



COPPE/UFRJ

AVALIAÇÃO DE QUALIDADE DE SINAIS DE ÁUDIO E VOZ UTILIZANDO
DWT E MARCAS D'ÁGUA DIGITAIS

Rafael Almeida de Jesus

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Luiz Wagner Pereira Biscainho

Rio de Janeiro
Dezembro de 2009

AVALIAÇÃO DE QUALIDADE DE SINAIS DE ÁUDIO E VOZ UTILIZANDO
DWT E MARCAS D'ÁGUA DIGITAIS

Rafael Almeida de Jesus

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO
ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE
ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE
JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A
OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA
ELÉTRICA.

Examinada por:

Prof. Luiz Wagner Pereira Biscainho, D.Sc.

Prof. Sergio Lima Netto, Ph.D.

Prof. Lisandro Lovisolo, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
DEZEMBRO DE 2009

Jesus, Rafael Almeida de

Avaliação de Qualidade de Sinais de Áudio e Voz Utilizando DWT e Marcas d'Água Digitais/Rafael Almeida de Jesus. – Rio de Janeiro: UFRJ/COPPE, 2009.

XIV, 92 p.: il.; 29, 7cm.

Orientador: Luiz Wagner Pereira Biscainho

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2009.

Referências Bibliográficas: p. 72 – 75.

1. Avaliação de Qualidade. 2. Áudio. 3. Voz. 4. DWT. 5. Marca d'Água. I. Biscainho, Luiz Wagner Pereira. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

*Aos meus pais, Paulo e
Marilene, por sua incansável luta
pela minha felicidade. Ao meu
querido irmão, Felipe, a pessoa
que eu mais amo no mundo.*

Agradecimentos

Em primeiro lugar, agradeço a Deus por tudo.

Agradeço ao meu pai Paulo, que sempre foi e sempre será o maior exemplo que tenho na vida; à minha mãe Marilene, que sempre foi o meu porto seguro, o meu equilíbrio e a minha consciência; ao meu irmão Felipe, sem o qual os meus dias não teriam a metade do brilho que possuem.

Agradeço à Ana Paula Novello, que esteve ao meu lado durante a maior parte da luta e que viveu comigo a maior parte das alegrias e tristezas da caminhada, sendo sempre um referencial de doçura, bondade e desvelo.

Agradeço à Denise Silva, pelo auxílio, apoio e compreensão que me permitiram concluir este trabalho.

Agradeço ao meu orientador Luiz Wagner, que sempre me ajudou a me tornar melhor, e sempre me impulsionou para grandes realizações, além de ter sido um amigo com quem sempre contei e que terei por toda a vida.

Agradeço aos companheiros do Projeto MuQuE, prof. Sergio L. Netto, Amaro A. Lima, Fábio P. Freeland, Paulo A. A. Esquef e Bruno C. Bispo, que me proporcionaram muito aprendizado e muita alegria e que tornaram o dia a dia de trabalho prazeroso. Em especial ao Fábio P. Freeland, que me ajudou muito durante todo o trabalho, e ao Paulo A. A. Esquef, que veio somar bastante ao final.

Agradeço também a: Igor Alves, Maria Lúcia Lima, Leandro Ramos, Letícia Souza, Felipe Pereira, Glória Novello, Rodolpho Rio; e aos companheiros do LPS: Tadeu Ferreira, Leonardo Nunes, Amanda Alves, Alan Tygel, Alexandre Leizor, Flávio Rainho, Wallace Martins, Markus Lima, Camila Gussen, Ana Fernanda Quaresma, Paulo César Martins Filho, Rafael Amaro, Marcos Magalhães e André Targino, pelo apoio, auxílio e presença sem os quais os meus dias estariam privados de todos os sorrisos que eles me proporcionaram.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte financeiro.

Agradeço, finalmente, a todos aqueles não citados nominalmente mas que, de uma forma ou de outra, tornaram possível essa realização.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

AValiação DE QUALIDADE DE SINAIS DE ÁUDIO E VOZ UTILIZANDO DWT E MARCAS D'ÁGUA DIGITAIS

Rafael Almeida de Jesus

Dezembro/2009

Orientador: Luiz Wagner Pereira Biscainho

Programa: Engenharia Elétrica

A avaliação objetiva de sinais de áudio e voz é um tema que vem crescendo em importância em função dos avanços tecnológicos ocorridos nos sistemas computacionais e de telecomunicações, da mudança nos hábitos de consumo de música, e do aumento da exigência de qualidade por parte tanto dos ouvintes de música quanto dos usuários de sistemas de comunicação.

Neste trabalho será detalhada e analisada uma família de métodos que utilizam a DWT (*Discrete Wavelet Transform*) e marcas d'água digitais para aferir a qualidade de sinais de áudio e voz. A motivação para o estudo dessa família de métodos é que alguns deles podem realizar a avaliação dos sinais de áudio e voz sem comparar diretamente o sinal que está sendo investigado com o sinal que deu origem a ele.

Primeiramente será apresentada uma medida de qualidade usando a DWT e, logo depois, sua combinação com a inserção de marcas d'água digitais. Em seguida serão apresentadas modificações sobre o método para melhorar o seu desempenho, assim como serão discutidas possíveis maneiras de converter a saída dos métodos analisados em métricas consagradas, como o MOS.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

SPEECH AND AUDIO QUALITY ASSESSMENT USING DWT AND DIGITAL WATERMARKING

Rafael Almeida de Jesus

December/2009

Advisor: Luiz Wagner Pereira Biscainho

Department: Electrical Engineering

The importance of speech and audio quality assessment has been growing along the years as a result of technological advances in computer and telecommunication systems, changes in music consumption habits and the increase of quality requirement by music listeners as well as communication system users.

This work will detail and analyze a family of methods that utilizes DWT and digital watermarking over DWT (Discrete Wavelet Transform) to perform quality assessment of speech and audio signals. The motivation for studying this specific family of methods is that some of them may evaluate the signal quality without directly comparing it to any reference.

First a measure of quality that uses DWT is presented, and next it is combined with digital watermarking insertion. After that, some modifications to improve the method's performance are proposed, followed by strategies to map the method output to well-known metrics, like MOS.

Sumário

Lista de Figuras	xi
Lista de Tabelas	xiv
1 Avaliação de qualidade de sinais de áudio e voz	1
1.1 Introdução	1
1.2 Motivações para o tema	2
1.3 A percepção dos sinais de áudio	3
1.3.1 Audibilidade	4
1.3.2 Mascaramento	5
1.4 Métodos Subjetivos	6
1.4.1 P.800	6
1.4.2 BS.1116-1	7
1.4.3 BS.1534-1	8
1.5 Métodos Objetivos	8
1.5.1 PEAQ	9
1.5.2 PESQ	10
1.6 Organização do trabalho	13
2 A DWT aplicada à avaliação de qualidade	14
2.1 Introdução	14
2.2 A DWT	15
2.3 Método de comparação de sinais de áudio no domínio da DWT	18
2.4 Testes	21
2.4.1 Voz	21
2.4.2 Áudio	22
2.4.3 Parâmetros	22
2.4.4 Métricas	24
2.5 Resultados	27
2.5.1 Voz – Ruído	27
2.5.2 Voz – Perda de Pacotes	29

2.5.3	Voz – Filtragem Passa-Baixas	30
2.5.4	Voz – Codificação G.729.1	31
2.5.5	Voz – Codificação Speex	32
2.5.6	Áudio – Codificação MP3	33
2.5.7	Resultados finais	34
2.6	Conclusões	34
3	Marcas d’água digitais na avaliação de qualidade de sinais de áudio e voz	36
3.1	Introdução	36
3.2	Avaliação de qualidade usando marcas d’água digitais	37
3.2.1	Inserção de marcas d’água digitais usando a DWT	38
3.2.2	Degradação dos sinais de áudio/voz	43
3.2.3	Extração da marca d’água	44
3.2.4	Comparação das marcas	44
3.3	Testes e Resultados	44
3.3.1	Parâmetros utilizados	45
3.3.2	Voz – Ruído	45
3.3.3	Voz – Perda de Pacotes	46
3.3.4	Voz – Filtragem Passa-Baixas	47
3.3.5	Voz – Codificação G.729.1 e Speex	48
3.3.6	Áudio – Codificação MP3	49
3.4	Discussão dos resultados da literatura	51
3.5	Conclusões	52
4	Aperfeiçoamento do método de avaliação de qualidade de áudio/voz que utiliza marcas d’água digitais	54
4.1	Introdução	54
4.2	<i>DWT-watermarking</i> modificado	55
4.2.1	Modificação A: Recuperação exata da marca	55
4.2.2	Modificação B: Aumento na robustez da marca	57
4.3	Mapeamento das notas	60
4.3.1	Metodologia de mapeamento	60
4.3.2	Resultados	62
4.4	Conclusão	67
5	Conclusões	68
5.1	Resultados	68
5.2	Contribuições deste trabalho	69
5.3	Trabalhos futuros	70

Referências Bibliográficas	72
A Otimização de Q_s	76
B Parâmetros obtidos para o mapeamento do PCEWM	78
C Gráficos do mapeamento do PCEWM	80

Lista de Figuras

1.1	Curvas de igual audibilidade.	4
1.2	Mascaramento na frequência.	5
1.3	Mascaramento temporal para um sinal de 200 ms de duração.	6
1.4	Diagrama simplificado dos blocos do PEAQ.	10
1.5	Diagrama de blocos do PESQ.	11
2.1	Cálculo dos coeficientes usando banco de filtros.	17
2.2	Rótulo para os coeficientes.	20
2.3	Ponderação de importância (“pesos”) dos níveis.	25
2.4	Comparação entre o WPESQ e o WAQA na avaliação de sinais de voz contaminados com ruído branco gaussiano.	27
2.5	WAQA com ajuste linear.	28
2.6	WAQA computado em bloco.	28
2.7	Comparação entre o WPESQ e o WAQA na avaliação de sinais de voz degradados por perda de pacotes.	29
2.8	Comparação entre o WPESQ e o WAQA na avaliação de sinais de voz degradados por filtragem passa-baixas.	30
2.9	Comparação entre o WPESQ e o WAQA na avaliação de sinais de voz degradados pelo codificador G.729.1.	31
2.10	Comparação entre o WPESQ e o WAQA na avaliação de sinais de voz degradados pelo codificador Speex.	32
2.11	Comparação entre o PEAQ e o WAQA na avaliação de sinais de áudio degradados pelo codificador MP3.	33
3.1	Diagrama de operação da avaliação de qualidade usando marcas d’água digitais construídas através da DWT.	38
3.2	Bloco de inserção da marca d’água em detalhes.	39
3.3	Fluxograma de otimização de Q_s	42
3.4	Comparação entre o WPESQ e o <i>DWT-watermarking</i> na avaliação de sinais de voz contaminados com ruído branco gaussiano.	46
3.5	<i>DWT-watermarking</i> com ajuste linear.	47

3.6	Comparação entre o WPESQ e o <i>DWT-watermarking</i> na avaliação de sinais de voz degradados por perda de pacotes.	48
3.7	Comparação entre o WPESQ e o <i>DWT-watermarking</i> na avaliação de sinais de voz degradados por filtragem passa-baixas.	49
3.8	Comparação entre o WPESQ e o <i>DWT-watermarking</i> na avaliação de sinais de voz degradados pelo codificador G.729.1.	50
3.9	Comparação entre o WPESQ e o <i>DWT-watermarking</i> na avaliação de sinais de voz degradados pelo codificador Speex.	50
3.10	Comparação entre o PEAQ e o <i>DWT-watermarking</i> na avaliação de sinais de áudio degradados pelo codificador MP3.	51
4.1	Ajuste original e novo ajuste dos coeficientes da DWT.	58
C.1	Mapeamento polinomial e divisão dos sinais — sinais de voz contaminados por ruído branco.	81
C.2	Mapeamento exponencial e divisão dos sinais — sinais de voz contaminados por ruído branco.	81
C.3	Mapeamento polinomial e divisão das degradações — sinais de voz contaminados por ruído branco.	82
C.4	Mapeamento exponencial e divisão das degradações — sinais de voz contaminados por ruído branco.	82
C.5	Mapeamento polinomial e divisão dos sinais — sinais de voz contaminados por perda de pacotes.	83
C.6	Mapeamento exponencial e divisão dos sinais — sinais de voz contaminados por perda de pacotes.	83
C.7	Mapeamento polinomial e divisão das degradações — sinais de voz contaminados por perda de pacotes.	84
C.8	Mapeamento exponencial e divisão das degradações — sinais de voz contaminados por perda de pacotes.	84
C.9	Mapeamento polinomial e divisão dos sinais — sinais de voz contaminados por filtragem passa-baixas.	85
C.10	Mapeamento exponencial e divisão dos sinais — sinais de voz contaminados por filtragem passa-baixas.	85
C.11	Mapeamento polinomial e divisão das degradações — sinais de voz contaminados por filtragem passa-baixas.	86
C.12	Mapeamento exponencial e divisão das degradações — sinais de voz contaminados por filtragem passa-baixas.	86
C.13	Mapeamento polinomial e divisão dos sinais — sinais de voz contaminados pelo codificador G.729.1.	87

C.14 Mapeamento exponencial e divisão dos sinais — sinais de voz contaminados pelo codificador G.729.1.	87
C.15 Mapeamento polinomial e divisão das degradações — sinais de voz contaminados pelo codificador G.729.1.	88
C.16 Mapeamento exponencial e divisão das degradações — sinais de voz contaminados pelo codificador G.729.1.	88
C.17 Mapeamento polinomial e divisão dos sinais — sinais de voz contaminados pelo codificador speex.	89
C.18 Mapeamento exponencial e divisão dos sinais — sinais de voz contaminados pelo codificador speex.	89
C.19 Mapeamento polinomial e divisão das degradações — sinais de voz contaminados pelo codificador speex.	90
C.20 Mapeamento exponencial e divisão das degradações — sinais de voz contaminados pelo codificador speex.	90
C.21 Mapeamento polinomial e divisão dos sinais — sinais de áudio degradados por codificação MP3.	91
C.22 Mapeamento exponencial e divisão dos sinais — sinais de áudio degradados por codificação MP3.	91
C.23 Mapeamento polinomial e divisão das degradações — sinais de áudio degradados por codificação MP3.	92
C.24 Mapeamento exponencial e divisão das degradações — sinais de áudio degradados por codificação MP3.	92

Lista de Tabelas

1.1	Notas e respectivos significados da SDG/ODG.	10
2.1	Sinais de referência.	23
2.2	Resultados finais obtidos com os parâmetros que apresentaram melhor desempenho médio.	34
4.1	Resultados da modificação A.	57
4.2	Resultados das modificações B-1 e B-2, comparados à modificação A.	59
4.3	Comparação entre <i>DWT-watermarking</i> Original e Modificado	60
4.4	Resultados do mapeamento para sinais de voz contaminados com ruído.	62
4.5	Resultados do mapeamento para sinais de voz degradados por perda de pacotes.	63
4.6	Resultados do mapeamento para sinais de voz degradados por filtragem passa-baixas.	64
4.7	Resultados do mapeamento para sinais de voz degradados pelo codificador G.729.1.	65
4.8	Resultados do mapeamento para sinais de voz degradados pelo codificador Speex.	65
4.9	Resultados do mapeamento para sinais de áudio degradados por codificação MP3.	66
B.1	Parâmetros obtidos para o mapeamento polinomial.	79
B.2	Parâmetros obtidos para o mapeamento exponencial.	79

Capítulo 1

Avaliação de qualidade de sinais de áudio e voz

1.1 Introdução

A avaliação de qualidade lida necessariamente com características específicas associadas ao objeto da avaliação. Entretanto, é possível classificar os métodos de avaliação de qualidade, em geral, como:

Subjetivos Fundamentados no julgamento humano. Nesses métodos as pessoas realizam a avaliação do objeto de análise sob condições controladas. Possuem a grande vantagem de gerarem resultados muito representativos, porém são extremamente custosos devido ao tempo e às condições controladas exigidas.

Objetivos Fundamentados em processamento computacional dos sinais. Nesses métodos procura-se prever o julgamento humano através de cálculos matemáticos. São pouco custosos se comparados aos métodos subjetivos, porém bastante sujeitos a previsões equivocadas.

Com referência O sinal cuja qualidade se pretende analisar é comparado com a sua versão original, ou seja, sem modificações/degradações.

Sem referência O sinal cuja qualidade se pretende analisar é avaliado de forma absoluta, ou seja, sem comparações.

Intrusivos O sistema de avaliação de qualidade interfere no sinal ou no sistema em que o sinal está inserido.

Não-Intrusivos O sistema de avaliação de qualidade não interfere no sinal ou no sistema em que o sinal está inserido.

A literatura [1] algumas vezes considera como não-intrusivos os métodos sem referência. A classificação apresentada nesse trabalho busca ser simples, objetiva e inambígua, e por isso diferencia os métodos sem referência dos métodos não-intrusivos.

Neste trabalho serão estudados métodos objetivos que, direta ou indiretamente, buscam efetuar a avaliação de forma comparativa, não importando se a qualidade do sinal é boa ou ruim em termos absolutos. Isso significa que a meta de “qualidade” aqui será o sinal sob avaliação ser igual ao sinal de referência.

A avaliação de qualidade de sinais de áudio/voz pode ser feita de diversas maneiras. A avaliação sem referência de uma pessoa sobre a qualidade de um sinal de áudio/voz provavelmente seria diferente da avaliação dessa mesma pessoa sobre a qualidade desse mesmo sinal (que pode ter sido acidentalmente degradado, deliberadamente modificado ou mesmo restaurado) com referência à sua versão livre de processamento.

Os diferentes interesses/enfoques na avaliação geraram padrões distintos para os métodos subjetivos. Alguns dos padrões mais importantes serão apresentados na Seção 1.4.

Na prática, os avaliadores objetivos de qualidade de sinais de áudio/voz costumam lançar mão de diversos fenômenos estudados na Psicoacústica para realizar a sua estimativa da qualidade dos sinais. Nas Subseções 1.5.1 e 1.5.2 serão apresentados métodos que se baseiam na Psicoacústica para a geração de sua estimativa de qualidade.

1.2 Motivações para o tema

A principal motivação para a investigação dentro do tema “avaliação de qualidade de sinais de áudio e voz” é o crescimento de sua importância desde meados da década de 1990. Esse crescimento foi impulsionado pelos avanços tecnológicos verificados nos sistemas de comunicação e computacionais, assim como, num segundo momento, pela mudança nos hábitos de consumo de música da sociedade, que migrou dos álbuns comprados em lojas para o consumo de músicas codificadas obtidas (legal ou ilegalmente) na *Internet*.

É importante destacar também que os avanços tecnológicos estão elevando cada vez mais o nível de exigência dos consumidores em todos os setores da indústria de bens de consumo. Já não é mais satisfatório assistir a um filme em uma TV pequena e com o som gerado por pequenos alto-falantes de baixa qualidade; anseia-se por assistir ao filme em uma tela plana, em formato *widescreen*, e com um sistema de som de alta fidelidade que permita a imersão da audiência no universo retratado no filme. Níveis de exigência mais elevados requerem um maior controle da qualidade

daquilo que se oferece aos consumidores.

Outro movimento importante a ser considerado é a evolução dos métodos de avaliação de qualidade de áudio para se ajustarem às demandas do mercado. Antes, as demandas se encontravam restritas apenas à comunicação de voz em canais telefônicos tradicionais (300 Hz a 3400 Hz) e à música codificada com alta qualidade; agora a comunicação por voz de alta qualidade na Internet e, no sentido oposto, a demanda por músicas para telefones celulares com altas taxas de compressão sem requisitos elevados de qualidade são exemplos de aplicações que tornam insuficientes os métodos hoje disponíveis.

Nesse cenário, a análise dos métodos de avaliação de qualidade de áudio e voz utilizando DWT se apresenta bastante interessante, na medida em que, conforme será detalhado ao longo dos próximos capítulos, eles não se restringem a uma única aplicação; ao contrário, seu estudo pode levar ao desenvolvimento de métodos direcionados para a avaliação de qualidade de sinais de áudio ou voz, em versões com e sem referência. Em resumo: a abrangência e a flexibilidade dessa família de métodos justificam o seu estudo.

1.3 A percepção dos sinais de áudio

Uma análise superficial poderia definir um sinal digital de áudio/voz como um simples conjunto de amostras. Uma modelagem mais criteriosa poderia defini-lo como uma função-amostra de um processo aleatório não-estacionário [2]. Entretanto, em se tratando de avaliação de qualidade de sinais de áudio/voz, não se está interessado na descrição matemático-estatística dos sinais, e sim na percepção humana deles.

Com a afirmação anterior não se pretende propalar a inutilidade ou ineficiência dos modelos matemáticos, e sim destacar a importância de um outro ramo da ciência: a Psicoacústica.

A Psicoacústica pode ser definida como a ciência que estuda as relações estatísticas entre os estímulos acústicos e as sensações auditivas. Ela não se preocupa com a física das ondas mecânicas ou com a biologia do ouvido humano. Seu foco de interesse é a percepção pelo sistema auditivo humano [3].

O entendimento de como se dá a percepção auditiva do ser humano é ferramenta indispensável para que se consiga construir um método objetivo de avaliação de qualidade de voz ou áudio que seja capaz de obter resultados análogos aos dos métodos subjetivos. A construção de uma medida ou de uma série de medidas que estejam correlacionadas com a qualidade do sinal sob avaliação depende da capacidade de se modelar computacionalmente os fenômenos que são caracterizados pela Psicoacústica.

Dois importantes conceitos da Psicoacústica serão apresentados a seguir: a **au-**

dibilidade e o **mascaramento**.

1.3.1 Audibilidade

O ouvido humano pode responder a uma faixa muito grande de intensidades sonoras, podendo chegar a 120 dB a diferença de intensidade entre o som mais intenso que pode ser ouvido sem causar dano imediato ao aparelho auditivo e o som mais suave que pode ser detectado por ele [4]. Isso não significa que o mais suave e o mais intenso dos sons possam ser detectados ao mesmo tempo; verifica-se uma adaptação da mecânica do ouvido, de acordo com a intensidade do som que se recebe.

