Universidade Estadual Paulista - Unesp
Faculdade de Engenharia -Campus De Ilha Solteira
Programa de Pós-Graduação Em Engenharia Elétrica

"REDUÇÃO DE RUÍDO EM SINAIS DE VOZ NO DOMÍNIO WAVELET"

Tese submetida à Faculdade de Engenharia de Ilha Solteira – UNESP – como parte dos requisitos necessários para obtenção do título de **Doutor em Engenharia Elétrica**.

Autor : Marco Aparecido Queiroz Duarte
Orientador : Francisco Villarreal Alvarado
Co-orientador : Jozué Vieira Filho

Ilha Solteira - SP
Fevereiro de 2005

À minha família.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SÍMBOLOS E ABREVIAÇÕES

R - Conjunto dos números reais

Z – Conjunto dos números inteiros

$L^2(\boldsymbol{R})$ – Espaço das funções mensuráveis a Lebesgue

$\psi(t)$ – Função Wavelet

$\phi(t)$ – função de escala

a – Parâmetro de escala da função wavelet

b – Parâmetro de translação da função wavelet

$\|\bullet\|$ - Norma de $\bullet$

$\hat{\bullet}$ - Transformada de Fourier de $\bullet$

$\bullet^*$ - Conjugado complexo de $\bullet$

$W_f(a,b)$ – Transformada Wavelet Contínua de f

$\langle *, \bullet \rangle$ - produto interno

V_j – Espaço de escala 2^j

W_j – Complemento de V_j

$\cup$ - Reunião de conjuntos

$\cap$ - Intersecção de conjuntos

$\oplus$ - soma direta de conjuntos

$\mathrm{Pr}oj_{V_j}(f)$ - Projeção ortogonal de f em V_j.

$\mathrm{Pr}oj_{W_j}(f)$ - Projeção ortogonal de f em W_j.

h_n – Coeficientes dos filtros gerados por $\phi(t)$

d_n – Coeficientes dos filtros gerados por $\psi(t)$

$(\downarrow 2)$ – Operador de dizimação de ordem 2

($\uparrow 2$) – Operador de inserção de zeros de ordem 2

AMR- Análise de Multirresolução

FWT – Transformada Wavelet Rápida

DWT – Transformada Wavelet Discreta

IDWT – Transformada Wavelet Discreta Inversa

$c_i[n]$ – Coeficientes de aproximação do nível i da DWT

$d_i[n]$ – Coeficientes de detalhes do nível i da DWT

$L_D[n]$ – Filtro passa baixa da DWT

$H_D[n]$ – Filtro passa alta da DWT

$L_R[n]$ – Filtro passa baixa da IDWT

$H_R[n]$ – Filtro passa alta da IDWT

λ - Limar usado para filtragem dos sinais

TRH – Treshold ou limiar

γ - parâmetro de inclinação da Sigmóide

β - fator de correção da potência do ruído

LS – Limiar Sigmóde

LSD – Limiar Sigmóide Direto

LSP – Limiar Sigmóide Ponderado

PLSD – Limiar Sigmóide Direto com Pré-filtragem

PLSP – Limiar Sigmóide Ponderado com Pré-filtragem

SMW – Método proposto por SHEIKHZADEH e ABUTALEBI

PESQ – Perceptual Evaluation of Speech Quality

RESUMO

Neste trabalho é feito um estudo sobre os métodos de redução de ruído aditivo em sinais de voz baseados em wavelets e, através deste estudo, propõe-se um novo método de redução de ruído em sinais de voz no domínio wavelet. O princípio básico da maioria dos métodos de redução de ruído baseados em wavelets é a determinação e aplicação de um limiar, que permite bons resultados para sinais contaminados por ruído branco, mas não são eficientes no processamento de sinais contaminados por ruído colorido, que é o tipo de ruído mais comum em situações reais. Nesses métodos, o limiar, geralmente, é calculado nos intervalos de silêncio e aplicado em todo o sinal. Os coeficientes no domínio wavelet são comparados com este limiar e aqueles que estão abaixo deste valor são eliminados, fazendo assim uma aplicação linear deste limiar. Esta eliminação acaba causando descontinuidades no tempo e na freqüência no sinal processado. Além disso, a forma com que o limiar é calculado pode degradar os trechos de voz do sinal processado, principalmente nos casos em que o limiar depende fortemente da última janela do último trecho de silêncio. O método proposto neste trabalho também é baseado em corte por limiar, mas em vez de uma aplicação linear do limiar, ele faz uma aplicação não-linear, o que evita as descontinuidades causadas por outros algoritmos. O limiar é calculado nos trechos de silêncio e não depende apenas da última janela do último trecho de silêncio, mas sim de todas as janelas, já que este limiar é uma média de todos os limiares calculados neste trecho. Isto faz com que a redução do ruído seja mais uniforme e introduza menos distorções no sinal processado. Além disso, nos trechos de voz ainda é calculado um novo limiar que também será usado, em conjunto com o limiar calculado no silêncio. Isto faz com que a energia da janela que está sendo processada seja também considerada. Desta forma ele acaba sendo eficaz para sinais que apresentam características não-estacionárias e

ainda elimina a decisão *voiced/unvoiced*, técnica muito usada para minimizar distorções, que é dependente da energia do sinal. Na aplicação do limiar, em vez de anular coeficientes que estão abaixo do limiar, estes coeficientes são apenas atenuados, de forma que o que se apresenta é uma nova versão de limiar suave no domínio wavelet. Além do método proposto, que é eficiente tanto para sinais contaminados por ruído branco quanto para sinais contaminados por ruído colorido, apresenta-se um novo algoritmo de detecção silêncio/voz baseado em wavelets.

Palavras-chave – Wavelets, Multirresolução, Limiar, Redução de Ruído, Transformada Wavelet Discreta, Medidas de Qualidade.

ABSTRACT

In this work a study of additive noise reduction in speech based on wavelets is presented and, based on this study a new noise reduction method in speech in the wavelet domain is proposed. The basic idea of most methods of noise reduction based on wavelets is the determination and application of a threshold, that produces good results for signals contaminated by white noise, but they are not very efficient in processing signals contaminated by colored noise, which is more common in real situations. In those methods, the threshold, generally, is calculated in the silence intervals and applied to the whole signal. The coefficients in the wavelet domain are compared with this threshold and those that are below this value are eliminated, making a linear application of this threshold. This elimination causes discontinuities in time and frequency of the processed signal. Besides, the way that the threshold is computed can degrade the voice segments of the processed signal, principally when the threshold depends strongly on the last window of the last silence segment. The proposed method in this work is also based in thresholding, but, instead of a linear application of the threshold, it makes a non-linear application, which avoids the discontinuities caused by other algorithms. The threshold is calculated in the silence segments and is not dependent only on the last window of the last silence segment, but of all the windows, since this threshold is an average of all thresholds calculated in this segment. It makes noise reduction more uniform and introduces less distortion in the processed signal. Besides, in the voice segments a new threshold is calculated that will be also used with the threshold calculated in the silence. It makes that the energy of the window that is being processed is also considered. This way, it is effective for signals with non-stationary characteristics and still eliminates the voiced/unvoiced decision, a very common technique used to minimize distortions,

dependent on the signal energy. In the threshold application, instead of canceling the coefficients that are below the threshold, these coefficients are just attenuated, so it presents a new soft threshold version in the wavelet domain. Besides the proposed method, which is efficient for signals contaminated by white or colored noise, it is presented a new voice activity detection based on wavelets.

Keywords – Wavelets, Multiresolution, Threshold, Noise Reduction, Discrete Wavelet Transform, Performance Measures.

CAPÍTULO 1

INTRODUÇÃO

A redução de ruído é importante nas mais variadas aplicações que envolvem processamento de sinais: como em sistemas de reconhecimento automático de fala, telefonia, sistemas de codificação, etc. Os métodos mais explorados nas últimas três décadas são baseados na Transformada de Fourier (AGUIAR, 1989; BOLL, 1979; EPHRAIM e MALAH, 1984; EPHRAIM, 1992; HÄNDEL, 1995; VIEIRA FILHO et al., 1995) e muitos apresentam bons resultados. Entretanto, ainda existem aplicações para as quais as técnicas atuais não conseguem apresentar resultados satisfatórios. Um exemplo típico é a aplicação em sistemas de codificação usados em comunicações sem fio, onde o nível de ruído pode degradar de maneira significativa o sinal transmitido.

Os estudos com base na Transformada de Fourier ainda são importantes, mas existem muitos estudos que tentam explorar a Transformada Wavelet Discreta (DWT) (DONOHO, 1995; SEOK e BAE, 1997; SHEIKHZADEH e ABUTALEBI, 2001).

Se do ponto de vista de codificação de sinais de voz com boa relação sinal/ruído (SNR) a Transformada Wavelet é de fácil aplicação e gera resultados considerados bons, o mesmo não pode ser dito com relação às aplicações de redução de ruído quando a SNR é baixa. As metodologias mais relevantes encontradas na literatura usam como base o cálculo de um limiar a partir do sinal ruidoso, com posterior eliminação de coeficientes wavelet, que é um tipo de codificação básico usando wavelets (DONOHO, 1995; KASTANTIN et al., 1994; RIOUL e VETERLLI, 1991). O problema é que o uso de um limiar resulta em ruídos residuais incômodos ao ouvido humano, apesar de muitos autores tentarem justificar seus métodos usando o nível de redução de ruído obtido. O ruído

residual aparece quando o limiar é aplicado aos coeficientes, gerando novos coeficientes que podem ser nulos ou não. Do ponto de vista auditivo, o ruído residual obtido em sistemas baseados na DWT, é parecido com o "ruído musical", que é um ruído residual característico dos métodos baseados na Transformada de Fourier (EPHRAIM e MALAH, 1984; VIEIRA FILHO, 1996), principalmente aqueles que usam o princípio da subtração espectral (BOLL, 1979).

Por outro lado, os métodos baseados em wavelets, na sua maioria, são eficientes na redução de ruído branco, deixando a desejar quando se trata de ruído colorido (BAHOURA e ROUAT, 2001(B); DUARTE et al., 2004(A); SEOK e BAE, 1997; SHEIKHZADEH e ABUTALEBI, 2001).

Neste trabalho são apresentados novos métodos baseados em wavelets para redução de ruído em sinais de voz, com o objetivo de reduzir o ruído residual e torná-los eficientes tanto para sinais contaminados por ruído branco quanto para sinais contaminados por ruído colorido.

O texto está organizado como segue:

- No capítulo 2 são apresentados aspectos da redução de ruído em sinais de voz, bem como as medidas para avaliação do desempenho dos métodos de melhoria em sinais de voz;
- No capítulo 3 é apresentado um estudo detalhado da teoria de wavelets;
- No capítulo 4, métodos de redução de ruído em sinais de voz baseados em wavelets são apresentados juntamente com suas vantagens e desvantagens;
- No capítulo 5 é apresentado e discutido um novo método para redução de ruído em sinais de voz usando wavelets.
- No capítulo 6 são apresentados os resultados da aplicação do método proposto para redução de ruído em sinais de voz e as avaliações deste método;
- Finalmente, no capítulo 7 são apresentadas as conclusões e as considerações finais deste trabalho.

CAPÍTULO 2

ASPECTOS GERAIS DA REDUÇÃO DE RUÍDO EM SINAIS DE VOZ

Introdução

Em muitos sistemas de comunicação, o ruído de fundo pode causar perda da qualidade ou inteligibilidade do sinal de voz. Quando a comunicação é feita num ambiente silencioso, a informação trocada é facilmente detectada e precisa. Entretanto, um ambiente ruidoso reduz a habilidade do ouvinte em entender o que é dito. Além disso, na comunicação interpessoal, o sinal de voz pode ser transmitido por telefone, canais de rádio e reproduzido por autofalantes ou fones de ouvido. A qualidade do sinal pode, assim, ser também influenciada na conversão dos dados (microfone), transmissão (ruído através dos canais) ou reprodução (autofalantes ou fones de ouvido). O objetivo principal de um algoritmo de redução de ruído em sinais de voz é melhorar a qualidade do sinal, seja reduzindo ruído de fundo, seja suprimindo o canal ou interlocutor que esteja gerando interferência.

O problema da redução de ruído em sinais de voz, ou melhoria de sinal contaminado por ruído aditivo, tem sido muito discutido nas últimas décadas. Muitos métodos têm sido propostos, cada um voltado para características ou restrições específicas, todos com diferentes graus de sucesso. O ruído considerado nesses métodos pode ser o ruído Gaussiano branco, ruído produzido por um avião, por um carro ou por máquinas no ambiente de uma fábrica. Dependendo da aplicação, um sistema pode ser direcionado a um ou mais objetivos, como melhoria na qualidade geral, aumento da inteligibilidade ou redução da fadiga do ouvinte, como exemplo de fadiga pode-se citar um ruído de fundo que não interfira na voz, mas que se torna incômodo por estar sempre presente. O objetivo

de alcançar melhor qualidade e/ou inteligibilidade de um sinal de voz ruidoso pode também contribuir para a melhoria de rendimento em outras aplicações, tais como compressão ou reconhecimento de fala e identificação de locutor.

Como em qualquer problema de engenharia, é necessário ter um entendimento claro de objetivos e habilidades para medir o desempenho do sistema e verificar se estes foram alcançados. Quando se considera a redução de ruído, normalmente se pensa em melhorar a relação sinal ruído (SNR). É importante perceber, entretanto, que este pode não ser o critério mais apropriado de rendimento para a melhoria em sinais de voz. Todo ouvinte tem um entendimento intuitivo de qualidade de voz, inteligibilidade e fadiga auditiva. Entretanto, estas áreas são difíceis de se quantificar na maioria das aplicações de melhoria de sinais de voz, pois elas são baseadas em avaliações subjetivas do sinal processado.

Neste capítulo, apresenta-se a classificação de métodos de melhoria de sinais de voz, com ênfase na redução de ruído, e as principais medidas usadas para avaliação de desempenho dos algoritmos de redução de ruído.

2.1 Classificação de Métodos para Melhoria de Sinais de Voz

Existem várias formas nas quais métodos de melhoria de sinais de voz podem ser classificados (DELLER et al., 1993). Um grupo abrangente de técnicas está concentrado na forma em que o sinal é modelado. Algumas técnicas são baseadas em modelos de processos estocásticos do sinal. Estas técnicas, geralmente, trabalham sobre um dado critério matemático. Existem sistemas que são baseados em critérios perceptivos e procuram melhorar aspectos importantes da percepção humana. Por exemplo, uma determinada técnica pode se concentrar apenas na melhoria da qualidade de consoantes, pois as consoantes são conhecidas por seu papel importante na inteligibilidade numa maneira desproporcional à energia de todo o sinal.

Os algoritmos de melhoria podem ser classificados, também, de acordo com o número de canais usados para captar/transmitir o sinal de voz. Para aplicações de um único canal, apenas um microfone é usado para captar voz e ruído. Neste caso, a estimação do perfil de ruído deve ser feita durante os períodos de silêncio entre as expressões vocais, assumindo-se, assim, um caráter estacionário do ruído de fundo. Sinais de voz em telefonia ou de rádio locução são captados por um único canal. Nos algoritmos de dois ou

mais canais (multicanais), as ondas sonoras chegam em cada um dos sensores em intervalos de tempo levemente diferentes, sendo que, normalmente, um é uma versão atrasada do outro. Técnicas para melhoria de sinais de voz em sistemas de multicanais são baseadas em dois cenários básicos: no primeiro, um canal primário contém o sinal de voz com ruído aditivo e um segundo canal contém uma amostra do sinal ruidoso (ruído de referência) relacionado ao ruído no canal primário. Normalmente, existe uma barreira acústica entre os sensores para garantir que o sinal de voz não se misture com o ruído de referência. No segundo, não existe barreira acústica e, então, o algoritmo empregado deve levar em consideração o cruzamento de informações entre os vários canais ou aplicar um mecanismo multisensor para detecção das formas de onda.

Além das classificações baseadas em detalhes específicos dos métodos, existem ainda quatro classes de métodos de melhoria em sinais de voz que diferem substancialmente na forma em que são aplicados. Cada uma dessas classes tem suas próprias suposições, vantagens e limitações. A primeira classe se concentra no domínio espectral de curto-prazo. Esta técnica reduz o ruído existente, subtraindo o ruído (no espectro de potência, transformada de Fourier ou domínio de autocorrelação) encontrado durante atividade sem fala em casos de um único microfone ou de um microfone de referência em sistemas com mais de um canal. A segunda classe é baseada no uso de métodos iterativos de modelagem do sinal de voz. Estes sistemas se concentram na estimação de parâmetros dos modelos que caracterizam o sinal de voz, seguidos por ressintetização do sinal sem ruído, baseado num filtro não-causal de Wiener. Estas técnicas estimam os parâmetros do sinal de voz em ruído baseado em modelos auto-regressivos, auto-regressivos com restrições ou média móvel autoregressivos (DELLER et al., 1993). Esta classe de técnicas de melhoria requer um conhecimento a priori do ruído e das estatísticas do sinal de voz e, geralmente, resulta em esquemas iterativos de melhoria. A terceira classe de sistemas é baseada no "cancelamento do ruído adaptativo" (ANC). O tradicional ANC é formulado usando um ambiente com dois canais, sendo um para captar o sinal de voz ruidoso e o outro para captar apenas o ruído. Nesse caso é usado um filtro adaptativo, atualizado com base na minimização do erro quadrático médio (Algoritmo *least mean squares*, LMS). A última classe é baseada na peridiocidade do sinal de voz. Estes métodos empregam a faixa da freqüência fundamental usando ou um ANC de único canal (aplicação especial) ou um filtro tipo pente adaptativo da magnitude do espectro harmônico (DELLER et al., 1993).

Os métodos abordados neste trabalho considerarão que: (1) apenas um microfone será usado para a captação do sinal de voz, (2) o ruído é aditivo; e (3) o ruído e o sinal são descorrelacionados.

Os ruídos considerados para as aplicações serão o ruído branco, ruído que possui sua potência uniformemente distribuída no espectro de freqüência, e o ruído colorido, ruído cuja densidade espectral de potência é proporcional ao inverso da freqüência (CONNOR, 1973). O ruído colorido é mais comum em situações reais. Além disso, os ruídos apresentados, branco ou colorido, serão estacionários, isto é, ruídos cujas propriedades estatísticas se mantêm constantes ao longo do tempo ou, pelo menos, por um período longo de tempo. Os ruídos não-estacionários são ruídos cujas propriedades estatísticas variam ao longo do tempo.

2.2 Medidas de Qualidade

Para estabelecer meios de comparação e avaliação de algoritmos de melhoria em sinais de voz, várias técnicas têm sido desenvolvidas ao longo da história do processamento de sinais. Geralmente, estes testes caem em duas classes: medidas subjetivas de qualidade e medidas objetivas de qualidade. As medidas subjetivas são baseadas em comparações entre o sinal original e o sinal processado por um ou mais ouvintes, que subjetivamente classificam a qualidade do sinal conforme uma escala predeterminada. As medidas objetivas de qualidade são baseadas em comparações matemáticas entre o sinal original e o sinal processado. Muitas medidas objetivas de qualidade quantificam a qualidade com uma medida de distância numérica ou um modelo de como o sistema auditivo interpreta qualidade (GRAY Jr. e MARKEL, 1976). Como as distorções introduzidas por algoritmos de melhoria em sinais de voz variam, vários testes e medidas têm surgido para diferentes aplicações.

2.2.1 Medidas Subjetivas de Qualidade

As medidas subjetivas são baseadas em comparações entre o sinal original e o sinal processado, feitas por um ouvinte ou um grupo de ouvintes (DELLER et al., 1993). Os sinais original e processado são apresentados para o grupo de ouvintes que emitem notas,

ou seja, classificam os sinais. Entretanto, para que testes subjetivos possam ser aplicados, as diferenças entre os sinais devem ser praticamente imperceptíveis ao ouvido humano.

Este tipo de medida, como se pode concluir facilmente, é altamente dependente do grupo de ouvintes, ou seja, diferentes grupos de ouvintes podem classificar de forma diferente os mesmos sinais, por isso as avaliações subjetivas devem ser planejadas de modo que não sejam dependentes do grupo de ouvintes. Além disso, este tipo de procedimento requer muito tempo para sua execução e nem sempre os resultados obtidos permitem chegar a uma conclusão sobre o que deve ser melhorado no esquema de processamento. Assim, considerando a diversidade de locutores, de ouvintes e ainda de frases, a aplicação de tais testes passa a ser algo difícil de ser organizado. Porém, desconsiderar os testes subjetivos significa fugir da metodologia mais completa que existe na avaliação de sistemas de melhoria de sinais de voz. Em DELLER et al. (1993) várias propostas de avaliação subjetiva são apresentadas. Para simplificação, dois testes subjetivos seriam suficientes:

a) Apresenta-se ao ouvinte o sinal original e o sinal processado, que deve decidir se os sinais são iguais ou se o sinal processado é melhor ou pior do que o sinal original;
b) Após a aplicação de vários métodos de melhoria em sinais de voz, as frases processadas são apresentadas aos ouvintes pelo menos duas vezes em ordem diferenciada. O objetivo é verificar se o ouvinte não privilegia um método pela sua posição em relação ao sinal teste (ruidoso). O ouvinte deverá apenas informar qual é a melhor frase.

2.2.2 Medidas Objetivas de Qualidade

Pensando nas limitações das medidas subjetivas, alguns algoritmos para estimação da qualidade do sinal processado foram criados. O critério de desempenho de uma medida objetiva de qualidade do sinal de voz é a sua correlação com estimativas subjetivas de qualidade. Dessa forma, como o objetivo da melhoria em sinais de voz é produzir um novo sinal que seja percebido pelo sistema auditivo de forma natural e livre de distorções, é compreensível que as medidas subjetivas sejam meios preferíveis para

classificação de qualidade. Entretanto, à medida que o algoritmo de melhoria de sinais de voz tem sua complexidade aumentada, torna-se necessário ser hábil para distinguir até mesmo as mais sutis diferenças na qualidade do sinal de voz processado.

A análise da correlação entre os sinais original e processado é aplicada para determinar a habilidade da medida objetiva de qualidade em predizer a qualidade julgada pelos ouvintes na avaliação subjetiva.

A classificação de qualidade através de medidas objetivas fornece meios quantitativos, repetitivos e precisos de comparação do desempenho de algoritmos de codificação e melhoria de voz (DELLER et al., 1993). As medidas objetivas fazem uma comparação direta entre o sinal original e sua versão processada. Por isso, é necessário que no processo de cálculo do algoritmo as formas de onda dos sinais de voz, original e processado, estejam sincronizadas.

Um método normalmente utilizado na implementação de medidas objetivas consiste em dividir os sinais de voz em blocos de curta duração, e então calcular a medida de distância / distorção de cada um dos blocos. Feito isto, pode-se formar uma medida final por meio da combinação das medidas de distorção de cada bloco ou observar os valores obtidos ao longo do processamento, apresentados em um gráfico de distâncias / distorção.

Várias medidas objetivas de qualidade são propostas na literatura especializada e a mais comum delas, apresentada a seguir, é a relação sinal ruído (DELLER et al., 1993).

2.2.2.1 Relação Sinal Ruído

A relação Sinal / Ruído (SNR) é a medida mais amplamente utilizada em sistemas analógicos de codificação de forma de onda. Ao longo dos anos foram desenvolvidas diferentes variações, incluindo a SNR clássica, SNR segmentada e SNR segmentada ponderada na freqüência, entre outras (DELLER et al., 1993).

É importante notar que as medidas baseadas na SNR são apropriadas somente para sistemas de codificação ou melhoria que visam reproduzir a forma de onda original, pois seu procedimento de cálculo mede o erro entre as amostras simultâneas dos sinais original e processado. A SNR é dada por (DELLER et al., 1993):

$$SNR = 10\log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right), \tag{2.27}$$

sendo σ_x^2 a variância do sinal original e σ_e^2 a variância do erro, calculado como a diferença entre o sinal original e o sinal processado.

2.2.2.2 Outras Medidas Objetivas (Baseadas em Distância)

Além da SNR, outras medidas baseadas em distância (métrica) entre os sinais podem ser utilizadas.

Em muitos sistemas de melhoria de sinal de voz, a forma de onda do sinal processado, às vezes, é muito diferente da forma de onda original, embora seja percebida da mesma maneira pelo ouvido humano. Medidas como as baseadas na SNR, que obtêm uma medida de distorção baseada em diferenças entre as amostras das formas de onda do sinal original e do sinal processado, não fornecem uma medida significativa de desempenho quando as duas formas de onda diferem em seus aspectos físicos. Portanto, medidas de distância que sejam sensíveis à variação no espectro do sinal de voz tornam-se necessárias.

Uma das medidas de distância muito aplicada, neste caso, é a medida de **Itakura-Saito** (DELLER et al., 1993; GRAY Jr. e MARKEL, 1976). Esta medida é baseada na diferença entre modelos do sinal de referência e da forma de onda do sinal processado. A medida de distância é calculada entre conjuntos de parâmetros de Predição Linear (LP) estimados a partir de blocos sincronizados (tipicamente com duração de 16ms a 32ms) dos sinais de referência e processado.

Outra medida muito usada é a distância cepstral (DELLER et al., 1993; GRAY Jr. e MARKEL, 1976) que representa uma distância euclidiana ponderada entre os sinais original e processado, usando como base seus cepstros. Na verdade, isto representa uma distância direta entre os coeficientes dos seus modelos de Predição Linear. O objetivo é mostrar o quanto os dois sinais estão espectralmente distantes. Assim, a distância cepstral mede o nível de distorção entre os sinais original e processado.

Além dessas medidas existem outras que tentam reproduzir a avaliação subjetiva feita por seres humanos, são as chamadas medidas objetivas baseadas em critérios psicoacústicos, que fazem a avaliação baseada em características do ouvido humano (FLANAGAN, 1972; HOWARD e ANGUS, 1996; SCHROEDER et al., 1979; TSOUKALAS et al., 1993).

Uma das medidas baseadas em critérios psicoacústicos que tem sido amplamente usada para avaliação de sistemas de melhoria de voz é a PESQ (Perceptual Evaluation of Speech Quality), medida proposta pela ITU (International Telecommunications Union) através da recomendação P.862 de fevereiro de 2001 (BEERENDS et al., 2002; ITU-T, 2001; Rix et al., 2000).

A avaliação feita através da PESQ é baseada em características do ouvido humano, onde os sinais original e processado são comparados com relação a atraso, distúrbios simétricos e assimétricos (BEERENDS et al., 2002; Rix et al., 2000). A pontuação final de qualidade, emitida pelo algoritmo da PESQ, é obtida calculando a distância entre os sinais original e processado. A pontuação varia entre –0,5 e 4,5, embora, na maioria dos casos, a faixa de saída assume valores semelhantes aos observados nos testes subjetivos do tipo MOS (Mean Square Opinion) (DELLER et al., 1993), entre 1,0 e 5,0, sendo esta a faixa normal de valores alcançados em experimentos ACR (Absolute Category Rating method) (RIX, et al., 2000). Assim, quanto maior for a semelhança entre os sinais comparados, maior o valor da pontuação obtida.

Não existe nenhuma referência na recomendação P.862 ou em qualquer artigo especializado, mas os profissionais que trabalham com equipamentos de medição PESQ, atribuem ao valor 3 a menor pontuação para um sinal de boa qualidade. Valores PESQ menores que 3, são considerados de baixa qualidade, não sendo adequados ao uso corporativo ou comercial, exceto quando não se tem opção com melhor qualidade ou se faça uma escolha consciente por solução de baixo custo (BEERENDS et al., 2002).

Para avaliação dos métodos propostos neste trabalho são usadas a SNR, a distância cepstral, a PESQ e uma avaliação subjetiva feita por um grupo de ouvintes.

No Apêndice B são apresentados mais detalhes sobre a distância cepstral e PESQ.

CAPÍTULO 3

ANÁLISE WAVELET

Introdução

O primeiro registro do termo "wavelet" data de 1909 em uma tese de Alfred Haar (HAAR, 1910), que apresentou uma função que décadas depois viria a ser conhecida como a primeira função wavelet. O conceito de wavelet, em sua forma teórica atual, foi proposto em meados dos anos oitenta por Jean Morlet (geofísico), Yves Meyer (matemático) e a equipe do Centro de Física Teórica de Marseille, trabalhando sob orientação de Alex Grossman (físico teórico) na França. Os métodos de análise wavelet foram desenvolvidos principalmente por Yves Meyer (MEYER, 1993) e seus colegas, que asseguraram a sua disseminação. A atenção da comunidade de processamento de sinais foi atraída quando Ingrid Daubechies (DAUBECHIES, 1988; DAUBECHIES, 1990; DAUBECHIES, 1992) e Stephane Mallat (MALLAT, 1989(A); MALLAT, 1989(B)), além de suas contribuições para a teoria de wavelets, estabeleceram a conexão entre os dois assuntos e obtiveram resultados via processamento de sinal discreto. O algoritmo de Mallat, apresentado em MALLAT, 1989(A), pode ser considerado um marco na área de processamento de sinais. Desde então, a pesquisa em wavelets tornou-se difundida internacionalmente. Tal pesquisa encontra atividade relevante, particularmente nos Estados Unidos, e vem sendo relatada nos trabalhos de cientistas como Ingrid Daubechies, Ronald Coifman e Victor Wickerhauser (COIFMAN et al., 1993; COIFMAN, 1990; RIOUL e VETERLLI, 1991) entre outros.

