo PESQ.

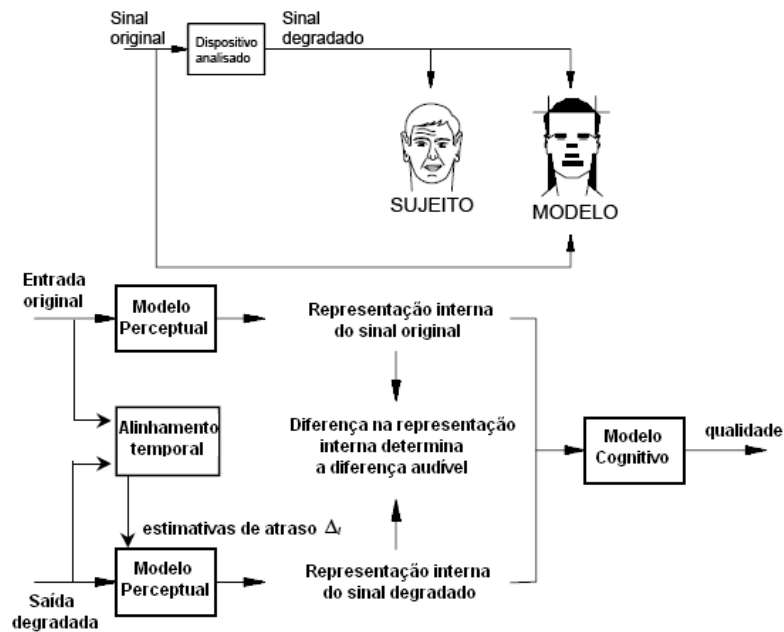


Figura 5.3: Diagrama algoritmo do PESQ

Tabela 5.1: Escala de qualidade de escuta ACR usada no desenvolvimento do PESQ

<i>Qualidade da voz</i>	<i>Pontuação</i>
Excelente	5
Boa	4
Média	3
Pobre	2
Ruim	1

## 6 METODOLOGIA

*Este capítulo descreve a metodologia seguida no estudo da DWT para a minimização de ruído aplicada a sinais de voz.*

### 6.1 INTRODUÇÃO

O projeto foi iniciado pelo estudo do som e sua geração, juntamente com as distorções a que este pode ser submetido, evoluindo até a análise do ruído nos sinais de áudio e a utilização da DWT como ferramenta para o *de-noising*. Foi definido um escopo para o projeto, que consistiu da análise de um banco de sinais de 22kHz, constituído por várias gravações de voz com diferentes locutores. Houve a proposta de utilização da eliminação de ruído como aplicação complementar a um sistema de reconhecimento de locutor. Passada a etapa de pesquisa, foram definidos os procedimentos seguidos e as ferramentas a serem utilizadas. Segue a descrição das etapas envolvidas no projeto:

- Aquisição do banco de sinais (considerados desprovidos de ruído) e escolha de uma quantidade suficiente de dados para a análise.
- Retirada de amostra reduzida dos sinais escolhidos para permitir a análise de parâmetros que serão utilizados nos sinais originais.
- Extrapolação com teste dos resultados obtidos na etapa anterior para os sinais inteiros.
- Aplicação dos parâmetros obtidos para o *de-noising* de sinais diferentes dos analisados. Proposta de análise com diferentes níveis de ruído.

A determinação de parâmetros como a função-base mais adequada à representação do sinal ou a quantidade de deslocamentos a serem utilizados e o tamanho da janela de corte utilizada para a decomposição é essencial para a implementação de um algoritmo eficiente no *de-noising* podendo, para o caso de vozes, permitir uma definição de uma configuração ideal razoavelmente generalizada.

### 6.2 DADOS

O banco de sinais utilizado consiste de uma grande quantidade de gravações de diferentes locutores (homens e mulheres) dos quais foram selecionados 10 sinais (5 de cada sexo) para o estudo neste trabalho. O critério de escolha foi aleatório, sem escuta prévia das gravações em questão. Os sinais foram originalmente gravados em formato WAV quantizados com 16 bits e amostrados a 22050 Hz. Para cada sinal

foi retirada aleatoriamente uma amostra de 10%(dez por cento) da duração para a determinação inicial de parâmetros. Posteriormente, os sinais completos foram utilizados. Para a verificação dos parâmetros foi escolhido um sinal extra do mesmo banco. Também foi utilizado um banco de sinais de 8kHz para a verificação de adequação dos parâmetros com uma taxa de amostragem diferente e objetivando a avaliação de qualidade do *de-noising* utilizando um algoritmo que vem se consolidando como um padrão nesse sentido, o PESQ (ITU-T P.862).

### 6.3 FERRAMENTAS COMPUTACIONAIS

Todo o procedimento foi realizado em computadores Pentium IV 3.2 e Athlon 2.6 com 512MB de memória RAM e utilizando a versão 2006 do software MATLAB(Mathworks), que serviu como ambiente de desenvolvimento dos algoritmos envolvidos no projeto. A escolha foi motivada pela já existência de primitivas pré-implementadas para a manipulação de wavelets e demais ferramentas de processamento de sinais.

### 6.4 TÉCNICAS UTILIZADAS

Diversas funções foram desenvolvidas neste projeto para a implementação das técnicas de minimização de ruído pretendidas. Segue uma descrição das técnicas, procedimentos e implementações:

#### 6.4.1 Adição de ruído branco

A adição de ruído branco de energia controlada é necessária para a determinação de parâmetros para o *de-noising* em diversos níveis de distorção de sinais, uma vez que o foco do projeto é a minimização de ruído branco, este algoritmo é específico para a adição deste tipo de ruído aos sinais. Devido à sua característica de extensão em todo o espectro de frequência, o ruído branco é muito utilizado no estudo prévio do projeto de sistemas robustos [19]. O parâmetro de controle do nível de ruído é a relação Sinal/Ruído em dB que se pretende obter no sinal ruidoso dada pela equação 5.1. Os parâmetros de entrada para a função são o sinal original e a relação sinal-ruído (SNR). Para obter a energia do sinal ruidoso com SNR dB, calcula-se a energia do sinal original e adiciona-se a cada ponto do sinal uma quantidade aleatória de ruído relacionada à energia obtida.

$$x_r = \sum_i x_i + n_i \sqrt{P_n} \quad (6.1)$$

Sendo  $x_i$  os vários pontos do sinal original,  $n_i$  um número aleatório no intervalo  $[0,1]$  e  $x_r$  o sinal com o ruído inserido.  $P_n$  (equação 6.3) se refere à energia do ruído, que será acrescida ao sinal analisado e é obtida a partir da própria energia do sinal (equação 6.2):

$$P_s = \frac{1}{N} \sum_{i=1}^N x_i(t)^2 \quad (6.2)$$

$$P_n = \frac{P_s}{10^{SNR/10}} \quad (6.3)$$

#### 6.4.2 De-noising Wavelet invariante ao deslocamento

A técnica TIWD (*Translation Invariant Wavelet De-Noising*) - *De-Noising Wavelet Invariante ao deslocamento* - concebida por Donoho e Coifman [25] foi a escolhida como base para aplicação no processo de *de-noising* realizado neste projeto. A minimização de ruídos utilizando a técnica TIWD envolve uma série de etapas: deslocamentos, decomposição Wavelet, *threshold*, reconstrução, inversão dos deslocamentos, média dos sinais resultantes.

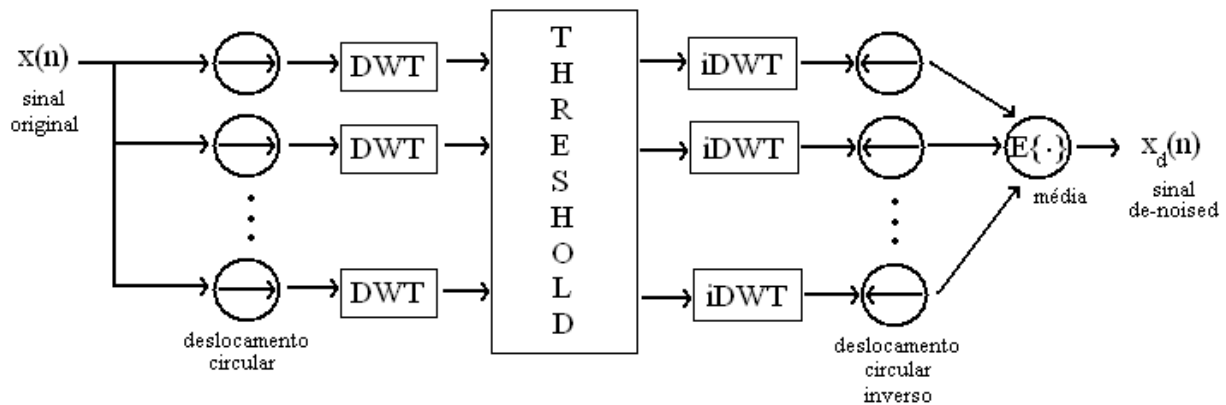


Figura 6.1: Função TIWD

A figura 6.1 apresenta um diagrama ilustrativo da aplicação da técnica TIWD, cujos elementos serão analisados no restante desta seção.

##### 6.4.2.1 Deslocamentos circulares

A TIWD foi desenvolvida como um refinamento dos processos de *de-noising* diretos com a Transformada de *Wavelets* com o intuito de aumentar a eficiência de tais processos imunizando-os contra um fenômeno intrínseco à realização da etapa de *threshold*, o pseudo-fenômeno de Gibbs. Para isso, o sinal é deslocado em blocos (compostos de uma determinada quantidade de amostras), decomposto, submetido ao esquema determinado de *threshold*, reconstruído e restaurado (inversão dos deslocamentos) para cada uma

das configurações submetidas, sendo o sinal resultante calculado pela média simples dos sinais obtidos para cada uma dessas configurações. Este tipo de alteração da configuração do sinal é chamado de deslocamento circular. Na sua implementação utiliza-se um parâmetro de ajuste da quantidade de deslocamentos a serem realizados objetivando-se a obtenção de um valor ótimo. Observe-se que este processo é iterativo, de modo a permitir que se proceda com quantos deslocamentos forem necessários para a melhor atenuação. A configuração do bloco de deslocamento (quantidade de amostras) é ajustada de acordo com a quantidade de deslocamentos. Os vários sinais resultantes obtidos pela aplicação da técnica TIWD serão utilizados para o cálculo de um valor médio conforme descrito na equação 6.4 a seguir:

$$x'_d = \frac{1}{N} \sum_i^N x_{d_i} \quad (6.4)$$

Uma observação importante se refere ao intervalo em que é limitada a quantidade de deslocamentos. Se a quantidade de deslocamentos for nula, tem-se a hipótese da técnica clássica de *de-noising wavelet* susceptível ao pseudo-fenômeno de Gibbs. Este seria o limite inferior do parâmetro. Também, o máximo de deslocamentos relevantes equivale a uma escolha de bloco com o tamanho de apenas uma amostra. Se procedendo desta forma, seria possível realizar a análise com todas as configurações de sinais possíveis para um deslocamento circular. Com base em ambas as ressalvas anteriores, pode-se determinar um intervalo para a quantidade de deslocamentos  $D = [0, N_a]$ , onde  $N_a$  é a quantidade de amostras presentes no sinal. Uma das vantagens do TIWD é que esta técnica aproveita a característica de ortogonalidade (entre si) do ruído branco, permitindo, num caso extremo (por exemplo, um sinal contínuo e infinitos deslocamentos), a completa eliminação do ruído. Na prática, busca-se tornar o número de deslocamentos máximo, limitado ao tamanho da janela de análise. Um fator de ponderação que deve ser considerado na utilização da técnica é o aumento no tempo de processamento, diretamente proporcional à redução no tamanho do bloco. Isso significa que operando-se nos dois extremos de  $D$ , obteria-se, respectivamente, o menor e o maior tempo de processamento para o caso.

#### 6.4.2.2 Decomposição Wavelet

A decomposição em coeficientes *wavelet* é conseguida através de processos de filtragem e sub-amostragem (*downsampling*) em múltiplos níveis, como já explicado nos capítulos 4 e 5. A decomposição realiza a especificação dos coeficientes de aproximação e detalhes para os diversos níveis, preservando a informação da configuração do sinal original antes da aplicação de cada etapa de filtragem (quantidade de amostras resultantes da aplicação das etapas de filtragem). Ambas informações são armazenadas em vetores que serão utilizados tanto pela etapa de *threshold* quanto pela reconstrução para o domínio de tempo.



### 6.4.2.3 Threshold

#### **Cálculo do nível de *threshold***

O escopo deste trabalho se limita a aplicações *off-line*, com pré-processamento dos sinais de áudio que serão inseridos em um sistema de reconhecimento de locutor. Nesse sentido, e como já discutido anteriormente (capítulo 5), a determinação do nível de *threshold* se dá de forma empírica através de um parâmetro de percentual que é variado nas diversas iterações da aplicação da técnica TIWD. O parâmetro é calculado a partir de um percentual do maior elemento do vetor de coeficientes gerado pela etapa de decomposição. O *threshold* é calculado conforme a seguinte regra:

$$t = \frac{p\%}{100} \max |C|, \quad (6.5)$$

onde  $t$  é o nível de *threshold* utilizado,  $C$  é o vetor de coeficientes da decomposição e  $p\%$  é o percentual do maior valor de  $C$  que se pretende usar como nível de *threshold*. Assim, é possível obter-se uma seletividade tão rigorosa quanto se deseje, ou, até mesmo proceder em uma verificação empírica do percentual ótimo para o conjunto de parâmetros analisado. Apesar de haver sido considerada isoladamente nesta seção, a estimação do nível de *threshold* não pode ser analisada isolada do esquema de *threshold* em si, uma vez que a combinação de ambas é o principal elemento do *de-noising*, e a sua perfeita determinação é responsável pela qualidade deste.

#### **Funções de *threshold***

Conforme já discutido em seções anteriores, o *threshold* é o bloco que seleciona e/ou aproxima os coeficientes a serem utilizados na recomposição do sinal  $X$ . Pode ser entendido como um estágio de filtro no domínio de *Wavelets* ao qual se submeterá o sinal transformado de modo a retirar dele a informação indesejada e/ou redundante. As funções implementadas são as mesmas apresentadas no capítulo de *de-noising* (equação 5.2). A implementação da técnica TIWD pode ser realizada com um dos dois esquemas de *threshold*, a depender da aplicação pretendida. A função de *threshold* definida será aplicada ao vetor de coeficientes *wavelet* obtido na decomposição de acordo com o limiar pretendido. Esta operação resultará em um vetor de coeficientes alterado (ou *thresholded*) que será utilizado na reconstrução do sinal.

### 6.4.3 Reconstrução Wavelet

A síntese do sinal a partir dos coeficientes da DWT submetidos ao *threshold* é conseguida através de um segundo processo de filtragem e re-amostragem (sobre-amostragem, *upsampling*). Para isso, são

utilizados dois vetores, o vetor de coeficientes alterado pela aplicação do *threshold* e o vetor de controle da quantidade de amostras resultantes de cada filtro. Nível a nível, refaz-se a recomposição dos coeficientes de aproximação e detalhes, até que se obtenha o novo sinal no “domínio do tempo”. Pode-se verificar que o procedimento para esta etapa é o inverso do procedimento de decomposição, com a diferença de estar-se reconstruindo o sinal com os coeficientes *wavelet* alterados pelo *threshold*. Nessa reconstrução, a depender da família de funções-base escolhida e, principalmente, do fato da função-base se comportar como uma base ortonormal ou ortogonal, se realizará a transformada inversa com a mesma função-base (caso ortonormal) ou com um par ortogonal<sup>1</sup>. O sinal resultante desta etapa é uma das configurações de deslocamento circular que serão utilizadas para a obtenção do sinal com ruído minimizado (*de-noised*).

#### 6.4.4 Manipulações da técnica TIWD

Novamente, no caso do processamento *off-line* do sinal e, também, como estudo prévio na configuração de um sistema em tempo real, é possível utilizar métodos iterativos para a verificação empírica dos melhores parâmetros para a técnica TIWD. Por isso, manipulou-se o algoritmo que a implementava no intuito de variar os parâmetros envolvidos no *de-noising* e acrescentar outros aspectos experimentais à análise. O diagrama da figura 6.2 ilustra a implementação da técnica TIWD na determinação de parâmetros ótimos para o *de-noising*. Um dos aspectos relevantes é a necessidade de corte do sinal em partes de

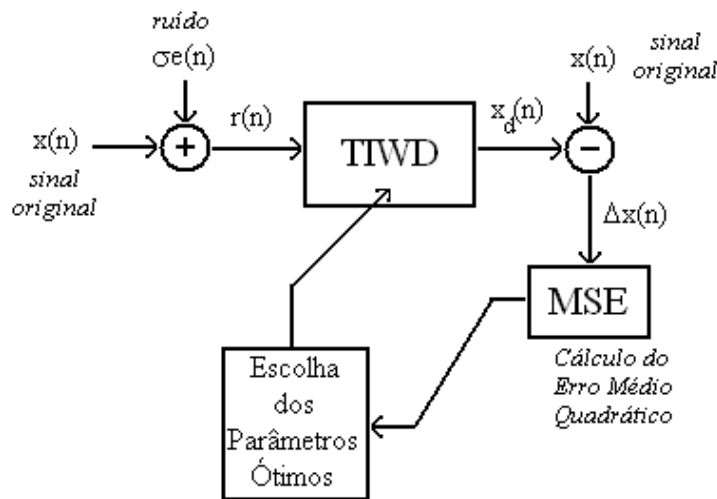


Figura 6.2: Processo de Escolha de Parâmetros Ótimos para o *De-Noising*

mesmo tamanho para proceder com o *de-noising* nas frações do sinal. O fracionamento do sinal permite, dentro de um certo nível de limitação, aproximar um aspecto de estacionariedade para um sinal que é intrinsecamente não-estacionário. A maior estacionariedade do sinal permite uma melhor caracterização do

<sup>1</sup>Este aspecto afeta a configuração dos filtros utilizados na síntese. Caso se trate de um par ortogonal, diferentes configurações serão utilizadas em relação à decomposição.

mesmo pela transformada e simplifica a análise realizada. Tal procedimento também é importante num contexto de transmissão de dados, de modo a simular uma rotina de recepção, *de-noising* e transmissão para um sistema que funcione através da transmissão de pacotes de tamanho limitado. Uma observação interessante se refere à verificação do efeito do fracionamento do sinal no processo de *de-noising* e na preservação da integridade das informações do sinal, uma vez que um fracionamento excessivo (ignorando a taxa de Nyquist) poderia gerar redução da banda do sinal no trecho sob análise, além de incorrer em uma seletibilidade restritiva de acordo com o princípio da incerteza de Heisenberg. Desta forma, o algoritmo que implementa a técnica TIWD permite a seletividade de um tamanho de janela de corte do sinal. O corte do sinal é realizado resultando em frações de mesmo tamanho, a não ser em relação ao último segmento, cujo tamanho é equivalente ao resto da divisão do tamanho do sinal pelo número de janelas. O algoritmo de manipulação também calcula os erros médios quadráticos (MSE) (seção 5.4.2), a relação sinal-ruído (SNR) e a relação sinal-ruído segmentar (SSNR) (seção 5.4.1), associados a cada conjunto de parâmetros utilizados no processo de denoising de modo a selecionar o melhor resultado com base em tais parâmetros objetivos (utilizados como figuras de mérito na análise). Além disso, para o banco de sinais de 8kHz foi implementada uma versão do algoritmo que avalia o *de-noising* com o algoritmo PESQ (seção 5.4.3).

