



Universidade Federal de Uberlândia

Programa de pós-graduação em Ciência
Faculdade de Engenharia Elétrica

**REDUÇÃO DE RUÍDO EM SINAIS DE VOZ
COMBINANDO FILTRO DE KALMAN E
TRANSFORMADA WAVELET.**

JUCELINO CARDOSO MARCIANO DOS SANTOS

MESTRADO

2015

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

S237r Santos, Jucelino Cardoso Marciano dos, 1979-
2015 Redução de ruído em sinais de voz combinando filtro de Kalman e
transformada Wavelet / Jucelino Cardoso Marciano dos Santos. - 2015.
98 f. : il.

Orientador: Gilberto Arantes Carrijo.
Dissertação (mestrado) - Universidade Federal de Uberlândia,
Programa de Pós-Graduação em Engenharia Elétrica.
Inclui bibliografia.

1. Engenharia elétrica - Teses. 2. Kalman, Filtragem de - Teses. 3.
Wavelets (Matemática) - Teses. 4. Algoritmos genéticos - Teses. I.
Carrijo, Gilberto Arantes, 1948- II. Universidade Federal de Uberlândia,
Programa de Pós-Graduação em Engenharia Elétrica. III. Título.

CDU: 621.3

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE ENGENHARIA ELÉTRICA
PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA



**REDUÇÃO DE RUÍDO EM SINAIS DE VOZ COMBINANDO FILTRO
DE KALMAN E TRANSFORMADA WAVELET**

JUCELINO CARDOSO MARCIANO DOS SANTOS

Uberlândia

2015

JUCELINO CARDOSO MARCIANO DOS SANTOS

**REDUÇÃO DE RUÍDO EM SINAIS DE VOZ COMBINANDO FILTRO
DE KALMAN E TRANSFORMADA WAVELET.**

Dissertação de Mestrado apresentada à
Faculdade de Engenharia Elétrica da
Universidade Federal de Uberlândia,
como exigência parcial para obtenção do
Título de Mestre em Engenharia Elétrica.

Área de concentração: Processamento de Sinais.

Orientador: Prof. Gilberto Arantes Carrijo, PhD (UFU)

Uberlândia

2015

DEDICATÓRIA

Dedico esse trabalho aos meus amados pais por ter me concedido o dom da vida. E por ter me ensinado os valores morais e sociais que homem deve ter.

A minha amada esposa Cristiane e os meus filhos Laura, Arthur e Heloisa por ser a minha fonte inesgotável de superação nos momentos difíceis.

AGRADECIMENTOS

Ao meu orientador e professor Gilberto Arantes Carrijo pelos ensinamentos e pela confiança depositada em mim para a realização deste trabalho.

Agradeço também ao grande amigo Leandro Aureliano do programa de pós-graduação da UFU pelos ensinamentos e por dividir seus conhecimentos e pelo incentivo na realização deste trabalho.

A professora Edna Lucia Flores por ter iluminado os caminhos tortuosos das disciplinas concluídas no curso de mestrado da UFU.

Aos colegas matemáticos Ricardo e Eduardo Vasconcelos por dividirem os problemas e também as soluções.

A Cristiane de Fátima dos Santos Cardoso pela paciência, compressão, colaboração e também pela amizade e o amor em nossa parceria de mais de 10 anos nos estudos na vida familiar e também em todas as nossas conquistas ao longo desses anos.

Ao CNPq pelo auxílio financeiro estritamente necessário para a realização desta dissertação de mestrado.

Santos, Jucelino C. M. **REDUÇÃO DE RUÍDO EM SINAIS DE VOZ COMBINANDO FILTRO DE KALMAN E TRANSFORMADA WAVELET**. Dissertação de Mestrado. Universidade Federal de Uberlândia. Uberlândia – MG.

RESUMO

Este trabalho tem como objetivo apresentar, combinar e comparar técnicas de redução de ruído aplicadas em sinais de voz contaminados por ruído branco gaussiano. Os algoritmos abordados nesta proposta são o filtro de Kalman clássico e a transformada wavelet. Após a aplicação do filtro de Kalman, o sinal é decomposto pela transformada wavelet (WT), aplicando-se a limiarização nos coeficientes da WT, como último passo tem-se a reconstrução do sinal. A WT pode decompor o sinal em diversos coeficientes, com níveis de profundidade diferentes, e a limiarização pode ser realizada em quaisquer destes coeficientes, surgem assim várias combinações entre o filtro de Kalman e a WT. Na busca da melhor combinação desses algoritmos, e dos melhores coeficientes da transformada wavelet discreta (DWT) e da transformada wavelet packet (WPD), é utilizado Algoritmo Genético (AG). O AG é aplicado na determinação dos coeficientes da transformada wavelet a serem filtrados. A aplicação do AG resultou em três métodos que são comparados de maneira objetiva utilizando a distância de Itakura-Saito e SNR segmentada, e de maneira subjetiva com um teste de reconhecimento de palavras realizado com voluntários. Determinado o melhor algoritmo, é feita uma comparação entre ele e a subtração espectral em que se verifica a superioridade do método proposto em SNRs de 0 a 15dB. O presente trabalho aborda aspectos teóricos sobre os dois métodos usados, e exibe com riqueza de detalhes os experimentos realizados.

Palavras-chave: Filtro de Kalman. Transformada Wavelet. Redução de ruído. Algoritmo Genético.

Santos, Jucelino C. M. **DENOISING IN SPEECH SIGNALS COMBINING KALMAN FILTER AND WAVELET TRANSFORM.** Master's Thesis. Federal University of Uberlândia. Uberlândia – MG

ABSTRACT

This work aims to present, combine and compare noise reduction techniques applied to voice signals contaminated by Gaussian white noise. The algorithms discussed in this proposal are the classic Kalman filter and wavelet transform. After application of the Kalman filter, the signal is decomposed by wavelet transform (WT) by applying the thresholding coefficients in the WT, as the last step has to signal reconstruction. The WT may decompose the signal into different coefficients with different depth levels, and thresholding can be performed in any of these coefficients, as well arise various combinations of the Kalman filter and WT. In search of the best combination of algorithms, and the best coefficients of discrete wavelet transform (DWT) and wavelet packet transform (WPD) is used Genetic Algorithm (GA). The AG is applied in the determination of wavelet transform coefficients to be filtered. The application of AG resulted in three methods are compared objectively using the distance Itakura-Saito and targeted SNR, and subjectively with a word recognition test performed on volunteers. Given the best algorithm, a comparison is made between it and the spectral subtraction in which there is the superiority of the method proposed in SNRs of 0 to -15dB. This paper discusses theoretical aspects of the two methods used, and displays in great detail the experiments.

Key- words: *Kalman Filter. Discret Wavelet Transform. Denoising. Genetic Algorithms.*

LISTA DE FIGURAS

Figura 2.1 Modelos dos Ressonadores Christian Gottlieb Kratzenstein	16
Figura 2.2 Máquina construída por Wolfgang Ritter von Kempelen nas três primeiras figuras da parte superior. Máquina construída a partir do modelo de Kempelen por Charles Wheatstone na parte inferior.....	17
Figura 2.3 Fonógrafo Cilíndrico de Thomas Edison. Fonte retirada de: http://www.memoriadaeletricidade.com.br	17
Figura 2.4 Diagrama de blocos do processo de produção da voz	18
Figura 2.5 Diagrama resumido do processo de produção da voz.....	18
Figura 2.6 Modelo de produção da fala (Silva, 2007).....	19
Figura 2.7 (a) Exemplo de um som sonoro e (b) exemplo de som não sonoro (Silva, 2007) ..	20
Figura 2.8 Tipos de ruídos. (a) Ruído Branco. (b) Ruído Marrom. (c) Ruído Rosa. O eixo horizontal indica o tempo e o vertical a amplitude do ruído. Fonte: (Vaseghi, 2000).	25
Figura 2.9 Sinal original e seu espectro antes e depois da contaminação por ruído branco gaussiano, SNR=0 dB.....	29
Figura 3.1 Comparação das funções base wavelet (Misiti, et al., 2015)	40
Figura 3.2 Distribuição Gaussiana. Fonte: (Vaseghi, 2000).....	42
Figura 3.3 Diagrama de blocos das equações do filtro de Kalman de tempo discreto	50
Figura 4.1 Decomposição multinível DWT	56
Figura 4.2 Decomposição WPD	56
Figura 4.3 Árvore de decomposição WPD e o sinal de origem. (a) Sinal de voz ‘bala’.....	57
Figura 4.4 Sinal de voz correspondente a palavra ‘bala’ à esquerda e seus respectivos espectrogramas à direita, utilizando as famílias Daubechies ‘dB5’ e coiflet ‘coif3’.	58
Figura 4.5 Algoritmos para redução de ruído. (a) Algoritmo baseado em decomposição DWT. (b) Algoritmo baseado em decomposição WPD	60
Figura 4.6 Diagrama de blocos do método proposto com refinamento por meio de algoritmo Genético.....	65
Figura 4.7 Limiarização x Kalman 0dB	66
Figura 4.8 Limiarização x Kalman 3dB	66
Figura 4.9 Limiarização x Kalman 5dB	66
Figura 4.10 Limiarização x Kalman 10dB	66
Figura 4.11 Limiarização x Kalman 15dB	66

Figura 4.12 Limiarização x Kalman 0dB	66
Figura 4.13 - Melhores coeficientes para SNR 0 dB.....	67
Figura 4.14 - Melhores coeficientes para SNR 3 dB.....	67
Figura 4.15 - Melhores coeficientes para SNR 0 dB.....	67
Figura 5.1 Comparativo entre DWT, WPD e DWTWPD para SNR segmentada e Itakura-Saito	81
Figura 5.2 Comparativo entre DWT e WPD em 0dB, ‘elétrica’ voz masculina	83
Figura 5.3 Espectrograma entre DWT e WPD em 0dB, ‘elétrica’ voz masculina	83
Figura 5.4 Comparativo entre DWT e WPD em 20dB, ‘elétrica’ voz masculina	83
Figura 5.5 Espectrograma entre DWT e WPD em 20dB, ‘elétrica’ voz masculina	83
Figura 5.6 Comparativo entre DWT e DWTWPD em 0dB, ‘elétrica’ voz feminina	84
Figura 5.7 Espectrograma entre DWT e DWTWPD em 0dB, ‘elétrica’ voz feminina	84
Figura 5.8 Comparativo entre DWT e DWTWPD em 20dB, ‘elétrica’ voz feminina	84
Figura 5.9 Espectrograma entre DWT e DWTWPD em 20dB, ‘elétrica’ voz feminina	84
Figura 5.10 Comparativo entre DWT e subtração espectral em 0dB, ‘bala’ voz masculina....	86
Figura 5.11 Comparativo entre DWT e subtração espectral em 20dB, ‘bala’ voz masculina..	86
Figura 5.12 Espectrogramas DWT e subtração espectral em 0dB, ‘bala’ voz masculina.....	87
Figura 5.13 Espectrogramas DWT e subtração espectral em 20dB, ‘bala’ voz masculina.....	87
Figura 5.14 Comparativo entre DWT e subtração espectral para SNR segmentada e Itakura-Saito,	90
Figura 5.15 Comparativo entre DWT e subtração espectral para SNR segmentada e Itakura-Saito,	91

LISTA DE TABELAS

Tabela 1 Variação da Frequência Fundamental	21
Tabela 3 - Cromossomo representando um possível esquema de filtragem.....	61
Tabela 4 Sons testados.....	72
Tabela 5 Comparativo DWT x WPD, SNRi 0dB.....	73
Tabela 6 Comparativo DWT x WPD, SNRi 5dB.....	73
Tabela 7 Comparativo DWT x WPD, SNRi 15dB.....	74
Tabela 8 Comparativo DWT x WPD, SNRi 20dB.....	74
Tabela 9 Comparativo DWT x DWTWPD, SNRi 0dB.....	75
Tabela 10 Comparativo DWT x DWTWPD, SNRi 5dB.....	76
Tabela 11 Comparativo DWT x DWTWPD, SNRi 15dB.....	76
Tabela 12 Comparativo DWT x DWTWPD, SNRi 20dB.....	77
Tabela 13 Comparativo WPD x DWTWPD, SNRi 0dB.....	78
Tabela 14 Comparativo WPD x DWTWPD, SNRi 5dB.....	78
Tabela 15 Comparativo WPD x DWTWPD, SNRi 15dB.....	79
Tabela 16 Comparativo WPD x DWTWPD, SNRi 20dB.....	79
Tabela 17 Teste de reconhecimento com ouvintes.....	81
Tabela 18 Comparativo WPD x subtração espectral, SNRi 0dB	88
Tabela 19 Comparativo WPD x subtração espectral, SNRi 20dB	89

LISTA DE SIGLAS E ACRÔNIMOS

AG	Algoritmos Genéticos
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
FFT	Fast Fourier Transform
ISD	Itakura-Saito Distance
LIT	Linearmente Invariante no Tempo
LPC	Linear Predictive Coding
MIT	Massachussetts Institute of Technology
PDF	Probability Density Function
PDS	Processamento Digital de Sinais
PMX	Partially Matched Crossover
SNR	Signal Noise Ratio
SNRseg	Signal Noise Ratio segmental
STFT	Short Time Fourier Transform
WPD	Wavelet Packet Decomposition
WT	Wavelet Transform

SUMÁRIO

1	Introdução.....	12
1.1	Considerações Preliminares e Motivação	12
1.2	Objetivos.....	13
1.3	Estrutura da dissertação	13
2	Conceitos básicos	15
2.1	Processamento de sinais.....	15
2.2	Sinais de Voz	16
2.2.1	Raízes históricas.....	16
2.2.2	Produção do sinal de voz.....	18
2.2.3	Modelagem Matemática dos sons da voz.....	20
2.2.4	Princípios Básicos do Modelo de predição Linear.....	22
2.3	Ruído.....	24
2.3.1	Ruído Branco.....	25
2.3.2	Ruído Colorido.....	25
2.4	Quantificando a qualidade da fala	26
2.4.1	Medidas de Distorção da Fala – distância de Itakura-Saito (ISD)	26
2.4.2	Medidas para a Quantização do Ruído.....	27
2.5	Espectrograma para sinais de voz	28
2.6	Análise da DFT para sinais de voz	29
2.7	Considerações finais	31
3	Algoritmos de redução de ruído	32
3.1	Revisão Literária.....	32
3.2	Introdução à Transformada Wavelet.....	33
3.3	Transformada Wavelet e suas famílias	35
3.3.1	Wavelet de Haar	37
3.3.2	Wavelet de Meyer	37
3.3.3	Wavelet de Morlet.....	38
3.3.4	Wavelet Daubechies.....	38
3.3.5	Wavelet Symlets.....	39
3.3.6	Wavelet Coifman.....	39
3.4	Introdução ao Filtro de Kalman	40
3.5	Conceitos de Processos Estocásticos para o Filtro de Kalman	41
3.5.1	Média e covariância	41

3.5.2	Distribuição Normal.....	42
3.5.3	Teorema de Bayes	43
3.6	Filtro de Kalman de Tempo Discreto (FKT)	44
3.7	Modelagem do Filtro de Kalman para Voz.....	48
3.8	Computação do Filtro de Kalman	50
3.9	Considerações finais	53
4	Capítulo – Aplicação de Algoritmos genéticos para determinar o método de redução de ruído.....	55
4.1	Decomposição wavelet	55
4.1.1	Famílias wavelets	57
4.1.2	Limiar	58
4.2	O método proposto, a priori.....	58
4.3	O método proposto com refinamento por algoritmos genéticos.....	60
4.3.1	Aplicação de AG ao problema	61
4.3.2	Resultados do refinamento por meio de AG	65
4.4	Considerações finais	69
5	Experimentos e resultados.....	71
5.1	O experimento.....	71
5.2	DWT <i>versus</i> WPD	72
5.3	DWT <i>versus</i> DWTWPD	75
5.4	WPD <i>versus</i> DWTWPD	77
5.5	Conclusões preliminares	80
5.6	Comparação entre DWT e subtração espectral.....	85
5.6.1	Comparação baseada em amostra do sinal.....	85
5.6.2	Comparação baseada em espectrograma.....	85
5.6.3	Comparação baseada em medidas.....	88
5.7	Considerações finais	89
6	Conclusões e trabalhos futuros.....	92
6.1	Trabalhos futuros	93
	REFERÊNCIAS BIBLIOGRÁFICAS	94

1 INTRODUÇÃO

1.1 Considerações Preliminares e Motivação

Á área de processamento de sinais obteve grande avanço nos últimos 70 anos possibilitando o advento de tecnologias integradoras e revolucionárias que estão mudando completamente a forma de agir e pensar das pessoas. Nunca a comunicação esteve tão evidente quanto nos últimos anos, o ser humano criou uma total dependência em relação a dispositivos eletrônicos que permitam a troca de mensagens, envio de e-mails, acesso a redes sociais e comunicação por voz. Com isso, os requisitos de qualidade na comunicação também se evidenciam, diante de tamanha variedade tecnológica a disposição, as pessoas não toleram falhas, indisponibilidade ou qualidade ruim de modo geral.

Para que o desenvolvimento e a qualidade se mantenham em patamares aceitáveis pelo mercado consumidor é preciso que haja a melhoria contínua nos sistemas de comunicação. Um dos grandes problemas em sistemas de comunicação por voz é a presença de ruído. Rabiner et al. (2011) definem ruído como um processo de interferência que degrada o sinal de voz prejudicando a inteligibilidade (Rabiner, et al., 2011).

O ruído está presente em quase todos os tipos de processamento e pode afetar consideravelmente sistemas de armazenamento, compressão, transmissão e reconhecimento de sinais. A remoção de ruído em sinais de fala pode ter finalidades distintas: a compreensão humana e o reconhecimento de fala feito por máquinas. Apesar da ideia de melhoria estar embutida nestas duas situações, tem se uma dicotomia, uma vez que a melhoria de sinais de fala destinados ao processamento por máquinas tem sua avaliação pautada em medidas objetivas, o que não necessariamente garante a inteligibilidade, que é a principal quantificação de qualidade atribuída por pessoas a um sinal de voz.

Além do contexto, a qualidade da redução de ruído também é dependente do tipo de ruído, que pode ser classificado em estacionários e não estacionários (Rabiner, et al., 2011). Os ruídos estacionários são aqueles que não sofrem mudanças em seus aspectos estatísticos (média e variância) ao longo do tempo mantendo assim as características espectrais bem definidas. No sentido oposto os ruídos não estacionários sofrem mudanças em seus aspectos estatísticos não preservando suas características temporais e espectrais tendo sua variabilidade relacionada ao tipo de ambiente. Embora ruídos não possam ser eliminados, eles podem ter seus efeitos diminuídos significativamente. A qualidade da redução do ruído depende da

técnica utilizada, do tipo de ruído e em alguns casos do correto ajuste de parâmetros. Pode ser utilizado hardware, software ou uma forma híbrida combinando estes dois elementos.

O filtro de Wiener, o filtro de Kalman, a subtração espectral e a transformada wavelet são exemplos de algoritmos que tem mostrado boas características na redução de ruído, e podem ser aplicados até mesmo em outros contextos na área de processamento de sinais (Rabiner, et al., 2011), (Abreu, 2013), (Boll, 1979), (Brown, et al., 1997), (Daubechies, 1990). Assim, com o propósito de obter métodos robustos de redução de ruído, estudos tem indicado a combinação entre métodos com objetivo de explorar as melhores características, por exemplo, o método apresentado por (Messaoud, et al., 2015) que aplicou PCA melhorado à transformada wavelet packet e obteve bons resultados, o método proposto por (Hassani, et al., 2011) que combinou subtração espectral e wavelets, dentre outros.

Por isso, este trabalho visa a redução de ruído em sinais de voz por meio de uma combinação de técnicas, observando o equilíbrio entre inteligibilidade e qualidade objetiva medida por meio da distância de Itakura-Saito e SNR segmentada (Itakura, 1975), (Vaseghi, 2008).

1.2 Objetivos

Este trabalho tem como objetivo geral a combinação de algoritmos de redução de ruído, com o propósito de prover melhorias na filtragem de sinais de voz ruidosos.

Quanto aos objetivos específicos, pode-se citar:

- a) Estudar e implementar o filtro de Kalman;
- b) Estudar e explorar a transformada wavelet e suas famílias;
- c) Unir filtro de Kalman e transformada wavelet;
- d) Aplicar algoritmo genético (AG) para ajustar a limiarização nos coeficientes wavelet;
- e) Avaliar o resultado da filtragem por meio de medidas objetivas, como, a distância de Itakura – Saito, a relação sinal ruído (SNR) e relação sinal ruído segmentada (SNRseg).

1.3 Estrutura da dissertação

A presente dissertação está dividida em 6 capítulos sendo o primeiro esta introdução. Os demais capítulos assume esta estrutura:

- O Capítulo 2 consiste dos conceitos básicos sobre processamento de sinais, sinais de voz, modelagem de sinais de voz, medidas objetivas para quantificação da qualidade do sinal.
- O Capítulo 3 é o cerne teórico deste trabalho, aborda os principais conceitos da transformada wavelet, com suas diversas famílias e o filtro de Kalman, contemplando o desenvolvimento das equações que originam o filtro.
- O Capítulo 4 discute o algoritmo para redução de ruído que combina o filtro de Kalman e a transformada wavelet e os refinamentos realizados para se chegar a um algoritmo mais preciso.
- O Capítulo 5 apresenta os testes realizados com os algoritmos gerados a partir do capítulo 4 e compara a melhor alternativa com subtração espectral.
- O Capítulo 6 traz a conclusão e discute os trabalhos futuros.

2 CONCEITOS BÁSICOS

Este capítulo tem como finalidade apresentar conceitos importantes para a compreensão deste trabalho. Por isso, segue uma breve revisão sobre as origens históricas do processamento dos sinais, as características da produção do sinal de voz e sua representação, definição de ruído, medidas de qualidade para o sinal de voz, espectrogramas e transformada discreta de Fourier.

2.1 Processamento de sinais

Para Oppenheim e Schafer (2012) o termo sinal é aplicado a algo que transmite informação e tem representação matemática na forma de funções de uma ou mais variáveis. A variável utilizada na representação do sinal pode ser discreta ou contínua caracterizando os sinais em analógico ou digital. No presente contexto, o termo sinal é usado como sinônimo de sinal digital.

Processamento de sinais é um ramo da ciência que estuda maneiras de tratar, extrair informações, ou comparar ondas eletromagnéticas que representem fatos do mundo físico, tais como imagens e sons. O objetivo de compreender o processamento de sinais é extrair informação suficiente do sinal com objetivo de realizar a classificação, ou o reconhecimento de padrões, ou melhoria, e até mesmo a síntese de sinais. Como aplicações práticas tem-se o áudio, a televisão, o rádio, o telefone celular móvel, o reconhecimento de voz, o radar, a tomografia computadorizada entre outros (Vaseghi, 2008).

Originalmente não se tem um início formal do estudo de processamento de sinais, mas sim, avanços que ocorreram em épocas distintas e que introduziram novas e fundamentais ferramentas para se trabalhar com sinais, por exemplo, De Moivre, introduziu a transformada Z em 1730 e é considerado por muitos um precursor do processamento de sinais. Posteriormente, Fourier introduziu a transformada de Fourier, que será descrita adiante. Contudo, o desenvolvimento massivo de processamento de sinais ocorreu a partir da década de 50, com o advento do computador digital. A flexibilidade dos computadores digitais tornou possível, realizar a análise de algoritmos ou sistema de processamento de sinais antes da sua implementação em hardware. Laboratórios como Lincoln Laboratory do Massachusetts Institute of Technology (MIT) e Bell Telephone Laboratories contribuíram de maneira

significativa para a simulação de muitos sistemas e projetos de filtros, como exemplo, temos as simulações do vocoder para sistemas de codificação de voz (Oppenheim, et al., 2012).

2.2 Sinais de Voz

Um sinal de voz é constituído por uma sequência de pulsos produzidos pelas pregas vocais e possui uma frequência fundamental que varia conforme as características biológicas do sujeito que o emitiu. A representação, a síntese e a análise de sinais de voz levam em consideração o processo de produção de voz em seres humanos e utilizam aproximações matemáticas. Por isso, nas próximas seções serão discutidas as raízes históricas do processamento de fala, com a criação dos primeiros dispositivos de síntese e reprodução de fala, a produção do sinal de voz, a modelagem matemática para representação da voz e o modelo de predição linear.

2.2.1 Raízes históricas

O início do desenvolvimento de processamento de sinais se deu principalmente em função da síntese, armazenamento e reprodução da fala. A proposta de produzir um modelo mecânico com base matemática que permitisse uma representação satisfatória da voz despertou o interesse de muitos pesquisadores e inventores. Uma das primeiras tentativas de reproduzir a fala artificialmente foi realizada ano de 1779 por Christian Gottlieb Kratzenstein que projetou e construiu uma máquina com o objetivo de sintetizar as vogais com base no modelo fisiológico de produção da fala humana. Para isso foram construídos ressoadores (Figura 2.1) com paletas capazes de vibrar, semelhante a pregas vocais (H.Juang, et al., 2005).

O modelo de Kratzenstein apresentava limitações e não era capaz de produzir consoantes, mas serviu como fonte de inspiração para a criação de outros modelos.

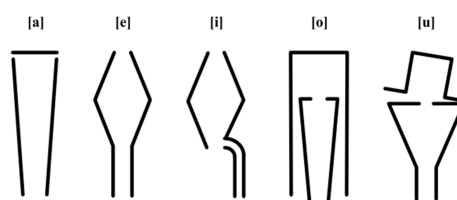


Figura 2.1 Modelos dos Ressoadores Christian Gottlieb Kratzenstein

Fonte: Nicodem (2006)

Wolfgang Ritter Von Kempelen fez importantes formulações sobre máquinas para síntese de fala contínua em meados de 1791, e chegou a publicar o livro “The Mechanism of

Human Speech Including the Description of His Speaking Machine”. No ano de 1835, Charles Wheatstone construiu uma máquina baseada nos estudos de Kempelen usando um fole para imitar o aparelho respiratório, a máquina era capaz de reproduzir as constrictões e produzir as consoantes (Nicodem, 2006). Na Figura 2.2 podem ser vistas as duas máquinas.

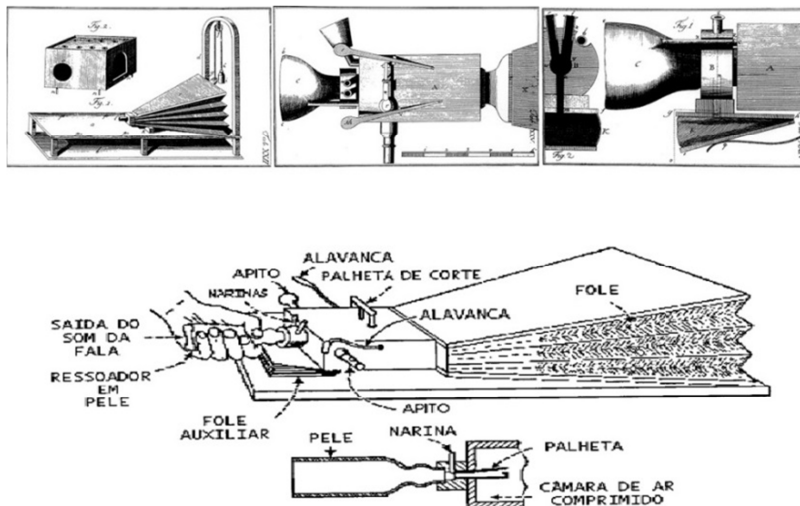


Figura 2.2 Máquina construída por Wolfgang Ritter von Kempelen nas três primeiras figuras da parte superior.

Máquina construída a partir do modelo de Kempelen por Charles Wheatstone na parte inferior.

Fonte: Nicodem (2006)

Outros inventores que também contribuíram para este princípio de processamento de sinais de fala foram Joshep Faber que construiu a Eufonia, ou Fabulosa máquina falante (H.Juang, et al., 2005). Alexander Gran Bell em 1876 apresentou a primeira formulação do telefone. E Thomas Alva Edison, que foi pioneiro na criação de equipamentos capazes de gravar sinais de áudio, criou o fonógrafo cilíndrico ¹ no ano de 1877, conforme ilustra a Figura 2.3.



Figura 2.3 Fonógrafo Cilíndrico de Thomas Edison. Fonte retirada de: <http://www.memoriadaeletricidade.com.br>

¹ Fonógrafo é um tambor cilíndrico rotatório, recoberto por uma folha de estanho.

A partir destes inventos, estudos e teorias foram formuladas e permitiram não só a síntese e a reprodução, mas também a análise de sinais de voz, é o caso da Transformada de Fourier, apresentada em 1811 por Jean-Baptiste Joseph Fourier (Rabiner, et al., 2011).