A teoria das funções wavelets se situa na intersecção entre a Matemática e o Processamento de Sinais (RIOUL e VETERLLI, 1991). Por outro lado, o Processamento de

Sinais é parte essencial das atividades científicas e tecnológicas contemporâneas. A análise de sinal hoje em dia é usada na televisão, nas telecomunicações, na telefonia, na medicina, em ultra-sonografias, tomografias, ressonâncias magnéticas nucleares, nas quais é necessário processar séries temporais (em uma dimensão) ou imagens (em duas dimensões) (BENEDETO e FRAZIER, 1993; MADHUKUMAR et al., 1997). Também, uma lista de preços de supermercado ou o registro de temperaturas ou variáveis climáticas são sinais que podem ser processados (NIEVERGELT, 1999). A gama de aplicações é muito grande e isto caracteriza a importância de se estudar, investigar e adquirir experiências no uso de técnicas e ferramentas mais atuais de processamento de sinais, com o intuito de aperfeiçoar a codificação, a análise, a transmissão e a reconstituição de sinais em uma ou mais dimensões. A realização de todas essas operações é importante, pois todo sinal possui informação "escondida" em sua representação gráfica (domínio natural), que só é colocada em evidência quando são usadas técnicas mais apuradas (domínio transformado) em sua análise.

3.1 Wavelets

O conjunto $L^2(\mathbf{R})$ é o espaço das funções, mensuráveis a Lebesgue, de quadrado integrável, denominado também de espaço das funções de energia finita, isto é, se $\psi(t) \in L^2(\mathbf{R})$ (HÖNIG, 1977), então:

$$\int_{-\infty}^{\infty} |\psi(t)|^2 \, dt < \infty . \tag{3.1}$$

Definição: Uma função $\psi(t) \in L^2(\mathbf{R})$ é denominada wavelet se, e somente se, sua Transformada de Fourier $\hat{\Psi}(\omega)$ satisfaz a condição

$$C_\psi = \int_{-\infty}^{\infty} \frac{\left|\hat{\Psi}(\omega)\right|^2}{|\omega|} d\omega < \infty . \tag{3.2}$$

A condição anterior, equação (3.2), é chamada de condição de admissibilidade (DAUBECHIES, 1992). Segue da condição de admissibilidade que $\lim_{\omega \to 0} \hat{\Psi}(\omega) = 0$. Assim, se $\hat{\Psi}(\omega)$ é contínua, então $\hat{\Psi}(0) = 0$, ou seja:

$$\int_{-\infty}^{\infty} \psi(t)dt = 0. \tag{3.3}$$

Geometricamente, a condição (3.3) estabelece que $\psi(t)$ deve oscilar de modo a cancelar as áreas positivas e negativas a fim de anular a integral. Portanto, o gráfico de $\psi(t)$ tem a forma de uma onda, conforme ilustra a figura 3.1, que é um exemplo de wavelet.

Como $\psi(t) \in L^2(\mathbf{R})$, então $\lim_{t \to \pm\infty} \psi(t) = 0$ e, como ψ deve estar bem localizada no tempo, este decaimento deve ser muito rápido. Assim, ela terá a forma de uma onda de duração curta.

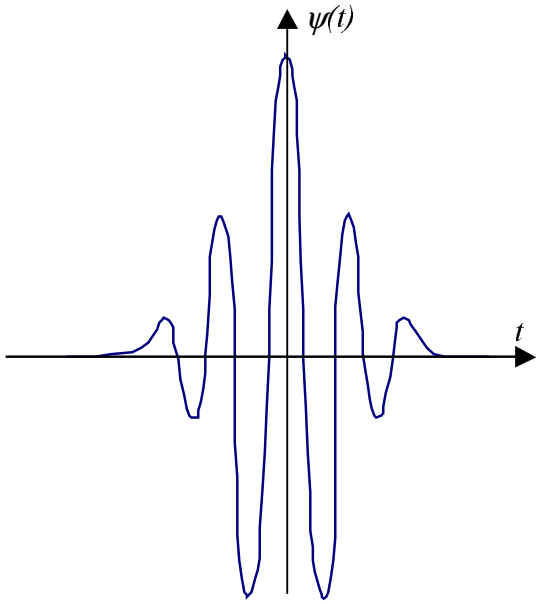


Figura 3.1 - Gráfico de uma Wavelet.

3.2 Transformada Wavelet

A Transformada de Fourier com janela introduz uma escala e analisa o sinal do ponto de vista dessa escala (CHUI, 1992; GOMES et al., 1997; GOMES et al., 2001). Se o sinal possui detalhes importantes fora dessa escala, teremos problemas na análise:

- Se os detalhes do sinal são menores que a largura da janela, tem-se um problema semelhante ao que encontramos com a Transformada de Fourier (GOMES et al., 2001), esses detalhes não serão localizados apesar de serem detectados;

- Se as características forem muito maiores que a largura da janela, tem-se o problema inverso.

Para resolver as deficiências da Transformada de Fourier e da Transformada de Fourier Janelada (Gomes et al., 2001), deve-se obter uma transformada que não use uma escala fixa para análise, mas sim variável para evitar compromisso com uma escala específica.

A escala é definida pela largura da função de modulação. Assim, surge a necessidade de se utilizar uma função de modulação que não tenha escala fixa e que tenha boa localização no tempo. Toma-se, então uma função wavelet $\psi(t)$ como função de modulação. Dados $p \geq 0$ e $0 < a \in \mathbf{R}$, define-se:

$$\psi_a(t) = |a|^{-p} \psi\left(\frac{t}{a}\right) \tag{3.4}$$

a é o parâmetro de escala e $\mathbf{R}$ é o conjunto dos números reais. $\psi(t)$ é uma função localizada no tempo, ou seja, uma função com seus valores concentrados numa determinada vizinhança de um ponto no tempo. Da mesma forma que se faz na Transformada de Fourier com Janela, translada-se a origem para o ponto b, $b \in \mathbf{R}$, ou seja, desliza-se ψ ao longo do eixo-t. Assim, considera-se a função:

$$\psi_{a,b}(t) = |a|^{-p} \psi\left(\frac{t-b}{a}\right). \tag{3.5}$$

Se $\psi_a \in L^2(\mathbf{R})$, então, $\psi_{a,b} \in L^2(\mathbf{R})$:

$$\left\| \psi_{a,b} \right\|^2 = | a |^{1-2p} \left\| \psi \right\|^2 . \tag{3.6}$$

Assim, se $p=1/2$, tem-se

$$\left\| \psi_{a,b} \right\|^2 = \left\| \psi \right\|^2 \tag{3.7}$$

desta forma, se ψ tem norma unitária, $\psi_{a,b}$ também terá.

Considere um sinal (função) $f(t) \in L^2(\mathbf{R})$. Utilizando as funções $\psi_{a,b}$ como funções moduladoras, a Transformada Wavelet Contínua é dada por (DAUBECHIES, 1992):

$$W_f(a,b) = \left\langle f, \psi_{a,b} \right\rangle = \int_{-\infty}^{\infty} f(t) \, | a |^{-p} \, \psi^* \left(\frac{t-b}{a} \right) dt \tag{3.8}$$

sendo ψ^* o conjugado complexo de ψ.

Assim, fazendo uma analogia do fator de escala a com a escala utilizada em mapas, se a escala aumenta, $\psi_{a,b}(t)$ "aumenta" no tempo, capturando comportamentos de tempo longo. Escala maior, "significa" visão global, enquanto que escala menor "significa" visualização de detalhes.

Assumindo $p=1/2$, (3.8) resulta em:

$$W_f(a,b) = \left\langle f, \psi_{a,b} \right\rangle = \frac{1}{\sqrt{| a |}} \int_{-\infty}^{\infty} f(t) \psi^* \left(\frac{t-b}{a} \right) dt . \tag{3.9}$$

3.2.1 Wavelets Filhas (Famílias de Wavelets)

A idéia fundamental da transformada wavelet é a operação de escalonamento pelo parâmetro a. O escalamento possibilita a compressão ($a<1$) ou expansão ($a>1$) da wavelet mãe $\psi(t)$.

A wavelet mãe escalonada, quando deslocada no tempo (translação), origina a família de wavelets ou wavelets "filhas":

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) \tag{3.10}$$

O termo $\frac{1}{\sqrt{|a|}}$ torna a energia das wavelets filhas igual a da wavelet mãe (DAUBECHIES, 1992).

A Transformada Wavelet Contínua ou simplesmente "transformada wavelet" possui várias interpretações. Pela própria definição $W_f(a,b) = \langle f, \psi_{a,b} \rangle$, ou seja, é uma operação de produto interno que pode ser interpretada como uma medida de semelhança entre *f(t)* e cada wavelet filha (DAUBECHIES, 1992). Do ponto de vista de processamento de sinais, a transformada wavelet comporta-se como uma operação de filtragem de um sinal *f(t)* por um filtro cujos coeficientes são gerados pela função wavelet que está sendo usada na análise (STRANG e NGUYEN, 1996).

3.3 Multirresolução

3.3.1 Escala

Tudo que se observa no cotidiano ocorre num processo de escala. Cada observação é feita utilizando uma escala adequada aos detalhes desejados para tal. Assim, quando se faz algum tipo de representação de um objeto, procura-se uma escala adequada na qual os detalhes de interesse possam ser bem representados (GOMES et al., 1997).

A Análise de Multirresolução (AMR) é o modelo matemático adequado para formalizar a representação por escala no universo físico (STRANG e NGUYEN, 1996). Esse problema está intimamente relacionado com as wavelets. Através da AMR encontra-se um método eficiente para construir wavelets (DAUBECHIES, 1992).

A idéia de escala está ligada ao problema da amostragem pontual de um sinal. A freqüência de amostragem é o número de amostras de um sinal por unidade de tempo.

Quando um sinal é amostrado numa freqüência 2^j, o que se faz é fixar uma escala para representa-lo. Detalhes do sinal menores que essa escala são perdidos na representação. Porém, todos os detalhes capturados nessa escala também serão capturados por uma amostragem numa resolução maior, 2^k, $k > j$.

3.3.2 Espaços de Escala

A relação da amostragem com a escala leva a um modelo matemático para formalizar a representação por escala do universo físico. Assim, para cada $j \in \mathbf{Z}$, cria-se um subespaço $V_j \subset L^2(\mathbf{R})$, formado pelas funções cujos detalhes estão na escala 2^j. Então, tais funções são bem amostradas numa freqüência 2^j. Surge a necessidade de se criar um operador que represente uma função $f(t) \in L^2(\mathbf{R})$ na escala 2^j. Pode-se utilizar a representação por projeção ortogonal, bastando, para isso, obter uma base ortonormal de V_j. Porém, existe uma função $\phi(t) \in L^2(\mathbf{R})$ tal que a família de funções:

$$\phi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t - k),\ j,k \in Z \tag{3.11}$$

é uma base ortonormal de V_j (GOMES et al., 1997). Assim, em V_j, a representação por projeção ortogonal de uma função $f(t) \in L^2(\mathbf{R})$ é dada por

$$\mathrm{Pr}oj_{V_j}(f) = \sum_k \langle f, \phi_{j,k} \rangle \phi_{j,k}\ . \tag{3.12}$$

A seqüência de representação $(\langle f(t), \phi_{j,k} \rangle)$ deve conter amostras de $f(t)$ na freqüência 2^j. Como esta seqüência, $(\langle f(t), \phi_{j,k} \rangle)_{j,k \in Z}$, é formada pelas amostras de uma versão filtrada do sinal, $f(t)$, isto é, $\langle f, \phi_{j,k} \rangle = F(k)$ (GOMES et al., 1997), ou seja, F é obtida pela filtragem de $f(t)$ com um filtro de núcleo $\phi_{j,k}$, isto é, $F(T) = \int_{-\infty}^{\infty} f(t)\phi_{2^j}(t - T)dt$, sendo $\phi_s(u) = \frac{1}{|s|^{1/2}}\phi_s(\frac{u}{s})$. Os elementos na seqüência de representação de $f(t)$ ficarão próximos das amostras pontuais de f quando $\phi_{j,k}$ definir um filtro passa-baixa. Isso pode ser obtido exigindo que $\hat{\phi}(0) = 1$ uma vez que $\hat{\phi}(\varpi)$ se aproxima de zero quando $\omega \to \pm\infty$. Sendo

que, ϕ é a função de escala, $\hat{\phi}$ é a transformada de Fourier de ϕ e V_j é o espaço de escala 2^j ou simplesmente espaço de escala.

As larguras das funções $\phi_{j,k}$ e ϕ são relacionadas por:

$largura(\phi_{j,k}) = 2^j largura(\phi)$.

O termo largura está relacionado à compressão ou expansão que a função sofre ao se aplicar o parâmetro de escala.

Assim, quando j decresce, a largura de $\phi_{j,k}$ diminui, refinando a escala, ou seja, aumentando a freqüência de resolução.

Os detalhes do sinal que aparecem na escala 2^j também estão presentes na escala 2^{j-1}, ou seja,

$$V_j \subset V_{j-1} \tag{3.13}$$

Assim, a respeito de tais espaços, tem-se:

a) Dada uma função $f(t) \in L^2(\mathbf{R})$, tem-se que

$f(t) \in V_j$ se, e somente se, $f(2t) \in V_{j-1}$

Isto significa que, escalonando a variável de $f(t)$ por 2, a "largura" da função $f(t)$ é reduzida pelo fator ½. Aplicando sucessivamente este resultado à equação 3.11, conclui-se que $\phi_{0,k}(t) = \phi(t-k)$ é uma base do espaço de escala V_0 (escala 1).

b) O espaço $L^2(\mathbf{R})$ contém todas as possíveis escalas, isto é:

$$\overline{\cup_{j \in Z} V_j} = L^2(R)$$

Por outro lado, a função identicamente nula é a única que pode ser representada em qualquer escala, ou seja,

$\bigcap_{j\in Z} V_j = \{0\}$.

Qualquer função constante pode ser representada em qualquer escala, porém, a função nula é a única função constante em $L^2(\boldsymbol{R})$ (DAUBECHIES, 1992).

3.3.3 Análise de Multirresolução

Definição: Uma Análise de Multirresolução (AMR) em $L^2(\boldsymbol{R})$ é uma seqüência de subespaços fechados $V_j \subset L^2(\boldsymbol{R})$, $j \in \boldsymbol{Z}$, satisfazendo as seguintes propriedades:

(M1) $V_j \subset V_{j-1}$

(M2) f$(t) \in V_j$ se, e somente se, $f(2t) \in V_{j-1}$

(M3) $\bigcap_{j\in Z} V_j = \{0\}$

(M4) $\overline{\bigcup_{j\in Z} V_j} = L^2(R)$

(M5) Existe uma função $\phi \in V_0$ tal que o conjunto $\{\phi(t-k):k\in \boldsymbol{Z}\}$ é uma base ortonormal de V_0

A visualização da seqüência encaixante formada pelos espaços de escala é melhor obtida no domínio da freqüência (STRANG e NGUYEN, 1996). Uma seqüência de intervalos é chamada de encaixante se cada intervalo desta seqüência contém o seguinte (GUIDORIZZI, 1987).

A projeção ortogonal de uma função $f(t) \in L^2(\boldsymbol{R})$ em V_j é obtida através da filtragem de $f(t)$ pelas funções $\phi_{j,k}$, $k \in \boldsymbol{Z}$, que definem filtros passa-baixa. A freqüência de corte desse filtro é α_j (ver figura 3.2). Assim, cada espaço V_j é constituído pelas funções cujas freqüências estão contidas no intervalo $[-\alpha_j, \alpha_j]$, $\alpha_j > 0$. Ao se passar de V_j para V_{j-1}, a escala diminui de 2^j para 2^{j-1}, aumentando a faixa de freqüência para um intervalo $[-\alpha_{j-1}, \alpha_{j-1}]$. O espaço de escala V_{j-1} é formado por todas as funções com espectro de freqüência em $[-\alpha_{j-1}, \alpha_{j-1}]$. O gráfico do espectro $\phi_{j-1,k}$ é a curva pontilhada na figura 3.3.

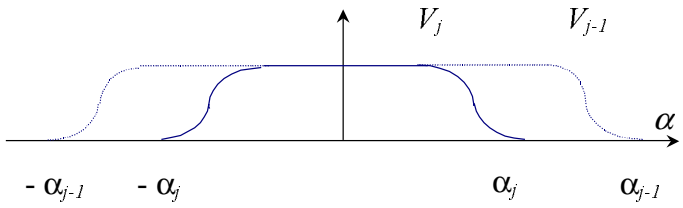


Figura 3.2 - Espectro da função de escala.

Para cada V_j, de escala 2^j, tem-se o operador de representação $R_j : L^2(\mathbf{R}) \rightarrow V_j$

$$R_j(f) = \mathrm{Pr}oj_{V_j}(f) = \sum_k \langle f, \phi_{j,k} \rangle \phi_{j,k} . \tag{3.14}$$

Pelas condições (M4) e (M1) tem-se:

$$\lim_{j \to \infty} R_j(f) = f$$

isto é, partindo da representação de $f(t)$ numa determinada escala $R_j(f) \in V_j$, é possível reobter $f(t)$ acrescentando detalhes em escalas mais altas.

A figura 3.3 ilustra o caso citado anteriormente, ou seja, o espaço V_{j-1} é obtido a partir do espaço V_j acrescentando-se todas as funções de $L^2(\mathbf{R})$ com freqüências na faixa $[\alpha_j, \alpha_{j-1}]$ do espectro.

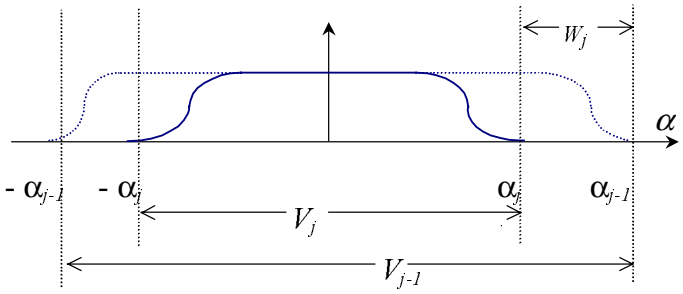


Figura 3.3 - Faixas de freqüências entre V_j e V_{j-1}.

O que diferencia V_j de V_{j-1} é o espaço que representa o complemento entre estes dois espaços. Este espaço será indicado por W_j . Assim, W_j é ortogonal a V_j, podendo-se escrever:

$$V_{j-1} = V_j \oplus W_j .$$

O espaço W_j contém detalhes do sinal na escala j e pode ser obtido usando uma filtragem passa-faixa, cuja faixa de passagem é exatamente o intervalo $[\alpha_j,\alpha_{j-1}]$, conforme a figura 3.4.

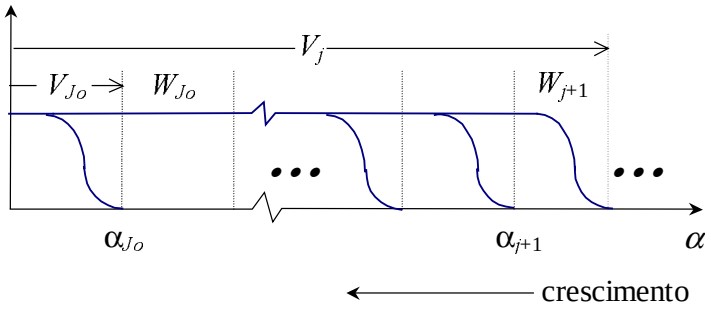


Figura 3.4 - Faixas de freqüências entre V_j e V_{j_0-1}

Além da análise de multirresolução apresentada nesta seção, pode-se destacar ainda as análises de multirresolução bidimensional, biortogonal, biortogonal bidimensional e a análise de multirresolução interpolante (DUARTE et al., 2002).

3.4 Multirresolução e Wavelets

Como foi visto anteriormente, o complemento ortogonal W_j de V_j em V_{j-1} pode ser obtido pelo uso de um filtro passa-faixa de $L^2(\boldsymbol{R})$. Pode-se mostrar que esse espaço complementar é gerado por uma base ortonormal de wavelets (DAUBECHIES, 1992).

Verifica-se que W_j é ortogonal a W_k, se $j \neq k$. Assim, fixando $J_0 \in \boldsymbol{Z}$, para todo $j < J_0$ tem-se:

$$V_j = V_{J_0} \oplus \left(\bigoplus_{k=0}^{J_0-j} W_{J_0-k} \right) \qquad (3.15)$$

isto é, os sinais cujos espectros estão na faixa de freqüência de V_j, representam a soma dos sinais com faixa de freqüência em V_{j_0} e W_{j_0}, W_{j_0-1},..., W_j. Todos estes subespaços são ortogonais. Quando $J_0, k \to \infty$, de (M3) e (M4) tem-se:

$$L^2(R) = \bigoplus_{j \in Z} W_j \ ,$$

ou seja, obtém-se uma decomposição de $L^2(\boldsymbol{R})$ como soma de subespaços ortogonais.

A projeção em cada subespaço W_j pode ser feita usando um filtro passa-faixa (STRANG e NGUYEN, 1996). Este processo pode ser obtido por projeção ortogonal usando uma base de wavelets. Isto se deve ao teorema apresentado a seguir (GOMES et al., 1997).

Teorema 3.1: Para cada $J \in \boldsymbol{Z}$ existe uma base ortonormal de wavelets $\{\psi_{j,k}, k \in Z\}$ do espaço W_j.

Este teorema é uma ferramenta importante para a construção de wavelets, por isso os passos mais importantes de sua demonstração são apresentados a seguir:

Base de W_0 – Os espaços W_j têm as mesmas propriedades de escala dos espaços de escala V_j. Em particular,

$$f(t) \in W_j \text{ se, e somente se, } f(2^j t) \in W_0 \qquad (3.16)$$

Deste modo, basta mostrar que existe uma wavelet $\psi \in W_0$ tal que o conjunto $\{\psi(t-k)\}$ é uma base ortonormal de W_0. Neste caso, de (3.16) tem-se que o conjunto

$$\{\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k)\}$$

é uma base ortonormal de W_j.

Filtro passa-baixa da função de escala – Como $\phi \in V_0 \subset V_{-1}$ e $\phi_{-1,n}$ é uma base ortonormal de V_{-1}, tem-se:

$$\phi = \sum_n h_n \phi_{-1,n} \qquad (3.17)$$

sendo $h_n = \langle \phi, \phi_{-1,n} \rangle$ e $\sum_{n \in Z} \|h_n\|^2 = 1$.

Substituindo $\phi_{-1,n}(t) = \sqrt{2}\phi(2t - n)$ em (3.17) obtém-se:

$$\phi(t) = \sqrt{2} \sum_n h_n \phi(2t - n) \tag{3.18}$$

e aplicando a Transformada de Fourier em (3.18), tem-se:

$$\hat{\phi}(\xi) = m_0(\frac{\xi}{2})\hat{\phi}(\frac{\xi}{2}) \tag{3.19}$$

sendo $m_0(\xi) = \frac{1}{\sqrt{2}} \sum_n h_n e^{-in\xi}$.

Na equação (3.19) $\hat{\phi}(\frac{\xi}{2})$ tem uma faixa de freqüência que é o dobro da faixa de freqüência de $\hat{\phi}(\xi)$. Então, de (3.18), a função m_0 é um filtro passa-baixa. A função m_0, que é periódica e de período 2π, é o filtro passa-baixa da função de escala ϕ.

Caracterização de W_0 – Dada uma função $f(t) \in W_0$, como $V_{-1} = V_0 \oplus W_0$, tem-se que $f(t) \in V_{-1}$ e $f(t)$ é ortogonal a V_0, portanto:

$$f(t) = \sum_{n \in Z} f_n(t)\phi_{-1,n} \tag{3.20}$$

sendo $f_n = \langle f, \phi_{-1,n} \rangle$.

Assim, tem-se:

$$\hat{f}(\xi) = m_f(\frac{\xi}{2})\hat{\phi}(\frac{\xi}{2}) \tag{3.21}$$

com $m_f(\xi) = \frac{1}{\sqrt{2}} \sum_n f_n e^{-in\xi}$. Após alguns cálculos, a equação 3.21 pode ser escrita na forma:

$$\hat{f}(\xi) = e^{i\frac{\xi}{2}}\overline{m_0(\frac{\xi}{2}+\pi)}v(\xi)\hat{\phi}(\frac{\xi}{2}) \tag{3.22}$$

sendo v uma função periódica de período 2π.

Escolha da wavelet – Em (3.22) as funções de W_0 são caracterizadas através da Transformada de Fourier, a menos de uma função periódica v. Uma escolha para a wavelet $\psi \in W_0$ é:

$$\hat{\psi}(\xi) = e^{-i\frac{\xi}{2}}\overline{m_0(\frac{\xi}{2}+\pi)}\hat{\phi}(\frac{\xi}{2}) \tag{3.23}$$

e, sendo assim, de (3.22), tem-se $\hat{f}(\xi) = (\sum_n v_k e^{-in\xi})\hat{\psi}(\xi)$.

Aplicando a transformada inversa de Fourier, tem-se $f(t) = \sum_k v_k \psi(t-k)$. Assim, definida ψ por (3.23), $\psi_{0,k}$ é uma base ortonormal de W_0 (DAUBECHIES, 1992).

A Transformada Wavelet faz uma filtragem passa-faixa e os espaços de escala permitem representar a função $f(t)$ em diferentes resoluções. Quando a representação de $f(t)$ numa escala 2^j é obtida, perdem-se detalhes do sinal com relação à sua representação na escala 2^{j-1}. Os detalhes são calculados pela projeção ortogonal do espaço W_j, isto é,

$$\mathrm{Pr}oj_{W_j}(f) = \sum_{k\in Z}\langle f, \psi_{j,k}\rangle \psi_{j,k}$$

Assim, a função $f(t)$ pode ser obtida de forma exata a partir do sinal de baixa resolução, somando-se as perdas dos detalhes.

3.5 Bancos de Filtros e Análise de Multirresolução

A função de escala de uma AMR é um filtro passa-baixa e a wavelet associada é um filtro passa-alta (STRANG e NGUYEN, 1996). Com base nisso, o que se verá nesta seção é a relação que há entre bancos de filtros e AMR no domínio discreto. No universo contínuo, uma AMR define um banco de filtros (STRANG e NGUYEN, 1996). Assim, a tarefa agora é

mostrar que esta relação também existe no universo discreto. Antes, porém, serão feitos alguns ajustes de notação, pensando em filtragem, em alguns conceitos de AMR para facilitar a discretização.

Se $\phi(t)$ é a função de escala de uma AMR,

$$\ldots \subset V_1 \subset V_0 \subset V_{-1} \subset \ldots$$

tem-se $V_{j-1} = V_j \oplus W_j$, sendo W_j o espaço de wavelets associado. Se $f(t) \in L^2(\mathbf{R})$, então

$$\Pr oj_{V_{j-1}}(f) = \Pr oj_{V_j}(f) + \Pr oj_{W_j}(f). \tag{3.24}$$

$\Pr oj_{V_j}(f)$ é uma filtragem passa-baixa da função $f(t)$ e $\Pr oj_{W_j}(f)$ representa uma filtragem passa-alta de $f(t)$ através do filtro definido pela função wavelet. $\Pr oj_{V_j}(f)$ fornece uma representação de $f(t)$ na escala 2^j e $\Pr oj_{W_j}(f)$ fornece detalhes perdidos na representação de $f(t)$ nessa escala. Pensando em filtros, usando as notações

$$L_j(f) = \Pr oj_{V_j}(f) \qquad \text{e} \qquad H_j(f) = \Pr oj_{w_j}(f)$$

tem-se:

$$L_{j-1}(f) = L_j(f) + H_j(f). \tag{3.25}$$

Decompondo $L_j(f(t))$ sucessivamente, obtém-se:

$$L_{j-1}(f) = L_k(f) + H_k(f) + H_{k-1}(f) + \ldots + H_j(f), \tag{3.26}$$

ou seja, o sinal $f(t)$ é decomposto num componente de baixa freqüência, $L_k(f)$, porém sem muitos detalhes, juntamente com componentes de altas freqüências, $H_n(f)$, $n=k,k-1,\ldots,j$, que contém os detalhes perdidos na representação em baixas freqüências.

De acordo com a equação (3.26), é possível reconstruir *f(t)* de forma exata a partir do componente de baixa freqüência, somando-se os detalhes $H_n(f)$. Assim, no domínio contínuo temos um banco de filtros com reconstrução perfeita.

Os dois diagramas a seguir, figuras 3.5 e 3.6, respectivamente, mostram os processos de análise (decomposição) e síntese (reconstrução) de *f(t)* através do banco de filtros definido por (3.26).

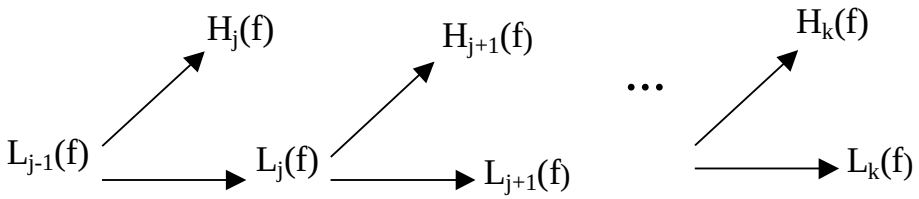


Figura 3.5 – diagrama de decomposição.