Dentro desse entendimento, a audibilidade pode ser vista como a sensibilidade variável do ser humano às diferentes faixas de frequência¹. Assim, por exemplo, é necessária uma energia mais elevada para que se comece a ouvir um sinal de 200 Hz que a necessária para que se comece a ouvir um sinal de 1000 Hz.

Através de testes realizados com grupos de indivíduos, foi possível obter [6] [7], em cada frequência, a pressão sonora necessária para se obter sensações auditivas equivalentes. Dessa maneira, chegou-se a um conjunto de curvas que pode ser visto na Fig. 1.1. Nela, pode-se notar que a numeração das curvas em ‘fones’ está relacionada aos níveis de pressão sonora em dB_{SPL} na frequência de 1 kHz. É importante destacar também que a curva tracejada localizada na parte inferior da Fig. 1.1 — rotulada como limiar de audibilidade — indica quais são os níveis de pressão sonora necessários para que um estímulo em cada frequência possa ser detectado pelo ouvido humano.

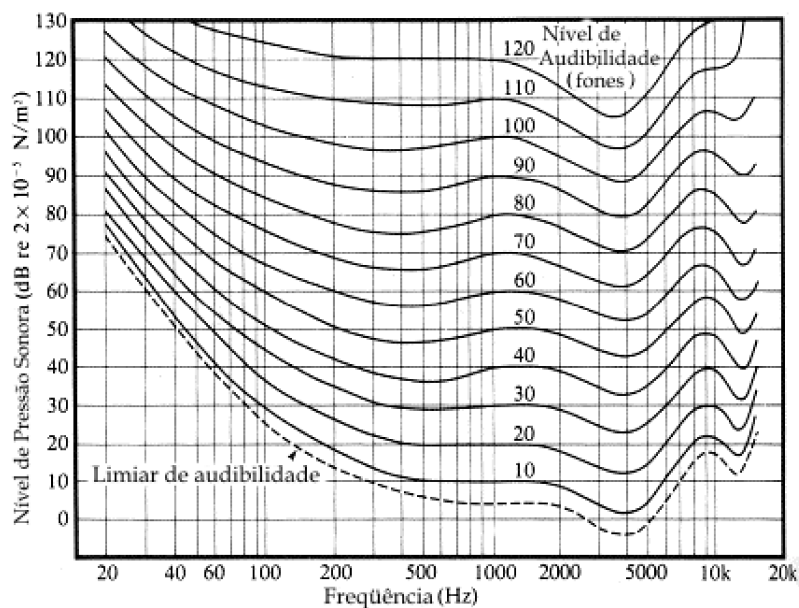


Figura 1.1: Curvas de igual audibilidade (adaptado de [7]).

¹Usualmente são adotados os limites da audição humana como 20 Hz e 20 kHz [5].

Através da análise do conceito da audibilidade, é possível identificar a necessidade de os avaliadores trabalharem com os sinais no domínio da frequência (ou algum domínio que permita algum tipo de separação das componentes na frequência), buscando ponderar a degradação detectada em uma dada faixa do espectro de frequências pela sensibilidade do ouvido humano nesta faixa.

1.3.2 Mascaramento

O fenômeno do mascaramento pode ser entendido como a capacidade de um som impedir que outro som menos intenso seja ouvido, sob determinadas condições.

A vivência do mascaramento de sons menos intensos por sons mais intensos é um fato comum no cotidiano das pessoas: é muito corriqueiro perder-se a inteligibilidade daquilo que se tenta ouvir por ocorrência de um som mais intenso simultâneo.

O estudo do fenômeno do mascaramento pode ser dividido em duas partes [4]: **mascaramento freqüencial** e **mascaramento temporal**.

Mascaramento Freqüencial Também denominado mascaramento na frequência, ocorre quando há dois sinais simultâneos de frequências muito próximas. Nessa situação, o de maior intensidade pode “encobrir” o (impedir a audição do) de menor intensidade.

Na Fig. 1.2 encontra-se ilustrado o mascaramento freqüencial, onde é possível observar a curva de mascaramento que o sinal de maior intensidade (ao centro) gera em torno de si, encobrindo os sinais de frequências vizinhas à sua.

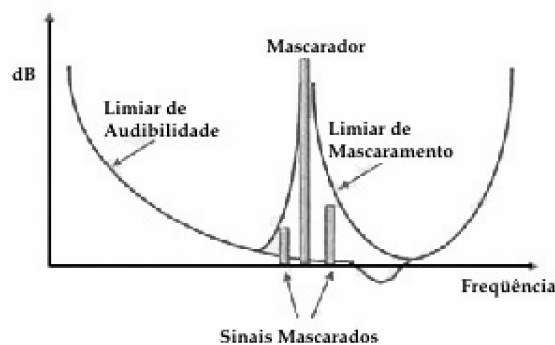


Figura 1.2: Mascaramento na frequência (adaptado de [8]).

Mascaramento Temporal Também denominado mascaramento no tempo, pode ser entendido como uma extensão do mascaramento freqüencial. No mascaramento temporal, o fenômeno descrito anteriormente deixa de ser observado apenas para eventos simultâneos e passa a estender-se ao longo do tempo. A Fig. 1.3 ilustra esse comportamento para um sinal de 200 ms de duração.

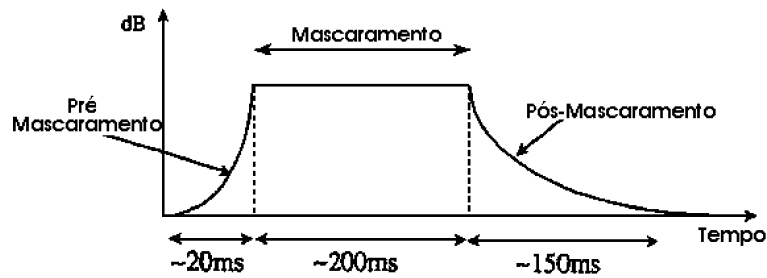


Figura 1.3: Mascaramento temporal para um sinal de 200 ms de duração. O trecho indicado como “Mascaramento” representa o intervalo de tempo em que o sinal mascarador (de aproximadamente 200 ms) se encontra presente (adaptado de [8]).

No mascaramento temporal podem ser identificados dois fenômenos: o **pré-mascaramento** que um sinal produz sobre sinais anteriores a ele e o **pós-mascaramento** que um sinal produz sobre sinais posteriores a ele.

1.4 Métodos Subjetivos

Por serem baseados na avaliação direta do som por pessoas, os chamados métodos subjetivos de avaliação de qualidade de áudio produzem resultados mais acurados e coerentes com a qualidade percebida pelo ouvinte médio. Já foram desenvolvidas diversas padronizações para esse tipo de avaliação (e.g. [9], [10] e [11]), que entretanto recaem em duas dificuldades: reunir um grupo numeroso de indivíduos para realizar os testes e garantir certas condições repetíveis para os testes e o ambiente. Tais fatores tornam os métodos subjetivos custosos e demorados. Entre os métodos da literatura, destacam-se:

- ITU-T P.800 [9] – Conjunto de métodos para avaliação subjetiva de qualidade de sinais de voz no contexto de transmissão.
- ITU-R BS.1116-1 [10] – Método para avaliação subjetiva de qualidade de sinais de áudio de alta qualidade, i.e., com baixos níveis de distorção.
- ITU-R BS.1534-1 [11] – Método para avaliação de sinais de áudio de qualidade intermediária, i.e., sujeitos a níveis moderados de distorção.

1.4.1 P.800

Na recomendação P.800 [9] se encontram três tipos de avaliação de sinais de voz, a saber:

ACR – *Absolute Category Rating* – Fundamentada na avaliação de um único estímulo, i.e., avaliação não comparativa. Os indivíduos devem ouvir o sinal que se deseja avaliar e em seguida classificá-lo em um de cinco níveis possíveis. Essa classificação pode ter três objetivos distintos: a avaliação da qualidade do sinal de voz, a avaliação do esforço para o entendimento da voz ou a avaliação do volume preferencial para audição do sinal de voz. Para esse teste, a recomendação define a necessidade de se utilizar de 24 a 36 indivíduos não treinados (leigos), com pelo menos 12 de cada gênero.

DCR – *Degradation Category Rating* – Fundamentada na avaliação de dois estímulos, i.e., avaliação comparativa. Os indivíduos devem classificar o sinal sob teste comparando-o a um sinal de referência entre 5 níveis de degradação, indo desde uma degradação imperceptível até uma degradação muito severa. Para esse teste, a recomendação define a necessidade de se utilizar de 24 a 32 indivíduos não treinados (leigos), com pelo menos 12 de cada gênero.

CCR – *Comparison Category Rating* – Fundamentada na avaliação de dois estímulos, i.e., avaliação comparativa. Os indivíduos devem ouvir os dois sinais, sendo um deles o próprio sinal de referência e o outro o sinal sob teste, que pode ser uma versão degradada ou melhorada do sinal de referência, e em seguida classificar os estímulos um contra o outro em uma escala de 7 níveis. Para esse teste, a recomendação define a necessidade de se utilizar de 24 a 32 indivíduos não treinados (leigos), com pelo menos 12 de cada gênero. Um diferença fundamental entre o CCR e o DCR é que no CCR utiliza-se um sinal de referência escondido, i.e., o indivíduo não sabe se o sinal que ele está ouvindo é o sinal sob teste ou o sinal de referência.

1.4.2 BS.1116-1

A recomendação BS.1116-1 [10] define um teste para sinais de áudio duplamente cego² e com referência escondida. Três estímulos são apresentados aos indivíduos, e é permitido que eles os escutem em qualquer ordem e quantas vezes lhes aprouver. Um dos estímulos é identificado como o de referência, e os outros dois são, não identificadamente, o sinal de referência escondido e o sinal sob teste. Os indivíduos devem avaliar a diferença entre os dois sinais não identificados e o sinal de referência, dando uma nota em uma escala de 5 níveis (com passo de 0,1), indo desde uma diferença imperceptível até uma diferença muito irritante.

Como se pode perceber pela metodologia do teste, o objetivo pretendido é a avaliação de diferenças muito pequenas no sinal de áudio ou, em outras palavras, a

²Tanto o aplicador do teste quanto o indivíduo testado desconhecem a natureza do objeto sob avaliação.

avaliação de sinais de áudio de alta qualidade. Esse método foi desenvolvido visando à avaliação de qualidade de sinais de áudio codificados por codificadores perceptivos, e normalmente é aplicado nesse contexto. Por esse motivo, são necessários ouvintes especialistas, que devem passar por uma pré-seleção, além de o método recomendar uma pós-seleção dos resultados após a realização dos testes. O número de indivíduos especializados selecionados deve ser de ao menos 20, havendo ainda na recomendação uma série de orientações para o dimensionamento do teste de acordo com o que se deseja avaliar.

1.4.3 BS.1534-1

A recomendação BS.1534-1 [11] define um teste para sinais de áudio duplamente cego e com referência e âncora escondidas. É possível a avaliação de até 15 degradações distintas para um mesmo sinal. Para tal, fornecem-se um sinal de referência e dois outros sinais escondidos entre os sinais sob teste, uma repetição do de referência e um âncora, sendo este último um sinal afetado por uma degradação padrão (a recomendação sugere uma versão passa-baixas do sinal de referência).

Os sinais sob teste e os sinais de referência e âncora escondidos são avaliados, em relação ao sinal de referência, com uma nota entre 0 e 100.

Assim como no teste definido na BS.1116-1, a audição dos sinais pode ser feita livremente, e deve ser feita uma pré-seleção e pós-seleção dos indivíduos (que também devem ser ao menos 20), porém não há a necessidade de serem indivíduos especialistas, mas apenas experientes, devido ao fato de esse teste ter sido definido para a avaliação de sinais de áudio sujeitos a níveis moderados de distorção.

1.5 Métodos Objetivos

As dificuldades inerentes aos métodos subjetivos de avaliação de qualidade de áudio/voz motivaram o desenvolvimento de métodos objetivos que realizassem a avaliação de qualidade dispensando a participação de pessoas, ou seja, através do processamento computacional dos sinais de áudio/voz. Medidas diretas tradicionais como a razão sinal-ruído (SNR) nem sempre são capazes de discernir ruído audível de ruído inaudível [4]. Por isso, os métodos objetivos de avaliação de qualidade de áudio foram aprimorados pela inserção de critérios psicoacústicos: a meta é avaliar a qualidade percebida, emulando o julgamento humano. Objeto da literatura desde a década de 1980 [12], esses métodos usualmente tentam medir a diferença percebida entre o sinal a avaliar e um sinal de referência (em vez de sua diferença absoluta).

Como este trabalho trata de uma família métodos objetivos de avaliação de qualidade de áudio e voz, seria interessante que fossem apresentados os métodos

objetivos disponíveis na literatura. Entretanto, como o número de métodos objetivos disponíveis é bastante elevado, é inviável que todos sejam apresentados. Como uma solução de compromisso, serão apresentados apenas os dois principais métodos objetivos de avaliação de qualidade de áudio e voz, respectivamente, ambos recomendados pela ITU (*International Telecommunication Union*).

1.5.1 PEAQ

Em 1998, a ITU publicou na recomendação BS.1387 [13] o avaliador PEAQ (*Perceptual Evaluation of Audio Quality*), combinando o trabalho de diversos pesquisadores na área.

O PEAQ se tornou referência na avaliação objetiva de sinais de áudio de alta qualidade (amostrados a 48 kbps). Assim como a maior parte dos avaliadores objetivos que fizeram parte do seu desenvolvimento [14], ele foi definido para avaliar a transparência de codificadores perceptivos de áudio, pela quantificação das distorções perceptíveis inseridas pelo processo de codificação. Por ser um sistema resultante de treinamento, o PEAQ garante elevada correlação com a avaliação subjetiva desse tipo de distorção. Mas pelo mesmo motivo, pode não ser adequado para avaliação de outros tipos de distorção.

Como a maioria dos métodos objetivos de avaliação de qualidade de áudio, o PEAQ é um avaliador com referência, i.e., avalia um sinal de áudio através da sua comparação com um sinal de referência.

Conforme pode ser visto na Fig. 1.4, o PEAQ realiza primeiramente o mapeamento tempo-freqüência dos sinais; em seguida, a saída do mapeamento é aplicada ao modelo psicoacústico, que fará a análise perceptual dos sinais considerando diversos efeitos, como o limiar absoluto de audibilidade e o mascaramento; no terceiro bloco, o resultado do modelo psicoacústico é processado de diversas formas, gerando as chamadas MOVs (*Model Output Variables*); esse conjunto de variáveis alimenta a rede neural (pré-treinada) que fornecerá a nota final. Deve-se ressaltar que nos três primeiros blocos os processamentos do sinal de teste e do sinal de referência são feitos de maneira independente.

A rede neural incluída no sistema é treinada para um extenso conjunto de arquivos degradados de áudio. À entrada da rede são apresentadas as MOVs calculadas para um dado arquivo e à sua saída, a nota média que a ele foi atribuída em testes subjetivos formais. A nota que a rede fornece segue a escala ODG (*Objective Difference Grade*) de diferença em relação a uma referência. A ODG tem correspondência direta com a SDG (*Subjective Difference Grade*), que se refere a testes subjetivos e cuja graduação (com os respectivos significados) é mostrada na Tab. 1.1. O treinamento da rede foi realizado conforme as SDGs.

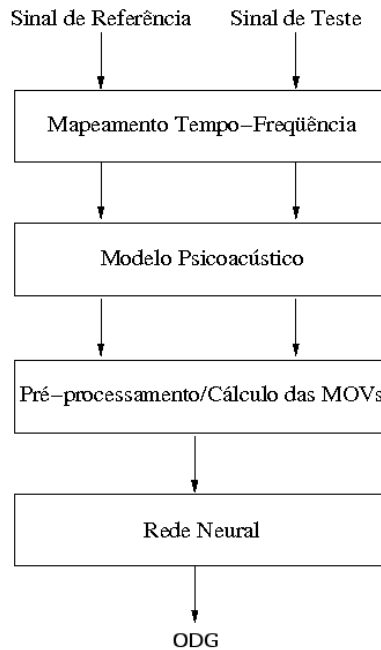


Figura 1.4: Diagrama simplificado dos blocos do PEAQ.

Tabela 1.1: Notas e respectivos significados da SDG/ODG.

Nota	Descrição da diferença
0	Imperceptível
-1	Perceptível, porém não incômoda
-2	Pouco incômoda
-3	Incômoda
-4	Bastante incômoda

O PEAQ possui dois modos de implementação, chamados básico e avançado. O modo básico é menos complexo, quatro vezes mais rápido que o avançado, e se destina a aplicações em tempo real; o modo avançado envolve processamento de maior complexidade, mas fornece resultados mais acurados, chegando a uma correlação de 95% com os resultados dos testes subjetivos [15].

1.5.2 PESQ

O método PESQ (*Perceptual Evaluation of Speech Quality*), que se encontra definido na recomendação P.862 [16] da ITU-T, foi desenvolvido a partir da combinação dos pontos principais de dois métodos anteriores, o PSQM99 e o PAMS (ambos descritos em [17]), além de receber o acréscimo de algumas inovações nas funções de equalização e na ponderação das distorções ao longo do tempo.

O PSQM99 foi desenvolvido pelo *KPN Research* em 1999 e as principais inovações que ele trazia eram o processamento assimétrico e a escalabilidade. O PAMS foi desenvolvido pela *PsyTechnics* em conjunto com a *British Telecommunications*, e pos-

sui como maior contribuição uma rotina bastante robusta de alinhamento temporal dos arquivos a serem comparados.

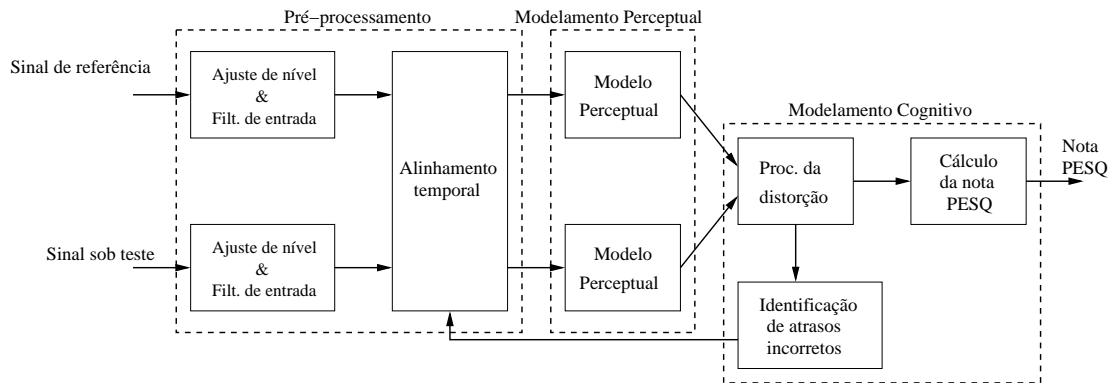


Figura 1.5: Diagrama de blocos do PESQ.

O PESQ [17–23] pode ser descrito através de três blocos principais: pré-processamento, modelagem perceptual e modelagem cognitiva. Cada um dos três blocos pode ser sub-dividido em estruturas menores, como se verá a seguir.

Pré-processamento Este bloco realiza a preparação dos sinais de referência e sob teste para o bloco de modelagem perceptual, sendo realizadas as seguintes operações:

- Ajuste de nível – Este bloco ajusta os sinais de referência e sob teste para um mesmo nível médio de potência.
- Filtragem de entrada – Este bloco realiza uma filtragem nos sinais para simular os efeitos tanto da banda telefônica (300 Hz a 3400 Hz) como das variações de sensibilidade dos aparelhos telefônicos com a frequência.
- Alinhamento temporal – Neste bloco os sinais de referência e sob teste são alinhados no tempo através do deslocamento temporal do sinal sob teste.

Modelagem perceptual Possui o objetivo de gerar uma mimetização da percepção pelo sistema auditivo humano dos sinais de referência e sob teste, sendo realizadas as seguintes operações:

- Mapeamento tempo-freqüência – Divide os sinais em janelas de 256 amostras (considerando sinais com 8000 Hz de taxa de amostragem), aplicando a FFT (*Fast Fourier Transform*) em cada uma dessas janelas.
- Remapeamento na freqüência – O espectro uniformemente distribuído resultante da FFT é remapeado para a escala Bark [8], que tenta reproduzir a acuidade da percepção de alguns fenômenos na freqüência pelo sistema auditivo humano.

- Remapeamento na intensidade – Os sinais já no domínio Bark têm a sua intensidade mapeada na escala Sone [8], que visa a emular aproximadamente a variação de percepção de intensidade sonora pelo sistema auditivo humano.

Modelagem cognitiva Usando as saídas dos blocos anteriores, esse bloco visa à geração de duas medidas (distorção simétrica e assimétrica) do ruído presente no sinal sob teste que serão combinadas linearmente para a geração da nota PESQ. Neste bloco se realizam as seguintes operações/medidas:

- Densidade de distorção – Representa a diferença entre as representações de ambos os sinais no domínio da frequência. Valores positivos indicam a presença de ruído aditivo, enquanto que valores negativos representam perda de informação.
- Mascaramento de pequenas distorções – Um conjunto de limiares ao longo da frequência é calculado para cada uma das janelas do mapeamento tempo-frequência. Em seguida, as distorções que se encontram abaixo do seu respectivo limiar são consideradas inaudíveis por terem sido mascaradas.
- Processamento assimétrico – Processando-se apenas as distorções classificadas como ruído aditivo, cada uma delas é multiplicada por um fator distinto, obtido a partir da densidade de distorção em cada uma das bandas/raias na frequência.
- Medida das distorções – Realiza-se o cálculo das medidas de distorção simétrica e assimétrica através da ponderação dos resultados ao longo do tempo, agregando primeiramente os resultados de 20 quadros com 50% de sobreposição (320 ms) e em seguida unindo todos os resultados obtidos ao longo do sinal.
- Identificação de atrasos incorretos – Atrasos incorretos nas janelas são identificados e estimados. Após o realinhamento, as medidas de distorção são recalculadas para as janelas que foram afetadas.
- Obtenção da nota PESQ – As medidas de distorção simétrica e assimétrica são combinadas linearmente para gerar a nota de saída do avaliador.