2.2.2 Produção do sinal de voz

O conhecimento sobre os fundamentos da produção da fala humana é essencial para compreensão e aplicação de muitas técnicas utilizadas no processo de telecomunicação e análise de voz, especialmente em técnicas de redução de ruído, compressão, reconhecimento e transmissão. Segundo Rabiner (Rabiner, et al., 2011) qualquer idioma leva em consideração um conjunto de símbolos responsáveis pela produção dos sons. No processo de produção da voz alguns parâmetros são muito importantes como a: ressonância, intensidade, frequência, articulação, velocidade de fala, pronúncia e aspectos referentes à emissão. Os parâmetros são variáveis e dependentes da escolha do idioma (Rabiner, et al., 2011). O diagrama de blocos da Figura 2.4 é uma representação esquemática do modelo de produção dos sons da fala.

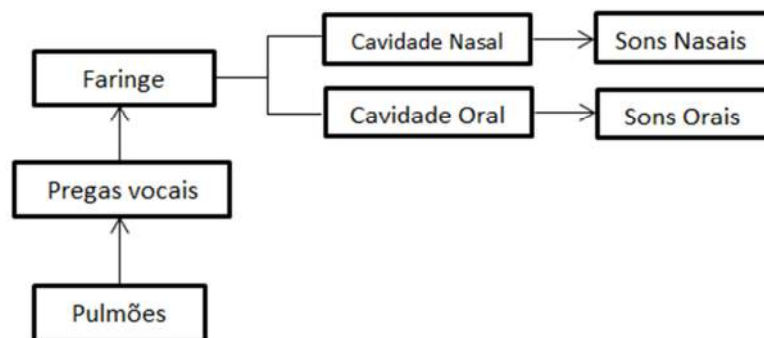


Figura 2.4 Diagrama de blocos do processo de produção da voz

A Figura 2.6 mostra os órgãos envolvidos no processo de produção da voz. Com base no diagrama da Figura 2.5 os órgãos envolvidos no processo de produção da fala ou voz, podem ser divididos em três grupos distintos, sendo (i) um grupo para fonte de excitação, (ii) um grupo para o trato vocal e (iii) um grupo para radiação. Estes três grupos podem ser classificados em subglotal, glotal e supraglotal (Rabiner, et al., 2011).

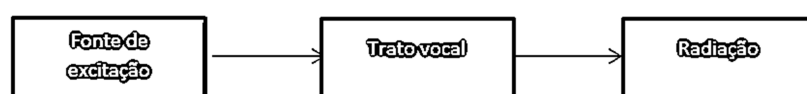


Figura 2.5 Diagrama resumido do processo de produção da voz

O grupo subglotal formado pelos pulmões, brônquios e traqueia funciona como uma fonte de energia para produção da fala. Inicialmente o ar é inalado para os pulmões como uma fonte de excitação do modelo, juntamente com os brônquios e a traqueia. Em seguida os sinais da fala são produzidos durante a fase de exalação do ar presente nos pulmões. No grupo glotal formado pelas cartilagens e pregas vocais, o ar exalado dos pulmões percorre o interior da laringe até chegar às pregas vocais, que vibram com a passagem do ar produzindo dois tipos específicos de sons: os sons sonoros e não sonoros (Rabiner, et al., 2011). As pregas são fibras elásticas que se distendem ou se relaxam pela ação dos músculos da laringe constituindo a parte essencial do trato vocal. A frequência de vibração das pregas que geralmente determina a frequência fundamental da voz de uma pessoa.

No grupo supraglotal estão os lábios, a língua, os dentes, o véu palatino e a boca. Este grupo funciona com um grupo de filtros ressonadores capazes de produzirem os fonemas de um determinado idioma, pois dependendo da configuração geométrica assumida por cada um desses órgãos é possível destacar (e em alguns casos até mesmo atenuar) diferentes frequências presentes nos sons produzidos.

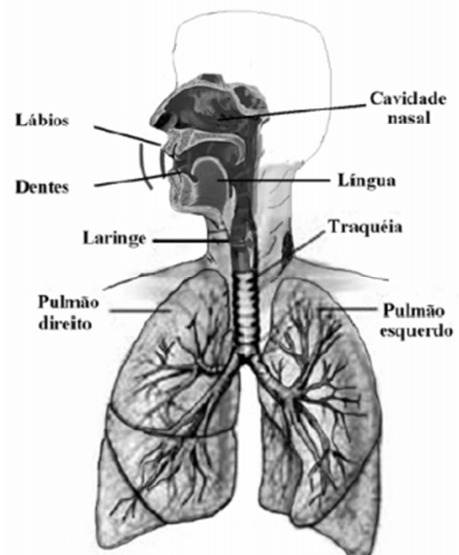


Figura 2.6 Modelo de produção da fala (Silva, 2007)

Em resumo, os sons são gerados com ou sem vibração das pregas vocais e podem ser classificados em três classes (Oppenheim, et al., 2012):

- **Sons sonoros:** são determinados como sendo o resultado da excitação de uma série de pulsos de ar com um comportamento quase periódico gerado pelas pregas vocais. Mostrado na Figura 2.7(a).
- **Sons não sonoros:** são determinados quando a glote está aberta. Assim a corrente de ar proveniente dos pulmões passa pelas cordas vocais sem provocar vibrações, o que produz um fluxo turbulento causado pelo deslocamento da fonte de excitação. Este tipo de turbulência é denominado de ruído do espectro. Mostrado na Figura 2.7(b).
- **Sons Mistos:** é uma combinação de sons sonoros e não sonoros. Por exemplo, o fonema /z/ na palavra zebra.

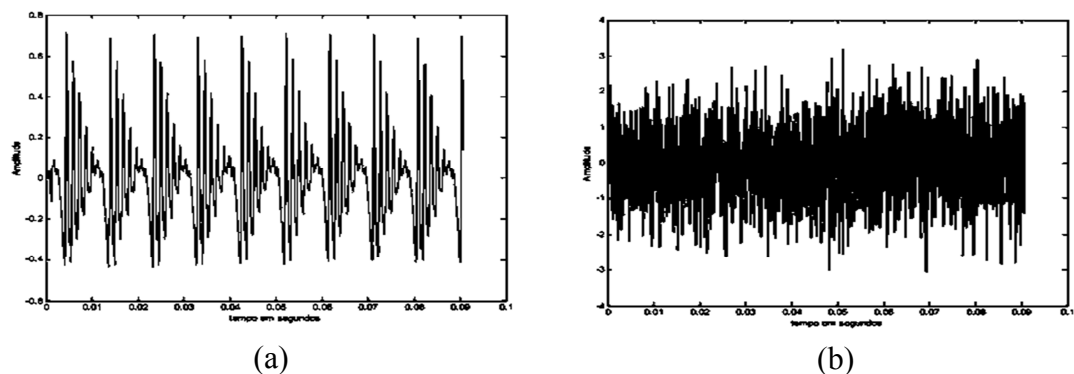


Figura 2.7 (a) Exemplo de um som sonoro e (b) exemplo de som não sonoro (Silva, 2007)

Utilizando um conjunto discreto de símbolos pode-se dar um significado para um determinado som independente do idioma que representa. O estudo e a classificação dos sons da voz são chamados de fonética. Assim, todo idioma é governado pelo estudo da linguística associada às regras da fonética (Rabiner, et al., 2011).

2.2.3 Modelagem Matemática dos sons da voz

Nas zonas vozeadas, à medida que as pregas vocais vibram, o grau de abertura da glote varia e consequentemente o volume de ar proveniente dos pulmões. Esta variação periódica na velocidade na glote excita o trato vocal produzindo sons com harmônicas da frequência de vibração das pregas vocais. Estes sons harmônicos também são frequentemente chamados de frequência fundamental (F_0), ou ainda, frequência da pitch, e são responsáveis pela produção

dos formantes (Rabiner, et al., 2011). A frequência fundamental (F_0) depende da dimensão e espessura da glote, na tabela 1 pode ser visualizada a classificação para os gêneros masculino e feminino (Rabiner, et al., 2011).

Gênero	Faixa de variação das cordas vocais (Hz)
Masculino	50 – 250
Feminino	120 – 300

Tabela 1 - Variação da Frequência Fundamental

A partir da definição da frequência fundamental, define-se a representação dos fonemas. Os fonemas sonoros podem ser representados por uma sequência de pulsos quase periódicos, espaçados em um intervalo da frequência fundamental. O trato vocal molda os pulsos periódicos de acordo com suas frequências resultando no sinal de voz. No caso dos sons não sonoros tem-se um fluxo de ar turbulento que caracteriza a fonte como um ruído branco gaussiano (Oppenheim, et al., 2012). Nas zonas não vozeadas é possível identificar de modo geral maior concentração de energia nas altas frequências, o mesmo não ocorrendo nas zonas vozeadas (Rabiner, et al., 2011).

Um fato que causa grande impacto nas técnicas de análise de som de fala é que não há variações bruscas num fonema ou entre fonemas (Rabiner, et al., 2011). Por isso, métodos de processamento em tempo curto (short-time) são aplicáveis, como por exemplo, a transformada de Fourier em tempo curto (STFT), na qual se analisam amostras de curta duração do sinal (*frames*) que são de aproximadamente de 10 a 30 milissegundos (Rabiner, et al., 2011). Cada segmento curto de voz pode ser visto como um som estável com propriedades fixas o que proporciona a análise das principais características do sinal, como a potência, SNR, envelope espectral entre outras.

Os *frames* (também chamados de quadros ou janelas) do sinal de fala devem ser classificados em som sonoro e não sonoro, uma vez que algoritmos de melhoramento de sinais da fala, principalmente codificadores que utilizam o modelo de predição linear (LPC) usam esta características para prover melhorias. Este modelo será retomado com mais detalhes na próxima seção. Assim, o sinal de voz pode ser considerado como um processo estocástico onde o trato vocal muda lentamente na voz contínua.

2.2.4 Princípios Básicos do Modelo de predição Linear

Uma característica muito importante das vibrações das pregas vocais é o fato de ser possível aproximar o modelo de vibrações por meio de um modelo matemático de polo duplo conhecido na literatura como modelagem paramétrica de sinais. Esta abordagem pode ser vantajosa no sentido de que com uma quantidade predeterminada de parâmetros é possível descrever características inerentes à voz (Oppenheim, et al., 2012).

O modelo paramétrico é designado por modelo de predição linear (LPC) e está intimamente relacionado com as bases dos modelos de síntese de voz. Um sinal de voz amostrado pode ser modelado como um sistema linear variante no tempo (Equações diferença) excitado por pulsos quase periódicos (duração da voz sonora) ou ruídos aleatórios (duração da voz não sonora). A modelagem do sinal de voz utilizando equação diferença sugere que cada amostra do sinal de voz pode ser aproximada por uma combinação linear de p amostras passadas (Rabiner, et al., 2011). Portanto, um sistema no qual o sinal $s[n]$ pode ser representado por um conjunto específico de parâmetros resulta num sinal de saída $\hat{s}[n]$ que seja o mais próximo do sinal de entrada $s[n]$.

Tais modelos hipotéticos do sinal são compostos pelo sinal de entrada $s[n]$ e por uma função de sistema $H(z)$ do sistema linear, tornam-se úteis com acréscimo de restrições que possibilitam encontrar os parâmetros de $H(z)$ dado o sinal a ser representado. Assumindo que a entrada $s[n]$ seja especificada, a função de sistema pode assumir a forma de uma função racional, que pode ser vista na Equação 2.1. Na área de processamento de sinais, a função $H(z)$ é denominada função de transferência, que a partir dos dados de entrada pode representar um grande número de amostras do sinal por um conjunto de parâmetros relativamente pequeno. Para obter o conjunto de parâmetros do sinal, faz-se uso da transformada z , que é uma generalização da transformada de Fourier² (Oppenheim, et al., 2012).

$$H(z) = \frac{\sum_{k=0}^r b_k z^{-k}}{1 - \sum_{k=1}^s a_k z^{-k}} \quad 2.1$$

A Equação 2.1 gera tanto polos quanto zeros. Existe uma grande gama de técnicas que tem como objetivo determinar um conjunto completo de coeficientes tanto do numerador

² Considerando a relação matemática $z = e^{j\omega}$. No caso da transformada de Fourier, vale a relação $z = e^{j\omega}$.

quanto do denominador. Os polos da função $H(z)$ podem assumir diferentes formatos dependendo do tipo de sinal e qual o objetivo pretende-se alcançar (Oppenheim, et al., 2012).

Quando r é igual à zero, gera-se um modelo apenas com polos e o numerador pode ser substituído pelo parâmetro de ganho G , (Equação 2.2). Em processamento de voz, o parâmetro G pode ser calculado a partir dos segmentos de duração finita do sinal em avaliação (Rabiner, et al., 2011).

$$H(z) = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} = \frac{G}{A(z)} \quad 2.2$$

O modelo só-polos pode ser associado a uma classe importante de sistemas LIT (Sistemas Linearmente Invariantes no Tempo). Na escolha dessa classe de sistemas a entrada $s[n]$ e a saída $\hat{s}[n]$ satisfazem um modelo de equações diferenças lineares de n -ésima ordem com os coeficientes constantes. Assim para o modelo só-polos, tem-se a equação:

$$\hat{s}[n] = \sum_{k=1}^p a_k \hat{s}[n-k] + Gu[n] \quad 2.3$$

A estrutura da Equação 2.3 indica que o modelo assumido para a saída do sistema em certo instante n pode ser visto como uma combinação linear de amostras passadas mais uma amostra de entrada multiplicada por um fator de ganho, que numericamente é um valor escalar. Na modelagem só-polos, a entrada $s[n]$ e os parâmetros G, a_1, a_2, \dots, a_p devem ser escolhidos de forma que $\hat{s}[n]$ seja uma aproximação do sinal de entrada $s[n]$ (Oppenheim, et al., 2012). Um Método muito usado na literatura é a aproximação por mínimos quadrados (Rabiner, et al., 2011). Por está técnica pode-se determinar valores para os parâmetros através da minimização da energia total do erro, dado pela equação:

$$e = (s[n] - \hat{s}[n]) \quad 2.4$$

Segundo Oppenheim (2012), para sinais determinísticos os parâmetros do modelo podem ser escolhidos de forma a minimizar o erro quadrático total dado por:

$$\sum_{n=-\infty}^{\infty} (s[n] - \hat{s}[n])^2 = \sum_{i=-\infty}^{\infty} \left(s[n] - \sum_{k=1}^p a_k \hat{s}[n-k] - Gv[n] \right)^2 \quad 2.5$$

No sentido de minimizar a função erro, pode ser empregada na Equação 2.5 a diferenciação em relação a cada um dos parâmetros seguida da igualdade com zero. Contudo, este processo gera um sistema de equações não-lineares, cuja solução em termos computacionais é tida como difícil. Assim, uma alternativa para resolver esse problema é utilizar um modelo inverso por mínimos quadrados que possui uma formulação simples. Uma abordagem com mais detalhes pode ser vista em (Oppenheim, et al., 2012).

2.3 Ruído

Em diversas situações, equipamentos sofrem interferências que se expressam na forma de ruído. Particularmente, os sinais de áudio sofrem com este problema uma vez que a captação do mesmo está suscetível a ruídos do ambiente e até mesmo ruídos do próprio equipamento. Assim, um sinal dificilmente é observado, gravado ou transmitido isoladamente. Existe sempre uma combinação de ruído ou interferência, que é uma questão fundamental em sistemas de comunicação, pois limita a capacidade de enviar e receber informações, corrompe as medidas de precisão em sistemas de controle e diminui a exatidão das decisões em reconhecimento de padrões (Vaseghi, 2008). Existe uma variedade de ruídos que degradam a qualidade do processo de comunicação e que estão presentes em quase todos os ambientes. Por exemplo, no sistema de telefonia móvel, o nível de ruído de fundo é extremamente dependente do lugar onde a comunicação é realizada (Vaseghi, 2008).

Por isso, são empregadas técnicas diversas na remoção ou redução de ruídos. O ruído só não é prejudicial nos contextos em que pode ser controlado. Assim, surge a preocupação com a eficácia da redução de ruídos, e a maneira mais direta de obtê-la é utilizar as características do próprio ruído para diferenciar o sinal de voz do ruído.

Compreender o ruído leva a diferentes classificações do mesmo, e uma classificação usual e bastante geral é dada por sua interpretação estatística: ruídos estacionários e não estacionários. Na classe dos ruídos estacionários uma das principais características é que suas representações estatísticas não sofrem mudança ao longo do tempo. No sentido inverso, os ruídos não estacionários variam suas características estatísticas ao longo do tempo, portanto, suas características temporais e espectrais mudam constantemente à medida que o ruído presente no ambiente muda (Silva, 2011).

O ruído também pode ser classificado em ruído acústico, ruído eletromagnético, ruído eletrostático, distorção do canal (eco) dentre outros (Vaseghi, 2000). Mas, em se tratando de redução de ruído em sinais de voz, é suficiente dizer que o ruído é branco ou colorido.

2.3.1 Ruído Branco

Trata-se de outra designação dada ao ruído estacionário. Um conceito teórico assumido para ruído branco é que ele é um processo aleatório de ruído não correlacionado que tem a mesma potência em todas as frequências, ou seja, a partir do seu registro histórico não é possível prever qual será seu próximo valor no futuro, mas é possível aproximá-lo. Uma boa aproximação do ruído branco pode ser feita gerando um processo aleatório gaussiano com média zero e desvio padrão igual a um. Na maioria dos algoritmos que simulam sistemas de melhoramento de sinais de voz, o ruído branco é aproximado por uma função densidade de probabilidade gaussiana, conforme ilustra a Figura 2.8(a).

O fato de assumir o ruído branco como uma aproximação gaussiana refere-se ao fato que o ruído branco é assumido na modelagem como um ruído aditivo ao sinal de voz. O ruído branco pode simular de maneira satisfatória os ruídos de fundo presente no sinal. Rabiner (2011) cita que o ruído branco é o principal tipo de ruído presente nos sinais reais.

2.3.2 Ruído Colorido

O ruído colorido é o ruído não estacionário, ou seja, que não segue uma propriedade estatística e, portanto, é difícil de ser caracterizado. Todos os ruídos que não possuam as características citadas para o ruído branco são considerados ruídos coloridos, por exemplo, som de carro passando, pessoas conversando, turbina de avião, ventilador, ruídos de ambientes como o metrô, rua etc.

A Figura 2.8(b) traz um exemplo de ruído marrom e a Figura 2.8(c) de ruído rosa. Como pode ser percebido, o ruído rosa varia bastante no decorrer do tempo, enquanto que o ruído marrom possui um comportamento mais previsível em relação a estados anteriores.

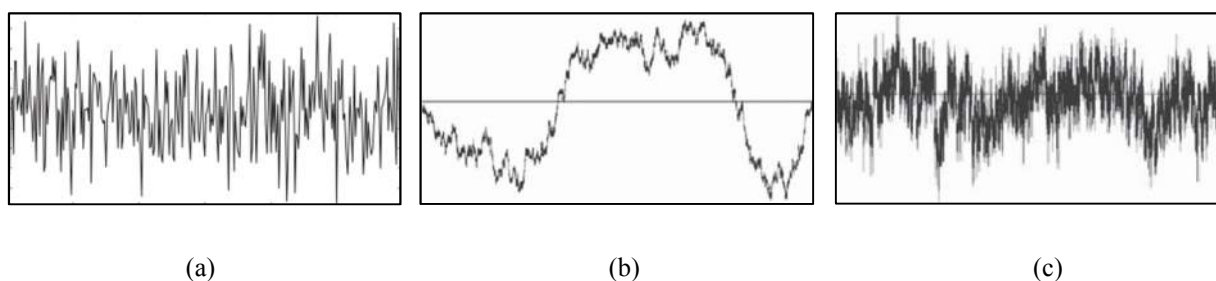


Figura 2.8 Tipos de ruídos. (a) Ruído Branco. (b) Ruído Marrom. (c) Ruído Rosa. O eixo horizontal indica o tempo e o vertical a amplitude do ruído. Fonte: (Vaseghi, 2000).

2.4 Quantificando a qualidade da fala

No processamento de sinais da fala um dos fatores fundamentais é a avaliação da qualidade do sinal processado. Independente do objetivo que se queira alcançar (compressão, melhoramento do sinal, codificação, redução de ruído) existe a necessidade de realizar uma avaliação que possibilite uma comparação do sinal após a etapa de processamento ou filtragem. A seguir serão discutidas algumas medidas.

2.4.1 Medidas de Distorção da Fala – distância de Itakura-Saito (ISD)

Ao longo do tempo, várias medidas foram desenvolvidas para realizar a avaliação do sinal de voz. Basicamente estas medidas de qualidade estão divididas em duas classes: as medidas de qualidade subjetivas e as medidas de qualidade objetivas (Rabiner, et al., 2011).

Medidas de qualidade subjetivas estão relacionadas com a avaliação de um indivíduo ou um grupo de ouvintes treinados para avaliar o sinal limpo e o sinal contaminado reconstruído. Cada ouvinte é responsável por emitir uma nota com base numa escala predeterminada após ouvir os sinais em análise. Posteriormente, agrupam-se os resultados obtendo os sinais com melhor desempenho. Um dos grandes problemas no uso de testes subjetivos é que o tempo de aplicação é extremamente longo e existe a necessidade de uma grande quantidade de ouvintes (Silva, 2011).

Medidas objetivas são baseadas em modelos matemáticos, e tem como finalidade uma comparação do sinal de entrada (original) e o sinal contaminado reconstruído (filtrado).

A distância de Itakura – Saito é uma das medidas mais utilizadas na literatura. Por meio desta medida é possível avaliar o quanto o sinal original e o sinal estimado diferem em seus espectros de amplitudes. (Chu, et al., 1982), (Itakura, 1975). A Modelagem ISD faz uso dos coeficientes de predição linear do modelo LPC, a principal vantagem desta escolha é que o parâmetro de ganho, G , e os coeficientes de filtro $\{a_k\}$ da Equação 2.2 podem ser estimados de uma forma muito simples e computacionalmente eficiente (Rabiner, et al., 2011).

$$d(a, b) = \log \left[\frac{aRa^T}{bRb^T} \right] \quad 2.6$$

Sendo (a) o vetor dos coeficientes LPC do sinal original, (R) a matriz de autocorrelação do sinal original e (b) o vetor dos coeficientes LPC do sinal estimado. Portanto, quanto menor a distância de Itakura-Saito melhor é a correlação entre o sinal original e o sinal filtrado, uma vez que ela reflete a dissimilaridade entre dois sinais.

2.4.2 Medidas para a Quantização do Ruído

Para um sistema de comunicação apresentar bons resultados é necessário um controle de alguns parâmetros indispensáveis na determinação da qualidade da voz. Uma ideia bastante intuitiva é medir a quantidade de ruído presente no sinal. Isso é possível por meio do uso da relação sinal ruído (SNR – Signal Noise Ratio), que é definida como a razão entre a potência do sinal e a potência do ruído:

$$SNR = 10 \log_{10} \left(\frac{\sum_{m=0}^n x^2(m)}{\sum_{m=0}^n r^2(m)} \right) dB = 20 \log_{10} \left(\frac{\sum_{m=0}^n x(m)}{\sum_{m=0}^n r(m)} \right) dB \quad 2.7$$

Onde x^2 é a potência do sinal, r^2 é a potência do ruído e dB indica uma escala adimensional dessa variação.

A SNR é uma medida bastante útil que pode ser usada para avaliar a qualidade do sinal tanto para percepção humana quanto processamento por máquinas, para o contexto deste trabalho, vale observar que:

- A SNR varia de ambiente para ambiente;
- Em ambientes silenciosos a SNR é maior que 30 dB;
- Um sinal de voz é pouco afetado pelo ruído se a SNR é alta;
- Um sinal de voz com níveis de SNR menores que 20 dB (incluindo níveis negativos) está com sua inteligibilidade comprometida, e nesse caso técnicas de redução de ruído se faz necessário, pois permitem reduzir o ruído.

O impacto da SNR na percepção é discutido por Rabiner (2011), que afirma que na percepção humana, ambientes de baixa SNR (por exemplo, SNR=-6) prejudicam bastante a compreensão de sons e até mesmo em ambientes com alta SNR (por exemplo, +12 dB) ocorre confusão na detecção de certas consoantes, particularmente aquelas produzidas da mesma maneira. Em sistemas de detecção automatizada o problema se torna ainda mais intenso, uma vez que a audição humana é dotada de mecanismos altamente sofisticados que permitem inferência a partir de dados incompletos, enquanto que sistemas automatizados normalmente não possuem esta característica.

Embora a SNR seja uma boa medida para quantificar a qualidade de uma filtragem de sinal, deve ser levado em consideração que a maioria das técnicas de redução de ruído realiza o processo de divisão do sinal em janelas para garantir que as características em frequência do

sinal sejam mantidas quando se aplica a transformada de Fourier em tempo curto. Assim, uma melhor formulação da SNR reflete o janelamento do sinal:

$$SNR_{seg} = \frac{1}{k} \sum_{k=0}^{k-1} 10 \log_{10} \left(\frac{\sum_{m=0}^{N-1} x_k^2(m)}{\sum_{m=0}^{N-1} (x_k(m) - \hat{x}_k(m))^2} \right) dB \quad 2.8$$

Em que $x_k(m)$ é o sinal original, $\hat{x}_k(m)$ é o sinal obtido após processamento, N é número total de janelas e k é o tamanho da janela. Esta formulação da relação sinal ruído é conhecida na literatura como relação sinal ruído segmentada, ou ainda SNR_{seg} (Vaseghi, 2000). Desta forma a relação sinal ruído segmentada expressa a comparação entre a potência do sinal de entrada e a potência do ruído em cada frame.

2.5 Espectrograma para sinais de voz

Na área de processamento de sinais existem inúmeras metodologias, algoritmos e fórmulas para tratar a informação. Em particular se o sinal de entrada $x(n)$ for representado por uma imagem, é possível visualizar de inúmeras maneiras o conteúdo de informação da imagem. Em sentido oposto, o trabalho envolvendo o processamento de voz não permite visualizar todas as informações simultaneamente. Uma proposta para visualizar as informações do processo de formação da voz com base na relação de tempo e frequência é chamada de espectrograma (Rabiner, et al., 2011).

Na década de 40 o espectrograma de um sinal de voz era uma ferramenta fundamental para compreensão de como os sons são produzidos e também com a informação fonética e codificada. Na década de 70 o espectrograma era feito por um dispositivo engenhoso composto de um laço de fita de áudio, um filtro de banda analógica eletricamente variável e papel sensível (Rabiner, et al., 2011). Atualmente o espectrograma é realizado por técnicas de processamento digital de sinais sendo possível atribuir uma escala de pseudo-cores para as variações de magnitude da STFT (Rabiner, et al., 2011). A Figura 2.9 mostra a plotagem de um sinal (palavra ‘elétrica’), seu espectrograma, e o espectrograma do sinal original contaminado com ruído na mesma potência ($SNR=0$).

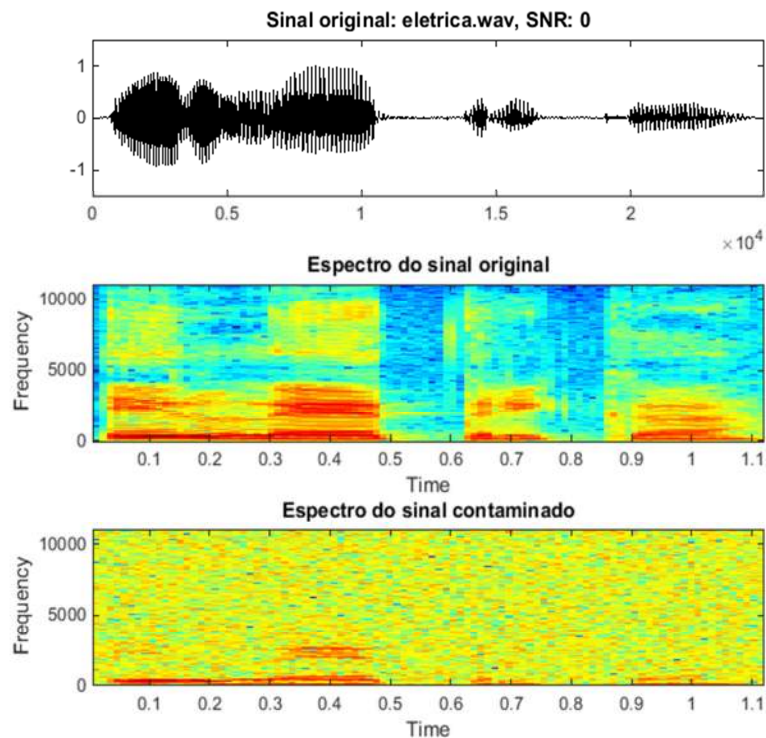


Figura 2.9 Sinal original e seu espectro antes e depois da contaminação por ruído branco gaussiano, SNR=0 dB

2.6 Análise da DFT para sinais de voz

Do ponto de vista matemático, uma Transformação é um procedimento que tem como principal objetivo a mudança ou mapeamento de um conjunto (domínio) em outro conjunto (imagem). No caso, a transformada de Fourier permite a visão geral do sinal a ser analisado no domínio da frequência, o que traz facilidade na hora de analisar o seu comportamento, e possibilita a aplicação de uma grande variedade de técnicas de filtragem digital com o objetivo de retirar alguma característica indesejável no sinal em estudo.