## 7 RESULTADOS EXPERIMENTAIS

*Este capítulo descreve os experimentos realizados e apresenta, em forma de tabelas e gráficos, os resultados obtidos nos mesmos.*

### 7.1 INTRODUÇÃO

O nível de ruído ao qual se deu maior ênfase neste projeto foi de 10dB. O estudo abrangeu exclusivamente o ruído gaussiano, apesar da utilização da transformada invariante ao deslocamento para atenuação do pseudo-fenômeno de Gibbs. Esse nível de ruído gaussiano foi adicionado a cada sinal inteiro, sendo as amostras retiradas dos mesmos trechos dos sinais originais e ruidosos. A técnica TIWD foi aplicada com variação de cinco parâmetros no algoritmo de implementação, o nível de resolução (decomposição), a quantidade de deslocamentos, a *wavelet*-mãe e o nível e função de *threshold*. Disto, iniciou-se a escolha empírica dos seguintes intervalos dos parâmetros.

- *wavelets*-mãe: daubechies 4, coifflets 4, symlets 8, discrete meyer e biortogonal 2.8. A escolha dessas funções foi motivada por haverem apresentado resultados relativamente superiores em simulações envolvendo a quase totalidade das *wavelets*-mãe passíveis de utilização com a DWT implementadas no Matlab.
- Nível de resolução: 4 e 8. Da mesma forma que a escolha das *wavelets*-mãe, a escolha dos níveis de decomposição foi motivada por testes com uma maior quantidade desses níveis. Para níveis de resolução superiores a 8 não se justifica o maior esforço computacional por ser o ganho de qualidade do *de-noising* muito reduzido. Preferiu-se utilizar índices que sejam potências de dois devido ao paradigma binário da computação.
- A quantidade de deslocamentos circulares foi escolhida inicialmente como 512 por haver-se constatado empiricamente que esta quantidade seria suficiente para o teste das amostras.
- Os dois esquemas de *threshold* (*hard* e *soft*) foram analisados e o nível do *threshold* foi variado de 0 a 100% do valor do maior coeficiente até que se encontrasse a melhor condição de *de-noising* para o conjunto do restante de parâmetros.

As amostras analisadas possuem 10% da duração dos sinais originais, o que significa uma duração média de 3,5s para as mesmas. Esse tamanho de amostra também foi determinado empiricamente após uma ampla quantidade de testes e ponderações com vários tamanhos dessas amostras. Pode-se observar que neste tamanho de trecho é possível captar uma grande quantidade de palavras (e silêncios entre estas)

de cada locutor, além de preservar características como a modelagem espectral do trato vocal (que pode ser preservada com um trecho curto), o tom (*pitch*) da voz falada e o ritmo de pronúncia das palavras, que são aspectos peculiares da fala e particulares para cada pessoa. A aleatoriedade foi provida para que os parâmetros obtidos não se tornassem viciados em relação a um texto específico falado por todos os locutores e assim tivessem maior potencial de generalização para os vários aparelhos fonadores. Um comentário importante é que os sinais amostrados foram decompostos por completo para a análise, não havendo janelamento, que será analisado na próxima etapa (ou seja, se a amostra tinha 3,5s, todo esse trecho foi analisado pela mesma decomposição). Após a obtenção dos parâmetros ótimos com o algoritmo descrito acima, passou-se ao estudo do sinal (ainda o amostrado e com os melhores parâmetros obtidos acima) segmentado em janelas de 5 tamanhos diferentes (inteira, 1/2, 1/4, 1/8, 1/16) em relação às durações das amostras. Além disso, foram variadas as quantidades de deslocamentos para cada um dos tamanhos de janela de modo a verificar estes outros dois parâmetros (tamanho da janela e quantidade de deslocamentos) para aplicação no sinal inteiro. Por fim, a totalidade de resultados culmina na análise dos 10 sinais inteiros originais para os quais foi verificada a possibilidade de extrapolação dos parâmetros determinados nas etapas anteriores com as duas melhores *wavelets*-mãe. Com o intuito de checar a validade dos resultados obtidos, foram utilizados os parâmetros obtidos para a análise do restante dos sinais do banco de 22kHz. Seguindo-se a esta verificação, a escolha de um 11º sinal que será analisado para vários níveis de ruído maiores e menores do que os 10dB tomados como base para este trabalho com o intuito de verificar a validade dos parâmetros determinados em situações diversas às estudadas.

Também como proposta de análise por um padrão novo e que vem se consolidando recentemente, o PESQ, utilizou-se o banco de sinais de 8kHz e foram refeitos os testes da etapa anterior com o intuito de verificar a variação dos parâmetros anteriores para uma diferente taxa de amostragem e com um critério de avaliação adicional. Este banco consiste de 11 sinais, 6 em vozes femininas e 5 masculinas, dos quais foram analisados 6 (3 masculinos e 3 femininos). A qualidade das gravações é bem inferior à anterior, uma vez que se utiliza uma banda passante de 4kHz. A quantidade de ruído inserida também se limitou aos 10dB de modo a permitir a comparação com a análise anterior.

## 7.2 ANÁLISE DO BANCO DE 22KHZ

Foram utilizados 10 sinais do banco de 22kHz nos quais foi gravada, por cada um dos locutores, a leitura das 10 frases a seguir:

“A questão foi retomada no Congresso. Leila tem um lindo jardim. O analfabetismo é a vergonha do país. A casa foi vendida sem pressa. Trabalhando com união rende muito mais.

Tabela 7.1: Listagem dos arquivos utilizados na análise das vozes em 22kHz

grupo	sinal	duração
<b>vozes femininas</b>	lf501ls11.wav	32s
	lf502ls11.wav	30s
	lf503ls11.wav	29s
	lf504ls11.wav	27s
	lf505ls11.wav	24s
<b>vozes masculinas</b>	lm101ls11.wav	27s
	lm102ls11.wav	31s
	lm103ls11.wav	27s
	lm104ls11.wav	25s
	lm105ls11.wav	31s

Recebi nosso amigo para jantar. A justiça é a única vencedora. Isso se resolverá de forma tranqüila. Os pesquisadores acreditam nesta teoria. Sei que atingiremos o objetivo.”

A tabela 7.1 apresenta a listagem dos arquivos utilizados juntamente com as durações em segundos dos mesmos.

## 7.2.1 Análise das amostras

Conforme já explanado, foram retiradas amostras aleatórias com 10% da duração dos sinais da tabela 7.1 para avaliação de parâmetros. Esta avaliação consiste em duas etapas, a primeira avalia a melhor função-base e níveis de decomposição, e a segunda avalia a quantidade de deslocamentos e o tamanho da janela de corte.

### 7.2.1.1 Escolha de Funções-Base

As tabelas III.1 a III.10 expõem os resultados obtidos nas simulações com as amostras dos respectivos sinais. A figura 7.1 ilustra as curvas do MSE para vários valores de percentual na análise do arquivo amostra\_lf501ls11. Cada uma das curvas se refere a uma função-base. Pode-se observar, para todas as tabelas, que a decomposição em 8 níveis apresenta resultado superior em comparação à decomposição em 4 níveis, usando como parâmetro o MSE. Isto ocorre porque a quantidade de coeficientes de detalhamento obtidos com 8 níveis é maior do que em 4, havendo uma melhor descrição do sinal por estes coeficientes de 8 níveis.<sup>1</sup> Uma comparação subjetiva ouvindo os dois sinais resultantes não permite uma diferenciação de

<sup>1</sup>Este tipo de resultado levaria à dedução de que uma decomposição em mais níveis do que 8 deveria ser utilizada. O problema é que, com o aumento da quantidade de níveis, há maior distribuição do sinal e do ruído pelos coeficientes da transformada e isso pode tornar o processo de *de-noising* menos eficiente. (Pelo menos do ponto de vista do MSE)

um melhor dentre os dois. Uma vez que não foi possível uma escolha ponderada com o critério subjetivo, optou-se exclusivamente pelo critério objetivo, no qual a maior decomposição prevalece. Ainda em relação a esta etapa, a escolha das *wavelets*-mãe foi feita estritamente pelas avaliações objetivas. Uma vez que não se pôde realizar um teste de auditório ou algum tipo de exposição a mais de um ouvinte, optou-se por não ponderar a escolha da *wavelet*-mãe pela escuta das amostras resultantes do processo de *de-noising*, sendo a avaliação subjetiva utilizada apenas para a percepção de divergências crônicas e eventuais inconsistências. Para a escolha das *wavelets*-mãe, foi atribuída uma nota 1 para a melhor *wavelet*-mãe em cada caso, sendo, para cada uma das outras restantes, atribuído o valor da razão entre a sua SNR e a da melhor. Estas notas foram agrupadas através de uma média do desempenho no grupo, sendo escolhidas as duas melhores médias para cada esquema de *threshold* e para cada grupo. Este procedimento pode ser observado na tabela 7.2, onde se pode verificar que os melhores desempenhos para o caso feminino foram, para o *hard threshold*, com as Symlets 8 e Biortogonais 2.8 e, para o *soft threshold*, com as Discrete Meyer e Biortogonais 2.8. Já no caso masculino, para o *hard threshold*, as Coifflets 4 e as Symlets 8 e, para o *soft threshold*, as Biortogonais 2.8 e as Discrete Meyer apresentaram os melhores resultados.

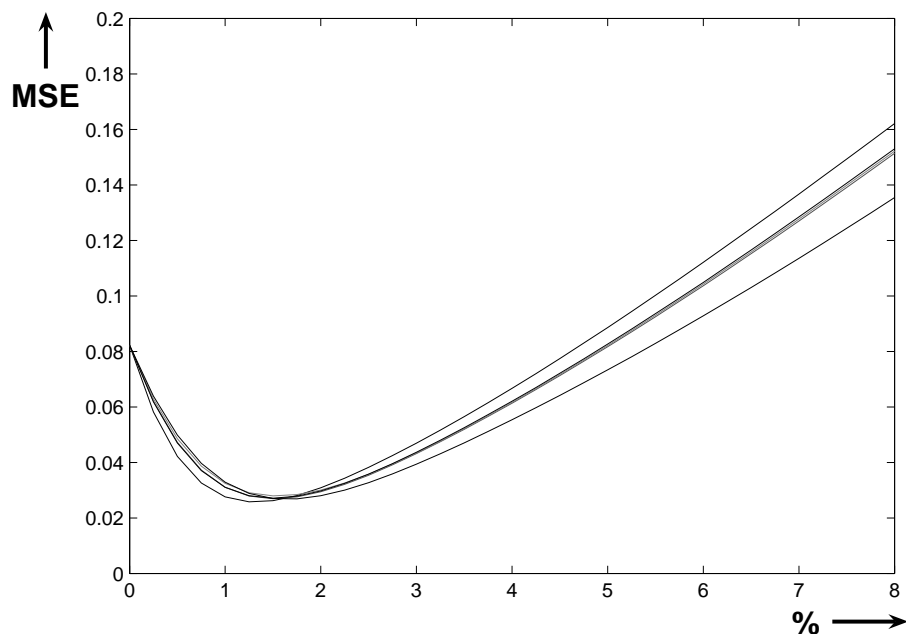


Figura 7.1: Curvas de MSE para a escolha de *wavelets*-mãe

#### 7.2.1.2 Verificação do Número de Deslocamentos e Tamanho da Janela de Corte

Os resultados dos testes para a verificação da quantidade ótima de deslocamentos e para o melhor tamanho da janela de corte encontram-se nas tabelas III.11 a III.20. Observando-se que só foi utilizado

Tabela 7.2: Escolha da melhor *Wavelet*-mãe dentre as amostras analisadas

femininas												
base	1		2		3		4		5		médias	
	h	s	h	s	h	s	h	s	h	s	h	s
<b>db4</b>	0,921	0,975	0,986	0,977	0,995	0,976	0,988	0,979	0,981	0,968	0,974	0,975
<b>coif4</b>	0,932	0,987	0,994	0,984	1,000	0,985	1,000	0,988	1,000	0,981	0,985	0,985
<b>sym8</b>	0,993	0,984	0,995	0,984	1,000	0,984	1,000	0,986	0,999	0,980	<b>0,997</b>	0,984
<b>dmey</b>	0,943	0,998	0,992	1,000	0,990	1,000	0,989	1,000	0,999	1,000	0,983	<b>1,000</b>
<b>bior2.8</b>	1,000	1,000	1,000	0,996	0,991	0,990	0,982	0,984	0,986	0,982	<b>0,992</b>	<b>0,991</b>
masculinas												
base	1		2		3		4		5		médias	
	h	s	h	s	h	s	h	s	h	s	h	s
<b>db4</b>	0,982	0,980	0,996	0,979	1,000	0,983	0,986	0,976	0,995	0,980	0,992	0,980
<b>coif4</b>	0,998	0,991	0,997	0,983	1,000	0,989	0,995	0,986	1,000	0,988	<b>0,998</b>	0,987
<b>sym8</b>	0,998	0,990	0,996	0,983	1,000	0,988	0,995	0,985	0,999	0,987	<b>0,997</b>	0,987
<b>dmey</b>	1,000	0,993	0,985	0,991	0,994	1,000	0,987	0,999	0,999	1,000	0,993	<b>0,997</b>
<b>bior2.8</b>	0,989	1,000	1,000	1,000	0,995	0,997	1,000	1,000	0,990	0,998	0,995	<b>0,999</b>

o esquema de *hard threshold*, por ser este o que possibilita melhor resultado em relação à avaliação pelo menor MSE e por ser possível compatibilizar os dois parâmetros analisados nesta etapa para o esquema de *soft threshold*. Os resultados foram agrupados na tabela 7.3 de modo a permitir uma análise do conjunto. O objetivo desta análise não era uma verificação precisa de um tamanho de janela específico e melhor quantidade de deslocamentos para cada sinal, mas sim uma verificação de valores aproximados para estes dois parâmetros que houvessem permitido uma análise razoável para cada sinal. O tamanho da janela de corte será escolhido como uma potência de dois, seguindo o modelo de raciocínio do número de deslocamentos. Uma questão adicional a ser colocada é em relação aos tempos de execução para cada iteração em relação ao tamanho de janela e quantidade de deslocamentos utilizados. A tabela 7.4 apresenta a média dos tempos por iteração para o algoritmo utilizado. Pode-se, desta forma, observar que uma grande quantidade de deslocamentos compromete por demais a velocidade de execução do algoritmo, da mesma forma, uma escolha de janela inadequada também aumenta o tempo de processamento. Em relação à quantidade de deslocamentos, pôde-se verificar pelos dados apresentados que a utilização de uma janela de 1/8 do tamanho da amostra resultou suficiente e eficientemente na maioria dos casos. Além disso, os resultados com 64 deslocamentos se destacaram tanto pela velocidade de execução quanto pela eficiência no *de-noising*. Para uma generalização, utilizar-se-á um tamanho de janela de 16384 amostras ( $2^{14}$ ) combinada a uma quantidade de 64 deslocamentos. A análise da extrapolação para os sinais inteiros, porém, ainda utilizará os valores ótimos de cada amostra, sendo a generalização deixada para os sinais de teste.