A última etapa do avaliador é converter a sua nota para uma nota equivalente

ao MOS³ que o sinal sob teste receberia. Essa conversão é feita através da equação

$$y = 0,999 + \frac{4,999 - 0,999}{1 + e^{-1,3669x+3,8224}}, \quad (1.1)$$

onde x representa a nota PESQ e y , a nota PESQ-MOS [16].

O PESQ foi desenvolvido para a avaliação de sinais de voz submetidos a distorções e erros de codificação, perda de pacotes, atraso constante ou variável, filtragem e ruído de redes analógicas.

Vale destacar a existência de uma versão de banda estendida do avaliador, chamada W-PESQ (também definida em [16]). Nela, o filtro de entrada é substituído por outro, que define a banda a ser avaliada do sinal para o intervalo entre 100 Hz e 7000 Hz.

1.6 Organização do trabalho

No Cap. 2, como forma de avaliar a qualidade de um sinal degradado em relação a um sinal de referência livre de degradação, é descrito um método de comparação de sinais de áudio no domínio da DWT. São definidos ainda os sinais para teste, as degradações impostas a eles e as métricas de avaliação dos resultados usados ao longo de todo o trabalho. Objetiva-se com esse capítulo fundamentar o uso da DWT em métodos de avaliação de qualidade de áudio e voz.

No Cap. 3, é apresentado e testado o método da literatura que utiliza marcas d'água digitais com base na DWT para fazer a avaliação de qualidade de sinais de áudio e voz [24] [25] [26] [27]. Serão utilizadas as metodologias de teste e avaliação de resultados estabelecidas no Cap. 2. O objetivo deste capítulo é analisar o método da literatura e qualificar de forma adequada o seu desempenho.

No Cap. 4 serão apresentados e avaliados alguns aperfeiçoamentos propostos ao método mostrado no Cap. 3. Novamente utilizar-se-ão as metodologias de teste e avaliação de resultados estabelecidas no Cap. 2. O objetivo deste capítulo é avaliar a possibilidade de maximizar o desempenho da avaliação com marcas d'água digitais e DWT.

No Cap. 5 serão feitas as considerações finais, traçadas as conclusões do trabalho e propostas suas possíveis continuações.

³MOS – *Mean Opinion Score* – nota média que um conjunto de ouvintes daria para avaliar a distorção em um sinal, podendo variar entre 1 (distorção muito irritante) e 5 (distorção imperceptível).

Capítulo 2

A DWT aplicada à avaliação de qualidade

2.1 Introdução

Uma maneira bastante simples de se avaliar a qualidade de sinais degradados é através de uma comparação direta com o sinal original livre de degradações. Além disso, conforme visto na Seção 1.3, a Psicoacústica indica que os sinais de áudio/voz devem ser analisados no domínio da frequência (ou em algum domínio que ofereça uma certa separação das componentes na frequência). Essa é a razão pela qual medidas comparativas das amostras no tempo, como a razão sinal-ruído, se mostraram inadequadas para a avaliação de qualidade de sinais de áudio desde as primeiras pesquisas nesse tema [12] [28].

Neste contexto, a DWT (*Discrete Wavelet Transform*), que é definida na próxima Seção, oferece a possibilidade de analisar sinais de áudio em domínios cujas unidades de representação (em contraste com as componentes frequenciais isoladas de Fourier) se assemelham a estruturas recorrentes no sinal, que assim pode ser bem representado por poucos coeficientes [29]. Tais estruturas, em geral, têm suporte finito, e a representação por DWT é “organizada” simultaneamente em deslocamentos temporais e escalas de detalhamento.

Assim, suponha-se que se deseja realizar uma comparação direta entre um sinal de áudio/voz degradado e a sua versão não-degradada no domínio da DWT e com isso gerar uma medida de diferença entre os sinais. Listam-se abaixo algumas perguntas que deveriam ser respondidas ao se idealizar essa comparação:

- Que tipo de diferença se deseja medir? Diferença absoluta com algum limiar mínimo ou máximo? Outra medida de diferença?
- Uma vez que a diferença entre os coeficientes seja medida, como ponderar os resultados entre os coeficientes em uma mesma escala? E entre escalas?

Como se pode perceber, há inúmeras formas possíveis de se realizar essa comparação. A solução que será avaliada neste capítulo é inspirada em propostas de avaliação de qualidade de áudio através de marcas d'água digitais utilizando a DWT [24] [25] [26] [27]. Duas principais motivações levam à análise que será realizada:

- Oferecer uma fundamentação para os métodos de avaliação de qualidade utilizando marcas d'água digitais sobre a DWT.
- Verificar a eficácia de se utilizar somente a DWT na avaliação de qualidade de sinais de áudio/voz (esta segunda motivação pode ser considerada secundária).

Durante a descrição do método que se fará na Seção 2.3, algumas escolhas poderão parecer pouco intuitivas. Elas se justificam pela necessidade de convertê-lo prontamente nos métodos que utilizam marcas d'água digitais.

2.2 A DWT

Uma transformada *wavelet* ortogonal—portanto com funções-base iguais para análise e síntese—de um sinal $x(t)$ pertencente ao espaço de funções quadraticamente integráveis é a decomposição de $x(t)$ através de sua projeção em um conjunto de funções-base (ou *wavelets*) formado a partir de expansões e translações de uma função-mãe $\psi(t)$. Assim, assumindo wavelets reais e definindo, por exemplo, as funções-base $\psi_{j,k}(t)$ através da equação

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \quad (2.1)$$

a transformada *wavelet* pode ser escrita de acordo com a equação

$$a(j, k) = a_{j,k} = \int_{-\infty}^{\infty} \psi_{j,k}(t) x(t) dt, \quad (2.2)$$

com $j, k \in \mathbb{Z}$ [30] [29].

O fator j da Eq. (2.1) manipula a **escala** da função base; com mais exatidão, j representa o logaritmo na base 2 da escala. Considerando suporte temporal finito, quanto maior for j , menor será o comprimento da função no domínio do tempo e, grosso modo, maior será sua banda no domínio da frequência [29]. O fator k , por sua vez, representa a localização temporal (ou espacial) da função-base. O fator $2^{j/2}$ mantém a energia constante, independentemente de j .

Analisando apenas a Eq. (2.2), verifica-se o mapeamento de uma função contínua unidimensional em um espaço discreto bidimensional que permite que a função original seja recuperada a partir apenas de um *grid* discreto [31]. Esse *grid* se encontra

embutido na Eq. (2.1) através do fator ‘2’ como base da escala e do fator omitido ‘1’ que multiplica k . Define-se assim a DWT (*Discrete Wavelet Transform*) com derivação diádica das funções-base.

A representação de sinais discretos no tempo—note-se a diferença entre a discretização no sinal original e na transformada, essa última citada anteriormente—pode ser feita através da equação

$$f(n) = \sum_{j,k} d_j(k) 2^{j/2} \psi(2^j n - k), \quad (2.3)$$

e na equação

$$d_j(k) = \langle f(n), \psi(2^j n - k) \rangle = \sum_n f(n) \psi(2^j n - k) \quad (2.4)$$

definem-se os novos coeficientes $d_j(k)$, equivalentes aos $a(j, k)$ da Eq. (2.2) [29].

A propriedade da multirresolução dos sistemas de *wavelets* define que existe um segundo conjunto de funções-base $\phi_{j,k}(t)$, chamadas funções de escalamento, que podem representar uma certa família de sinais em uma dada “escala” j_0 e que uma família mais ampla (que inclui a primeira) pode ser representada pela “escala” $j_0 + 1$. Mais do que isso, a complementação entre os conjuntos pode ser feita pelas *wavelets* da “escala” j_0 . Utilizando essa propriedade, a Eq. (2.3) pode ser reescrita como

$$f(n) = \sum_{k=-\infty}^{\infty} c_{j_0,k} 2^{j_0/2} \phi(2^{j_0} n - k) + \sum_{k=-\infty}^{\infty} \sum_{j=j_0}^{J-1} d_{j,k} 2^{j/2} \psi(2^j n - k), \quad (2.5)$$

podendo $c_{j,k}$ ser obtido conforme

$$c_j(k) = \langle f(n), \phi(2^j n - k) \rangle = \sum_n f(n) \phi(2^j n - k). \quad (2.6)$$

Ressalta-se que J , na Eq.(2.5), define o número de escalas que são necessárias para a representação do sinal discreto $f(n)$. Uma maneira simples de se entender a possibilidade de se realizar a representação utilizando um número limitado de escalas pode ser vista no desenvolvimento (2.7), onde se parte da representação de $f(n)$ utilizando apenas as funções de escalamento na escala necessárias para tal, chegando-se à Eq. (2.5) através da decomposição em escalas menores até j_0

(usualmente se arbitra $j_0 = 0$).

$$\begin{aligned}
f(n) &= \sum_{k=-\infty}^{\infty} c_{J,k} 2^{J/2} \phi(2^J n - k) \\
&= \sum_{k=-\infty}^{\infty} c_{J-1,k} 2^{J-1/2} \phi(2^{J-1} n - k) + \sum_{k=-\infty}^{\infty} d_{J-1,k} 2^{J-1/2} \psi(2^{J-1} n - k) \\
&= \sum_{k=-\infty}^{\infty} c_{J-2,k} 2^{J-2/2} \phi(2^{J-2} n - k) + \sum_{k=-\infty}^{\infty} \sum_{j=J-2}^{J-1} d_{j,k} 2^{j/2} \psi(2^j n - k) \\
&= \sum_{k=-\infty}^{\infty} c_{j_0,k} 2^{j_0/2} \phi(2^{j_0} n - k) + \sum_{k=-\infty}^{\infty} \sum_{j=j_0}^{J-1} d_{j,k} 2^{j/2} \psi(2^j n - k)
\end{aligned} \tag{2.7}$$

Essa representação de sinais discretos, oriunda da DWT, também é conhecida como DTWT (*Discrete-Time Wavelet Transform*), e será utilizada ao longo deste trabalho. Destaca-se que na literatura é muito comum chamar a representação DTWT de DWT e que aqui também se tomará essa liberdade.

A última característica que se deseja destacar na DWT é a possibilidade de implementar seu cálculo através de um banco de filtros, conforme ilustrado na Fig. 2.1. Nela, os filtros FIR h e g são obtidos a partir de ψ e ϕ [29]. Essa particularidade permite o cálculo dos coeficientes de sua representação de uma forma bastante rápida e simples.

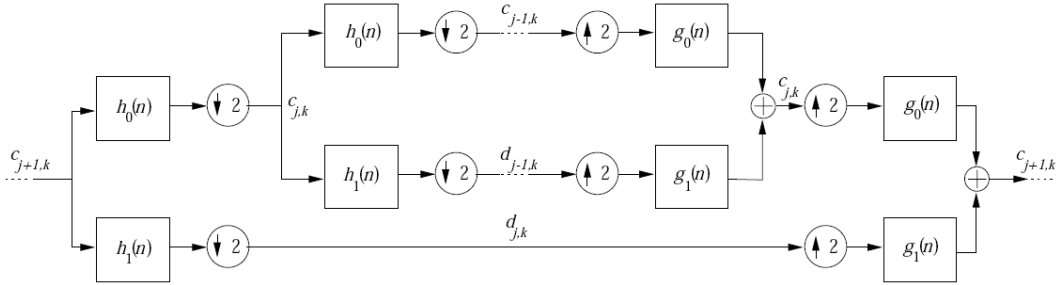


Figura 2.1: Cálculo dos coeficientes usando banco de filtros.

Uma grande motivação para a utilização da DWT para a avaliação de qualidade de áudio/voz é a flexibilidade que ela proporciona na escolha da função-mãe. Escolher uma função-mãe que leve o sinal para um domínio no qual as diferenças perceptivas se tornem evidentes pode permitir o desenvolvimento de métodos eficientes para avaliação de qualidade.

2.3 Método de comparação de sinais de áudio no domínio da DWT

Deixando de lado o emprego da marca d'água nos métodos da literatura previamente citados (e.g. [26]), aqui busca-se realizar a avaliação de qualidade de um sinal de voz/áudio apenas através da diferença entre ele e sua versão de referência no domínio da DWT. Para a medida dessa diferença, adaptar-se-á convenientemente todo o processo descrito em [26] e [27] (método baseado em marcas d'água). A seguir será feito o detalhamento do método desenvolvido, que doravante será chamado WAQA (*Wavelet-based Audio Quality Assessment*).

Número de níveis da DWT: 10; tomando-se um sinal com 48 kHz de taxa de amostragem e interpretando-se a divisão em níveis da DWT como filtros em oitavas, poder-se-ia afirmar que com 10 níveis, a menor banda correspondente teria aproximadamente 46 Hz, coerente com os limites de frequência da percepção humana (aproximadamente entre 20 Hz e 20 kHz [32]) e contendo praticamente a sua oitava inferior (de 20 Hz a 40 Hz). Em sinais com taxa de amostragem menor (e.g. 16 kHz) a escolha permanece adequada, pois a largura da menor banda é ainda menor que 46 Hz.

Escolha dos coeficientes: a comparação entre os sinais se realiza sobre um número fixo dos maiores coeficientes de cada nível. A estratégia para escolher quantos coeficientes serão comparados, em cada nível da DWT, é a seguinte:

1. Escolhe-se um número total de coeficientes a ser avaliado. Uma boa estratégia encontrada foi definir um número de coeficientes por segundo, que torna o número coerente com o tamanho do sinal que está sendo avaliado. Como a escolha de um valor para esse parâmetro não pode ser feita de forma direta e a literatura é pouco explícita na sua definição, alguns valores distintos terão o seu desempenho avaliado posteriormente.
2. Cada nível (ou escala) da DWT pode receber uma prioridade diferenciada. Além de prioridades distintas, cada um deles possui um número distinto de coeficientes: como estamos trabalhando com a DWT com derivação diádica, as escalas imediatamente inferiores possuem a metade dos coeficientes das escalas imediatamente superiores, e.g., se a escala 1 tiver 10 coeficientes, a escala 2 terá 20 coeficientes, a escala 3 terá 40 coeficientes e assim sucessivamente. Disso resultam três possibilidades básicas: ponderar o número de coeficientes escolhidos em cada nível pela prioridade que se atribui ao nível; ponderar pela quantidade de coeficientes no nível;

ponderar por ambos. Todas as três estratégias serão adotadas e terão o seu desempenho avaliado posteriormente.

3. De posse do número N de coeficientes que serão avaliados num dado nível, selecionam-se os N maiores coeficientes nesse nível. Sem que seja realizada uma reordenação dos coeficientes de acordo com seu nível, i.e., mantendo sua ordenação temporal, comparam-se os coeficientes dos dois sinais de referência (A) e sob teste (B). Isso significa que, num dado nível, o primeiro coeficiente selecionado (primeiro no tempo) do sinal A será comparado com o primeiro coeficiente selecionado do sinal B, o segundo coeficiente selecionado do sinal A será comparado com o segundo coeficiente selecionado do sinal B, e assim sucessivamente até o N -ésimo coeficiente selecionado dos sinais A e B.

Como se pode perceber, essa estratégia é bastante vulnerável à eventual variação excessiva de um coeficiente, que poderia colocá-lo ou retirá-lo do conjunto dos coeficientes sob avaliação e conseqüentemente mudaria o índice temporal dos coeficientes subseqüentes. Um erro desse tipo acarretaria um grande impacto na medida que se deseja realizar, já que se passaria a comparar coeficientes diferentes nos dois sinais.

Para contornar esse problema, uma possibilidade seria verificar o índice (no tempo) do coeficiente antes de realizar a comparação. Apesar de eficiente, essa estratégia requer os dois sinais, o de referência e o sob teste. Essa necessidade é plenamente atendida no WAQA, e portanto será adotada.

Vale ser destacado que essa necessidade não é atendida nos métodos que utilizam marca d'água, que não dispõem do sinal original para a comparação, mas apenas da marca d'água utilizada. Nesse casos, a solução proposta é impor redundância aos bits de marca d'água. Essa estratégia será explicada no Cap. 3.

Comparação: uma vez escolhidos os coeficientes que serão comparados, a comparação de cada par (coeficiente do sinal original e coeficiente do sinal degradado) é feita de acordo com as seguintes etapas:

1. Define-se para cada nível N um valor Δ_N com base no tamanho dos coeficientes daquele nível e numa constante Q_s , pela relação:

$$\Delta_N = \frac{C_{N\text{máx}} - C_{N\text{mín}}}{Q_s}, \quad (2.8)$$

onde $C_{N\text{máx}}$ é o maior coeficiente do nível considerado, $C_{N\text{mín}}$ é o menor coeficiente do nível considerado. Conforme será detalhado no Cap. 3,

os métodos que utilizam marcas d'água digitais usam Q_s para ajustar o grau de distorção que a marca d'água vai impor ao sinal. No WAQA, alguns valores distintos de Q_s serão testados e terão o seu desempenho comparado. A Eq. (2.8) foi concebida para lidar com o caso mais comum, onde o maior coeficiente é positivo e o menor, negativo. Destaca-se que as rotinas que foram desenvolvidas neste trabalho tratam adequadamente os casos especiais onde o maior e o menor coeficiente possuem o mesmo sinal. Esse tratamento não será explicitado porque foi considerado um detalhe de implementação.

2. Conforme definido na Eq. (2.9), cada coeficiente (coef) recebe um rótulo (r) que pode ser '0' ou '1', a depender do resultado da sua divisão pelo valor fixo de cada nível Δ_N . Se o resultado da divisão estiver em um intervalo par (entre $-0,5$ e $0,5$, entre $1,5$ e $2,5$ etc.), recebe o rótulo '0'; em caso contrário, recebe o rótulo '1'. Esse processo se encontra ilustrado na Fig. 2.2.

$$r \equiv \left\lfloor \frac{\sqrt{(\text{coef})^2}}{\Delta_N} + 0,5 \right\rfloor \pmod{2}. \quad (2.9)$$

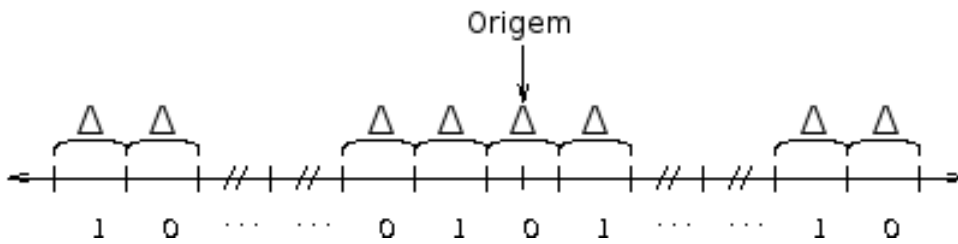


Figura 2.2: Rótulo para os coeficientes.

Parece pouco lógica a utilização desses rótulos para realizar a comparação, em vez de se fazer uma comparação direta dos valores dos coeficientes; entretanto, isso se deve novamente ao fato buscarmos preservar maior semelhança com os trabalhos que serão avaliados no Cap. 3.

Medida: para a medida da qualidade, calcula-se o percentual de coincidência dos rótulos dos coeficientes. Essa medida levará o mesmo nome do método — WAQA — e por ser uma razão, poderá variar de 0 a 1. Na prática essa medida variará entre 0,5 e 1, tendo o valor 0,5 quando não houver correlação entre os rótulos originais dos coeficientes e os novos rótulos (50% de chance de acerto), e tendo o valor 1 quando forem todos iguais.

Para testar o método WAQA proposto, resta definir os testes a serem feitos e os demais parâmetros a serem utilizados. Essas informações e os resultados obtidos serão apresentados a seguir.

2.4 Testes

2.4.1 Voz

Para o teste do WAQA na avaliação de sinais de voz, escolheu-se um conjunto de 20 sinais de voz amostrados a 16 kHz que foram extraídos da base de dados da ITU Suplemento 23 [33], experimento 1. Cada um dos sinais possui 8 segundos, sendo constituídos por dois trechos de fala (frases) intercalados com silêncio, havendo silêncio também no início e no fim do sinal. A banda dos sinais se estende até aproximadamente 7 kHz.

A fim de se realizar a avaliação do WAQA, será feita a sua comparação com o método já consagrado para a avaliação de sinais de voz, o PESQ [18] [20], na sua versão de banda estendida (até aproximadamente 7 kHz): WPESQ [16].

Na degradação dos sinais, serão utilizadas contaminações para as quais o PESQ é indicado e que além disso aparecem citadas nos trabalhos sobre avaliação de qualidade utilizando marcas d'água digitais baseadas na DWT [27]. As degradações escolhidas foram:

Ruído: adição de ruído branco gaussiano (*AWGN – Additive White Gaussian Noise*) aos sinais, gerando valores de SNR de 5 a 50 dB, em passos de 5 dB.

Perda de Pacotes: assumindo que o sinal de voz é constituído de conjuntos de pacotes que carregam 30 ms de sinal, simula-se a perda aleatória de um percentual desses pacotes. Para isso, divide-se o sinal em trechos de 30 ms e zera-se aleatoriamente um número de trechos proporcional ao percentual de perdas estabelecido, que foi definido entre 5% e 50%, com passos de 5%.

Filtragem passa-baixas: filtram-se os sinais utilizando-se filtros passa-baixas nas frequências de 1 kHz até 7 kHz, com passos de 1 kHz.

Codificação – G.729.1: codificação dos sinais de voz utilizando o codificador de voz G.729.1, padronizado pela ITU, em taxas variando entre 14 kbps e 32 kbps, com passos de 2 kbps. Não foram utilizadas taxas mais baixas devido ao fato de o codificador limitar a banda em 4 kHz para taxas menores que 14 kbps.