A transformada de Fourier é uma transformada integral que expressa uma função em termos de funções de base senoidal, como soma ou integral de funções senoidais multiplicadas por coeficientes (amplitudes). Pode-se destacar ainda que a transformada de sinais tem duas representações tradicionais:

- A representação no domínio do tempo, na qual a amplitude do sinal é representada como uma função do tempo;
- Representação no domínio da frequência, na qual a função que representa o sinal mostra a amplitude de cada frequência que o compõe.

Segundo a teoria de Fourier, toda função periódica de período 2π pode ser escrita com uma soma infinita de senos e cossenos, como ilustra a equação 2.9:

$$f(x) = a_0 + \sum_{k=1}^{\infty} [a_k \cos(kx) + b_k \text{sen}(kx)] \quad 2.9$$

Onde os coeficientes a_0 , a_k e b_k são calculados segundo as equações a seguir:

$$a_0 = \frac{1}{\pi} \int_0^{2\pi} f(x) dx \quad 2.10$$

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(kx) dx \quad 2.11$$

$$b_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \text{sen}(kx) dx \quad 2.12$$

A transformada de Fourier é construída a partir da sua série de Fourier dada pela Equação 2.9 e é descrita conforme a equação:

$$F(\omega_p) = \frac{1}{2N} \sum_{p=0}^{2N-1} f(t_k) e^{-i\omega_p t_k} = \frac{1}{2N} \sum_{p=0}^{2N-1} f(t_k) [\cos(\omega_p t_k) - i \text{sen}(\omega_p t_k)] \quad 2.13$$

Esta transformada recebe um sinal de entrada no domínio do tempo (t_k), e retorna um sinal no domínio da frequência (ω_p). Para se realizar o cálculo da Transformada de Fourier de um sinal é necessário saber qual será a sua taxa de amostragem.

A taxa de amostragem indica quantas vezes num segundo se podem retirar amostras de voz ou música analógica (Oppenheim, et al., 2012). A taxa de amostragem é dada em hertz, isto é, número de amostras por segundo, sendo as taxas mais comuns 5.500, 8.000, 11.025, 22.050 e 44.100 Hertz (Rabiner, et al., 2011).

Tanto a taxa como a precisão da amostragem afetam a qualidade do som assim como o número de bits. Quanto mais elevada for a taxa e a precisão da amostragem, melhor será a qualidade do som. Contudo, um maior número de bits requer mais espaço de armazenamento e maior capacidade de transmissão de dados (Rabiner, et al., 2011).

O sinal de voz na sua forma pura é considerado como um processo aleatório não estacionário. Dessa forma, a determinação das características presentes no sinal não é trivial. Por isso, um método para contornar esse problema é assumir que as características do sinal

permanecem constantes em intervalos de tempo de 10 ms a 30 ms (Rabiner, et al., 2011). Dessa forma, o sinal pode ser considerado como localmente estacionário. Neste caso pode ser aplicada a transformada de Fourier de tempo curto (STFT) para o processamento no domínio da frequência. A transformada de Fourier de tempo curto é dada por:

$$X_{\hat{n}}(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} w[\hat{n} - m]x[m]e^{j\hat{\omega}m} \quad 2.14$$

Onde $w[\hat{n} - m]$ determina o tamanho da janela que se desloca pelo sinal em particular no índice de tempo \hat{n} . O resultado proveniente da transformada de Fourier gera um sinal de componentes complexas. Nesse sentido, a função complexa da transformada de Fourier é dependente de duas variáveis: o índice de tempo dado por \hat{n} e a frequência determinada por $\hat{\omega}$ que é periódica de período 2π .

2.7 Considerações finais

Este capítulo apresentou conceitos importantes para compreensão deste trabalho, tais como a produção e representação de um sinal de voz, fundamentos históricos. Foram definidos e exemplificados o ruído estacionário e não estacionário, ferramentas importantes como o espectrograma e a transformada de Fourier também foram apresentadas e discutidas. Outra ferramenta importante apresentada, e que é a chave para a implementação do método proposto é a quantificação da qualidade do sinal por meio de medidas como a SNR, SNRseg e distância de Itakura-Saito. No Capítulo 4 é mostrado como estas medidas auxiliaram no refinamento de um algoritmo para redução de ruído. No próximo capítulo serão discutidos algoritmos para a redução de ruído.

3 ALGORITMOS DE REDUÇÃO DE RUÍDO

O ruído prejudica a compreensão e o reconhecimento (por máquinas) dos sons de fala e por isso é importante melhorar a qualidade dos mesmos enfatizando a fala e diminuindo a potência do ruído presente no sinal. Inúmeras técnicas podem ser usadas, sendo elas baseadas em filtros, transformadas, limiarização etc. A redução de ruído também pode ser feita por *hardware* ou, ainda, de forma híbrida. Dentre tais técnicas foram escolhidos o filtro de Kalman e a transformada wavelet que serão discutidos neste capítulo que apresenta também uma revisão literária.

3.1 Revisão Literária

Com base no contexto da redução de ruído mantendo a inteligibilidade do sinal de voz contaminado por ruído aditivo, alguns trabalhos podem ser encontrados. Um trabalho bastante relevante foi apresentado por Boll (1979), que propôs uma técnica denominada de subtração espectral que realiza a melhoria do sinal de voz calculando uma estimativa do espectro de magnitude do sinal limpo por meio da subtração espectral da magnitude do sinal ruidoso e do espectro de magnitude do ruído estimado em trechos em que não existe atividade de voz.

A subtração espectral também é utilizada no trabalho de Hassani e Mollaei (2011) no qual é proposta uma nova abordagem para melhoria do sinal de voz ruidoso combinando dois métodos: a transformada wavelet e a subtração espectral. A ideia é aplicar a subtração espectral nos coeficientes de aproximação e detalhe da transformada wavelet. Deepa et al. (2011) propuseram um método de redução de ruído baseado na variância e função de ganho modificada que utiliza um *threshold* adaptativo, atualizado a partir da faixa de frequência.

Seok e Bae (1997) realizam a redução de ruído aditivo no domínio wavelet. Para remover as componentes de ruído do sinal é realizada uma estimação dos momentos de silêncio no qual estão concentradas as componentes de alta frequência, assim é aplicada uma função de limiarização semi-soft para reduzir o ruído preservando a qualidade do sinal em relação à distorção. Seguindo essa linha, Hu e Loizou (2004) propõem um método para reduzir o ruído musical gerado por algoritmos de melhoramento de sinais de fala que atuam no domínio da frequência realizando uma estimativa mais precisa da variância espectral no domínio wavelet multitaper.

Silva (2011) realiza a redução de ruído em sinais de voz utilizando uma função de ganhos adaptativa para o filtro de Wiener, uma vez que este introduz o ruído musical distorcendo o sinal. Xia e Bao (2014) também apresentam um trabalho bastante semelhante ao trabalho de Silva (2011), com a diferença de que utilizaram um auto-encoder ponderado (Rede Neural Artificial com o objetivo de aprender sobre um conjunto de dados) e classificação de ruído.

Sunny et al. (2013) apresentam um algoritmo para suavização adaptativa do sinal de fala utilizando *soft thresholding* nos coeficientes da Wavelets. Messaoud et al. (2015) utilizam a wavelet packet (WPD) combinada com uma versão melhorada de análise de componentes principais (PCA). O método apresentado usa a habilidade da PCA para determinar um relacionamento linear nos coeficientes WPD e derivar um vetor de características que é usado para melhoria do sinal de voz.

Como pode ser visto, o uso de wavelet na remoção de ruído é bastante difundido, enquanto que o filtro de Kalman, é utilizado para propósitos variados devido a sua natureza intrínseca de predição.

3.2 Introdução à Transformada Wavelet

A transformada de Fourier criada por Jean-Baptiste Fourier é considerada a base fundamental do processamento de sinais, permitindo a mudança de um sinal com representação no domínio do tempo, para uma representação no domínio frequência. As funções base de Fourier são representadas por exponenciais complexas, que teoricamente podem representar qualquer sinal estacionário como uma combinação linear formada por senos e cossenos. Contudo, a transformada de Fourier em tempo curto (ou seja, em quadros, ou janela), quando aplicada ao sinal de voz, introduz uma escala única na qual a análise é realizada do ponto de vista desta escala (Chun-Li, 2010).

A representação de um sinal a partir de uma única escala introduz o inconveniente de que se a largura da janela for menor que os detalhes importantes durante o processamento do sinal, os detalhes não serão localizados mesmo sendo detectados (Duarte, 2005). Outro inconveniente é a hipótese inversa, onde a largura da janela é maior que os detalhes importantes do sinal não revelando as características em análise. Por esse ponto de vista a transformada de Fourier em tempo curto não é capaz de mostrar todas as informações do sinal (Chui, 1992). Uma abordagem para contornar o problema de investigar sinais em escalas

diferentes foi apresentada pela primeira vez na tese de Alfred Haar no ano de 1909. Trata-se de uma função que posteriormente ficaria conhecida como Wavelets (Duarte, 2005).

As funções determinadas por Haar formam a primeira base ortonormal das wavelets conhecidas atualmente. A técnica de se utilizar funções $\phi(\cdot, \cdot)$ a partir de uma única função inicial ϕ não é nova, tanto em matemática quanto na física, ela já existe há algum tempo. Mas, apesar do potencial, as ideias de Haar ficaram no esquecimento por um longo período.

Somente na década de 70 a Wavelet chamou a atenção dos pesquisadores. Isso aconteceu quando o francês Jean Morlet propôs modificações na transformada de Fourier com a finalidade de tratar sinais geofísicos, mais especificamente dados sísmicos. Ele chamou as wavelets de funções de forma constante fazendo uma analogia com as funções base da transformada de Fourier. Jean Morlet solicitou a Alex Grossmann um estudo mais detalhado dos resultados numéricos que obteve. E por isso, Alex Grossmann foi o primeiro a notar a importância do conceito de ‘frame’ ou quadro para transformada wavelet. O matemático Yves Meyer depois de analisar e aprender sobre os resultados de Jean Morlet e Alex Grossmann indicou uma relação entre os operadores integrais singulares e o método de análise existente (Daubechies, 1990).

Os estudos desenvolvidos mostravam uma unificação de várias teorias e sugeria uma nova ferramenta para análise de processamento de sinais que teoricamente poderia substituir as series de Fourier em aplicações numéricas (Faria, 1997). Em 1985, Stephane Mallat trabalhando com o conteúdo informativo em imagens notou que as propriedades do operador que aproxima o sinal de uma dada resolução, com representações vizinhas, poderia ser extraído por decomposição de funções do espaço $L^2(\mathbb{R}^n)$. Sendo que em $L^2(\mathbb{R})$ uma base wavelet ortonormal é uma família de funções da forma $2^{\frac{1}{2}j}\psi(2^j t - n)_{(j,n) \in \mathbb{Z}^2}$ que é construída por meio de escalonamento e translação de uma função exclusiva $\psi(t)$. Este tipo de wavelet também é conhecida como wavelet “diádica”, por fazer uso de uma base binária. Tais estudos foram aplicados na codificação de imagens, discriminação de textura e análise fractal, mostrando a versatilidade da abordagem wavelets (Mallat, 1989).

No início da década de noventa, uma das maiores contribuições para as wavelets atuais foi dada por Ingrid Daubechies, que a partir dos conceitos do trabalho de Mallat, construiu um conjunto de bases ortonormais de wavelets suaves com suporte compacto.

3.3 Transformada Wavelet e suas famílias

Considere o espaço $L^p(\mathbb{R})$ como sendo o espaço das funções ' f ' que estão definidas no conjunto dos números reais (\mathbb{R}), e assume valores no conjunto dos números complexos (\mathbb{C}) (ou no conjunto dos números reais \mathbb{R}) (Lima, 2003) tais que:

$$\|g\| = \left(\int_{-\infty}^{+\infty} |g(x)|^p dx \right)^{\frac{1}{p}} < \infty \quad 3.1$$

No caso particular onde $p = 2$ tem-se o espaço das funções mensuráveis $L^2(\mathbb{R})$ a Lebesgue, que são de quadrado integrável, segundo Chui (1992) pode ser analogamente associado ao espaço das funções de energia finita, ou seja, neste caso estamos considerando a wavelet como uma função contínua, assim se $\psi(t) \in \mathcal{R}$ temos:

$$\int_{-\infty}^{+\infty} |\psi(t)|^2 dt < \infty \quad 3.2$$

Definição 1: Uma função $\psi(t) \in L^2(\mathbb{R})$ é denominada wavelet se, e somente se, sua transformada de Fourier $\hat{\psi}(\omega)$ satisfaz a condição de admissibilidade de Daubechies (Daubechies, 1990):

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\omega)|}{|\omega|} d\omega < \infty \quad 3.3$$

A condição de admissibilidade relacionada com a integral (3.2.2) diz que:

$$\lim_{\omega \rightarrow 0} \hat{\psi}(\omega) = 0 \quad 3.4$$

Em outras palavras, a condição de admissibilidade diz que a transformada de Fourier do sinal na frequência se anula quando a frequência é zero (Daubechies, 1990).

A transformada wavelet superar a transformada de Fourier em tempo curto, no sentido de fornecer uma avaliação do sinal em mais de uma escala (Chun-Li, 2010). A introdução de uma escala na abordagem da wavelet introduz a necessidade de uma função de controle. A

função de controle é chamada na literatura como função moduladora, ou ainda wavelet mãe³ que é dada por:

$$\psi_{a,b}(t) = |a|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right) \quad a, b \in \mathbb{R} \quad 3.5$$

Pode se obter uma família wavelets simplesmente alterando seus parâmetros de escalonamento e translações, usando como base a wavelet mãe que é dada na Equação 3.5. Uma normalização geralmente escolhida é que $\|\psi_{a,b}\| = \|\psi\| = 1$ para todo $a, b \in \mathbb{R}, a \neq 0$ (Faria, 1997).

Como salientado anteriormente, a transformada Fourier em tempo curto realiza uma análise da distribuição de energia deslocando uma janela de mesmo tamanho por todo o sinal, mostrando o comportamento da distribuição de energia para cada quadro. Sendo que a escolha da janela é de suma importância na análise do sinal, pois a transformada de Fourier em janelas depende exclusivamente da janela para revelar seu conteúdo em frequência.

Com o uso da transformada wavelet complexa, pode-se fazer uma restrição à função $\psi(t)$ com objetivo de obter somente componentes de frequência positiva, ou seja, $\psi(\omega = 0) = 0$ de forma que a parte imaginária seja a transformada de Hilbert da parte real (Faria, 1997).

A transformada wavelet na sua forma contínua é dada por Daubechies (Daubechies, 1990) na forma:

$$W_f(a, b) = \langle f, \psi_{a,b} \rangle = \int_{-\infty}^{+\infty} f(x) |a|^{-\frac{1}{2}} \psi^*\left(\frac{t-b}{a}\right) dt \quad 3.6$$

Onde $\psi^*(\cdot)$ representa o complexo conjugado da função ψ , e os parâmetros a e b são responsáveis pelo escalonamento e translação respectivamente.

O principal objetivo de se trabalhar com a transformada wavelet, seja ela na matemática na física ou na engenharia, é a mudança dos parâmetros a e b dados na Equação 3.6. O parâmetro a representa o escalonamento e o parâmetro b representa a translação da função wavelet mãe $\psi(t)$. A título de exemplo, pode-se pensar em um processamento de sinal de

³ É uma função com dois parâmetros (escala e translação) que permite a criação de novas famílias wavelets a partir das escolhas dos parâmetros de escala e translação.

voz, no qual está definida uma wavelet como função de manipulação, em que o objetivo é a compressão do sinal. Se o parâmetro $a < 1$, então a wavelet resultante provoca compressão. No sentido oposto, se $a > 1$, a wavelet causa expansão (Duarte, 2005).

As mudanças dos parâmetros a e b tornam possível uma infinidade de famílias wavelets com particularidades diferentes. Algumas famílias de wavelets apresentam bases ortogonais e suporte compacto como as wavelets Daubechies. Outras famílias não apresentam bases ortogonais nem tem suporte compacto, como é o caso das wavelets de Meyer. No entanto, são continuamente diferenciáveis. Nesse sentido serão descritas, a seguir, algumas das wavelets mais citadas na literatura.

3.3.1 Wavelet de Haar

Sendo a primeira wavelet a surgir ao longo da história, a transformada wavelet de Haar tem uma formulação simples cuja função $\phi(t)$ é definida como:

$$\phi(t) = \begin{cases} 1, & \text{se } 0 \leq t \leq 1 \\ 0, & \text{caso contrário} \end{cases} \quad 3.7$$

A wavelet de Haar é construída por meio das operações de mudança de escala e translação aplicada na função wavelet mãe $\phi(t)$ (Soares, 2009). Neste caso a função $\psi(t)$ é definida como $\psi(t) = \phi(2t) - \phi(2t - 1)$, e pode ser representada da seguinte forma:

$$\psi(t) = \begin{cases} 1, & \text{se } 0 \leq t \leq \frac{1}{2} \\ -1, & \text{se } \frac{1}{2} \leq t \leq 1 \\ 0, & \text{caso contrário} \end{cases} \quad 3.8$$

Nesse sentido, as funções bases da wavelet de Haar apresentam suporte compacto⁴. Contudo não são continuamente diferenciáveis, ou seja, suas funções bases não exibem uma transição suave na relação escala/tempo.

3.3.2 Wavelet de Meyer

Yves Meyer construiu a primeira wavelet diferente da wavelet de Haar. A wavelet de Meyer é continuamente diferenciável, determinando uma base ortonormal.

⁴ Diz-se que um sinal é de suporte compacto se existe um intervalo fechado e limitado fora do qual a função se anula (Lima, 1976).

$$\hat{\phi}(\omega) = \begin{cases} \frac{1}{2\pi^{\frac{1}{2}}} & \text{se } |\omega| \leq \frac{2\pi}{3} \\ \frac{1}{2\pi^{\frac{1}{2}}} \cos \left[\frac{\pi}{2} v \left(\frac{3}{2\pi} |\omega| - 1 \right) \right] & \text{se } \frac{2\pi}{3} \leq \omega \leq \frac{4\pi}{3} \\ 0 & \text{se } |\omega| > \frac{4\pi}{3} \end{cases} \quad 3.9$$

Onde $v(a) = a^4(35 - 84a + 70a^2 - 20a^3)$, $a \in [0, 1]$ (Misiti et al. 1996)

$$\hat{\psi}(\omega) = \begin{cases} (2\pi)^{-\frac{1}{2}} e^{j\omega} \sin \left[\frac{\pi}{2} v \left(\frac{3}{2\pi} |\omega| - 1 \right) \right], & \text{se } \frac{2\pi}{3} \leq |\omega| \leq \frac{4\pi}{3} \\ (2\pi)^{-\frac{1}{2}} e^{j\omega} \cos \left[\frac{\pi}{2} v \left(\frac{3}{2\pi} |\omega| - 1 \right) \right], & \text{se } \frac{2\pi}{3} \leq |\omega| \leq \frac{4\pi}{3} \\ 0, & \text{caso contrário} \end{cases} \quad 3.10$$

3.3.3 Wavelet de Morlet

A função wavelet de Morlet é da família não-ortogonal, sendo assim não possui função escala, e é modelada por uma função gaussiana modulada, levemente ajustada de forma que $\psi(0) = C$. Assim sua função wavelet é dada por:

$$\psi(t) = C e^{-\frac{t^2}{2}} \cos(5t) \quad 3.11$$

3.3.4 Wavelet Daubechies

Ingrid Daubechies construiu uma família de wavelets ortonormais todas de suporte compacto. Para garantir que $\psi(t)$ tenha suporte compacto deve-se escolher a função escala $\phi(t)$ também de suporte compacto.

$$\phi(t) = \sum_{k \in \mathbb{Z}} \alpha_k \phi(2t - k) \quad 3.12$$

Onde os α_k são os coeficientes de filtro.

A escolha de certas condições sobre os coeficientes garantem propriedades para um conjunto de função wavelet, como por exemplo, a ortonormalidade. Fatores que garantem que as wavelets Daubechies tenha suporte compacto são:

- a) Um número finito de coeficientes de filtro deve ser diferente de zero;

- b) O número de coeficientes α_k diferentes de zero, determinam diferentes famílias wavelets Daubechies;
- c) O número N é metade do número de coeficientes do filtro utilizado na equação (3.13);
- d) As funções escalas assumem valores diferentes de zero no intervalo $[0, 2N - 1]$, também chamado de suporte da função.

Para as wavelets da Ingrid Daubechies pode se definir as famílias de wavelets com valores atribuídos para a variável N , os valores determinados ficam entre 2 e 10 (Daubechies, 1990). Portanto, dado um número N qualquer, o polinômio de grau igual ou inferior a $N - 1$ pode ser representado de uma forma exata pela combinação linear da função escala $\psi(t)_N$ e de todas as suas translações inteiras. Dessa forma, a wavelet primária de cada uma das famílias associadas pode ser definida por meio de:

$$\psi(t)_N = \sum_{k=0}^{2N-1} b_k \phi_N(2t - k) = \sum_{k=0}^{2N-1} (-1)^k \alpha_{2N-1-k} \phi(2t - k) \quad 3.13$$

A escolha dos coeficientes $b_k = (-1)^k \alpha_{2N-1-k}$ é determinada pela necessidade de garantir a ortonormalidade entre a função escala e a wavelet primária.

Um dos aspectos importantes das wavelets ortogonais de Daubechies e que pode ser considerado como grande vantagem sobre outras famílias de wavelets é fato que o erro no sinal de entrada não cresce com a transformação, proporcionando estabilidade numérica e garantindo agilidade no cálculo da transformada.

3.3.5 Wavelet Symlets

Ingrid Daubechies conseguiu com seu trabalho a generalização das wavelets de Haar, resolvendo o problema de determinar bases que sejam continuamente diferenciáveis. Além disso, ela também propôs uma melhoria sobre a família de wavelet Daubechies, “dbN” fazendo algumas modificações mas garantindo propriedades similares e uma tendência simétrica.

3.3.6 Wavelet Coifman

Além das famílias Daubechies e symlets, Ingrid Daubechies construiu uma série de wavelets ortogonais projetadas para satisfazer condições específicas como simetria e fase

linear, a Coiflet, que é uma importante família semelhante a família Daubechies. Sua principal característica é que limita os momentos de decaimento na função wavelet e na função escala. Assim uma grande vantagem é assumir que os coeficientes de aproximação podem ser representados pelas amostras do sinal. A Figura 3.1 mostra um comparativo das funções base das principais wavelets.

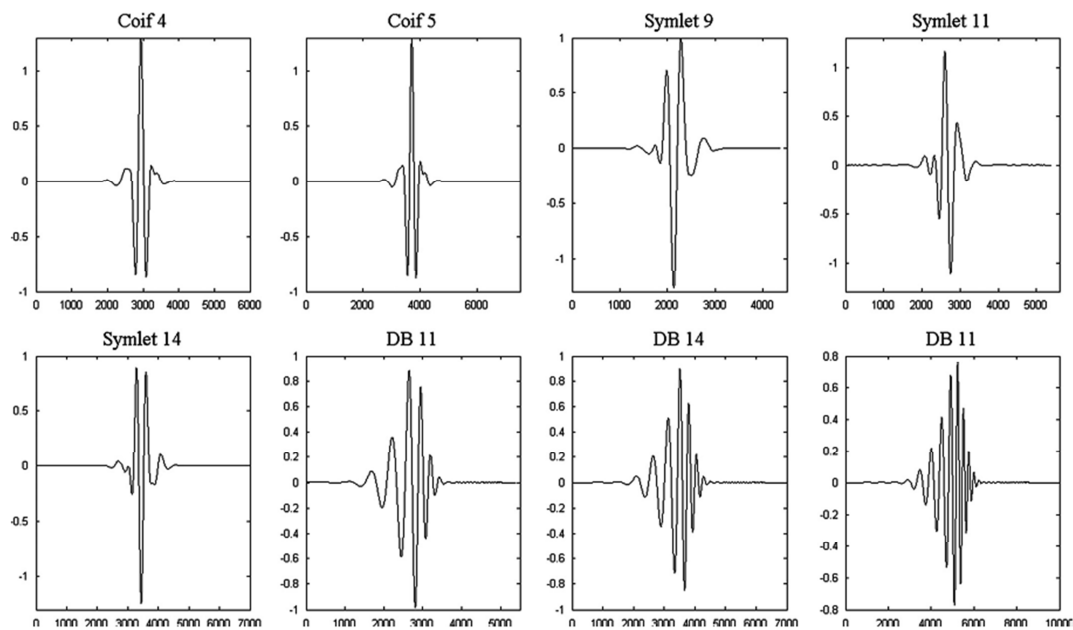


Figura 3.1 Comparação das funções base wavelet (Misiti, et al., 2015)

3.4 Introdução ao Filtro de Kalman

Rudolf E. Kalman publicou seu trabalho no ano de 1960 mostrando um método de espaço de estados que permite fazer estimativas de valores futuros usando amostras anteriores. Este algoritmo recursivo considerado muito eficiente é capaz de fazer estimativas das variáveis de estado a partir de medição ruidosas de um sistema. Uma das principais características do filtro de Kalman é prever o estado de um sistema a partir de dados passados, ou seja, existe um modelo de espaço de estados que permite a minimização do erro quadrático médio entre o sinal real e o sinal estimado (Kalman, 1960). O filtro de Kalman foi proposto para superar os problemas da teoria de predição de Noberth Wiener desenvolvida durante a década de 40, que só é aplicável para séries de tempo discreto estacionárias e possui elevado custo computacional. O filtro de Kalman reduz a complexidade computacional do sistema e também pode ser modelado para séries não estacionárias.

Em resumo, o filtro de Kalman independente da aplicação tem como objetivo fornecer estimativas no sentido ótimo dos estados do sistema, prever valores futuros de determinadas variáveis, ou ainda, melhorar a estimativa de variáveis em instantes anteriores (Junior, 2003).

3.5 Conceitos de Processos Estocásticos para o Filtro de Kalman

No algoritmo do filtro de Kalman é utilizada uma estimação recursiva, ou seja, a estimação é obtida por meio de um caso base, que quando atingido durante uma sequência de chamadas sucessivas, faz com que dados sejam disponibilizados promovendo uma atualização constante do vetor de parâmetros do modelo. Portanto, no desenvolvimento do filtro de Kalman é necessário algum conhecimento referente aos modelos estocásticos que funcionam como ferramentas essenciais no processo de estimação recursiva. Uma vantagem dos modelos estocásticos em relação aos modelos determinísticos é sua capacidade de modelar as incertezas na forma de variáveis aleatórias (Junior, 2003).

3.5.1 Média e covariância

Conforme Fukunaga (1990), os conceitos de média e covariância podem ser dados da seguinte forma:

- Uma variável aleatória X é uma função que associa a cada elemento de um espaço amostral um número real.
- Uma variável aleatória discreta é tal que seus valores pertencem a um conjunto finito ou infinito enumerável⁵ de números reais.

Para uma variável aleatória vetorial $X \in \mathbb{R}^n$ é possível determinar a sua média, ou seu valor esperado como:

$$E[x] = \bar{x} = \int_{\mathbb{R}^n} xf(x)dx \quad 3.14$$

Onde $f(x)$ é uma função densidade de probabilidade da variável x .

O conceito de média ou valor esperado é comumente utilizado na definição de momentos centrais. Um momento central é definido em torno de um valor médio (Fukunaga, 1990), e pode ser dado pela fórmula:

⁵ Um conjunto infinito é enumerável quando existe uma correspondência um a um entre os seus elementos e os números naturais.

$$E\{[(x - \bar{x})^n]\} \quad 3.15$$

A Equação 3.15 representa um momento de ordem n . Uma das ordens mais utilizadas em processos estocásticos é dada por $n = 2$. O Momento central de ordem dois ou momento central de segunda ordem recebe o nome de variância no caso unidimensional, e covariância no caso multidimensional (Junior, 2003). Neste caso, tem-se a seguinte equação para representar a covariância:

$$cov[x] = \{E[(x - E[x])[E(x - E[x])^T]\} \quad 3.16$$

Onde T representa a operação de transposição do vetor.

3.5.2 Distribuição Normal

Um dos modelos com maior relevância no cálculo de probabilidades é o modelo normal ou modelo Gaussiano ilustrado na Figura 3.2, que foi proposto pelo matemático Carl Friedrich Gauss. O modelo Gaussiano representa de forma adequada o comportamento de uma variável que se distribui de forma simétrica em torno de um valor central. Os dois parâmetros que caracterizam essa distribuição são a média e o desvio padrão⁶ que definem a variabilidade da distribuição (Pinheiro, et al., 2012).