Tabela 7.3: Resumo das análises de quantidade de deslocamentos e tamanho da janela de corte

<b>Sinal</b>	<b>w-mãe</b>	<b>melhor tam. janela</b>	<b>melhor qtd deslocs</b>
<b>lf501ls11</b>	bior2.8	17891	64
<b>lf502ls11</b>	bior2.8	16622	64
<b>lf503ls11</b>	coif4	16130	512
<b>lf504ls11</b>	coif4	15178	512
<b>lf505ls11</b>	coif4	13258	512
<b>lm101ls11</b>	dmey	29768	64
<b>lm102ls11</b>	bior2.8	17365	64
<b>lm103ls11</b>	db4	15160	256
<b>lm104ls11</b>	bior2.8	14057	64
<b>lm105ls11</b>	coif4	34343	512
<b>média</b>		18977	262,4

Tabela 7.4: Tempos de execução em segundos para o teste de deslocamentos e janelas

<b>qtd deslocs</b>	<b>tamanho da janela</b>				
	<b>4473</b>	<b>8946</b>	<b>17891</b>	<b>35782</b>	<b>71563</b>
1	0,76	0,47	0,32	0,27	0,23
4	2,82	1,67	1,12	0,88	0,73
16	11,08	6,48	4,22	3,19	2,74
64	44,06	25,71	16,72	12,52	10,74
256	199,70	102,75	66,88	50,79	42,87
512	351,92	205,57	133,27	102,35	85,57

## 7.2.2 Análise dos sinais inteiros

Para a análise dos sinais inteiros, de modo a permitir a verificação dos parâmetros obtidos, estes foram aplicados utilizando-se as duas melhores *wavelets*-mãe obtidas com cada amostra. A tabela 7.5 apresenta a avaliação do *de-noising* de todo o conjunto de sinais analisados. Para a avaliação da extrapolação, foram elaboradas as tabelas 7.6 e 7.7 na qual se atribui uma diferença entre os parâmetros obtidos nas duas situações (amostra e sinal inteiro) de modo a quantificar o erro na extrapolação. Pela análise das tabelas em questão, pode-se verificar que a extrapolação dos parâmetros permitiu a obtenção de uma margem aceitável de precisão na análise dos sinais inteiros, enfatizando que o objetivo final do processo de análise das amostras era a obtenção de parâmetros para o *de-noising* dos sinais inteiros sem que fosse necessário realizar os testes com os sinais inteiro propriamente (Um procedimento inviável). A tabela 7.8 aponta os valores médios para o MSE e a SNRt, além dos intervalos de percentuais observados. Também se estende a análise para os dois grupos, onde pode-se observar que, em média, o procedimento de *de-noising* foi mais eficiente no grupo masculino, resultando em um valor médio de MSE inferior ao caso feminino. Este tipo de conclusão é consolidada no processamento de áudio relacionado à voz. Observa-se que a voz feminina possui uma maior variação de afinação (*pitch*) em relação à masculina, tornando mais difícil uma generalização. Também é possível, por meio desta análise, observar que os valores de *threshold* ótimos se encontram nos intervalos aproximados de (1,25 ; 2,25) para o *soft threshold* e (3,25 ; 5,5) para o *hard threshold*. Esta última constatação é particularmente útil para a agilização do desenvolvimento de algoritmos para o *de-noising*, podendo-se supor que quaisquer limiares escolhidos dentro desses intervalos permitiriam um *de-noising*, se não ótimo, ao menos adequado. Pode-se observar, na figura 7.2 as formas de onda do sinal lm101ls11.wav original, com ruído e submetido à técnica TIWD.

A avaliação subjetiva (audição dos sinais) dos *de-noisings* permite algumas observações importantes. Uma vez que foi utilizada uma janela de tamanho inferior à duração do sinal, e, aliado ao fato de que os sinais constam de uma seqüência de frases alternada por uma série de pausas na leitura, pode-se perceber que nos momentos de silêncio o ruído branco continua presente de forma intensa, ou seja, a transformada interpreta o ruído como sinal em alguns desses pontos. Tal fato é observado em todos os sinais resultantes do processo de *de-noising*. A solução para esta situação seria a utilização de um algoritmo detector de silêncio que faria o ajuste do nível do sinal para zero nesses pontos (esta aplicação seria interessante nos períodos longos de silêncio). A avaliação subjetiva também permite um certo nível de comparação entre os dois esquemas de *threshold*. Os sinais submetidos ao esquema rígido (*hard*) apresentam maior frequência de trechos com distorções nas vozes dos locutores do que aqueles submetidos ao esquema suave (*soft*). Isto ocorre devido às descontinuidades geradas pelo *hard-thresholding*, que tem uma tendência a preservar o formato de onda original, em contrapartida ao *soft-thresholding*, que tende a uma reprodução mais fiel do som original, com o sacrifício da forma de onda. Uma vez que a característica utilizada para a avaliação do

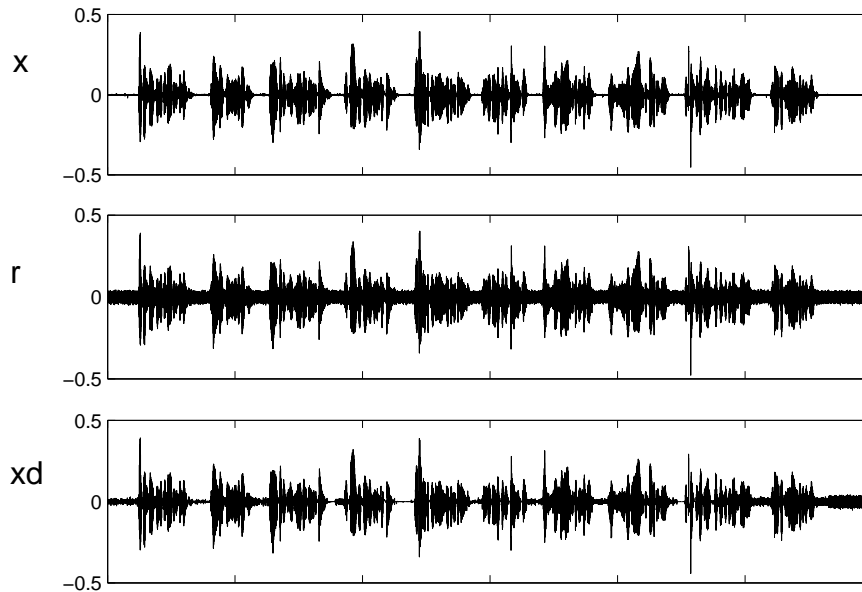


Figura 7.2: *De-noising* do sinal lm101ls11.wav

*de-noising* é, basicamente, o MSE (e a SNR, que é a inversa prática), que é uma medida de correlação entre o sinal original e o *de-noised*, há que se concluir que o *soft threshold* não está sendo escolhido da forma mais adequada. Isso se reflete nos valores do MSE médio obtidos no esquema *soft*, que se encontram em um patamar inferior em relação aos obtidos com o *hard threshold*. Da mesma forma que para as amostras do sinal, uma análise subjetiva mais precisa na tentativa de escolher um melhor *de-noising* não é viável ou mesmo possível, uma vez que os sinais resultantes não apresentam uma diferenciação marcante. Seguir-se-á utilizando o MSE como parâmetro de diferenciação. O MSE privilegia a preservação da forma de onda, por isso é justificada a tendência à escolha pelo *hard threshold*. Uma análise subjetiva provavelmente apontaria o *soft threshold* como esquema mais adequado.

### 7.2.3 Análise para diferentes níveis de ruído

Neste experimento, as amostras que foram retiradas do sinal o foram em relação ao mesmo trecho, de modo a preservar as características deste trecho e proceder à análise referente apenas à variação do nível de ruído. O arquivo utilizado nesta fase foi o lf508ls14.wav, gravado por uma locutora diferente daquelas escolhidas para as etapas passadas. A seguinte seqüência de frases foi lida pela locutora:

“A velha leoa ainda aceita combater. É hora do homem se humanizar mais. Ela ficou na fazenda por uma hora. Seu crime foi totalmente encoberto. A escuridão da garagem assustou a criança. Ontem não pude fazer minha ginástica Comer quindim é sempre uma boa pedida.

Hoje eu irei precisar de você. Sem ele o tempo flui num ritmo suave. A sujeira lançada nos rios contamina os peixes.”

Para o sinal, foram inseridas os níveis de 0, 5, 10 e 15dB e repetido o procedimento de escolha de wavelets-mãe para as amostras executado no restante dos sinais do banco anteriormente estudados. Em relação a este procedimento, optou-se por utilizar apenas 64 deslocamentos circulares, uma vez que este foi um parâmetro ótimo anteriormente obtido (será, portanto, verificado se o mesmo caberia para os diversos níveis de ruído). Os resultados obtidos estão expostos nas tabelas 7.9 e 7.10.

Para estas tabelas, pôde-se verificar uma certa invariância nas funções-base ideais para cada uma das situações. A tabela 7.11 foi elaborada com o intuito de facilitar a visualização desse fato.

Pode-se, portanto, concluir que é possível a extrapolação dos parâmetros anteriormente obtidos para outros níveis de ruídos, não sendo os mesmos restritos aos 10dB (o que reduziria a utilidade dos mesmos).

Pode-se realmente verificar que as funções-base tiveram uma variação mínima para os vários níveis de ruído, destacando-se, para o *hard threshold*, as Coiflets 4 e Symlets 8, e, para o *soft threshold*, as Biorotogonais 2.8 e as Discrete Meyer. O resultado também encontra-se em conformidade com as análises previamente realizadas.

#### 7.2.4 Utilização dos parâmetros obtidos

Por fim, a tabela III.29 apresenta o resultado da análise de todos os sinais de um dos locutores masculinos (20 sinais). Para estes sinais, foram utilizadas as *wavelets*-mãe Coiflets 4 para o *hard threshold* e as Biorotogonais 2.8 para o *soft threshold*. Para estes sinais, o MSE médio obtido foi:

- Para o *hard threshold*,  $MSE_h = 0,0525$ ; no intervalo de [3,25;4,5]
- Para o *soft threshold*,  $MSE_s = 0,0460$ ; no intervalo de [1;1,25]

### 7.3 ANÁLISE DO BANCO DE 8KHZ

Como proposta de utilização do algoritmo do PESQ no resultado do *de-noising*, foi utilizado um segundo banco de sinais amostrados a 8kHz. O procedimento da etapa anterior foi repetido para esta etapa, sendo, desta vez, a escolha de *wavelets*-mãe realizada com apenas 64 deslocamentos das amostras, de modo a agilizar o processo. Os sinais são gravações de leituras de textos diferentes pelos locutores, sendo tais textos relacionados entre os anexos.

### 7.3.1 Análise das amostras

Da mesma forma que na análise do banco de 22kHz, foram retiradas amostras aleatórias dos sinais com duração de 10% da duração das gravações originais de modo a agilizar (e até mesmo viabilizar) a obtenção de parâmetros. O procedimento inicial foi a obtenção das melhores funções-base seguido pela verificação da quantidade ótima de deslocamentos, com a diferença de que todas estas etapas obtiveram os melhores parâmetros tanto para o PESQ quanto para o MSE e, também, restringindo as análises de deslocamento a apenas duas das amostras, uma de cada gênero. As análises para escolha de funções-base foram realizadas com 2, 4 e 8 níveis de decomposição, justificado pela utilização do PESQ, não se podendo utilizar a experiência adquirida com a análise para o MSE. Também diferentemente da etapa anterior, os sinais completos foram analisados na totalidade com os parâmetros ótimos obtidos para as duas figuras de mérito (e não com os melhores valores particulares de cada sinal).

#### 7.3.1.1 Escolha de Funções-Base

A tabela 7.12 relaciona os arquivos utilizados juntamente às durações dos mesmos. As tabelas III.21 a III.26 apresentam os dados obtidos na escolha da função base e do nível de decomposição para cada um dos sinais, sendo um resumo com os melhores resultados apresentado nas tabelas 7.13 a 7.15. Essas tabelas foram elaboradas de forma análoga às da análise anterior, utilizando-se o melhor resultado como pontuação unitária (percentual máximo) e expressando os demais como proporções deste em uma média. A análise das tabelas III.21 a III.26 permite a comparação entre os *de-noisings* nos diferentes níveis de decomposição, podendo-se reparar que os melhores resultados em termos do PESQ ocorreram, em geral, para a decomposição em 16 níveis, enquanto para o MSE se destacou a decomposição em 8 níveis (situação que coincide com os resultados do banco de 22k). Dentre as *wavelets*-mãe ensaiadas, destacaram-se principalmente as Discrete Meyer, resultando nos melhores *de-noisings* para praticamente todos os casos, com excessão do PESQ com *hard threshold* para as vozes masculinas, onde a mesma foi suavemente superada pela Coiflets 4. Nesse aspecto também há algum nível de coincidência com os resultados anteriores, nos quais a *wavelet*-mãe Discrete Meyer também se destacou entre as representações. Uma observação de interesse surgida se refere à melhor avaliação do PESQ para o esquema de *soft threshold*, enquanto para o MSE prevalece o *hard threshold*, ratificando uma observação anterior e agregando uma informação extra para a condição de análise dos sinais inteiros.

### 7.3.1.2 Análise de deslocamentos e janela de corte

Uma vez que a banda passante de um sinal de 8kHz é bastante limitada para a modelagem de um sinal de voz de qualidade em comparação com os 22kHz utilizados anteriormente, não se pode intuir uma hipótese de extrapolação direta dos resultados anteriormente obtidos para os dois parâmetros desta seção. Além disso, uma vez que a banda passante do sinal se tornou mais limitada, a possibilidade de reduzir o tamanho da janela de corte para 8192 amostras se mostra como um atrativo, justificando a repetição do procedimento anterior. Apenas uma ressalva em relação a esta repetição: a análise não foi realizada para todas as amostras, sendo escolhida uma voz feminina e uma masculina e limitando às duas o procedimento. Os resultados podem ser observados nas tabelas III.27 e III.28. As tabelas mostram que as janelas de um oitavo do tamanho da amostra permitiram uma análise aceitável, incorrendo em uma decisão de utilizar 8192 amostras para compor a janela de corte na análise dos sinais inteiros. A quantidade de deslocamentos será novamente fixada em 64, uma vez que a velocidade de análise e os valores experimentais obtidos corroboram esta escolha.

### 7.3.2 Análise dos sinais inteiros

Para a análise dos sinais inteiros, utilizou-se a decomposição em 8 níveis de resolução associada ao esquema de *hard threshold* uma vez que este é o que permite os melhores valores de MSE e uma decomposição em 16 níveis para o esquema de *soft threshold*, que é o que viabiliza os melhores valores do PESQ. Além disso, foi utilizada uma quantidade de 64 deslocamentos circulares e, para todos os sinais, a função-base foi a Discrete Meyer, parâmetros compatíveis com os utilizados para a análise dos sinais inteiros do banco de 22kHz. Os resultados encontram-se indicados na tabela 7.16.

Em relação ao tamanho da janela de corte, foi realizado inicialmente o procedimento para os dois primeiros sinais masculino e feminino com dois dos tamanhos de janela, 8192 e 16382 amostras. Observou-se que, ao contrário do que se esperava, o tamanho de janela de 8192 resultou em valores inferiores àqueles que se podem obter com uma janela de 16382. Esses resultados iniciais levaram à decisão de continuar a análise com uma janela de 16382 amostras. O valor médio do MSE obtido para a análise dos sinais inteiros foi  $MSE_h = 0,0358$  para o *hard threshold* e  $MSE_s = 0,04089$  para o *soft threshold*. Para o PESQ, obteve-se  $PESQ_h = 2,204$  e  $PESQ_s = 2,349$ . O intervalo de percentual em que se obteve menor MSE foi [5,5;6,5] e o intervalo com melhor PESQ [1,25;5,5]. Observa-se que há uma maior regularidade associada à determinação do MSE, por ter proporcionado um intervalo mais curto de determinação. Na análise subjetiva observa-se uma distorção acentuada dos sinais aos quais se aplicou o *hard threshold* e uma diminuição do volume nos submetidos ao *soft threshold*. Uma observação imediata quando da comparação dos resultados obtidos nesta etapa àqueles obtidos com o banco de 22kHz é a de que o *de-noising* fica

comprometido pela largura da banda do sinal. Esse resultado era esperado, no sentido de que quando se utiliza uma maior largura de banda, como em 22kHz, a energia do ruído se concentra também em regiões do espectro que não são utilizadas pela voz falada, o que não ocorre com 8kHz, situação em que há ocupação de quase toda a largura de banda.