Codificação – Speex: codificação dos sinais de voz utilizando codificador *open-source* Speex [34], com fator de qualidade variando entre 0 e 10, com passos de 1. A principal razão de se utilizar esse codificador de voz nos testes é a impossibilidade de o codificador G.729.1 gerar sinais de voz com menor qualidade mantendo a banda de 7 kHz.

2.4.2 Áudio

Para o teste do WAQA na avaliação de sinais de áudio, escolheu-se um conjunto de 20 sinais de áudio típicos, contendo música clássica e popular, com voz feminina, voz masculina e música instrumental, assim como diversos ritmos. Na Tab. 2.1 se encontram listados os arquivos, acompanhados de uma pequena descrição.

Todos os sinais foram extraídos de CDs e depois convertidos para ficarem armazenados com taxa de amostragem de 48 kHz.

A fim de se realizar a avaliação do WAQA, será feita a sua comparação com o método já consagrado para a avaliação de sinais de áudio, o PEAQ [18] [20], na sua versão básica. A implementação utilizada foi a do prof. Peter Kabal e se encontra disponível no seu *website* [35]. A razão de utilizar os sinais com taxa de amostragem de 48 kHz é o fato de o PEAQ requerer que os sinais a serem avaliados estejam nessa taxa de amostragem.

Como o PEAQ foi desenvolvido para avaliação de sinais de áudio contaminados por ruído oriundo de codificação perceptiva, o método de codificação MP3 foi o escolhido para a contaminação dos sinais de áudio. A codificação foi realizada utilizando taxas que variavam entre 32 e 256 kbps, com passo de 32 kbps.

2.4.3 Parâmetros

Como o teste proposto não é computacionalmente muito pesado, resolveu-se realizar uma busca dos parâmetros que gerariam resultados mais coerentes com os métodos consagrados, WPESQ e PEAQ. Essa busca foi realizada entre os seguintes parâmetros:

- Wavelets-mãe: Coiflet 5 e Daubechies 3 [29]. Esse parâmetro será chamado de *wfam* e as duas opções ‘coif5’ e ‘db3’, respectivamente. Ambas foram escolhidas por terem apresentado melhores resultados em testes preliminares informais.
- Número de coeficientes da DWT a serem analisados por segundo: 450, 800 e 1250. Este parâmetro será chamado *wbps*. Novamente, a escolha desses parâmetros se deveu a testes preliminares.
- Constante de divisão dos coeficientes, Q_s : 100, 200 e 400. Define o número de divisões dos coeficientes, conforme ilustrado na Fig. 2.2. Esses valores foram escolhidos em torno de valores citados na literatura [25].
- Ponderação de importância (“pesos”) dos níveis: encontra-se ilustrada na Fig. 2.3. Com os pesos define-se quantos coeficientes serão utilizados em cada nível, por exemplo: se temos os pesos 1, 3, e 9 para os níveis a, b e c, respectivamente, os coeficientes terão que ser distribuídos de maneira que o nível c

Tabela 2.1: Sinais de referência.

Nome	Descrição	Duração(s)
apla.wav	Aplausos de platéia.	11
berim.wav	Solo de violão imitando o som de um berimbau.	11
blues.wav	<i>Blues</i> com bateria, baixo e vibrafone.	11
voivv.wav	Música popular brasileira, cantada por uma cantora acompanhada por dois violões.	8
castn.wav	Solo de castanholas.	8
harps.wav	Trecho de acordes repetidos de música barroca para cravo.	14
carni.wav	Frevo cantado por cantor, acompanhado de bateria, baixo e trompete.	12
funky.wav	Música <i>pop</i> contendo bateria, baixo, teclado e instrumentos de corda.	6
scats.wav	Canção com voz feminina em vocalise, acompanhada por bateria e piano.	11
synco.wav	Música instrumental com saxofone, piano, baixo e bateria.	11
gaffs.wav	Samba de gafeira com instrumentos de sopro, percussão e baixo.	10
glass.wav	Música lenta tocada em uma harmônica de vidro.	11
ethic.wav	Música popular brasileira com voz masculina acompanhada por instrumentos regionais de percussão.	10
balla.wav	Balada <i>pop</i> interpretada por um cantor acompanhado de bateria e violão.	12
organ.wav	Música para órgão de tubos, majoritariamente constituída de acordes.	8
orche.wav	Música orquestral em tempo lento.	13
samba.wav	Samba instrumental utilizando cavaquinho e diversos instrumentos de percussão.	13
tango.wav	Tango cantado por cantora acompanhada de piano.	11
svega.wav	Solo <i>a capella</i> de voz feminina.	9
fvráp.wav	<i>Rap</i> cantado por cantora acompanhada de grupo vocal.	8

tenha 9 vezes mais coeficientes que o nível a, e que o nível b tenha 3 vezes mais coeficientes que o nível a.

Para a distribuição dos pesos, utilizaram-se: três elevações e quedas exponenciais com bases 1,5, 2 e 3, ponderadas e não ponderadas pelo número de coeficientes em cada nível; e a curva média obtida com dois termos: a energia média dos sinais ao longo dos níveis, e os valores dos maiores coeficientes ao longo dos níveis, também ponderada e não ponderada pelo número de coeficientes em cada nível. Disso resultam 8 conjuntos distintos de ponderações (“pesos”) para os níveis. Este parâmetro será doravante chamado **pesos** e as suas opções serão: ‘exp1.5ponderado’, ‘exp2ponderado’, ‘exp3ponderado’, ‘exp1.5naoPonderado’, ‘exp2naoPonderado’, ‘exp3naoPonderado’, ‘energiaPonderado’, ‘energiaNaoPonderado’. As elevações e quedas, que dão maior importância para as escalas intermediárias, foram recomendadas na literatura (e.g. [25]) por atingirem mais a faixa média do espectro dos sinais, considerada de maior relevância (responsável pela inteligibilidade dos sinais de voz, por exemplo). As ponderações pelo número de coeficientes e a distribuição pela energia foram propostas deste trabalho por distribuírem os pesos de forma mais equilibrada.

A combinação de todos os parâmetros gera um total de 432 conjuntos distintos que devem ser testados com cada arquivo sob teste. Lembrando que cada uma das amostras de áudio sofrerá diversos níveis de degradação, é possível perceber que o número total de testes que devem ser realizados é bastante elevado, o que resulta num alto custo computacional. Por exemplo, uma máquina com processador de 3 GHz leva em média 12 horas para encontrar cada um dos conjuntos ótimos de parâmetros para cada uma das degradações avaliadas.

2.4.4 Métricas

Para que seja definido qual conjunto de parâmetros apresenta melhor resultado, é necessário escolher alguma métrica para avaliar os resultados obtidos em cada conjunto.

Uma escolha óbvia é medir a correlação entre os resultados através do seu coeficiente de correlação

$$r = \frac{\sum_{\forall i} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{\forall i} (x_i - \bar{x})^2 \sum_{\forall i} (y_i - \bar{y})^2}}, \quad (2.10)$$

onde x e y representam cada uma das amostras que estão sendo comparadas, e \bar{x} e \bar{y} suas respectivas médias resultantes. Um coeficiente de correlação igual a 1 indica

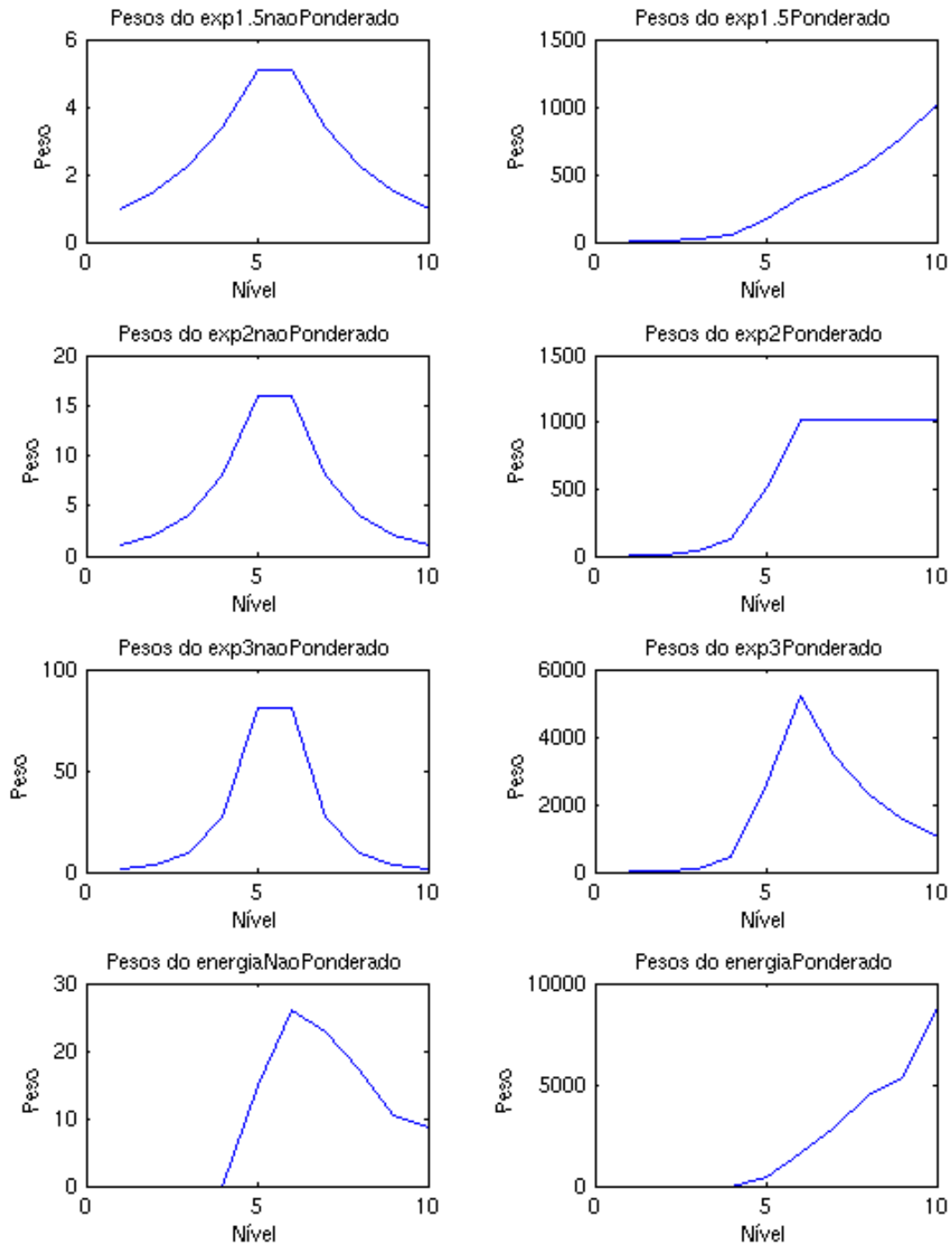


Figura 2.3: Ponderação de importância (“pesos”) dos níveis.

uma correlação direta perfeita entre as duas variáveis, -1 indica uma correlação negativa perfeita, e 0 indica que as variáveis são descorrelacionadas.

Entretanto, durante a realização do trabalho foi possível perceber que a medida da correlação não é suficiente para garantir uma boa avaliação dos resultados. Isso se deve à sua incapacidade de medir algumas características das distribuições, como, por exemplo, se as medidas obtidas são pouco precisas, i.e., se algumas notas que deveriam ser equivalentes apresentam uma diferença muito grande entre si.

Uma alternativa para aferir esse tipo de diferença é realizar uma avaliação direta da diferença entre as notas do WAQA e as notas do método de referência (PEAQ ou WPESQ). Para isso, propôs-se uma segunda medida que tenta quantificar essa diferença. Como as notas estão distribuídas em intervalos diferentes, mapeiam-se as notas do WAQA no intervalo no método de referência utilizando um polinômio do primeiro grau cujos coeficientes angular e linear são simultaneamente otimizados por mínimos quadrados. Em seguida, calcula-se a soma das diferenças absolutas entre as notas do método de referência e as notas mapeadas do WAQA, como expresso em

$$\text{dist} = \frac{\sum_{\forall i} \text{abs}(A_i - R_i)}{I}, \quad (2.11)$$

onde A_i é a i -ésima nota do WAQA ajustada pelo polinômio de primeiro grau, R_i é a i -ésima nota do método de referência e I é simplesmente o número de notas do WAQA avaliadas.

A união das duas medidas foi feita da seguinte maneira: tomam-se o conjunto de parâmetros com a maior correlação e os conjuntos cujas correlações estejam distantes até 5% da maior correlação, elegendo-se como melhor conjunto aquele que tiver a menor diferença encontrada pela segunda medida.

Serão apresentados a seguir os resultados obtidos para todas as contaminações de voz e áudio. Em todas as figuras que serão apresentadas, os resultados foram reordenados de maneira a deixar as notas dos métodos de referência numa ordem crescente. No topo das figuras se encontram também representadas as duas métricas descritas, tanto a correlação (chamada `corr` nas figuras) quanto a medida de desvio médio (ou de distância média) em relação à curva com ajuste de primeiro grau (chamada `dist` nas figuras).

2.5 Resultados

2.5.1 Voz – Ruído

Na Fig. 2.4 encontram-se representadas as notas obtidas com o WAQA e as notas do WPESQ no teste de avaliação de qualidade de sinais de voz contaminados com ruído branco gaussiano. A descrição dos diferentes contaminações foi apresentada na Seção 2.4.1.

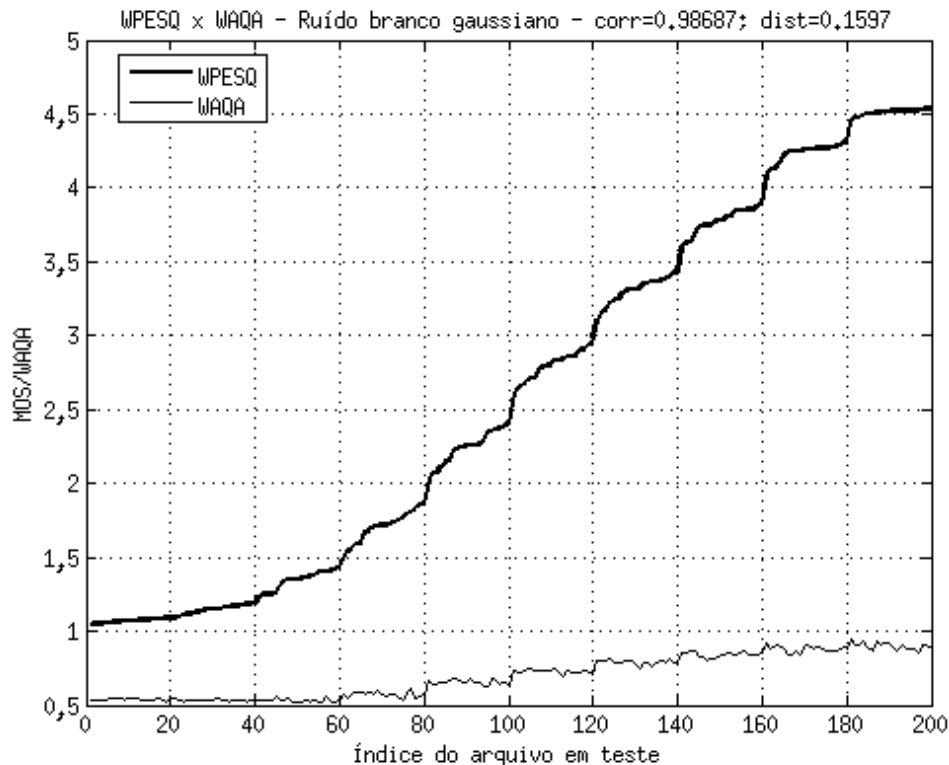


Figura 2.4: Comparação entre o WPESQ e o WAQA na avaliação de sinais de voz contaminados com ruído branco gaussiano.

Os parâmetros utilizados na geração do resultado apresentado na Fig. 2.4 foram: $Q_s = 200$; $wbps = 1250$; $wfam = 'db3'$; $pesos = 'exp2naoPonderado'$.

O resultado obtido foi bastante satisfatório. Além de a correlação ter sido muito elevada (aproximadamente 0,987), o desvio médio foi baixo. Além disso, por inspeção visual da Fig. 2.4 é possível perceber a eficiência do método, que reproduz todos os degraus correspondentes aos diferentes níveis de contaminação.

A fim de ilustrar, ao menos com um exemplo, o efeito do ajuste linear através do qual a medida de desvio médio ($dist$) é calculada, apresentam-se na Fig. 2.5 as notas do método de referência (WPESQ) e as notas WAQA ajustadas.

Outra ilustração que merece ser feita é a de uma métrica cuja avaliação dos rótulos dos coeficientes é feita em blocos. Inspirada na avaliação que será utilizada

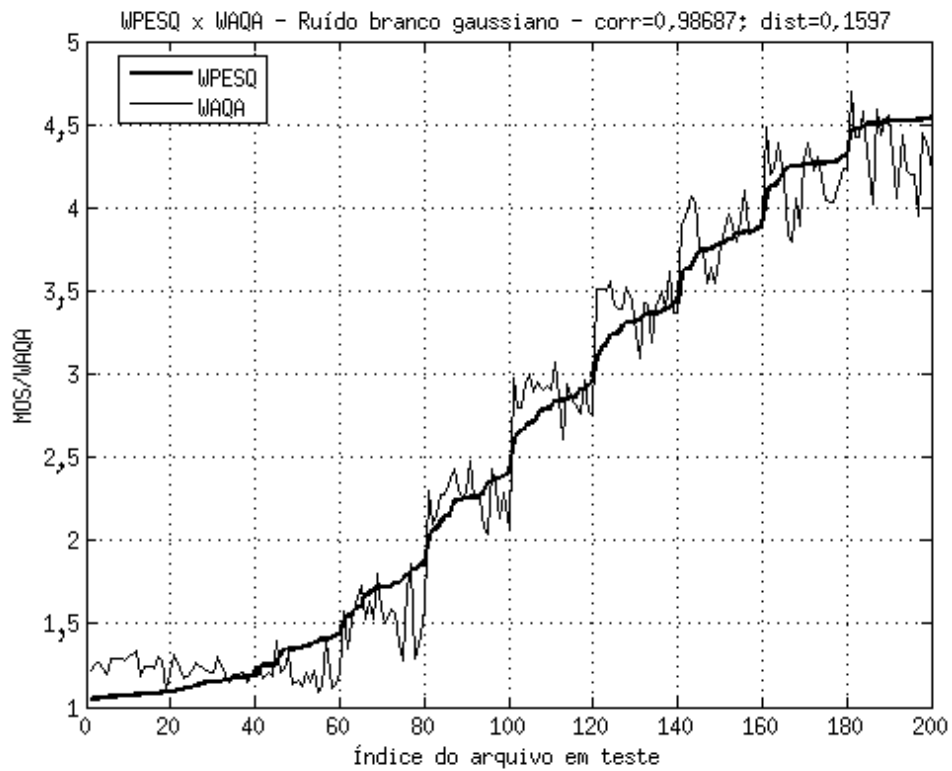


Figura 2.5: WAQA com ajuste linear.

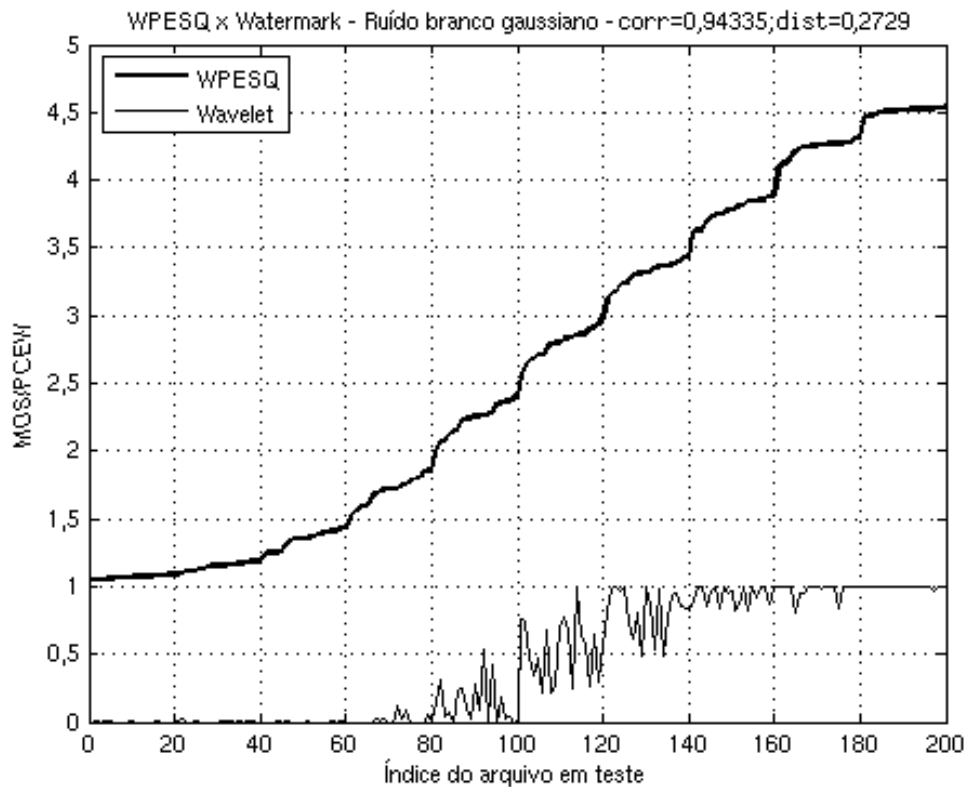


Figura 2.6: WAQA computado em bloco.

no Cap. 3, nela um certo número de coeficientes em seqüência (20 nessa ilustração) forma um bloco, que é considerado válido se um determinado percentual (60% nessa ilustração) dos rótulos for válido. O resultado obtido está representado na Fig. 2.6, na qual é possível constatar que houve perdas significativas em ambas as medidas de avaliação em relação aos resultados apresentados na Fig. 2.4. Essa avaliação em bloco foi realizada para todos os testes planejados, com diversos tamanhos de blocos e percentuais de acerto, e o empobrecimento dos resultados foi observado em todos os casos. Devido a isso, decidiu-se omitir os resultados correspondentes nos demais testes, já que ela nada acrescentaria à discussão do desempenho do WAQA.