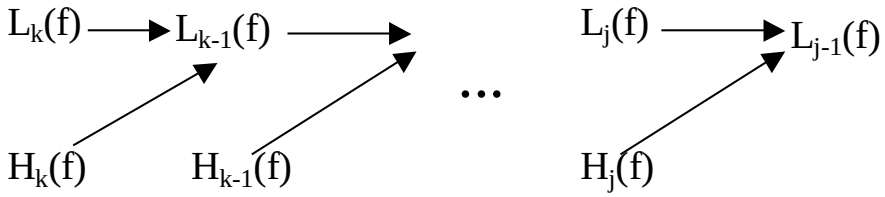


Figura 3.6 – Diagrama da reconstrução.

3.5.1 Análise de Multirresolução Discreta

Se $\psi(t)$ é uma wavelet associada a uma AMR, tem-se:

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k) \tag{3.27}$$

e a relação de escala dupla associada a $\psi(t)$ é

$$\psi(t) = \sum_n d(n)\phi_{-1,n}(t) = \sqrt{2}\psi(t) = \sum_n d(n)\phi_{-1,n}(2t - n) \tag{3.28}$$

sendo que $\phi(t)$ é a função de escala e $d(n) = \langle \psi(t), \phi_{-1,n}(t) \rangle$. Usando esta relação na equação 3.27 tem-se:

$$\begin{aligned} \psi_{j,k}(t) &= 2^{-j/2} \sum_n d(n) 2^{1/2} \phi(2^{-j+1}t - 2k - n) \\ &= \sum_n 2^{\frac{-j+1}{2}} d(n) \phi(2^{-j+1}t - (2k+n)) \\ &= \sum_n d(n) \phi_{j-1,2k+n}(t) \end{aligned} \tag{3.29}$$

e fazendo $n=n-2k$, obtém se

$$\psi_{j,k}(t) = \sum_n d(n-2k) \phi_{j-1,n} \tag{3.30}$$

Daí, conclui-se que:

$$\langle f(t), \psi_{j,k}(t) \rangle = \sum_n \overline{d(n-2k)} \langle f, \phi_{j-1,n} \rangle . \tag{3.31}$$

Usando os operadores de filtragem, a equação (3.31) pode ser representada da seguinte forma (GOMES et al., 1997):

$$\langle f(t), \psi_{j,k}(t) \rangle = (\downarrow 2)[\langle f, \phi_{j-1,n} \rangle * \overline{d(-n)}] \tag{3.32}$$

sendo que $(\downarrow 2)$ é o operador de dizimação de ordem 2 (STRANG e NGUYEN, 1996).

Tomando a função de escala $\phi(t)$ da AMR, a sua relação de escala dupla é, conforme a equação 3.17:

$$\phi(t) = \sum_n h(n) \phi_{-1,n}(t)$$

e da mesma forma que para a wavelet, tem-se:

$$\phi_{j,k}(t) = \sum_n h(n-2k) \phi_{j-1,n} . \tag{3.33}$$

Donde:

$$\left\langle f(t),\phi_{j,k}(t)\right\rangle = \sum_n \overline{h(n-2k)}\left\langle f,\phi_{j-1,n}\right\rangle \tag{3.34}$$

e, na linguagem de filtros:

$$\left\langle f(t),\phi_{j,k}(t)\right\rangle = (\downarrow 2)[\left\langle f,\phi_{j-1,n}\right\rangle * \overline{h(-n)}] \tag{3.35}$$

As equações (3.31) e (3.34) são as equações de decomposição do banco de filtros associados a AMR. O que se apresenta a seguir é uma síntese do funcionamento destas duas equações.

A função *f(t)* é, inicialmente, definida numa escala $f^0(t)$. A função $f^0(t)$ está associada a uma determinada freqüência de amostragem na qual *f(t)* é representada por uma seqüência de amostras. Suponha que $f^0 = \Pr oj_{V_0}(f)$, isto é, $f^0(t)$ é a discretização de *f(t)* no espaço de escala V_0. Como $V_0 = V_{-1} \oplus W_{-1}$, tem-se:

$$f^0 = f^1 + g^1, \quad \text{com} \quad f^1 \in V_{-1} \text{ e } g^1 \in W_{-1}. \tag{3.36}$$

Além disso,

$$\mathrm{f}^0 = \sum_{\mathrm{n}} \mathrm{c}_{\mathrm{n}}^0 \phi_{0,\mathrm{n}} ;$$

$$f^1 = \sum_n c_n^1 \phi_{1,n} ;$$

$$g^1 = \sum_n d_n^1 \psi_{1,n} .$$

De (3.34) e (3.31), obtém-se:

$$c_k^1 = \sum_n h(n-2k) c_n^0 \tag{3.37}$$

$$d_k^1 = \sum_n d(n-2k) c_n^0 \tag{3.38}$$

As equações (3.37) e (3.38) permitem obter a representação de $f(t)$ na escala mais fina V_{-1} a partir da seqüência de representação $(c_n^0)_{n\in Z}$ inicial na escala V_0.

Na verdade, faz-se uma mudança de base, de $\{\phi_{0,n}\}_{n\in Z}$ de V_0 para $\{\phi_{1,n},\psi_{1,n}\}_{n\in Z}$.

Sejam L a matriz do operador em (3.32) e H a matriz em (3.35). Então

$$(c_n^1) = L(c_n^0) \quad e \quad (d_n^1) = H(c_n^0).$$

Da mesma forma, $f^1 \in V_{-1} = V_{-2} \oplus W_{-2}$

e então, $f^1 = f^2 + g^2$ com $f^2 \in V_{-2}$ e $g^2 \in W_{-2}$.

Logo

$$f^2 = \sum_n c_n^2 \phi_{2,n}\,,\ g^2 = \sum_n d_n^2 \psi_{2,n}$$

com

$$(c_n^2) = L(c_n^1) \quad e \quad (d_n^2) = H(c_n^1).$$

Tem-se assim, um banco de filtros de análise ou de decomposição, ilustrado na figura 3.7.

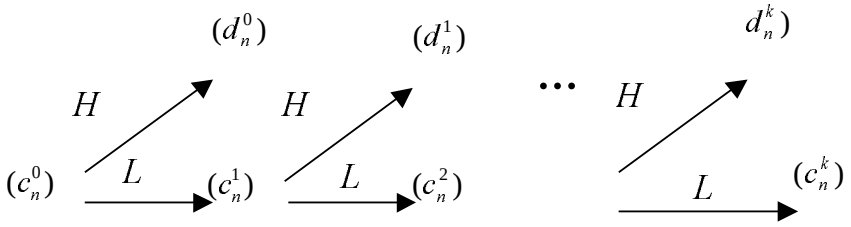


figura 3.7 – Diagrama do Filtro de decomposição discreta

Este tipo de banco de filtros é chamado de banco piramidal (STRANG e NGUYEN, 1996).

Em notação matricial, o banco de filtros S é representado da seguinte forma:

$$S = \begin{pmatrix} L_k & | & 0 \\ H_k & | & 0 \\ - & - & - \\ 0 & | & I \end{pmatrix} \cdots \begin{pmatrix} L_1 & | & 0 \\ H_1 & | & 0 \\ - & - & - \\ 0 & | & I \end{pmatrix} \begin{pmatrix} L \\ -- \\ H \end{pmatrix}, \; I \text{ é a matriz identidade,} \tag{3.39}$$

sendo que a ordem de cada bloco $\begin{pmatrix} L_k \\ -- \\ H_k \end{pmatrix}$ é a metade da ordem do bloco anterior $\begin{pmatrix} L_{k-1} \\ -- \\ H_{k-1} \end{pmatrix}$.

A ordem do bloco da matriz identidade dobra em cada matriz. Esse produto de matrizes é a representação matricial do banco de filtros da AMR. Aplicando este filtro no vetor inicial (c_n^0) tem-se toda a decomposição. O vetor desta decomposição $S((c_n^0))$ é

$$(c^{j_0}, d^{j_0}, d^{j_{0-1}}, ..., d^0) \tag{3.40}$$

sendo

$$d^k = (d_0^k, d_1^k, ..., d_m^k), \quad m \in Z, \quad k = j_0, j_{0-1}, ..., 0 .$$

Como exemplo, apresenta-se a seguir a parte da AMR de *Haar* (GOMES et al., 1997), onde as relações de escala dupla são dadas por:

$$\phi(t) = \phi(2t) + \phi(2t-1)$$
$$\psi(t) = \psi(2t) - \psi(2t-1)$$

Se os sinais forem representados por quatro amostras, a matriz do banco de filtros correspondente, de ordem 4 no primeiro nível, é dada por:

$$\begin{pmatrix} L \\ \\ H \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

ortogonalizando esta matriz, e fazendo $r = 1/\sqrt{2}$ tem-se

$$\begin{pmatrix} r & r & 0 & 0 & 0 \\ 0 & 0 & 0 & r & r \\ r & -r & 0 & 0 & 0 \\ 0 & 0 & 0 & r & -r \end{pmatrix}$$

no próximo nível de escala, a matriz será

$$\begin{pmatrix} r & r \\ r & -r \end{pmatrix}.$$

Assim, a matriz do banco de filtros será dada por

$$\begin{pmatrix} r & r & 0 & 0 \\ r & -r & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r & r & 0 & 0 & 0 \\ 0 & 0 & 0 & r & r \\ r & -r & 0 & 0 & 0 \\ 0 & 0 & 0 & r & -r \end{pmatrix}.$$

Por se tratar de um algoritmo piramidal, o número de operações decresce de nível para nível, na razão 1/2. Assim, conclui-se que a complexidade computacional no cálculo da decomposição pelo banco de filtros associado a AMR é linear no comprimento do sinal de entrada e também em virtude da convolução linear aplicada.

3.5.2 Banco de Reconstrução

As bases formadas pelas wavelets e pelas funções de escala são ortonormais e os operadores envolvidos na decomposição são ortogonais (STRANG e NGUYEN, 1996). Assim, através das matrizes adjuntas, é possível obter as operações inversas. O que leva a uma reconstrução perfeita do banco de filtros. Portanto, pode-se obter as expressões de reconstrução. Sendo:

$$f^{j-1} = f^j + d^j = \sum_k c_k^j \phi_{j,k} + \sum_k d_k^j \psi_{j,k} \tag{3.41}$$

tem-se:

$$
\begin{aligned}
c_n^{j-1} &= \left\langle f^{j-1}, \phi_{j-1,n} \right\rangle \\
&= \sum_k c_k^j \left\langle \phi_{j,k}, \phi_{j-1,n} \right\rangle + \sum_k d_k^j \left\langle \psi_{j,k}, \phi_{j-1,n} \right\rangle, \\
&= \sum_k [h(n-2k)c_k^j + d(n-2k)d_k^j]
\end{aligned}
\tag{3.42}
$$

obtendo assim, a equação de síntese (reconstrução) do banco de filtros associado à AMR: (c_n^{j-1}) é obtido numa escala mais fina, a partir de (c_n^j) e (d_n^j) numa escala com menor resolução. Conforme o diagrama representado na figura 3.8.

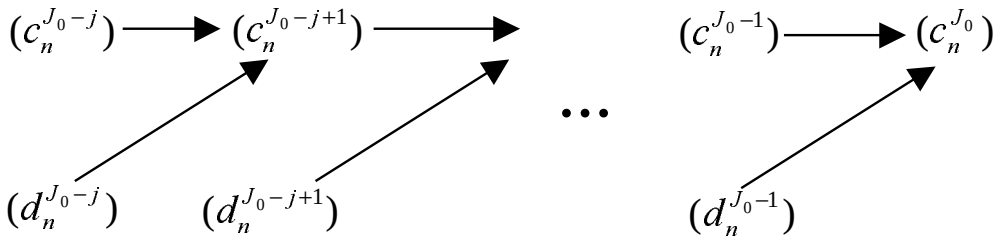


Figura 3.8 - Diagrama de reconstrução discreta.

Analisando o diagrama acima, conclui-se: $(c_n^{j_0})$ foi reconstruído de forma perfeita na escala 2^{j_0}, através da representação de *f(t)* em baixa resolução $(c_n^{j_{0-j}})$ e dos detalhes do sinal em diferentes resoluções intermediárias $(d_n^{j_{0-j}}), (d_n^{j_{0-j+1}}), ..., (d_n^{j_{0-1}})$.

A matriz de reconstrução é a transposta conjugada da matriz *S* em (3.39), já que o operador de análise é ortogonal. O esforço computacional da reconstrução é o mesmo da decomposição, ou seja, linear no comprimento do sinal de entrada.

3.6 Transformada Wavelet Discreta

A transformada wavelet discreta (DWT) envolve a escolha de escalas e posições baseadas em potências de 2, as conhecidas escalas diádicas e posições. A wavelet mãe é reescalonada ou "expandida", por potências de 2 e transladada por inteiros. Especialmente, uma função $f(t) \in L^2(\mathbf{R})$ pode ser representada, em função da função wavelet $\psi(t)$ e da função de escala $\phi(t)$, como (SARKAR, 1998):

$$\mathrm{Pr}oj_{V_{j-1}}(f) = \sum_{j=1}^{L}\sum_{k=-\infty}^{\infty} d(j,k)\psi(2^{-j}t-k) + \sum_{k-\infty}^{\infty} c(l,k)\phi(2^{-l}t-k). \tag{3.43}$$

Lembrando que o conjunto de funções

$$\{\sqrt{2^{-L}}\phi(2^{-L}t-k), \sqrt{2^{-j}}\psi(2^{-j}t-k) / j \le L, j, k, L \in Z\}$$

é uma base ortonormal de $L^2(\mathbf{R})$ e, que $c(l,k)$ são os coeficientes de aproximação na escala L e $d(j,k)$ são os coeficientes de detalhes na escala j. Os coeficientes de aproximação e detalhes são expressos, respectivamente, como:

$$c(l,k) = \frac{1}{\sqrt{2^l}}\int_{-\infty}^{\infty} f(t)\phi(2^{-l}t-k)dt \tag{3.44}$$

e

$$d(j,k) = \frac{1}{\sqrt{2^j}}\int_{-\infty}^{\infty} f(t)\psi(2^{-j}t-k)dt. \tag{3.45}$$

Para entender o papel dos coeficientes expressos em (3.44) e (3.45) na equação (3.43), considere a projeção $f_l(t)$ da função $f(t)$ que proporciona a melhor aproximação para $f(t)$ na escala l, no sentido de minimizar o erro. Esta projeção pode ser construída a partir dos coeficientes $c(l,k)$, usando a equação :

$$f_l(t) = \sum_{k=-\infty}^{\infty} c(l,k)\phi(2^{-l}t+k).$$

Conforme a escala l decresce, a aproximação se torna melhor, convergindo para $f(t)$ quando $l \to 0$. A diferença entre a aproximação na escala $l+1$ e na escala l, $f_{l+1}(t) - f_l(t)$, é completamente descrita pelos coeficientes $d(j,k)$ usando a equação:

$$f_{l+1}(t) - f_l(t) = \sum_{k=-\infty}^{\infty} d(l,k)\psi(2^{-l}t-k).$$

Usando estas relações, dados $c(l,k)$ e $\{d(j,k)/j \le l\}$, fica claro que se pode construir uma aproximação de $f(t)$ em qualquer escala. Assim, a transformada wavelet quebra um sinal

numa aproximação menos refinada $f_l(t)$, dados $c(l,k)$ e um número de camadas de detalhes $\{f_{l+1}(t) - f_l(t) \,/\, j \leq l\}$, dados por $\{d(j,k) \,/\, j \leq l\}$. Conforme cada camada de detalhes é adicionada, a aproximação numa escala melhor é obtida.

3.6.1 Momentos Nulos

O número de momentos nulos de uma wavelet indica a suavidade da função wavelet, assim como o achatamento dos filtros wavelet de resposta à freqüência, filtros esses usados no cálculo da DWT (STRANG e NGUYEN, 1996). Por definição, uma wavelet com p momentos nulos satisfaz a seguinte equação (DAUBECHIES, 1992):

$$\int_{-\infty}^{\infty} t^m \psi(t) dt = 0 \quad \text{para } m=0,1,...,p-1,$$

ou equivalentemente:

$$\sum_k (-1)^k k^m h(k) = 0 \text{ para } m = 0,1,...,p-1.$$

Para a representação de sinais suaves, um número de momentos nulos maior leva a uma taxa de decaimento rápido dos coeficientes wavelet. Assim, wavelets com um alto número de momentos nulos levam a uma representação mais compacta do sinal e são muito úteis em aplicações de codificação (KASTANTIN et al., 1994). Entretanto, em geral, o comprimento do filtro aumenta com o número de momentos nulos e a complexidade no cálculo dos coeficientes da DWT aumenta com o tamanho dos filtros wavelet (KASTANTIN et al., 1994).

3.7 Transformada Wavelet Rápida

Os coeficientes da transformada wavelet discreta (DWT) podem ser computados usando o algoritmo da transformada wavelet rápida (FWT) de Mallat (MALLAT, 1989(A); MALLAT, 1989(B)). Este algoritmo é também conhecido como *codificador subfaixa de dois canais* (STRANG e NGUYEN, 1996) e envolve a filtragem do sinal de entrada baseada na função wavelet usada.

3.7.1 Implementação Usando Filtros

Para melhor entendimento da implementação do algoritmo da transformada wavelet rápida são consideradas as equações:

$$\phi(t) = \sum_k h(k)\phi(2t-k) \tag{3.46}$$

$$\psi(t) = \sum_k (-1)^k h(1-k)\phi(2t-k) \tag{3.47}$$

$$\sum_k h_k h_{k-2m} = 2\delta_{0,m} \tag{3.48}$$

A equação (3.46) é a relação de escala dupla e define a função de escala ϕ. A equação (3.47) expressa a wavelet ψ em função da função de escala ϕ. A equação (3.48) é a condição requerida para que a wavelet seja ortogonal à função de escala e suas translações. Os coeficientes $h(k)$ ou $\{h_0,...,h_{2N-1}\}$ nas equações acima representam os coeficientes de resposta ao impulso para um filtro passa-baixa de comprimento $2N$, com soma 1 e norma unitária. O filtro passa-alta é obtido através do filtro passa-baixa usando a relação $g_k = (-1)^k h(1-k)$ com k variando no intervalo $(1,(2N-1))$. A equação (3.46) mostra que a função de escala é essencialmente um filtro passa-baixa e é usada para definir aproximações dos sinais filtrados. A função wavelet definida na equação (3.47) é um filtro passa-alta e define os detalhes dos sinais analisados (STRANG e NGUYEN, 1996).

Começando com um sinal de entrada $s[n]$, discreto, o primeiro estágio do algoritmo da FWT decompõe o sinal em dois conjuntos de coeficientes. Estes dois conjuntos são os coeficientes de aproximação $c_1[n]$ (informações de baixa freqüência) e os coeficientes de detalhes $d_1[n]$ (informações de alta freqüência), conforme ilustra a figura 3.9.

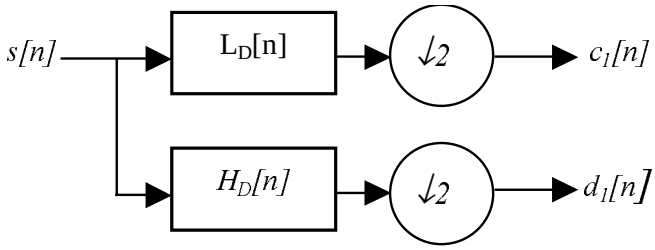


Figura 3.9 - Operação de Filtragem da DWT.

O vetor dos coeficientes de aproximação, $c_1[n]$, é obtido através da convolução de $s[n]$ com o filtro passa-baixa $L_D[n]$ e o vetor dos coeficientes de detalhes, $d_1[n]$, é obtido através da convolução de $s[n]$ com o filtro passa-alta $H_D[n]$. A operação de filtragem é seguida por uma dizimação diádica ou subamostragem por um fator 2, isto porque, após serem feitas as convoluções, o número de coeficientes é dobrado em relação ao sinal de entrada. Assim o que operador de dizimação faz é eliminar todas as amostras de ordem ímpar dos vetores $c_1[n]$ e $d_1[n]$, mantendo assim um número de coeficientes igual ao do vetor original.

Matematicamente, a filtragem de dois canais de um sinal discreto $s[n]$ é representada pelas expressões (SARKAR et al., 1998):

$$c_1(k) = \sum_n l(2n-k)s[n] \quad \text{e} \quad d_1(k) = \sum_n h(2n-k)s[n]. \tag{3.49}$$

Estas equações representam uma convolução mais dizimação de ordem 2 e representam o primeiro estágio da transformada wavelet rápida (FWT).

3.7.2 Decomposição em Multinível

O processo de decomposição pode ser iterado, com os sucessivos vetores de aproximação $c_j[n]$ sendo decompostos a cada nível $j+1$. De forma que um sinal $s[n]$ pode ser decomposto em vários níveis de resolução. O número máximo de níveis de resolução para um sinal $s[n]$ de comprimento n é $j = \log_2 n$. Em cada nível de decomposição j, o vetor de detalhes $d_{j-1}[n]$ é mantido e são feitas novas convoluções do vetor de aproximações $c_{j-1}[n]$ com os filtros passa-baixa e passa-alta, respectivamente, obtendo assim os vetores $c_j[n]$ e $d_j[n]$. O processo de decomposição sucessivo da FWT forma uma estrutura em cascada, ilustrada na figura 3.10.

A decomposição wavelet de um sinal $s[n]$ analisado no nível j tem a seguinte estrutura $[c_j[n], d_j[n],..., d_1[n]]$. Cada um dos novos sinais da decomposição wavelet revela importantes informações a respeito do sinal original (SARKAR et al., 1998). Assim, estes sinais são usados para se fazer as operações desejadas no sinal original, como compressão, codificação, redução de ruído, etc.

Como o processo de análise é iterativo, teoricamente ele pode ser continuado indefinidamente. Porém, na realidade, a decomposição pode ser feita apenas até que o vetor tenha apenas uma amostra. Normalmente, há pouca ou nenhuma vantagem em se decompor um sinal além de um certo nível. A seleção do nível de decomposição ótimo depende da natureza do sinal que está sendo analisado ou de algum parâmetro do sistema, por exemplo, a freqüência de corte do filtro passa-baixa.

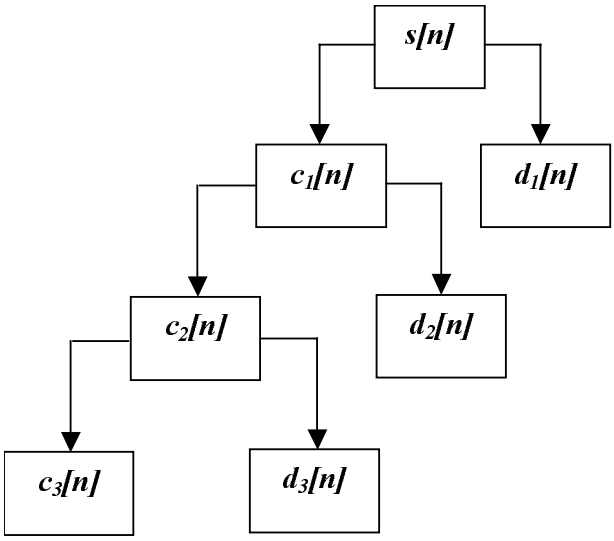


Figura 3.10 - Decomposição dos coeficientes da DWT.

3.7.3 Reconstrução do Sinal

O sinal original pode ser reconstruído ou sintetizado usando a transformada wavelet discreta inversa (IDWT).

A síntese começa com os coeficientes de aproximação e detalhes $c_j[n]$ e $d_j[n]$ e, então, reconstrói $c_{j-1}[n]$ através de inserção de zeros(*upsampling*) e filtragem com os filtros de reconstrução, conforme ilustra a figura 3.11.

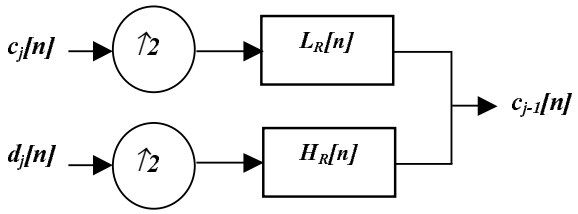


Figura 3.11 – Um passo da Reconstrução da IDWT.

Os filtros de reconstrução são projetados de forma a cancelar os efeitos de *"aliasing"* provocados pelo processo de decomposição wavelet. Os filtros de reconstrução, $L_R[n]$ e $H_R[n]$, juntos com os filtros usados na decomposição $L_D[n]$ e $H_D[n]$ formam um sistema conhecido como *quadrature mirror filters* (QMF) (STRANG e NGUYEN, 1996). Quando a função wavelet usada na decomposição e reconstrução do sinal é uma função ortogonal, os filtros usados na reconstrução são os mesmos usados na decomposição, porém em ordem reversa (SARKAR et al., 1998).

O processo de reconstrução, a cada iteração, passagem do nível *j* para o nível *j-1*, produz aproximações sucessivas em resoluções cada vez melhores do sinal original até que finalmente, no último nível de resolução, sintetiza o próprio sinal original.

Quando os filtros são aplicados através de convolução circular, as operações de subamostragem na decomposição e inserção de zeros na reconstrução não são efetuadas. A convolução circular atenua também o problema gerado na aplicação da DWT nas bordas do sinal (NIEVERGELT, 1999).

CAPÍTULO 4

WAVELETS E O PROCESSAMENTO DE SINAIS DE VOZ

Introdução

A transformada wavelet tem sido usada intensivamente em vários campos de processamento de sinais. Ela tem a vantagem de usar janelas variáveis no tempo para diferentes faixas de freqüência. Isto resulta numa alta resolução nas baixas freqüências (e baixa resolução no tempo) e baixa resolução nas freqüências mais altas (SARKAR et al., 1998). Conseqüentemente, a transformada wavelet é uma ferramenta eficiente para modelar sinais não-estacionários como sinais de voz (STRANG e NGUYEN, 1996). Além disso, quando se está restrito a usar apenas um canal (como a melhoria de um sinal captado por microfone), geralmente o uso de processamento em subfaixas pode resultar num rendimento melhor (SHEIKHZADEH e ABUTALEBI, 2001). Assim, a transformada wavelet pode proporcionar um modelo apropriado para remoção (ou redução) de ruído em sinal de voz.

A remoção de componentes de ruído por limiar dos coeficientes wavelet é baseada na observação de que em muitos sinais (como voz), a energia está concentrada principalmente num número pequeno de coeficientes wavelet (MORETTIN, 1999). Os valores destes coeficientes são relativamente maiores, se comparados aos outros coeficientes ou a algum outro sinal (especialmente voz) que tenha sua energia num grande número de coeficientes. Assim, fazendo os coeficientes com valores absolutos muito pequenos iguais a zero, pode-se eliminar as redundâncias do sinal no domínio wavelet, enquanto se preserva a informação importante do sinal original.

Nas seções seguintes apresentam-se:

- o método wavelet básico (DONOHO e JOHNSTONE, 1994).
- um estudo sobre as funções wavelet e a concentração da energia de um sinal de voz, baseado nas propriedades do sinal;
- os problemas apresentados na aplicação do limiar wavelet básico (DONOHO e JOHNSTONE, 1994) em sistemas de redução de ruído em sinais de voz;
- a apresentação e a discussão de métodos conhecidos na literatura especializada para redução de ruído em sinais de voz usando wavelets.

4.1 Método Wavelet Básico

O método wavelet básico foi proposto por DONOHO e JOHNSTONE (1994) e é explicado a seguir.

Seja y uma seqüência de observação de comprimento finito do sinal x que está contaminado por ruído aditivo r, ou seja:

$$y = x + r. \tag{4.1}$$

Os sinais x e r são considerados processos aleatórios independentes e o objetivo é recuperar o sinal x a partir de y. Se W é a matriz da transformada wavelet discreta (DWT), a equação (4.1), que está no domínio do tempo, pode ser escrita no domínio wavelet como:

$$Y = X + R$$

sendo

$$Y = Wy, \qquad X = Wx, \qquad R = Wr.$$

Seja X_{est} a melhor estimativa do sinal limpo X, baseada na estimação ruidosa Y. No domínio wavelet, o sinal limpo x_{est} pode ser estimado por

$$x_{est} = W^{-1}X_{est} = W^{-1}Y_{thr}$$

sendo que Y_{thr} representa os coeficientes wavelet após a redução do ruído com base no uso do limiar (DONOHO e JOHNSTONE, 1994).