Tabela 7.5: Resultados do *de-noising* dos sinais inteiros com a extrapolação dos parâmetros

Sinal	thresh	wmãe	tam jan	qtd jan	%	MSE	SNRt	%	SNRseg	SNRt
lf501ls11	s	bior2.8	17891	40	1,25	0,04095	13,8774	1,5	12,4202	13,8146
		dmey			1,75	0,040874	13,8855	1,75	12,4538	-
	h	bior2.8			3,75	0,040779	13,8957	3,75	12,7404	-
		sym8			4,25	0,040102	13,9683	4,25	12,7625	-
lf502ls11	s	bior2.8	16622	40	1,25	0,04722	13,2588	1,5	10,5517	13,1501
		dmey			1,5	0,047256	13,2554	1,75	10,5781	13,2268
	h	bior2.8			3,5	0,048062	13,182	4	10,7093	13,1267
		sym8			4,25	0,048843	13,112	4,75	10,6485	13,0424
lf503ls11	s	bior2.8	15160	43	1	0,049005	13,0976	1,25	11,8974	12,9838
		dmey			1,25	0,047396	13,2426	1,75	12,0955	13,03
	h	coif4			3,5	0,052707	12,7813	4,25	12,0138	12,6364
		sym8			3,5	0,052905	12,765	4,25	12,0007	12,7053
lf504ls11	s	coif4	15178	40	1,5	0,047809	13,2049	1,75	10,4303	13,1947
		dmey	17365	35	2	0,046948	13,2838	2	10,5269	-
	h	coif4			4,25	0,043242	13,641	4,75	11,5953	13,6111
		sym8			4,25	0,043467	13,6184	4,5	11,5674	13,6143
lf505ls11	s	bior2.8	13258	40	1,5	0,042534	13,7127	1,5	12,3776	-
		dmey			2	0,040742	13,8995	2,25	12,5786	13,8842
	h	coif4			4,75	0,038923	14,0979	5	12,9367	14,086
		sym8			4,75	0,039143	14,0735	5	12,9014	14,0548
lm101ls11	s	bior2.8	29768	20	1	0,035648	14,4797	1,25	13,3771	14,3908
		dmey			1,5	0,03827	14,1714	1,5	13,1069	-
	h	dmey			3,75	0,035566	14,4897	3,75	13,6128	-
		sym8			3,25	0,033789	14,7122	3,5	13,8511	14,7027
lm102ls11	s	bior2.8	17365	40	1,5	0,034627	14,6059	1,5	12,1508	-
		dmey			1,75	0,034879	14,5743	1,75	12,1123	-
	h	bior2.8			3,75	0,033667	14,728	3,5	12,6478	14,7253
		coif4			4,25	0,034763	14,5888	4,25	12,5035	-
lm103ls11	s	bior2.8	15160	40	1,25	0,037521	14,2573	1,75	13,2936	14,0762
		dmey			1,75	0,037591	14,2492	2	13,2779	14,14
	h	db4			4,75	0,03746	14,2643	5,5	13,7075	14,1501
		sym8			4,75	0,037155	14,2998	5,5	13,7284	14,1807
lm104ls11	s	bior2.8	17365	33	1,25	0,04294	13,6714	1,25	11,9942	-
		dmey			1,5	0,042538	13,7122	1,75	12,0116	13,6951
	h	bior2.8			3	0,042143	13,7527	3,25	12,2402	13,7386
		coif4			3,75	0,041262	13,8445	4	12,3116	13,8316
lm105ls11	s	bior2.8	34343	20	1,25	0,036304	14,4005	1,25	13,9514	-
		dmey			1,5	0,038826	14,1088	1,75	13,7285	13,9667
	h	coif4			3,75	0,036164	14,4173	4	14,1161	14,3786
		sym8			3,75	0,036289	14,4022	4	14,0933	14,3641



Tabela 7.6: Quantificação dos erros da extrapolação para o grupo feminino

Sinal	thresh	wmãe	MSE
lf501ls11	s	bior2.8	0,016
		dmey	0,016
	h	bior2.8	0,021
		sym8	0,019
lf502ls11	s	bior2.8	0,031
		dmey	0,031
	h	bior2.8	0,035
		sym8	0,035
lf503ls11	s	bior2.8	0,011
		dmey	0,011
	h	coif4	0,021
		sym8	0,017
lf504ls11	s	coif4	-0,001
		dmey	0,000
	h	coif4	0,002
		sym8	0,002
lf505ls11	s	bior2.8	0,007
		dmey	0,008
	h	coif4	0,009
		sym8	0,009

Tabela 7.7: Quantificação dos erros da extrapolação para o grupo masculino

Sinal	thresh	wmãe	MSE
lm101ls11	s	bior2.8	0,009
		dmey	0,011
	h	dmey	0,012
		sym8	0,010
lm102ls11	s	bior2.8	0,007
		dmey	0,006
	h	bior2.8	0,010
		coif4	0,011
lm103ls11	s	bior2.8	0,017
		dmey	0,017
	h	db4	0,021
		sym8	0,020
lm104ls11	s	bior2.8	0,015
		dmey	0,015
	h	bior2.8	0,021
		coif4	0,020
lm105ls11	s	bior2.8	0,012
		dmey	0,014
	h	coif4	0,016
		sym8	0,016

Tabela 7.8: Análises dos grupos de sinais

grupos		$mse_m$	$SNR_m$	< %	> %
geral	h	0,041	13,932	3,25	5,5
	s	0,041	13,847	1,25	2,25
fem	h	0,045	13,514	3,75	5
	s	0,045	13,472	1,25	2,25
masc	h	0,037	14,350	3,25	5,5
	s	0,038	14,223	1,25	2

Tabela 7.9: Análise de amostras com vários níveis de ruído

wmãe		0 dB (2 dB)			5 dB (7 dB)		
		SNRseg	MSE	%	SNRseg	MSE	%
db4	s	8,4497	0,1429	5,5	12,0723	0,062053	2,75
	h	9,4104	0,11454	12	12,8459	0,051928	6,5
coif4	s	<b>8,6619</b>	<b>0,13608</b>	<b>5,25</b>	<b>12,2712</b>	<b>0,05928</b>	<b>2,5</b>
	h	<b>9,5579</b>	<b>0,11072</b>	<b>11,25</b>	<b>13,0754</b>	<b>0,04926</b>	<b>6</b>
sym8	s	8,6181	0,13746	5,25	12,2411	0,059688	2,5
	h	<b>9,5283</b>	<b>0,11147</b>	<b>11,5</b>	<b>13,0407</b>	<b>0,04965</b>	<b>6,25</b>
dmey	s	<b>8,7223</b>	<b>0,13421</b>	<b>5,75</b>	<b>12,4023</b>	<b>0,05751</b>	<b>2,75</b>
	h	9,4421	0,11371	12,25	<b>13,0592</b>	<b>0,04944</b>	<b>6,25</b>
bior2.8	s	<b>8,8247</b>	<b>0,13108</b>	<b>4,5</b>	<b>12,3245</b>	<b>0,05855</b>	<b>2,25</b>
	h	<b>9,4589</b>	<b>0,11327</b>	<b>10,25</b>	12,7402	0,053208	5,5

Tabela 7.10: Análise de amostras com vários níveis de ruído

wmãe		10 dB (12 dB)			15 dB (17 dB)		
		SNRseg	MSE	%	SNRseg	MSE	%
db4	s	15,879	0,025829	1,25	19,8513	0,010348	0,75
	h	16,4927	0,022425	3,5	20,2668	0,009404	2
coif4	s	<b>16,0649</b>	<b>0,02475</b>	<b>1,25</b>	<b>19,9738</b>	<b>0,01006</b>	<b>0,5</b>
	h	<b>16,6816</b>	<b>0,02147</b>	<b>3,25</b>	<b>20,5833</b>	<b>0,00874</b>	<b>1,75</b>
sym8	s	16,0414	0,02488	1,25	19,9857	0,010033	0,75
	h	<b>16,6643</b>	<b>0,02156</b>	<b>3,5</b>	<b>20,5337</b>	<b>0,00884</b>	<b>1,75</b>
dmey	s	<b>16,1358</b>	<b>0,02435</b>	<b>1,25</b>	<b>20,1154</b>	<b>0,00974</b>	<b>0,75</b>
	h	16,5877	0,02194	3,5	<b>20,4661</b>	<b>0,00898</b>	<b>1,75</b>
bior2.8	s	<b>16,0547</b>	<b>0,0248</b>	<b>1</b>	<b>20,0409</b>	<b>0,00991</b>	<b>0,5</b>
	h	<b>16,2993</b>	<b>0,02345</b>	<b>3</b>	20,2146	0,009518	1,5

Tabela 7.11: Melhores funções-base para cada nível de ruído

		nível de ruído			
		0 dB	5dB	10dB	15dB
h	1	coif4	coif4	coif4	coif4
	2	sym8	dmey	sym8	sym8
	3	bior2.8	sym8	bior2.8	dmey
s	1	bior2.8	dmey	bior2.8	dmey
	2	dmey	bior2.8	dmey	bior2.8
	3	coif4	coif4	coif4	coif4

Tabela 7.12: Arquivos utilizados do banco de 8kHz

Arquivo	Duração (s)
fernandaAfogamento.wav	25
flaviaDivas.wav	36
izumiPoesia.wav	33
fredMagoa.wav	34
ortis-grua.wav	37
tiagoPalocci.wav	35

Tabela 7.13: Melhores resultados para todos os locutores

	wmãe	4		8		16	
		s	h	s	h	s	h
PESQ	db4	0,977	0,980	0,976	0,984	0,964	0,970
	coif4	<b>0,992</b>	<b>0,993</b>	0,994	<b>0,991</b>	<b>0,983</b>	0,987
	sym8	0,992	0,992	<b>0,996</b>	<b>0,997</b>	0,983	<b>0,989</b>
	dmey	<b>0,999</b>	<b>0,999</b>	<b>0,997</b>	0,991	<b>1,000</b>	<b>0,992</b>
	bior2.8	0,986	0,982	0,985	0,980	0,969	0,973
MSE	db4	0,996	0,999	0,961	0,995	0,942	0,989
	coif4	<b>0,998</b>	<b>0,999</b>	0,990	<b>0,997</b>	<b>0,973</b>	0,992
	sym8	0,998	0,999	<b>0,993</b>	<b>0,997</b>	0,972	<b>0,991</b>
	dmey	<b>0,999</b>	<b>0,999</b>	<b>0,995</b>	1,000	<b>1,000</b>	<b>0,999</b>
	bior2.8	0,999	0,998	0,975	0,998	0,949	0,992

Tabela 7.14: Melhores resultados para locutoras femininas

	wmãe	4		8		16	
		s	h	s	h	s	h
<b>PESQ</b>	<b>db4</b>	0,967	0,973	0,968	0,977	0,956	0,970
	<b>coif4</b>	<b>0,987</b>	<b>0,989</b>	0,990	<b>0,988</b>	<b>0,976</b>	0,978
	<b>sym8</b>	0,987	0,989	<b>0,991</b>	<b>0,994</b>	0,979	<b>0,990</b>
	<b>dmey</b>	<b>1,000</b>	<b>0,999</b>	<b>0,997</b>	0,988	<b>1,000</b>	<b>0,995</b>
	<b>bior2.8</b>	0,978	0,971	0,978	0,971	0,962	0,969
<b>MSE</b>	<b>db4</b>	0,996	0,998	0,947	0,994	0,928	0,987
	<b>coif4</b>	<b>0,997</b>	<b>0,999</b>	0,983	<b>0,997</b>	<b>0,961</b>	0,991
	<b>sym8</b>	0,997	0,999	<b>0,985</b>	<b>0,996</b>	0,966	<b>0,989</b>
	<b>dmey</b>	<b>1,000</b>	<b>1,000</b>	<b>0,995</b>	0,999	<b>1,000</b>	<b>0,998</b>
	<b>bior2.8</b>	0,998	0,997	0,962	0,997	0,938	0,990

Tabela 7.15: Melhores resultados para os locutores masculinos

	wmãe	4		8		16	
		s	h	s	h	s	h
<b>PESQ</b>	<b>db4</b>	0,987	0,988	0,985	0,990	0,973	0,970
	<b>coif4</b>	<b>0,998</b>	<b>0,998</b>	0,998	<b>0,994</b>	<b>0,990</b>	0,995
	<b>sym8</b>	0,997	0,996	<b>1,000</b>	<b>1,000</b>	0,986	<b>0,988</b>
	<b>dmey</b>	<b>0,998</b>	<b>1,000</b>	<b>0,997</b>	0,993	<b>1,000</b>	<b>0,990</b>
	<b>bior2.8</b>	0,994	0,992	0,993	0,990	0,976	0,976
<b>MSE</b>	<b>db4</b>	0,997	0,999	0,976	0,996	0,956	0,991
	<b>coif4</b>	<b>0,998</b>	<b>0,999</b>	0,996	<b>0,998</b>	<b>0,984</b>	0,993
	<b>sym8</b>	0,998	0,999	<b>1,000</b>	<b>0,998</b>	0,978	<b>0,993</b>
	<b>dmey</b>	<b>0,999</b>	<b>0,998</b>	<b>0,995</b>	1,000	<b>1,000</b>	<b>1,000</b>
	<b>bior2.8</b>	1,000	1,000	0,989	0,999	0,961	0,994

Tabela 7.16: Análise dos sinais inteiros do banco de 8kHz

**wmãe = dmey**

Sinal	thresh	tam jan	qtd jan	%	MSE	%	PESQ
fernanda	s	8192	34	0,75	0,044249	3	2,318
	h			6,5	0,050854	3,5	2,147
	s	16382	17	1	0,039053	5	2,546
	h			5,75	0,03042	3,75	2,302
flavia	s	16382	18	1	0,034481	2,25	1,907
	h			6,75	0,029757	4	1,701
izumi	s	16382	17	0,75	0,040898	1,25	2,376
	h			6	0,035122	2,75	2,278
fred	s	8192	34	0,75	0,047721	1,75	2,347
	h			6,5	0,061483	2,75	2,131
	s	16382	17	0,75	0,0428	3	2,451
	h			6	0,038152	3	2,227
ortis	s	16382	19	0,75	0,045396	5,5	2,525
	h			5,5	0,040785	3,25	2,359
tiago	s	16382	18	0,75	0,042708	3,25	2,292
	h		18	5,5	0,040785	3,25	2,359

## 8 CONCLUSÕES

Foi apresentada e analisada uma implementação de um sistema que utiliza a TIWD para minimização de ruído gaussiano em sinais de voz. Mais do que a preocupação com a avaliação da técnica de *de-noising* utilizada (este aspecto não foi ignorado), este projeto teve como foco a obtenção de parâmetros gerais que possam ser considerados para a implementação de etapas de minimização de ruído em quaisquer tipos de sistemas de processamento de voz que convirjam com as condições estudadas. Neste sentido, os parâmetros obtidos refletem os melhores resultados para a aplicação da técnica TIWD a locutores de ambos os sexos. Não foi possível, porém, uma verificação subjetiva da superioridade de um dos esquemas de *threshold* para aplicações de *de-noising*. Os resultados só permitiram a conclusão de que o ceifamento rígido (*hard threshold*) é mais conservativo das características da forma de onda do sinal e o ceifamento suave (*soft threshold*) apresenta resultados com uma menor perceptividade auditiva de descontinuidades (apesar de que estes resultados não necessariamente implicam uma melhor qualidade auditiva para o ceifamento suave). Além da contribuição fornecida pelos resultados obtidos, este projeto também apresentou uma metodologia que pode ser seguida para o estudo de sinais em condições diversas às aquelas apresentadas. Apesar da especificidade de aplicação para a qual se concebeu este projeto, o procedimento é suficientemente genérico para ser aplicado a uma grande quantidade de situações de processamento de voz que requeiram uma etapa de tratamento de ruído. O banco de sinais de 8kHz foi colocado como uma situação genérica e permitiu a comprovação da eficiência e eficácia da metodologia utilizada. Alguns dos obstáculos surgidos no desenvolvimento e geração de resultados se referem à implementação lenta decorrente da utilização do ambiente de desenvolvimento do Matlab. Esta peculiaridade se refletiu no tempo de processamento dos algoritmos desenvolvidos, limitando a quantidade e qualidade das análises realizadas. A implementação em alguma linguagem de programação mais rápida aumentaria a eficiência computacional permitindo uma maior abrangência para as análises tanto em relação ao número de funções-base utilizado nos testes quanto no que se refere à quantidade de sinais propriamente analisados. Além da restrição de implementação, foi identificado um refinamento que poderia aumentar a qualidade dos sinais resultantes do processo de *de-noising*. Este refinamento seria algum tipo de algoritmo de detecção de silêncio nos intervalos da fala dos locutores, uma vez que a técnica não se mostrou suficiente para a remoção do ruído nesses trechos, por vezes comprometendo o resultado final da aplicação. Também se destacou um fator como limitante da qualidade das análises realizadas; uma vez que para o presente trabalho foram utilizados predominantemente (e quase que exclusivamente) critérios objetivos como figuras de mérito na avaliação da qualidade da técnica, e em decorrência da quantidade limitada desses critérios que pôde ser analisada, verifica-se que os parâmetros obtidos podem não corresponder às características ótimas para quaisquer

tipos de aplicação. No caso do reconhecimento de locutor, algum critério que permitisse agregar a técnica utilizada para a aplicação e as figuras de mérito utilizadas no *de-noising* (como, por exemplo, algum tipo de correlação entre os coeficientes mel-cepstrais) favoreceria a especialização e maior eficiência e eficácia de ambos os processos. As sugestões para continuidade e complementação deste trabalho incluem a detecção de silêncio e a verificação de outras possíveis figuras de mérito para a avaliação do *de-noising*.

## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] PARSONS, T. *Voice and Speech Processing*. [S.l.]: McGraw-Hill, 1986.
- [2] HALLIDAY, D.; RESNICK, R.; WALKER, J. *Fundamentos da Física 2*. [S.l.]: LTC, 2001.
- [3] STARK, S. H. *Live Sound Reinforcement*. [S.l.]: MixBooks, 1996.
- [4] AINSLEY, R. *Enciclopedia de la Música Clásica*. [S.l.]: Parramon Espanha, 2006.
- [5] MED, B. *Teoria da Música*. [S.l.]: Musimed, 1996.
- [6] FLETCHER, H. *Speech and Hearing in Communication*. [S.l.]: D. Van Nostrand Company, Inc., 1953.
- [7] SANTOS, C. R. de M. *Aplicação do Espectro Auditivo no Reconhecimento de Locutor*. Dissertação (Mestrado) — Universidade de Brasília, 1998. Dissertação de mestrado.
- [8] HAZAS, M. *Processing of Non-Stationary Audio Signals*. Dissertação (Mestrado) — University of Cambridge, 1999. Dissertação de mestrado.
- [9] KOTHE, I. *Minimização de Ruído em Sinais de Áudio utilizando Transformada de Wavelets*. 2003. Projeto final de graduação.
- [10] SCHREMMER, C. *Wavelets in real time digital audio processing: Analysis and sample implementations*. Dissertação (Mestrado) — University of Mannheim, 2000. Dissertação de mestrado.
- [11] VALENS, C. *A Really Friendly Guide to Wavelets*. 1999. Não publicado. Disponível em: <[www.cs.unm.edu/williams/cs530/arfgtw.pdf](http://www.cs.unm.edu/williams/cs530/arfgtw.pdf)>.
- [12] POLIKAR, R. *The Wavelet Tutorial*. 2001. Não publicado. Disponível em: <<http://users.rowan.edu/polikar/WAVELETS/WTtutorial.html>>.
- [13] KAISER, G. *A Friendly Guide to Wavelets*. [S.l.]: Birkhäuser, 1994.
- [14] SHENG, Y. *The transforms and applications handbook - The Electrical Engineering Handbook Series*. [S.l.]: CRC Press, 1994.
- [15] BURRUS, C. S.; GOPINAH, R. A.; GUO, H. *A Primer Introduction to Wavelets and Wavelet Transforms*. [S.l.]: Prentice Hall, 1998.
- [16] DAUBECHIES, I. *Ten Lectures on Wavelets*. [S.l.]: CBMS-NSF regional conference series in applied mathematics 61, 1992.