2.5.2 Voz – Perda de Pacotes

Na Fig. 2.7 encontram-se representadas as notas obtidas com o WAQA e as notas do WPESQ no teste de avaliação de qualidade de sinais de voz degradados por perda de pacotes. A descrição dos diferentes níveis de perdas foi apresentada na Seção 2.4.1.

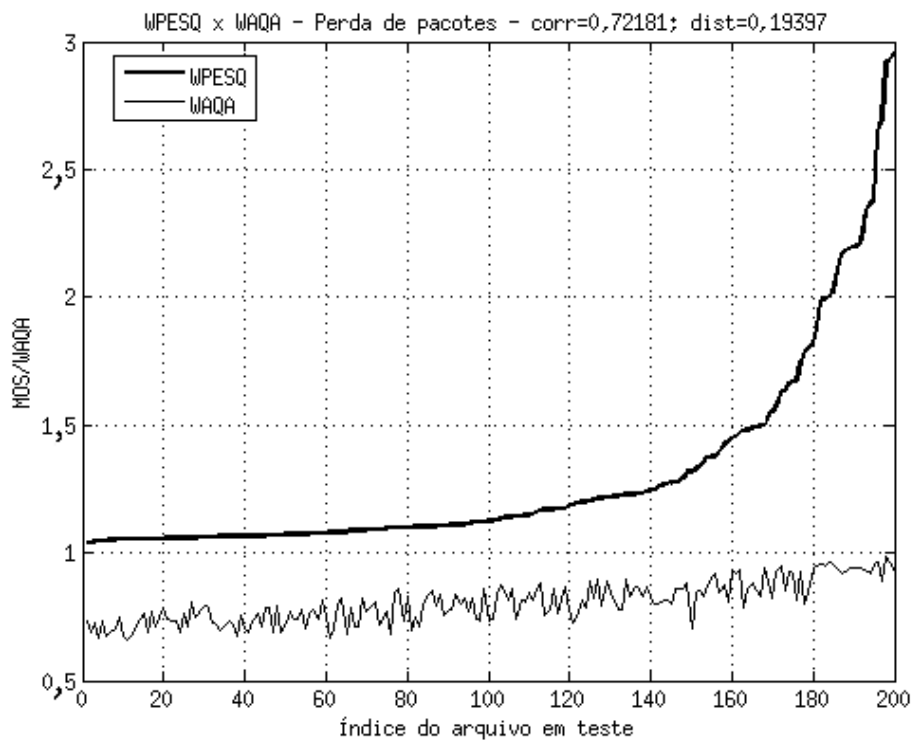


Figura 2.7: Comparação entre o WPESQ e o WAQA na avaliação de sinais de voz degradados por perda de pacotes.

Os parâmetros utilizados na geração do resultado apresentado na Fig. 2.7 foram: $Q_s = 200$; $wbps = 1250$; $wfam = 'db3'$; $pesos = 'exp2Ponderado'$.

O teste apresentou um resultado mediano, onde a correlação obtida não foi muito elevada, porém a medida de proximidade foi baixa. Além disso, por inspeção visual da Fig. 2.7 é possível constatar que o comportamento médio do WAQA acompanhou

a variação do nível de contaminação detectado pelo WPESQ, porém não tendo reproduzido uma curva idêntica: as notas do WPESQ parecem seguir uma curva exponencial, enquanto que as notas do WAQA parecem seguir uma reta. Poder-se-ia pensar em realizar um mapeamento exponencial a fim de maximizar a correlação, mas isso não será feito por não se enquadrar nos objetivos desta parte do trabalho.

2.5.3 Voz – Filtragem Passa-Baixas

Na Fig. 2.8 se encontram representadas as notas obtidas com o WAQA e as notas do WPESQ no teste de avaliação de qualidade de sinais de voz degradados por filtragem passa-baixas. A descrição das diferentes filtragens realizadas foi apresentada na Seção 2.4.1.

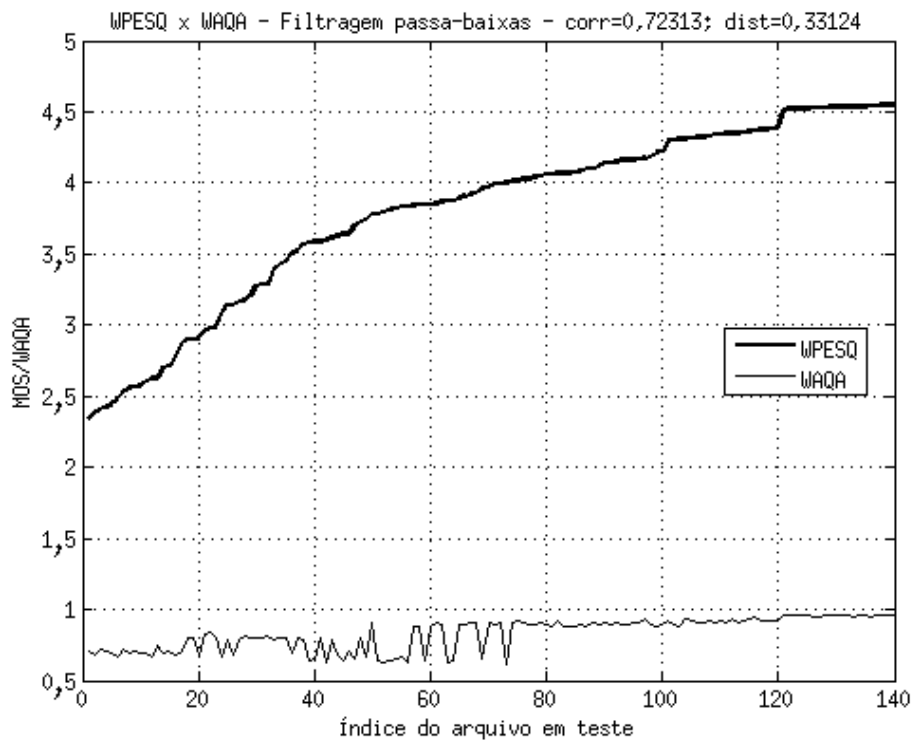


Figura 2.8: Comparação entre o WPESQ e o WAQA na avaliação de sinais de voz degradados por filtragem passa-baixas.

Os parâmetros utilizados na geração do resultado apresentado na Fig. 2.8 foram: $Q_s = 400$; $wbps = 1250$; $wfam = 'db3'$; $pesos = 'energiaNaoPonderado'$.

As métricas de medição do resultado deste teste ficaram muito próximas às do teste anterior (de degradação por perda de pacotes), com um desempenho um pouco inferior na medida $dist$. Assim como foi comentado no teste anterior, a correlação e a medida de proximidade indicam um desempenho mediano do WAQA na avaliação desse tipo de degradação.

Além disso, na Fig. 2.8 é possível notar uma mudança bastante abrupta nas notas, havendo um trecho inicial de notas predominantemente baixas, um segundo trecho onde notas baixas e altas se alternam, e finalmente um trecho de notas elevadas. Esse resultado indica que o avaliador não teve sensibilidade suficiente para realizar a medição de níveis intermediários de degradação.

2.5.4 Voz – Codificação G.729.1

Na Fig. 2.9 encontram-se representadas as notas obtidas com o WAQA e as notas do WPESQ no teste de avaliação de qualidade de sinais de voz degradados pelo codificador de voz G.719.1. A descrição das diferentes taxas de codificação utilizadas foi apresentada na Seção 2.4.1.

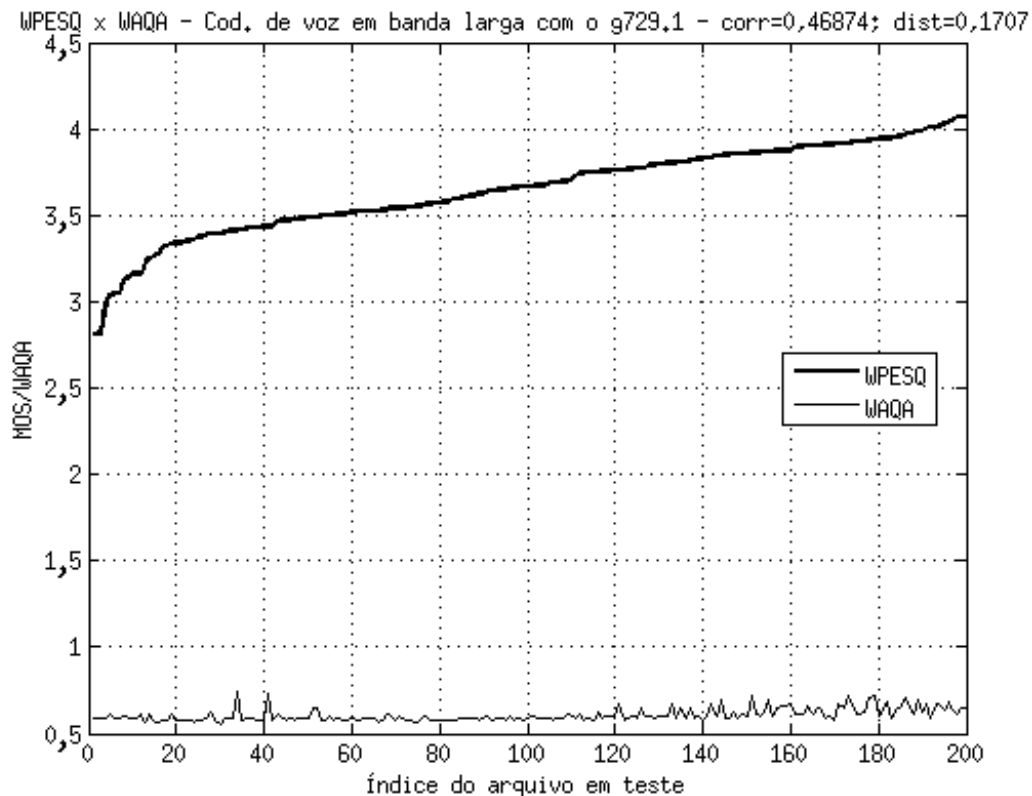


Figura 2.9: Comparação entre o WPESQ e o WAQA na avaliação de sinais de voz degradados pelo codificador G.729.1.

Os parâmetros utilizados na geração do resultado apresentado na Fig. 2.9 foram: $Q_s = 100$; $wbps = 1250$; $wfam = 'db3'$; $pesos = 'exp3naoPonderado'$.

O resultado obtido nesse teste foi insatisfatório, o que fica evidenciado tanto pelas medidas de correlação e proximidade quanto através da inspeção visual da Fig. 2.9.

É importante destacar que nesse teste o intervalo de notas do WPESQ consi-

derado foi menor que o considerado nos testes anteriores. A razão desse intervalo reduzido é a impossibilidade, descrita na Seção 2.4.1, de o codificador G.729.1 trabalhar com taxas mais baixas.

Fica demonstrada, mais uma vez, uma fragilidade do método na avaliação de qualidade sinais que sofreram determinados tipos de degradação. Vale ressaltar que essa fragilidade gera notas baixas incorretamente, porém nunca notas altas incorretamente. Nesse cenário, uma nota que indicasse alta qualidade poderia ser considerada válida, independentemente da degradação imposta ao sinal.

2.5.5 Voz – Codificação Speex

Na Fig. 2.10 encontram-se representadas as notas obtidas com o WAQA e as notas do WPESQ no teste de avaliação de qualidade de sinais de voz degradados pelo codificador de voz Speex. A descrição das diferentes taxas de codificação utilizadas foi apresentada na Seção 2.4.1.

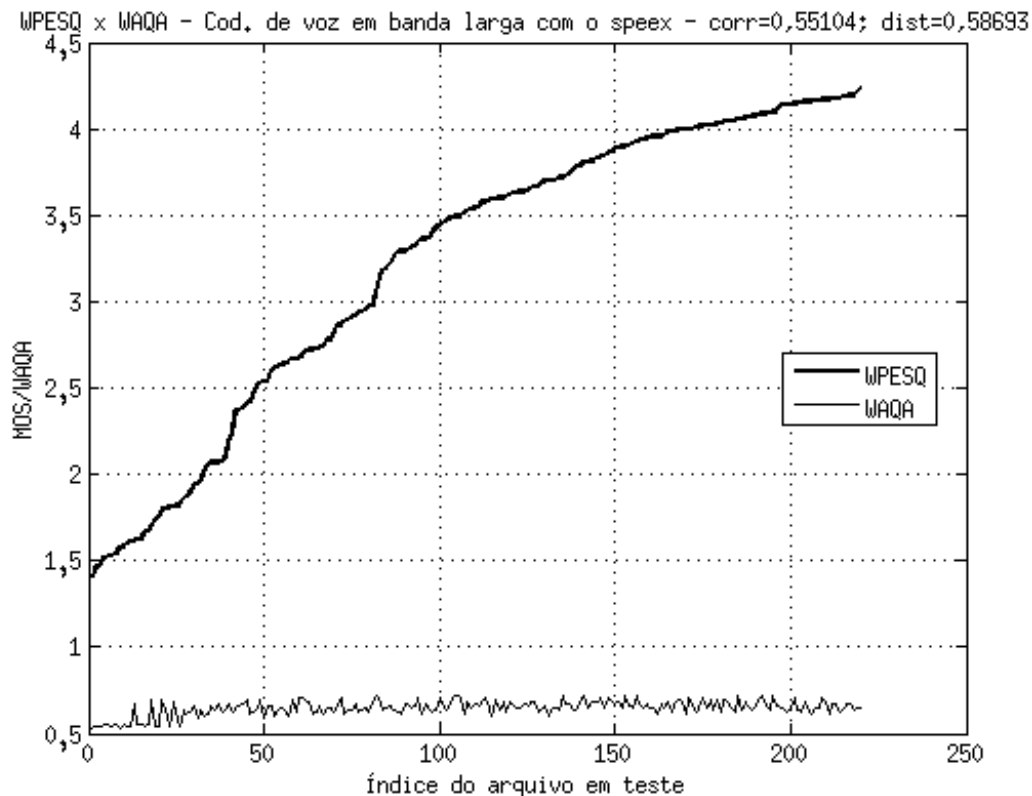


Figura 2.10: Comparação entre o WPESQ e o WAQA na avaliação de sinais de voz degradados pelo codificador Speex.

Os parâmetros utilizados na geração do resultado apresentado na Fig. 2.10 foram: $Q_s = 100$; $wbps = 1250$; $wfam = 'coif5'$; $pesos = 'exp3naoPonderado'$.

Os dois resultados obtidos nesse teste de codificação também foram insatis-

fatórios, tanto do ponto de vista das medidas objetivas (corr e dist), como pela inspeção visual dos resultados.

O maior objetivo desse teste era validar o resultado obtido no teste com o codificador G.729.1, visando a provar que o menor intervalo de notas obtido naquele teste não era o responsável pelo baixo desempenho do método estudado. Os resultados insatisfatórios obtidos corroboram essa hipótese, indicando, portanto, que o método é inadequado para realizar a medida de qualidade de sinais degradados por processos de codificação de voz.

2.5.6 Áudio – Codificação MP3

Na Fig. 2.11 encontram-se representadas as notas obtidas com o WAQA e as notas ODG do PEAQ no teste de avaliação de qualidade de sinais de áudio degradados pelo codificador MP3. A descrição das diferentes taxas de codificação utilizadas foi apresentada na Seção 2.4.2.

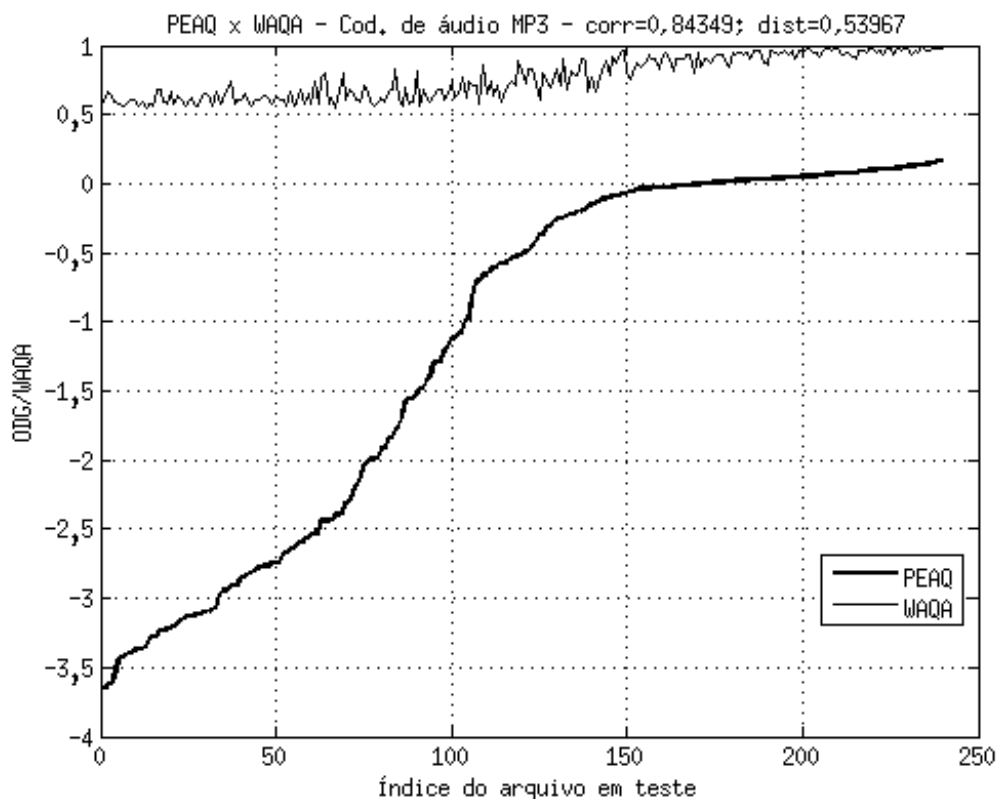


Figura 2.11: Comparação entre o PEAQ e o WAQA na avaliação de sinais de áudio degradados pelo codificador MP3.

Os parâmetros utilizados na geração do resultado apresentado na Fig. 2.10 foram: $Q_s = 100$; $wbps = 1250$; $wfam = 'coif5'$; $pesos = 'exp3naoPonderado'$.

O resultado desse teste pode ser considerado satisfatório, pois se obteve uma

correlação acima de 0,8 ($\approx 0,84$); porém, a medida *dist* mais elevada que a obtida em quase todos os outros testes (acima de 0,55). Por inspeção visual da Fig. 2.11, é possível notar que o WAQA acompanhou, na média, o comportamento do PEAQ; entretanto houve uma grande variação entre notas que deveriam ser próximas, i.e., o resultado aparentemente foi acurado, porém pouco preciso.

2.5.7 Resultados finais

Utilizando todos os resultados parciais até agora obtidos, o seguinte conjunto de parâmetros foi gerado, tomando como base ou o maior número de ocorrências registradas (no caso dos parâmetros *wfam*, *pesos* e *wbps*) ou o valor testado que ficou mais próximo da média (no caso do parâmetro Q_s). São eles: $Q_s = 200$; *wbps* = 1250; *wfam* = ‘db3’; *pesos* = ‘exp2naoPonderado’.

Com eles, foram obtidas as medidas objetivas apresentadas na Tab 2.2.

Tabela 2.2: Resultados finais obtidos com os parâmetros que apresentaram melhor desempenho médio.

Teste realizado	corr	dist
WPESQ x WAQA – Ruído branco gaussiano	0,985	0,163
WPESQ x WAQA – Perda de pacotes	0,711	0,198
WPESQ x WAQA – Filtragem passa-baixas	0,699	0,360
WPESQ x WAQA – Cod. de voz com o G.729.1	0,433	0,178
WPESQ x WAQA – Cod. de voz com o speex	0,476	0,628
PEAQ x WAQA – Cod. de áudio MP3	0,781	0,666

A tabela mostra que o conjunto de parâmetros escolhido apresenta um desempenho similar ao obtido nos testes com os melhores parâmetros de cada teste.

2.6 Conclusões

Com a análise de todos os resultados, destaca-se que:

- A eficácia do WAQA variou bastante conforme o tipo de degradação avaliada. Isso demonstra que ele deve ser utilizado apenas na avaliação de sinais que sofreram alguns tipos determinados de degradação.
- Merece grande destaque a correlação muito alta apresentada pelo WAQA na avaliação da degradação causada pela adição de ruído branco gaussiano, mesmo sem nenhum tipo de ajuste de curvas.

- A avaliação dos sinais de áudio que sofreram codificação perceptiva também foi razoavelmente bem sucedida, apesar de não ter tido um desempenho tão espetacular quanto o encontrado nos testes com ruído branco.
- O teste de perda de pacotes apresentou um resultado satisfatório, apesar de indicar que um mapeamento alternativo poderia gerar resultados ainda melhores.
- Os testes de filtragem passa-baixas e de codificação de voz apresentaram resultados que indicam a incapacidade do WAQA em avaliar certos tipos de degradação. Vale destacar que a fragilidade na avaliação do WAQA sempre se dá com a geração de notas baixas para arquivos que deveriam ter notas mais altas, o que fica evidente na Fig. 2.8, que mostra vales claros no gráfico. Essa característica permite que o WAQA não precise ser descartado como possível entrada para métodos mais robustos que utilizam diversas métricas para gerar a sua avaliação (e.g. PEAQ). Uma boa nota do WAQA poderia indicar, para o método mais robusto, uma boa qualidade do sinal que está sob avaliação.