O valor do limiar pode ser determinado de várias formas (MORETTIN, 1999). DONOHO e JOHNSTONE (1994) sugerem um limiar λ de energia como:

$$\lambda = \sigma\sqrt{2\log_{10}(N)}\,, \tag{4.2}$$

sendo N o número de amostras (comprimento) do sinal ruidoso. A variável σ representa o desvio padrão do ruído e é dada por:

$$\sigma = mediana(|Y|)/0{,}6745\,. \tag{4.3}$$

O processamento com o uso de um limiar pode ser realizado com um *limiar duro (hard)* ou um *limiar suave (soft)* conforme definido, respectivamente, a seguir:

$$X_{est} = THR_H(Y,\lambda) = \begin{cases} Y & , se\ |Y| > \lambda \\ 0 & , se\ |Y| \le \lambda \end{cases}$$

$$X_{est} = THR_S(Y,\lambda) = \begin{cases} \text{sgn}(Y)(|Y| - \lambda), & se\ |Y| > \lambda \\ 0 & , se\ |Y| \le \lambda \end{cases}$$

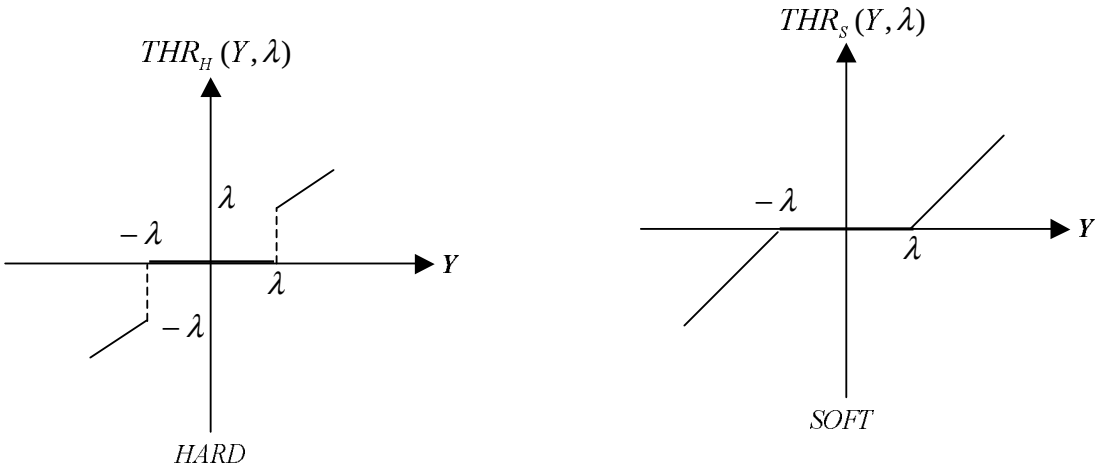


Figura 4.1 - Limar duro e Limiar Suave.

A figura 4.1 mostra como estes dois esquemas agem. O limiar duro é do tipo "elimina" ou "preserva", enquanto que o limiar suave é do tipo "elimina" ou "encolhe" (reduz o tamanho na quantidade λ).

4.2 Wavelets e a Concentração da Energia de um Sinal de Voz

Uma função wavelet concentra informações do sinal de voz (energia) em poucos coeficientes (MORETTIN, 1999). Assim, ao se decompor um sinal usando a transformada wavelet, muitos coeficientes ou serão iguais a zero ou terão magnitudes pouco significantes. Por outro lado, um sinal de voz tem a maior parte de sua energia concentrada nas freqüências mais baixas. Assim, uma wavelet apropriada para processamento de sinal de voz deverá ser uma função que mantenha as propriedades do sinal, principalmente a distribuição de energia do sinal original.

Wavelets com mais momentos nulos proporcionam melhor reconstrução do sinal processado, pois elas introduzem menos distorção no sinal e concentram sua energia em coeficientes numa pequena vizinhança. Porém, a complexidade da transformada wavelet discreta aumenta com o número de momentos nulos, o que deve inviabilizar certas aplicações em tempo real (KASTANTIN et al., 1994; VISWANATHAN et al., 1994).

Vários critérios podem ser usados na seleção da função wavelet ótima para tratamento de um sinal de voz. O objetivo é minimizar o erro na reconstrução do sinal e maximizar a relação sinal / ruído (SNR). Mas, em geral, as melhores wavelets podem ser selecionadas baseadas na propriedade de conservação da energia nos coeficientes de baixa freqüência, já que a maior porção da energia de um sinal de voz está localizada nas baixas freqüências (DELLER et al., 1993).

A seguir são apresentadas, através de estudo feito neste trabalho, três tabelas que trazem o percentual de energia contida na primeira metade do espectro de um sinal de voz decomposto pela transformada wavelet.

Na tabela 4.1 tem-se o percentual da energia nos primeiros $N/2$ coeficientes wavelet de um sinal de voz sem ruído (mais baixas freqüências). A tabela 4.2 apresenta a mesma análise para o mesmo sinal, porém, contaminado por ruído branco. O comprimento do sinal é de 18432 amostras e está dividido em 18 janelas de 1024 amostras, cada uma. A

janela usada neste caso é a janela retangular e é aplicada em sobreposição. A oitava janela é um trecho *unvoiced*, trecho com energia mais bem distribuída ao longo do espectro do sinal. As janelas 7, 12 e 13 são trechos mistos e as três últimas janelas são de silêncio. Todos os outros trechos são do tipo *voiced*, que significa uma maior concentração de energia nas baixas freqüências. A figura 4.2 apresenta a forma de onda do sinal limpo usado para análise. A figura 4.3 apresenta a forma de onda do mesmo sinal contaminado por ruído branco de variância 1 e SNR 11,7335dB.

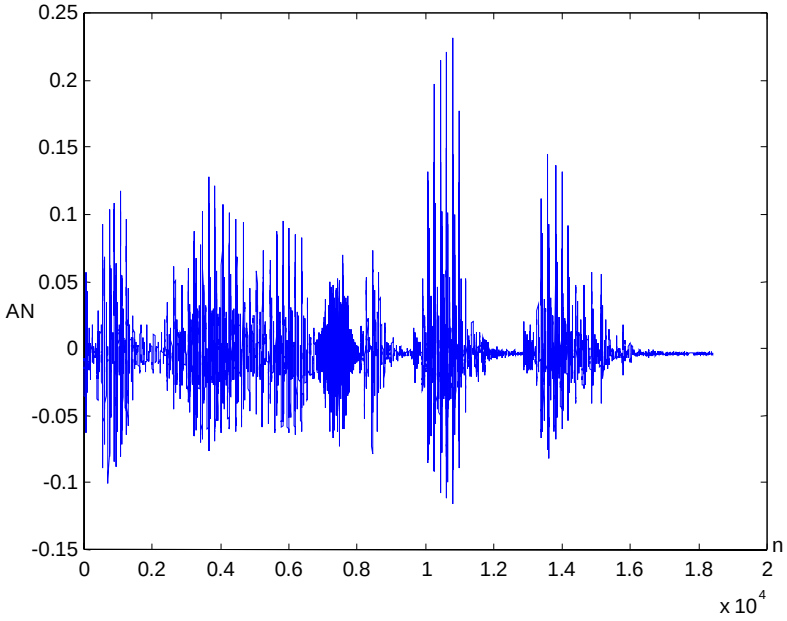


Figura 4.2 - Sinal de voz sem ruído

Nas figuras 4.2, 4.3 e 4.4, AN é a Amplitude Numérica e n é o número de amostras do sinal.

Na tabela 4.3, apresenta-se a mesma análise feita nas tabelas anteriores para o mesmo sinal, porém, agora, contaminado por ruído colorido (ruído de carro gravado numa rodovia asfaltada, a 90km / h) e SNR 10,5845dB.

Para os três sinais, a análise foi feita usando a wavelet de Battle-Lemarie (BL) de ordem 3, a wavelet de Haar, e as de Daubechies (db) de ordem 4, 6, 8 e 10. AGBYNIA (1996) mostra que a wavelet de Battle-Lemarie concentra mais de 97% da energia do sinal de voz na sua primeira metade (faixas de mais baixa freqüência). A wavelet de Battle-Lemarie de ordem 3 tem 4 momentos nulos e seu filtro tem 9 coeficientes.

O que se percebe, principalmente para sinais ruidosos, é que o desempenho das wavelets de Daubechies é praticamente o mesmo da wavelet de Battle-Lemarie, além de serem mais fáceis de implementar, pois a wavelet de Battle-Lemarie é biortogonal. Filtros biortogonais não preservam a energia do sinal, a análise da energia nos coeficientes pode não refletir a energia do sinal reconstruído.

Tabela 4.1 - Percentual da energia nos primeiros N/2 coeficientes wavelet num sinal de voz sem ruído.

Janela	BL	Haar	db4	db6	db8	db10	db20
1	99,66	98,44	99,45	99,62	99,72	99,78	99,88
2	99,50	98,59	99,34	99,42	99,49	99,57	99,67
3	99,42	97,79	98,76	99,03	99,11	99,12	99,58
4	98,95	94,34	96,81	97,70	98,15	98,45	99,11
5	99,48	96,88	98,53	99,02	99,27	99,45	99,67
6	99,43	98,02	98,96	99,23	99,37	99,45	99,64
7	87,03	90,15	89,96	89,59	89,29	89,11	88,92
8	7,80	27,45	19,78	16,31	14,28	12,71	9,79
9	99,53	98,47	99,39	99,52	99,56	99,61	99,72
10	99,26	96,70	98,54	98,85	99,05	99,27	99,31
11	98,75	93,27	96,82	97,85	98,33	98,59	98,93
12	94,11	93,47	94,55	95,29	95,56	95,22	95,13
13	89,65	91,27	90,39	90,06	90,25	90,42	89,68
14	98,47	94,61	96,73	97,31	97,68	97,97	98,52
15	99,49	98,19	99,18	99,36	99,43	99,47	99,65
16	99,31	98,50	99,28	99,56	99,61	99,51	99,76
17	99,41	99,89	99,93	99,94	99,95	99,95	99,95
18	99,33	99,89	99,91	99,91	99,89	99,89	99,90
Média total	92,70	92,55	93,13	93,20	93,22	93,20	93,16
Média de voz	91,37	91,18	91,81	91,88	91,90	91,88	91,83

Conforme observado nas tabelas 4.1, 4.2 e 4.3, todas as wavelets testadas preservam a característica do sinal usando apenas metade dos coeficientes wavelets (freqüências mais baixas). Observa-se, também, no resultado da linha 8 que os trechos unvoiced apresentam valores menores. Isto sugere uma atenção especial para o processamento destes segmentos, visando a obtenção de melhores resultados na redução de ruído.

Observa-se, também, que o percentual de energia nas baixas freqüências é menor para o sinal contaminado por ruído branco. Isto é bastante natural, pois este tipo de ruído possui uma distribuição de energia quase que uniforme em todas as faixas de freqüências

do sinal. Mesmo assim, as wavelets usadas ainda conseguem manter as características do sinal, mantendo apenas os $N/2$ coeficientes iniciais.

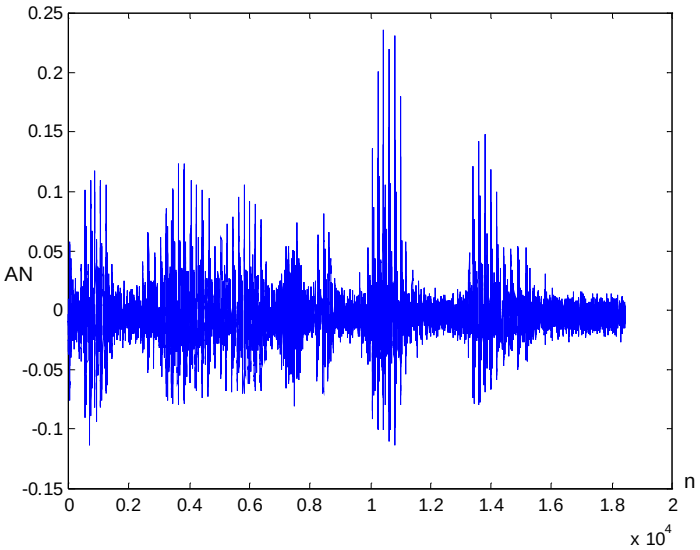


Figura 4.3 - Sinal anterior contaminado por ruído branco.

Tabela 4.2 - Percentual da energia nos primeiros N / 2 coeficientes wavelet num sinal de voz contaminado por ruído branco.

Janela	BL	Haar	db4	db6	db8	db10	db20
1	97,14	95,96	97,19	97,38	97,39	97,40	97,53
2	95,17	94,58	94,99	95,01	95,18	95,35	95,25
3	93,20	90,67	91,95	92,44	92,38	92,04	92,80
4	97,28	92,29	95,15	96,37	96,79	96,88	97,51
5	96,61	94,02	95,45	96,09	96,54	96,71	96,77
6	96,78	95,31	96,31	96,51	96,60	96,65	96,87
7	82,88	86,81	86,32	85,29	84,63	84,57	84,72
8	10,59	26,99	19,03	16,04	14,80	13,86	10,39
9	93,92	93,05	93,63	93,83	93,89	93,79	94,00
10	94,84	91,56	93,92	94,38	94,42	94,36	94,62
11	97,67	91,96	95,85	96,96	97,31	97,45	97,94
12	69,08	68,79	68,46	68,10	68,48	69,08	68,12
13	72,55	70,01	70,58	70,52	69,90	68,96	70,05
14	96,09	92,99	94,61	94,68	94,96	95,49	95,99
15	91,48	90,64	91,10	90,91	90,92	91,16	91,33
16	61,90	62,07	63,05	62,69	61,68	61,15	62,45
17	50,42	57,08	53,24	49,58	49,13	51,24	49,66
18	50,33	51,47	50,36	50,58	51,48	52,05	50,99
Média total	80,44	80,35	80,62	80,41	80,36	80,46	80,39
Média de voz	85,69	85,04	85,64	85,63	85,61	85,58	85,59

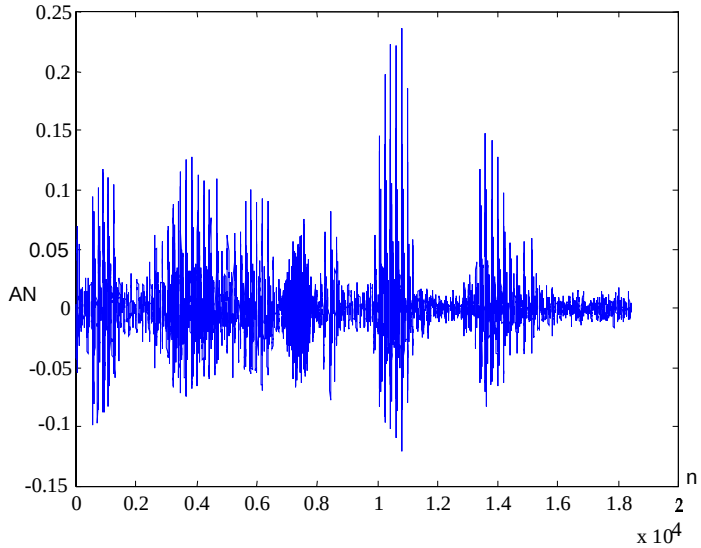


Figura 4.4 - Sinal de voz contaminado por ruído colorido

Tabela 4.3 - Percentual da energia nos primeiros N/2 coeficientes wavelet num sinal de voz contaminado por ruído colorido

Janela	BL	Haar	db4	db6	db8	db10	db20
1	99,66	98,48	99,48	99,64	99,73	99,78	99,88
2	99,51	98,57	99,31	99,40	99,47	99,56	99,65
3	99,43	97,87	98,84	99,06	99,12	99,15	99,57
4	98,97	94,54	96,91	97,75	98,18	98,47	99,11
5	99,49	96,85	98,53	99,00	99,25	99,42	99,66
6	99,44	97,92	98,91	99,20	99,35	99,44	99,63
7	87,75	90,20	90,20	89,92	89,62	89,39	89,22
8	17,12	32,43	25,71	22,63	20,68	19,09	16,50
9	99,56	98,38	99,38	99,52	99,57	99,60	99,73
10	99,35	96,86	98,68	99,06	99,23	99,46	99,45
11	98,79	93,25	96,82	92,87	98,34	98,60	98,95
12	94,99	93,30	94,80	95,58	95,82	95,48	95,49
13	92,90	93,01	92,58	92,45	92,72	92,94	92,32
14	98,52	94,74	96,84	97,42	97,78	98,06	98,59
15	99,48	97,85	99,01	99,23	99,32	99,38	99,59
16	99,33	97,91	99,05	99,30	99,35	99,32	99,62
17	99,64	98,64	99,50	99,66	99,72	99,75	99,84
18	99,64	98,41	99,44	99,63	99,72	99,78	99,84
Média total	93,53	92,73	93,56	93,68	93,72	93,70	93,70
Média de voz	92,33	91,62	92,40	92,51	92,55	92,52	92,49

Concluindo, as funções analisadas são eficientes para o processamento de sinais de voz, tanto para compressão, no caso do sinal limpo apresentado na figura 4.2, quanto para redução de ruído, no caso dos sinais apresentados nas figuras 4.3 e 4.4. Como estas funções têm desempenhos muito próximos quanto à concentração da energia do sinal então para a escolha de uma wavelet que melhor se ajuste ao processamento de um determinado sinal, deve-se levar em conta a SNR, o nível de distorções que esta wavelet introduz no sinal processado e o tempo de processamento.

4.3 Problemas na Aplicação do Limiar Wavelet para Sinais de Voz

Alguns problemas surgem quando o método de limiar wavelet básico, proposto por DONOHO e JOHNSTONE (1994) é aplicado em sinais de voz com ruídos reais (ruído de carro, serra, avião, etc.).

Primeiro, o método básico assume que o sinal está contaminado por ruído branco. Entretanto, na maioria das situações práticas, depara-se com ruídos coloridos. Como resultado, este método não produz uma qualidade satisfatória para muitos dos tipos de ruídos reais. Além disso, o método encontra problemas na remoção de ruídos não-estacionários, como ruído de multilocutores (murmúrio), pois nenhum mecanismo de adaptação no tempo é proposto no algoritmo.

Um outro problema está relacionado com a compressão wavelet dos segmentos *unvoiced* (DELLER et al., 1993) do sinal de voz. Como as partes *unvoiced* do sinal contêm muitos componentes de ruído de alta freqüência, eliminá-las, no domínio wavelet, pode degradar seriamente a qualidade do sinal processado (saída). O uso de um único limiar para todas as faixas de freqüência é outra desvantagem para aplicações em sinais de voz. A aplicação dos métodos *hard* ou *soft* sempre resulta em descontinuidades no tempo e na freqüência, observadas como falhas no espectrograma do sinal melhorado. Isto leva a um certo desconforto auditivo e outras degradações no sinal de saída.

O método de limiar apresentado por DONOHO e JOHNSTONE (1994) foi proposto inicialmente para compressão de sinais e mostrou-se bastante satisfatório. Isto não é surpresa, considerando-se a capacidade de concentração da energia do sinal em poucos coeficientes wavelets. Assim, pode-se eliminar boa parte dos coeficientes no domínio

wavelet sem comprometer a reconstrução do sinal. Baseados nesses princípios, vários métodos de compressão e codificação de sinais foram propostos (DUARTE et al., 2003; KASTANTIN et al., 1994; KINSNER e LANGI, 1993; LOUIS et al., 1998).

Quando se trata de redução de ruído, a aplicação de um limiar wavelet da mesma forma que é usado para compressão deverá eliminar apenas as redundâncias do ruído e do sinal de voz. Assim, para que este limiar seja eficiente na redução de ruído, é necessário que ele seja multiplicado por um fator de correção que leve em consideração a potência do ruído. O problema é que este fator de correção acaba eliminando também coeficientes do sinal de voz com valores próximos aos valores dos coeficientes do ruído, causando uma deterioração de trechos de voz no sinal reconstruído.

As análises anteriores mostram que um método de limiar aplicado à redução de ruído em sinal de voz deve evitar as descontinuidades causadas pelos limiares propostos por DONOHO e JOHNSTONE (1994) e evitar a deterioração dos trechos de voz. A seguir, são apresentados alguns métodos propostos na literatura especializada para redução de ruído em sinais de voz usando a transformada wavelet. Os métodos apresentados em SEOK e BAE (1997) e SHEIKHZADEH e ABUTALEBI (2001) são apresentados com mais detalhes, pois estes, em seu desenvolvimento, tratam todas as características que envolvem sinais de voz ruidosos.

4.4 Melhoria de Sinal de Voz com Redução de Componentes Ruidosos no Domínio Wavelet (SEOK e BAE, 1997)

Este trabalho (SEOK e BAE, 1997) trata o sinal de voz através de corte por limiar, mas em vez de usar um dos tipos de corte por limiar proposto por DONOHO e JOHNSTONE (1994) (*soft* ou *hard*). Ele propõe um método chamado semisoft (semi-suave) no qual se trabalha com dois limiares: λ_1 e λ_2.

Sendo Y o sinal transformado (coeficientes de wavelet) e $\hat{Y}$ o sinal obtido após o limiar, tem-se:

$$\hat{Y} = \begin{cases} 0, & se \ \ |Y| \leq \lambda_1 \\ \text{sgn}(Y)\left(\dfrac{\lambda_2 |Y| - \lambda_1}{\lambda_2 - \lambda_1}\right), & se \ \ \lambda_1 < |Y| \leq \lambda_2, \\ Y, & se \ \ |Y| \geq \lambda_2 \end{cases}$$

sendo que λ_1 é calculado conforme a equação (4.2) e λ_2 é fixado como sendo $\sqrt{2}\lambda_1$. Este tipo de limiar é vantajoso sobre os tradicionais com relação a variância e ponderação do ruído estimado. A figura 4.5 mostra como este método atua sobre o sinal.

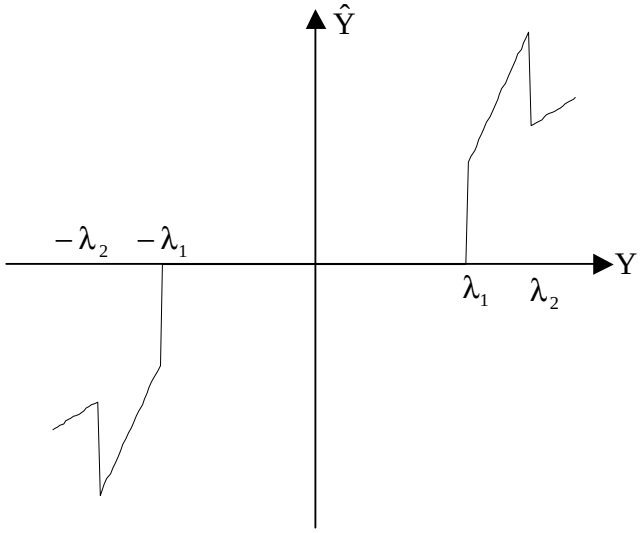


Figura 4.5 - Características de entrada e saída para o limiar semi-suave.

Em sinais de voz, existem sons que se confundem com ruído, como os sons produzidos pelas consoantes *s* e *x*, que são denominados *unvoiced*. Todos os outros são denominados *voiced* (DELLER et al., 1993). Assim, qualquer método para processamento de sinal de voz deve levar em conta o tratamento diferenciado para trechos *unvoiced*. Como estes sons estão localizados quase em todo espectro, com um pouco mais de energia nas altas freqüências, eliminá-los no domínio wavelet pode causar degradação severa de inteligibilidade no sinal reconstruído. Assim, é necessário, primeiro separar regiões *unvoiced* do ruído e, então, aplicar o método de limiar de maneira diferenciada para regiões *voiced* e *unvoiced*.

Para separação de regiões *unvoiced*, a distribuição dos coeficientes wavelet em cada faixa é examinada. Primeiro o sinal de voz é dividido em 4 faixas diferentes como

ilustrado na figura 4.6, e a energia média de cada faixa no domínio wavelet é calculada. O segmento de voz é, então, classificado como *unvoiced* se as duas condições seguintes são satisfeitas: (1) A energia da mais alta faixa de freqüência no domínio wavelet é maior que a energia das outras faixas somadas. (2) A razão entre a energia da mais baixa faixa e a da mais alta faixa é menor que 0,9.

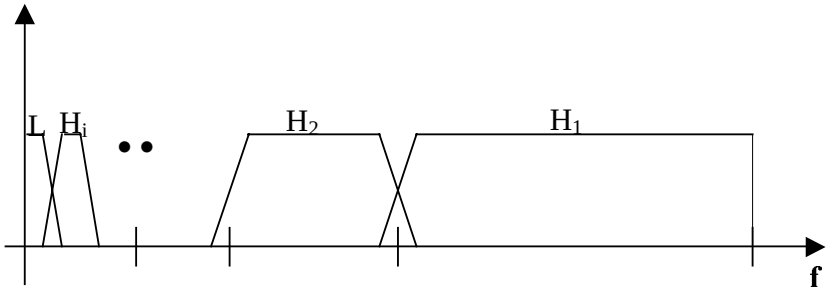


Figura 4.6 - Esquema de Freqüências de um banco de filtros.

A estrutura do algoritmo proposto por SEOK e BAE (1997) é mostrada na figura 4.7 e resumida a seguir, com a aplicação sendo realizada a cada janela analisada do sinal de voz:

a) Calcula-se a transformada wavelet discreta (DWT) do sinal de entrada;
b) Normalizam-se os coeficientes wavelet com a energia da mais alta faixa de freqüência (H_1 na figura 4.6, considerando apenas 4 faixas de freqüências);
c) Determinam-se as regiões *unvoiced*, observando a distribuição de energia dos coeficientes wavelet;
d) Se a região classificada for do tipo *unvoiced*, usa-se o método de limiar apenas para os coeficientes wavelet da mais alta faixa de freqüência. Caso contrário, o limiar é aplicado para todos os coeficientes wavelets;
e) Calcula-se a transformada wavelet discreta inversa (IDWT).

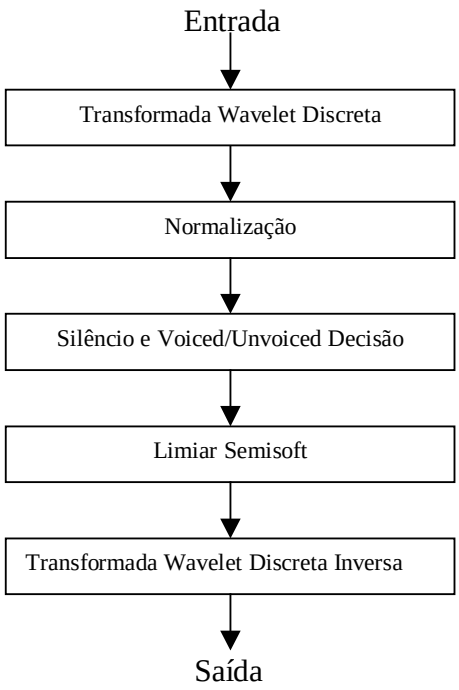


Figura 4.7 - Diagrama de blocos do algoritmo de melhoria de sinal de voz.

Embora este método trate as principais características de um sinal de voz ruidoso, os resultados ainda não são os melhores. Um problema identificado é que, ao deixar de aplicar o limiar nas faixas de baixa freqüência para os trechos *unvoiced*, acaba deixando ruído em excesso. Uma forma de evitar este problema seria ponderar o limiar para cada tipo de trecho (*voiced* ou *unvoiced*).

4.5 Um Sistema de Melhoria em Sinais de Voz Baseado em Wavelets (SHEIKHZADEH e ABUTALEBI, 2001)

Este trabalho (SHEIKHZADEH e ABUTALEBI, 2001) apresenta uma versão modificada do método wavelet básico (DONOHO e JOHNSTONE, 1994), com o objetivo de eliminar os problemas que surgem na sua aplicação. O diagrama de blocos do método proposto é ilustrado na figura 4.8. Os componentes básicos do sistema proposto serão discutidos a seguir.

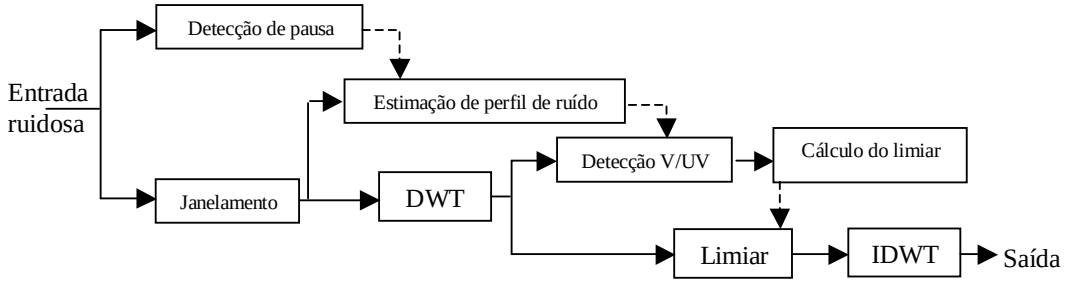


Figura 4.8 - Diagrama de blocos do sistema proposto por SHEIKHZADEH e ABUTALEBI (2001).

4.5.1 Bloco de detecção de pausa

Pausas longas (com duração de algumas centenas de milissegundos) ocorrem naturalmente em uma conversação e podem ser usadas para obter uma estimativa do perfil de ruído (DELLER et al., 1993). Como o ruído, apesar de não-estacionário, apresenta variações espectrais mais lentas do que as variações de uma conversação (pelo menos dez vezes), assume-se que o perfil estimado permanece constante entre duas longas pausas. Estes longos segmentos de pausa podem ser detectados de várias formas possíveis (DELLER et al., 1993) e, normalmente, estão associados às metodologias de redução de ruído.

4.5.2 Estimação de perfil de ruído

Durante uma longa pausa, pode-se obter uma estimação do perfil espectral do ruído. Especificamente, em cada faixa wavelet, calcula-se a variância dos coeficientes ruidosos. Usando-se a fórmula proposta por DONOHO e JOHNSTONE (1994), cada variância pode, então, ser usada para fixar o limiar para um valor baseado na energia do ruído da faixa.

Em situações práticas, esta variância é geralmente encontrada com ruído colorido em vez de ruído branco. A estimação de perfil de ruído proposta capacita o algoritmo a dominar ruídos coloridos.

Seja c_i a seqüência de coeficientes da i-ésima faixa wavelet de ruído. Assumindo ruído gaussiano de média zero, os coeficientes serão variáveis aleatórias gaussianas de

média zero e variância σ_i^2. O desvio padrão σ_i é, então, estimado por (DONOHO e JOHNSTONE, 1994), conforme a equação (4.3):

$$\sigma_i = (1/0{,}6745)mediana(|c_i|).$$

O conjunto de desvios padrões pode agora ser usado como "perfil de ruído" para ajuste de limiar.