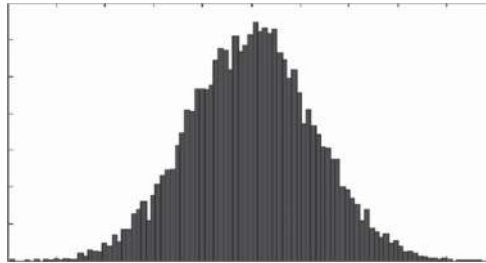


Figura 3.2 Distribuição Gaussiana. Fonte: (Vaseghi, 2000)

Dados os valores da média e da variância, pode-se dizer que uma variável X tem distribuição normal com os parâmetros média e variância se a sua função de densidade é dada por:

⁶ Desvio padrão é definido como sendo a raiz quadrada da variância no caso unidimensional ou covariância no caso multidimensional, ou seja, $\sigma = \sqrt{cov[x]}$.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2} \frac{(x - \bar{x})^2}{\sigma^2}\right\} \quad \text{para todo } x \text{ real.} \quad 3.17$$

Para os propósitos de processamento de sinais, muitos autores usam como modelagem do ruído uma distribuição gaussiana. A justificativa por essa escolha está pautada em princípios físicos, pois geralmente o ruído é composto por um somatório de pequenas fontes de perturbação. Com base nos conceitos matemáticos da distribuição normal, pode-se demonstrar que certa quantidade de variáveis aleatórias somadas pode ser representada por uma função densidade de probabilidade gaussiana (PDF). Na prática, os ruídos do mundo real não são exatamente gaussianos, mas o conceito de se usar uma distribuição de probabilidade gaussiana funciona como uma boa aproximação. (Fukunaga, 1990).

3.5.3 Teorema de Bayes

O teorema de Thomas Bayes é a principal ferramenta para problemas que envolvem estimativas baseadas em probabilidade, fazendo uso dos conceitos estatísticos de probabilidade condicional e probabilidade total de um evento (Fukunaga, 1990). A probabilidade condicional de um evento X_n , dado o evento A , é expressa por:

$$P(X_n|A) = \frac{P(X_n \cap A)}{P(A)} \quad 3.18$$

Sendo que o índice n refere-se ao n -ésimo evento mutuamente excludente do conjunto X , e $P(X_n \cap A)$ representa a probabilidade de interseção entre o evento X_n e o evento A dado, sendo que a probabilidade do evento A é necessariamente diferente de zero, em outras palavras, $P(A) \neq 0$. De maneira análoga pode-se descrever a probabilidade condicional do evento A , dado o evento X_n , sendo expresso por:

$$P(A|X_n) = \frac{P(A \cap X_n)}{P(X_n)} \quad 3.19$$

Novamente tem-se a restrição de que a probabilidade dos eventos X_n seja diferente de zero, ou seja, $P(X_n) \neq 0$. Das probabilidades determinadas pelo teorema de Bayes, apresentadas nas Equações 3.18 e 3.19 ficam implícitas que a probabilidade do primeiro evento depende do segundo evento. Nesse sentido, a definição de probabilidade total de um determinado evento A é dado por:

$$P(A) = \sum_{n=1}^N P(A|X_n)P(X_n) \quad 3.20$$

Com base nas Equações 3.18 e 3.19 é possível obter uma das formas do teorema de Bayes dada por (Pinheiro, et al., 2012):

$$P(X_n|A) = \frac{P(A|X_n)P(X_n)}{P(A)} \quad 3.21$$

Agora substituindo a Equação 3.20 na Equação 3.21 é possível encontrar mais uma versão do teorema de Bayes (Pinheiro, et al., 2012):

$$P(X_n|A) = \frac{P(A|X_n)P(X_n)}{\sum P(A|X_n)P(X_n)} \quad 3.22$$

A equação 3.22 mostra que o teorema de Bayes calcula novas probabilidades de ocorrência dos eventos X_n em função do conhecimento adquirido pela realização do evento A. Os termos $P(X_n)$ e $P(X_n|A)$ representam respectivamente a probabilidade a priori e a probabilidade a posteriori (Fukunaga, 1990). Assim, para a probabilidade a priori, tem-se a probabilidade de ocorrência do evento X_n sem considerar o evento A. No caso da probabilidade a posteriori, tem-se a probabilidade de ocorrência dos eventos X_n considerando a ocorrência do evento A (Junior, 2003).

3.6 Filtro de Kalman de Tempo Discreto (FKT)

O filtro de Kalman pertence à classe dos métodos de filtragem adaptativa. Os filtros adaptativos podem ser utilizados em muitas aplicações envolvendo problemas de processamento de sinais, por exemplo: equalização de canal ruidoso, estimação de trajetória, sistema de identificação, redução de ruído entre outros (Vaseghi, 2008), (Kalman, 1960).

Com o propósito de facilitar a compreensão do algoritmo do filtro de Kalman, é realizada uma breve descrição da solução de um sistema unidimensional com apenas uma variável para estimação, com essa abordagem, pretende-se fazer uso da simplicidade ao transformar variáveis vetoriais em valores escalares, o que facilita o processo de generalização da solução recursiva proposta por Kalman.

Considere uma variável aleatória escalar determinada por x_k satisfazendo a equação linear dada por:

$$x_k = ax_{k-1} + w_{k-1} \quad 3.23$$

onde

- x_k é o estado atual do sistema.
- a é um valor conhecido que relaciona x_k e x_{k-1} .
- w_{k-1} é um ruído branco gaussiano com média zero e variância q .

Considere que a melhor estimativa inicial para a variável x_0 é conhecida e neste caso será representada por \hat{x}_0 . Como na prática nenhum sistema real fornece uma estimativa absoluta sobre a variável em estudo, admite-se que exista um erro entre o valor real de x_0 e o valor estimado de \hat{x}_0 cuja variância pode ser expressa por:

$$p_0 = E[x_0 - \hat{x}_0]^2 \quad 3.24$$

A partir das condições iniciais é possível realizar o cálculo da estimativa seguinte ao qual nomeamos de \tilde{x}_1 . Substituindo o índice $k = 1$ na Equação 3.23 encontramos o valor real para esta amostra, dado por:

$$x_1 = ax_0 + w_0 \quad 3.25$$

Evidentemente, a melhor estimação para a variável x_1 é considerar que a média da variável w_0 seja nula, pois existe pouca informação até o momento. Assim o valor para a estimativa x_1 pode ser representado pela equação:

$$\tilde{x}_1 = a\hat{x}_0 \quad 3.26$$

Novamente podemos calcular a variância do erro de estimação da medida \tilde{x}_1 dada por:

$$\tilde{p}_1 = E[x_1 - \tilde{x}_0]^2 \quad 3.27$$

Da Equação 3.25 e 3.26 obtemos:

$$\hat{p}_1 = E[ax_0 + w_0 - a\hat{x}_0]^2 \quad 3.28$$

Da propriedade da esperança de uma variável aleatória (Garcia, 2008) obtemos:

$$\tilde{p}_1 = a^2 E[x_0 - \hat{x}_0]^2 + E[w_0]^2 + 2aE[(x_0 - \hat{x}_0)w_0] \quad 3.29$$

Utilizando novamente as propriedades da variância, sabe-se que duas variáveis aleatórias não correlacionadas têm suas esperanças nulas. Assim não existe uma correlação

entre as variáveis w e x o que simplifica a Equação 3.29. Lembrando que q é a variância do ruído gaussiano branco w obtendo-se:

$$\tilde{p}_1 = a^2 p_0 + q \quad 3.30$$

Na próxima etapa do filtro de Kalman é possível obter uma medida da variável x contaminada por ruído, com base na seguinte equação:

$$y_1 = cx_1 + e_1 \quad 3.31$$

- y_1 é a amostra do sinal medido.
- c é uma constante conhecida.
- e_1 é um ruído branco gaussiano com variância r .

Se o valor da variável y_1 fosse estimado antes de realizar o processo de medição, teríamos:

$$\tilde{y}_1 = c\tilde{x}_1 \quad 3.32$$

Observando as Equações 3.26 e 3.32, chega-se a conclusão que as variáveis \tilde{x}_1 e \tilde{y}_1 são obtidas de valores calculados no passado, neste caso os valores são dados pela variável \hat{x}_0 . Agora com a possibilidade de medir a variável y_1 , o algoritmo de Kalman abre a possibilidade de uma nova estimativa para a variável x_1 dada pela equação:

$$\hat{x}_1 = \tilde{x}_1 + k_1(y_1 - \tilde{y}_1) \quad 3.33$$

Substituindo a relação determinada pela Equação 3.32 na Equação 3.33, obtemos a nova equação de estimativa para variável \tilde{x}_1 dada por:

$$\hat{x}_1 = \tilde{x}_1 + k_1(y_1 - c\tilde{x}_1) \quad 3.34$$

Observando a Equação 3.34, tem-se uma nova variável k_1 denominada como ganho de Kalman no instante $k = 1$. Consequentemente, o erro ao estimar a variável y_1 dado pela expressão $(y_1 - \tilde{y}_1)$ e multiplicado pelo ganho de Kalman e posteriormente somada ao termo \tilde{x}_1 inserindo assim uma estimativa corrigida. A Equação 3.33 e o termo $(y_1 - \tilde{y}_1)$ são determinados por Kalman como medida de inovação e equação de correção respectivamente.

A determinação do ganho k_1 leva em consideração a variância do erro no processo de estimação da nova medida, assim temos:

$$E[x_1 - \hat{x}_1]^2 = E[x_1 - \tilde{x}_1 - k_1(y_1 - c\tilde{x}_1)]^2 \quad 3.35$$

Substituindo o valor da variável y_1 determinado na Equação 3.31 obtemos:

$$E[x_1 - \hat{x}_1]^2 = E[x_1 - \tilde{x}_1 - k_1(cx_1 + e_1 - c\tilde{x}_1)]^2 \quad 3.36$$

$$E[x_1 - \hat{x}_1]^2 = E[x_1 - \tilde{x}_1 - k_1cx_1 - k_1e_1 + k_1c\tilde{x}_1]^2 \quad 3.37$$

$$E[x_1 - \hat{x}_1]^2 = E[(x_1 - \tilde{x}_1)(1 - k_1c) - k_1e_1]^2 \quad 3.38$$

Aplicando mais uma vez a propriedades de esperança para uma variável aleatória (Garcia, 2008) e lembrando que a variância do ruído gaussiano branco e_1 é igual a r , a Equação 3.38 pode ser escrita como:

$$p_1 = \tilde{p}_1(1 - ck_1)^2 + rk_1^2 \quad 3.39$$

Se o objetivo é obter \hat{x} como uma estimativa no sentido ótimo, uma alternativa é encontrar um valor capaz de minimizar p_1 na Equação 3.39. Fazendo uso dos conceitos de Calculo Diferencial e Integral (Leithold, 1998), derivamos p_1 em relação a variável k_1 e igualamos o resultado a zero, assim obtemos:

$$\frac{\partial p_1}{\partial k_1} = \frac{\partial}{\partial k_1} [\tilde{p}_1(1 - ck_1)^2] + \frac{\partial}{\partial k_1} [rk_1^2] = 0 \quad 3.40$$

$$-2c\tilde{p}_1 + 2c^2\tilde{p}_1k_1 + 2rk_1 = 0 \quad 3.41$$

$$k_1(c^2\tilde{p}_1 + r) = c\tilde{p}_1 \quad 3.42$$

Lembrando que na suposição inicial os valores das variáveis em estudo são escalares, uma alternativa para isolar a variável k_1 é multiplicar ambos os lados da Equação 3.42 pelo inverso do número $(c^2\tilde{p}_1 + r)$. Desta forma obtemos:

$$k_1 = c\tilde{p}_1(c^2\tilde{p}_1 + r)^{-1} \quad 3.43$$

Organizando as Equações 3.26, 3.30, 3.32, 3.34, 3.39 e 3.43 em sequência, obtemos o algoritmo recursivo proposto por Kalman (1960). Contudo, podemos generalizar cada uma das equações para um tempo k qualquer. Nas formulações matemáticas a seguir, temos cada uma das equações citadas e sua versão generalizada:

$$\tilde{x}_1 = a\hat{x}_0 \quad \Rightarrow \quad \tilde{x}_k = a\hat{x}_{k-1} \quad 3.44$$

$$\tilde{p}_1 = a^2 p_0 + q \quad \Rightarrow \quad \tilde{p}_k = a^2 p_{k-1} + q \quad 3.45$$

$$\tilde{y}_1 = c\tilde{x}_1 \quad \Rightarrow \quad \tilde{y}_k = c\tilde{x}_k \quad 3.46$$

$$\hat{x}_1 = \tilde{x}_1 + k_1(y_1 - c\tilde{x}_1) \quad \Rightarrow \quad \hat{x}_k = \tilde{x}_k + k_k(y_k - c\tilde{x}_k) \quad 3.47$$

$$p_1 = \tilde{p}_1(1 - ck_1)^2 + rk_1^2 \quad \Rightarrow \quad p_k = \tilde{p}_k(1 - ck_k)^2 + rk_k^2 \quad 3.48$$

$$k_1 = c\tilde{p}_1(c^2\tilde{p}_1 + r)^{-1} \quad \Rightarrow \quad k_k = c\tilde{p}_k(c^2\tilde{p}_k + r)^{-1} \quad 3.49$$

Com base no desenvolvimento das equações para o filtro de Kalman no caso unidimensional, conclui-se que o filtro de Kalman é um conjunto de equações matemáticas que podem descrever o estado de uma variável a partir de dados passados. O filtro de Kalman também é descrito por alguns autores (Haykin, 2001), (Vaseghi, 2008) como um filtro Bayesiano que emprega as distribuições de probabilidade a priori do sinal e do ruído. Como o objetivo é aplicar o filtro de Kalman em sinais de voz, as equações deduzidas anteriormente precisam ser adaptadas para sua versão multidimensional onde operações matriciais possam ser executadas. Isso será discutido na próxima seção.

$$\tilde{x}_k = A\hat{x}_{k-1} \quad 3.50$$

$$\tilde{P}_k = AP_{k-1}A^T + Q \quad 3.51$$

$$\tilde{y}_k = C\tilde{x}_k \quad 3.52$$

$$\hat{x}_k = \tilde{x}_k + K_k(y_k - C\tilde{x}_k) \quad 3.53$$

$$P_k = (1 - K_k C)\tilde{P}_k(1 - K_k C)^T + K_k R K_k^T \quad 3.54$$

$$K_k = \tilde{P}_k C^T (C \tilde{P}_k C^T + R)^{-1} \quad 3.55$$

3.7 Modelagem do Filtro de Kalman para Voz

Como o objetivo deste trabalho é utilizar o filtro de Kalman aplicado a sinais de voz contaminados por ruído, inicialmente é feita a modelagem do sinal de voz por um processo autoregressivo de ordem \mathbf{P} utilizando um sistema de equações de diferenças (Vaseghi, 2008):

$$x_n = \sum_{k=1}^P a_k x_{n-k} + w_n \quad 3.56$$

Sendo que x_n é o sinal de voz, a_p são os coeficientes LPC de predição linear de ordem p e, por fim, w_n é o erro de predição, que está associado à excitação do modelo fonte-filtro de produção da fala. Uma maneira de representar a Equação 3.56 é dada por um modelo de espaço de estados descritos em Kalman (1960):

$$\begin{bmatrix} x_n \\ x_{n-1} \\ x_{n-1} \\ \vdots \\ x_{n-p+1} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_{p-1} & a_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} x_{n-1} \\ x_{n-2} \\ x_{n-3} \\ \vdots \\ x_{n-p} \end{bmatrix} + \begin{bmatrix} w_n \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad 3.57$$

Usando a notação matricial, a Equação 3.56 pode ser escrita da seguinte maneira:

$$x_n = A_{n-1} x_{n-1} + w_n \quad 3.58$$

Na prática, o sinal de voz x_n é medido na presença de algum tipo de ruído aditivo sendo incorporado ao sinal. Como x_n está na forma de um vetor no espaço de estados, pode-se escrever a equação de observação do sinal:

$$y_n = H^T x_{n-1} + w_n \quad 3.59$$

As Equações 3.58 e 3.59 representam respectivamente a equação de estado e equação de observação do modelo do sinal de voz contaminado por ruído. Sendo:

- x_n é um vetor de estado do sinal de voz com dimensão $P \times 1$.
- A_{n-1} é a matriz transição de estado com dimensão $P \times P$, que relaciona o estado de tempo n com $n - 1$.
- w_n é um vetor de excitação de entrada da equação de estado de ordem $P \times 1$ modelado por ruído branco.
- y_n é o vetor de observação de dimensão $M \times 1$.
- H_n é uma matriz variante no tempo de ordem $P \times P$.
- v_n é um vetor de ruído branco aditivo de dimensão $M \times 1$.

A Figura 3.3 apresenta o diagrama de blocos da teoria do filtro de Kalman de tempo discreto com base nas equações desenvolvidas na Seção 3.5:

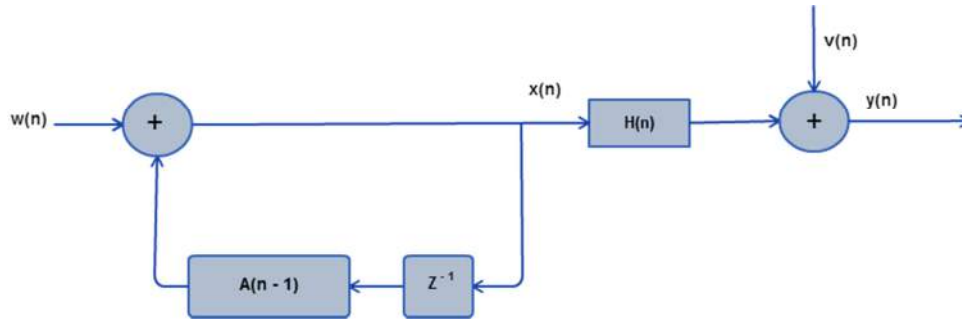


Figura 3.3 Diagrama de blocos das equações do filtro de Kalman de tempo discreto

Os processos w_n e v_n são ruídos gaussianos brancos e independentes, cuja correlação pode ser calculada utilizando a Equação 3.16 apresentado na Seção 3.4, assim tem-se:

$$E[w_n w_k^T] = \begin{cases} Q_n; & k = n \\ 0; & k \neq n \end{cases} \quad 3.60$$

$$E[v_n v_k^T] = \begin{cases} R_n; & k = n \\ 0; & k \neq n \end{cases} \quad 3.61$$

3.8 Computação do Filtro de Kalman

O desenvolvimento das equações do filtro de Kalman para um sinal de voz pode ser feito de maneira análoga ao processo descrito na Seção 3.5. Assim, para o desenvolvimento das equações no caso multidimensional, considera-se a existência inicial do estado x_0 que pode ser calculado utilizando a matriz de covariância do erro de estimação que é dada por:

$$P_0 = E[e_0 e_0^T] = E[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T] \quad 3.62$$

Definida a estimativa inicial para cada observação y_n disponível, o processo é repetido com objetivo de determinar a estimativa \hat{x}_n . Para determinar a estimativa \hat{x}_n , estima-se primeiro \tilde{x}_n sem utilizar uma nova medida y_n , e posteriormente quando y_n é disponibilizado, calcula-se \hat{x}_n . Contudo, não existe nenhuma medida disponível para calcular \tilde{x}_n . Assim, a melhor informação para estimar \tilde{x}_n é \tilde{x}_{n-1} sem considerar a observação y_n . Logo temos a seguinte equação:

$$\tilde{x}_n = A_{n-1} \hat{x}_{n-1} \quad 3.63$$

Novamente o erro associado com a estimação de \tilde{x}_n pode ser determinado calculando a matriz de covariância dada por:

$$\tilde{e}_n = x_n - \tilde{x}_n = A_{n-1}x_{n-1} + w_n - A_{n-1}\hat{x}_{n-1} = A_{n-1}\tilde{e}_{n-1} + w_n \quad 3.64$$

Como o erro \tilde{e}_{n-1} é descorrelacionado com o ruído de excitação w_n , logo a matriz de covariância relacionada com a estimativa \tilde{x}_n é da forma (Brown, et al., 1997):

$$\begin{aligned} \tilde{P}_n &= E[\tilde{e}_{n-1}\tilde{e}_{n-1}^T] \\ \tilde{P}_n &= E[\{A_{n-1}e_{n-1} + w_n\}\{A_{n-1}e_{n-1} + w_n\}^T] \\ \tilde{P}_n &= A_{n-1}P_{n-1}A_{n-1}^T + Q_n \end{aligned} \quad 3.65$$

Sendo Q_n a matriz de covariância associada ao ruído de estado definida na Equação 3.60. Determinada as condições iniciais para o filtro de Kalman, pode-se usar a medida y_n para determinar a medida \hat{x}_n da seguinte forma:

$$\hat{x}_n = \tilde{x}_n + K_n[y_n - H_n\tilde{x}_n] \quad 3.66$$

A Equação 3.66 é denominada como equação de correção e o termo $[y_n - H_n\tilde{x}_n]$ é determinado como medida de inovação do filtro de Kalman e representa a discrepância entre a medida atual de y_n . O termo K_n e H_n representam o ganho do filtro de Kalman, e a matriz de distorção do canal. Analisando a equação de correção, percebe-se que a medida \hat{x}_n é formada por um vetor contendo as observações passadas, o ruído e parte do sinal. Para obter uma estimativa no sentido ótimo, o filtro de Kalman faz uso do critério do erro quadrático médio. Assim o sinal de inovação deve ser descorrelacionado e ortogonal aos vetores de observação passadas (Brown, et al., 1997).

Para obter uma estimativa ótima de x_n utilizando o método dos mínimos quadrados é preciso determinar o ganho k_n de Kalman ótimo. Este ganho que pode ser encontrado determinando a matriz de covariância associada à estimativa atualizada, sendo expressa pela equação:

$$P_n = E[e_n e_n^T] = E[(x_n - \hat{x}_n)(x_n - \hat{x}_n)^T] \quad 3.67$$

Substituindo a Equação 3.59 em 3.66 e em seguida substituindo o resultado na Equação 3.67 obtém-se:

$$P_n = E\{[(x_n - \tilde{x}_n) - K_n[H_n x_n + v_n - H_n \tilde{x}_n]][(x_n - \tilde{x}_n) - K_n[H_n x_n + v_n - H_n \tilde{x}_n]]^T\} \quad 3.68$$

Como apresentado na Seção 3.5, o valor esperado de $(x_n - \tilde{x}_n)$ associado ao erro anterior não apresenta correlação com a medida v_n . Portanto o cálculo da sua esperança é zero. Logo, a matriz de covariância da Equação 3.68 pode ser reescrita de forma simplificada. Aplicando a distributiva à direita do ganho K_n , e evidenciando os termos em comum e aplicando as propriedades de álgebra linear para obter uma matriz toeplitz⁷, temos (Brown, et al., 1997):

$$P_n = E[I - k_n H_n] \tilde{P}_n [I - K_n H_n]^T + k_n R_n K_n^T \quad 3.69$$

A Equação 3.69 é responsável pela atualização da matriz de covariância para algum ganho K_n no sentido ótimo (Kalman, 1960). Sendo que os termos P_n representam a matriz do erro de covariância associada à estimativa a priori, I é a matriz identidade, K_n é o ganho de Kalman, \tilde{P}_n é a matriz de covariância do erro associada a estimativa a priori, H_n é a matriz de distorção do canal e R_n é a matriz de covariância associada ao ruído aditivo gaussiano branco.

A determinação do ganho ótimo de Kalman, K_n , é realizada de diferentes maneiras. No presente contexto, o processo de otimização do ganho K_n é feito ao minimizar as variâncias de erro associadas à estimação dos elementos do vetor de estado utilizando as fórmulas para diferenciação de matrizes dadas por Brown (1997). Assim, a equação que fornece o ganho ótimo de Kalman segundo Brown é:

$$K_n = \tilde{P}_n H_n^T [H_n \tilde{P}_n H_n^T + R_n]^{-1} \quad 3.70$$

Assim a matriz de covariância associada à estimativa ótima pode ser obtida substituindo a Equação 3.70 na Equação 3.69 e reorganizando os termos obtemos a expressão simplificada para P_n segundo Brown (Brown, et al., 1997):

$$P_n = [I - K_n H_n] \tilde{P}_n \quad 3.71$$

As Equações 3.63, 3.65, 3.66, 3.70 e 3.71 são as equações que realizam as atualizações de tempo e medida para o filtro de Kalman. A Tabela 2 apresenta a disposição das equações para o algoritmo:

⁷ É uma matriz em que cada diagonal descendente da esquerda para direita tem valor constante (Lang, 2003).

Equações de atualização de tempo do filtro de Kalman de Tempo Discreto	Equações de atualização de medida do filtro de Kalman de Tempo Discreto
$\tilde{x}_n = A_{n-1}\hat{x}_{n-1}$	$K_n = \tilde{P}_n H_n^T [H_n \tilde{P}_n H_n^T + R_n]^{-1}$
$\tilde{P}_n = A_{n-1} P_{n-1} A_{n-1}^T + Q_n$	$\hat{x}_n = \tilde{x}_n + K_n [y_n - H_n \tilde{x}_n]$
-----	$P_n = [I - K_n H_n] \tilde{P}_n$

Quadro 1 - Equações de atualização de tempo e medida do filtro de Kalman

Sumarizando toda a construção matemática exposta até então, o algoritmo recursivo de Kalman é o seguinte:

<pre> /*Calculo das estimativas iniciais:*/ $\hat{x}_0 = E[x_0]$ e $\tilde{P}_0 = E[e_0 \tilde{e}^T]$ Inicio Para n = 1, 2, 3,..., faça $\tilde{x}_n = A_{n-1}\hat{x}_{n-1}$ $\tilde{P}_n = A_{n-1}P_{n-1}A_{n-1}^T + Q_n$ $K_n = \tilde{P}_n H_n^T [H_n \tilde{P}_n H_n^T + R_n]^{-1}$ $\hat{x}_n = \tilde{x}_n + K_n [y_n - H_n \tilde{x}_n]$ $P_n = [I - K_n H_n] \tilde{P}_n$ Fim </pre>

Quadro 2 Algoritmo do filtro de Kalman

3.9 Considerações finais

O presente capítulo abordou as técnicas utilizadas no desenvolvimento do método de redução de ruído proposto neste trabalho: transformada wavelet e filtro de Kalman. A base matemática de ambos foi trabalhada passo a passo e as particularidades foram discutidas. Contudo, restou uma questão importante a ser resolvida quanto à elaboração do método,

conforme foi dito ao longo deste capítulo, a transformada Wavelet decompõe o sinal em diversos coeficientes, e embora determinar a família wavelet a ser usada na aplicação seja trivial uma vez que pode ser feita em função das propriedades das famílias, saber o coeficientes que contem a informação a ser filtrada, no caso, o ruído, é um passo importante e deve ser feito de maneira cuidadosa. Assim, o próximo capítulo trata da aplicação de algoritmos genéticos para determinar os coeficientes que serão utilizados na limiarização. A implementação conta com duas diferentes maneiras de realizar a decomposição em coeficientes wavelets e testa 26 coeficientes, ao todo.

4 CAPÍTULO – APLICAÇÃO DE ALGORITMOS GENÉTICOS PARA DETERMINAR O MÉTODO DE REDUÇÃO DE RUÍDO

Sinais de voz são altamente sensíveis. A qualquer interferência eles se perdem e se tornam totalmente ininteligíveis. Por isso, deve-se cuidar para que os componentes do sinal não sejam removidos durante o processo de redução, ao passo que a redução precisa ser efetiva. Portanto, existe uma linha muito tênue entre um bom algoritmo de redução de ruído e um algoritmo ruim. Recomenda-se que três aspectos sejam avaliados constantemente durante o processo de redução: o ruído está sendo reduzido? O sinal está sendo removido? O sinal possui distorção? Não é suficiente garantir que o ruído seja reduzido é preciso garantir também que o sinal resultante tenha qualidade suficientemente boa para a aplicação fim, seja ela a audição humana ou reconhecimento feito por máquinas.

Por isso, neste trabalho propõe-se um método híbrido de redução de ruído baseado em filtragem Kalman e limiarização dos coeficientes wavelets. Embora a ideia inicial seja relativamente simples, a sua execução deve ser criteriosa uma vez que os coeficientes wavelets proporcionam inúmeras possibilidades e a redução de ruído nem sempre é algo trivial. Nas próximas seções serão discutidas possibilidades de filtragem nos coeficientes wavelets e como algoritmos genéticos podem ser usados para auxiliar na tarefa de encontrar um bom método de redução de ruído baseado em wavelet. No desenvolvimento foi utilizada a ferramenta MatLab@2015 e sua *toolbox* para wavelet.

4.1 Decomposição wavelet

A decomposição via transformada wavelet possui duas maneiras distintas de ser feita. A primeira é a Wavelet discreta (Discret Wavelet Transform - DWT) e a segunda a Wavelet Packets (Wavelet Packet Decomposition - WPD). A principal diferença entre elas é que a transformada Wavelet Packets filtra o sinal por meio de mais filtros que a DWT: cada nível da DWT é gerado a partir do coeficiente de aproximação do nível anterior, enquanto que na WPD os coeficientes de aproximação e detalhe são decompostos (Coifman, et al., 1992). A decomposição DWT pode ser vista na Figura 4.1 e a decomposição WPD na Figura 4.2.