Capítulo 3

Marcas d'água digitais na avaliação de qualidade de sinais de áudio e voz

3.1 Introdução

De forma simplista, é possível definir uma marca d'água como uma informação a ser embutida em outra. Em sinais de áudio, busca-se que as marcas d'água inseridas não possam ser detectadas pelo ouvinte, i.e., que para ele o sinal original e o sinal marcado soem idênticos. Nesse contexto, num sinal de áudio só seria possível detectar a marca inserida com o processamento daquele sinal.

A idéia de utilizar marcas d'água digitais como instrumento para a aferição da qualidade de sinais vem do seguinte raciocínio: após um sinal marcado passar por um processo de degradação, se for possível medir com precisão a degradação da marca e ela estiver correlacionada com a degradação do sinal, será possível inferir a degradação do sinal através da degradação da marca. A grande vantagem dessa abordagem é poder construir um método de avaliação de qualidade sem referência.

O objetivo deste capítulo é apresentar e testar os métodos da literatura que utilizam marcas d'água digitais com base na DWT para fazer a avaliação de qualidade de sinais de áudio e voz [24] [25] [26] [27]. Nesse capítulo, utilizar-se-á o termo *DWT-watermarking* para fazer referência a esses métodos.

É preciso dizer que na literatura é possível encontrar outros métodos de avaliação de qualidade de sinais de áudio e voz utilizando marcas d'água digitais que não utilizam a DWT [36] [37]. Ambos os métodos realizam a inserção da marca d'água com auxílio da DCT e somente necessitam conhecer a marca inserida para efetuar a avaliação de qualidade.

Em [36], é mostrado que—em quatro sinais de áudio degradados com três taxas de

compressão MP3—a marca inserida é degradada juntamente com o sinal. Os autores não buscaram estabelecer nenhuma relação entre uma dada medida de degradação da marca d’água e um dado nível de degradação imposto (talvez porque os gráficos mostravam medidas muito diferentes para níveis equivalentes de degradação).

Em [37], são utilizados apenas dois sinais de áudio para a realização dos testes (apesar da alegação de que o método proposto poderia ser utilizado tanto para áudio como para voz). Os dois sinais de áudio são submetidos a três degradações simultâneas (ruído branco, *jitter* e compressão). Os autores apenas mostraram: que quanto mais forte a marca inserida, menor a SNR obtida; e que os dois sinais de teste degradados apresentaram notas altas ($\approx 0,9$ num intervalo de 0 a 1) com o método. Não foi efetuado nenhum estudo da correlação entre as notas e alguma medida subjetiva ou objetiva de qualidade, ou mesmo da variação das notas conforme o aumento da degradação.

Conforme o exposto, fica clara a dificuldade de comparar diretamente os resultados destes trabalhos mais recentes da literatura com o *DWT-watermarking*. Apesar da implementação e análise desses métodos ser possível, não será efetuada por ultrapassar o escopo deste trabalho, que seria o estudo e aprimoramento dos métodos com base na DWT.

Voltando a tratar do *DWT-watermarking*, de acordo com os resultados obtidos no Cap. 2, seria possível determinar a qualidade dos sinais de áudio/voz através da comparação direta entre os coeficientes da DWT dos sinais original e degradado. A abordagem que aqui será apresentada consiste na avaliação da qualidade através da comparação entre uma marca d’água inserida no sinal original e o que resta dela no sinal após ele ter sofrido algum tipo de degradação.

A primeira parte do capítulo apresentará o método e os conceitos teóricos envolvidos, muitos dos quais já foram apresentados e explicados no Cap. 2. Em paralelo, serão apresentados e comentados os testes reportados na literatura. Em seguida, serão apresentados os testes realizados e resultados obtidos neste trabalho e se fará uma comparação com aqueles da literatura. O capítulo termina traçando considerações finais sobre o método estudado.

3.2 Avaliação de qualidade usando marcas d’água digitais

O processo de avaliação de qualidade usando marcas d’água digitais com base na DWT se encontra representado na Fig. 3.1. O sinal é decomposto usando a DWT, para em seguida ser modificado no domínio da DWT (insere-se a marca d’água nos coeficientes da DWT), e retornado para o seu domínio original para sofrer de-

gradações. Para a avaliação da marca d'água do sinal degradado utiliza-se novamente o domínio da DWT. Todo o processo de avaliação será detalhado a seguir.

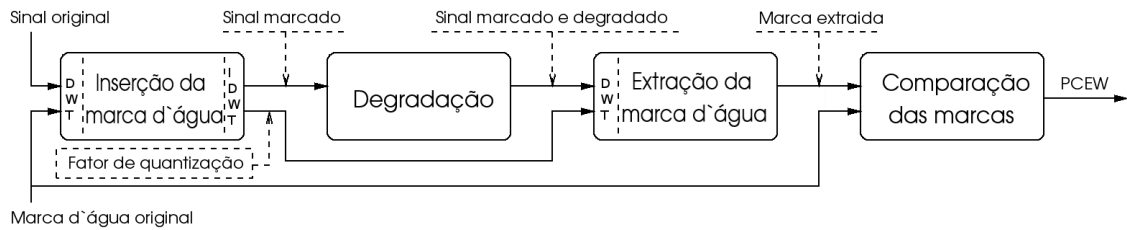


Figura 3.1: Diagrama de operação da avaliação de qualidade usando marcas d'água digitais construídas através da DWT. No diagrama, a DWT é utilizada nos blocos de inserção e extração da marca d'água.

Inserção da marca d'água: Este bloco recebe o sinal original e a marca que se deseja inserir e produz o sinal marcado (sinal com marca d'água inserida) e o fator de quantização (Q_s —constante já citada no Cap. 2 e que doravante receberá esse nome, conforme a literatura) que foi utilizado para a inserção da marca.

Degradação: Bloco de degradação do sinal marcado com algum tipo de contaminação.

Extração da marca d'água: Este bloco recebe o sinal marcado e degradado e o fator de quantização utilizado para a inserção da marca e devolve a marca d'água encontrada nesse sinal.

Comparação das marcas: Este bloco compara a marca d'água original com a marca d'água extraída e produz a nota de saída (PCEW — apresentada posteriormente nesse capítulo).

As etapas expostas serão detalhadas a seguir.

3.2.1 Inserção de marcas d'água digitais usando a DWT

O bloco de inserção da marca d'água se encontra ilustrado em detalhes na Fig. 3.2. Nele é possível observar as diversas etapas envolvidas no processo.

Para a compreensão do funcionamento do bloco ilustrado na Fig. 3.2, os sub-blocos presentes na figura terão o seu funcionamento detalhado, a começar pelo de entrada.

Inserir marca com determinado Q_s

Conforme apresentado na Seção 2.3, é possível definir um rótulo para cada coeficiente da DWT através de sua divisão por um valor fixo Δ_N (Eq. (2.9)), que por sua vez

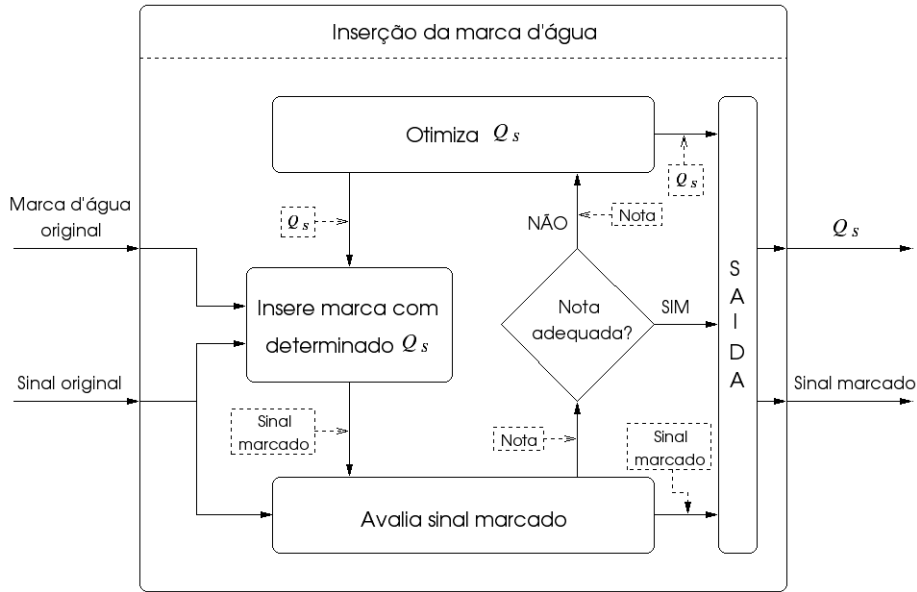


Figura 3.2: Bloco de inserção da marca d'água em detalhes.

é calculado em cada nível a partir dos coeficientes de maior módulo e de um valor fixo Q_s (Eq. (2.8)).

Sabendo-se disso, o processo de inserção da marca d'água com um determinado Q_s pode ser facilmente entendido com sua divisão em 4 etapas:

1. Recebem-se o sinal original, o fator de quantização (Q_s) e um vetor binário de comprimento arbitrário que será tomado como a marca d'água original.
2. A fim de aumentar a robustez do método, insere-se uma redundância em cada um dos bits de marca (v. Seção 2.3, logo antes da apresentação da Eq. (2.8)). Assim, se a marca originalmente era 0101001 e se impõe uma redundância de 3, a nova marca equivalente será 000111000111000000111.
3. Escolhem-se os coeficientes que serão utilizados para a inserção da marca conforme descrito na Seção 2.3.
4. Cada bit da marca passa a definir o rótulo do coeficiente a que corresponde. Se o rótulo atual do coeficiente não for igual ao seu bit de marca equivalente, adiciona-se ao módulo do coeficiente o mesmo Δ_N utilizado no cálculo do seu rótulo.

Ao término desse processo, o sinal está com todos os rótulos dos coeficientes iguais aos bits de marca d'água.

O grande problema associado à inserção da marca através das 4 etapas é saber quão “forte” ela deve ser, i.e., o quanto se deve modificar os coeficientes da DWT. De

fato, se deseja saber qual deve ser o Q_s utilizado para gerar um Δ_N que seja suficientemente grande para medir a qualidade do sinal, e ao mesmo tempo suficientemente pequeno para que a marca inserida não degrade muito o sinal original.

A medida proposta para contornar esse problema é a utilização de um processo de otimização em que as 4 etapas são realizadas com diferentes valores de Q_s até que o sinal marcado apresente um determinado grau de degradação. Os demais sub-blocos da Fig. 3.2 que serão apresentados a seguir são os responsáveis por essa otimização.

Avalia sinal marcado

Ao receber o sinal marcado e o sinal original para comparação, este sub-bloco utiliza algum avaliador objetivo de qualidade para medir a distorção imposta pela inserção da marca. Na literatura, foram testadas algumas medidas como SNR e PSNR [25] e nos trabalhos mais recentes (e.g. [27]), o PESQ (v. Seção 1.5.2) com sinais de fala.

Para essa categoria de sinais, o desempenho do bloco utilizando o WPESQ (versão de banda estendida do PESQ) foi, conforme testes informais realizados, muito superior ao que se obteve com uso das comparações de cunho mais matemático (SNR e PSNR). Ao utilizar o WPESQ o impacto da marca sobre os sinais era muito mais uniforme no conjunto dos sinais. Esse resultado obtido era esperado pois o WPESQ, é um método perceptivo padronizado que possui alta correlação com o julgamento humano. Após esses testes preliminares informais, confirmou-se a decisão inicial de se utilizar o WPESQ nesse bloco para avaliar o desempenho do *DWT-watermarking* para os sinais de voz.

Enquanto os trabalhos da literatura se restringiram quase sempre à avaliação de sinais de voz, neste capítulo o *DWT-watermarking* também será utilizado para a avaliação de qualidade de sinais de áudio. Para tal, é necessário utilizar nesse sub-bloco de avaliação do sinal marcado um método objetivo que seja capaz de avaliar a qualidade de sinais de áudio. Dado o bom resultado apresentado pelo WPESQ, optou-se por também utilizar um método de fundamentação perceptiva padronizado pela ITU, o PEAQ (v. Seção 1.5.1).

Em resumo, para este bloco utilizar-se-á o WPESQ para avaliação sinais de voz e o PEAQ para avaliação de sinais de áudio.

Nota adequada?

Nesse bloco a nota do método utilizado no bloco anterior é avaliada com o fim de se decidir sobre a adequação da marca aplicada. Para tal, dois aspectos são avaliados:

1. Verificar se a marca não deteriorou demasiadamente o sinal original.

2. Avaliar se a marca original pode ser recuperada adequadamente a partir do sinal marcado e do fator de quantização Q_s , i.e., se ela altera o sinal “suficientemente”.

Para a primeira avaliação compara-se a nota obtida no bloco anterior com alguma nota pré-definida. Caso a nota obtida esteja distante da nota pré-definida, o que indica que a qualidade não corresponde ao valor estabelecido, define-se que o fator de quantização utilizado não é adequado e a nota obtida é enviada para o bloco de otimização do Q_s .

Na segunda avaliação simula-se a recuperação da marca, i.e., utiliza-se um bloco de extração como o da Fig. 3.1 para obter a marca sem que o sinal tenha sofrido nenhuma outra distorção a não ser a imposta pela inserção da própria marca. Caso a marca recuperada seja muito diferente da marca original, envia-se essa diferença para o bloco de otimização do Q_s .

Os valores-alvo utilizados para as notas mínimas dos métodos (primeiro quesito) e para distorção máxima da marca extraída (segundo quesito) foram: 4,2 para o WPESQ, conforme indicado em [27]; $-0,5$ para o PEAQ, valor arbitrado após a realização de testes informais; e 10% para a distorção máxima, valor que será explicado a seguir.

Apesar de a duração dos sinais de voz utilizados não ser uma informação explícita na literatura, em [27] há referência a sinais com 1 minuto de duração, muito maior que a dos sinais que estão sendo utilizados nesse trabalho e que foram descritos no Cap. 2. Com isso, utilizaram-se muitos menos bits de marca d'água que nos trabalhos da literatura. Com poucos bits, o valor de 1% recomendado na literatura para a distorção máxima da marca extraída não pode ser utilizado, pois esse limiar seria facilmente ultrapassado. Optou-se, então, com bom resultado, por um nível de distorção máxima mais elevado (10%).

Otimiza Q_s

Como nos trabalhos da literatura não era explicitada a forma de otimização de Q_s , desenvolveu-se um algoritmo *ad hoc*, cujo pseudocódigo é apresentado no Apêndice A, e cujo fluxograma se vê na Fig. 3.3.

O valor de Q_s é ajustado de acordo com a diferença entre a nota obtida e a nota desejada: quanto maior a diferença, maior o passo¹ de ajuste utilizado (v. pseudocódigo no Apêndice A). Para garantir a convergência da otimização utilizam-se limiares ajustáveis para máximo e mínimo Q_s , implementados da seguinte forma: a cada vez que o bloco de ajuste reduz Q_s , o valor anterior (Q_s original) é tomado

¹Os passos foram escolhidos de forma a minimizar a chance de a otimização ficar presa num ciclo de valores.

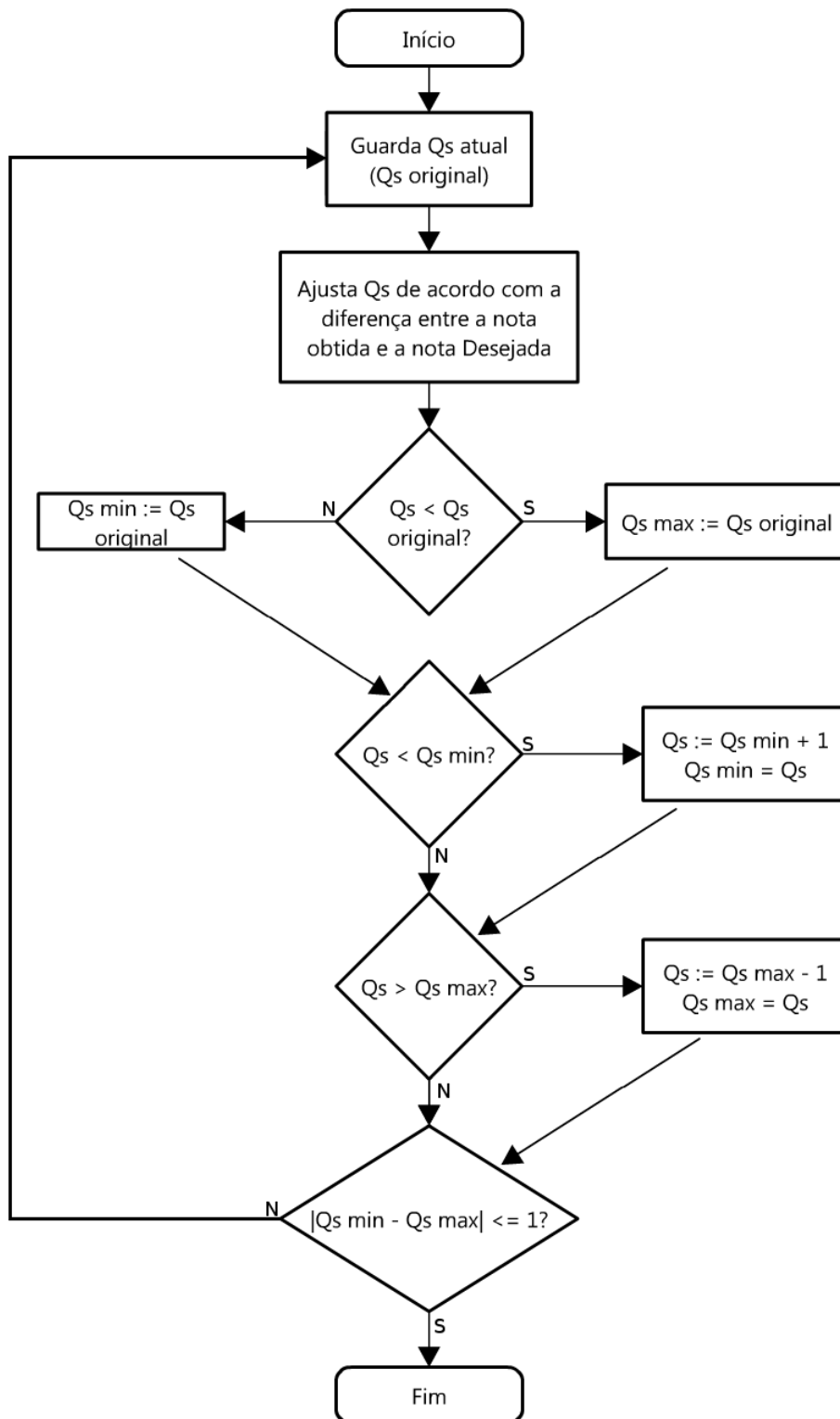


Figura 3.3: Fluxograma de otimização de Q_s .

como o Q_s máximo; analogamente, redefine-se Q_s mínimo a cada vez que o bloco de ajuste aumenta Q_s . Com esse algoritmo foi possível reduzir substancialmente a carga computacional do método, se comparado ao método de busca linear (partindo com $Q_s = 1$, incrementando de 1 em 1).

Saída

Este sub-bloco apenas fornece as saídas do bloco quando a otimização termina.

3.2.2 Degradação dos sinais de áudio/voz

Esse bloco realiza as mesmas degradações que estão descritas em detalhes na Seção 2.4. Relembrando, temos para os sinais de voz as seguintes degradações:

- ruído aditivo branco gaussiano,
- perda de pacotes,
- filtragem passa-baixas,
- codificação de voz;

e para os sinais de áudio, a degradação pelo codificador MP3.

Vale abrir aqui uma discussão sobre algumas degradações utilizadas em [26] [27].

Codificação MP3

Em [27], um codificador MP3 foi utilizado para realizar a degradação dos sinais de voz que estavam sendo avaliados pelo *DWT-watermarking* e pelo PESQ.

Cabe aqui lembrar que o PESQ não é adequado para avaliação de sinais contaminados por ruído oriundo de processos de codificação perceptiva de áudio como o MP3. O codificador MP3 não fazia parte da categoria de codificadores utilizados para calibração do PESQ, e portanto a medida de qualidade do PESQ para os sinais contaminados por esse codificador não é confiável. Não se poderia, portanto, dizer da eficiência de um novo método de avaliação de qualidade como o *DWT-watermarking* através da comparação de seus resultados com os resultados do PESQ na avaliação de sinais contaminados através da codificação MP3.

Filtragem passa-baixas

Para avaliar o impacto da filtragem passa-baixas, foi proposto em [27] e [26] que se utilizasse uma série de filtros passa-baixas com frequência de corte de 1 kHz até 29 kHz, em passos de 4 kHz, para contaminar os sinais; experimento não adequado.

Mesmo que se estivessem utilizando sinais de voz amostrados a mais de 58 kHz (taxa de amostragem mínima para a representação de sinais a 29 kHz), o que já constituiria um conjunto de sinais de voz pouco usual, o PESQ somente avaliaria os primeiros 4 kHz do espectro. Dessa forma, todas as filtragens acima de 5 kHz dariam as mesmas notas quando avaliadas no PESQ, inviabilizando o teste. Os autores dos trabalhos citados encontraram exatamente os mesmos resultados para as filtragens superiores a 4 kHz, sem entretanto identificar a causa dessa ocorrência.

3.2.3 Extração da marca d'água

O processo de extração da marca, dado o sinal marcado e o fator de quantização (Q_s), é bastante simples e já foi explicado no Cap. 2.

A única diferença a ser destacada é a utilização da redundância nos bits de marca, conforme apresentado nas Seções 2.3 e 3.2.1. Com a redundância, o bit será decidido pela seguinte regra:

$$W_{\text{bit}} = \begin{cases} 0 & \text{se } S/N \leq 0,4 \\ 1 & \text{se } S/N \geq 0,6 \\ 2 & \text{nos demais casos} \end{cases}, \quad (3.1)$$

onde S representa a soma dos bits obtidos em cada seqüência de bits repetidos, e N o número de bits que cada bloco possui, i.e., a redundância utilizada.