4.5.3 Decisão *Voiced/Unvoiced*

Para prevenir a degradação de trechos unvoiced, inicialmente devem-se classificar os segmentos de voz em categorias voiced ou unvoiced (V/UV). O problema da decisão V/UV tem sido discutido extensivamente na literatura (DELLER et al., 1993; SEOK e BAE, 1997). Em SHEIKHZADEH E ABUTALEBI (2001), é implementado um algoritmo simples baseado na transformada wavelet para detecção V/UV. O método usa a distribuição de freqüência da energia média em função da freqüência para cada segmento de tempo. Primeiro, para cada faixa wavelet, a energia média é calculada. Acumulando a energia média das faixas abaixo de 2kHz, pode-se calcular a energia das faixas baixas (EL) deste segmento de tempo. Da mesma forma, a energia das faixas altas (EH) do segmento de tempo pode ser calculada através da acumulação da energia média das faixas acima de 2 kHz. A razão de EL para EH (EL/EH) é o parâmetro fundamental usado na decisão V/UV.

Quando uma pausa longa é detectada, a razão de EL para EH para esta pausa (EL_p/ EH_p) é calculada. Para os segmentos entre esta pausa e a consecutiva, EL_p/ EH_p constitui o limiar de decisão. Se EL/EH é maior que EL_p/EH_p, o segmento é classificado como *Voiced*; caso contrário, ele é considerado *Unvoiced*. Como este método de detecção V/UV considera as propriedades espectrais do ruído, ele é robusto para variações de ruído e funciona para diferentes tipos de ruído.

4.5.4 Adaptação de limiar para segmentos *Unvoiced*

O ajuste de limiar baseado na estimação de perfil de ruído é uma parte importante do sistema, que o capacita a lidar com ruídos coloridos e não-estacionários. Entretanto, o

problema de super filtragem de componentes de alta freqüência dos segmentos *unvoiced* deve ser considerado. Empregando a informação obtida dos dois blocos adicionados (estimação de perfil de ruído e decisão *Voiced/Unvoiced*), os valores de limiar podem ser adaptados para aliviar o problema. Esta adaptação é motivada pelo fato de que em muitos sistemas de melhoria de sinal de voz, os segmentos *voiced* e *unvoiced* são tratados de maneira diferente (DELLER, 1993; SEOK e BAE, 1997).

Como foi discutido na seção 4.5.2, para cada faixa, o valor do limiar λ_i é calculado conforme a equação 4.2, com σ, calculado para todo o sinal transformado, sendo substituído por σ_i, calculado apenas para a faixa de freqüência *i*, equação (4.3). Quando uma nova longa pausa é detectada, atualiza-se o perfil de ruído e, subseqüentemente, o valor do limiar. Geralmente, os segmentos *voiced* estão mais localizados nas faixas de baixa freqüência do que os segmentos *unvoiced*. Para segmentos *unvoiced*, o limiar é reduzido visando uma menor redução de ruído nas freqüências mais altas. Assim, para os outros segmentos, com energia mais significativa nas baixas freqüências, o valor do limiar é aumentado. Dessa forma, obtém-se um sinal com menos ruído e menos distorções.

Este método, (SHEIKHZADEH e ABUTALEBI, 2001) embora faça a redução do ruído por faixa de freqüência nos trechos de voz, usa o limiar calculado no último trecho de voz. Este limiar é baseado apenas na última janela deste trecho de silêncio. Basear o limiar apenas em uma janela pode significar pouca redução de ruído nos trechos de voz, se esta tem pouco ruído, ou forte redução se esta janela tem muito ruído, em ambos casos a inteligibilidade do sinal pode ser comprometida.

4.5.5 Modificação do algoritmo de limiar duro (*Hard Thresholding*)

Como outra melhoria, usa-se uma versão refinada da função de limiar duro em vez da forma padrão proposta por DONOHO e JOHNSTONE (1994). Mais precisamente, em vez de fixar alguns coeficientes wavelet em zero (o que causa descontinuidades tempo-freqüência no espectro do sinal de voz), atenuam-se os coeficientes que são menores que o valor do limiar de uma maneira não-linear para evitar mudanças abruptas.

A forma padrão de limiar duro possui uma característica de entrada e saída mostrada na linha cheia da figura 4.9. No método estudado, a linha tracejada da figura 4.9

é usada como característica de entrada e saída. A parte não-linear desta curva pode ser aproximada por uma função exponencial.

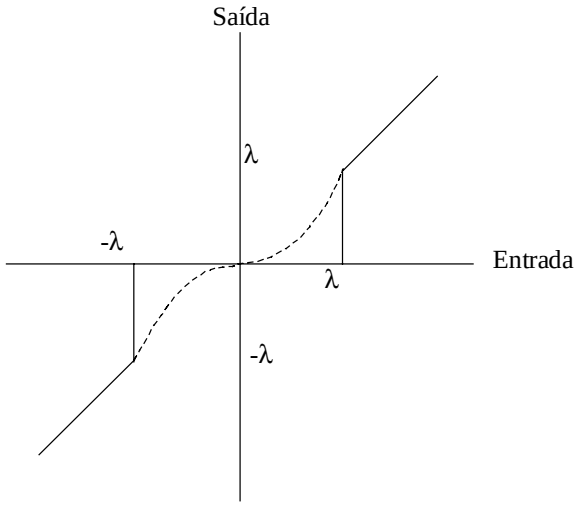


Figura 4.9 - Características de entrada e saída para o limiar duro modificado.

4.6 Outros Métodos Para redução de Ruído em Sinais de Voz (Baseados em Wavelets)

Além do método original proposto DONOHO e JOHNSTONE (1994) e dos métodos apresentados nas duas seções anteriores, (SEOK e BAE, 1997; SHEIKHZADEH e ABUTALEBI, 2001), diversos métodos baseados em wavelets têm sido propostos para redução de ruído em sinais de voz. A maioria deles mostra preocupação na estimação do ruído e na forma de filtragem do ruído com o objetivo de evitar a deterioração dos trechos de voz (AGBINYA, 1996; ZYLANY et al., 2002).

Os métodos propostos por PINTÉR em 1995 e em 1996 são baseados na transformada wavelet perceptiva e usam funções não-lineares para aplicação de limiar. Após a decomposição de wavelet do sinal original, os coeficientes wavelet são multiplicados pelos coeficientes desta função não-linear com o objetivo de reduzir o ruído e evitar as descontinuidades causadas pelo limiar wavelet básico (DONOHO e JOHNSTONE, 1994).

Os métodos propostos por BIJAOUI (2002) e MADHUKUMAR et al. (1997) apresentam algoritmos de redução de ruído baseados no uso de filtros de Wiener (DELLER et al., 1993) no domínio wavelet.

Nos métodos propostos por HU e LOIZOU (2004) e VISSER et al. (2003), aplica-se um limiar no espectro do sinal no domínio wavelet com o objetivo principal de reduzir o ruído musical e efeitos causados por multilocutores (CAPPÉ, 1994; VIEIRA FILHO, 1996).

O método proposto por BAHOURA e ROUAT (2001(A)) aplica o limiar baseado na energia do sinal no domínio wavelet. Esta técnica não requer uma estimação explícita do nível de ruído ou um conhecimento a priori da SNR.

BAHOURA e ROUAT (2000), BAHOURA e ROUAT (2001(B)) e BLACK e ZEYTINOGLU (1995) apresentam métodos para tratamento de sinais de voz baseados na wavelet packet (WP) (MISITI, et al., 1996). Porém, como a WP decompõe também as faixas de alta freqüência, estes métodos não são eficientes para redução de ruído, pois devido ao fato de sinais de voz terem a energia concentrada nas faixas de baixa freqüência, decompor sucessivamente as faixas de alta pode causar a eliminação total dos coeficientes dessas faixas, fazendo com que o sinal reconstruído tenha um tom muito grave.

DAE-SUNG, et al. (2002) propõem uma filtragem passa-baixa nos coeficientes wavelet localizados nas faixas de baixa freqüência do sinal. Para sinais de voz, este método não é eficiente porque não reduz o ruído presente nas faixas de alta freqüência e a aplicação do filtro passa-baixa deteriora muito a voz, mesmo nas faixas de baixa freqüência.

No método apresentado por QIANG e WAN (2003) primeiro faz-se uma subtração espectral no sinal de voz original, atenuando boa parte do ruído, depois decompõe-se este novo sinal usando a transformada wavelet perceptiva . Como limiar, usa-se um vetor de coeficientes que é criado pela relação do limiar segundo DONOHO e JOHNSTONE (1994) e o coeficiente da faixa de freqüência para a qual aquele limiar foi calculado. A seguir, faz-se um produto vetorial, ponto a ponto, entre o sinal no domínio wavelet e este vetor de limiares de forma que haja atenuação e não eliminação de coeficientes, numa tentativa de evitar descontinuidades no sinal reconstruído e o ruído musical. A grande deficiência deste método está na relação limiar / coeficiente, pois isto já atenua o ruído e faz com que o sinal apresente fortes distorções quando o produto vetorial é efetuado.

CAPÍTULO 5

MÉTODO PROPOSTO

Introdução

A redução de ruído em sinais de voz, baseada em wavelets, conforme estudado no capítulo 4, deve ser feita, geralmente, seguindo alguns procedimentos básicos. Estes procedimentos são:

a) Detecção de silêncio: é uma primeira etapa importante, pois quando o sinal está contaminado por ruído aditivo, o perfil do ruído é mais bem estimado nos intervalos de silêncio, ou seja, nas pequenas pausas existentes na fala (DELLER et al. , 1993; SHEIKHZADEH e ABUTALEBI, 2001).

b) Janelamento: os sinais de voz, geralmente, têm comprimento muito extenso e quando se trata de sinais em tempo real é praticamente impossível prever o comprimento final do sinal. Por isso é necessário que o sinal seja segmentado (ou janelado) em intervalos de tempo de alguns milissegundos, principalmente para fins de processamento.

c) Transformação para o domínio wavelet: após o janelamento e a identificação do trecho do sinal como sendo um trecho de silêncio ou de voz, o sinal é decomposto pela transformada wavelet e a operação de redução de ruído é realizada.

d) Retorno para o domínio do tempo: após a redução do ruído, o sinal é reconstruído pela transformada wavelet inversa, retornando para o domínio do tempo.

Com base no processo descrito acima, um bom algoritmo de redução de ruído em sinal de voz deve começar por um bom detector de silêncio / voz, pois no momento da

eliminação do ruído, trechos de silêncio e trechos de voz devem ser tratados de forma diferenciada.

Além da detecção silêncio/voz, deve ser feita também a detecção de segmentos *voiced* e *unvoiced* (SEOK e BAE, 1997; SHEIKHZADEH e ABUTALEBI, 2001), pois de acordo com as tabelas 4.1 a 4.3 apresentadas na seção 4.1, nestes trechos a energia do sinal é distribuída de forma diferente. Assim, o algoritmo de redução de ruído deve atuar de forma diferenciada em segmentos *voiced* e *unvoiced* do sinal.

Um outro ponto importante é a forma com que se aplica o limiar, que pode ser efetuado considerando-se todos os coeficientes wavelets como um único bloco ou considerando-se cada faixa de freqüência. Estudos mostram que a eliminação por faixa de freqüência é mais eficiente (BAHOURA e ROUAT, 2001(B); DUARTE et al., 2004(B); SHEIKHZADEH e ABUTALEBI, 2001), pois a energia do sinal não está distribuída de maneira uniforme nas diferentes faixas de freqüências do sinal no domínio wavelet. Geralmente num trecho de voz comum, tipo *voiced*, 90% da energia está concentrada nas faixas de baixa freqüência, ou seja, na primeira metade dos coeficientes (AGBINYA, 1996). Para que a redução de ruído seja feita por faixa de freqüência, o perfil de ruído, que é estimado nos intervalos de silêncio, deve ser calculado também por faixa de freqüência ou, como uma alternativa, pode-se criar uma curva de ponderação para as diferentes faixas e obter o perfil de ruído a partir da janela inteira.

Nas seções seguintes, serão apresentadas, com detalhes, todas as etapas do algoritmo de redução de ruído em sinal de voz proposto neste trabalho.

5.1 Objetivos do Método Proposto

Os algoritmos de redução de ruído baseados em wavelets apresentados na literatura especializada, em sua maioria, são eficientes para redução de ruído branco, mas não têm a mesma eficiência quando se trata de sinais contaminados por ruído colorido. Por essa razão, o que se propõe é um método que seja eficiente não apenas para a redução de ruído branco, mas também para a redução de ruído colorido. Os passos do algoritmo proposto neste trabalho, que serão descritos a seguir, são, de acordo com o diagrama de blocos apresentado na figura 5.1, os seguintes: pré-filtragem (opcional), janelamento, detecção

silêncio / voz, aplicação da DWT, cálculo do limiar, decisão *voiced/unvoiced*, aplicação do método de limiar e a aplicação da IDWT.

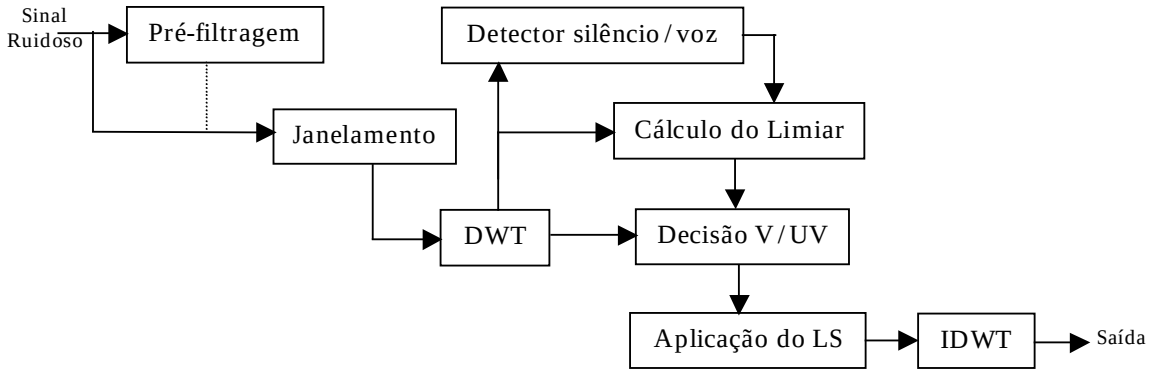


Figura 5.1 – Diagrama de blocos do Método Proposto

5.2 Aplicação da DWT

Considerando um sinal de comprimento N, o número máximo de níveis que a DWT pode decompor este sinal é $n = \log_2 N$. Assim, de acordo com o tipo de processamento que se deseja fazer no sinal, escolhe-se o número de níveis da DWT. A maioria dos métodos de redução de ruído em sinais de voz propostos na literatura especializada não faz menção ao número de níveis usados na decomposição do sinal. Em SEOK e BAE (1997) são usados apenas 4 níveis de decomposição e o sinal processado não apresenta boa qualidade (DUARTE et al., 2004(A)). Neste trabalho, após vários testes feitos na fase de implementação computacional do algoritmo proposto, chegou-se à conclusão que, para redução de ruído, o sinal de voz deve ser decomposto no número máximo de níveis de decomposição. Por isso, o número de níveis de decomposição usado nos testes é $n = \log_2 N$, sendo N, neste caso, o comprimento da janela usada para o janelamento do sinal.

5.3 Detector de Silêncio/Voz

A detecção silêncio / voz de um sinal é uma etapa importante no processamento de sinais de voz, tanto para algoritmos de melhoria quanto, para algoritmos de compressão e codificação (GAZOR e SHANG, 2003; RAMÍREZ et al., 2004; SOHN e SUNG, 1998). Por isso, neste trabalho, propõe-se um detector de silêncio / voz baseado na transformada wavelet.

Uma fórmula muito utilizada para o cálculo do perfil de ruído de um sinal de voz foi proposta por DONONHO e JOHNSTONE (1994). Se Y é um sinal no domínio wavelet, então, o perfil (desvio padrão) de Y será expresso por:

$$\sigma = mediana(|Y|)/0{,}6745. \tag{5.1}$$

Após o janelamento do sinal, é possível calcular um perfil de ruído para cada janela, o que, no final, gerará um vetor de perfis relativos a cada uma das janelas do sinal que está sendo processado. Observando a forma de onda de um sinal de voz e a curva de perfis, apresentadas na figura 5.2, é possível distinguir os trechos de silêncio e os trechos de voz do sinal. Assim, verifica-se que é possível criar um detector silêncio / voz eficiente usando os próprios perfis de ruído das janelas. Na figura 5.2, AN é a amplitude numérica, n é o número de amostras do sinal e J é o número de janelas do sinal.

A execução da detecção de silêncio / voz baseado nos perfis de ruído, proposto neste trabalho, consiste nos seguintes passos:

a) cálculo da média dos perfis nas n primeiras janelas do sinal, que devem ser janelas de silêncio. Normalmente, isto é válido para os primeiros 200ms;

b) a partir da n-ésima janela considerada, cada vez que um novo perfil é calculado, ele é comparado com a média dos perfis de silêncio anteriores. Se for menor que a média mais um desvio, que neste caso é considerado um décimo deste mesmo perfil, esta janela será considerada janela de silêncio, caso contrário, ela será considerada janela de voz;

c) toda vez que uma janela de silêncio é detectada, a média dos perfis é atualizada;

d) repetem-se os passos b e c até que seja alcançada a última janela do sinal.

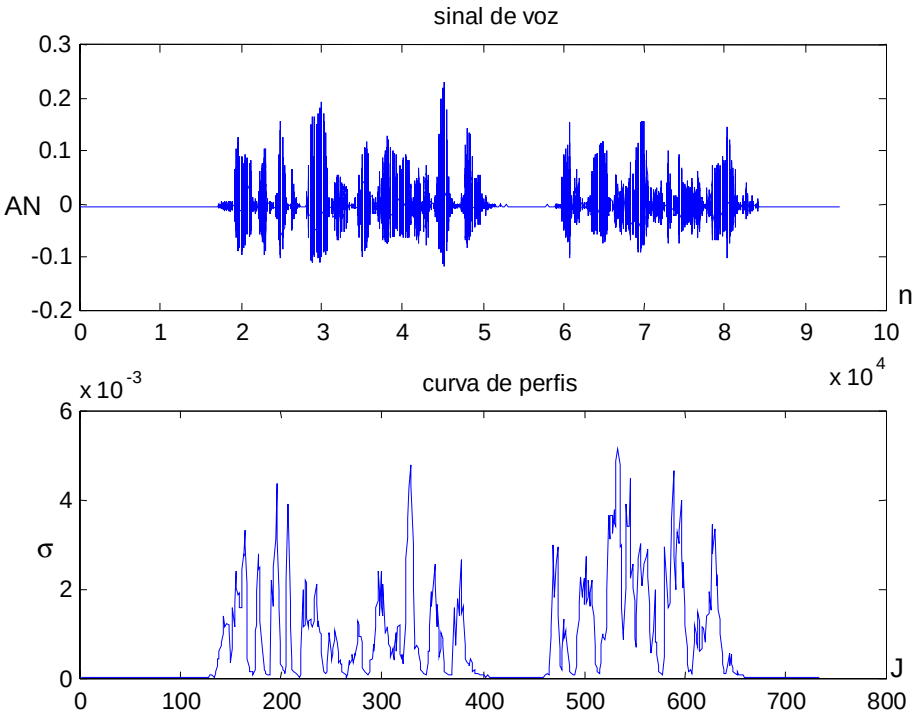


Figura 5.2 – Sinal de voz e sua curva de perfis.

Nas figuras 5.3 e 5.4 tem-se, respectivamente, o mesmo sinal da figura 5.2, porém contaminado por ruído branco (figura 5.3) e por ruído colorido (figura 5.4), com seus respectivos vetores de detecção silêncio / voz. Nas figuras 5.3 e 5.4, AN é a amplitude numérica do sinal, n é o número de amostras, J é o número de janelas e S / V é o resultado da detecção silêncio / voz do sinal.

O detector de silêncio/voz proposto neste trabalho é de simples implementação computacional, pois usa o sinal no domínio wavelet. Este método é robusto e eficiente, pois, como constatado nas figuras 5.3 e 5.4, ele funciona tanto para sinais contaminados por ruído branco como para sinais contaminados por ruído colorido. Além disso, o detector não confunde voz com silêncio, o que evita que, na hora do processamento (redução de ruído), o perfil de ruído seja atualizado nos trechos de voz. Não haverá problemas no processamento se um detector trocar silêncio por voz, o que ele não pode é confundir voz com silêncio.

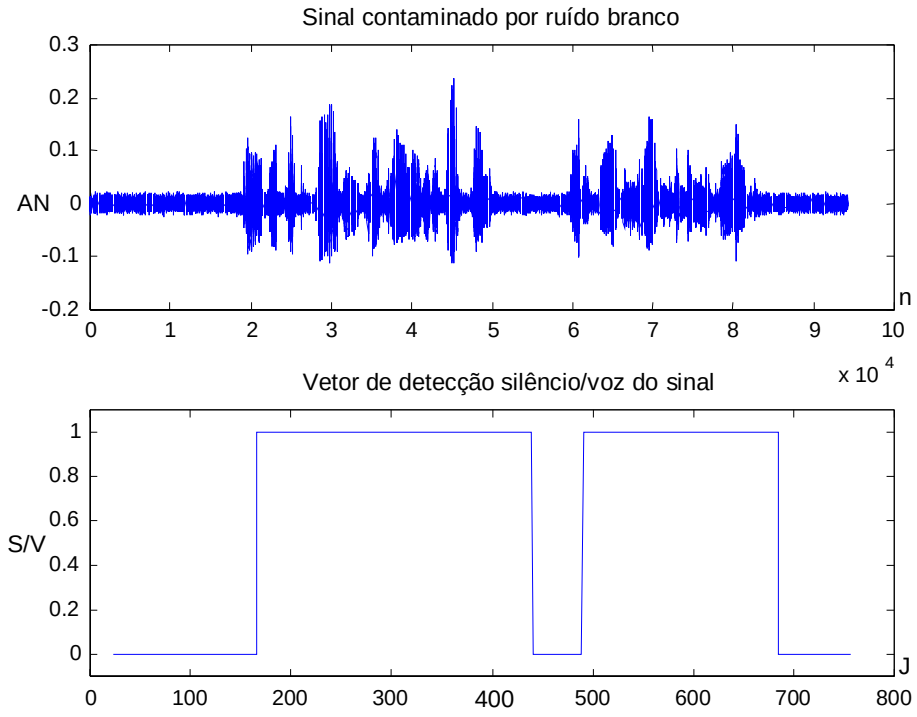


Figura 5.3 – Sinal de voz com ruído branco e seu vetor de detecção silêncio / voz

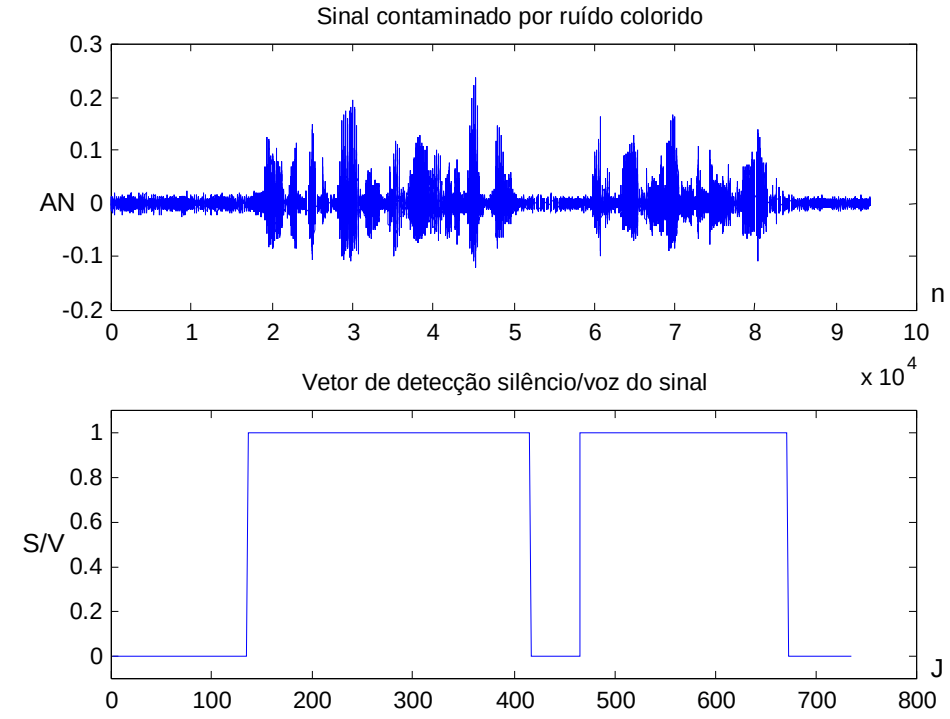


Figura 5.4 – Sinal de voz com ruído colorido e seu vetor de detecção silêncio / voz

5.4 Decisão *Voiced*/*Unvoiced*

Num sinal de voz, os segmentos classificados como *voiced* têm mais de 90% de sua energia localizada nas faixas de baixa freqüência enquanto que os segmentos classificados como *unvoiced* têm a energia melhor distribuída nas faixas de freqüência do sinal (SEOK e BAE, 1997; SHEIKHZADEH e ABUTALEBI, 2001).

Conforme se verifica nas tabelas 4.1 a 4.3, nas janelas *unvoiced*, a energia nas faixas de baixa freqüência está abaixo de 40%, o que significa que a filtragem não pode ser feita da mesma forma para janelas *voiced* e janelas *unvoiced*. Por isso, é necessário que antes do processamento do sinal, já janelado, este segmento seja identificado como *voiced* ou *unvoiced*, para que a filtragem seja feita de acordo com suas características.

Além dos trechos *voiced* e *unvoiced*, existem os trechos mistos, que são, geralmente, janelas que estão entre uma janela *voiced* e uma *unvoiced*. Como exemplo, pode-se citar a janela 7 na tabela 4.2. Nestas janelas, porém, a energia nas faixas de baixa freqüência é algo próximo de 90%, podendo, assim, ser tratadas da mesma forma que as janelas *voiced*.

A decisão entre *voiced* ou *unvoiced* é feita verificando-se o percentual de energia nas faixas de baixa freqüência. Se este percentual for maior que 40% a janela deve ser tratada como *voiced,* caso contrário a janela deve ser considerada *unvoiced*.

5.5 Cálculo do Limiar

O perfil de ruído é calculado nos trechos de silêncio usando a equação (5.1) e depois o limiar é calculado usando a equação (4.2).

Estudos comprovam que o método de limiar gera bons resultados quando a aplicação do limiar é feita por faixa de freqüência (DUARTE et al., 2004(B); SHEIKHZADEH e ABUTALEBI, 2001). Nestes métodos, o perfil de ruído e o limiar são calculados nos trechos de silêncio e aplicado para os trechos de silêncio e também para os trechos de voz. Assim, toda vez que é detectada uma janela de silêncio, o perfil de ruído é atualizado para cada faixa de freqüência da seguinte forma:

$$\sigma_i = mediana(|c_i|)/0{,}6745, \tag{5.2}$$

sendo que *i* é a faixa de freqüência para a qual o perfil de ruído está sendo atualizado e c_i é o vetor de coeficientes wavelet da faixa de freqüência *i*. O limiar é atualizado conforme a equação (5.3), que é uma versão da equação (4.2) para a faixa de freqüência *i*, de comprimento N_i.

$$\lambda_i = \sigma_i \sqrt{2\log(N_i)} . \tag{5.3}$$

Pelo método proposto neste trabalho, para cada janela de silêncio, o limiar é calculado conforme a equação (5.3), por faixa de freqüência. Este valor é aplicado, tanto nos trechos de silêncio quanto nos trechos de voz, também por faixa de freqüência. Porém, quando se chega numa nova janela, o que se tem é um vetor de limiares, cujos componentes correspondem às médias dos limiares calculados para cada uma das faixas de freqüência de todas as janelas do trecho de silêncio. Tanto nos trechos de silêncio quanto de voz, cada componente do vetor de limiares será aplicado na sua devida faixa de freqüência. Assim, nos trechos de silêncio, o limiar é calculado da seguinte forma:

$$\lambda_{M,j} = (1-\alpha)\lambda_{i,j} + \alpha\lambda_{MA,j} \tag{5.4}$$

sendo:

$\lambda_{M,j}$, o limiar médio para a faixa de freqüência j;

$\lambda_{i,j}$, o limiar calculado na janela de silêncio *i* para a faixa de freqüência *j*;

$\lambda_{MA,j}$, a média dos limiares das janelas de silêncio anteriores para a faixa de freqüência *j*

α, um valor real entre 0 e 1.

Toda vez que um novo trecho de silêncio é detectado, os limiares são atualizados para todas as faixas de freqüências. O parâmetro α faz com que a média seja calculada de forma suave e que nenhum valor de limiar muito grande mude de maneira abrupta o valor médio.

5.6 Aplicação do Limiar

Como já foi descrito na seção anterior, o limiar é aplicado tanto nos trechos de silêncio quanto nos trechos de voz por faixa de freqüência.