- [17] MALLAT, S. *A Wavelet Tour of Signal Processing*. [S.l.]: Academic Press, 1998.
- [18] BORGES, F. B. *Codificadores de Voz a Baixas Taxas Operando em Ambientes Ruidosos e Redes IP*. Dissertação (Mestrado) — PUC-Rio, 2005. Dissertação de mestrado.
- [19] DIAS, M. H. C. *Estimação das Respostas do Canal Real de Propagação rádio-móvel nos domínios Espacial e Temporal, Análise da supressão de Ruído por decomposição Wavelet como técnica complementar de processamento*. Dissertação (Mestrado) — PUC-Rio, 2003. Dissertação de mestrado.
- [20] DONOHO, D. L. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*.
- [21] DONOHO, D. L.; JOHNSTONE, I. M. Ideal spatial adaptation via wavelet shrinkage. 1992.
- [22] DONOHO, D. L.; JOHNSTONE, I. M. Ideal denoising in an orthonormal basis chosen from a library of bases. 1994.
- [23] LUNDEEN, J. *Mean Squared Error*. 2006. Não publicado. Disponível em: <[http://en.wikipedia.org/wiki/Mean\\_square\\_error](http://en.wikipedia.org/wiki/Mean_square_error)>.
- [24] BEERENDS, J. et al. Perceptual evaluation of speech quality (pesq), the new itu standard for end-to-end speech quality assessment. *J. Audio Eng. Soc.*, Part II - Psychoacoustic model, 2001.
- [25] COIFMAN, R.; DONOHO, D. Translation-invariant de-noising. 1995.



# I. DESCRIÇÃO DO CONTEÚDO DO CD

No CD incluído em anexo ao projeto, encontram-se alguns dos sinais utilizados para o de-noising e aqueles que resultaram do processo.

## **BANCO 22050Hz**

- Exemplos de amostras

- 1.

- aleat-lf501ls11.wav

- aleat-lf501ls11-ruído10dB.wav

- aleat-lf501ls11-denoised.wav

- 2.

- aleat-lm103ls11.wav

- aleat-lm103ls11-ruído10dB.wav

- aleat-lm103ls11-denoised.wav

- 4 sinais de teste

- 1.

- lm101ls11.wav

- lm101ls11-ruído10dB.wav

- lm101ls11-hard-sym8.wav

- lm101ls11-soft-bior2.8.wav

- 2.

- lm102ls11.wav

- lm102ls11-ruído10dB.wav

- lm102ls11-hard-bior2.8.wav

- lm102ls11-soft-dmey.wav

- 3.

- lf501ls11.wav

- lf501ls11-ruído10dB.wav

- lf501ls11-hard-bior2.8.wav

- lf501ls11-soft-bior2.8.wav

– 4.

lf502ls11.wav

lf502ls11-ruido10dB.wav

lf502ls11-soft-bior2.8.wav

lf502ls11-hard-sym8.wav

- 2 sinais com o *de-noising* extrapolado

– 1.

lm102ls29.wav

lm102ls29-ruido10dB.wav

lm102ls29-hard-denoised.wav

lm102ls29-soft-denoised.wav

– 2.

lm103ls19.wav

lm103ls19-ruido10dB.wav

lm103ls19-hard-denoised.wav

lm103ls19-soft-denoised.wav

## **BANCO 8000HZ**

- 2 sinais de teste

– 1.

fredMagoa.wav

fredMagoa-ruido10dB.wav

fredMagoa-hard-dmey.wav

fredMagoa-soft-dmey.wav

– 2.

fernandaAfogamento.wav

fernandaAfogamento-ruido10dB.wav

fernandaAfogamento-hard-dmey.wav

fernandaAfogamento-soft-dmey.wav

## II. TEXTOS

### TEXTOS UTILIZADOS NAS GRAVAÇÕES DOS SINAIS DO BANCO DE 8KHz

#### **fernandaAfogamento.wav**

Para o corpo de bombeiros a falta de informação é o principal motivo do grande índice de afogamentos. Como não há razões externas identificadas vamos trabalhar com educação, explica o Capitão. Além de prevenir, os bombeiros também vão ensinar como proceder em caso de afogamento. As mortes do córrego Taguatinga ocorreram porque as pessoas não tinham conhecimento de técnicas de salvamento. Nos dois casos de acidentes, um dos banhistas afogou-se e o outro tentou salvá-lo. Por isso vamos ensinar as pessoas, explica o Capitão Geraldo. Antes de nadar, não ingerir bebidas alcoólicas. Procure conhecer o local an...

#### **flaviaDivas.wav**

Com o objetivo de unir o popular e o erudito em uma única apresentação, o Centro Cultural Banco do Brasil dá sequência ao projeto Divas. Nesta terça a mezzo-soprano Magda Paino e a cantora popular Vânia Bastos presenteiam o público da capital com um repertório que exalta o lirismo da música brasileira de todos os tempos. Canções inspiradas no folclore de vários países e algumas das mais belas criações de Tom Jobim também compõem o roteiro. Acompanhadas dos pianistas Fábio Caramuru e Marco Antônio Bernardo e de Pedro Baldanza (Baixo Elétrico).

#### **izumiPoesia.wav**

A Poesia em debate

No dia da poesia, comemoramos nesta terça a posição do poeta está no centro do debate. E nada melhor do que fomentar a discussão com o próprio resultado do trabalho. Em "o poema quer ser útil", Angélica Torres Lima coloca lenha na fogueira e inicia o falatório sobre a sobrevivência e o ofício de um artista do verbo. Mesmo sendo um segmento de arte pouco consumido ela acredita que o povo brasileiro não vai se perder da poesia. Tudo por conta do seu legado literário e de força que ajuda na aproximação com o verbo.

#### **fredMagoa.wav**

Mágoa

Em entrevista por telefone à GE.net outro candidato ao cargo de treinador do timão, justamente o ex-demitido Márcio Bittencourt enfatizou que não foi procurado por ninguém da diretoria da MSI. Márcio é o nome favorito de pessoas ligadas à direção, mas demonstra ressentimentos ao lembrar da forma como saiu

no último campeonato brasileiro. - O trabalho que fiz lá foi incontestável.

#### **ortiz-grua.wav**

Um homem de 300kg teve que ser tirado, com a ajuda de uma grua, de seu apartamento, situado no 4º andar de um edifício de Hanover, anunciaram os bombeiros locais. A operação de resgate durou 3h e mobilizou 35 bombeiros. 10 bombeiros conseguiram colocar, com grande esforço, o homem de 300kg em uma cesta especial fixada a uma grua de 60 toneladas depois que instalaram em sua varanda uma rampa com chapas metálicas e táboas de madeira. Uma primeira tentativa, com uma escada mecânica dos bombeiros fracassou.

#### **tiagoPalocci.wav**

O ministério da fazenda divulgou nesta terça-feira uma nota em que o ministro Antonio Palocci nega que tenha frequentado uma casa no Lago Sul, em Brasília, onde provavelmente seus ex-assessores da prefeitura de Ribeirão Preto, Rogério Buratti e Wladimir Boleto, fizeram a repartição de dinheiro proveniente de doações ilegais a campanhas do PT. O ministro reafirma o que disse no depoimento à CPI dos bingos, que nunca foi à casa e que não tem conhecimento das atividades que aconteciam lá.

### III. TABELAS

Tabela III.1: Escolha wavelet-mãe para amostra do arquivo lf501ls11.wav

Níveis		4			8		
wmãe		> SNRseg	< MSE	%	> SNRseg	< MSE	%
db4	s	15,3691	0,029046	1,75	15,5988	0,02755	1,5
	h	16,238	0,023779	4	16,7029	0,021365	3,75
coif4	s	<b>15,4652</b>	<b>0,02841</b>	<b>1,5</b>	<b>15,7827</b>	<b>0,026408</b>	<b>1,5</b>
	h	<b>16,2674</b>	<b>0,023619</b>	<b>4</b>	<b>16,7991</b>	<b>0,020897</b>	<b>3,75</b>
sym8	s	15,4517	0,028499	1,5	15,737	0,026687	1,5
	h	<b>16,2763</b>	<b>0,02357</b>	<b>4</b>	<b>16,8044</b>	<b>0,020872</b>	<b>3,75</b>
dmey	s	<b>15,5802</b>	<b>0,027668</b>	<b>1,5</b>	<b>15,9592</b>	<b>0,025356</b>	<b>1,75</b>
	h	16,1034	0,024528	3,75	16,5487	0,022138	4
bior2.8	s	<b>15,7202</b>	<b>0,026791</b>	<b>1,25</b>	<b>15,9939</b>	<b>0,025154</b>	<b>1,25</b>
	h	<b>16,4911</b>	<b>0,022433</b>	<b>3,5</b>	<b>16,9298</b>	<b>0,020278</b>	<b>3,25</b>

Tabela III.2: Escolha wavelet-mãe para amostra do arquivo lf502ls11.wav

Níveis		4			8		
wmãe		> SNRseg	< MSE	%	> SNRseg	< MSE	%
db4	s	17,2297	0,018925	1	17,5105	0,01774	0,75
	h	18,0673	0,015605	2,25	18,4751	0,014207	1,75
coif4	s	<b>17,3884</b>	<b>0,018246</b>	<b>1</b>	<b>17,6439</b>	<b>0,017203</b>	<b>0,75</b>
	h	<b>18,2181</b>	<b>0,015073</b>	<b>2,25</b>	<b>18,6204</b>	<b>0,013739</b>	<b>1,75</b>
sym8	s	17,3588	0,01837	1	17,64	0,017219	0,75
	h	<b>18,2363</b>	<b>0,015009</b>	<b>2,25</b>	<b>18,6413</b>	<b>0,013673</b>	<b>1,75</b>
dmey	s	<b>17,5922</b>	<b>0,017409</b>	<b>1</b>	<b>17,9276</b>	<b>0,016115</b>	<b>0,75</b>
	h	18,2101	0,0151	2,25	18,591	0,013832	1,75
bior2.8	s	<b>17,5951</b>	<b>0,017398</b>	<b>0,75</b>	<b>17,86</b>	<b>0,016368</b>	<b>0,5</b>
	h	<b>18,3745</b>	<b>0,014539</b>	<b>2,25</b>	<b>18,7401</b>	<b>0,013366</b>	<b>1,5</b>

Tabela III.3: Escolha wavelet-mãe para amostra do arquivo lf503ls11.wav

Níveis		4			8		
wmãe		> SNRseg	< MSE	%	> SNRseg	< MSE	%
db4	s	13,718	0,042482	1,75	14,0498	0,039357	1,5
	h	14,1892	0,038114	4,25	14,9212	0,032202	4
coif4	s	<b>13,8099</b>	<b>0,041592</b>	<b>1,75</b>	<b>14,1716</b>	<b>0,038269</b>	<b>1,5</b>
	h	<b>14,2842</b>	<b>0,037289</b>	<b>4</b>	<b>15,0021</b>	<b>0,031607</b>	<b>4</b>
sym8	s	13,8139	0,041554	1,75	14,1612	0,03836	1,5
	h	<b>14,2614</b>	<b>0,037485</b>	<b>4</b>	<b>14,9989</b>	<b>0,031631</b>	<b>4</b>
dmey	s	<b>13,8828</b>	<b>0,0409</b>	<b>1,75</b>	<b>14,3906</b>	<b>0,036386</b>	<b>1,75</b>
	h	14,1682	0,038299	4,25	14,8538	0,032706	4
bior2.8	s	<b>13,9613</b>	<b>0,040167</b>	<b>1,5</b>	<b>14,2409</b>	<b>0,37662</b>	<b>1,25</b>
	h	<b>14,2988</b>	<b>0,037164</b>	<b>3,5</b>	<b>14,8685</b>	<b>0,032595</b>	<b>3,25</b>

Tabela III.4: Escolha wavelet-mãe para amostra do arquivo lf504ls11.wav

Níveis		4			8		
wmãe		> SNRseg	< MSE	%	> SNRseg	< MSE	%
db4	s	12,8773	0,051555	2	13,0381	0,049681	1,75
	h	13,1866	0,048011	5,25	13,6617	0,043036	4,75
coif4	s	<b>12,9758</b>	<b>0,050399</b>	<b>2</b>	<b>13,1568</b>	<b>0,048341</b>	<b>1,75</b>
	h	<b>13,3226</b>	<b>0,046531</b>	<b>5,25</b>	<b>13,8294</b>	<b>0,041406</b>	<b>4,5</b>
sym8	s	12,9741	0,050419	2	13,1296	0,048645	1,75
	h	<b>13,3402</b>	<b>0,046343</b>	<b>5,25</b>	<b>13,8309</b>	<b>0,041391</b>	<b>4,5</b>
dmey	s	<b>13,0581</b>	<b>0,049453</b>	<b>2</b>	<b>13,3111</b>	<b>0,046654</b>	<b>1,75</b>
	h	13,2659	0,047142	5,25	13,6792	0,042862	4,5
bior2.8	s	<b>12,9992</b>	<b>0,050128</b>	<b>1,75</b>	<b>13,102</b>	<b>0,048955</b>	<b>1,5</b>
	h	<b>13,1476</b>	<b>0,048445</b>	<b>4,25</b>	<b>13,5852</b>	<b>0,043801</b>	<b>3,5</b>



Tabela III.5: Escolha wavelet-mãe para amostra do arquivo lf505ls11.wav

Níveis		4			8		
wmãe		> SNRseg	< MSE	%	> SNRseg	< MSE	%
db4	s	14,1142	0,038777	2	14,3059	0,037103	1,5
	h	14,5686	0,034925	4,75	15,0006	0,031618	4
coif4	s	<b>14,2857</b>	<b>0,037276</b>	<b>2</b>	<b>14,4976</b>	<b>0,035501</b>	<b>1,5</b>
	h	<b>14,7938</b>	<b>0,03316</b>	<b>5</b>	<b>15,2858</b>	<b>0,029608</b>	<b>4</b>
sym8	s	14,2755	0,037364	2	14,483	0,035621	1,5
	h	<b>14,7919</b>	<b>0,033175</b>	<b>4,75</b>	<b>15,2732</b>	<b>0,029694</b>	<b>4</b>
dmey	s	<b>14,4846</b>	<b>0,035608</b>	<b>2</b>	<b>14,7835</b>	<b>0,033239</b>	<b>1,75</b>
	h	14,8117	0,033024	4,75	15,2726	0,029699	4,5
bior2.8	s	<b>14,3964</b>	<b>0,036338</b>	<b>1,5</b>	<b>14,5235</b>	<b>0,03529</b>	<b>1,25</b>
	h	<b>14,6602</b>	<b>0,034196</b>	<b>3,5</b>	<b>15,0784</b>	<b>0,031057</b>	<b>3,25</b>

Tabela III.6: Escolha wavelet-mãe para amostra do arquivo lm101ls11.wav

Níveis		4			8		
wmãe		> SNRseg	< MSE	%	> SNRseg	< MSE	%
db4	s	15,1955	0,030231	1,5	15,3898	0,028908	1
	h	15,3455	0,029204	3,75	15,9158	0,02561	2,25
coif4	s	<b>15,3423</b>	<b>0,029226</b>	<b>1,5</b>	<b>15,5558</b>	<b>0,027824</b>	<b>1</b>
	h	<b>15,538</b>	<b>0,027938</b>	<b>3,75</b>	<b>16,1698</b>	<b>0,024156</b>	<b>2,5</b>
sym8	s	15,3458	0,029202	1,5	15,5435	0,027903	1
	h	<b>15,5201</b>	<b>0,028054</b>	<b>3,75</b>	<b>16,1778</b>	<b>0,024111</b>	<b>2,25</b>
dmey	s	<b>15,4341</b>	<b>0,028615</b>	<b>1,5</b>	<b>15,5885</b>	<b>0,027616</b>	<b>1</b>
	h	15,6788	0,027047	3,75	16,2088	0,02394	2,5
bior2.8	s	<b>15,5315</b>	<b>0,02798</b>	<b>1,25</b>	<b>15,6996</b>	<b>0,026918</b>	<b>0,75</b>
	h	<b>15,4418</b>	<b>0,028564</b>	<b>3,25</b>	<b>16,0315</b>	<b>0,024937</b>	<b>2</b>

Tabela III.7: Escolha wavelet-mãe para amostra do arquivo lm102ls11.wav

Níveis		4			8		
wmãe		> SNRseg	< MSE	%	> SNRseg	< MSE	%
db4	s	15,0231	0,031455	1,75	15,2361	0,02995	1,5
	h	15,7099	0,026854	4,25	16,2377	0,023781	3,25
coif4	s	<b>15,0568</b>	<b>0,031212</b>	<b>1,75</b>	<b>15,2958</b>	<b>0,029541</b>	<b>1,25</b>
	h	<b>15,7251</b>	<b>0,02676</b>	<b>4,25</b>	<b>16,2476</b>	<b>0,023727</b>	<b>3</b>
sym8	s	15,0678	0,031133	1,75	15,2972	0,029531	1,25
	h	<b>15,704</b>	<b>0,026891</b>	<b>4,25</b>	<b>16,2298</b>	<b>0,023824</b>	<b>3</b>
dmey	s	<b>15,1477</b>	<b>0,030565</b>	<b>1,75</b>	<b>15,4287</b>	<b>0,02865</b>	<b>1,5</b>
	h	15,6082	0,02749	4	16,0581	0,024785	3
bior2.8	s	<b>15,3804</b>	<b>0,02897</b>	<b>1,25</b>	<b>15,5623</b>	<b>0,027782</b>	<b>1</b>
	h	<b>15,9358</b>	<b>0,025493</b>	<b>2,75</b>	<b>16,3016</b>	<b>0,023433</b>	<b>2,5</b>