Destaca-se que com essa proposta alguns bits poderão, durante a extração, receber um rótulo inválido (2). Esses bits de rótulo inválido forçosamente receberão avaliação negativa.

3.2.4 Comparação das marcas

A nota PCEW (*Percentage of Correctly Extracted Watermark bits* — percentual de bits de marca d'água extraídos corretamente) é obtida através do cálculo simples do percentual de acertos dos bits da marca d'água extraída em relação aos bits da marca d'água original. Por exemplo, se há uma coincidência de 70% entre os bits da marca original e da marca extraída, o PECW resultante será de 0,7.

3.3 Testes e Resultados

Conforme discutidos na seção 3.2.2, nem todos os testes da literatura foram considerados adequados para que se fizesse uma comparação de desempenho e resultados. Assim, apresentar-se-ão todos os resultados dos testes realizados neste trabalho e serão feitas comparações com os resultados da literatura quando isso for possível.

Além disso, para análise dos resultados serão utilizadas as mesmas medidas apresentadas e descritas na Seção 2.4.4.

3.3.1 Parâmetros utilizados

Os parâmetros utilizados no teste (descritos detalhadamente na Seção 2.4.3) foram:

- parâmetros citados na literatura: $fTarVal$ (nota mínima do sinal marcado) usando o WPESQ = 4,2;
- parâmetros utilizados nos testes feitos no Cap. 2: N (número de níveis usado na DWT) = 10;
- parâmetros que se destacaram nos testes feitos no Cap. 2: $wbps$ (bits de marca d'água por segundo) = 50²; $wfam$ (família *wavelet*) = 'db3'; $pesos$ = 'exp2naoPonderado';
- parâmetros arbitrariamente escolhidos: $fTarVal$ (nota mínima do sinal marcado) usando o PEAQ = -0,5, valor apresentado na Seção 3.2.1; $watRed$ (redundância utilizada nos bits da marca d'água) = 25; $maxWatDist$ (distorção máxima da marca recuperada) = 10%.

Vale relembrar o parâmetro Q_s , não pode ser pré-fixado por ter, no método *Watermaking*, a função de ajustar a “força” da marca inserida de acordo com os limites impostos por $fTarVal$ e $maxWatDist$.

Além disso, fez-se uma modificação no parâmetro $wbps$ para os testes dos sinais de áudio devido à sua maior taxa de amostragem. Como a taxa de amostragem deles é 3 vezes maior que a taxa utilizada nos sinais de voz (48 kHz x 16 kHz), utilizou-se um número 3 vezes maior para o $wbps$, 150.

3.3.2 Voz – Ruído

Na Fig. 3.4 encontram-se representadas as notas obtidas com o *DWT-watermarking* e as notas do WPESQ no teste de avaliação de qualidade de sinais de voz contaminados com ruído branco gaussiano. A descrição dos diferentes contaminações foi apresentada na Seção 2.4.1.

Assim como no Cap. 2, os resultados obtidos foram bastante satisfatórios, com alto valor de correlação e baixa medida de distância (valores na Fig. 3.4). Também como no Cap. 2, os degraus correspondentes aos diferentes níveis de degradação foram reproduzidos. Isso pode ser mais facilmente visto na Fig. 3.5, que ilustra o ajuste linear feito para o cálculo da medida *dist*.

²Com a redundância de 25 chega-se a 1250, valor utilizado no Cap. 2.

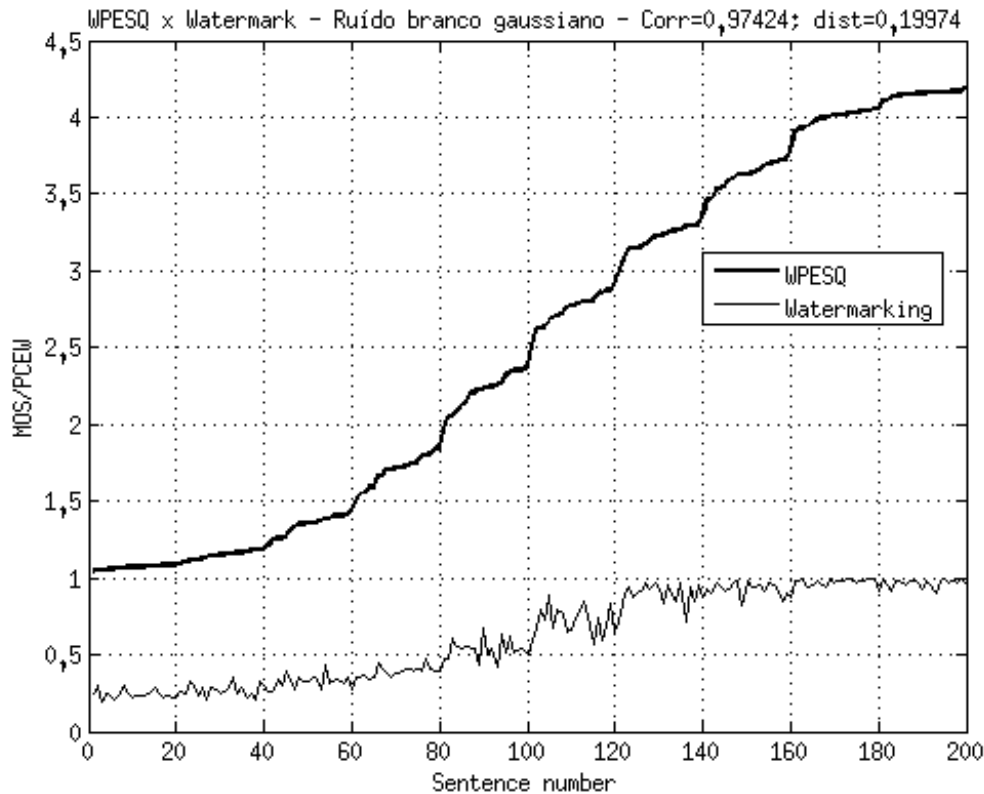


Figura 3.4: Comparação entre o WPESQ e o *DWT-watermarking* na avaliação de sinais de voz contaminados com ruído branco gaussiano.

Apesar de a Fig. 3.5 apresentar um comportamento um pouco mais “ruidoso” que a sua equivalente no Cap. 2 (Fig. 2.5), e de as medidas escolhidas para a avaliação dos resultados também terem sido ligeiramente menos expressivas, pode-se considerar que o desempenho obtido com o *DWT-watermarking* foi equivalente ao ótimo desempenho obtido com o WAQA para avaliação de qualidade de sinais contaminados com ruído.

Esse resultado é significativo, principalmente se considerarmos que o *DWT-watermarking* não é um método de avaliação de qualidade com referência, já que a única informação sobre o sinal original de que ele necessita é o fator de quantização (Q_s) utilizado na inserção da marca (além da marca utilizada, que pode ser fixada previamente).

3.3.3 Voz – Perda de Pacotes

Na Fig. 3.6 encontram-se representadas as notas obtidas com o *DWT-watermarking* e as notas do WPESQ no teste de avaliação de qualidade de sinais de voz degradados por perda de pacotes. A descrição dos diferentes níveis de perda foi apresentada na Seção 2.4.1.

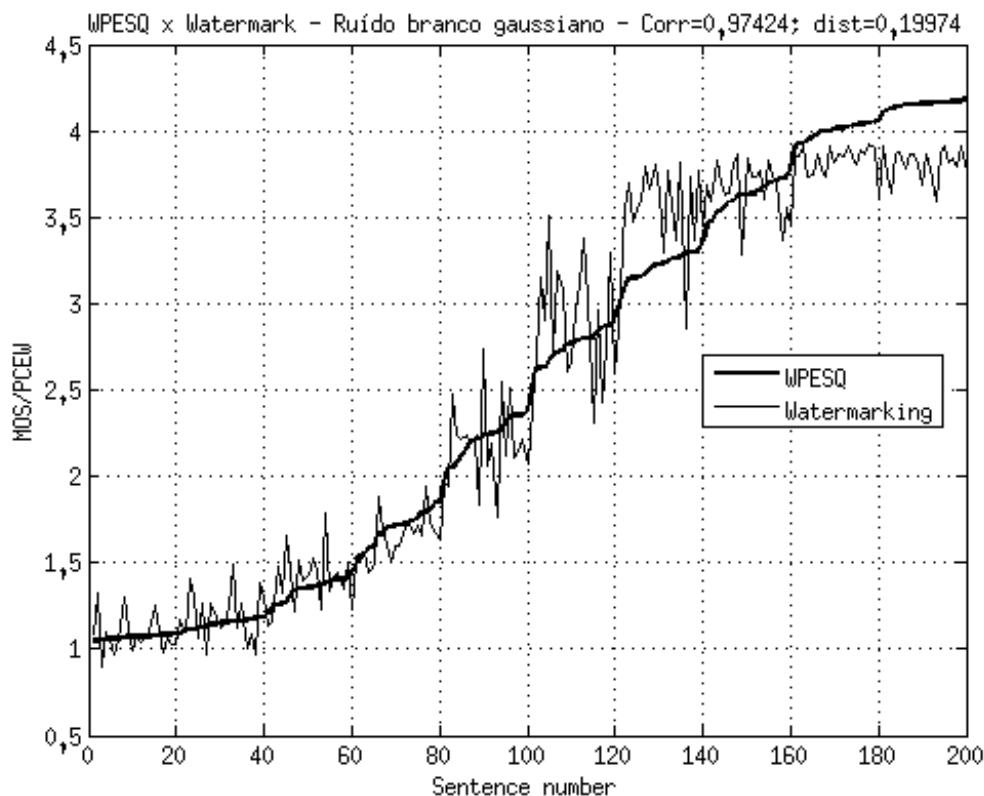


Figura 3.5: *DWT-watermarking* com ajuste linear.

Novamente o resultado obtido foi similar ao encontrado no Cap. 2, ligeiramente pior. A correlação obtida foi medíocre, porém a medida de proximidade foi boa (valor baixo). Além disso, por inspeção visual da Fig. 3.6 é possível observar que o comportamento médio do *DWT-watermarking* acompanhou a variação do nível de contaminação detectado pelo WPESQ, porém não com a mesma “intensidade”: as notas do WPESQ parecem seguir uma curva exponencial, enquanto que as notas do *DWT-watermarking* parecem seguir uma reta. Vale lembrar a possibilidade de se realizar um mapeamento exponencial das notas do *DWT-watermarking* a fim de tentar maximizar a correlação, o que ainda não é o objetivo desta parte do trabalho.

3.3.4 Voz – Filtragem Passa-Baixas

Na Fig. 3.7 se encontram representadas as notas obtidas com o *DWT-watermarking* e as notas do WPESQ no teste de avaliação de qualidade de sinais de voz degradados por filtragem passa-baixas. A descrição das diferentes filtrações realizadas foi apresentada na Seção 2.4.1.

Ao contrário do ocorrido nos testes anteriormente relatados (ruído e perda de pacotes), no teste com filtragem passa-baixas os resultados obtidos com o *DWT-watermarking* foram superiores aos obtidos com o WAQA. As medidas de correlação

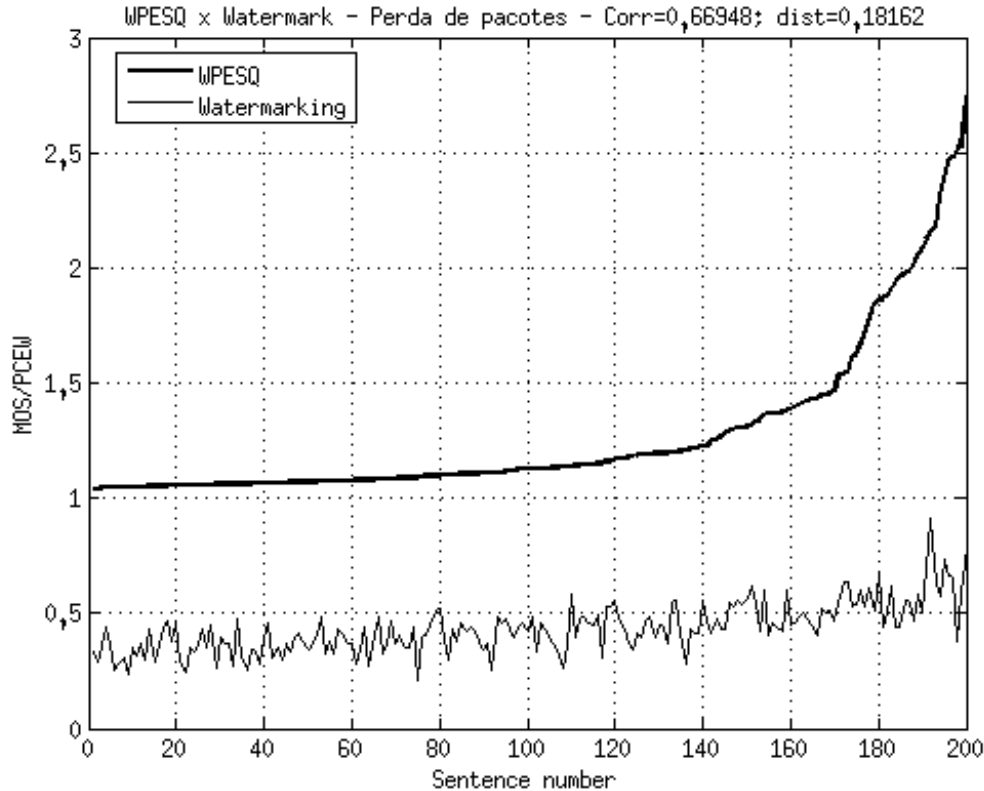


Figura 3.6: Comparação entre o WPESQ e o *DWT-watermarking* na avaliação de sinais de voz degradados por perda de pacotes.

e distância obtidas indicam que o *DWT-watermarking* é mais adequado para a avaliação de qualidade de sinais degradados com esse tipo de contaminação.

O resultado do *DWT-watermarking* foi superior ao WAQA, pois enquanto na Fig. 2.8 é possível identificar uma região inicial de notas baixas, uma região intermediária de notas oscilando entre baixas e elevadas, e uma região de notas elevadas, a Fig. 3.7 apresenta uma curva com um crescimento mais suave, mesmo havendo ainda alguma oscilação.

Ao contrário do WAQA, o *DWT-watermarking* manipula os coeficientes *wavelet* diretamente e otimiza o fator de quantização Q_s a fim de garantir uma nota elevada para a decodificação. Essa diferença pode gerar um avaliador mais sensível, o que explicaria esse melhor desempenho.

3.3.5 Voz – Codificação G.729.1 e Speex

Nas Figs. 3.8 e 3.9 se encontram representadas as notas obtidas com o *DWT-watermarking* e as notas do WPESQ no teste de avaliação de qualidade de sinais de voz degradados pelos codificadores de voz G.729.1 e Speex, respectivamente. A descrição das diferentes taxas de codificação utilizadas em cada um dos codificadores

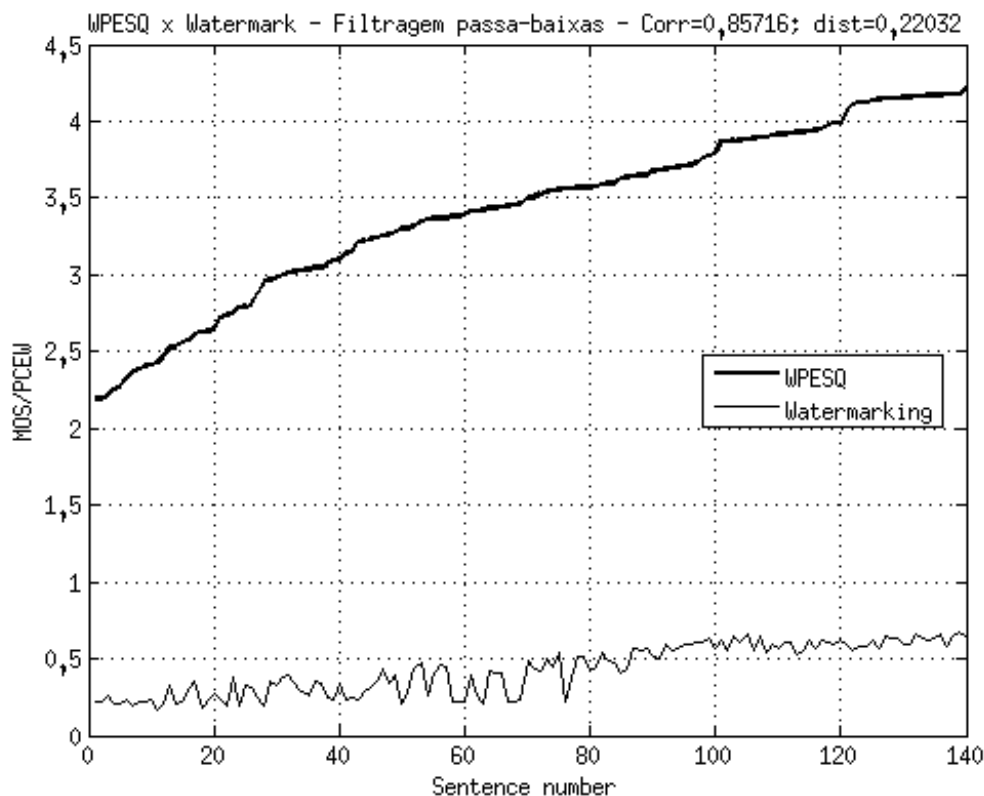


Figura 3.7: Comparação entre o WPESQ e o *DWT-watermarking* na avaliação de sinais de voz degradados por filtragem passa-baixas.

foi apresentada na Seção 2.4.1.

Os resultados encontrados confirmam o que já havia sido identificado no Cap. 2, no teste do WAQA: que não é possível realizar a avaliação de qualidade de sinais degradados por processos de codificação de voz através da avaliação dos coeficientes da DWT. Ambos os testes evidenciaram um resultado pífio, tanto em relação à medida de correlação quanto em relação à medida de distância. É possível concluir que, apesar de o *DWT-watermarking* ser mais robusto que o WAQA (conforme discutido na Seção 3.3.4), existem alguns tipos de degradações que, mesmo nos seus níveis mais leves, destroem a marca inserida, impedindo assim a avaliação da qualidade.

3.3.6 Áudio – Codificação MP3

Na Fig. 3.10 encontram-se representadas as notas obtidas com o *DWT-watermarking* e as notas ODG do PEAQ no teste de avaliação de qualidade de sinais de áudio degradados pelo codificador MP3. A descrição das diferentes taxas de codificação utilizadas foi apresentada na Seção 2.4.2.

Assim como o WAQA, o *DWT-watermarking* apresentou um desempenho que pode ser considerado satisfatório na avaliação de qualidade de sinais de áudio conta-

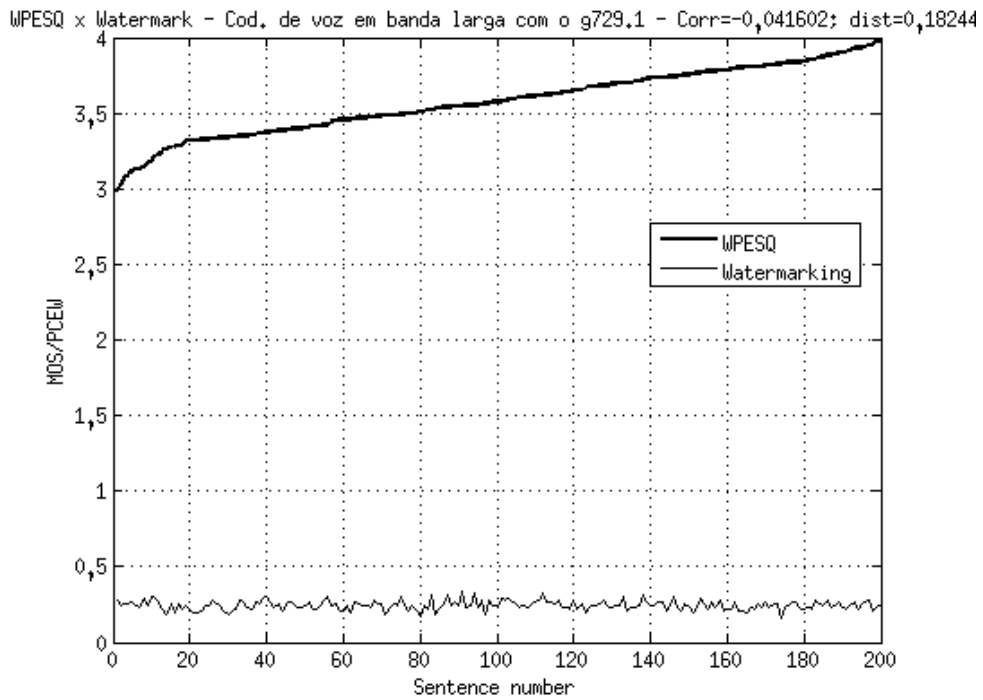


Figura 3.8: Comparação entre o WPESQ e o *DWT-watermarking* na avaliação de sinais de voz degradados pelo codificador G.729.1.