A energia de um sinal no domínio wavelet está concentrada em poucos coeficientes (MORETTIN, 1999; STRANG e NGUYEN, 1996), isto é, muitos coeficientes no domínio wavelet têm valores absolutos muito pequenos, de forma que a eliminação destes coeficientes não afetaria a reconstrução do sinal pela transformada wavelet inversa, gerando um sinal praticamente idêntico ao sinal original. Este é o princípio básico usado na compressão e codificação de sinais usando a transformada wavelet (AGBINYA, 1996; DUARTE et al. 2003; KINSNER e LANGI, 1993). Estes coeficientes, porém, num sinal contaminado por ruído, podem interferir na estimação do perfil de ruído, causando a eliminação de coeficientes importantes do sinal de voz e gerando degradação, principalmente nos trechos de voz do sinal reconstruído.

A aplicação original dos métodos de limiar propostos por DONOHO e JOHNSTONE (1994) é para compressão, ou seja, o limiar é aplicado usando o método de limiar duro ou suave apenas para a eliminação das redundâncias do sinal. Se o objetivo é a redução ou eliminação de ruído, o limiar deve ser multiplicado por um fator de correção que leva em consideração a potência do ruído de fundo presente no sinal de voz. Isto, porém, acaba gerando descontinuidades no tempo e na freqüência do sinal reconstruído, que embora faça uma boa redução ou, às vezes, uma eliminação total do ruído, também deteriora os trechos de voz, tornando o sinal ininteligível.

Com base na discussão apresentada no parágrafo anterior, propõe-se um método que faça a redução de ruído de forma a tornar o sinal inteligível após sua reconstrução. Neste método, usa-se uma versão de limiar suave, chamado de limiar sigmóide (LS), que em vez de eliminar os coeficientes que têm valor absoluto abaixo de um determinado valor, apenas os atenua, evitando assim descontinuidades no sinal reconstruído e, conseqüentemente, evitando a presença do ruído musical.

O uso de uma função sigmóide para a eliminação do ruído se justifica pela sua não-linearidade, pois um método linear como o proposto por DONOHO e JOHNSTONE (1994) causa degradação das faixas de voz do sinal reconstruído (SEOK e BAE, 1997; SHEIKHZADEH e ABUTALEBI, 2001). A função sigmóide usada, expressa na equação

(5.5), faz com que não haja eliminação de coeficientes, ou seja , que não haja coeficientes que sejam igualados a zero, mas sim que os memos sejam atenuados.

$$Sigmóide(x) = \frac{1 - e^{-\gamma x}}{1 + e^{-\gamma x}} \tag{5.5}$$

O parâmetro γ controla a inclinação da sigmóide. A figura 5.5 apresenta gráficos da sigmóide para vários valores de γ.

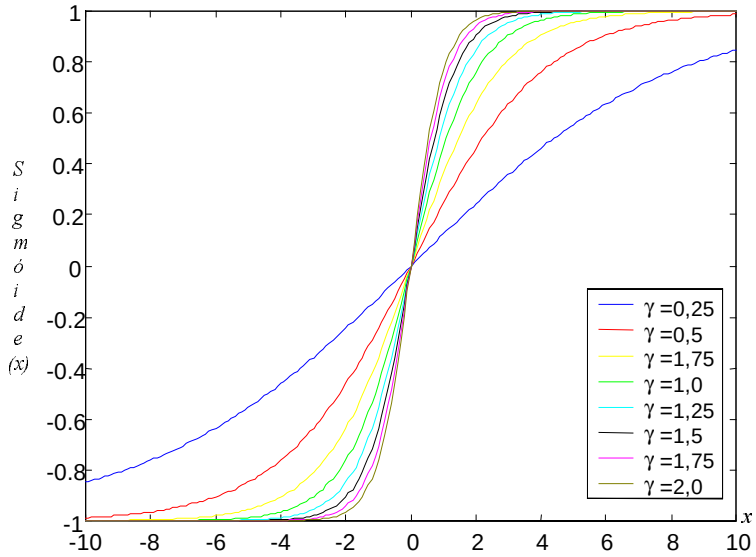


Figura 5.5 – Sigmóides para vários valores de γ.

A aplicação do LS é proposta de duas maneiras:

a) Limiar Sigmóide Direto – LSD

Tanto para os trechos de silêncio quanto para os trechos de voz, para cada faixa de freqüência, levando em consideração a aplicação para segmentos *voiced* e *unvoiced* que será explicada na seção 5.6.1, o limiar é aplicado da seguinte forma:

$$\hat{Y} = \begin{cases} Y \quad , \text{se} \mid Y \mid > \beta\lambda \\ Y. \mid \text{sigmóide}(Y) \mid \quad , \text{se} \mid Y \mid \leq \beta\lambda \end{cases} \tag{5.6}$$

sendo β um fator de correção baseado na potência do ruído.

b) Limiar Sigmóide Ponderado – LSP

Nos trechos de silêncio, o limiar é aplicado da mesma maneira que no LSD, enquanto que nos trechos de voz, para cada janela, usa-se como limiar para cada faixa de freqüência, um valor que é a razão entre um limiar calculado para a faixa de freqüência em questão na própria janela de voz e o valor que representa a média dos limiares calculados para esta faixa no último trecho de silêncio. Este valor é usado para que o ruído em cada janela de voz seja considerado, além do ruído já estimado no último trecho de silêncio. Assim, nos trechos de voz, para cada faixa de freqüência, o valor de λ usado na equação (5.6) é substituído por:

$$\lambda_r = \frac{\lambda_{voz,i}}{\lambda_i} \tag{5.7}$$

sendo que λ_i é o limiar médio calculado no trecho de silêncio para a faixa de freqüência i e $\lambda_{voz,i}$ é o limiar calculado na janela de voz que esta sendo analisada para a faixa de freqüência i.

O limiar proposto na equação (5.7) apresenta duas grandes vantagens sobre os outros métodos que aplicam limiar, inclusive sobre o LSD, que são:

1) O limiar calculado na equação (5.7) pode ser usado tanto para sinais contaminados por ruído estacionário, quanto para sinais contaminados por ruído não-estacionário. Os limiares usados no LSD, em BAHOURA E ROUAT (2001(B)), DONOHO e JOHNSTONE (1994), SEOK e BAE (1997) e SHEIKHZADEH e ABUTALEBI (2001), assumem que o ruído é estacionário ou quase-estacionário, pois uma vez que o ruído é estimado num trecho de silêncio, ele só será atualizado no próximo trecho de silêncio, o que não acontece no método LSP. Embora a equação (5.4) faça um acompanhamento das variações do ruído, ela só o faz para os trechos de silêncio, já que λ_i é calculado nos trechos de silêncio. Assim, se houver alguma variação no ruído no trecho de voz, ela será detectada por λ_r.

2) Usando um limiar conforme proposto neste método, não há necessidade de se fazer a decisão *voiced / unvoiced*, pois λ_r, da forma como é calculado, já leva em conta a energia do sinal para os diferentes tipos de trechos de voz do sinal. Assim, quando o sinal é processado usando o LSP, o bloco de decisão *voiced / unvoiced* apresentado na figura 5.1 não deve ser considerado. Depois do cálculo do limiar passa-se direto para a aplicação do limiar.

A eliminação da decisão *voiced / unvoiced* se deve ao fato de que λ_i é calculado no silêncio e, por isso, é baseado apenas na potência do ruído. $\lambda_{voz,i}$ é baseado também na potência da voz, mas se for aplicado direto, acabará eliminando voz também daí o uso de λ_r. Porém, este fato, faz que no próprio cálculo de $\lambda_{voz,i}$ já se faça uma detecção *voiced / unvoiced*, o torna desnecessário um mecanismo apenas para este fim.

Na figura 5.6 observa-se as características de entrada e saída para o método de limiar sigmóide.

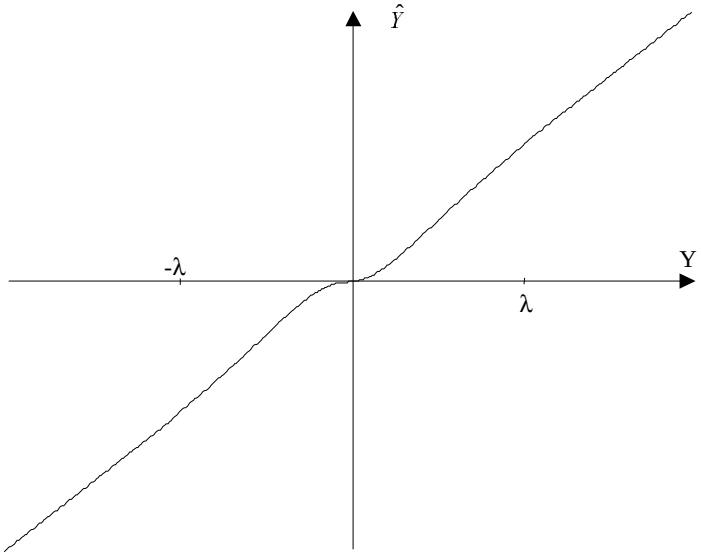


Figura 5.6 – Características de entrada e saída do limiar sigmóide.

Comparando o gráfico da figura 5.6 com os gráficos da figura 4.1, pode-se comentar o seguinte:

- O método de limiar duro apresenta uma transição abrupta quando se passa dos coeficientes que estão abaixo do limiar para os coeficientes que estão acima, o que é percebido pela descontinuidade apresentada no seu gráfico. Esta transição abrupta é a causa das distorções apresentadas no sinal de voz quando este método é aplicado para redução de ruído.
- O método de limiar suave não apresenta um desnível quando se passa dos coeficientes que estão abaixo do limiar para os que estão acima, o que também se percebe pelo seu gráfico. Mas, ao reduzir os coeficientes que estão acima do limiar, ele reduz a amplitude do sinal reconstruído e isso acaba mascarando o ruído, dando uma falsa impressão de que houve maior redução de ruído.
- O limiar sigmóide, além de não eliminar os coeficientes que estão abaixo do limiar, mas sim atenuá-los, apresenta uma transição menos abrupta quando se passa dos coeficientes que estão abaixo do limiar para os que estão acima.

Uma outra alternativa para a aplicação do limiar seria, na equação (5.6), trocar a expressão $Y.|sigmóide(Y)|$ por $sigmóide(Y)$, ou seja, em vez de multiplicar o coeficiente que está abaixo do limiar pela sua sigmóide, o coeficiente seria apenas trocado pela sua sigmóide, fazendo assim, uma aplicação direta da sigmóide. O que também geraria uma transição abrupta, conforme observado na figura 5.7.

Desta forma, vê-se que a aplicação do limiar sigmóide, como é proposto nesta seção, acaba sendo uma boa opção para evitar as distorções causadas pelo método de limiar duro ou a falsa impressão da eliminação de ruído causada pelo método de limiar suave, ambos propostos por DONOHO e JOHNSTONE (1994). A simples troca do coeficiente pela sua sigmóide, atenua as distorções causadas pelo limiar duro, conforme observado na figura 5.7, mas mesmo assim acaba gerando mais distorções que o método de limiar proposto neste trabalho.

No apêndice A, é apresentado um estudo mais aprofundado da aplicação de limiar proposta neste trabalho.

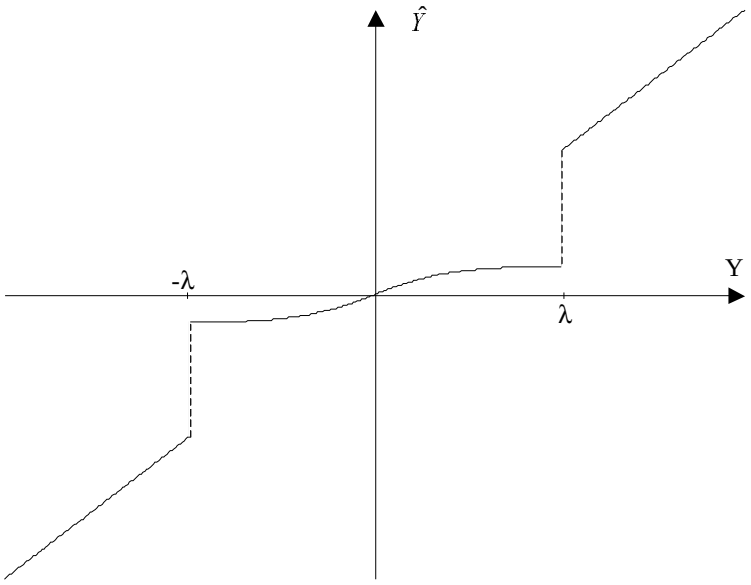


Figura 5.7 – Características de entrada e saída da aplicação direta da sigmóide.

5.6.1 Aplicação do Limiar para Trechos *Voiced* e *Unvoiced*

Nos trechos *voiced*, mais de 90% da energia do sinal está concentrada nas faixas de baixa freqüência do sinal de voz enquanto que nos trechos *unvoiced*, este percentual não passa de 40%. Os trechos cujo percentual de energia nas baixas freqüências está entre 40% e 90% são considerados mistos (*voiced/unvoiced*) (AGBINYA, 1996). Alguns métodos aplicam o limiar apenas nas faixas de baixa freqüência para trechos *voiced* e nas faixas de alta para trechos *unvoiced*, deixando a outra porção do sinal como está (SEOK e BAE, 1997). Isto acaba deixando ruído na porção que não passou pela filtragem que, mesmo sendo pouco, pode se tornar desagradável ao ouvinte, normalmente como "estampidos".

Com as considerações anteriores, propõe-se que, nos trechos considerados *voiced*, o limiar seja aplicado de forma normal nas faixas de baixa freqüência e de forma atenuada nas faixas de alta freqüência. Já para os trechos *unvoiced* deve-se fazer o contrário, ou seja, atenuar o limiar para as faixas de baixa freqüência e o aplicá-lo de forma normal para as faixas de alta freqüência. Esta atenuação consiste em aumentar ou diminuir o parâmetro β de acordo com o segmento de voz. Os trechos mistos são tratados como *voiced*.

5.7 Pré-filtragem

No domínio wavelet, a energia de um sinal está concentrada em poucos coeficientes (MORETTIN, 1999; STRANG e NGUYEN, 1996). Por isso, propõe-se, antes do uso do limiar sigmóide direto ou ponderado, a eliminação das redundâncias do sinal, ou seja, eliminação dos coeficientes pouco significativos no domínio wavelet, uma vez que a ausência dos mesmos não afetaria a reconstrução do sinal (DONOHO e JOHNSTONE, 1994). Entretanto, isto evita que estes coeficientes influenciem na estimativa do perfil de ruído e, principalmente, que se processem coeficientes redundantes. Isto minimiza significantemente o ruído residual. Após a eliminação das redundâncias do sinal, que pode ser feita aplicando o método de limiar duro proposto por DONOHO e JOHNSTONE (1994), o método LS é aplicado para a redução efetiva do ruído. Desta forma, quando se faz a opção da pré-filtragem, tem-se, na realidade, o uso de um limiar misto: Limiar duro mais limiar sigmóide.

O limiar duro foi escolhido para a pré-filtragem, pois o objetivo é apenas eliminar as redundâncias do sinal de voz. Caso se escolhesse, por exemplo, o limiar suave os coeficientes que não são considerados redundantes também sofreriam alterações.

CAPÍTULO 6

RESULTADOS

Introdução

Neste capítulo são apresentados os resultados das implementações do método proposto e suas variantes. Estes resultados são comparados com o método de melhoria em sinais de voz usando a transformada wavelet, proposto por SHEIKHZADEH e ABUTALEBI (2001).

Todos os métodos, inclusive o método apresentado em SHEIKHZADEH e ABUTALEBI (2001) foram implementados da forma proposta originalmente e também usando a pré-filtragem proposta neste trabalho.

Para verificar a qualidade dos sinais processados, foram avaliados os níveis de redução de ruído e de distorção usando medidas objetivas e subjetivas.

Para a apresentação dos resultados, serão usadas as seguintes siglas:

- SMW – Método apresentado em SHEIKHZADEH e ABUTALEBI (2001)
- PSMW - Método apresentado em SHEIKHZADEH e ABUTALEBI (2001) com pré-filtragem
- LSD - Limiar sigmóide direto
- LSP - Limiar sigmóide ponderado
- PLSD - Limiar sigmóide direto com pré-filtragem
- PLSP - Limiar sigmóide ponderado com pré-filtragem

A transformada wavelet discreta foi aplicada usando a wavelet db10. Além do estudo apresentado na seção 3.1, para a escolha desta wavelet foi feito, ainda, o

processamento de um dos sinais-teste, sinal A apresentado na seção 6.1, contaminado por ruído branco. Algumas wavelets de Daubechies foram testadas para verificar qual seria a melhor, levando em conta o nível de redução de ruído (relação sinal/ruído), distorção (distância cepstral média) e o tempo de processamento em segundos. O método usado foi o PLSP. Os resultados destes testes são apresentados na tabela 6.1.

Tabela 6.1 – SNR, Distância Cepstral e tempo de processamento para algumas wavelets de Daubechies.

	db2(Haar)	db10	db14	db16	db18	db20
SNR	34,7192	38,8490	34,7618	37,7510	33,3964	33,4904
Distância Cepstral	1,5011	1,3528	1,6243	1,6571	1,7390	1,6842
Tempo(s)	4,4680	5,5160	5,5160	4,8750	4,5470	4,5470

O sinal limpo usado no teste tem SNR inicial de 41,9595dB e o sinal ruidoso tem SNR de 11,7335dB. A SNR é calculada de acordo com a equação (6.1), seu cálculo é feito tomando a razão entre um segmento de voz e um segmento de silêncio de cada sinal, conforme (DELLER et al., 1992):

$$SNR = 10\log_{10}\left(\frac{\sum_{i=1}^{N} x^2[i]}{\sum_{i=1}^{N} r^2[i]}\right) \tag{6.1}$$

sendo N o número de amostras do segmento escolhido, x as amostras no trecho de voz e r as amostras no trecho de silêncio. Neste trabalho, baseando-se nos sinais utilizados, adotou-se N=*5000*. O valor de SNR apresentado nas tabelas é uma média dos valores de SNR calculados para vários segmentos do sinal de voz.

Se após o processamento, a SNR do sinal for baixa em relação a SNR do sinal limpo, significa que houve pouca redução de ruído. Se, após o processamento a SNR do sinal processado for muito maior que a SNR do sinal limpo, significa que houve grande redução de ruído e, possivelmente, o sinal deve estar bastante distorcido. Por isso, após o processamento, sinais que têm SNR próxima da SNR do sinal original são mais atrativos.

A distância cepstral indica o nível de distorção do sinal processado em relação ao sinal original e quanto menor for o seu valor menor será o nível de distorção.

Assim, embora não apresente o menor tempo de processamento, conforme se verifica na tabela 6.1, o sinal processado com a db10 apresenta SNR mais próxima da SNR do sinal original e menor distância cepstral média. Por isso, a db10 foi escolhida para o processamento dos sinais processados neste trabalho. Além disso, como se verificou na seção 4.2, a db10 preserva a distribuição de energia do sinal de acordo com o trecho de voz analisado.

A aplicação do limiar sigmóide foi feita usando $\gamma = 1$. Neste caso, para a escolha, também se considerou a SNR e a distância cepstral média. A tabela 6.2 apresenta os valores de SNR e distância cepstral para vários valores de γ, para o sinal A contaminado por ruído branco, também processado pelo PLSP. Verifica-se que a melhor SNR e a menor distância cepstral média ocorrem quando $\gamma = 1$, daí, a opção por este valor.

Tabela 6.2 – SNR e Distância Cepstral para diferentes inclinações da sigmóide.

	γ=0,25	γ=0,5	γ=0,75	γ=1,0	γ=1,25	γ=1,5	γ=1,75	γ=2,0
SNR	37,8266	37,8205	37,8072	38,8490	37.7648	37,7359	37,7021	37,6633
D Cep.	1,4575	1,4446	1,4345	1,3528	1,4176	1,4100	1,4027	1,3955

Para o método LSD, foi usado β=*5,5* para aplicação do limiar no domínio wavelet. Este valor foi obtido após várias simulações, onde se buscou um melhor compromisso entre os níveis de redução de ruído e distorção no sinal processado. Para isto, o valor *3,0* foi tomado como valor inicial para β, este é o valor nos algoritmos de redução de ruído desenvolvidos pela equipe de processamento digital de sinais da *Rice University* (http://www.dsp.rice.edu).

6.1 Sinais e Resultados

Para implementação dos métodos e apresentação dos resultados foram usados três sinais de voz, aqui referenciados como A, B e C, gravados por voz masculina, sendo dois em português e um em inglês. As frases dos sinais são as seguintes:

Sinal A - *"FOI DETECTADO UM PROBLEMA EM SEU CARTÃO, ELE DEVE SER SUBSTITUÍDO"*;

Sinal B - *"À MEDIDA QUE O TEMPO PASSA, MAIS NOS CONVENCEMOS DA EFICÁCIA DO CONVÊNIO";*

Sinal C - *"GOOD SERVICE SHIP YOU ORDER BY BIG TIPS".*

Todos os sinais foram amostrados a 16kHZ e quantizados com 16 bits por amostra. Nos testes, foram usados arquivos no formato *wave*, que apresenta a amplitude normalizada no intervalo (-1,1). Os segmentos de voz foram obtidos usando-se uma janela de *Hanning* com sobreposição de 50% entre os segmentos. O comprimento das janelas foi de 256 amostras, que no domínio wavelet, permite a obtenção de até 8 níveis na decomposição do sinal.

6.1.1 Avaliações Objetivas

Para os testes dos algoritmos, os sinais foram contaminados por dois tipos de ruído: ruído Gaussiano branco e ruído colorido (ruído de carro).

Para verificar o nível de redução de ruído nos sinais processados pelos métodos implementados, a SNR foi calculada para o sinal limpo, antes da contaminação por ruído, para o sinal ruidoso e para o sinal processado.

Além da SNR, foram usadas mais duas medidas objetivas para avaliar os sinais processados: a distância cepstral, que mede o nível de distorção entre os sinais original e processado, baseado nos espectros dos dois sinais e a PESQ, que faz uma avaliação objetiva a partir dos sinais original e processado, baseando-se num modelo psicoacústico da audição humana .

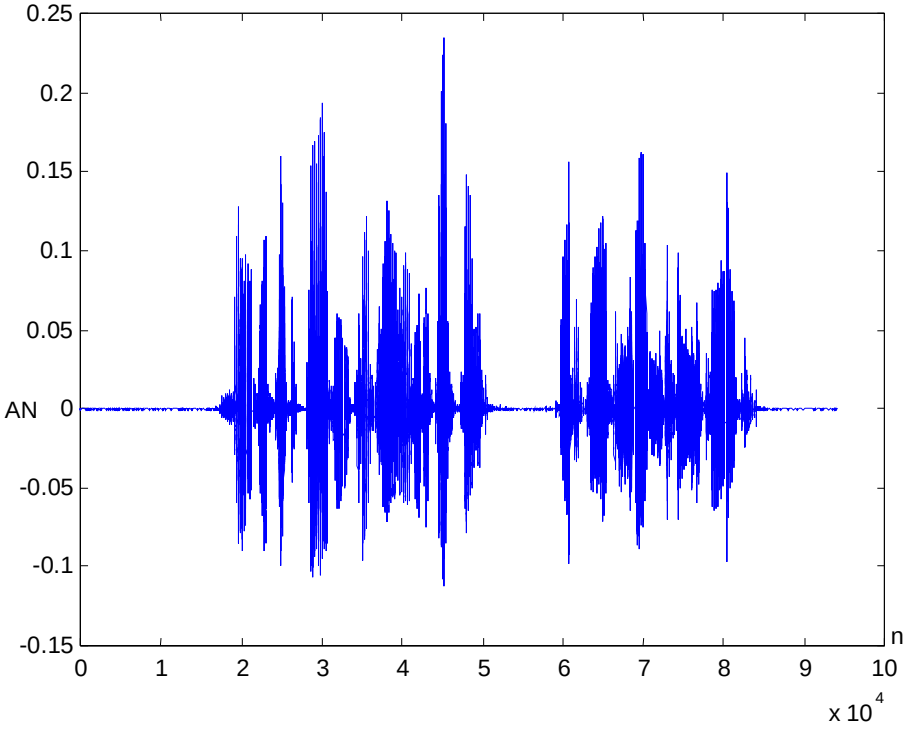
Na tabela 6.3 são apresentadas as SNRs, em dB, dos sinais limpos e contaminados pelos ruídos mencionados acima. Percebe-se que o nível de ruído, tanto branco quanto colorido, é forte, o que causa degradações significativas das faixas de voz dos sinais.

Nas figuras 6.1, 6.2 e 6.3 são apresentadas, respectivamente, as formas de onda dos sinais A, B e C e os mesmos sinais contaminados por ruído branco e por ruído colorido, sendo n o número de amostras e AN a amplitude numérica do sinal.

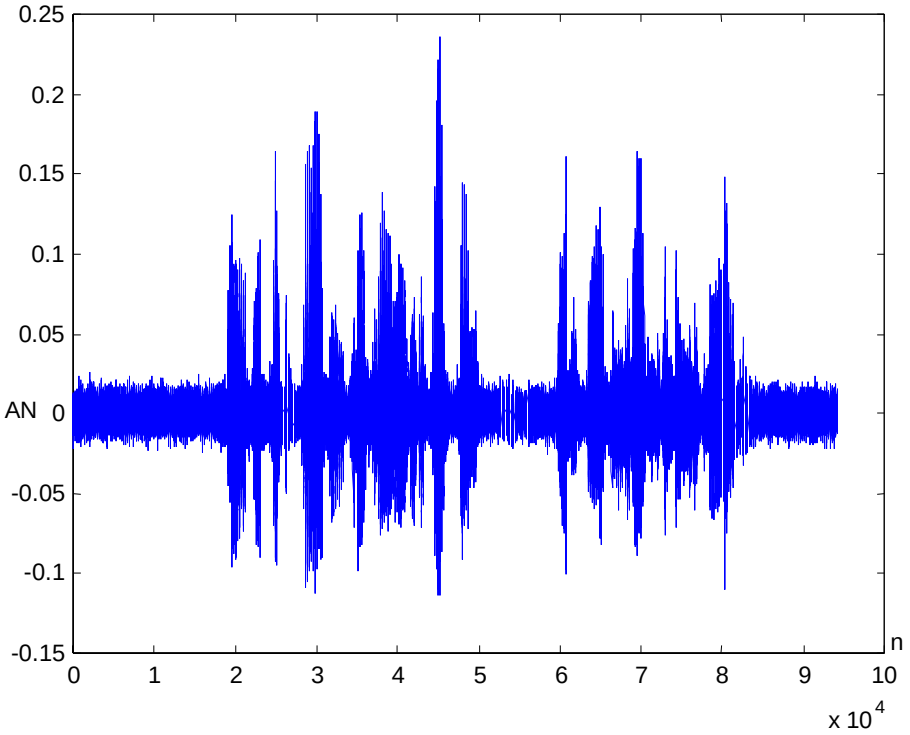
Tabela 6.3 – SNR dos sinais limpos e contaminados.

	Sinal limpo	Sinal com ruído branco	Sinal com ruído colorido
Sinal A	41,9595	11,7335	10,5845
Sinal B	36,8879	11,7442	10,5803
Sinal C	31,3989	16,8460	15,6357

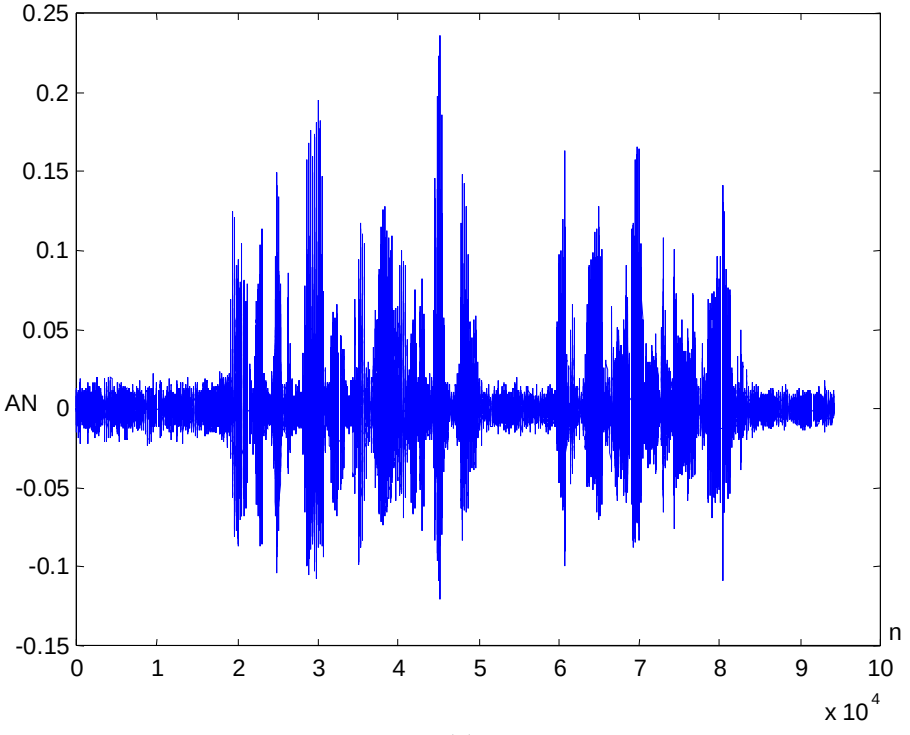
As tabelas 6.4 e 6.5 apresentam, respectivamente, as SNRs dos sinais processados com ruído branco e com ruído colorido pelo método proposto neste trabalho, suas variantes e pelo método de SMW.