Tabela III.8: Escolha wavelet-mãe para amostra do arquivo lm103ls11.wav

Níveis		4			8		
wmãe		> SNRseg	< MSE	%	> SNRseg	< MSE	%
db4	s	16,2212	0,023871	1,5	16,5585	0,022088	1
	h	16,9066	0,020387	3,75	17,7283	0,016872	2,5
coif4	s	<b>16,2543</b>	<b>0,02369</b>	<b>1,5</b>	<b>16,6524</b>	<b>0,021615</b>	<b>1</b>
	h	<b>16,9366</b>	<b>0,020246</b>	<b>4</b>	<b>17,7225</b>	<b>0,016895</b>	<b>2,5</b>
sym8	s	16,2857	0,02352	1,5	16,6435	0,021659	1
	h	<b>16,9178</b>	<b>0,020334</b>	<b>3,75</b>	<b>17,7253</b>	<b>0,016884</b>	<b>2,5</b>
dmey	s	<b>16,412</b>	<b>0,022845</b>	<b>1,75</b>	<b>16,8375</b>	<b>0,020714</b>	<b>1</b>
	h	16,8991	0,020422	4	17,6155	0,017316	2,5
bior2.8	s	<b>16,4941</b>	<b>0,022417</b>	<b>1,25</b>	<b>16,779</b>	<b>0,020994</b>	<b>1</b>
	h	<b>16,9963</b>	<b>0,01997</b>	<b>3,25</b>	<b>17,6339</b>	<b>0,017243</b>	<b>2,25</b>

Tabela III.9: Escolha wavelet-mãe para amostra do arquivo lm104ls11.wav

Níveis		4			8		
wmãe		> SNRseg	< MSE	%	> SNRseg	< MSE	%
<b>db4</b>	s	14,9307	0,032132	1,75	15,196	0,030228	1
	h	15,8595	0,025945	4,25	16,5226	0,022271	2,5
<b>coif4</b>	s	<b>15,0546</b>	<b>0,031228</b>	<b>2</b>	<b>15,3499</b>	<b>0,029175</b>	<b>1</b>
	h	<b>15,9599</b>	<b>0,025352</b>	<b>4,75</b>	<b>16,6741</b>	<b>0,021508</b>	<b>2,5</b>
<b>sym8</b>	s	15,0431	0,031311	2	15,3417	0,02923	1
	h	<b>15,957</b>	<b>0,025369</b>	<b>4,5</b>	<b>16,6616</b>	<b>0,02157</b>	<b>2,5</b>
<b>dmey</b>	s	<b>15,2217</b>	<b>0,030049</b>	<b>2</b>	<b>15,5536</b>	<b>0,027838</b>	<b>1</b>
	h	15,903	0,025686	4,75	16,5414	0,022175	2,5
<b>bior2.8</b>	s	<b>15,3831</b>	<b>0,028953</b>	<b>1,5</b>	<b>15,5711</b>	<b>0,027726</b>	<b>1</b>
	h	<b>16,1674</b>	<b>0,024169</b>	<b>3,75</b>	<b>16,7517</b>	<b>0,021127</b>	<b>2</b>

Tabela III.10: Escolha wavelet-mãe para amostra do arquivo lm105ls11.wav

Níveis		4			8		
wmãe		> SNRseg	< MSE	%	> SNRseg	< MSE	%
<b>db4</b>	s	15,6304	0,02735	1,75	15,8122	0,026229	0,5
	h	16,4172	0,022818	4	16,7906	0,020938	1,25
<b>coif4</b>	s	<b>15,7181</b>	<b>0,026804</b>	<b>1,75</b>	<b>15,9498</b>	<b>0,025411</b>	<b>0,5</b>
	h	<b>16,426</b>	<b>0,022772</b>	<b>4</b>	<b>16,8709</b>	<b>0,020555</b>	<b>1,25</b>
<b>sym8</b>	s	15,7153	0,026821	1,75	15,9259	0,025551	0,5
	h	<b>16,455</b>	<b>0,02262</b>	<b>4</b>	<b>16,8568</b>	<b>0,020621</b>	<b>1,25</b>
<b>dmey</b>	s	<b>15,8644</b>	<b>0,025916</b>	<b>1,75</b>	<b>16,1377</b>	<b>0,024335</b>	<b>0,5</b>
	h	16,396	0,02293	4	16,8476	0,020665	1,25
<b>bior2.8</b>	s	<b>15,8497</b>	<b>0,026004</b>	<b>1,5</b>	<b>16,1065</b>	<b>0,02451</b>	<b>0,5</b>
	h	<b>16,4095</b>	<b>0,022859</b>	<b>3,5</b>	<b>16,6952</b>	<b>0,021403</b>	<b>1,25</b>

Tabela III.11: Deslocamentos para amostra do arquivo lf501ls11.wav

<b>bior2.8</b>									
<b>janela</b>	<b>desloc</b>	<b>%</b>	<b>SNRseg</b>	<b>mse</b>	<b>SNRt</b>	<b>%</b>	<b>MSE</b>	<b>snrseg</b>	<b>SNRt</b>
<b>4473</b>	<b>1</b>	2,75	9,06	-	12,41	2,75	0,0574	-	-
	<b>4</b>	4,75	10,76	-	14,03	4,75	0,0395	-	-
	<b>16</b>	5,5	11,72	0,0330	14,81	5,25	0,0330	11,72	14,82
	<b>64</b>	5,25	11,87	-	14,94	5,25	0,0320	-	-
	<b>256</b>	5,25	11,85	-	14,93	5,25	0,0322	-	-
	<b>512</b>	5	10,75	0,0403	13,94	4,75	0,0402	10,73	13,96
<b>8946</b>	<b>1</b>	3,75	10,51	-	13,17	3,75	0,0482	-	-
	<b>4</b>	4	11,97	-	14,44	4	0,0359	-	-
	<b>16</b>	4,5	12,55	0,0322	14,92	4,75	0,0322	12,55	14,93
	<b>64</b>	4,5	12,70	-	15,05	4,5	0,0312	-	-
	<b>256</b>	5	12,22	-	14,66	5	0,0342	-	-
	<b>512</b>	4,5	12,67	0,0315	15,02	4,75	0,0315	15,02	15,02
<b>17891</b>	<b>1</b>	4	11,96	-	13,43	4	0,0454	-	-
	<b>4</b>	3,5	13,02	0,0363	14,40	3,75	0,0360	13,02	14,44
	<b>16</b>	4,75	14,05	-	15,48	4,75	0,0283	-	-
	<b>64</b>	4,5	14,57	0,0250	16,02	4,75	0,0250	14,57	16,02
	<b>256</b>	4,5	14,56	-	16,01	4,5	0,0250	-	-
	<b>512</b>	4,75	14,12	0,0278	15,56	5	0,0278	14,11	15,56
<b>35782</b>	<b>1</b>	3,5	14,33	-	14,45	3,5	0,0359	-	-
	<b>4</b>	3,25	16,17	-	16,31	3,25	0,0234	-	-
	<b>16</b>	3,5	16,08	-	16,14	3,5	0,0243	-	-
	<b>64</b>	3,5	16,78	-	16,91	3,5	0,0204	-	-
	<b>256</b>	3,5	16,79	-	16,91	3,5	0,0204	-	-
	<b>512</b>	3,5	16,80	-	16,92	3,5	0,0203	-	-
<b>71563</b>	<b>1</b>	3,5	14,40	-	-	3,5	0,0363	-	-
	<b>4</b>	3,5	16,38	-	-	3,5	0,0230	-	-
	<b>16</b>	3,25	16,01	-	-	3,25	0,0250	-	-
	<b>64</b>	3,5	16,71	-	-	3,5	0,0213	-	-
	<b>256</b>	3,25	16,92	-	-	3,25	0,0203	-	-
	<b>512</b>	3,25	16,93	-	-	3,25	0,0203	-	-

Tabela III.12: Deslocamentos para amostra do arquivo lf502ls11.wav

<b>bio2.8</b>									
<b>janela</b>	<b>desloc</b>	<b>%</b>	<b>SNRseg</b>	<b>mse</b>	<b>SNRt</b>	<b>%</b>	<b>MSE</b>	<b>snrseg</b>	<b>SNRt</b>
<b>4156</b>	<b>1</b>	2,5	8,54	-	14,08	2,5	0,0391	-	-
	<b>4</b>	2,5	9,69	-	15,09	2,5	0,0310	-	-
	<b>16</b>	2,75	10,15	0,0287	15,42	2,5	0,0287	10,14	15,42
	<b>64</b>	2,5	9,16	0,034787	14,59	2	0,0344	9,1546	14,6379
	<b>256</b>	2,5	9,19	0,034694	14,60	2	0,0343	9,1725	14,6453
	<b>512</b>	2,5	9,23	0,0344	14,64	2	0,0341	9,20	14,67
<b>8311</b>	<b>1</b>	2,5	11,18	-	14,64	2,5	0,0343	-	-
	<b>4</b>	2,5	12,38	-	15,68	2,5	0,0270	-	-
	<b>16</b>	2,5	12,97	-	16,19	2,5	0,0240	-	-
	<b>64</b>	2,5	13,02	-	16,26	2,5	0,0236	-	-
	<b>256</b>	2,25	12,26	-	15,63	2,25	0,0273	-	-
	<b>512</b>	2,25	12,30	-	15,68	2,25	0,0271	-	-
<b>16622</b>	<b>1</b>	1,75	10,71	0,039185	14,07	1,5	0,0389	10,7008	14,1022
	<b>4</b>	2,5	12,61	0,0263	15,79	2,25	0,0262	12,59	15,81
	<b>16</b>	2,5	12,05	0,030104	15,21	2	0,0298	12,0023	15,2617
	<b>64</b>	2,25	13,06	-	16,19	2,25	0,0240	-	-
	<b>256</b>	1,75	11,79	-	15,09	1,75	0,0309	-	-
	<b>512</b>	1,75	11,76	-	15,06	1,75	0,0312	-	-
<b>33243</b>	<b>1</b>	2	14,30	0,038732	14,12	1,25	0,0364	14,2353	14,3875
	<b>4</b>	2,5	16,92	0,02344	16,30	2	0,0216	16,7605	16,6577
	<b>16</b>	2,5	17,56	0,020786	16,82	1,75	0,0188	17,1919	17,267
	<b>64</b>	2,5	17,67	0,02013	16,96	2	0,0182	17,5181	17,398
	<b>256</b>	2,5	17,72	0,019886	17,01	2	0,0179	17,7088	17,2987
	<b>512</b>	2,25	16,26	0,026529	15,76	1,75	0,0244	16,123	16,1235
<b>66486</b>	<b>1</b>	1,25	15,94	-	-	1,25	0,0254	-	-
	<b>4</b>	1,25	18,21	-	-	1,25	0,0151	-	-
	<b>16</b>	1,5	18,66	-	-	1,5	0,0136	-	-
	<b>64</b>	1,25	17,55	-	-	1,25	0,0176	-	-
	<b>256</b>	1,25	18,71	-	-	1,25	0,0135	-	-
	<b>512</b>	1,5	18,74	-	-	1,5	0,0134	-	-

Tabela III.13: Deslocamentos para amostra do arquivo lf503ls11.wav

<b>coif4</b>									
<b>janela</b>	<b>desloc</b>	<b>%</b>	<b>SNRseg</b>	<b>mse</b>	<b>SNRt</b>	<b>%</b>	<b>MSE</b>	<b>snrseg</b>	<b>SNRt</b>
<b>4033</b>	<b>1</b>	5,5	9,94	-	11,75	5,5	0,0668	-	-
	<b>4</b>	6	10,95	0,055736	12,54	5,75	0,0557	10,9493	12,5441
	<b>16</b>	6	11,57	-	13,12	6	0,0487	-	-
	<b>64</b>	6	11,95	-	13,45	6	0,0452	-	-
	<b>256</b>	6	11,97	0,045186	13,45	6,25	0,0452	11,9539	13,4506
	<b>512</b>	6	11,95	0,0453	13,44	6,25	0,0453	11,94	13,44
<b>8065</b>	<b>1</b>	4	10,29	-	12,18	4	0,0605	-	-
	<b>4</b>	4,5	11,00	-	12,76	4,5	0,0529	-	-
	<b>16</b>	4,75	11,32	-	13,05	4,75	0,0496	-	-
	<b>64</b>	5	11,77	-	13,46	5	0,0451	-	-
	<b>256</b>	5	11,88	-	13,56	5	0,0441	-	-
	<b>512</b>	5	11,87	-	13,55	5	0,0441	-	-
<b>16130</b>	<b>1</b>	3,75	12,75	-	12,96	3,75	0,0506	-	-
	<b>4</b>	4,5	13,75	-	13,95	4,5	0,0403	-	-
	<b>16</b>	4,5	14,00	-	14,20	4,5	0,0380	-	-
	<b>64</b>	4,5	14,38	-	14,56	4,5	0,0350	-	-
	<b>256</b>	4,5	14,82	-	14,99	4,5	0,0317	-	-
	<b>512</b>	4,5	14,85	-	15,03	4,5	0,0314	-	-
<b>32259</b>	<b>1</b>	4	12,87	-	13,07	4	0,0493	-	-
	<b>4</b>	4	13,64	-	13,85	4	0,0412	-	-
	<b>16</b>	4	13,96	-	14,16	4	0,0384	-	-
	<b>64</b>	4	14,23	-	14,43	4	0,0360	-	-
	<b>256</b>	4	14,62	-	14,82	4	0,0330	-	-
	<b>512</b>	4	15,00	-	15,21	4	0,0302	-	-
<b>64517</b>	<b>1</b>	4	13,03	-	-	4	0,0497	-	-
	<b>4</b>	3,75	14,61	-	-	3,75	0,0346	-	-
	<b>16</b>	3,75	14,15	-	-	3,75	0,0384	-	-
	<b>64</b>	3,75	14,33	-	-	3,75	0,0369	-	-
	<b>256</b>	4	14,66	-	-	4	0,0342	-	-
	<b>512</b>	4	15,00	-	-	4	0,0316	-	-

Tabela III.14: Deslocamentos para amostra do arquivo lf504ls11.wav

<b>coif4</b>									
<b>janela</b>	<b>desloc</b>	<b>%</b>	<b>SNRseg</b>	<b>mse</b>	<b>SNRt</b>	<b>%</b>	<b>MSE</b>	<b>snrseg</b>	<b>SNRt</b>
<b>3795</b>	<b>1</b>	3,5	5,05	0,10193	9,92	3	0,1001	5,0219	9,9951
	<b>4</b>	4,75	5,98	-	10,83	4,75	0,0826	-	-
	<b>16</b>	4,75	6,49	0,0763	11,18	4,5	0,0762	6,48	11,18
	<b>64</b>	5,25	6,65	0,072983	11,37	5	0,0730	6,6529	11,3697
	<b>256</b>	5	6,51	-	11,24	5	0,0752	-	-
	<b>512</b>	5	6,67	-	11,38	5	0,0727	-	-
<b>7589</b>	<b>1</b>	4,75	7,82	-	10,64	4,75	0,0863	-	-
	<b>4</b>	6	9,34	0,064941	11,87	5,25	0,0641	9,2673	11,9287
	<b>16</b>	5,75	9,67	0,0628	12,02	5	0,0615	9,57	12,11
	<b>64</b>	6	9,69	0,061977	12,08	5,25	0,0607	9,6086	12,171
	<b>256</b>	6	10,02	0,057738	12,39	5,25	0,0568	9,9375	12,4592
	<b>512</b>	6	9,74	0,0612	12,13	5,25	0,0601	9,65	12,21
<b>15178</b>	<b>1</b>	5,25	10,71	0,069597	11,57	4,75	0,0688	10,6884	11,6266
	<b>4</b>	5,5	12,15	0,0536	12,71	4,75	0,0523	12,03	12,82
	<b>16</b>	5,5	12,01	0,054253	12,66	5	0,0537	11,9289	12,6997
	<b>64</b>	5,5	12,89	0,0451	13,46	5	0,0445	12,81	13,52
	<b>256</b>	5,5	12,92	0,044546	13,51	5	0,0441	12,8379	13,557
	<b>512</b>	5,5	12,96	0,0439	13,58	5,25	0,0435	12,93	13,61
<b>30355</b>	<b>1</b>	3,75	11,76	-	12,07	3,75	0,0621	-	-
	<b>4</b>	5	12,77	0,048987	13,10	4,75	0,0489	12,7389	13,105
	<b>16</b>	5	13,45	0,042363	13,73	4,75	0,0420	13,4287	13,7655
	<b>64</b>	5	13,03	0,045821	13,39	4,75	0,0456	12,9962	13,4148
	<b>256</b>	5	13,13	0,044925	13,48	4,75	0,0446	13,1096	13,5115
	<b>512</b>	5	13,49	0,042101	13,76	4,75	0,0415	13,4869	13,8168
<b>60709</b>	<b>1</b>	4,75	11,99	-	-	4,75	0,0633	-	-
	<b>4</b>	4,25	13,61	-	-	4,25	0,0435	-	-
	<b>16</b>	4,5	13,84	-	-	4,5	0,0413	-	-
	<b>64</b>	4,5	13,53	-	-	4,5	0,0444	-	-
	<b>256</b>	4,5	14,05	-	-	4,5	0,0394	-	-
	<b>512</b>	4,5	13,83	-	-	4,5	0,0414	-	-