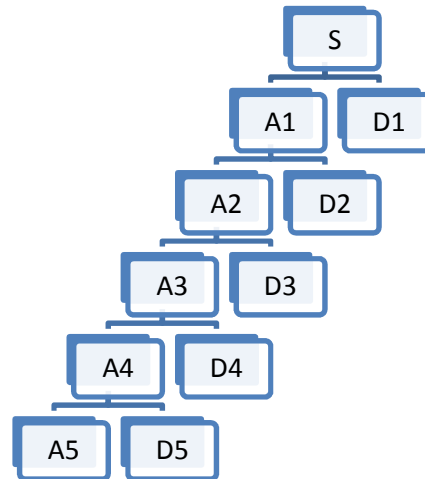


Figura 4.1 Decomposição multinível DWT

A WPD é uma generalização da decomposição wavelet caracterizada pelas inúmeras possibilidades de análises. Nela, o sinal unidimensional é dividido em uma estrutura binária e então pode ser analisado e posteriormente reconstruído. A Figura 4.3 mostra a decomposição de dois sinais, o primeiro é um sinal real (a palavra ‘bala’) gravado a partir de uma voz masculina. E o segundo é um sinal sintético.

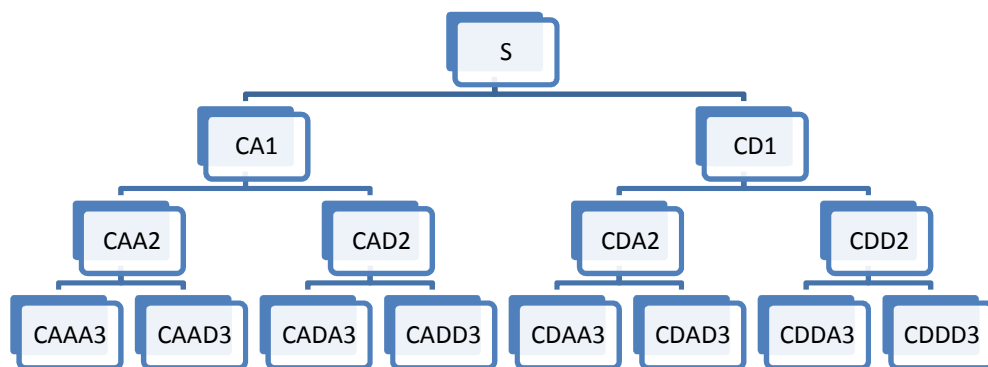


Figura 4.2 Decomposição WPD

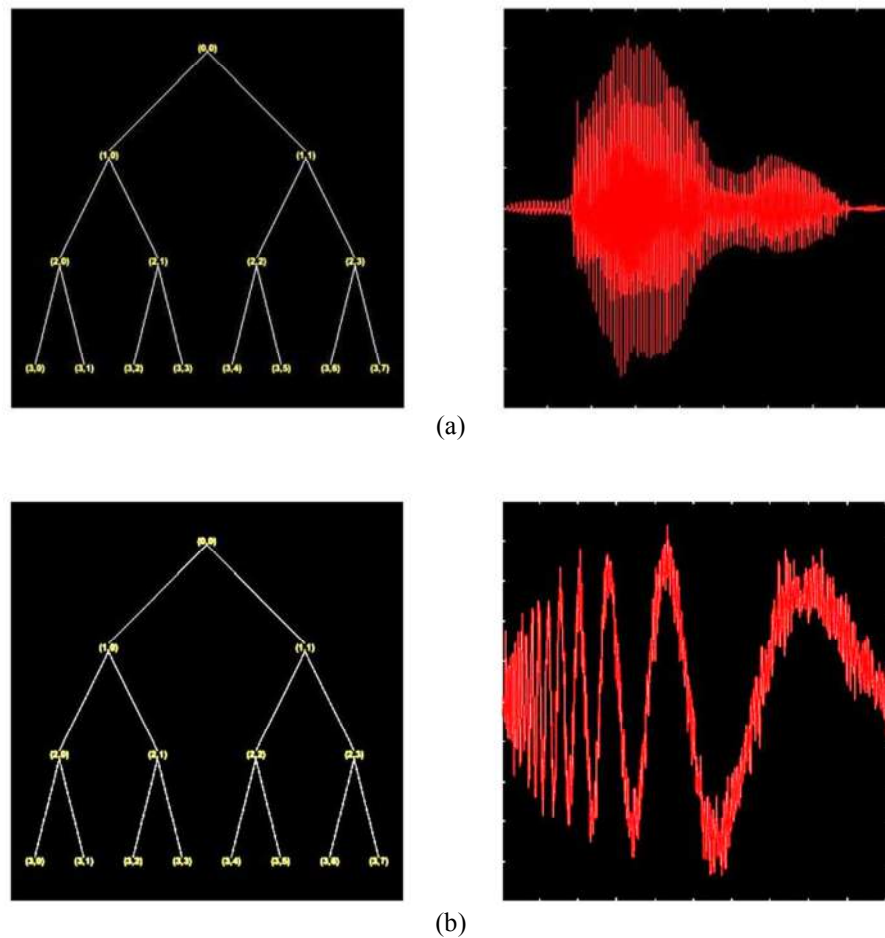


Figura 4.3 Árvore de decomposição WPD e o sinal de origem. (a) Sinal de voz ‘bala’.
(b) Sinal sintético.

4.1.1 Famílias wavelets

Diversos testes de redução de ruído foram realizados e as famílias Dabeuchies e coiflets proporcionaram resultados melhores. Devido ao fato de que possuem suporte compacto e por tanto preservam grande parte da energia do sinal. Com efeito, o resultado observado é demonstrado por Soares (2009) por meio de testes de energia que comprovam a superioridade destas duas famílias quando a aplicação fim é remoção de ruído ou compactação de sinal.

Na Figura 4.4, tem-se as famílias Dabeuchies e Coiflets aplicadas no mesmo sinal de voz.

4.1.2 Limiar

A escolha do limiar pode ser determinada de diversas formas. Neste trabalho foi usando o limiar universal proposto por Donoho (1994) que sugere o limiar λ de energia como:

$$\lambda = \sigma \sqrt{2 \log_{10}(N)} \quad 4.1$$

$$\sigma = \frac{\text{mediana}(Y)}{0.6745} \quad 4.2$$

Sendo N o número de amostras do sinal ruidoso e σ o desvio padrão dado pela Equação 4.2. (Donoho, et al., 1994) ressaltam que o uso do limiar pode ser realizado na forma (*hard*) limiar duro ou na forma (*soft*) limiar suave definidos no quadro 2:

$X_{est} = TH_{hard}(Y, \lambda) = \begin{cases} Y, & \text{se } Y > \lambda \\ 0, & \text{se } Y \leq \lambda \end{cases}$	$X_{est} = TH_{soft}(Y, \lambda) = \begin{cases} \text{sgn}(Y), & \text{se } Y > \lambda \\ 0, & \text{se } Y \leq \lambda \end{cases}$
---	---

Quadro 3 Equações do Limiar Hard e Soft

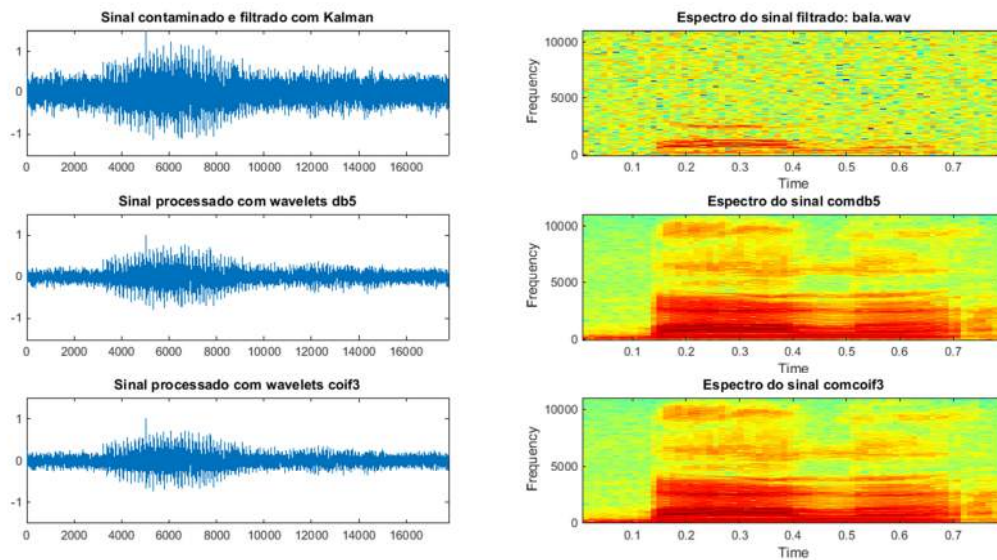


Figura 4.4 Sinal de voz correspondente a palavra ‘bala’ à esquerda e seus respectivos espectrogramas à direita, utilizando as famílias Daubechies ‘db5’ e coiflet ‘coif3’.

4.2 O método proposto, a priori

Como descrito anteriormente, dois métodos eficientes para a redução de ruído são a filtragem Kalman e a limiarização nos coeficientes wavelets. Assim, uma boa ideia é juntar

estes dois métodos em um método híbrido. A fim de explorar o grande número de possibilidades da transformada wavelet é necessário testar todos os coeficientes DWT e WPD e ainda possíveis combinações entre eles.

De antemão, surgem dois algoritmos para a redução de ruído: o algoritmo 1 (Figura 4.5(a)), baseado nos coeficientes DWT e o algoritmo 2 (Figura 4.5(b)) baseado em coeficientes WPD. No algoritmo 1 o sinal é contaminado com ruído e em seguida filtrado utilizando o filtro de Kalman no tempo⁸. No passo seguinte, o sinal é decomposto e reconstruído utilizando a transformada wavelet em 5 níveis, gerando 5 coeficientes de detalhe e 5 coeficientes de aproximação. A limiarização pode ser aplicada em um ou mais coeficientes, feito isto, o sinal é reconstituído por simples soma dos coeficientes. O algoritmo 2 é baseado em WPD e também prevê contaminação do sinal, filtragem Kalman e decomposição em coeficientes wavelets, só que utilizando a WPD. A limiarização ocorre nos coeficientes antes da reconstrução do sinal e como último passo é feita a reconstrução. Na WPD de profundidade 4 são gerados 16 coeficientes de nível 4, sendo que os coeficientes originados do ramo direito (detalhe) devem ser usados na limiarização. Supõe-se também o uso do filtro de Kalman como uma possível melhoria no lugar da limiarização.

Diante destes dois algoritmos iniciais surgem dois questionamentos: qual coeficiente deve ser filtrado? Qual algoritmo é melhor?

Coeficientes de aproximação contêm detalhes de alta frequência e coeficientes de detalhe, de baixa frequência, por isso, é importante observar que nem todo coeficiente oferece boas características para a filtragem com o objetivo de reduzir ruído. Espera-se que a melhor filtragem para ruído branco ocorra nos coeficientes de detalhe, e ainda, espera-se mudanças conforme a quantidade de ruído presente no sinal.

Testes exaustivos foram realizados a fim de obter a melhor opção de filtragem. No entanto, para uma decomposição em 4 níveis, 16 coeficientes são gerados e ainda existem os coeficientes da transformada wavelet DWT, que pode ser realizada em até 5 níveis na *toolbox* do matlab, resultando em 26 coeficientes, nos quais podem ser aplicada a limiarização ou até mesmo o filtro de Kalman. Sendo que na limiarização também podem ser usados diferentes limiares. Testes contemplando estas informações foram realizados, porém, a análise se tornou

⁸ O filtro de Kalman na frequência apresentou resultados ligeiramente inferiores e possui custo computacional bem mais elevado.

impraticável ocasionando a necessidade de um método mais conciso para certificação a cerca da melhor combinação de coeficientes. Por isso, foi escolhido o algoritmo genético (AG) que será discutido a seguir.

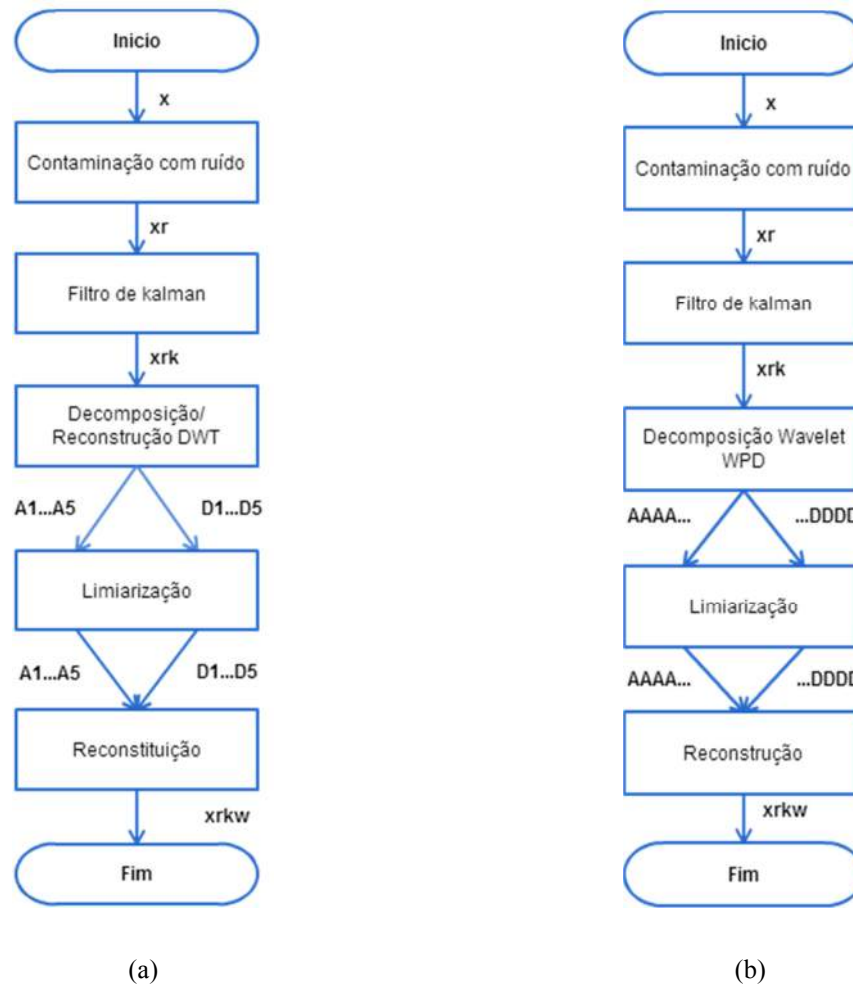


Figura 4.5 Algoritmos para redução de ruído. (a) Algoritmo baseado em decomposição DWT. (b) Algoritmo baseado em decomposição WPD

4.3 O método proposto com refinamento por algoritmos genéticos

Em muitas situações os problemas possuem um grande número de variáveis e ajustes sutis que podem resultar em grandes melhorias. Mas nem sempre é viável realizar isto manualmente uma vez que é preciso comparar diversos resultados numéricos. No presente contexto é preciso verificar diversos coeficientes e combinações entre eles a fim de determinar a melhor opção para a filtragem/limiarização.

Para contornar o referido problema, foi criado um Algoritmo Genético cujo cromossomo consiste de uma combinação de *flags* para os coeficientes. Cada *flag* informa se

o coeficiente vai ser filtrado (Kalman) ou limiarizado. Pré-testes foram realizados a fim de eliminar algumas situações e observou-se que a aplicação do filtro de Kalman nos coeficientes reconstruídos da DWT pouco contribui para uma efetiva remoção de ruído. Assim, esta possibilidade foi eliminada de antemão. Por outro lado, a limiarização e a filtragem Kalman nos coeficientes da transformada WPD se mostraram uma incógnita, hora sendo bastante proveitosos e hora sendo inútil. Por isso, foi preciso determinar um padrão para a sua aplicação por meio do AG.

É importante dizer que o AG é um método de aproximação e não um método exato, uma vez que não há garantias que o melhor resultado será obtido. Situações como a que está em questão são permissivas o suficientes para admitirem uma solução aproximada sem que haja perda, afinal, não se espera obter um cromossomo que reduza absurdamente todos os tipos e intensidades de ruído e sim um cromossomo que seja bom na maioria dos casos.

O que garante o desempenho dos AGs é que eles partem de combinações válidas no contexto em questão e adaptam à medida que evoluem as gerações, quanto maior o tamanho da população inicial, mais “rápido” o algoritmo tende a convergir. No entanto, populações grandes demais podem transformar o AG em um simples teste exaustivo, o que não chega a ser rápido. Logo, é sempre uma boa premissa encontrar um equilíbrio entre estes dois aspectos.

Para quantificar quão bom é um cromossomo gera-se um valor de *fitness* que é calculado por meio da distância de Itakura Saito e SNR, discutidas anteriormente. Ao final da execução do AG, os cromossomos resultantes informam quais as melhores opções de limiarização e filtragem nos coeficientes da transformada wavelet. É obtido um cromossomo para cada arquivo de áudio sendo testado.

4.3.1 Aplicação de AG ao problema

Dois pontos se fazem fundamentais para o desenvolvimento de um bom AG: a determinação do cromossomo e da função objetiva. Por isso, o primeiro passo para a resolução da questão foi o desenvolvimento de um cromossomo que permitisse a implementação das restrições requeridas. Chegou-se ao seguinte modelo de cromossomo:

A1	A2	A3	A4	A5	D1	...	D5	Cf1	...	Cf16	Limi	Kalm	Percentual
1	0	1	1	0	0	...	1	1	...	1	1	1	0.8

Tabela 2 - Cromossomo representando um possível esquema de filtragem

O modelo de cromossomo possui 29 genes, sendo que cada gene representa uma informação:

- A1, A2, A3, A4, A5 – coeficientes de aproximação da transformada wavelet DWT (ver Figura 4.1);
- D1, D2, D3, D4, D5 – coeficientes de detalhe da transformada wavelet DWT (ver Figura 4.1);
- Cf1, Cf2,...,Cf16 – coeficientes da transformada wavelet WPD em profundidade 4 (ver Figura 4.3(a));
- Limi – *flag* informando se será realizada a limiarização dos coeficientes Cf1 a Cf16;
- Kalm – *flag* informando se será realizada a filtragem Kalman dos coeficientes Cf1 a Cf16;
- Percentual – indica o percentual do limiar, calculado via Quadro 2, a ser utilizado.

Para cada coeficiente no cromossomo é atribuído um valor 0 ou 1. Quando 0, o coeficiente não será utilizado na filtragem e caso contrário será usado. O *flag* ‘Limi’ informa se vai haver limiarização nos coeficientes Cf1 a Cf16 e o *flag* ‘Kalm’ informa se vai haver filtragem Kalman nestes mesmos coeficientes.

Com isso a função objetiva usada para calcular a aptidão gerando informações que quantifiquem o quanto aquele indivíduo se aproxima da solução do problema, que no presente contexto é a redução de ruído sem que haja perdas significativas do sinal. Duas medidas foram escolhidas para quantificar a filtragem realizada pelo indivíduo representado pelo cromossomo. Uma delas é a distância de Itakura e a outra é a relação Sinal-Ruído (SNR). A SNR sozinha se mostrou insuficiente uma vez que mesmo com bons valores, permitia grande distorção do sinal. Já a distância de Itakura sozinha quantifica a semelhança do sinal filtrado com o sinal original sem, no entanto, levar em consideração a melhoria do sinal em relação ao ruído. Considerando também que as duas medidas são incompatíveis, foram atribuídos pesos e a distância de Itakura foi invertida para que o problema seja tratado como uma questão de maximização da função objetiva.

$$\begin{cases} \text{se } (SNR_o - SNR_I) < 0, \text{fator} = 0 \\ \text{caso contrário,} & \text{fator} = 1 \end{cases}$$

$$fitness = fator * (SNRO - SNRI) + 10 * \left(\frac{1}{ISD} \right)$$

O algoritmo geral utilizado foi o seguinte, (Linden, 2006):

1. Gere a população inicial. Deve ser garantida a criação do cromossomo que não irá filtrar nenhum coeficiente, do cromossomo que irá filtrar todos os coeficientes, do cromossomo que irá filtrar somente os coeficientes reconstruídos DWT, do cromossomo que irá filtrar somente os coeficientes da WPD e o cromossomo que irá filtrar somente o coeficiente de detalhe reconstruído;
2. Avalie a população obtendo seu *fitness*;
3. Faça o *crossover*: gere um número randômico *v* seguindo uma distribuição uniforme no intervalo [0 1]. Se este número estiver abaixo da taxa de mutação *pc* faça o cruzamento. Senão, selecione os próprios pais para a nova população. Em caso de cruzamento, selecione duas posições onde ocorrerá o cruzamento. Estas posições se referem ao intervalo de coeficientes a serem trocadas pelos dois pais;
4. Faça a mutação: para cada cromossomo gere um valor seguindo uma distribuição uniforme e se ele for menor que a taxa de mutação *pm* (definida pelo usuário) faça a mutação – que pode ser *swap*, *shift* ou *scramble*.
5. Faça o elitismo: selecione os dois melhores indivíduos e os inclua na nova população;
6. Selecione *n-número_de_sobreviventes* pares de pais por meio de roleta, torneio ou *ranking*, conforme especificado pelo usuário;
7. Atualize a população;
8. Repita os passos de 2 a 8 até atingir o número de gerações especificado.

Sobre as formas de seleção de indivíduos para cruzamento, tem-se o seguinte (Linden, 2006):

- *Roleta*: calcule a soma acumulada das probabilidades. Gere um número randômico de distribuição uniforme r para cada pai a ser encontrado e verifique a qual intervalo da soma acumulada ele pertence, o intervalo define o índice do pai dentro da população;
- *Ranking*: é semelhante a roleta, porém, utiliza ordenação de *fitness* para garantir maior uniformidade na roleta. Após ordenar os indivíduos, atribua a cada um o valor relativo a sua posição na ordenação de forma que o indivíduo mais bem qualificado fique com o valor n (tamanho da população) e o menos qualificado com o valor 1. Calcule a probabilidade acumulada com base no *ranking* construído e faça a roleta;
- Torneio: selecione uma quantidade definida de indivíduos e escolha o que apresentar melhor *fitness*.

Sobre os mecanismos de mutação, tem-se o seguinte:

- *Swap*: escolha dois genes no cromossomo e inverta-os;
- *Shift*: escolha duas posições no cromossomo, coloque o primeiro gene na posição do segundo e desloque os outros para frente;
- *Scramble*: escolha duas posições no cromossomo, troque aleatoriamente as posições dos genes dentro do intervalo especificado;

Diante desta nova maneira de encarar o problema, o diagrama de blocos da função objetivo do AG é exibido na Figura 4.6. Inicialmente, o sinal é contaminado com ruído e em seguida filtrado utilizando o filtro de Kalman no tempo. No passo seguinte, o sinal filtrado é decomposto utilizando a DWT em 5 níveis gerando 5 coeficientes de detalhe e 5 coeficientes de aproximação. Estes coeficientes são reconstruídos e a limiarização é realizada ou não, se realizada, pode ocorrer em quaisquer coeficientes. Feito isto, o sinal é novamente decomposto, desta vez por WPD, em profundidade 4 e então são gerados 16 coeficientes (os nós folha da árvore binária, conforme a Figura 4.2), que podem ser filtrados e limiarizados ou não. Por fim, o sinal é reconstruído, a SNR e a distância de Itakura são calculadas gerando o valor de *fitness*.

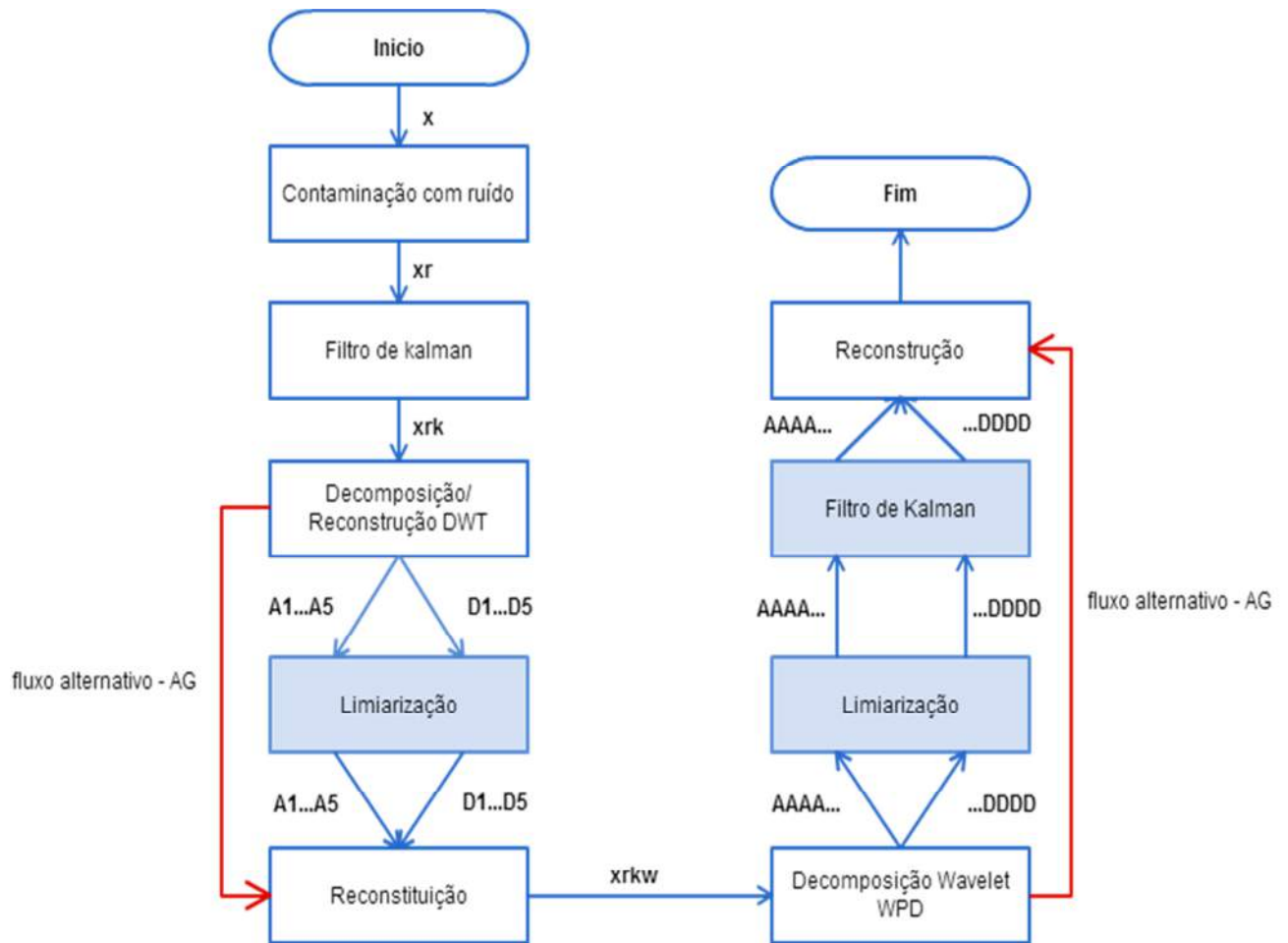


Figura 4.6 Diagrama de blocos do método proposto com refinamento por meio de algoritmo Genético.

Com a aplicação de AG, a filtragem e limiarização se tornam bastante flexíveis e com as condições iniciais da geração da população, garante-se que o espaço de possibilidades possuirá uma boa cobertura. Portanto, a adoção de algoritmos genéticos como parâmetro de refinamento torna a limiarização e a filtragem opcionais. Na próxima seção são discutidos os melhores indivíduos, sendo que estes conduzirão a filtragem/limiarização nos coeficientes wavelets nos experimentos. A escolha dos melhores indivíduos

4.3.2 Resultados do refinamento por meio de AG

Quanto aos parâmetros do AG, foi definida uma taxa de mutação de 80%, de *crossover* 40%, o tipo de *crossover* é PMX, a mutação é *swap* e a seleção é pelo método *rank*. Foi adotado o elitismo (6 pais foram preservados). A população inicial foi determinada com 30 indivíduos e o número de gerações 25.

Para os refinamentos por meio de AG foram utilizados 49 sinais de voz, sendo 25 femininos e 24 masculinos. As SNRs testadas foram: 0, 3, 5, 10, 15 e 20dB. Os resultados obtidos por sucessivas execuções do AG em diferentes níveis de intensidade de ruído podem ser vistos nas figuras a seguir.

Conforme Figuras 4.7 a 4.12, a filtragem Kalman não obteve nenhuma expressividade sendo recomendada somente a limiarização dos coeficientes, seja da DWT ou da WPD, sendo que isso pode ser observado em todos os níveis de ruído testados.