(a)

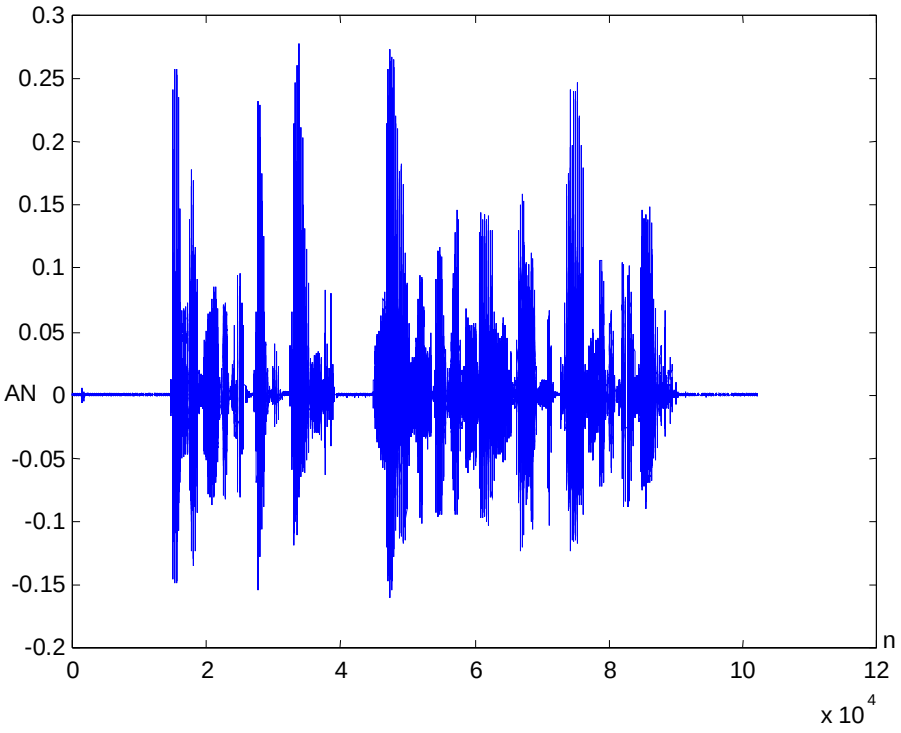


(b)

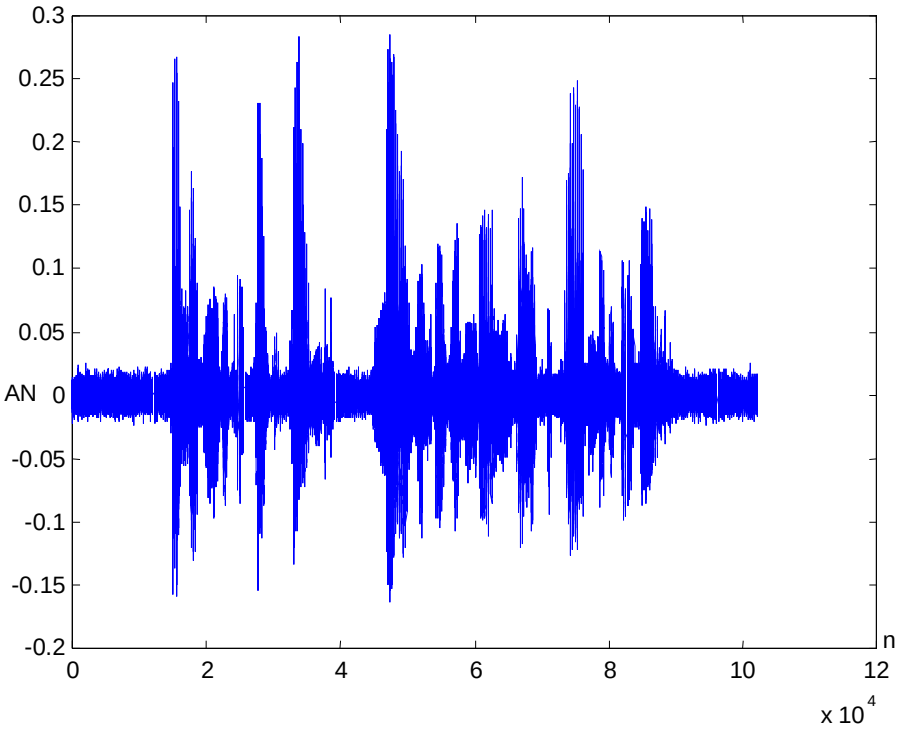


(c)

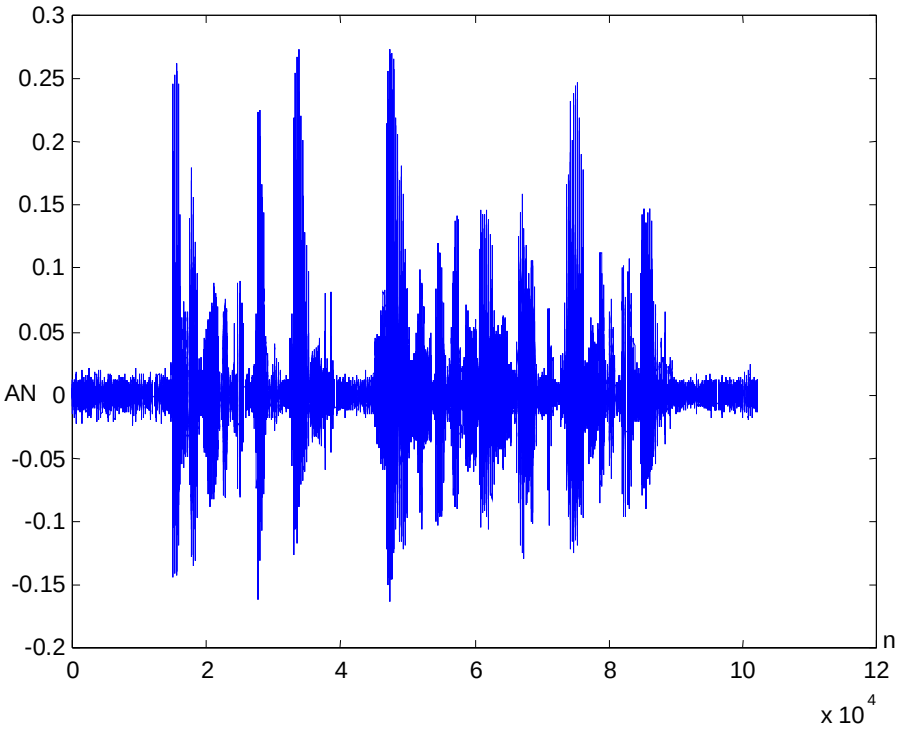
Figura 6.1 - (a) forma de onda do sinal A; (b) sinal A com ruído branco; (c) sinal A com ruído colorido.


0.3
0.25
0.2
0.15
0.1
0.05
AN 0
-0.05
-0.1
-0.15
-0.2
0
2
4
6
8
10
12
n
x 10^4

(a)

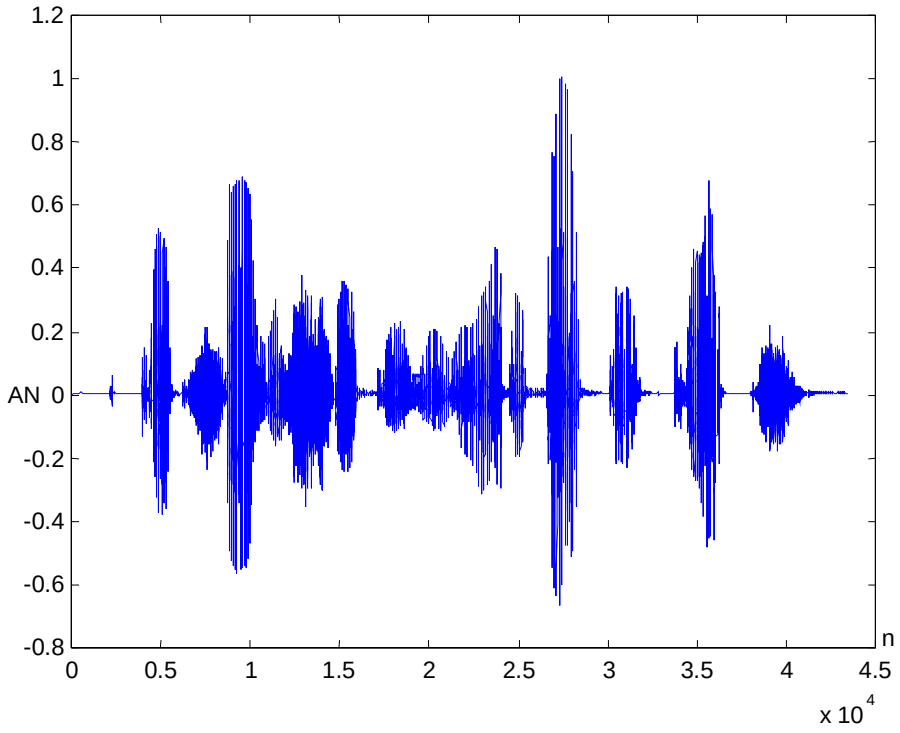

0.3
0.25
0.2
0.15
0.1
0.05
AN 0
-0.05
-0.1
-0.15
-0.2
0
2
4
6
8
10
12
n
x 10^4

(b)

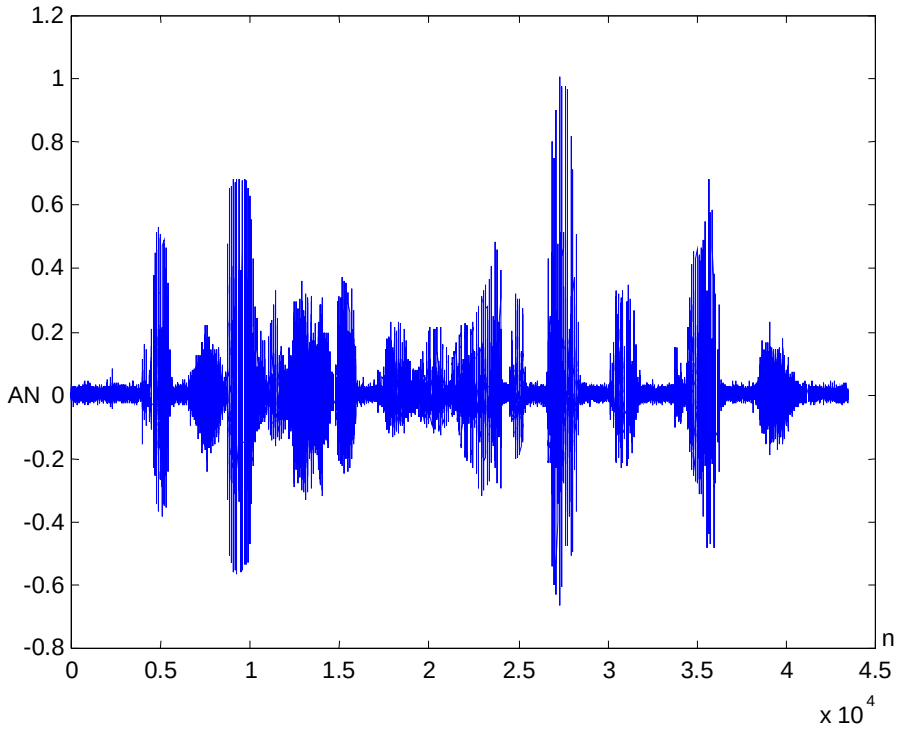


(c)

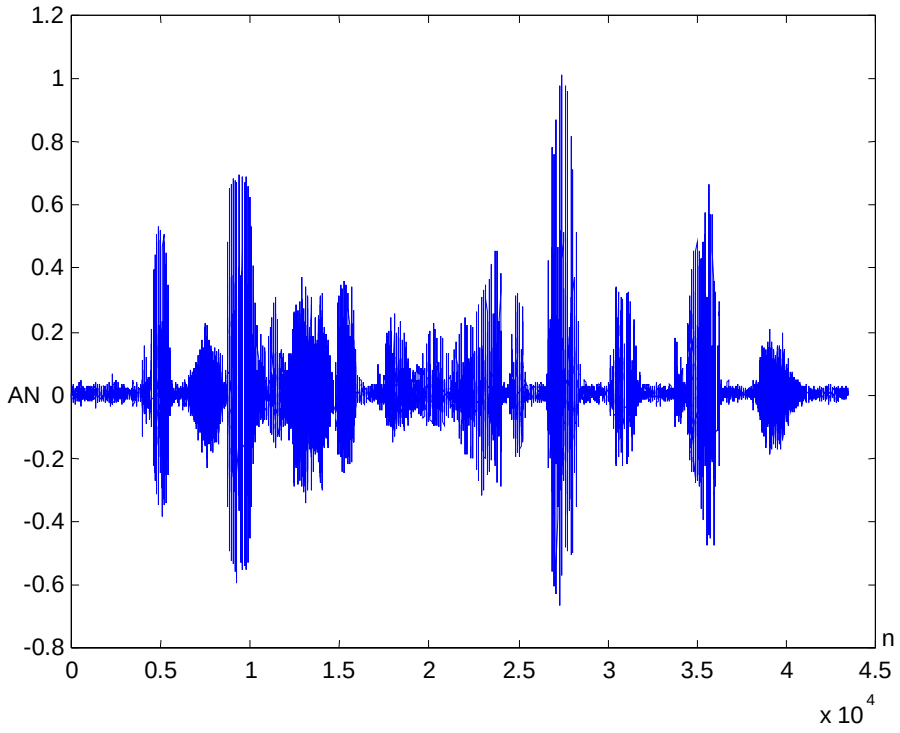
Figura 6.2 - (a) forma de onda do sinal B; (b) sinal B com ruído branco; (c) sinal B com ruído colorido.


1.2
1
0.8
0.6
0.4
0.2
AN 0
-0.2
-0.4
-0.6
-0.8
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
n
x 10^4

(a)


1.2
1
0.8
0.6
0.4
0.2
AN 0
-0.2
-0.4
-0.6
-0.8
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
n
x 10^4

(b)



(c)

Figura 6.3 - (a) forma de onda do sinal C; (b) sinal C com ruído branco; (c) sinal C com ruído colorido.

Tabela 6.4 – SNR dos sinais processados com ruído branco.

	SMW	PSMW	LSD	PLSD	LSP	PLSP
Sinal A	77,6360	77,6231	34,4016	37,5589	37,8057	38,8490
Sinal B	65,7640	64,7430	36,5582	36,5633	37,0713	37,0774
Sinal C	31,8234	31,6455	32,4300	32,4549	32,7665	32,7929

Tabela 6.5 – SNR dos sinais processados com ruído colorido.

	SMW	PSMW	LSD	PLSD	LSP	PLSP
Sinal A	15,0335	14,9501	44,9677	44,9480	46,8437	46,8269
Sinal B	15,3062	15,2225	40,4749	40,4528	43,1145	43,0994
Sinal C	18,4194	18,4078	31,4371	32,4606	32,0344	32,0586

Para a apresentação das distâncias cepstrais dos sinais processados, foi calculada a média da distância cepstral de cada sinal para verificar as distorções apenas nos trechos de voz e depois tomando também os trechos de silêncio.

Na tabela 6.6, são apresentadas as médias das distâncias nos trechos de voz, para os sinais processados com ruído branco e na tabela 6.7, são apresentadas as médias das distâncias cepstrais para estes mesmos sinais, porém tomando também os trechos de silêncio.

Tabela 6.6 – Distância cepstral nos trechos de voz dos sinais processados com ruído branco.

	SMW	PSMW	LSD	PLSD	LSP	PLSP
Sinal A	1,5592	1,5587	1,4362	1,4494	1,4199	1,4190
Sinal B	1,6322	1,6354	1,4217	1,4244	1,3976	1,3958
Sinal C	1,1965	1,1961	1,0702	1,0700	1,0995	1,0756

Tabela 6.7 – Distância cepstral total dos sinais processados com ruído branco.

	SMW	PSMW	LSD	PLSD	LSP	PLSP
Sinal A	1,4090	1,4086	1,3426	1,3495	1,3604	1,3528
Sinal B	1,5182	1,5204	1,3532	1,3531	1,3760	1,3728
Sinal C	1,2062	1,2064	0,9958	0,9958	0,9344	0,9338

Nas tabelas 6.8 e 6.9 os resultados são apresentados, respectivamente, da mesma maneira que nas tabelas 6.6 e 6.7, porém, os resultados são obtidos considerando-se sinais contaminados por ruído colorido.

Tabela 6.8 – Distância cepstral nos trechos de voz dos sinais processados com ruído colorido.

	SMW	PSMW	LSD	PLSD	LSP	PLSP
Sinal A	1,5763	1,5780	0,8098	0,8106	1,0355	1,0334
Sinal B	1,6386	1,6377	0,9585	0,9589	1,1736	1,1735
Sinal C	1,3236	1,3235	0,7577	0,7579	0,8433	0,8424

Tabela 6.9 – Distância cepstral total dos sinais processados com ruído colorido.

	SMW	PSMW	LSD	PLSD	LSP	PLSP
Sinal A	1,9317	1,9386	1,2814	1,2793	1,4509	1,4477
Sinal B	1,8222	1,8162	1,2115	1,2113	1,4170	1,4166
Sinal C	1,4947	1,1944	0,9348	0,9348	0,9506	0,9482

Nas tabelas 6.10 e 6.11 são apresentados, respectivamente, os resultados da PESQ para os sinais processados com ruído branco e para os sinais processados com ruído colorido. Conforme apresentado na seção 2.2.2.2, a maior nota que a PESQ atribui a um sinal é 4,5 e sinais que recebem nota a partir de 3 são considerados de boa qualidade.

Tabela 6.10 - PESQ dos sinais processados com ruído branco.

	SMW	PSMW	LSD	PLSD	LSP	PLSP
Sinal A	1,738	1,738	2,969	2,969	3,175	3,199
Sinal B	2,083	2,085	3,153	3,155	3,380	3,380
Sinal C	2,584	2,586	4,136	4,145	4,359	4,376

Tabela 6.11 -PESQ dos sinais processados com ruído colorido.

	SMW	PSMW	LSD	PLSD	LSP	PLSP
Sinal A	1,909	1,909	2,883	2,884	3,027	3,027
Sinal B	2,169	2,171	3,043	3,044	3,273	3,273
Sinal C	2,031	2,031	3,478	3,480	3,737	3,737

6.1.2 Avaliações Subjetivas

Além das avaliações feitas pela SNR e pela distância cepstral, foram feitas avaliações subjetivas informais, nas quais os sinais processados foram apresentados para um grupo de ouvintes para as seguintes verificações:

1) Se o ouvinte prefere o sinal ruidoso ou o sinal processado;
2) Dizer, dentre os sinais processados pelos métodos implementados, qual é o melhor, para ruído branco e para ruído colorido.

O grupo de ouvintes foi composto por 10 pessoas com idades entre 20 e 40 anos, sendo 5 mulheres e 5 homens.

Para a melhor concentração dos ouvintes e ausência de interferência do ambiente, foram usados fones de ouvidos para a audição dos sinais.

Na primeira verificação os sinais eram apresentados em pares, sinal ruidoso e sinal processado, e o ouvinte respondia qual deles é o melhor. Cada par de sentenças era tocado duas vezes em ordem trocada e em nenhuma delas o ouvinte era avisado qual sentença seria apresentada primeiro, para evitar favorecimento a um dos sinais.

Na segunda verificação, onde o ouvinte fazia comparações entre os sinais processados, primeiro eram apresentados três pares de sentenças para o ouvinte e de cada par ele escolhia a melhor. As três melhores sentenças eram de novo apresentadas ao ouvinte e, dessas, ele elegia a melhor. Nesta verificação, nas duas etapas, as sentenças também eram tocadas duas vezes aleatoriamente.

Tanto na primeira quanto na segunda verificação, sempre que o ouvinte apresentava dúvidas, o processo era repetido. Esta repetição era feita tantas vezes fossem necessárias para que o ouvinte se decidisse pela melhor sentença.

Os resultados são apresentados nas tabelas 6.12 a 6.15.

Na tabela 6.12 são apresentados os percentuais de preferência dos ouvintes pelos sinais processados com ruído branco, quando comparados com o sinal ruidoso. Na tabela 6.13 é feita a mesma avaliação, porém para os sinais contaminados por ruído colorido.

Tabela 6.12 – Preferência (em %) pelos sinais processados com ruído branco, quando comparados com o sinal ruidoso

	SMW	PSMW	LSD	PLSD	LSP	PLSP
Sinal A	50	60	80	80	100	100
Sinal B	40	50	80	90	100	100
Sinal C	70	80	100	100	100	100

Tabela 6.13 – Preferência (em %) pelos sinais processados com ruído colorido, quando comparados com o sinal ruidoso

	SMW	PSMW	LSD	PLSD	LSP	PLSP
Sinal A	0	0	60	70	100	100
Sinal B	0	0	70	90	100	100
Sinal C	0	0	100	100	100	100

Na tabela 6.14 são apresentados os percentuais de preferência dos ouvintes pelos sinais processados com ruído branco, quando os métodos são comparados entre si. Na tabela 6.15 é feita a mesma avaliação, porém para os sinais contaminados por ruído colorido

Tabela 6.14 – Preferência (em %) entre os sinais processados com ruído branco, comparados entre si.

	SMW	PSMW	LSD	PLSD	LSP	PLSP
Sinal A	0	0	0	0	20	80
Sinal B	0	0	0	0	10	90
Sinal C	0	0	0	0	20	80

Tabela 6.15 – Preferência (em %) entre os sinais processados com ruído colorido, comparados entre si.

	SMW	PSMW	LSD	PLSD	LSP	PLSP
Sinal A	0	0	0	0	30	70
Sinal B	0	0	0	0	20	80
Sinal C	0	0	0	0	10	90

6.2 Conclusões

Embora a SNR não seja uma medida para se avaliar a inteligibilidade do sinal processado, mas sim o nível de redução de ruído, ela serve para dar uma noção da qualidade do sinal processado, pois após a redução do ruído aditivo, é de se esperar que a SNR deste sinal seja algo próximo da SNR do sinal limpo (sinal de referência). Os resultados apresentados nas tabelas 6.4 e 6.5 mostram que todos os sinais processados pelo

método proposto têm SNR próxima da do sinal limpo, enquanto que os sinais processados com o método SMW apresentam uma redução de ruído ou muito alta ou muito baixa, com exceção do sinal C quando contaminado por ruído branco, o que de qualquer forma acaba deteriorando muito os trechos de voz destes sinais.

As distâncias cepstrais médias, apresentadas nas tabelas 6.6 a 6.9, mostram que o método proposto e suas variantes apresentam, em média, menos distorções que o método SMW, pois apresentam valores menores, independente de se considerar trechos de silêncio e voz juntos ou apenas trechos de voz. Para os sinais com ruído colorido, tanto na tabela 6.8, onde são apresentadas as distâncias cepstrais médias considerando apenas os trechos de voz dos sinais processados, quanto na tabela 6.9, onde são apresentadas as distâncias cepstrais médias dos trechos de voz e silêncio, verifica-se que o método proposto e suas variantes são superiores em todos os três sinais testados ao método SMW, pois também apresentam valores menores.

Os resultados da PESQ mostram que o método proposto e suas variantes são eficazes na redução de ruído, pois, são considerados de boa qualidade sinais que têm nota PESQ a partir de 3. Nos resultados apresentados nas tabelas 6.10 e 6.11, sinais contaminados por ruído branco e por ruído colorido, respectivamente, apenas o LSD e o PLSD apresentam notas inferiores a 3 para o sinal A, já o SMW tem notas inferiores a 3 para todos os sinais. Além destes testes foi feita uma comparação com o método apresentado por QIANG e WAN (2003) onde o sinal C é contaminado por ruído branco com forte nível de distorção e apresenta SNR de 2,37dB e, após processado apresenta nota PESQ igual a 0,636, o processamento deste sinal com a mesma SNR pelo método PLSP apresenta nota PESQ igual a 2,690. O sinais de referência e processado usados por QIANG e WAN (2003) estão disponíveis no endereço eletrônico http://cslu.cse.ogi.edu/research/nsel.htm.

As avaliações subjetivas apresentadas na seção 6.1.2 comprovam que o método proposto e suas variantes são melhores do que o método SMW, pois, quando os sinais processados foram comparados com os sinais ruidosos a preferência por um dos métodos propostos para sinais contaminados por ruído branco foi no mínimo de 80% enquanto que para o método SMW a preferência foi de no máximo 80% para o sinal C, quando contaminado por ruído branco, sendo que para os outros métodos este sinal teve 100% de preferência.

Para os sinais contaminados por ruído colorido, o método proposto e suas variantes foram superiores ao método SMW, o que já havia sido verificado nas avaliações objetivas. Conforme comentado por alguns ouvintes, os sinais processados com ruído colorido pelo método SMW ficam muito piores que os sinais ruidosos.

Quando os sinais são comparados entre si, tanto com ruído branco quanto com ruído colorido, a preferência dos ouvintes, conforme apresentado nas tabelas 6.14 e 6.15, é unânime pelo método PLSP. Embora não tenha sido apresentada esta comparação, quando os sinais processados pelo método LSP foram retirados da amostra e foram comparados apenas o método SMW e o LSD, a preferência dos ouvintes foi unânime pelo método LSD.

A pré-filtragem, que foi implementada para todos os métodos, embora não tenha se destacado nos testes objetivos, mostrou-se relevante nos testes subjetivos, onde, de acordo com as tabelas 6.14 e 6.15, verifica-se a preferência dos ouvintes pelos sinais processados com pré-filtragem.

CAPÍTULO 7

CONCLUSÕES

Neste trabalho foram estudadas técnicas de redução de ruído em sinais de voz com o objetivo de propor um método eficiente baseado na transformada wavelet discreta, principalmente para o processamento de sinais contaminados por ruído colorido.

Por isso, foram estudadas as características básicas dos métodos de redução de ruído em sinais de voz, a teoria da análise wavelet e alguns métodos de redução de ruído baseados em wavelets, além de um estudo detalhado sobre as melhores wavelets para processamento de sinais de voz.

A transformada wavelet discreta é uma ferramenta potente para o processamento de sinais de voz, tanto para compressão e codificação, quanto para melhoria do sinal. Quando se trata de melhoria do sinal de voz, o problema é determinar o melhor processo de filtragem para reduzir o ruído de fundo presente no sinal sem deteriorar os trechos de voz. Os métodos de corte por limiar são os mais comuns nas técnicas baseadas em wavelets, sendo que o maior problema é determinar a melhor forma de se calcular e aplicar o limiar.

O método proposto neste trabalho é também um método de atenuação por limiar. Porém, o limiar é calculado e aplicado de maneira diferente de todos os métodos até então propostos na literatura especializada, e resulta numa melhoria significativa dos sinais processados quando comparados com outros métodos.

Além de propor um novo método para redução de ruído, foi apresentado também um algoritmo simples e eficiente para detecção silêncio/voz baseado na transformada wavelet discreta, muito útil nas aplicações práticas envolvendo sinais de voz. Este algoritmo se mostrou eficiente tanto para sinais contaminados por ruído branco quanto para sinais contaminados por ruído colorido.

O método proposto neste trabalho foi testado para três sinais de voz diferentes e comparado com um método clássico da literatura especializada que considera todos os passos que devem ser seguidos para o bom funcionamento de um algoritmo de redução de ruído em sinais de voz. Para a avaliação destes métodos foram usados testes objetivos e subjetivos.

Sob todos os aspectos testados, o método proposto se mostrou melhor para os sinais processados, inclusive para os sinais contaminados por ruído colorido, que é o tipo de ruído mais comum em sinais reais. O método proposto reduz mais ruído e introduz menos distorções nos sinais processados.

Subjetivamente, os testes mostraram que este método é muito melhor do que o método clássico da literatura, usado para comparação neste trabalho, principalmente para sinais contaminados por ruído colorido.

Em geral, o método proposto apresentou as seguintes vantagens:

- O sinal é pré-filtrado, fazendo uma eliminação de redundâncias, antes da redução efetiva do ruído, o que faz com que somente os coeficientes significativos no domínio wavelet influenciem na estimação do perfil do ruído;
- O algoritmo de detecção silêncio / voz funciona já no sinal ruidoso, sem a necessidade de um prévio conhecimento do sinal limpo, que facilitará futuras implementações práticas do mesmo;
- O limiar é o obtido a partir da média dos limiares calculados para todas as janelas do último trecho de silêncio e é aplicado por faixa de freqüência. Os métodos tradicionais também aplicam o limiar por faixa de freqüência, mas o mesmo é calculado usando-se somente a última janela do último trecho de silêncio;
- Nos trechos de voz, para cada janela e para cada faixa de freqüência, também é calculado um limiar. Usa-se como limiar um valor que é a razão entre o limiar calculado para aquela faixa de freqüência na voz e o limiar calculado para aquela faixa de freqüência no último trecho de silêncio. Isto faz com que características não-estacionárias do ruído sejam consideradas, além de evitar

ter que fazer a detecção *voiced/unvoiced* do sinal, pois este valor aplicado já considera a energia em cada faixa de freqüência.

Com base nos resultados apresentados no capítulo 6, conclui-se que o método proposto neste trabalho é eficiente na redução de ruído em sinais de voz, na redução de ruído branco ou de ruído colorido. Isto significa baixa distorção no sinal processado e uma redução de ruído suficiente para melhorar a inteligibilidade do sinal.

A redução de ruído em sinais de voz baseada em wavelets continua sendo um campo amplo para estudos e, como sugestões para trabalhos futuros, pode-se citar:

- Apresentação de novos métodos de filtragem que não sejam baseados em corte por limiar;
- Estudo de wavelets que melhor se adaptem às características do sinal de voz;
- Usar a propriedade da linearidade da transformada wavelet para se fazer uma estimação a priori de ruído.