Tabela III.15: Deslocamentos para amostra do arquivo lf505ls11.wav

<b>coif4</b>									
<b>janela</b>	<b>desloc</b>	<b>%</b>	<b>SNRseg</b>	<b>mse</b>	<b>SNRt</b>	<b>%</b>	<b>MSE</b>	<b>snrseg</b>	<b>SNRt</b>
<b>3315</b>	<b>1</b>	3,75	8,37	0,071402	11,46	3,5	0,0714	8,3651	11,4634
	<b>4</b>	5,5	9,03	0,060966	12,15	4,5	0,0597	8,9746	12,2368
	<b>16</b>	5,75	9,81	0,0522	12,83	5	0,0515	9,76	12,88
	<b>64</b>	5,75	9,74	0,053313	12,73	4,75	0,0523	9,6634	12,8138
	<b>256</b>	6	9,44	0,057155	12,43	4,75	0,0555	9,3561	12,5539
	<b>512</b>	6	9,64	0,0545	12,63	5	0,0532	9,58	12,74
<b>6629</b>	<b>1</b>	3	9,09	-	11,75	3	0,0668	-	-
	<b>4</b>	4	10,43	0,050813	12,94	4,25	0,0507	10,4074	12,9483
	<b>16</b>	4,75	10,72	-	13,25	4,75	0,0473	-	-
	<b>64</b>	4,75	11,12	0,043751	13,59	5	0,0437	11,1216	13,5944
	<b>256</b>	4,75	11,12	0,04367	13,60	5	0,0436	11,1164	13,6019
	<b>512</b>	5	10,62	-	13,16	5	0,0482	-	-
<b>13258</b>	<b>1</b>	3	12,63	-	12,91	3	0,0512	-	-
	<b>4</b>	4,25	13,70	0,0404	13,94	4	0,0404	13,69	13,94
	<b>16</b>	5	13,79	0,039326	14,05	4,5	0,0388	13,7507	14,1109
	<b>64</b>	5	14,68	0,0321	14,94	4,75	0,0319	14,66	14,96
	<b>256</b>	5	14,69	0,032017	14,95	4,75	0,0319	14,6648	14,9601
	<b>512</b>	5	14,70	0,0320	14,95	4,75	0,0319	14,67	14,97
<b>26516</b>	<b>1</b>	3,25	12,42	-	12,65	3,25	0,0543	-	-
	<b>4</b>	3,75	13,97	-	14,21	3,75	0,0379	-	-
	<b>16</b>	4,25	14,67	-	14,91	4,25	0,0323	-	-
	<b>64</b>	4,5	14,30	-	14,57	4,5	0,0349	-	-
	<b>256</b>	4,5	14,77	-	15,02	4,5	0,0315	-	-
	<b>512</b>	4,5	14,81	-	15,06	4,5	0,0312	-	-
<b>53031</b>	<b>1</b>	2,75	13,14	-	-	2,75	0,0485	-	-
	<b>4</b>	3	14,85	-	-	3	0,0327	-	-
	<b>16</b>	3,5	15,07	-	-	3,5	0,0311	-	-
	<b>64</b>	4	14,79	-	-	4	0,0332	-	-
	<b>256</b>	4	15,26	-	-	4	0,0298	-	-
	<b>512</b>	4	15,29	-	-	4	0,0296	-	-



Tabela III.16: Deslocamentos para amostra do arquivo lm101ls11.wav

<b>dmey</b>									
<b>janela</b>	<b>desloc</b>	<b>%</b>	<b>SNRseg</b>	<b>mse</b>	<b>SNRt</b>	<b>%</b>	<b>MSE</b>	<b>snrseg</b>	<b>SNRt</b>
<b>4686</b>	<b>1</b>	5	8,25	-	12,67	5	0,0541	-	-
	<b>4</b>	5,5	9,60	-	13,70	5	0,0426	-	-
	<b>16</b>	5,5	9,59	-	13,69	5,5	0,0428	-	-
	<b>64</b>	5,75	10,03	-	14,04	5,75	0,0395	-	-
	<b>256</b>	6	10,02	-	14,03	6	0,0396	-	-
	<b>512</b>	5,75	10,11	-	14,09	5,75	0,0390	-	-
<b>7442</b>	<b>1</b>	4,75	8,98	0,052977	12,76	5,25	0,0530	8,969	12,7605
	<b>4</b>	4,75	10,00	0,043543	13,61	5,25	0,0433	9,9801	13,6334
	<b>16</b>	5	10,63	0,0383	14,17	5,25	0,0382	10,62	14,18
	<b>64</b>	5	10,33	0,040957	13,88	5,25	0,0409	10,318	13,8872
	<b>256</b>	4,75	10,66	0,038381	14,16	5,25	0,0382	10,6463	14,1833
	<b>512</b>	5	10,41	0,0403	13,95	5,25	0,0401	10,40	13,96
<b>14884</b>	<b>1</b>	3,25	10,78	0,050311	12,98	3,75	0,0497	10,746	13,0359
	<b>4</b>	3,5	11,86	0,0402	13,96	3,75	0,0398	11,85	14,00
	<b>16</b>	4	11,93	0,0399	13,99	4,5	0,0394	11,8514	14,0427
	<b>64</b>	4	11,61	0,0421	13,75	4,75	0,0414	11,53	13,83
	<b>256</b>	4	11,97	0,039319	14,05	5	0,0383	11,8415	14,1627
	<b>512</b>	4	12,22	0,0370	14,31	5	0,0363	12,06	14,40
<b>29768</b>	<b>1</b>	3,5	14,90	0,034837	14,58	3,25	0,0346	14,8341	14,6116
	<b>4</b>	3,5	16,15	0,026515	15,77	3,25	0,0258	16,1426	15,8759
	<b>16</b>	3,75	15,84	0,028777	15,41	3,25	0,0279	15,7551	15,5517
	<b>64</b>	3,25	16,64	0,023139	16,36	3	0,0229	16,5631	16,3967
	<b>256</b>	3,5	15,91	0,027642	15,58	3,25	0,0274	15,8435	15,6251
	<b>512</b>	3,5	16,35	0,025083	16,01	3,25	0,0247	16,2982	16,0748
<b>59536</b>	<b>1</b>	2	15,03	-	-	2	0,0314	-	-
	<b>4</b>	2,5	16,06	-	-	2,5	0,0248	-	-
	<b>16</b>	2,5	16,95	-	-	2,5	0,0202	-	-
	<b>64</b>	2,5	16,68	-	-	2,5	0,0215	-	-
	<b>256</b>	2,5	15,85	-	-	2,5	0,0260	-	-
	<b>512</b>	2,5	16,21	-	-	2,5	0,0239	-	-

Tabela III.17: Deslocamentos para amostra do arquivo lm102ls11.wav

<b>bio2.8</b>									
<b>janela</b>	<b>desloc</b>	<b>%</b>	<b>SNRseg</b>	<b>mse</b>	<b>SNRt</b>	<b>%</b>	<b>MSE</b>	<b>snrseg</b>	<b>SNRt</b>
<b>4342</b>	<b>1</b>	3,75	7,59	0,062067	12,07	3,25	0,0614	7,5471	12,1158
	<b>4</b>	4,25	7,94	0,056407	12,49	3,75	0,0558	7,9128	12,5355
	<b>16</b>	4,75	8,38	0,0526	12,79	4	0,0517	8,31	12,86
	<b>64</b>	5	8,52	0,052316	12,81	4,25	0,0515	8,4761	12,8813
	<b>256</b>	4,75	7,82	0,05968	12,24	4	0,0590	7,7084	12,2891
	<b>512</b>	5,25	7,89	0,0593	12,27	4,25	0,0581	7,80	12,36
<b>8683</b>	<b>1</b>	3,75	9,44	0,059207	12,28	3,25	0,0588	9,2952	12,3053
	<b>4</b>	4,5	10,29	0,048238	13,17	3,25	0,0458	10,1521	13,3936
	<b>16</b>	4,5	10,55	0,0447	13,49	3,75	0,0436	10,46	13,61
	<b>64</b>	4,5	10,76	0,043327	13,63	3,75	0,0422	10,6696	13,752
	<b>256</b>	4,75	10,80	0,043755	13,59	3,75	0,0419	10,6995	13,7798
	<b>512</b>	4,25	10,29	0,0485	13,15	3,5	0,0468	10,13	13,30
<b>17365</b>	<b>1</b>	3,5	13,37	0,043856	13,58	3	0,0430	13,2407	13,6667
	<b>4</b>	3,75	14,91	0,0314	15,03	3,5	0,0311	14,83	15,08
	<b>16</b>	3,5	15,49	0,027541	15,60	3,25	0,0271	15,4424	15,6705
	<b>64</b>	3,5	15,56	0,0271	15,67	3,25	0,0267	15,49	15,73
	<b>256</b>	3,75	15,54	0,027975	15,53	3,25	0,0270	15,4294	15,6815
	<b>512</b>	3,75	15,54	0,0280	15,53	3,25	0,0270	15,45	15,69
<b>34729</b>	<b>1</b>	3	13,45	0,041106	13,86	2,75	0,0409	13,3794	13,8821
	<b>4</b>	3,25	15,10	0,028564	15,44	3	0,0281	15,0964	15,5157
	<b>16</b>	3	15,33	0,026979	15,69	2,75	0,0269	15,2297	15,6978
	<b>64</b>	3	15,39	0,026634	15,75	2,75	0,0265	15,3017	15,764
	<b>256</b>	3	15,60	0,025571	15,92	2,75	0,0254	15,5167	15,9554
	<b>512</b>	3	15,60	0,025623	15,91	2,75	0,0254	15,5142	15,9507
<b>69458</b>	<b>1</b>	2,5	13,99	-	-	2,5	0,0399	-	-
	<b>4</b>	2,5	15,36	-	-	2,5	0,0291	-	-
	<b>16</b>	2,5	16,20	-	-	2,5	0,0240	-	-
	<b>64</b>	2,25	16,31	-	-	2,25	0,0234	-	-
	<b>256</b>	2,25	16,31	-	-	2,25	0,0234	-	-
	<b>512</b>	2,5	16,30	-	-	2,5	0,0234	-	-

Tabela III.18: Deslocamentos para amostra do arquivo lm103ls11.wav

<b>db4</b>									
<b>janela</b>	<b>desloc</b>	<b>%</b>	<b>SNRseg</b>	<b>mse</b>	<b>SNRt</b>	<b>%</b>	<b>MSE</b>	<b>snrseg</b>	<b>SNRt</b>
<b>3790</b>	<b>1</b>	4,75	9,56	-	13,45	4,75	0,0452	-	-
	<b>4</b>	4,75	10,75	0,037148	14,30	4,5	0,0370	10,7326	14,3167
	<b>16</b>	4,5	10,83	-	14,43	4,5	0,0360	-	-
	<b>64</b>	4,5	11,38	-	14,88	4,5	0,0325	-	-
	<b>256</b>	4,75	11,17	-	14,69	4,75	0,0340	-	-
	<b>512</b>	4,5	11,40	-	14,87	4,5	0,0326	-	-
<b>7580</b>	<b>1</b>	5	12,47	0,038218	14,18	4,5	0,0376	12,4407	14,2506
	<b>4</b>	5	13,85	0,029817	15,26	4,5	0,0292	13,7965	15,3536
	<b>16</b>	4,5	14,42	0,0251	16,00	4,25	0,0251	14,38	16,00
	<b>64</b>	5	14,25	0,026301	15,80	4,75	0,0262	14,22	15,8158
	<b>256</b>	4,75	14,54	0,024976	16,02	4,25	0,0248	14,4758	16,0514
	<b>512</b>	5	14,26	0,0265	15,64	4,5	0,0263	14,20	15,79
<b>15160</b>	<b>1</b>	3,75	15,06	-	15,23	3,75	0,0300	-	-
	<b>4</b>	4,25	16,83	-	16,99	4,25	0,0200	-	-
	<b>16</b>	4	17,42	0,0175	17,57	3,75	0,0174	17,4161	17,5922
	<b>64</b>	4	16,79	-	16,97	4	0,0201	-	-
	<b>256</b>	3,75	17,51	-	17,69	3,75	0,0170	-	-
	<b>512</b>	4	17,50	0,0172	17,65	3,75	0,0171	17,50	17,68
<b>30319</b>	<b>1</b>	3,75	14,94	0,031575	15,01	3,5	0,0314	14,9191	15,0297
	<b>4</b>	3,75	17,02	0,019551	17,09	3,5	0,0195	16,9855	17,0927
	<b>16</b>	3,5	17,29	-	17,40	3,5	0,0182	-	-
	<b>64</b>	3,5	17,49	0,01753	17,56	3,25	0,0175	17,4639	17,5701
	<b>256</b>	3,5	17,26	-	17,35	3,5	0,0184	-	-
	<b>512</b>	3,5	17,44	0,017762	17,51	3,25	0,0177	17,4194	17,5216
<b>60638</b>	<b>1</b>	2,5	15,68	-	-	2,5	0,0270	-	-
	<b>4</b>	2,75	17,39	-	-	2,75	0,0182	-	-
	<b>16</b>	2,75	17,99	-	-	2,75	0,0159	-	-
	<b>64</b>	2,5	18,11	-	-	2,5	0,0155	-	-
	<b>256</b>	2,5	17,59	-	-	2,5	0,0174	-	-
	<b>512</b>	2,5	17,73	-	-	2,5	0,0169	-	-

Tabela III.19: Deslocamentos para amostra do arquivo lm104ls11.wav

<b>bio2.8</b>									
<b>janela</b>	<b>desloc</b>	<b>%</b>	<b>SNRseg</b>	<b>mse</b>	<b>SNRt</b>	<b>%</b>	<b>MSE</b>	<b>snrseg</b>	<b>SNRt</b>
<b>3515</b>	<b>1</b>	3,5	8,26	-	11,73	3,5	0,0671	-	-
	<b>4</b>	4,75	9,95	-	13,27	4,75	0,0471	-	-
	<b>16</b>	4,75	10,41	-	13,60	4,75	0,0436	-	-
	<b>64</b>	4,75	10,24	-	13,48	4,75	0,0449	-	-
	<b>256</b>	4,75	10,47	-	13,66	4,75	0,0431	-	-
	<b>512</b>	4,75	10,25	-	13,49	4,75	0,0448	-	-
<b>7029</b>	<b>1</b>	3,25	9,57	-	12,18	3,25	0,0606	-	-
	<b>4</b>	4	11,33	0,043196	13,65	4,25	0,0429	11,3315	13,6721
	<b>16</b>	4	11,66	0,0412	13,85	4,5	0,0412	11,62	13,86
	<b>64</b>	4	11,79	0,040233	13,95	4,25	0,0401	11,7691	13,9638
	<b>256</b>	4	11,78	0,040428	13,93	4,25	0,0403	11,7615	13,942
	<b>512</b>	4	11,79	0,0404	13,94	4,25	0,0403	11,77	13,95
<b>14057</b>	<b>1</b>	4,5	13,39	0,04452	13,51	4	0,0437	13,3613	13,599
	<b>4</b>	3,75	15,06	0,0297	15,27	3,5	0,0295	15,01	15,30
	<b>16</b>	3,75	15,55	0,026621	15,75	3,5	0,0266	15,4802	15,7529
	<b>64</b>	4	15,89	0,0252	15,99	3,75	0,0248	15,89	16,06
	<b>256</b>	4	15,65	0,026422	15,78	3,75	0,0260	15,6448	15,8473
	<b>512</b>	4	15,87	0,0254	15,95	3,75	0,0250	15,86	16,03
<b>28114</b>	<b>1</b>	3,5	13,22	0,04492	13,48	3	0,0445	13,1403	13,5192
	<b>4</b>	3,75	14,64	0,032399	14,89	3,5	0,0321	14,6158	14,9328
	<b>16</b>	3,25	15,41	-	15,68	3,25	0,0270	-	-
	<b>64</b>	3,25	15,43	0,027075	15,67	3	0,0268	15,3651	15,7125
	<b>256</b>	3,5	15,49	0,027245	15,65	3	0,0265	15,412	15,7692
	<b>512</b>	3,5	14,94	0,030455	15,16	3,25	0,0301	14,9	15,2135
<b>56228</b>	<b>1</b>	3,25	14,11	-	-	3,25	0,0388	-	-
	<b>4</b>	2,5	16,20	-	-	2,5	0,0240	-	-
	<b>16</b>	2,25	16,10	-	-	2,25	0,0246	-	-
	<b>64</b>	2,25	16,16	-	-	2,25	0,0242	-	-
	<b>256</b>	2	16,76	-	-	2	0,0211	-	-
	<b>512</b>	2	16,75	-	-	2	0,0211	-	-

Tabela III.20: Deslocamentos para amostra do arquivo lm105ls11.wav

<b>coif4</b>									
<b>janela</b>	<b>desloc</b>	<b>%</b>	<b>SNRseg</b>	<b>mse</b>	<b>SNRt</b>	<b>%</b>	<b>MSE</b>	<b>snrseg</b>	<b>SNRt</b>
<b>4293</b>	<b>1</b>	5,25	6,73	0,061573	12,11	4,25	0,0599	6,5541	12,2236
	<b>4</b>	4,5	7,64	0,048849	13,11	3,75	0,0485	7,5282	13,1402
	<b>16</b>	4,75	8,14	0,0453	13,44	4,5	0,0451	8,12	13,45
	<b>64</b>	4,75	8,44	0,043089	13,66	4,5	0,0429	8,4236	13,6704
	<b>256</b>	4,75	7,89	0,047348	13,25	4,25	0,0470	7,8398	13,2824
	<b>512</b>	4,75	8,06	0,0458	13,39	4,5	0,0456	8,03	13,41
<b>8586</b>	<b>1</b>	3,75	6,89	0,063039	12,00	3,5	0,0629	6,8502	12,0117
	<b>4</b>	4	8,12	0,04989	13,02	3,75	0,0495	8,1182	13,0576
	<b>16</b>	4	8,20	0,0488	13,11	3,75	0,0486	8,19	13,13
	<b>64</b>	4	8,47	0,046804	13,30	3,75	0,0466	8,4622	13,3151
	<b>256</b>	3,75	8,75	-	13,51	3,75	0,0446	-	-
	<b>512</b>	3,75	8,16	0,0491	13,09	3,5	0,0490	8,14	13,10
<b>17172</b>	<b>1</b>	3	7,23	-	12,31	3	0,0587	-	-
	<b>4</b>	3,25	8,17	-	13,08	3,25	0,0492	-	-
	<b>16</b>	3	8,72	-	13,47	3	0,0449	-	-
	<b>64</b>	3	8,24	-	13,12	3	0,0487	-	-
	<b>256</b>	3	8,76	-	13,51	3	0,0446	-	-
	<b>512</b>	3	8,76	-	13,51	3	0,0446	-	-
<b>34343</b>	<b>1</b>	1,75	14,29	0,045371	13,43	3	0,0409	13,9598	13,8821
	<b>4</b>	2,5	15,80	0,028546	15,44	3	0,0273	15,7372	15,6428
	<b>16</b>	2,25	16,02	0,028519	15,45	3	0,0265	15,9028	15,7709
	<b>64</b>	1,75	15,42	0,034917	14,57	2,75	0,0304	15,2821	15,176
	<b>256</b>	2,5	16,08	0,026956	15,69	3	0,0260	15,9884	15,8542
	<b>512</b>	2,5	16,35	0,024749	16,06	3	0,0239	16,284	16,2146
<b>68685</b>	<b>1</b>	1,5	14,87	-	-	1,5	0,0326	-	-
	<b>4</b>	1,5	16,59	-	-	1,5	0,0219	-	-
	<b>16</b>	1,25	16,47	-	-	1,25	0,0225	-	-
	<b>64</b>	1,25	17,08	-	-	1,25	0,0196	-	-
	<b>256</b>	1,25	16,51	-	-	1,25	0,0224	-	-
	<b>512</b>	1,25	16,87	-	-	1,25	0,0206	-	-