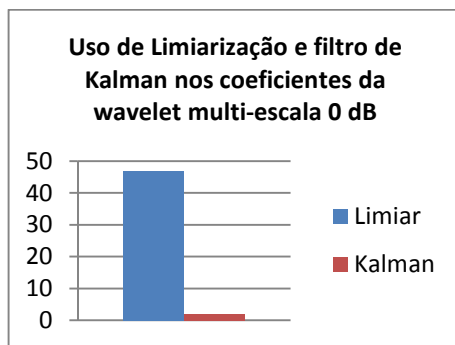


Figura 4.7 Limiarização x Kalman 0dB

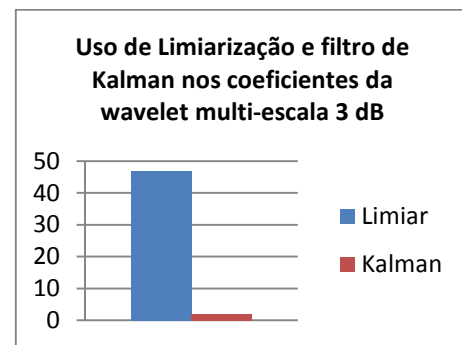


Figura 4.8 Limiarização x Kalman 3dB

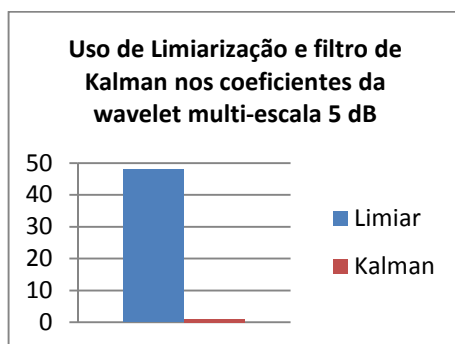


Figura 4.9 Limiarização x Kalman 5dB

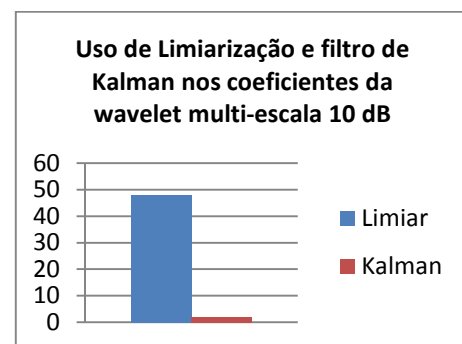


Figura 4.10 Limiarização x Kalman 10dB

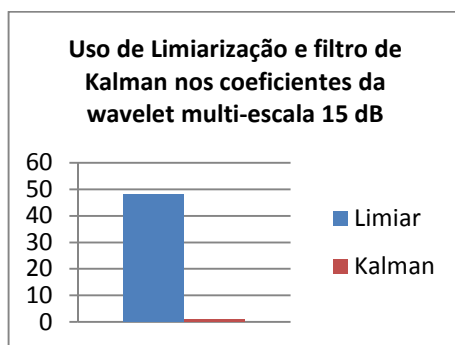


Figura 4.11 Limiarização x Kalman 15dB

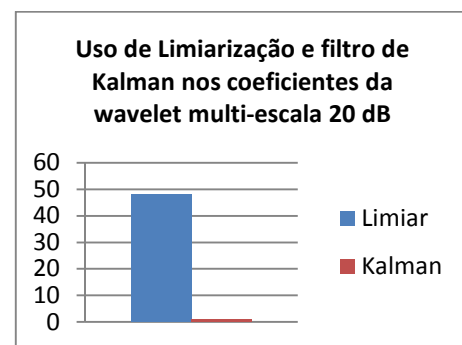


Figura 4.12 Limiarização x Kalman 0dB

Quanto aos coeficientes, observa-se (Figuras 4.13 a 4.18) que os valores da WPD foram mais úteis, principalmente os que carregam mais informações de detalhes que de aproximação. O único coeficiente que se destacou na DWT foi o coeficiente de detalhe D1, que teve desempenho bastante similar aos coeficientes da WPD.

Observa-se também que quanto menor a intensidade do ruído (alta SNR) diminui-se a quantidade de coeficientes a serem filtrados, notadamente, os coeficientes extremos do ramo direito (detalhe) são os mais utilizados em todos os casos, conforme era esperado.

Quanto ao limiar, a mutação do AG ajustou os valores e observou-se que uma taxa de 100% do valor calculado pelo Quadro 2 é a melhor alternativa.

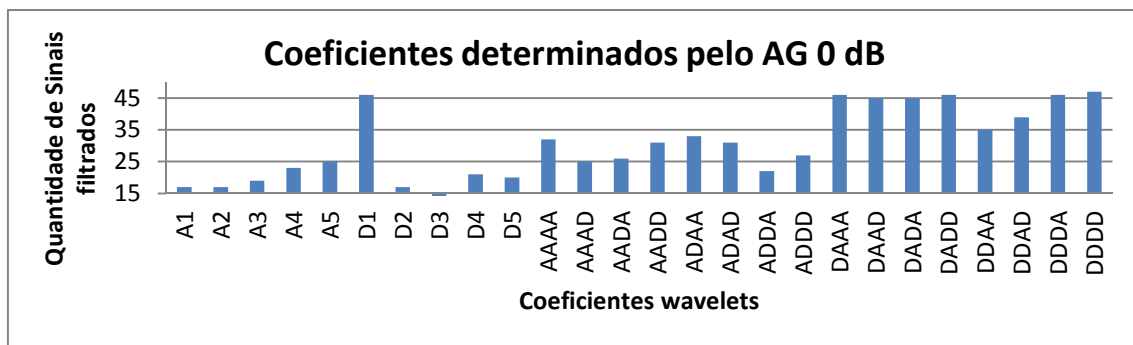


Figura 4.13 - Melhores coeficientes para SNR 0 dB

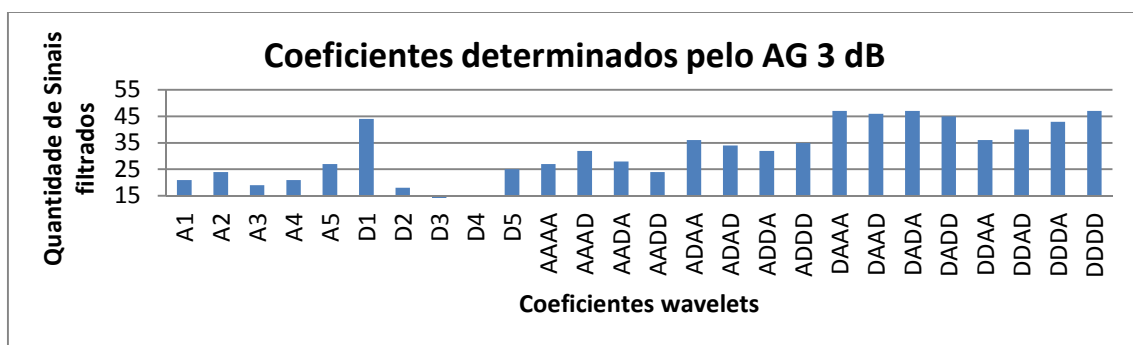


Figura 4.14 - Melhores coeficientes para SNR 3 dB

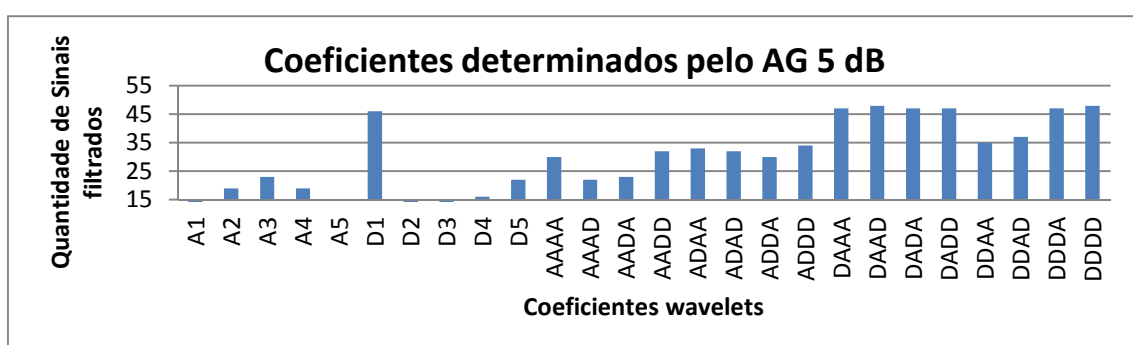


Figura 4.15 - Melhores coeficientes para SNR 0 dB

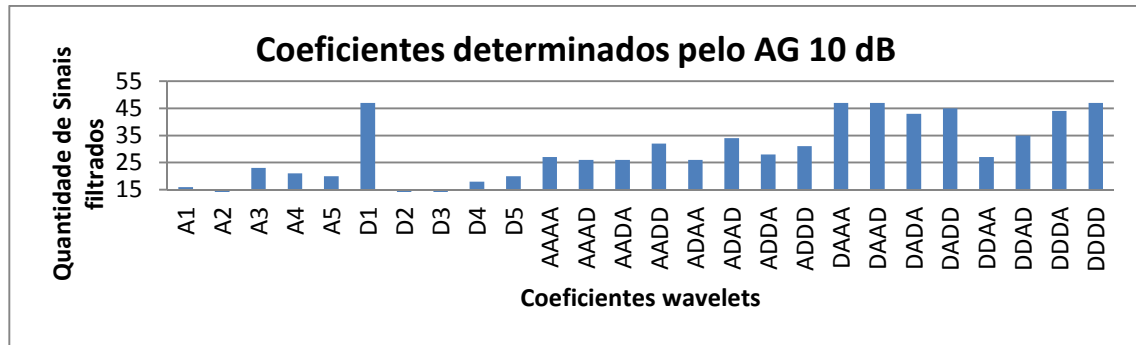


Figura 4.16 - Melhores coeficientes para SNR 10 Db

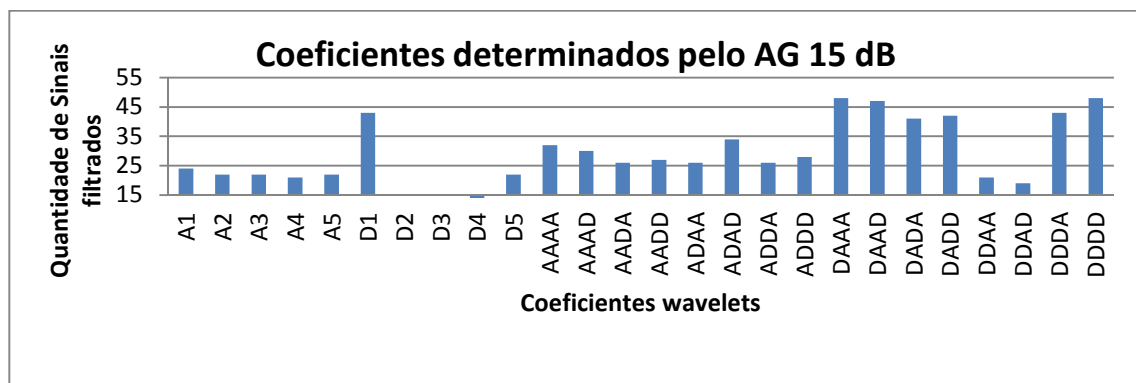


Figura 4.17 - Melhores coeficientes para SNR 15 dB

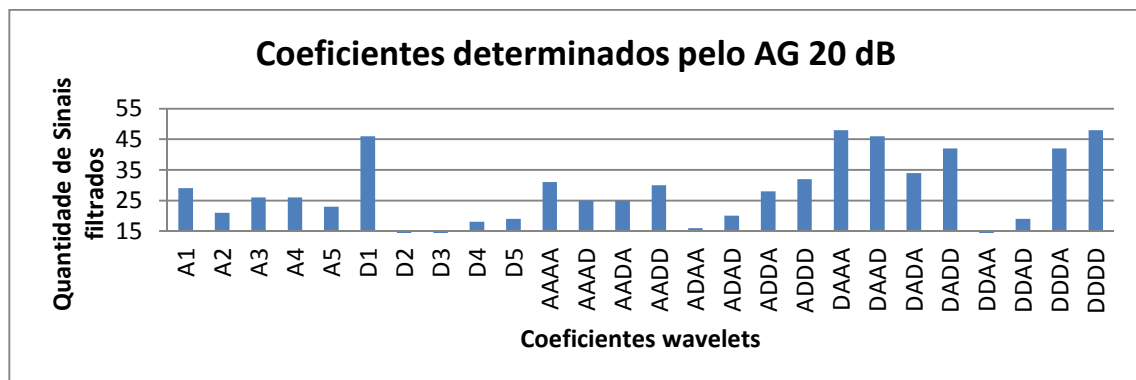


Figura 4.18 - Melhores coeficientes para SNR 20 dB

4.4 Considerações finais

O método proposto neste trabalho é um método híbrido composto por filtragem de Kalman e decomposição por meio de transformada wavelet com aplicação de limiarização nos coeficientes da transformada wavelet para a remoção de ruído. Em função da grande gama de possibilidades de combinações entre estes coeficientes foi aplicado um algoritmo genético no intuito de definir quais coeficientes limiarizados produzem melhor resultado.

Com a realização de 6 experimentos, cada um com uma SNR e contando com 49 sinais de voz, o AG determinou a melhor combinação de coeficientes, observando que quanto menor a quantidade de ruído, menos coeficientes devem ser filtrados. Como resultado observa-se que:

- A filtragem Kalman não obteve bom desempenho ao ser aplicada diretamente nos coeficientes da transformada Wavelet;
- O melhor coeficiente da DWT é o D1;
- Os melhores coeficientes da WPD são DAAA e DDDD (coeficientes 9 e 16), porém os coeficientes [10, 11, 14 e 15] tiveram bons desempenhos e podem ser bastante úteis;
- O melhor limiar é o valor encontrado no Quadro 2, (limiar *hard*), sendo que foram testados percentuais deste valor e chegou-se que 100% é a melhor alternativa.

No próximo capítulo são mostrados os testes de remoção de ruído em diversos sinais utilizando os coeficientes determinados neste capítulo. Além dos algoritmos mostrados anteriormente uma combinação dos dois é testada (Figura 4.19). Nesta nova combinação são usados o coeficiente de detalhe D1 da DWT e os melhores coeficientes da WPD determinados pelo AG em sequência, sendo que duas decomposições wavelets são realizadas. Estes três algoritmos são comparados entre si e por fim o melhor é comparado com subtração espectral.

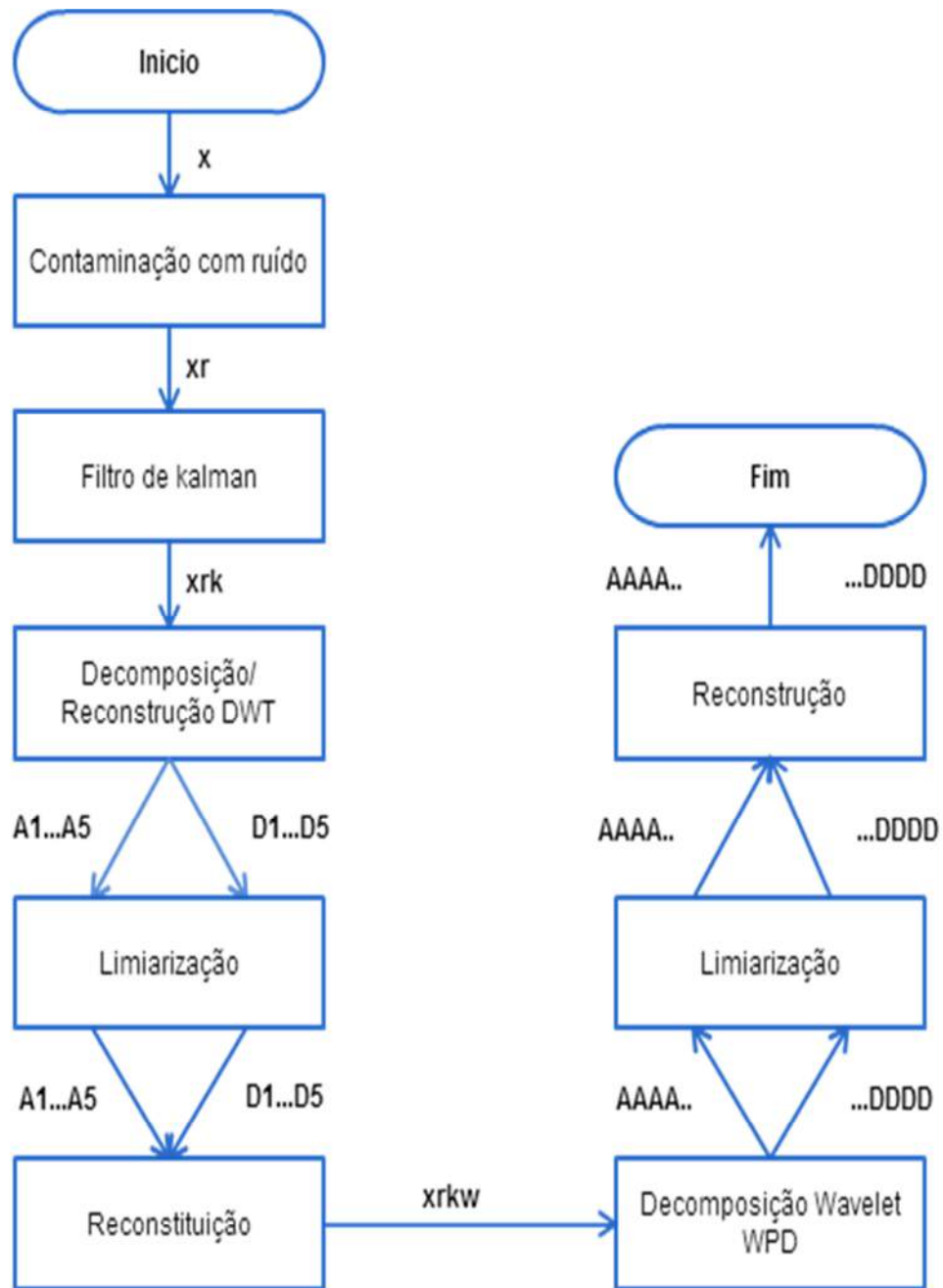


Figura 4.19 Algoritmo 3 gerado a partir da combinação dos algoritmos 1 e 2

5 EXPERIMENTOS E RESULTADOS

Como primeira etapa para definição do método proposto foi feita a decisão a respeito dos coeficientes wavelets a serem usados na limiarização utilizando algoritmos genéticos. Neste capítulo serão discutidos os experimentos realizados bem como os resultados obtidos para os três algoritmos apresentados anteriormente. Com a confrontação entre os mesmos, espera-se obter um algoritmo eficiente de redução de ruído, que ao final do capítulo é comparado à subtração espectral.

5.1 O experimento

Para os experimentos apresentados neste capítulo foram selecionados dez sinais de vozes masculinas e dez sinais de vozes femininas, gravados em diferentes frequências conforme mostra a Tabela 4. Os seis testes contemplaram as SNRs 0dB, 3dB, 5dB, 10dB, 15dB e 20dB⁹. Porém, apenas as SNRs de: 0dB, 5dB, 15dB e 20dB serão mostradas para que o capítulo não se torne demasiadamente grande. Os algoritmos são confrontados aos pares para melhor visualização dos resultados. Lembrando que a diferença entre um algoritmo e outro é somente quanto à transformada wavelet utilizada para decomposição wavelet e os coeficientes sendo filtrados. Tanto os coeficientes como o limiar foram determinados pelo AG, (Capítulo 4). Os algoritmos são:

- DWT – filtragem de Kalman, decomposição DWT e limiarização do coeficiente de Detalhe D1, (Figura 4.5(a));
- WPD – filtragem de Kalman, decomposição WPD e limiarização dos coeficientes [9 10 11 14 15 16], (Figura 4.5(b));
- DWTWPD – filtragem de Kalman, decomposição DWT e limiarização de D1, decomposição WPD e limiarização de [9 10 11 14 15 16], (Figura 4.19).

Utilizou-se apenas o ruído gaussiano branco nos testes iniciais, sendo que a cada teste entre dois algoritmos é gerado um ruído que seta a SNR de entrada (Equação 2.7) no sinal ruidoso.

⁹ SNR global, sendo que para SNR global a partir de 20dB o sinal é considerado bom e não necessita de filtragem.

Tabela 3 Sons testados

Masculinos		Femininos	
Arquivo	Freq.	Arquivo	Freq
bala	22050	bala_f	16000
bar	22050	bar_f	16000
casa	22050	casa_f	16000
elefante	22050	elefante_f	16000
eletrica	22050	eletrica_f	16000
fala	22050	fala_f	16000
janela	22050	janela_f	16000
maio	22050	maio_f	16000
mao	22050	mao_f	16000
melão	22050	melao_f	16000

5.2 DWT versus WPD

O primeiro teste contempla os algoritmos DWT e WPD. A Tabela 5 mostra a comparação dos algoritmos para os sinais de voz de entrada (vistos na Tabela 4) e a SNR de entrada é 0dB. São dadas a SNR segmentada de saída e distância de Itakura –Saito entre o sinal original e o sinal filtrado. É possível perceber que o algoritmo com decomposição DWT foi bastante superior ao algoritmo com decomposição WPD. Para todos os sinais a distância de Itakura-Saito foi melhor, obtendo inclusive um valor bastante satisfatório, afinal, como foi mencionado anteriormente, quanto menor a ISD, menor a dissemelhança entre o sinal original e o sinal filtrado. Já em relação a SNR de entrada, DWT foi apenas um pouco melhor, possivelmente por ter perdido parte da potência da sinal, enquanto que WPD perdeu pouco da potência do sinal, mas também não removeu o ruído.

A audição do sinal de saída deixa bastante evidente a superioridade de DWT, e ainda, a partir da SNR segmentada é possível ver o quanto o algoritmo conseguiu melhorar o sinal.

Para SNR de entrada igual a 5dB e 15dB (Tabelas 6 e 7), DWT resulta em uma distância de Itakura melhor, mas com SNR segmentada pior que WPD. Neste caso, o desempate é pela distância de Itakura, e de fato, é possível constatar ouvindo o sinal filtrado que a DWT produz é mais limpo. Já o sinal com SNR de 20dB de entrada (Tabela 8), as coisas se invertem e aparentemente WPD se sai melhor, mas isso não se observa na prática, pois a diferença na distância de Itakura é muito pequena, produzindo os sons muito parecidos.

Tabela 4 Comparativo DWT x WPD, SNRi 0dB

ARQUIVO	Algoritmo 1			Algoritmo 2	
	SNRIseg	SNROseg_DWT	ITAKURA_DWT	SNROseg_WPD	ITAKURA_WPD
bala	-10	1	0,74	1	1,55
bar	-9	3	0,77	3	1,77
casa	-11	4	0,24	2	0,53
elefante	-12	1	0,47	1	0,77
eletrica	-14	3	0,38	3	0,98
fala	-9	4	0,49	2	1,24
janela	-7	2	0,34	3	0,99
maio	-8	0	0,43	1	1,24
mao	-8	3	0,59	4	1,38
melão	-7	2	0,22	2	1,29
bala_f	-8	4	0,89	3	1,77
bar_f	-12	3	1,03	1	1,70
casa_f	-7	4	0,59	3	1,36
elefante_f	-9	3	0,64	3	0,85
eletrica_f	-7	3	0,43	4	1,30
fala_f	-8	5	0,77	3	1,67
janela_f	-5	3	0,35	4	1,00
maio_f	-6	1	0,63	3	1,45
mao_f	-5	3	0,57	4	1,56
melao_f	-5	1	0,63	2	1,48
Média	-8,35	2,95	0,56	2,7	1,29

Tabela 5 Comparativo DWT x WPD, SNRi 5dB

ARQUIVO	Algoritmo 1			Algoritmo 2	
	SNRIseg	SNROseg_DWT	ITAKURA_DWT	SNROseg_WPD	ITAKURA_WPD
bala	-7	3	0,66	2	1,31
bar	-6	3	0,61	2	1,43
casa	-8	3	0,25	3	0,39
elefante	-9	3	0,58	2	0,57
eletrica	-11	1	0,37	0	0,80
fala	-6	5	0,47	3	1,02
janela	-4	2	0,30	3	0,77
maio	-5	0	0,41	1	1,00
mao	-5	1	0,54	3	1,15
melão	-4	1	0,18	2	1,02
bala_f	-5	3	0,78	3	1,47
bar_f	-9	5	0,98	3	1,46
casa_f	-4	1	0,58	3	1,10
elefante_f	-6	5	0,41	4	1,20
eletrica_f	-4	4	0,36	3	1,07
fala_f	-5	5	0,70	5	1,38
janela_f	-3	2	0,40	2	0,81
maio_f	-4	1	0,60	4	1,16
mao_f	-3	2	0,48	3	1,29
melao_f	-3	0	0,63	2	1,16
Média	-5,55	2,5	0,51	2,65	1,08

Tabela 6 Comparativo DWT x WPD, SNRi 15dB

	Algoritmo 1			Algoritmo 2	
ARQUIVO	SNRIseg	SNROseg_DWT	ITAKURA_DWT	SNROseg_WPD	ITAKURA_WPD
bala	-3	2	0,54	1	0,82
bar	-2	3	0,56	3	0,92
casa	-3	1	0,25	1	0,22
elefante	-5	3	0,51	2	0,29
eletrica	-7	0	0,26	0	0,43
fala	-2	3	0,44	2	0,65
janela	-1	-1	0,29	2	0,45
maio	-2	0	0,35	2	0,59
mao	-2	0	0,39	3	0,71
melão	-2	1	0,18	2	0,60
bala_f	-2	1	0,71	1	0,95
bar_f	-4	3	0,95	3	0,99
casa_f	-1	1	0,71	2	0,67
elefante_f	-2	4	0,47	2	0,77
eletrica_f	-1	1	0,35	3	0,65
fala_f	-2	2	0,70	2	0,87
janela_f	-1	0	0,60	1	0,53
maio_f	-1	-1	0,74	2	0,74
mao_f	0	0	0,48	1	0,73
melao_f	-1	-1	0,76	1	0,77
Média	-2,20	1,10	0,51	1,8	0,67

Tabela 7 Comparativo DWT x WPD, SNRi 20dB

	Algoritmo 1			Algoritmo 2	
ARQUIVO	SNRIseg	SNROseg_DWT	ITAKURA_DWT	SNROseg_WPD	ITAKURA_WPD
bala	-2	2	0,53	2	0,64
bar	-1	2	0,56	2	0,70
casa	-2	-1	0,27	1	0,14
elefante	-3	2	0,46	1	0,18
eletrica	-5	1	0,25	0	0,30
fala	-1	2	0,44	1	0,50
janela	0	2	0,32	1	0,34
maio	-1	-1	0,34	1	0,42
mao	-1	-1	0,37	2	0,54
melão	-1	0	0,20	2	0,42
bala_f	-1	1	0,71	1	0,75
bar_f	-2	2	0,97	2	0,84
casa_f	0	-1	0,78	1	0,54
elefante_f	-1	3	0,49	1	0,61
eletrica_f	-1	0	0,37	1	0,50
fala_f	-1	1	0,73	1	0,70
janela_f	0	-2	0,63	1	0,43
maio_f	0	-2	0,83	1	0,59
mao_f	0	0	0,52	1	0,54
melao_f	0	-1	0,77	1	0,62
Média	-1,15	0,25	0,53	1,2	0,51

5.3 DWT versus DWTWPD

O comparativo entre DWT e DWTWPD mostrou algo bastante interessante: os números são praticamente idênticos uma vez que a diferença só ocorre a partir da 3ª casa decimal mesmo considerando todas as SNRs de entrada avaliadas (Tabelas 9, 10, 11 e 12). A justificativa para este fato é que a filtragem WPD não está sendo efetiva, com isso o algoritmo DWTWPD recai em DWT. Para averiguar esta hipótese, testes com limiares variando de 10% a 120% do valor original foram testados e ainda assim os resultados obtidos foram totalmente equivalentes.

Portanto, o algoritmo DWTWPD precisa de melhorias para que possa se beneficiar de todas as possibilidades oferecidas pela WPD.