REFERÊNCIAS BIBLIOGRÁFICAS

AGBINYA, J. I., *Discrete Wavelet Transform Techniques in Speech Processing,* IEEE TENCON-Digital Signal Processing Applications, vol. 2, pp. 514-519, November 1996.

AGUIAR, B. G., Melhoramento de Voz Degradada por Método Baseado em Subtração Espectral Adaptativa, *Anais do 7º Simpósio Brasileiro de Telecomunicações*, Florianópolis, pp. 54-59, setembro de 1989.

BAHOURA, M. and ROUAT, J., *Wavelet Noise Reduction: Application to Speech Enhancement*, Canadian Acoustics, vol. 28, No 3, pp 158-159, setembro de 2000.

BAHOURA, M. and ROUAT, J., *Wavelet Speech Enhancement Based on the Teager Energy Operator*, IEEE SPL, Vol. 8, pp. 10-12, January 2001 (A).

BAHOURA, M. and ROUAT, J., A New Approach for Wavelet Speech Enhancement, *EUROSPEECH 2001*, pp. 1937-1940, September 2001 (B).

BEERENDS, J. G., HEKSTRA, A. P., RIX, A. W. and HOLLIER, M. P., *Perceptual Evaluation of Speech Quality (PESQ) The New ITU Standard for end-to-end Speech Quality Assessment Part II – PSYCHOACOUSTIC MODEL,* AES Journal, Vol. 50, No. 10, pp. 765-778, October 2002.

BENEDETO, J. J. and FRAZIER, M. W., Wavelets: Mathematics and Applications, *John J., N. W. CRC Press Inc.* Boca Raton, 1993.

BIJAOUI, A., *Wavelets, Gaussian Mixtures and Wiener Filtering*, Elsevier Science, Signal Processing, Vol. 82, No. 4, pp 709-712, 2002.

BLACK, M. and ZEYTINOGLU, M., Computationally Efficient Wavelet Packet Coding of Wide-Band Stereo Audio Signals, *ICASSP-95*, pp: 3075-3078, 1995.

BOLL, S.F. *Suppression of Acoustic Noise in Speech using Spectral Subtraction*, IEEE Trans. Acoust. Speech and Signal Processing (ASSP), vol. 29, pp. 113-120, April 1979.

CAPPÉ, O., *Elimination of Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor*, IEEE Transactions on Acoustic, Speech and Signal Processing, pp: 345-349, April 1994.

CHUI, C. K., An introduction to wavelets, *Texas A&M University*, 1992.

COIFMAN, R., MEYER Y., QUAKE, S. and WICKERHAUSER, M. V., Signal Processing and Compression with Wavelet Packets, *Wavelet analysis and applications (Toulouse, 1992)*, pp. 77-93. Frontières, Gif, 1993.

COIFMAN, R., Adapted Multiresolution Analysis, Computation, Signal Processing and Operator Theory, *ICM 90, Springer-Verlag*, pp: 879-887, 1990.

COIFMAN, R. and WICKERHAUSER M. V., Wavelets and Adapted Waveform Analysis: A Toolkit for Signal Processing and Numerical Analysis, *A.K. Peters Ltd*., MA, 1993.

CONNOR, F. R., Noise, *Edward Arnold*, 1973.

DAE-SUNG, K., JUNG-GO, C., GEUN-TAEK, R., MUN-SEOB, B. and HYEON-DEOK, B., Noise Reduction Using Coefficients Smoothing Wavelet Transform, *Proceedings of the International Conference on Signal Processing, Application and Technology (ICSPAT 2002)*, Vol. I, pp. 029-033, San Jose, September 2002.

DAUBECHIES, I., Othonormal Bases of Compactly Supported Wavelets, *Communications on Pure and Applied Mathematics XLI*, pp. 909-996,1998.

DAUBECHIES, I., *The Wavelet Transform, Time Frequency Localization and Signal Analysis*, IEEE Trans. Inf. Theory, Vol. 36,pp. 961-1005, 1990.

DAUBECHIES, I., Ten Lectures on Wavelets, *SIAM Book*s, Philadelphia, 1992.

DELLER, J.L., PROAKIS, J.G and HANSEN, J. H. L, Discrete-Time Processing of Speech Signals, *Macmillam*, New York, 1993.

DONOHO, D. L. and JOHNSTONE, I. M., *Ideal Spatial Adaptation via Wavelet Shrinkage*, Biometrika, Vol. 81: No. 3, pp. 425-455, 1994.

DONOHO, D. L., *De-Noising by Soft-thresholding,* IEEE Transactions on Information Theory, vol. 41, N 3, pp. 613-627, maio de 1995.

DUARTE, M. A. Q., VILLARREAL, F. e DÍAZ, L. A., Sobre Análise Wavelet Biortogonal, *55º Seminário Brasileiro de Análise*, pp. 357-365, Uberlândia MG, maio de 2002.

DUARTE, M. A. Q., OLIVEIRA, L. C. O., VILLARREAL, F. e DÍAZ, L. A., *Compressão de Sinais Elétricos Usando A Transformada de Wavelet*, TEMA – Tendências em Matemática Aplicada e Computacional, vol. 4, No 1, pp. 21-30, 2003.

DUARTE, M. A. Q., VIEIRA FILHO, J. e VILLARREAL, F., Implementação e Avaliação de Técnicas de Redução de Ruído em Sinais de Voz Baseadas em Wavelets, *XV Congresso Brasileiro de Automática, CBA 2004*, 6p.(CD), Gramado RS, setembro de 2004(A).

DUARTE, M. A. Q., VIEIRA FILHO, J. e VILLARREAL, F., *Um Novo Método de Redução de Ruído em Sinais de Voz Baseado em Wavelets*, *XXI Simpósio Brasileiro de Telecomunicações, SBT 2004*, 5p.(CD), Belém PA, setembro de 2004(B).

EPHRAIM, Y, *Statistical-Model-Based Speech Enhancement Systems*, IEEE Trans. On Acoust., Speech and Signal Processing, Vol. 80, No. 10, p.1525-1555, October 1992.

EPHRAIM, Y. and MALAH, D., *Speech Enhancement Using Minimum Mean Square Error Short-Time Spectral Amplitude Estimator,* IEEE Trans. On Acoust., Speech and Signal Processing, Vol. 32, No. 6, pp.1109-1121, October 1984.

FLANAGAN, J. L., Speech Analysis Synthesis and Perception, *Springer-Verlag*, 1972.

GAZOR, S. and SHANG, W., *A Soft Voice Activity Detector Based on a Laplacian-Gaussian Model*, IEEE Transactions on Speech and Audio Processing, vol. 11, No. 5, pp. 498-505, September 2003.

GOMES, J., VELHO, L. e GOLDSTEIN, S., Wavelets: Teoria, Software e Aplicações, *IMPA*, 1997.

GOMES, S. M., DOMINGUES, M. O. e KAIBARA, M. K., Navegando de Fourier a Wavelets, *I Escola Brasileira de Aplicações em Dinâmica e Controle*, Uberlândia – MG – 2001.

GUIDORIZZI, H. L., Um Curso de Cálculo, vol. 1, 2 ed., *LTC LTDA*, São Paulo, 1987.

GRAY Jr., A. H. and MARKEL, J. D., *Distance Measures for Speech Processing,* IEEE Trans. On Acoustics, Speech and Signal Processing, Vol. 24, No. 5 - October 1976.

HAAR, A., *Zur Theorie der Orthogonalen Functionen-Systeme*, Mathematische Annalen, 69, pp. 331-371, 1910.

HÄNDEL, P., Low Distortion Spectral Subtraction for Speech Enhancement, *IV EUROSPEECH, European Conference on Speech Communication and Technology*, pp. 1549-1552, Madrid, September 1995.

HÖNIG, S. C., A Integral de Lebesgue e suas Aplicações, *11º Colóquio Brasileiro de Matemática (Editora IMPA)*, Poços de Calda, Julho de 1977.

HOWARD, D. M. and ANGUS, J., Acoustics and Psychoacoustics, *Focal Press*, 1996.

HU, Y,. and LOIZOU, P. C., *Speech Enhancement Based on Wavelet Thresholding the Multitaper Spectrum*, IEEE Transactions on Speech and Audio Processing, Vol. 12, No 1, pp. 59-67, January 2004.

ITU-T rec. P.862, Perceptual Evaluation of Speech quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs, *International Telecommunications Union*, GENEVA, Switzerland, February 2001.

KASTANTIN, R., STEFAOIU, D., FENG, G., MARTIN, N. and MRAYATI M., Optimal Wavelets for High Quality Speech Coding, *Proc. Of EUSIPCO-94, European Association for Signal Processing*, Edinburgh, Scotland, pp. 399-402, 1994.

KINSNER, W. and LANGI, A., Speech and Image Signal Compression with Wavelets, *IEEE Wescanex Conference Proceeding, IEEE*, New York, NY, pp. 368-375, 1993.

LOUIS, A. K., MAAB, P. and RIEDER, A., Wavelets Theory and Applications, John Wiley & Sons, 1998.

McAULAY, J. R. and MALPASS, M., *Speech Enhancement Using Soft - Decision Noise Suppression Filter*, IEEE Trans. On Acoust., Speech and Signal Processing, Vol. 28, No. 2, April 1980.

MADHUKUMAR, A. S., PREMKUMAR, A. B. and ABUT, H., Wavelet Quantization Of Noisy Speech Using Constrained Wiener Filtering, *ASILOMAR 1997, Proceedings*, pp. 2-5, Pacific Grover, C.A., November 1997.

MALLAT, S., *A theory for Multiresolution Representation Signal Decomposition: the Wavelet Representation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11 No. 7, pp. 674-693, 1989(A).

MALLAT, S., *Multiresolution Approximation and Wavelets*, Trans. Amer. Math. Soc., 315, pp. 69-88, 1989(B).

MEYER, Y., Wavelets: Algorithms and Applications, *SIAM Books*, Philadelphia, 1993.

MISITI, M., MISITI, Y., OPPENHEIM, G. and POGGI, J., Matlab: Wavelet Toolbox User's Guide, *Natick, MA: Math Works, Inc.*, 1996.

MORETTIN, P. A., Ondas e Ondaletas, Da Análise de Fourier à Análise de Ondaletas, *Edusp*, São Paulo, SP, 1999.

NIEVERGELT, Y., Wavelets Made Easy, *Bikhäuser*, Boston, 1999.

PINTÉR, S., Speech Enhancement by Soft Thresholding in the Perceptual Wavelet Domain, *IEEE Workshop on Nonlinear Signal and Image Processing II*, pp. 666-669, Neos Marmaras, June 1995.

PINTÉR, S., *Perceptual Wavelet-Representation of Speech Signals and its Application to Speech Enhancement*, Computer Speech and Language, vol. 10, pp. 1-22, 1996.

QIANG, F. and WAN, E. A., A Novel Speech Enhancement System Based on Wavelet Denosing, *Center of Spoken Language understanding, OGI, School of Science and Engineering at OHSU*, 2003.

RAMÍREZ, J., SGURA, J.C., BENÍTEZ, C., DE LA TORRE, A. and RUBIO, A., *Efficient Voice Activity Detection Algorithms using Long-term Speech Information*, Speech Communication, Vol. 42, No. 3-4, pp. 271-287, April 2004.

RIOUL, O. and VETTERLI, M, *Wavelets and Signal Processing*, IEEE Signal Processing Magazine, Vol. 8 No. 4, pp. 14-38, 1991.

RIX, A. W., BEERENDS, J. G., HOLLIER, M. P. and HEKSTRA, A. P., PESQ – The New ITU Standard for end-to-end Speech Quality Assessment, *AES 109th Convention*, 18p., Los Angeles, September 2000.

SARKAR, K., SU, C., ADVE, R., PALMA, M. S., CASTILHO, L. G. and BOIX, R. R., *A Tutorial on Wavelets from an Electrical Engineering Perspective, Part 1: Discrete Wavelet Techniques,* IEEE Antennas and Propagation Magazine, Vol. 40, No. 5, pp. 49-70, October 1998.

SCHROEDER, M. R., ATAL, B. S. and HALL, J., L., *Optimizing Digital Speech Coders by Exploiting Masking Properties of Human Ear*, Journal of Acoustical Society of America, p. 1647-1652, 1979.

SEOK, J. W. and BAE, K. S., Speech Enhancement with Reduction of Noise Components in the Wavelet Domain, *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97)*, vol. 2, pp. 1323-1326, April 1997.

SHEIKHZADEH, H. and ABUTALEBI H. R., An Improved Wavelet-Based Speech Enhancement System, *Eurospeech 2001*, pp. 1855-1858, September 2001.

SOHN, J. and SUNG, W., A Voice Activity Detector Employing Soft Decision Based Noise Spectrum Adaptation, *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'98)*, vol. 1, pp. 365-368, May 1998.

STRANG, G. and NGUYEN, T., *Wavelets and Filter Banks*, *Wellesley-Cambridge Press*, 1996.

TSOUKALAS, D., PARASKEVAS, M. and MOURJOPOULOS, J., Speech Enhancement Using Psychoacoustics Criteria, *IEEE Int. Conf. On Acoustic., Speech and Signal Processing*, pp. II.359-II.362, Minneapolis, 1993.

VASECHI, S. V. and FRAYLING, C. R., *Restoration of Old Gramophone Recordings*, Journal of Audio Engineering Society (JAEC), Vol. 40, No. 10, pp. 791-800, October 1992.

VIEIRA FILHO, J., SCART, P. e CHIQUITO, J. G., Redução de Ruído em Sinais de Voz: Análise e Avaliação de Técnicas Clássicas Baseadas na Subtração Espectral a Curto-termos (SECT), *13º Simpósio Brasileiro de Telecomunicações- SBT 95*, pp. 685-689, Águas de Lindóia, SP, setembro de 1995.

VIEIRA FILHO, J., "Redução de Ruído em Sinais de Voz nos Sistemas Rádio Móveis Veiculares", *Tese de Doutorado*, FEEC-DC-UNICAMP, Campinas 1996.

VISSER, E., LEE, T. and OTSUKA, M., Speech Enhancement in a Noisy Car Environment, *International Workshop on Independent Component Analysis. (ICA' 01)*, San Diego, pp.272-277, December 2001.

VISWANATHAN, V., ANDERSON, W., ROWLANDS, J., ALI, M. and TEWFIK, A., Real-Time Implementation of a Wavelet-Based Audi-Coder on the T1 TMS320C31 DSP Chip, *5th International Conference on Signal Processing Applications & Technology (ICSPAT)*, Dallas, October 1994.

ZILANY, M. S. A., HASAN, M. K and KHAN, M. R., Efficient Hard and Soft Thresholding for Wavelet Speech Enhancement, *in proceedings of the XI European Signal Processing Conference (EUSIPCO 2002)*, pp. 507-510, Toulouse, September 2002.

APÊNDICE A

ESTUDO DA APLICAÇÃO DO LIMIAR

A aplicação de limiar proposta no trabalho, para um sinal no domínio wavelet, Y, é

$$\hat{Y} = \begin{cases} Y \quad , \text{se} \mid Y \mid > \beta\lambda \\ Y. \mid \text{sigmóide}(Y) \mid \quad , \text{se} \mid Y \mid \leq \beta\lambda \end{cases}, \tag{A.1}$$

sendo

$$Sigmóide(x) = \frac{1 - e^{-\gamma x}}{1 + e^{-\gamma x}}.$$

Assim, da equação (A.1) pode-se tomar a função

$$f(x) = x \mid sigmóide(x) \mid$$

com gráfico apresentado na figura A.1, que é praticamente o mesmo gráfico apresentado na figura 5.7, para as características de entrada e saída do limiar sigmóide. Isto se explica devido ao fato de que

$$\lim_{x \to \infty} x. \mid sigmóide(x) \mid = x$$

e

$$\lim_{x \to -\infty} x. \mid sigmóide(x) \mid = x,$$

ou seja, para x com módulo suficientemente grande, $f(x) \approx x$, como se pode observar na figura A.2.

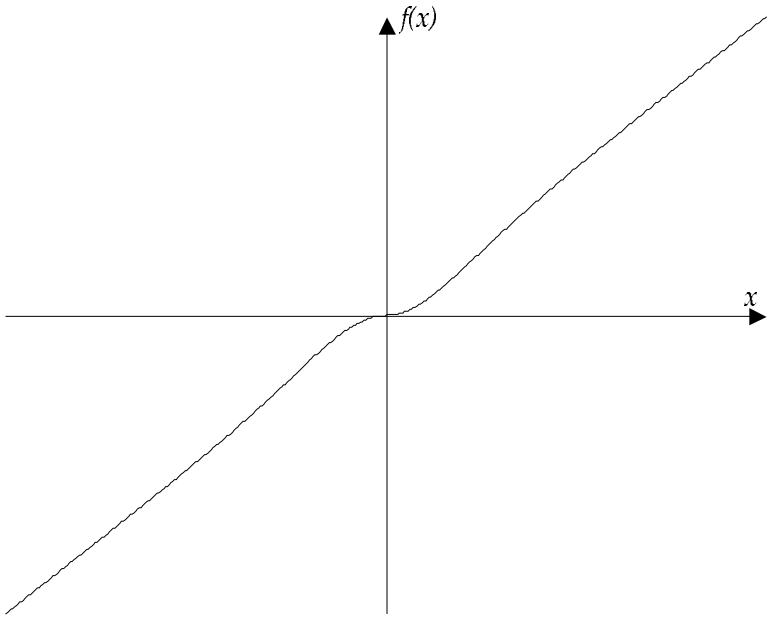


Figura A.1 – gráfico de $x.|\,sigmóide(x)\,|$.

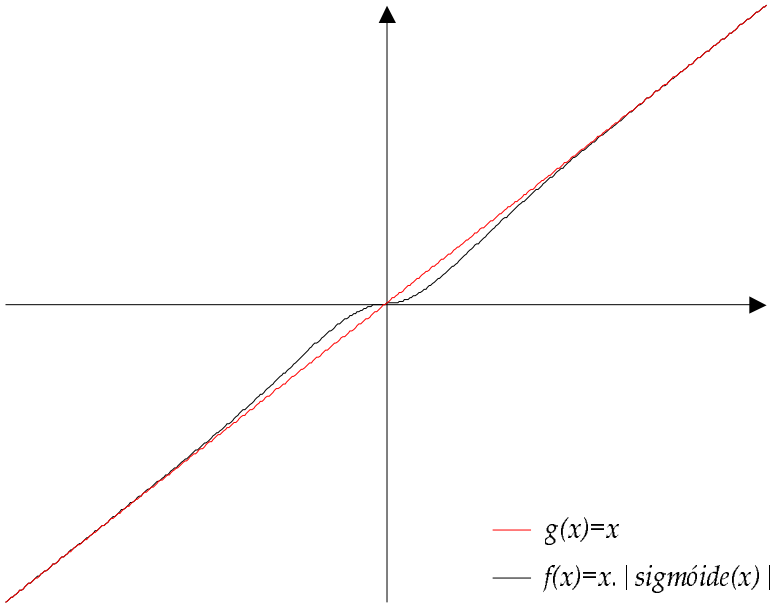


Figura A.2 – Gráficos $g(x)=x$ versus $f(x)=x.\,|\,sigmóide(x)\,|$.

Apesar das semelhanças entre os gráficos apresentados nas figura 5.7 e A.1, não se pode aplicar $f(x)$ direto para a redução de ruído em sinais de voz, sem o uso de um limiar, pois para sinais com formato *wave*, sinais com amplitude compreendida no intervalo (-1,1), isto causaria total eliminação do sinal. Por isso, $f(x)$ é aplicada somente para valores do sinal, no domínio wavelet, que estão abaixo de um limiar especificado.

O limiar aplicado da forma que se propõe na equação (A.1) também apresenta desnível, matematicamente descontinuidade, quando se passa dos coeficientes que estão abaixo do limiar para os que estão acima, mas este desnível é bem mais suave quando comparados com outros métodos já existentes.

APÊNDICE B

PESQ E DISTÂNCIA CEPSTRAL

B.1 PESQ

A PESQ compara um sinal original com um sinal degradado, resultante da passagem através de um sistema de comunicação. A saída PESQ é uma predição da percepção de qualidade que seria obtida por um indivíduo, num teste subjetivo de escuta.

Numa primeira etapa, uma série de atrasos entre os sinais original e degradado é computada, um para cada intervalo de tempo onde os atrasos são significativamente diferentes entre os dois sinais. Para cada intervalo são calculados os pontos de início e fim. O algoritmo de alinhamento tem condições de tratar alterações de atraso tanto nos períodos de silêncio como nos de fala.

Baseado no conjunto de atrasos encontrados, os sinais original e degradado já alinhados são comparados utilizando-se um modelo perceptivo, conforme ilustrado na figura B.1. A chave deste processo é transformar ambos os sinais numa forma de representação, semelhante à representação psicofísica do sinal no sistema auditivo humano, levando em conta a freqüência perceptiva (*Bark*) e a intensidade (*Sone*). Isto é alcançado em vários estágios: alinhamento de tempo, alinhamento de nível para calibragem do nível de escuta, mapeamento tempo-freqüência, deformação de freqüência e compressão da escala de intensidade.

O resultado da avaliação da PESQ é uma nota que varia entre –0,5 e 4,5. Porém, em sistemas de telecomunicações o método mais usado é o ACR (Absolute Category Rating method) que mapea a escala PESQ na MOS (Mean Square Opinion) e neste método a nota varia entre 1,0 e 5,0.

Se for usado o padrão ACR, sinais com notas maiores ou iguais a 4 são considerados de boa qualidade, caso contrário, sinais com notas maiores que 3 são considerados de boa qualidade.

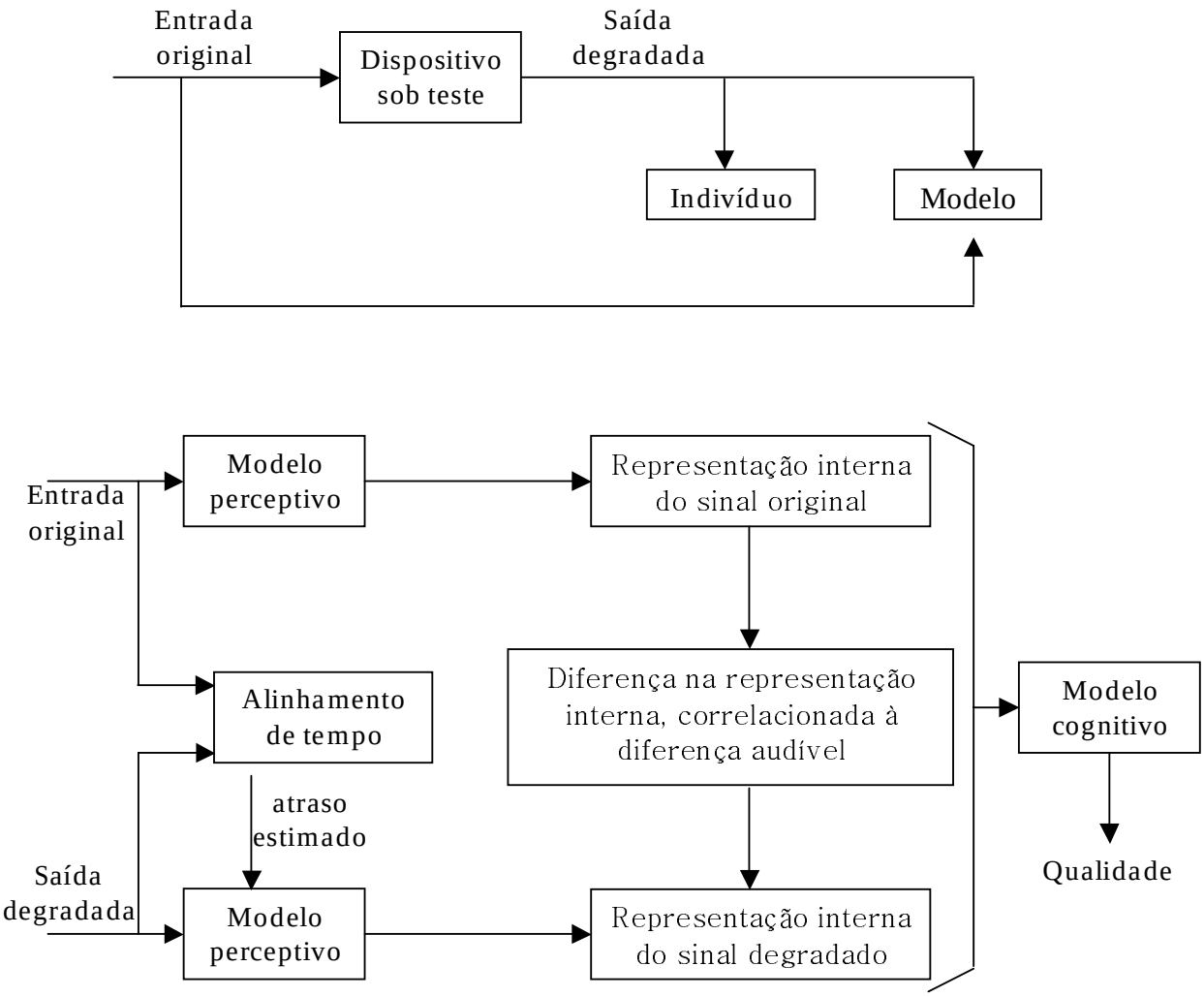


Figura B.1 – Filosofia básica usada na PESQ.

B.2 Distância Cepstral

A distância cepstral representa um medida Euclidiana ponderada entre dois sinais, usando como base seus cepstros. Na verdade, isto representa uma distância direta entre os

coeficientes e seu modelo LP (predição linear). O objetivo é mostrar o quanto os dois sinais estão distanciados espectralmente.

De um modo geral, sejam:

$$\frac{\sigma}{S(z)} \quad \text{e} \quad \frac{\sigma'}{S'(z)} \tag{B.1}$$

os modelos espectrais dos dois sinais. A diferença entre os logaritmos destes dois modelos fornece uma distância *D*, que é dada por:

$$D(\theta) = \ln\left(\frac{\sigma}{S(e^{j\theta})}\right) - \ln\left(\frac{\sigma'}{S'(e^{j\theta})}\right) \tag{B.2}$$

sendo θ a freqüência normalizada no plano z e π representa a metade da freqüência de amostragem.

A distância cepstral logarítmica é então definida como:

$$(d_p)^p = \frac{1}{2\pi}\int_{-\pi}^{\pi} |D(\theta)|^p \, d\theta \tag{B.3}$$

sendo que p pode ser definindo como um parâmetro de ajuste da medida. Para grandes valores de p, a distância obtida reflete apenas as grandes distorções.

De acordo com a equação B.1, se $S(z)$ e $S'(z)$ são polinômios em z^{-1} com raízes no interior de um círculo unitário e $S(\infty) = 1$, mostra-se através de uma expansão em série de Taylor (GRAY, 1976) que:

$$\ln[S(z)] = -\sum_{k=1}^{\infty} c_k z^{-k} \tag{A2.4}$$

onde os c_k representam os coeficientes cepstrais. Conseqüentemente, e considerando a equação B.4, tem-se que:

$$\ln\left[\frac{\sigma^2}{|S(e^{j\theta})|^2}\right] = -\sum_{k=-\infty}^{\infty} c_k z^{-j\theta k} \qquad \text{(B.5)}$$

com $c_0 = \ln(\sigma^2)$ e $c_{-k} = c_k$.

Das equações B.2 e B.5 tem-se que a distância cepstral quadrática é dada por:

$$(d_2)^2 = \sum_{k=-\infty}^{\infty} (c_k - c'_k)^2 = (c_0 - c'_0)^2 + \sum_{k=1}^{\infty} (c_k - c'_k)^2 \qquad \text{(B.6)}$$

A implementação da equação B.6 é baseada na limitação do número de coeficientes cepstrais. Apesar de parecer um "truncamento" da medida, o que se faz de fato é suavizar a distância de acordo com o número de coeficientes escolhidos. Assim, tomando-se L coeficientes cepstrais, a distância cepstral final é dada por:

$$(d_{cep})^2 = (c_0 - c'_0)^2 + \sum_{k=1}^{L} (c_k - c'_k)^2 \qquad \text{(B.7)}$$

Naturalmente, quanto maior for o valor de L, maior será a definição da distância. Mas, para o caso de se usar um modelo LPC de ordem x para a obtenção dos coeficientes cepstrais, um valor razoável para L será $L=2x$.

Bibliografia

BEERENDS, J. G., HEKSTRA, A. P., RIX, A. W. and HOLLIER, M. P., *Perceptual Evaluation of Speech Quality (PESQ) The New ITU Standard for end-to-end Speech Quality Assessment Part II – PSYCHOACOUSTIC MODEL,* AES Journal, Vol. 50, No. 10, pp. 765-778, October 2002.

GRAY Jr., A. H. and MARKEL, J. D., *Distance Measures for Speech Processing,* IEEE Trans. On Acoustics, Speech and Signal Processing, Vol. 24, No. 5 - October 1976.

ITU-T rec. P.862, Perceptual Evaluation of Speech quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and

Speech Codecs, *International Telecommunications Union*, GENEVA, Switzerland, February 2001.

RIX, A. W., BEERENDS, J. G., HOLLIER, M. P. and HEKSTRA, A. P., PESQ – The New ITU Standard for end-to-end Speech Quality Assessment, *AES 109th Convention*, 18p., Los Angeles, September 2000.