Tabela III.21: Escolha da wavelet-mãe para arquivo fernandaAfogamento.wav

wmãe		4				8				16			
		PESQ	%	MSE	%	PESQ	%	MSE	%	PESQ	%	MSE	%
db4	s	2,291	3,25	0,042176	2,25	2,304	3,25	0,041431	2	2,293	2,5	0,056126	0,5
	h	2,215	4,25	0,040293	5,75	2,251	4,25	0,036353	5,25	2,275	4	0,050951	1,25
coif4	s	2,371	3,5	0,039874	2,25	2,394	3,5	0,038468	2	2,396	3	0,049404	0,5
	h	2,259	4,25	0,038176	5,25	2,294	4,25	0,034206	5,25	2,353	3,75	0,044887	1,75
sym8	s	2,364	3,75	0,040034	2,25	2,392	3,75	0,038711	2	2,39	2,75	0,053832	0,25
	h	2,25	4,25	0,038404	5,5	2,3	4,25	0,033987	5,25	2,348	3,5	0,04763	1,25
dmey	s	2,433	3,25	0,037403	2,25	2,456	3,25	0,035399	2,25	<b>2,467</b>	<b>3,25</b>	0,041098	1
	h	2,318	4	0,036245	5,5	2,341	4	<b>0,03323</b>	<b>5,5</b>	2,418	4	0,034323	3,25
bior2.8	s	2,29	3	0,040525	1,75	2,307	2,75	0,039366	1,75	2,287	2,25	0,05106	0,25
	h	2,184	4	0,042356	4,5	2,222	4	0,038818	4,25	2,273	3,5	0,045216	1

Tabela III.22: Escolha wavelet-mãe para o arquivo flaviaDivas.wav

wmãe		4				8				16			
		PESQ	%	MSE	%	PESQ	%	MSE	%	PESQ	%	MSE	%
db4	s	2,449	6,75	0,033034	3,25	2,516	5	0,03078	1,75	2,494	4	0,038677	0,25
	h	2,344	7,25	0,025834	8,5	2,413	5	0,023421	5,25	2,443	5,75	0,034532	0,75
coif4	s	2,458	6,75	0,032152	3	2,534	5	0,029316	2	2,53	4,5	0,035832	0,25
	h	2,355	7,25	0,025374	8,25	2,412	4,75	0,022695	5,25	2,441	5,75	0,030542	0,5
sym8	s	2,456	6,75	0,032195	3	2,536	4,75	0,029577	1,75	2,534	3,75	0,038587	0,25
	h	2,351	7	0,025377	8,5	2,429	4,75	0,02282	5	2,466	4,75	0,030515	0,5
dmey	s	2,478	6,25	0,030273	3	2,522	4,5	0,027095	2,25	<b>2,586</b>	<b>4</b>	0,030057	1
	h	2,348	6	0,025222	7,75	2,355	4,25	<b>0,023609</b>	<b>5,5</b>	2,46	4	0,024329	2,75
bior2.8	s	2,468	5	0,030507	2	2,523	3,75	0,028366	1,5	2,51	3,25	0,038884	0,25
	h	2,319	5,5	0,026328	5,5	2,374	4,25	0,023909	3,75	2,401	4,25	0,030669	0,5

Tabela III.23: Escolha da wavelet-mãe para arquivo izumiPoesia.wav

wmãe		4				8				16			
		PESQ	%	MSE	%	PESQ	%	MSE	%	PESQ	%	MSE	%
db4	s	2,828	3,25	0,042319	1,75	2,862	3,25	0,041436	1,5	2,863	2,25	0,049811	0,25
	h	2,682	4	0,038308	4,75	2,732	4,25	0,035907	4,5	2,838	3,75	0,044804	1
coif4	s	2,894	3,25	0,041011	1,75	2,932	3,25	0,039642	1,75	2,878	2,5	0,04633	0,25
	h	2,748	4	0,0373	4,75	2,776	4	0,035245	4,75	2,813	3,75	0,03947	1
sym8	s	2,904	3,25	0,041072	1,75	<b>2,943</b>	<b>3,25</b>	0,040032	1,75	2,909	2	0,046499	0,25
	h	2,759	4,25	0,037267	4,75	2,799	4	0,034895	4,5	2,897	3,5	0,040945	0,75
dmey	s	2,911	3,25	0,039201	1,75	2,929	3,25	0,037267	2	2,942	2,75	0,041237	0,75
	h	2,767	3,75	0,037276	4,75	2,784	3,75	<b>0,035386</b>	<b>4,75</b>	2,857	3,5	0,036307	2,5
bior2.8	s	2,899	3,25	0,039828	1,5	2,938	3,25	0,03909	1,5	2,91	2	0,044944	0,25
	h	2,725	4	0,039073	4,25	2,762	4	0,036967	3,75	2,879	3,5	0,039632	0,75

Tabela III.24: Escolha da wavelet-mãe para arquivo fredMagoa.wav

wmãe		4				8				16			
		PESQ	%	MSE	%	PESQ	%	MSE	%	PESQ	%	MSE	%
db4	s	2,459	4,75	0,056998	2,5	2,491	4,25	0,056045	2	2,474	2,5	0,064168	0,25
	h	2,338	6,25	0,048776	6,5	2,423	6	0,044903	5,25	2,45	3,5	0,056201	0,5
coif4	s	2,5	4,75	0,055306	2,5	2,533	4,25	0,053069	2,25	2,533	2,25	0,063662	0,25
	h	2,378	5,5	0,048091	6,25	2,43	5,5	<b>0,04411</b>	<b>5,75</b>	2,517	3,75	0,054125	1
sym8	s	2,501	4,75	0,055385	2,5	2,543	4,25	0,053483	2	2,504	2,5	0,061422	0,25
	h	2,371	5,75	0,047972	6,25	2,451	5,5	0,043742	5,25	2,457	3,75	0,051927	0,5
dmey	s	2,499	4,25	0,053789	2,5	2,531	4	0,05073	2,5	<b>2,564</b>	<b>3</b>	0,054503	0,75
	h	2,38	5,25	0,050359	6,25	2,413	5,25	0,046222	6	2,536	4,75	0,046443	2
bior2.8	s	2,48	4,5	0,052493	2	2,514	4,25	0,051291	1,75	2,46	2,5	0,059589	0,25
	h	2,347	5,25	0,04762	5,25	2,399	5,25	0,044337	4,5	2,404	4	0,049282	0,5

Tabela III.25: Escolha da wavelet-mãe para arquivo ortis-grua.wav

wmãe		4				8				16			
		PESQ	%	MSE	%	PESQ	%	MSE	%	PESQ	%	MSE	%
db4	s	2,597	3	0,037205	1,5	2,617	2,75	0,036657	1,25	2,604	2,75	0,045471	0,25
	h	2,503	3,25	0,031264	3,75	2,525	3,25	0,029654	3,5	2,529	3,5	0,042718	0,75
coif4	s	2,618	2,75	0,037095	1,5	2,655	2,75	0,035526	1,5	2,646	2,5	0,042699	0,25
	h	2,518	3	0,03223	3,5	2,543	3	<b>0,0302</b>	<b>3,5</b>	2,578	3,25	0,039226	1
sym8	s	2,617	2,75	0,03711	1,5	2,66	2,75	0,035834	1,25	2,648	2,5	0,044667	0,25
	h	2,515	3	0,032059	3,5	2,547	3	0,029838	3,5	2,597	3,25	0,041352	0,5
dmey	s	2,619	2,75	0,036709	1,25	2,653	2,75	0,034772	1,5	<b>2,669</b>	<b>2,5</b>	0,039081	0,75
	h	2,517	2,75	0,034032	3,25	2,544	3	0,031972	3,5	2,575	2,5	0,033492	2
bior2.8	s	2,629	2,5	0,03496	1	2,658	2,5	0,034155	1	2,641	2	0,043252	0,25
	h	2,52	2,75	0,030883	2,75	2,542	2,75	<b>0,02957</b>	<b>2,75</b>	2,574	3	0,038973	0,5

Tabela III.26: Escolha da wavelet-mãe para arquivo tiagoPalocci.wav

wmãe		4				8				16			
		PESQ	%	MSE	%	PESQ	%	MSE	%	PESQ	%	MSE	%
db4	s	2,612	3,5	0,037861	1,75	2,637	3,25	0,036634	1,5	2,625	2,75	0,045455	0,25
	h	2,461	4,5	0,030036	4,25	2,497	4,25	0,028272	4	2,558	4	0,041459	0,75
coif4	s	2,635	3	0,036489	1,5	2,654	3	0,034973	1,5	2,661	2,25	0,043713	0,25
	h	2,48	3,5	0,029302	4,25	2,503	3,5	0,027517	4	2,635	3,25	0,039537	0,75
sym8	s	2,625	3	0,036564	1,75	2,658	3	0,03517	1,5	2,658	2	0,043356	0,25
	h	2,476	4	0,02934	4,25	2,522	4,25	<b>0,0274</b>	<b>3,75</b>	2,623	3,25	0,038752	0,75
dmey	s	2,64	3	0,034818	1,75	2,655	3	0,032872	1,75	<b>2,683</b>	<b>2,75</b>	0,036233	1
	h	2,493	3,5	0,029597	4	2,511	3,5	0,027514	4	2,581	3,5	0,028585	2,5
bior2.8	s	2,612	3,75	0,035482	1,75	2,634	3,5	0,034659	1,5	2,627	2,5	0,04301	0,25
	h	2,471	4,25	0,029434	4,25	2,502	4,25	0,028074	4	2,607	3,75	0,038005	0,5



Tabela III.27: Deslocamentos para amostra do arquivo fernandaAfogamento.wav

<b>dmey</b>							
<b>janela</b>	<b>deslocs</b>	<b>PESQ</b>	<b>mse</b>	<b>%</b>	<b>MSE</b>	<b>pesq</b>	<b>%</b>
<b>83252</b>	1	2,331	-	0,75	0,039292	-	0,75
	4	2,581	0,19934	5,75	0,037543	2,297	1
	16	2,667	0,24713	6	0,041097	2,258	0,75
	64	2,58	0,19464	6,25	0,041098	2,26	1
	256	2,692	0,2078	6,5	0,038985	2,28	1
	512	2,467	0,092851	3,25	0,041098	2,291	1
<b>41626</b>	1	2,25	0,052235	0,75	0,050751	2,248	1
	4	2,511	0,29631	6,5	0,043399	2,253	0,75
	16	2,676	0,17951	6	0,039731	2,302	1,25
	64	2,657	0,20086	6,5	0,039684	2,276	1
	256	2,643	0,21799	7	0,040476	2,273	1
	512	2,677	0,22721	7,25	0,039367	2,279	1
<b>20813</b>	1	2,232	-	0,75	0,054538	-	0,75
	4	2,495	0,31265	7,25	0,043038	2,284	0,75
	16	2,597	0,18235	5,75	0,043699	2,263	1
	64	2,67	0,19985	6	0,040685	2,287	1
	256	2,642	0,21296	6,5	0,041425	2,286	1
	512	2,505	2,505	6,25	0,042477	2,273	1
<b>10407</b>	1	2,209	0,065642	0,25	0,06028	2,102	0,5
	4	2,392	0,2852	7	0,046231	2,294	1
	16	2,602	0,24716	7	0,042315	2,279	1
	64	2,596	0,22404	7	0,043279	2,282	1
	256	2,431	0,15933	5	0,043265	2,269	1
	512	2,555	0,22934	7	0,042718	2,276	1
<b>5204</b>	1	2,2	0,075099	0,25	0,073938	2,001	0,5
	4	2,354	0,18196	4,5	0,052009	2,236	0,75
	16	2,412	0,1365	3,5	0,49325	2,274	1
	64	2,436	0,3766	10,75	0,046159	2,269	1
	256	2,371	0,42032	12,25	0,048521	2,255	1
	512	2,424	0,41495	12	0,047596	2,264	1

Tabela III.28: Deslocamentos para amostra do arquivo fredMagoa.wav

<b>dmey</b>							
<b>janela</b>	<b>deslocs</b>	<b>PESQ</b>	<b>mse</b>	<b>%</b>	<b>MSE</b>	<b>pesq</b>	<b>%</b>
<b>81869</b>	1	2,311	0,082878	0,5	0,059183	2,287	0,25
	4	2,428	0,13753	1,75	0,053995	2,313	0,5
	16	2,454	0,20868	2,75	0,056055	2,288	0,5
	64	2,485	0,21503	3,75	0,054503	2,312	0,75
	256	2,503	0,19697	3,75	0,054803	2,298	0,75
	512	2,564	0,16689	3	0,054503	2,332	0,75
<b>40935</b>	1	2,208	0,080065	0,25	0,079668	2,207	0,5
	4	2,385	0,07259	1	0,058025	2,28	0,5
	16	2,48	0,19667	3,25	0,05747	2,299	0,75
	64	2,477	0,19995	3,5	0,056705	2,297	0,75
	256	2,492	0,19348	3,75	0,055968	0,055968	0,75
	512	2,49	0,21097	4	0,055789	2,289	0,75
<b>20468</b>	1	2,304	-	0,5	0,067262	-	0,5
	4	2,459	0,08661	1,5	0,059411	2,356	0,75
	16	2,465	0,18982	3,5	0,056373	2,316	0,75
	64	2,48	0,16771	3,25	0,056765	2,302	0,75
	256	2,495	0,19707	3,75	0,056326	2,302	0,75
	512	485	0,18464	3,5	0,056187	2,302	0,75
<b>10234</b>	1	2,223	-	0,5	0,075288	-	0,5
	4	2,433	0,11894	2	0,060341	2,328	0,75
	16	2,44	0,18328	3,25	0,058494	2,307	0,75
	64	2,482	0,14983	3	0,056657	2,307	0,75
	256	2,474	0,16864	3,25	0,056459	2,302	0,75
	512	2,475	0,17037	3,25	0,05662	2,307	0,75
<b>5117</b>	1	2,211	-	0,5	0,075574	-	0,5
	4	2,339	0,070217	1	0,065847	2,311	0,75
	16	2,446	0,13414	2,25	0,06235	2,314	0,75
	64	2,452	0,1389	2,5	0,060207	2,305	0,75
	256	2,452	0,13862	2,5	0,060295	2,306	0,75
	512	2,454	0,13858	2,5	0,060286	2,307	0,75

Tabela III.29: Análise de vários sinais de um locutor masculino

Sinal	thresh	wmãe	qtd jan	%	MSE	Sinal	thresh	wmãe	qtd jan	%	MSE
<b>lm101ls11</b>	s	bior2.8	37	1,25	0,0455	<b>lm101ls21</b>	s	bior2.8	42	1,25	0,0456
	h	coif4		3,75	0,0511		h	coif4		4,25	0,0507
<b>lm101ls12</b>	s	bior2.8	34	1,25	0,0492	<b>lm101ls22</b>	s	bior2.8	38	1,25	0,0439
	h	coif4		3,75	0,0561		h	coif4		4	0,0511
<b>lm101ls13</b>	s	bior2.8	40	1,25	0,0512	<b>lm101ls23</b>	s	bior2.8	41	1,25	0,0460
	h	coif4		3,75	0,0587		h	coif4		4,25	0,0518
<b>lm101ls14</b>	s	bior2.8	37	1,25	0,0500	<b>lm101ls24</b>	s	bior2.8	38	1,25	0,0445
	h	coif4		3,25	0,0618		h	coif4		4	0,0505
<b>lm101ls15</b>	s	bior2.8	41	1,25	0,0460	<b>lm101ls25</b>	s	bior2.8	41	1,25	0,0430
	h	coif4		4	0,0515		h	coif4		4	0,0468
<b>lm101ls16</b>	s	bior2.8	41	1	0,0493	<b>lm101ls26</b>	s	bior2.8	40	1	0,0455
	h	coif4		3,5	0,0576		h	coif4		3,75	0,0522
<b>lm101ls17</b>	s	bior2.8	42	1,25	0,0468	<b>lm101ls27</b>	s	bior2.8	35	1,25	0,0472
	h	coif4		3,5	0,0555		h	coif4		3,75	0,0547
<b>lm101ls18</b>	s	bior2.8	41	1,25	0,0449	<b>lm101ls28</b>	s	bior2.8	38	1,25	0,0463
	h	coif4		4	0,0525		h	coif4		3,75	0,0529
<b>lm101ls19</b>	s	bior2.8	42	1,25	0,0427	<b>lm101ls29</b>	s	bior2.8	38	1,25	0,0419
	h	coif4		4,5	0,0483		h	coif4		4,25	0,0463
<b>lm101ls20</b>	s	bior2.8	41	1,25	0,0436	<b>lm101ls30</b>	s	bior2.8	38	1,25	0,0469
	h	coif4		4,5	0,0473		h	coif4		4	0,0523