Tabela 8 Comparativo DWT x DWTWPD, SNRi 0dB

ARQUIV	Algoritmo 1			Algoritmo 3	
	SNRIse	SNROseg_DW	ITAKURA_DWTWP	SNROseg_WP	ITAKURA_WPDWP
bala	-10	4	0,73	4	0,73
bar	-9	3	0,84	6	0,84
casa	-11	2	0,27	4	0,27
elefante	-12	1	0,65	1	0,65
eletrica	-14	1	0,43	1	0,43
fala	-9	8	0,49	8	0,49
janela	-7	3	0,32	3	0,32
maio	-7	0	0,44	0	0,45
mao	-8	3	0,84	3	0,84
melão	-7	3	0,21	3	0,21
bala_f	-8	2	0,88	2	0,88
bar_f	-12	5	1,03	5	1,03
casa_f	-7	5	0,59	5	0,59
elefante_f	-9	4	0,49	4	0,50
eletrica_f	-7	7	0,39	7	0,39
fala_f	-8	3	0,82	3	0,82
janela_f	-5	1	0,37	1	0,36
maio_f	-6	2	0,62	2	0,62
mao_f	-5	5	0,59	5	0,59
melao_f	-5	2	0,67	2	0,67
Média	-8,3	3,2	0,58	3,5	0,58

Tabela 9 Comparativo DWT x DWTWPD, SNRi 5dB

Algoritmo 1				Algoritmo 3	
ARQUIV	SNRIse	SNROseg_DW	ITAKURA_DWTWP	SNROseg_WP	ITAKURA_WPDWP
bala	-7	2	0,65	2	0,65
bar	-6	4	0,61	4	0,61
casa	-7	3	0,23	3	0,23
elefante	-9	3	0,63	3	0,63
eletrica	-11	1	0,36	1	0,36
fala	-6	4	0,46	4	0,46
janela	-4	1	0,29	1	0,29
maio	-5	0	0,41	0	0,41
mao	-5	2	0,53	2	0,53
melão	-4	1	0,18	1	0,18
bala_f	-5	5	0,77	5	0,77
bar_f	-9	5	0,95	5	0,95
casa_f	-4	3	0,58	3	0,58
elefante_f	-6	5	0,41	5	0,41
eletrica_f	-4	3	0,35	3	0,35
fala_f	-5	4	0,70	4	0,69
janela_f	-3	1	0,38	1	0,38
maio_f	-4	0	0,60	0	0,60
mao_f	-3	3	0,47	3	0,47
melao_f	-3	1	0,64	1	0,64
Média	-5,5	2,55	0,51	2,55	0,51

Tabela 10 Comparativo DWT x DWTWPD, SNRi 15dB

Algoritmo 1				Algoritmo 3	
ARQUIV	SNRIse	SNROseg_DW	ITAKURA_DWTWP	SNROseg_WP	ITAKURA_WPDWP
bala	-3	2	0,54	2	0,54
bar	-2	3	0,57	3	0,57
casa	-3	0	0,24	0	0,24
elefante	-5	2	0,51	2	0,51
eletrica	-7	1	0,27	1	0,27
fala	-2	3	0,44	3	0,44
janela	-1	-1	0,29	-1	0,29
maio	-2	0	0,35	0	0,35
mao	-2	0	0,39	0	0,39
melão	-2	1	0,18	1	0,18
bala_f	-2	2	0,71	2	0,71
bar_f	-4	3	0,95	3	0,95
casa_f	-1	1	0,74	1	0,74
elefante_f	-2	3	0,47	3	0,47
eletrica_f	-1	1	0,35	1	0,35
fala_f	-2	2	0,71	2	0,71
janela_f	0	-1	0,60	-1	0,60
maio_f	-1	-1	0,76	-1	0,76
mao_f	0	1	0,48	1	0,48
melao_f	-1	0	0,75	0	0,75
Média	-2,2	1,1	0,52	1,1	0,52

Tabela 11 Comparativo DWT x DWTWPD, SNRi 20dB

ARQUIV	Algoritmo 1			Algoritmo 3	
	SNRIse	SNROseg_DW	ITAKURA_DWTWP	SNROseg_WP	ITAKURA_WPDWP
bala	-2	3	0,53	3	0,53
bar	-1	2	0,56	2	0,56
casa	-2	0	0,27	0	0,27
elefante	-3	2	0,47	2	0,47
eletrica	-5	1	0,25	1	0,25
fala	-1	2	0,44	2	0,44
Janela	0	-2	0,32	-2	0,32
maio	-1	0	0,34	0	0,34
mao	-1	-1	0,36	-1	0,36
melão	-1	1	0,21	1	0,21
bala_f	-1	1	0,71	1	0,71
bar_f	-2	2	0,97	2	0,96
casa_f	0	0	0,78	0	0,78
elefante_f	-1	2	0,49	2	0,49
eletrica_f	-1	0	0,36	0	0,36
fala_f	-1	1	0,74	1	0,74
janela_f	0	-2	0,63	-2	0,63
maio_f	0	-2	0,82	-2	0,82
mao_f	0	0	0,52	0	0,52
melao_f	0	-1	0,79	-1	0,80
Média	-1,15	0,5	0,53	0,5	0,53

5.4 WPD versus DWTWPD

Por último é feita a comparação entre o WPD e o DWTWPD. Conforme os resultados obtidos anteriormente, espera-se que DWTWPD se saia melhor, uma vez que a efetividade da WPD está comprometida. De fato, para SNR de entrada igual a 0 e 5dB (Tabelas 13 e 14) este algoritmo é consideravelmente melhor. Já com 15dB (Tabela 15), a SNR segmentada é melhor no WPD. Ou seja: a DWT introduz perda do sinal em relação ao ruído conforme foi verificado em um teste anterior. Ainda assim, a distância de Itakura é melhor para DWTWPD o que deixa o sinal de saída bem melhor.

Já com a SNR de entrada em 20 dB (Tabela 16), novamente a situação se inverte. Mas da mesma maneira observada anteriormente, os valores próximos de distância de Itakura não precisam exatamente a qualidade do sinal de fala resultante após filtragem. E novamente uma audição do sinal filtrado leva a DWTWPD como sendo o melhor algoritmo, mas não por causa da limiarização nos coeficientes da WPD, mas sim por causa da DWT.

Tabela 12 Comparativo WPD x DWTWPD, SNRi 0dB

ARQUIV	Algoritmo 2			Algoritmo 3	
	SNRIse	SNROseg_WP	ITAKURA_WP	SNROseg_DWTWP	ITAKURA_DWTWP
bala	-10	2	1,56	0	0,75
bar	-9	1	1,71	2	0,74
casa	-11	1	0,56	2	0,26
elefante	-12	1	0,77	2	0,67
eletrica	-14	4	0,94	7	0,44
fala	-9	2	1,25	5	0,48
janela	-7	3	0,98	3	0,32
maio	-7	0	1,26	-1	0,45
mao	-8	4	1,37	3	0,59
melão	-7	5	1,25	7	0,21
bala_f	-8	3	1,81	3	0,90
bar_f	-12	1	1,67	3	1,02
casa_f	-7	5	1,34	5	0,60
elefante_f	-9	3	1,44	4	0,56
eletrica_f	-7	4	1,29	5	0,39
fala_f	-8	4	1,68	6	0,79
janela_f	-5	3	1,02	1	0,36
maio_f	-6	2	1,41	1	0,61
mao_f	-5	3	1,57	3	0,58
melao_f	-5	3	1,46	2	0,65
Média	-8,3	2,7	1,32	3,2	0,57

Tabela 13 Comparativo WPD x DWTWPD, SNRi 5dB

ARQUIV	Algoritmo 2			Algoritmo 3	
	SNRIse	SNROseg_WP	ITAKURA_WP	SNROseg_DWTWP	ITAKURA_DWTWP
bala	-7	2	1,30	3	0,66
bar	-6	2	1,40	2	0,61
casa	-8	2	0,40	4	0,24
elefante	-9	2	0,56	3	0,63
eletrica	-11	1	0,78	3	0,35
fala	-6	3	1,01	5	0,46
janela	-4	3	0,77	2	0,30
maio	-5	1	1,01	0	0,40
mao	-5	3	1,15	2	0,52
melão	-4	3	1,04	2	0,19
bala_f	-5	1	1,45	2	0,78
bar_f	-9	3	1,44	6	0,97
casa_f	-4	4	1,07	3	0,57
elefante_f	-6	4	1,15	6	0,67
eletrica_f	-4	4	1,08	3	0,35
fala_f	-5	2	1,38	3	0,70
janela_f	-3	1	0,80	1	0,38
maio_f	-4	2	1,17	0	0,58
mao_f	-3	2	1,29	3	0,46
melao_f	-3	2	1,16	1	0,63
Média	-5,6	2,4	1,07	2,7	0,52

Tabela 14 Comparativo WPD x DWTWPD, SNRi 15dB

ARQUIV	Algoritmo 2			Algoritmo 3	
	SNRIse	SNROseg_WP	ITAKURA_WP	SNROseg_DWTWP	ITAKURA_DWTWP
bala	-3	2	0,84	3	0,55
bar	-2	2	0,91	2	0,56
casa	-3	2	0,23	1	0,26
elefante	-5	2	0,29	2	0,51
eletrica	-7	0	0,43	0	0,26
fala	-2	2	0,64	3	0,44
janela	-1	2	0,45	-1	0,29
maio	-2	2	0,59	0	0,35
mao	-2	3	0,71	0	0,40
melão	-2	2	0,60	1	0,18
bala_f	-2	2	0,94	1	0,71
bar_f	-4	3	0,99	3	0,95
casa_f	-1	1	0,68	0	0,71
elefante_f	-2	2	0,77	3	0,47
eletrica_f	-1	1	0,66	1	0,35
fala_f	-2	2	0,88	2	0,71
janela_f	-1	1	0,53	-1	0,60
maio_f	-1	2	0,74	-1	0,76
mao_f	0	2	0,73	0	0,48
melao_f	-1	1	0,77	0	0,74
Média	-2,2	1,8	0,67	1,0	0,51

Tabela 15 Comparativo WPD x DWTWPD, SNRi 20dB

ARQUIV	Algoritmo 2			Algoritmo 3	
	SNRIse	SNROseg_WP	ITAKURA_WP	SNROseg_DWTWP	ITAKURA_DWTWP
bala	-2	2	0,63	2	0,53
bar	-1	2	0,69	2	0,56
casa	-2	1	0,14	-1	0,27
elefante	-3	1	0,18	2	0,47
eletrica	-5	0	0,30	0	0,25
fala	-1	1	0,49	2	0,45
janela	0	1	0,34	-2	0,32
maio	-1	2	0,43	-1	0,34
mao	-1	2	0,54	-1	0,37
melão	-1	2	0,42	1	0,20
bala_f	-1	1	0,75	1	0,71
bar_f	-2	2	0,82	2	0,97
casa_f	0	1	0,53	0	0,76
elefante_f	-1	1	0,62	3	0,50
eletrica_f	-1	1	0,51	0	0,36
fala_f	-1	2	0,69	1	0,74
janela_f	0	1	0,43	-2	0,63
maio_f	0	1	0,59	-1	0,82
mao_f	0	1	0,53	0	0,52
melao_f	0	1	0,62	-1	0,78
Média	-1,2	1,3	0,51	0,4	0,53

5.5 Conclusões preliminares

Conforme discutido nas seções anteriores, entre as 3 alternativas de algoritmo de redução de ruído baseado em filtro de Kalman e transformada wavelet, a melhor alternativa é utilizar a WDP (transformada discreta de wavelet) para decompor o sinal a fim de limiarizar os coeficientes gerados da decomposição. Seus resultados mostraram superioridade tanto em medidas de qualidade, (usando as métricas de Itakura-Saito e SNR segmentada), como em testes de audição. E isso ocorreu para quase todas as SNRs de entrada. Em altas SNRs, o ganho na SNR segmentada de saída é menos relevante que nos outros algoritmos, mas ainda assim, o som resultante é melhor, pois a distância de Itakura se mantém em uma boa margem.

A Tabela 17 apresenta o teste de reconhecimento de palavras realizado com 10 pessoas, sendo que os sons foram passados em ordem aleatória. Cada som foi repetido 4 vezes, aleatoriamente. Todas as pessoas afirmaram desconhecer qualquer problema auditivo. As idades dessas pessoas variam de 9 a 65 anos. O ambiente utilizado para o teste possui um isolamento acústico suficiente e para se ter maiores garantias de isolamento foi utilizado um fone de ouvido do tipo *headphone*. O ouvinte escutava o som e informava o que tinha escutado. Foi testado apenas o sinal filtrado após a adição de ruído gaussiano branco, com SNR global de entrada de 0dB uma vez que estes testes são bastante limitados em virtude do tempo e diversidade de ouvintes que demandam, (Rabiner, et al., 2011).

Apesar, deste teste possuir uma subjetividade muito grande, os números trazem informações que confirmam as análises realizadas pelas medidas de qualidade e pelos gráficos. O algoritmo que utiliza a decomposição por DWT se saiu melhor na maioria dos casos e menos erros foram cometidos pelas pessoas. A WPD apresentou um desempenho ligeiramente inferior e DWTWPD apresentou desempenho idêntico a DWT, conforme esperado. Algumas palavras, devido a semelhança com outras palavras da língua portuguesa, causaram maior confusão. É o caso da palavra ‘bar’ onde algumas pessoas confundiram com ‘par’. Já a palavra ‘elefante’ que foi confundida com ‘elegante’. Ao final foram mostrados sons filtrados por DWT e DWTWPD para que as pessoas optassem pelo melhor. Embora não fosse consensual, a maioria tinha decidido por DWT – aproximadamente 70%.

Tabela 16 Teste de reconhecimento com ouvintes

ARQUIVO	DWT	WPD	DWTWPD
bala	97,5	95	97,5
bar	92,5	90	92,5
casa	92,5	90	92,5
elefante	97,5	90	95
eletrica	100	100	100
fala	90	82,5	87,5
janela	100	100	100
maio	97,5	92,5	97,5
mao	95	95	95
melão	90	92,5	90
bala_f	92,5	87,5	90
bar_f	90	85	90
casa_f	87,5	85	87,5
elefante_f	90	90	90
eletrica_f	100	97,5	100
fala_f	87,5	82	87,5
janela_f	100	100	100
maio_f	100	85	100
mao_f	100	97,5	100
melao_f	100	95	100

A Figura 5.1 mostra um comparativo entre os três algoritmos considerando as medidas de SNR segmentada e a distância de Itakura-Saito para todas as SNR de entrada testadas. Quanto à SNRseg de saída, a variação é muito pequena e oscilante. Hora um algoritmo é melhor, hora outro se destaca. No entanto, quando se considera a distância de Itakura-Saito, DWT e DWTWPD se destacam muito em relação à WPD e isso reflete na qualidade do sinal resultante, que é bem melhor. Esses algoritmos conseguem de fato reduzir o ruído ao passo que mantém o sinal com pouca distorção.

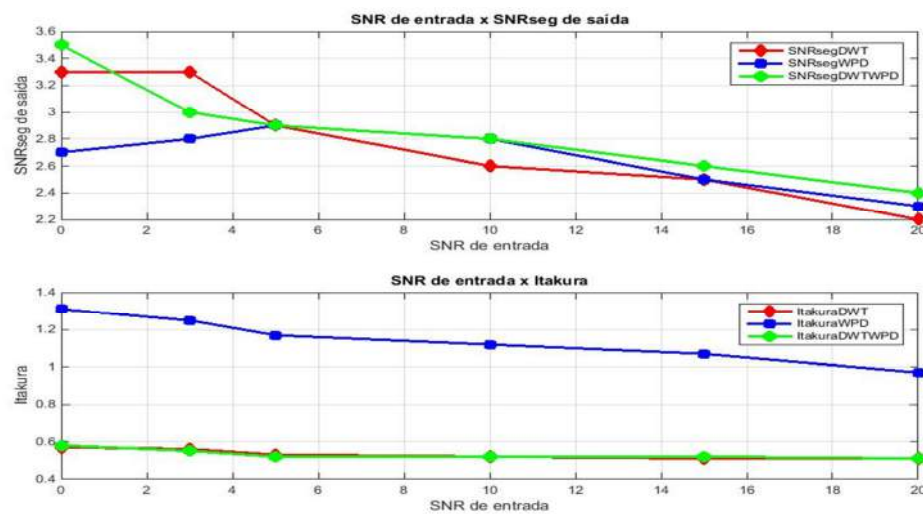


Figura 5.1 Comparativo entre DWT, WPD e DWTWPD para SNR segmentada e Itakura-Saito

As Figuras 5.2 e 5.3 apresentam, respectivamente, as amostras do sinal de fala e o espectrograma do sinal de voz da palavra ‘elétrica’ contaminada por ruído branco com SNR de entrada 0dB. As figuras foram divididas em quatro partes: sinal original, sinal contaminado, sinal filtrado DWT, sinal filtrado WPD.

Na Figura 5.2 tem-se a visualização das amostras do sinal de fala, e os efeitos da contaminação do ruído (SNR de entrada 0dB) e da filtragem com os algoritmos DWT e WPD. É possível perceber que ambos removeram bastante ruído, e que criaram pequenas distorções ao longo do sinal. DWT removeu mais ruído que WPD, e por isso as suas distorções estão mais aparentes. Tais distorções são provenientes do filtro de Kalman. A Figura 5.4 também apresenta o mesmo tipo de gráfico, só que com SNR 20dB. O sinal resultante dos dois métodos ficou bastante semelhante, porém, o sinal filtrado com DWT ficou mais parecido com o sinal original, já que esta remove melhor o ruído.

Na Figura 5.3, observando o espectrograma do sinal original é possível perceber trechos sem atividade de voz e trechos nos quais o sinal de voz possui mais intensidade, que são os tons mais quentes. O mesmo não pode ser dito quando o sinal é contaminado com ruído, pois existe um espalhamento da intensidade do sinal dificultando sua identificação no tempo. O espectrograma do sinal filtrado utilizando a DWT promove um espalhamento da intensidade, o que não ocorre com a filtragem WPD que reduz tanto intensidade do ruído quanto do sinal de voz. Ainda assim, o nível de espalhamento produzido pela DWT não é intenso e por isso permite boa inteligibilidade, uma vez que reduz muito mais o ruído que a WPD. Isso nítido na audição do sinal processado. O mesmo pode ser observado para SNR de entrada alta, Figura 5.5, em que se tem uma relação sinal ruído de entrada de 20 dB com a diferença que os momentos de silêncio são melhor visualizados para a DWT.

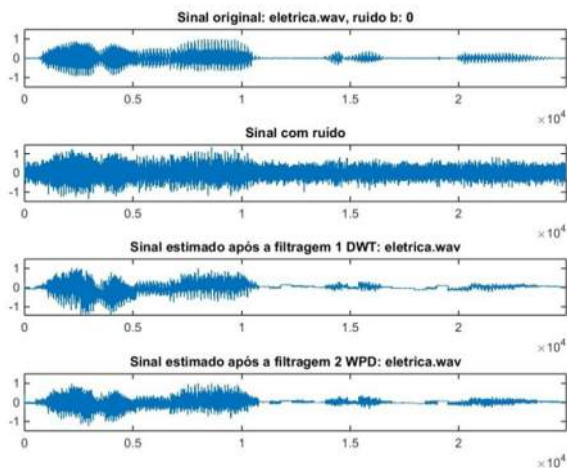


Figura 5.2 Comparativo entre DWT e WPD em 0dB, 'elétrica' voz masculina

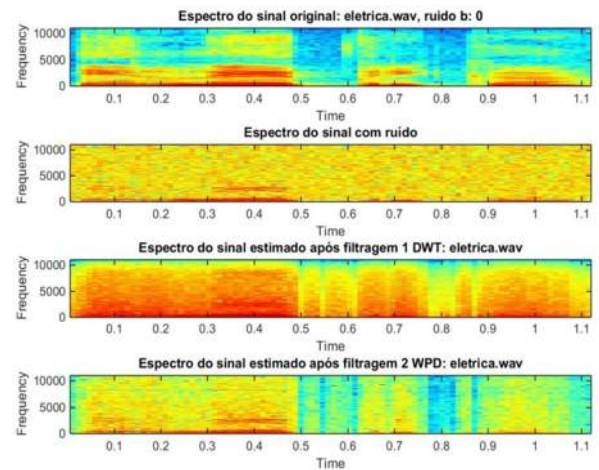


Figura 5.3 Espectrograma entre DWT e WPD em 0dB, 'elétrica' voz masculina

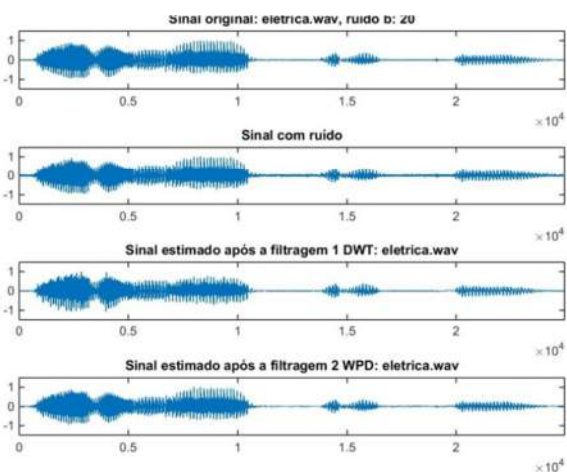


Figura 5.4 Comparativo entre DWT e WPD em 20dB, 'elétrica' voz masculina

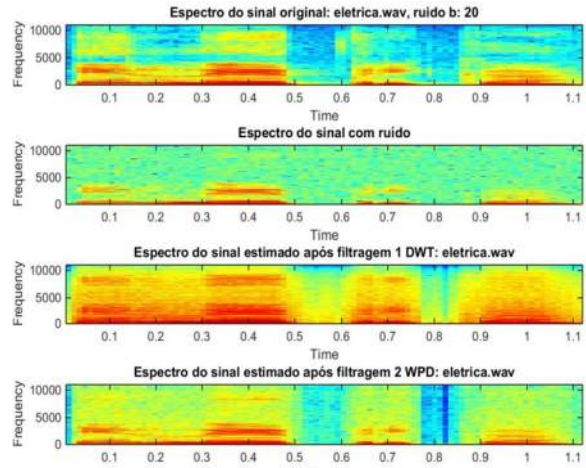


Figura 5.5 Espectrograma entre DWT e WPD em 20dB, 'elétrica' voz masculina

As Figuras 5.6 e 5.8 mostram o sinal de voz 'elétrica' gravado a partir de voz feminina, com SNR 0dB e 20dB respectivamente e filtragem usando os algoritmos DWT e DWTWPD. Observa-se que em 0dB houve intensa redução de ruído, porém, houve também a remoção de parte do sinal de voz, e isso ocorreu de maneira semelhante para os dois algoritmos. Enquanto, que para a SNR 20dB praticamente não existe diferença entre o sinal original e os sinais provenientes da filtragem. Nos espectrogramas deste sinal, Figuras 5.7 e 5.9, correspondentes à SNR 0dB e 20 dB respectivamente. Observa-se que o sinal filtrado com DWTWPD gerou menos espalhamento nas duas SNRs de entrada que DWT.

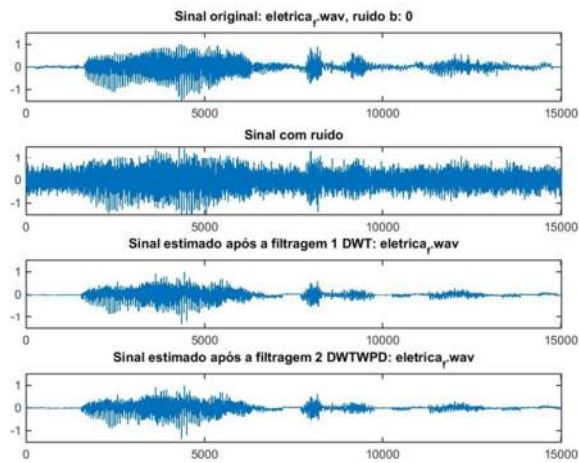


Figura 5.6 Comparativo entre DWT e DWTWPD em 0dB, 'elétrica' voz feminina

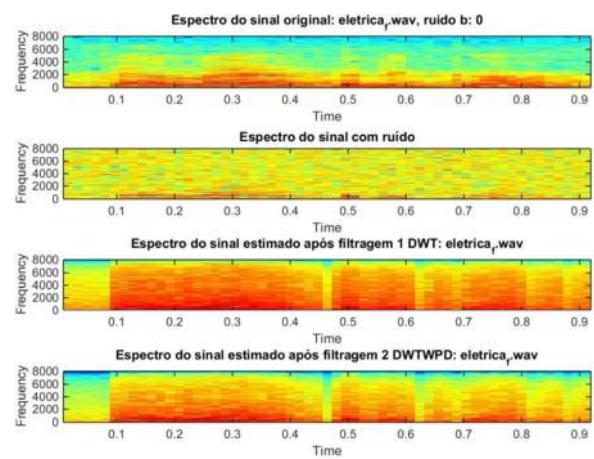


Figura 5.7 Espectrograma entre DWT e DWTWPD em 0dB, 'elétrica' voz feminina

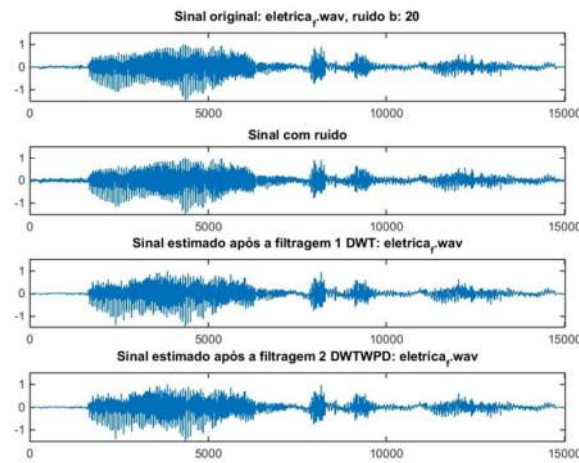


Figura 5.8 Comparativo entre DWT e DWTWPD em 20dB, 'elétrica' voz feminina

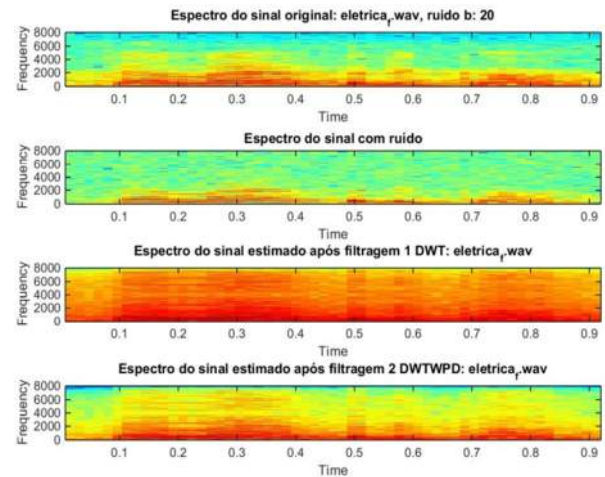


Figura 5.9 Espectrograma entre DWT e DWTWPD em 20dB, 'elétrica' voz feminina

5.6 Comparação entre DWT e subtração espectral

Como teste final foi feita a comparação entre DWT e subtração espectral que é um método bastante conhecido. Nos testes anteriores foi verificado que a decomposição por DWT gera os melhores resultados, assim como DWTWPD que é basicamente idêntica. A seguir, são apresentadas três comparações: por amostra, por espectrogramas e por medidas, considerando as SNRs de entrada 0dB e 20dB. O sinal de entrada é a palavra ‘bala’ gravada por uma voz masculina.

5.6.1 Comparação baseada em amostra do sinal

As Figuras 5.10 e 5.11 mostram o resultado da redução de ruído para as SNRs 0dB e 20dB, respectivamente. São mostrados quatro sinais: o sinal original, o sinal ruidoso, o sinal filtrado utilizando DWT e o sinal filtrado utilizando subtração espectral. Para SNR 0dB, o sinal filtrado por DWT possui distorções. Porém, teve quase todo o ruído removido, enquanto que o sinal filtrado por meio de subtração espectral possui bastante ruído, até mesmo em momentos de silêncio percebe-se a presença do ruído. Por meio de audição, percebe-se que o ruído resultante da subtração espectral é mais incomodo que o ruído remanescente da DWT. Alguns autores classificam o ruído resultante da subtração espectral como ruído musical (Fujimoto, et al., 2000).

Já na SNR de entrada 20dB, aparentemente, a subtração espectral se saiu melhor pois o sinal é bastante parecido com o sinal original. No entanto, também foi detectada a presença de ruído musical.

5.6.2 Comparação baseada em espectrograma

As Figuras 5.12 e 5.13 mostram os espectrogramas para SNR de entrada 0dB e 20dB. O sinal de entrada possui bastante energia ao longo do tempo, sendo ela concentrada até 4000 Hz. A filtragem por DWT conseguiu remover as frequências espúrias espalhadas relativamente bem, mas espalhou a energia do sinal. Já a subtração espectral removeu estas frequências espúrias, mas acrescentou outros artefatos de frequência, conforme observado anteriormente (trata-se de ruído musical). Quanto à energia do sinal, a subtração espectral a concentrou em até 2500 Hz.

Quanto a filtragem do sinal com SNR global de entrada 20dB, percebe-se grande melhora proporcionada pela DWT, que removeu o ruído quase completamente e manteve a energia do

sinal nas faixas de frequência originais. Enquanto que a subtração espectral, eliminou grande parte da energia do sinal e inseriu novamente o ruído musical.

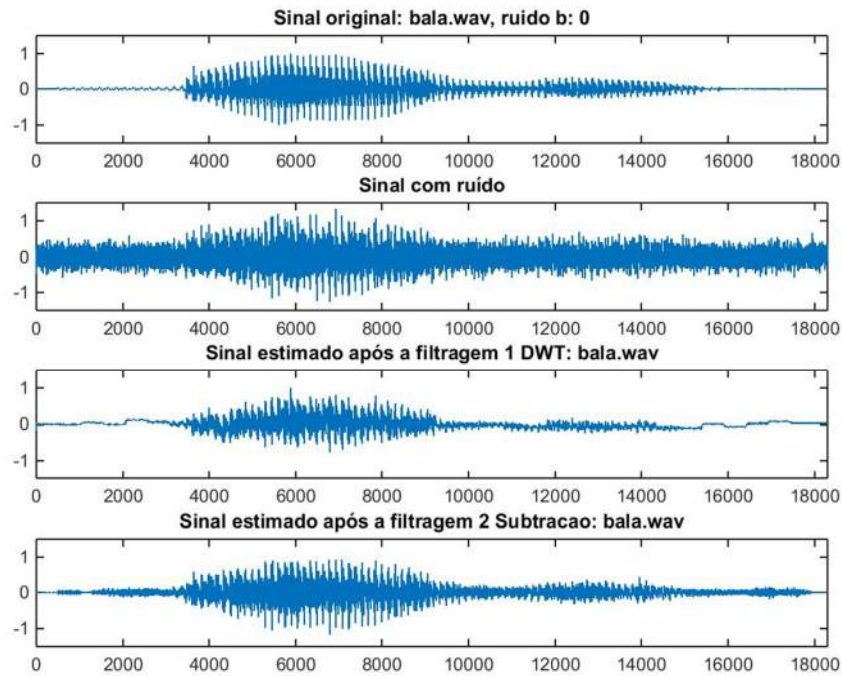


Figura 5.10 Comparativo entre DWT e subtração espectral em 0dB, ‘bala’ voz masculina

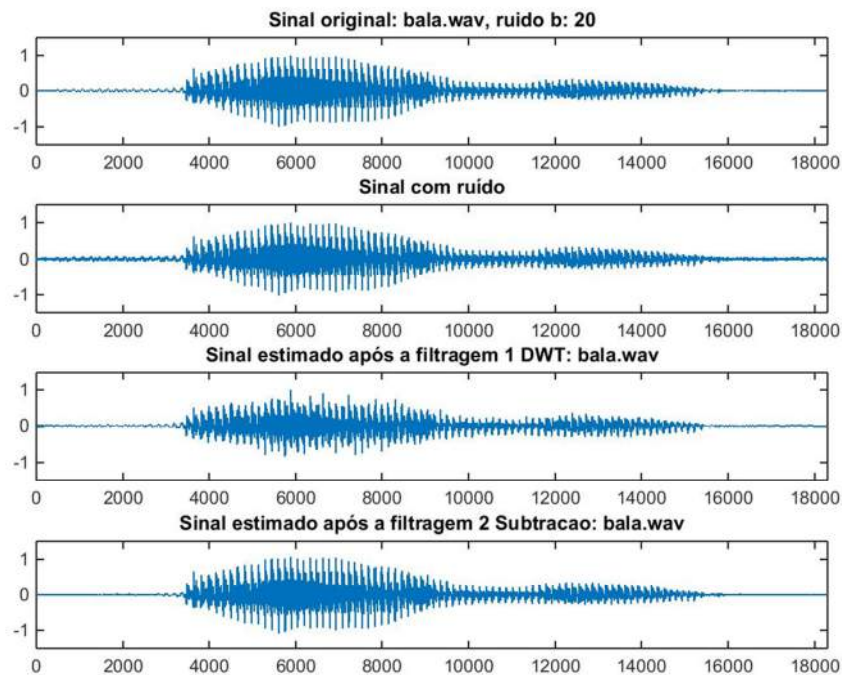


Figura 5.11 Comparativo entre DWT e subtração espectral em 20dB, ‘bala’ voz masculina

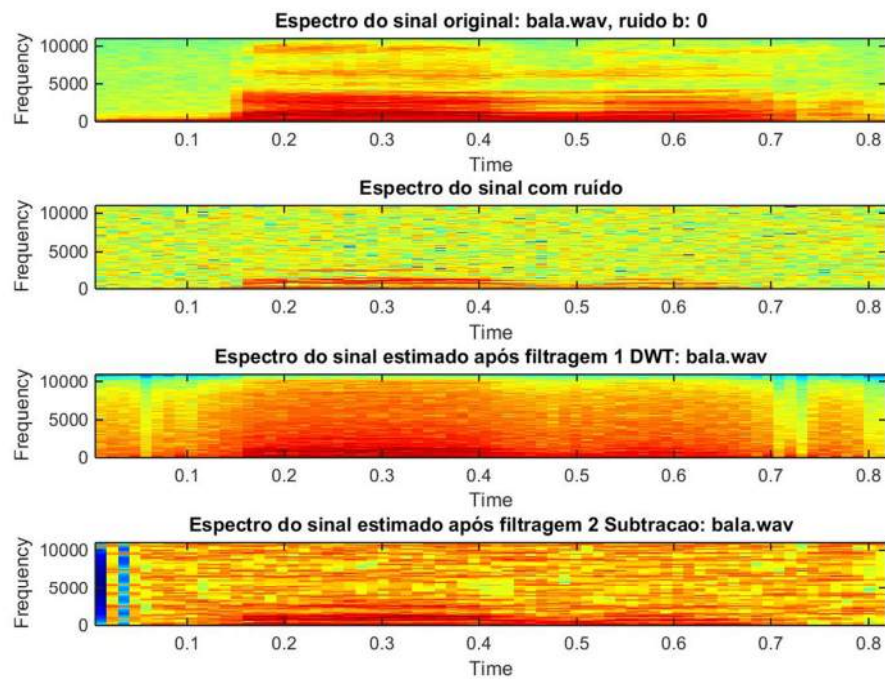


Figura 5.12 Espectrogramas DWT e subtração espectral em 0dB, 'bala' voz masculina

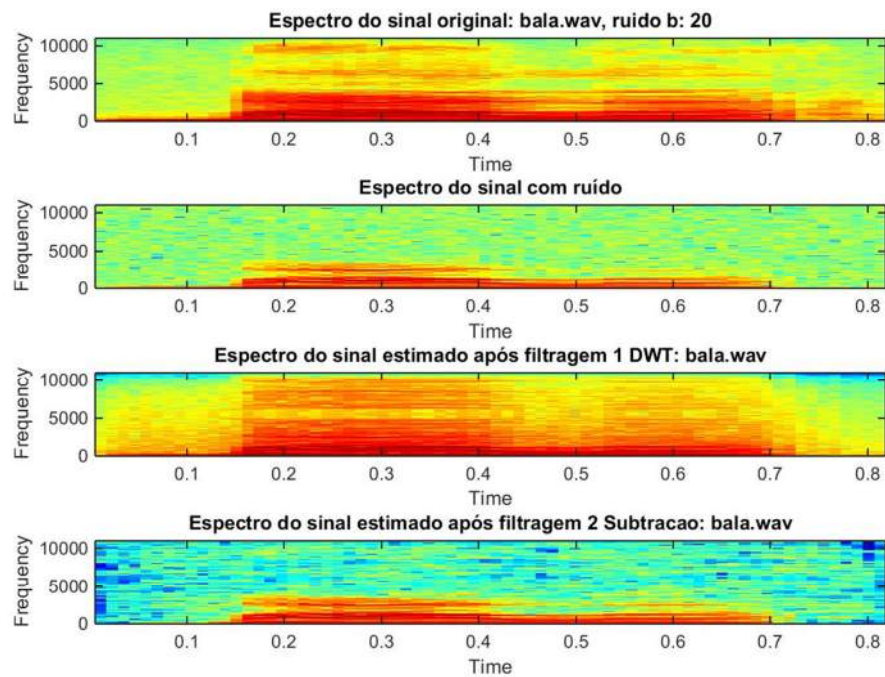


Figura 5.13 Espectrogramas DWT e subtração espectral em 20dB, 'bala' voz masculina

5.6.3 Comparação baseada em medidas

Por fim, realiza-se a comparação entre DWT e a subtração espectral por meio de medidas objetivas. Em SNR de entrada 0dB, a filtragem utilizando DWT é superior gerando uma IDS média de 0,57, enquanto que a subtração espectral gera 1,11 de IDS média. Ou seja, o sinal filtrado é mais dissemelhante ao sinal de entrada para a subtração. Quando a SNR de entrada é alta, 20dB, a situação se inverte e a subtração espectral produz uma distância de Itakura-Saito menor que a DWT.

Considerando a SNR de entrada 0dB, a subtração espectral produz um resultado pífió, enquanto que a DWT conseguiu levar a SNR segmentada média de -8,35 para 2,6, resultando em um ganho considerável. Em SNR de entrada 20dB, o desempenho de ambas se assemelha, mas com ligeira vantagem para a subtração espectral.

Para SNR alta, a subtração espectral obteve números melhores que a DWT. Ainda assim, nota-se a presença do ruído musical nos testes de audição, enquanto que a DWT produz sons muito semelhantes ao original.

Tabela 17 Comparativo WPD x subtração espectral, SNRi 0dB

arquivo	snriseg	Algoritmo 1		Algoritmo 3	
		snro_DWT	itakura_DWT	snro_Subtracao	itakura_Subtracao
bala	-10	2	0,73	0	1,26
bar	-9	2	0,75	0	1,48
casa	-11	2	0,28	0	0,60
elefante	-12	1	0,63	0	0,70
eletrica	-14	0	0,40	0	0,94
fala	-9	5	0,51	0	1,17
janela	-7	2	0,33	0	0,74
maio	-8	0	0,44	0	1,04
mao	-8	3	0,62	0	1,28
melão	-7	3	0,21	-2	1,03
bala_f	-8	1	0,87	0	1,45
bar_f	-12	6	1,02	0	1,44
casa_f	-7	-2	0,74	1	1,02
elefante_f	-9	6	0,49	-2	1,26
eletrica_f	-7	3	0,41	-2	1,13
fala_f	-8	7	0,78	0	1,33
janela_f	-5	2	0,36	-2	0,82
maio_f	-6	3	0,61	0	1,15
mao_f	-5	3	0,64	2	1,28
melao_f	-5	3	0,61	1	1,14
Media	-8,35	2,6	0,57	-0,2	1,11

Tabela 18 Comparativo WPD x subtração espectral, SNRi 20dB

arquivo	snriseg	Algoritmo 1		Algoritmo 3	
		snro_DWT	itakura_DWT	snro_Subtracao	itakura_Subtracao
bala	-2	2	0,53	1	0,34
bar	-1	2	0,56	2	0,49
casa	-2	0	0,27	4	0,05
elefante	-3	2	0,46	0	0,07
eletrica	-5	1	0,25	0	0,21
fala	-1	2	0,45	4	0,28
janela	0	0	0,32	1	0,16
maio	-1	0	0,34	2	0,19
mao	-1	0	0,36	3	0,20
melão	-1	1	0,20	-2	0,24
bala_f	-1	1	0,71	1	0,52
bar_f	-2	2	0,96	1	0,57
casa_f	0	0	0,78	2	0,31
elefante_f	-1	2	0,49	-2	0,52
eletrica_f	0	0	0,36	-2	0,35
fala_f	-1	1	0,74	3	0,58
janela_f	0	0	0,62	-2	0,30
maio_f	0	0	0,82	1	0,42
mao_f	0	0	0,52	3	0,45
melao_f	0	0	0,80	1	0,47
Média	-1,1	0,8	0,53	1,05	0,34

5.7 Considerações finais

Este capítulo apresentou diversos experimentos envolvendo as técnicas determinadas no Capítulo 4: DWT, WPD e DWTWPD. No primeiro momento estas técnicas foram comparadas aos pares e também todas juntas (Figura 5.1), em que se observaram os seguintes aspectos:

- A DWT
 - Gera melhor distância de Itakura-Saito tanto para baixas como para altas SNRs de entrada;
 - Não é a melhor alternativa quando o objetivo é melhorar a SNR; segmentada de saída, principalmente em altas SNR de entrada;
 - Consegue remover bastante ruído mas tende a espalhar a energia do sinal;
- A WPD
 - Possui desempenho fraco quanto a Itakura e testes de audição;
 - Gera boa SNR segmentada de saída;
 - Produz menos espalhamento da energia do sinal;

- A DWTWPD

- Possui desempenho semelhante a DWT para IDS e SNR segmentada de saída, afinal a WPD pouco contribui;
- No espectrograma conseguiu agregar características de ambas, removeu ruído e espalhou menos a energia do sinal;
- Não foi tão bem quando a DWT em testes de audição.

Por isso, o método escolhido foi a combinação entre filtro de Kalman e decomposição DWT com limiarização no coeficiente de detalhe D1, valendo destacar que DWTWPD possui um grande potencial uma vez que une o melhor de cada uma das duas decomposições. Contudo, um estudo mais aprofundado precisa ser feito para melhorar a qualidade audível do sinal resultante, que não consegue ser melhor que o sinal gerado pela DWT.

No segundo momento, a técnica utilizando DWT foi comparada à subtração espectral a fim de validar a qualidade do método. A Figura 5.14 sintetiza a comparação por meio das medidas IDS e SNR segmentada, das quais se constata que para SNRs de entrada até 15dB a DWT é melhor que a subtração espectral para a remoção de ruído branco gaussiano. A subtração espectral remove o ruído branco, mas insere ruído musical, enquanto que a DWT remove o ruído, mas tende a espalhar a energia do sinal. Considerando o ruído colorido, (ver Figura 5.15) a DWT não teve um bom desempenho, já que a combinação com o filtro de Kalman faz com que haja certa dificuldade em estimar o ruído.

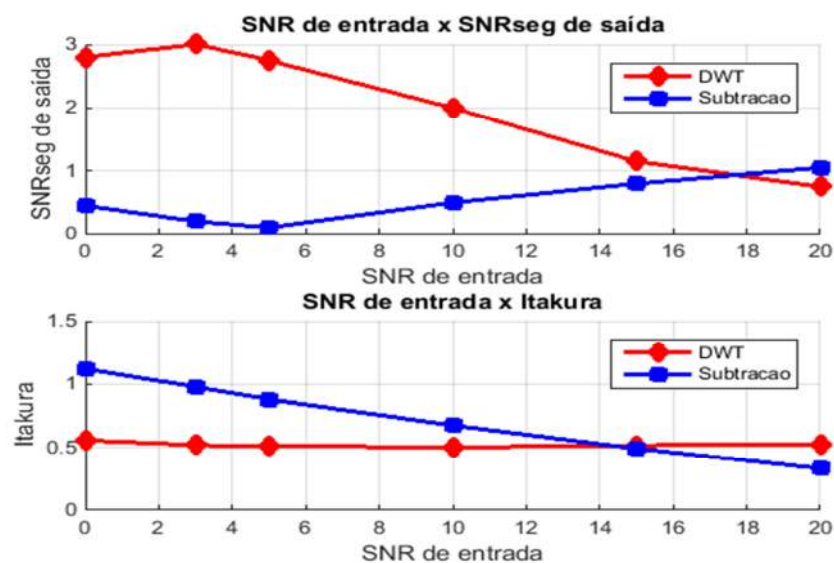


Figura 5.14 Comparativo entre DWT e subtração espectral para SNR segmentada e Itakura-Saito, ruído branco

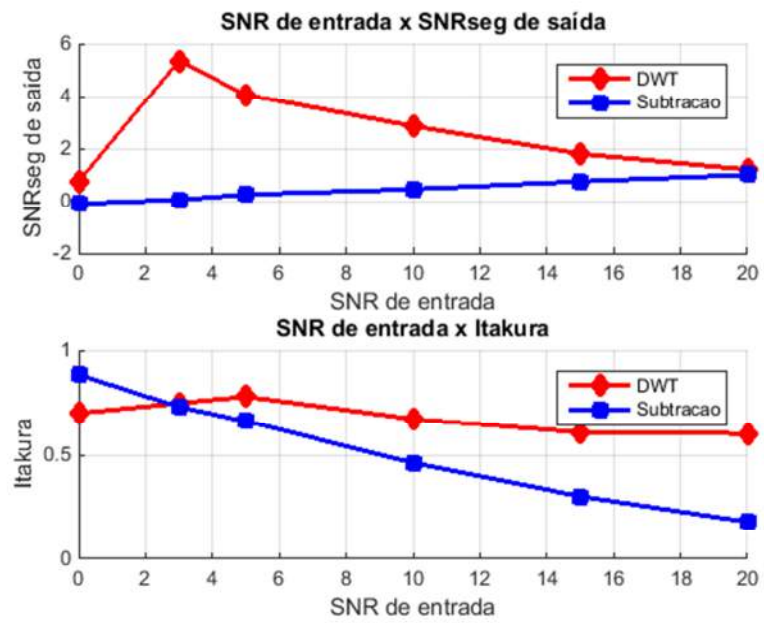


Figura 5.15 Comparativo entre DWT e subtração espectral para SNR segmentada e Itakura-Saito,

6 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho foi proposto um método para redução de ruído em sinais de voz utilizando o filtro de Kalman e a transformada wavelet. Após a aplicação do filtro de Kalman, o sinal é decomposto e os coeficientes da WT são limiarizados. Como a WT pode decompor o sinal em diversos coeficientes e de duas formas diferentes (DWT e WPD), o primeiro passo para garantir a qualidade do método foi o refinamento por meio de algoritmo genético, ao que se chegou a três algoritmos semelhantes.

Para o *fitness*, o AG utilizou uma combinação de duas medidas bem aceitas na quantificação da qualidade do sinal de voz: a distância de Itakura-Saito e a SNR segmentada. O cromossomo usado determinava a combinação de coeficientes (DWT e WPD) a serem filtrados, além de tentarem também o ajuste do limiar e a possibilidade de aplicar o filtro de Kalman diretamente nos coeficientes da WT. Para resultados mais efetivos, 49 sinais foram usados que resultou em algumas recomendações:

- coeficiente D1 da DWT;
- coeficientes 9,10,11,14,15,16 da WPD;
- 100 % do limiar *hard* do Quadro 2 (este valor era mutado durante as gerações do AG);
- Limiarização e não filtragem Kalman nos coeficientes da WPD.

A partir destes refinamentos, os 3 algoritmos foram concebidos e testados, e verificou-se a superioridade do algoritmo utilizando DWT, o que também gerou bons resultados para o DWTWPD. No entanto, em audições do sinal filtrado, observou-se ligeira vantagem para o método DWT puro. Ainda que medidas e espectrogramas mostrassem boas características no DWTWPD, ele espalha menos a energia do sinal e mantém bons valores para SNR segmentada de saída. Assim, é boa ideia para um trabalho futuro explorar a fundo este método. Há que se esclarecer sobre os testes de audição, foi feito um teste de reconhecimento de palavras com 10 voluntários e aproveitando a ocasião, de maneira bastante informal eles foram inqueridos sobre qual dos sinais estava melhor entre a DWT e DWTWDP, apurando o resultado verificou-se DWT como sendo melhor.

Para certificar a qualidade do método foi feita uma comparação com subtração espectral em que se verificou que em baixas SNR de entrada, a qualidade do método proposto

é bastante superior, uma vez que produz medidas melhores, remove as frequências espúrias e garante a inteligibilidade do sinal. Já em altas SNRs de entrada DWT não tem medidas tão boas quanto a subtração espectral, mas garante melhor inteligibilidade, uma vez que a subtração espectral introduz ruído musical no sinal.

6.1 Trabalhos futuros

Diversas possibilidades puderam ser vislumbradas durante a realização deste trabalho e podem ser executadas futuramente. Dentre as principais, tem-se as seguintes:

- Desenvolver um limiar adaptativo para a DWTWPD – este algoritmo tem boas características, mas perde em inteligibilidade, dessa forma espera-se obter resultados mais satisfatórios;
- Testar os sinais obtidos em um sistema de reconhecimento automático – como os testes com pessoas são muito subjetivos, espera-se maior confiança nas respostas dadas aos testes;
- Utilizar um auto-encoder – pessoas são melhores em reconhecer sons ruidosos do que sistemas automatizados, pois conseguem fazer inferências a partir de conhecimentos prévios, se um sistema de redução de ruído for dotado desta característica, esperam-se resultados ainda melhores.

REFERÊNCIAS BIBLIOGRÁFICAS

Abreu, Caio Cesar Enside de. 2013. Uso de equações de diferenças na obtenção de filtros para redução de ruído em sinais de voz no domínio wavelet . Ilha Solteira : UNESP, 2013.

Boll, Steven F. 1979. *Suppression of Acoustic Noise in Speech Using Spectral Subtraction*. s.l. : IEEE Transactions on Acoustics, Speech, And Signal Processi, 1979. pp. 113 - 119. Vols. ASSP-27.

Brown, Robert Grover e Hwang, Patrick Y. C. 1997. *Introduction to Random Signals and Applied Kalman Filtering*. s.l. : John Wiley & Sons, 1997.

Chu, Peter L. e Messerschitt, David G. 1982. A frequency Weighted Itakura - Saito Spectral Distance Measure. *IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING*. AUGUST de 1982, Vols. ASSP-30, pp. 545 - 560.

Chui, Charles K. 1992. *Wavelet Analysis And Its Applications*. Texas : ACADEMIC PRESS, 1992.

Chun-Li, Liu. 2010. A tutorial of the Wavelet Transform. february de 2010.

Coifman, Ronald R. e Wickerhauser, Mladen Victor. 1992. Entropy-Based Algorithms for Best Basis Selection. *IEEE Transactions On Information Theory*. 1992, Vols. 38-(2), pp. 713-718.

Cox, Pedro Henrique. 2004. Análise e Síntese de um Processador Digital Wavelet. Ilha Solteira I : Universidade Estadual Paulista Faculdade de Engenharia de Ilha Solteira, 2004.

Daubechies, Ingrid. 1990. The Wavelet Transform, Time-Frequency Localization and Signal Analysis. *IEEE TRANSACTIONS ON INFORMATION THEORY*. Swptember de 1990, pp. 961 - 1005.

Deepa, D., et al. 2011. Enhancement of Noisy Speech Signal Based on Variance and Modified Gain Function. *IEEE COMPUTER SOCIETY*. 2011, pp. 155 - 159.

Diniz, Henrique Carlos. 2013. Efeito da Fase em Técnicas de Redução de Ruídos em Sinais de Voz Baseado na SNR a Priori. Ilha Solteira : Unesp Ilha Solteira, 2013.

Donoho, David L. e Johnstone, Iain M. 1994. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*. 1994, pp. 425 - 455.

Duarte, Marco Aparecido Queiroz. 2005. *Redução de Ruído em Sinais de Voz no Domínio Wavelet (Dissertação)*. Ilha Solteira : Unesp, 2005.

Ephraim, Yariv e Malah, David. 1984. Speech Enhancement Using a Minimum Mean-Square Error Short-time Spectral Amplitude Estimator. *IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, VOL. ASSP-32*. DECEMBER de 1984, pp. 1109 -1121.

Faria, Regis Rossi Alves. 1997. Aplicação de Wavelets na Análise de Gestos Musicais em Timbres de Instrumentos Acusticos Tradicionais. São Paulo : s.n., 1997.

Fujimoto, M. e Ariki, Y. 2000. NOISY SPEECH RECOGNITION USING NOISE REDUCTION METHOD BASED ON KALMAN FILTER. *Acoustics, Speech, and Signal Processing*. 2000.

Fukunaga, Keinosuke. 1990. *Introduction to Statistical Patter Recognition*. Tokyo : Werner Rheinboldt, 1990.

Gabrea, Marcel. 2004. Robust Adaptive Kalman Filtering-Based Speech Enhancement Algorithm. *Acoustics, Speech and Signal Processing (ICASSP), 2004 IEEE International Conference on*. 2004, pp. 301 - 304.

Garcia, Alberto Leon. 2008. *Probability Statistics and Randon Processes for Electrical Engineering* . s.l. : William Opaluch, 2008.

H.Juang, B. e R.Rabiner, L. 2005. Automatic Speech Recognition – A Brief History of the Technology Development. *Atlanta Rutgers University and the University of California - Georgia Institute of Technology*. 2005.

Hassani, Malihe e Mollaei, M. R. Karami. 2011. Speech Enhancement Based on Spectral Subtraction. *IEEE 7th International Colloquium on Signal Processing and its Applications*. 2011, pp. 366 - 370.

Haykin, Simon. 2001. *Kalman Filtering and Neural Networks*. s.l. : John Wiley & Sons, 2001.

Hu, Yi e Loizou, Philipos C. 2004. Speech Enhancement Based on Wavelet Thresholding the Multitaper Spectrum. *IEEE TRANSACTION ON SPEECH AND AUDIO PROCESSING*. 26 de JANUARY de 2004, pp. 59 - 67.

Hwang, Juang Biing e Rabiner, Lawrence R. 2005. Automatic Speech Recognition - A Brief History of the Technology Development. *Encyclopedia of Language and Linguistics*. 2005, pp. 1 - 24.

Itakura, Fumitada. 1975. Minimum Prediction Residual Principle Applied to Speech Recognition. *IEEE TRANSACTION ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING*. ASSP - 23, FEBRUARY de 1975, Vol. 1, pp. 67 - 72.

Junior, Hudson Bitencourt. 2003. Métodos de Estimaco Recursiva Baseados no Filtro de Kalman Aplicados a Sistemas No Lineares . s.l. : Universidade Federal de Minas Gerais , 2003.

Kalman, Rodolf E. 1960. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME - Journal of Basic Engineering*. 24 de february de 1960, pp. 35 - 45.

Kalman, Rudolf Emily. 1960. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME - Journal of Basic Engeneering*. 24 de February de 1960, pp. 35 - 45.

Lang, Serge. 2003. *Álgebra Linear* . s.l. : Editora Cincia Moderna, 2003.

Leithold, Louis. 1998. *EL CÁLCULO*. Stima Edio. s.l. : Oxford University Press, 1998.

Lima, Elon Lages. 1976. *Curso de Anlise*. s.l. : SBM - Sociedade Brasileira de Matemtica, 1976.

Lima, Paulo Cupertino de. 2003. Wavelets: uma introduo. 29 de julho de 2003.

Linden, Ricardo. 2006. *Algoritmos Genticos: Uma importante ferramenta da Inteligncia*. Rio de Janeiro : Brasport, 2006.

Liu, Jia e Krzymienf, Witold A. 2006. A Minimum Mean-Square Error Criterion Based Nonlinear Joint Transmitter-Receiver Processing algorithm for the Downlink of Multi-User MIMO Systems. *Vehicular Technology Conference*. 7 - 10 de maio de 2006, pp. 1982 - 1987.

Mallat, Stephane. 1989. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. 30 de July de 1989. Vol. II, pp. 674 - 693.

Messaoud, Mohamed anouar Ben, Bouzid, Aïcha e Ellouze, Noureddine. 2015. Speech Enhancement Based on Wavelet Packet of an Improved Principal Component Analysis. *Computer Speech and Language*. 2015, Vol. 35, pp. 58-72.

Misiti, Michel, et al. 2015. User's Guide, Wavelet Toolbox MatLab R2015a. 2015.

Nicodem, Monique Vitório. 2006. Detecção e Tratamento de Cliques Naturais em Bancos de Fala Visando Síntese Concatenativa de Alta Qualidade (Dissertação). Florianópolis : Universidade Federal de Santa Catarina , 2006.

Oppenheim, Alan V. and Schafer, Ronald W. 2012. *Processamento em tempo discreto de sinais*. [trans.] Daniel Vieira. 3 . São Paulo : Pearson Education do Brasil, 2012.

Pinheiro, João Ismael D., et al. 2012. *Probabilidade e Estatística: Quantificando a incerteza*. Rio de Janeiro : Elsevier, 2012.

Rabiner, Lawrence R e Schafer, Ronald W. 2011. *Theory and Applications of Digital Speech Processing*. 5. Upper Saddle River : Pearson, 2011.

Seok, Jong Won e Bae, Keun Sung. 1997 . Speech Enhancement With Reduction of Noise Componentes in the Wavelet Domain . *Computer org.* 1997 , pp. 1323 - 1326.

Silva, Leandro Aureliano da. 2007. Filtros de Kalman no tempo e frequência discretos combinados com subtração Espectral (Dissertação de Mestrado). São Carlos : Universidade de São Paulo - USP, 2007.

Silva, Luiz Felipe da. 2011. Redução de ruído em sinais de voz utilizando uma função de ganhos adaptativa para o filtro de Wiener. (Dissertação). Florianópolis : Universidade Federal de Santa Catarina, 2011.

Soares, Wendel Cleber. 2009. *Um Método Não-Limiar para Redução de Ruído em Sinais de Voz no Domínio Wavelet*. Ilha Solteira : Unesp, 2009.

Sunny, Sonia, S, David Peter e Jacob, K Poulouse. 2013. A New Algorithm for Adaptive Smoothing of Signals in Speech Enhancement. *International Conference Eletronic Enginneering and Computer Sciense*. 2013, p. 337 343.

Vaseghi, Saed V. 2000. NOISE AND DISTORTION. *Advanced Digital Signal Processing and Noise Reduction, Second Edition*. s.l. : Copyright © 2000 John Wiley & Sons Ltd, 2000, pp. 29 - 43.

Vaseghi, Saeed V. 2008. *Advanced digital signal processing and noise reduction*. India : John Wiley & Sons Ltd., 2008.

Xia, Bingyin e Bao, Changchun. 2014. Wiener filtering based speech enhancement with Weighted Denoising Auto-encoder and noise classification. *Speech Communication*. 20 de February de 2014, pp. 13 - 29.