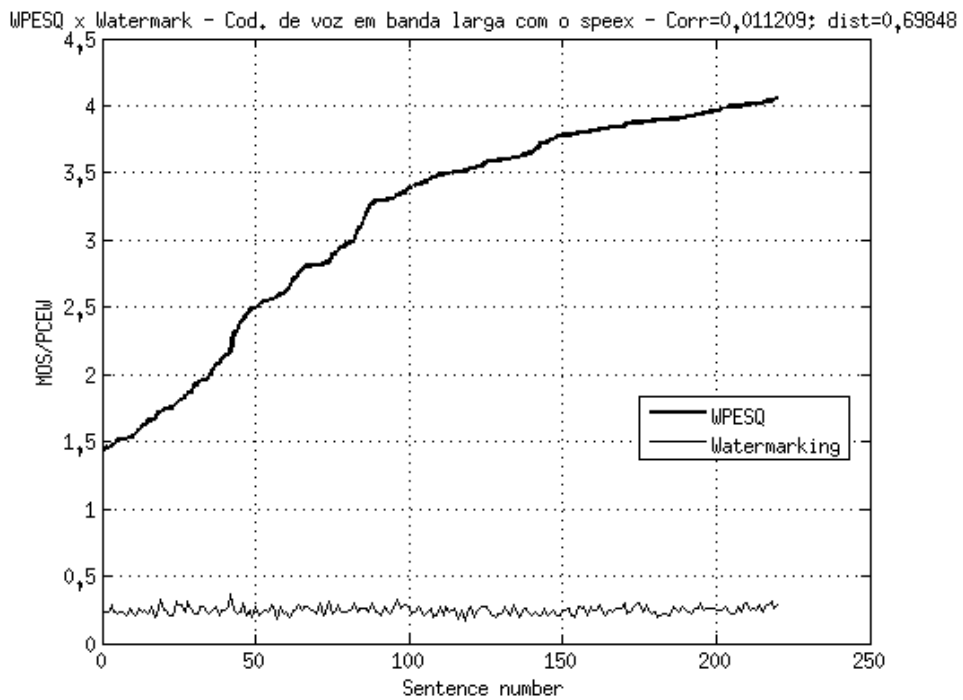


Figura 3.9: Comparação entre o WPESQ e o *DWT-watermarking* na avaliação de sinais de voz degradados pelo codificador Speex.

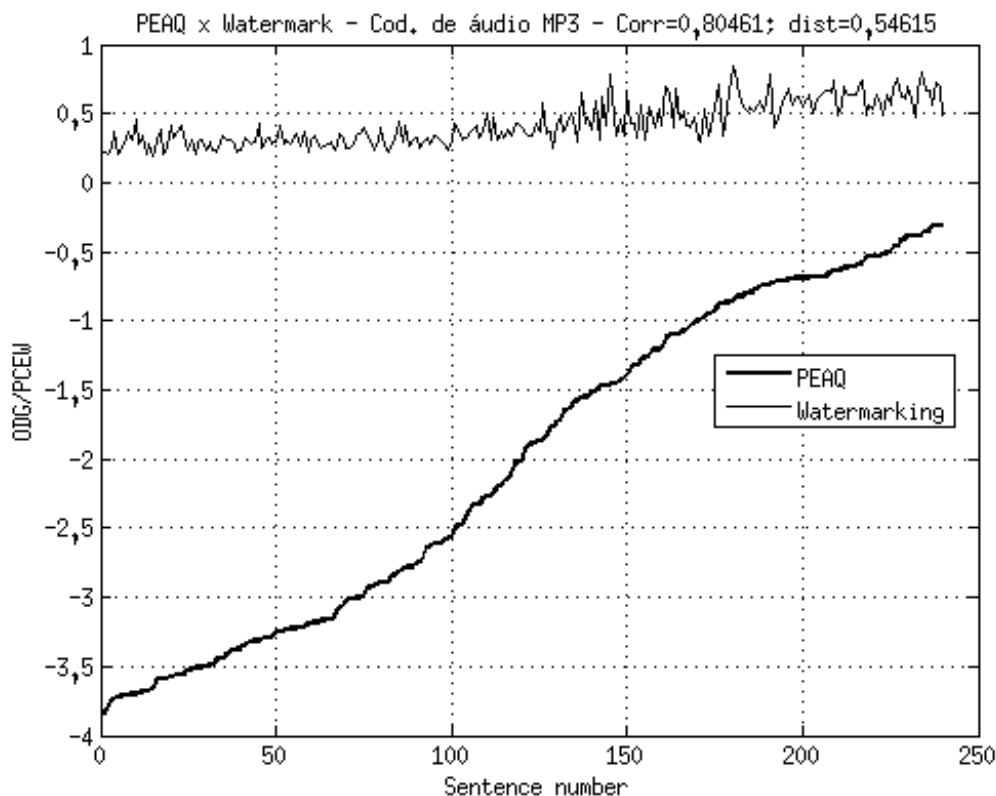


Figura 3.10: Comparação entre o PEAQ e o *DWT-watermarking* na avaliação de sinais de áudio degradados pelo codificador MP3.

minados com ruído oriundo de processos de codificação perceptiva. Obteve-se uma correlação acima de 0,8, porém a medida *dist* foi bastante ruim (muito elevada, se comparada aos outros resultados). Por inspeção visual da Fig. 3.10, é possível notar que o *DWT-watermarking* acompanhou, na média, o comportamento do PEAQ; entretanto, assim como ocorrido no WAQA, houve uma grande variação entre notas que deveriam ser próximas, i.e., o resultado aparentemente foi acurado, porém pouco preciso.

3.4 Discussão dos resultados da literatura

A definição dos testes de alguns trabalhos da literatura apresenta uma série de problemas que foram citados anteriormente e serão listados a seguir.

Mapeamento para o MOS Em [26] e [27] utiliza-se um mapeamento do PCEW para a nota MOS e todos os resultados divulgados foram obtidos através desse mapeamento. Além de essa estratégia não dar a medida clara da adequação do PCEW em si para a avaliação de qualidade, o mapeamento proposto também é incomum: realizam-se 10 mapeamentos distintos, cada um deles para um

intervalo de 0,1 nas notas do PCEW. Ademais, cada um dos testes realizados recebeu um conjunto distinto de mapeamentos. Isso pode ter mascarado resultados originalmente não muito bons.

Testes com MP3 Conforme explicado na Seção 3.2.2, o PESQ não é adequado para a avaliação de qualidade de sinais contaminados com ruído oriundo de codificação perceptiva MP3, tendo sido utilizado dessa forma em [27].

Testes com filtragem passa-baixas Conforme também citado na Seção 3.2.2, em [26] e [27] foram utilizadas filtragens passa-baixas incompatíveis com a taxa de amostragem dos sinais utilizados e com o método de avaliação de qualidade de referência que foi utilizado (PESQ).

Levando todas essas características em conta, ainda se podem destacar os seguintes pontos:

- O resultado obtido com ruído branco no nosso teste foi comparável ao resultado obtido na literatura, mesmo não tendo sido realizado nenhum mapeamento para o PCEW;
- Além dos resultados do teste com ruído, ainda que menos expressivos, houve resultados bastante positivos para a avaliação de sinais contaminados com filtragem passa-baixas e com codificação perceptiva MP3, mesmo sem nenhum tipo de mapeamento;
- Não foi encontrado teste com codificadores de voz na literatura para que se pudesse fazer uma comparação de desempenho. Entretanto, pelo mau resultado que foi obtido neste trabalho, é possível imaginar que não tenha sido publicado devido ao seu fraco desempenho;
- O teste com perda de pacotes na literatura teve um bom desempenho, resultado diferente do obtido neste trabalho. Entretanto, destaca-se que dos testes publicados foi o que teve desempenho mais fraco, além de ser o que requer a curva de mapeamento mais não linear (tabela listada em [27]), que afinal pode ter sido responsável pelo bom resultado.

3.5 Conclusões

Apesar de o *DWT-watermarking* ser um método que não necessita do sinal original para comparação, o seu desempenho nos testes foi comparável ao desempenho do WAQA. Com desempenhos equivalentes, as mesmas conclusões e considerações feitas no Cap. 2 são igualmente aplicáveis ao *DWT-watermarking*.

Apesar de não ser possível realizar uma comparação direta com os resultados apresentados na literatura, é possível identificar a correspondência com estes em alguns casos.

Analisando de uma forma ampla os resultados obtidos neste capítulo, o *DWT-watermarking* pode ser considerado um métrica de avaliação de qualidade eficaz em diversas das contaminações de sinais de áudio/voz que foram analisadas.

Capítulo 4

Aperfeiçoamento do método de avaliação de qualidade de áudio/voz que utiliza marcas d'água digitais

4.1 Introdução

O objetivo deste capítulo é apresentar e avaliar algumas modificações propostas aos métodos da literatura que utilizam marcas d'água digitais com base na DWT para fazer a avaliação de qualidade de sinais de áudio e voz [24] [25] [26] [27].

Conforme demonstram os resultados obtidos nos Caps. 2 e 3, é possível determinar, para determinados tipos de degradação, a qualidade dos sinais de áudio/voz através de métodos que se utilizam da DWT e de marcas d'água digitais.

A busca da melhoria de desempenho do método *DWT-watermarking* (apresentado no Cap. 3) será feita através de duas estratégias principais:

- Modificações no método original a fim de melhorar o desempenho da medida PCEW¹.
- Melhoria do desempenho geral através de um novo mapeamento do PCEW no MOS/ODG correspondente.

Neste capítulo, primeiramente apresentar-se-ão duas modificações propostas para o método original: avalia-se primeiramente o impacto de uma delas e em seguida o impacto de ambas no desempenho do *DWT-watermarking*. Na segunda metade do capítulo, apresenta-se o mapeamento do PCEW e o seu desempenho conforme a estratégia adotada.

¹Saída do avaliador do Cap. 3.

4.2 *DWT-watermarking* modificado

Conforme discutido e analisado no Cap. 3, o *DWT-watermarking* possui alguns pontos fracos. Dentre os pontos fracos que podem ser identificados, destacam-se:

- Não há garantia de recuperação exata da marca d'água, mesmo sem a degradação do sinal. Como o passo de quantização (Δ) é definido através do fator de quantização e do maior e menor coeficiente de cada nível (ver Cap. 2, Seção 2.3) ele fica sujeito às modificações desses coeficientes que podem ocorrer durante a própria inserção da marca.
- Não há robustez nos rótulos dos coeficientes (ver Cap. 2, Seção 2.3). Quando redefinidos, eles são modificados através da soma simples de um Δ , sem que exista a preocupação de que o coeficiente após a modificação se encontre distante da fronteira entre os rótulos. Com isso, alguns rótulos podem ficar posicionados perto da fronteira, o que poderá acarretar uma mudança de rótulo para o coeficiente, mesmo quando submetido a uma modificação muito pequena.
- Existe a necessidade do conhecimento do fator de quantização (Q_s) utilizado na inserção da marca para que a sua extração possa ser realizada (ver Cap. 3, Seção 3.2.1). Essa necessidade não permite que o *DWT-watermarking* seja um método sem referência, já que além do próprio sinal sob avaliação é necessário enviar o fator de quantização.

Buscaram-se formas de reduzir ou remover esses pontos fracos através de modificações no *DWT-watermarking*. Obteve-se sucesso e serão apresentadas a seguir as modificações que corrigem os dois primeiros pontos citados, relativos à garantia de recuperação e à robustez da marca. Em contrapartida, não foi possível obter algum esquema que conseguisse remover a necessidade do conhecimento do fator de quantização; as tentativas de estimação cega do fator de quantização foram mal-sucedidas.

4.2.1 Modificação A: Recuperação exata da marca

Conforme apresentado no Cap. 2, no *DWT-watermarking* os rótulos dos coeficientes são definidos através da Eq. (4.1), na qual Δ é definido através da Eq. (4.2).

$$r \equiv \left\lfloor \frac{\sqrt{(\text{coef})^2}}{\Delta_N} + 0,5 \right\rfloor \pmod{2}, \quad (4.1)$$

sendo

$$\Delta_N = \frac{C_{N\text{máx}} - C_{N\text{mín}}}{Q_s}. \quad (4.2)$$

Analisando as equações apresentadas, é possível perceber que qualquer mudança nos maiores ou menores coeficientes de cada nível da DWT ($C_{N\text{máx}}$ e $C_{N\text{mín}}$, respectivamente) gera uma mudança no Δ que pode gerar, por sua vez, uma mudança no rótulo (r) do coeficiente.

Conforme apresentado na Seção 3.2.1, escolhem-se alguns dos maiores coeficientes em módulo em cada nível para que possam ter seus rótulos avaliados e modificados conforme a necessidade. Nesse processo, é possível que os coeficientes utilizados no cálculo de Δ sejam modificados, e com isso o próprio Δ seja alterado.

Para minimizar esse problema, utiliza-se a rotina de otimização do passo de quantização (Q_s), apresentada no Apêndice A, que avalia o grau de recuperabilidade da marca d'água, i.e., mede qual o percentual de bits de marca inseridos que deixam de ser recuperados corretamente por conta da inexatidão no cálculo do Δ e modifica Q_s a fim de minimizar esse percentual.

Esse processo de minimização não é muito eficaz, pois, conforme se avaliou em testes preliminares, em muitos casos não é possível descobrir um passo de quantização que tanto mantenha a degradação imposta pela marca fixada no nível desejado, quanto minimize o problema da não-recuperabilidade da marca.

A solução encontrada foi a remoção dessa não-recuperabilidade através da modificação do método. A modificação implementada consiste em:

- excluir os coeficientes utilizados para o cálculo de Δ do conjunto usado para a inserção da marca d'água;
- após o cálculo de Δ , aumentar em Δ o módulo dos coeficientes utilizados para calculá-lo. Isso garante que a inserção da marca não tornará nenhum outro coeficiente maior (em módulo) no nível. Se isso ocorresse, outro coeficiente passaria a ser utilizado no cálculo do Δ durante a extração da marca.

Essa modificação impõe que na recuperação do Δ para extração da marca leve-se em conta a modificação dos coeficientes $C_{N\text{máx}}$ e $C_{N\text{mín}}$, caso eles tenham sinais diferentes (caso mais comum) e, portanto, sua diferença tenha sido aumentada em 2Δ . Nesse caso, um ajuste de Δ passa a ser necessário:

$$\Delta_{\text{real}} = \frac{\Delta_{\text{obtido}}}{1 + \frac{2}{Q_s}}. \quad (4.3)$$

O custo dessa modificação é impor ao sinal certa degradação adicional decorrelacionada da marca. O resultado dessa degradação pode ser a utilização de um passo de quantização maior—i.e., uma marca mais fraca—para que o sinal marcado fique dentro do nível de degradação pré-estabelecido.

Assim como na Seção 2.3, vale destacar que os casos especiais em que o maior e o menor coeficiente possuem o mesmo módulo foram adequadamente tratados na

implementação.

A seguir serão apresentados os resultados obtidos com essa estratégia.

Resultados da modificação A

Na Tab. 4.1 se encontram listadas as medidas de eficiência obtidas com a modificação proposta e as medidas obtidas no Cap. 3. Não serão apresentados os gráficos com os resultados ilustrados (conforme feito no Cap. 3), pois os gráficos obtidos com a modificação proposta são muito semelhantes aos anteriormente apresentados, não acrescentando nenhuma informação adicional.

Tabela 4.1: Resultados da modificação A.

Teste Realizado	Original		Modificação A	
	corr	dist	corr	dist
Ruído branco gaussiano	0,974	0,200	0,973	0,201
Perda de pacotes	0,669	0,182	0,766	0,163
Filtragem passa-baixas	0,857	0,220	0,844	0,222
Cod. de voz com o G.729.1	0,041	0,182	0,036	0,200
Cod. de voz com o speex	0,011	0,698	0,086	0,697
Cod. de áudio MP3	0,805	0,546	0,796	0,582

Conforme se pode observar na Tab. 4.1, à exceção do teste para sinais degradados através de perda de pacotes, todos os resultados obtidos foram bastante similares aos do Cap. 3. O teste com a degradação por perda de pacotes, diferentemente dos demais, apresentou um resultado bastante melhor que o obtido com o método original, tanto do ponto de vista da correlação quanto do ponto de vista da distância para o ajuste linear.

Com esses resultados é possível afirmar que a modificação proposta foi bem sucedida e traz melhoria ao *DWT-watermarking*.

4.2.2 Modificação B: Aumento na robustez da marca

Conforme apresentado na Seção 3.2.1, modificam-se os rótulos de alguns coeficientes da DWT para se inserir a marca d'água.

Sendo o rótulo de cada coeficiente dado pela Eq. (4.1), propõe-se a seguinte modificação: em vez de simplesmente adicionar Δ ao módulo dos coeficientes com rótulo diferente do bit de marca que se deseja inserir, adiciona-se ou subtrai-se o menor valor possível para que o coeficiente fique centralizado no rótulo. É possível ver uma ilustração da diferença proposta na Fig. 4.1.

Entretanto, essa modificação não pode ser feita de forma direta pelas seguintes razões:

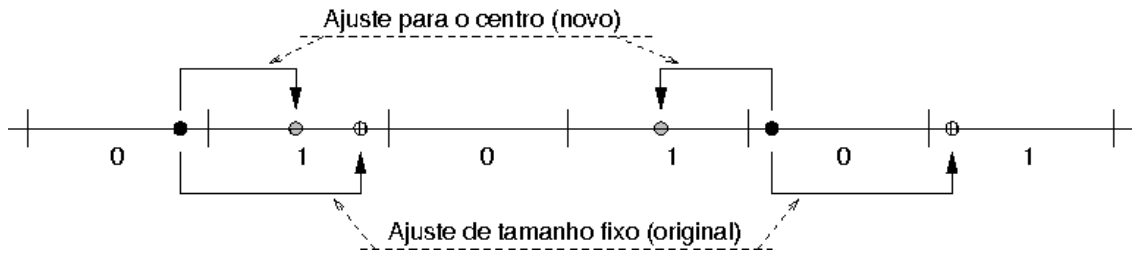


Figura 4.1: Ajuste original e novo ajuste dos coeficientes da DWT.

- Ao reduzir o módulo de um coeficiente, é necessário garantir que o mesmo siga no conjunto dos maiores coeficientes. Para tal, realizou-se um ajuste condicionado: caso a melhor solução seja reduzir o coeficiente, ela só é adotada se o coeficiente reduzido ainda for maior que o menor coeficiente escolhido para ajuste naquele nível; caso contrário, o ajuste é feito através do aumento do coeficiente.
- Como a modificação A previamente descrita está sendo mantida juntamente com a modificação B, é necessário garantir que uma mudança nos coeficientes não os torne os maiores daquele nível (já que os maiores de cada nível são utilizados para o cálculo de Δ). Para tal, alterou-se um pouco a modificação A aumentando o passo utilizado nos maiores coeficientes do nível de Δ para $1,5\Delta$.

A fim de testar a eficiência da modificação, idealizaram-se dois testes distintos:

Teste B-1: Ajuste restrito Neste teste alteraram-se apenas os coeficientes cujos rótulos precisavam ser mudados para que a marca fosse inserida. Em outras palavras, neste teste realizou-se a mesma operação anteriormente descrita, modificando-se apenas o método de ajuste dos coeficientes.

Teste B-2: Ajuste total Neste teste todos os coeficientes que armazenam informações da marca d'água são reajustados—não importando seu rótulo original—de forma a ter seu rótulo centralizado de acordo com a marca d'água que se pretende inserir.

O impacto esperado com a primeira alternativa (B-1) é uma modificação menor dos coeficientes com rótulo inadequado (com exceção de poucos casos anteriormente expostos). Assim, espera-se utilizar um passo de quantização menor (uma marca mais “fraca”), que potencialmente apresentaria um desempenho superior.

Com a segunda alternativa (B-2), coeficientes que antes não eram modificados passam a sê-lo. Dessa maneira, espera-se utilizar um passo de quantização maior (uma marca mais “forte”) que a obtida apenas com a modificação A, pois o número

de coeficientes alterados será muito maior. O que se deseja investigar é se a desvantagem dessa marca mais “forte” não seria compensada pela vantagem de se minimizarem os erros de coeficientes que originalmente se encontravam próximos das fronteiras dos rótulos.

A seguir serão apresentados os resultados obtidos com cada uma das estratégias.

Resultados das modificações B-1 e B-2

Na Tab. 4.2 se encontram listadas as medidas de eficiência obtidas com a segunda modificação proposta e as medidas obtidas com a modificação A. Assim como na seção anterior, não houve preocupação em apresentar os gráficos com os resultados ilustrados (conforme feito no Cap. 3), pois novamente os gráficos obtidos com a modificação são muito semelhantes aos anteriormente apresentados.

Tabela 4.2: Resultados das modificações B-1 e B-2, comparados à modificação A.

Teste Realizado	Modif. A		Modif. B-1		Modif. B-2	
	corr	dist	corr	dist	corr	dist
Ruído branco gaussiano	0,973	0,201	0,973	0,205	0,969	0,221
Perda de pacotes	0,766	0,163	0,727	0,161	0,682	0,178
Filtragem passa-baixas	0,844	0,222	0,848	0,221	0,848	0,220
Cod. de voz com o G.729.1	0,036	0,200	0,029	0,202	0,087	0,189
Cod. de voz com o speex	0,086	0,697	0,040	0,697	0,015	0,692
Cod. de áudio MP3	0,796	0,582	0,845	0,509	0,849	0,494

Conforme é possível observar na Tab. 4.2, as modificações B-1 e B-2 trouxeram ganho de desempenho para a avaliação de qualidade de sinais de áudio contaminados com ruído de codificação MP3. Foi obtido desempenho equivalente entre a modificação A e as modificações B-1 e B-2 para os demais testes (com sinais de voz), com exceção da perda de pacotes, em que se obteve pequena redução.

Além disso, observa-se que a modificação B-2 não apresentou desempenho superior à modificação B-1, resultando num esforço computacional adicional inútil. Descarta-se, portanto, a modificação B-2. O método “*DWT-watermarking*+modificação A+modificação B-1”, será doravante chamado simplesmente de *DWT-watermarking* Modificado.

A Tab. 4.3 apresenta uma comparação entre os resultados obtidos com o *DWT-watermarking* Original e os obtidos com o *DWT-watermarking* Modificado.

Diante dos resultados expostos, é possível afirmar que houve sucesso na tentativa de melhorar o método originalmente proposto na literatura.

Tabela 4.3: Comparação entre *DWT-watermarking* Original e Modificado

Teste Realizado	Original		Modificado		Desempenho Alcançado
	corr	dist	corr	dist	
Ruído branco gaussiano	0,974	0,200	0,973	0,205	Equivalente
Perda de pacotes	0,669	0,182	0,727	0,161	Melhor
Filtragem passa-baixas	0,857	0,220	0,848	0,221	Equivalente
Cod. de voz G.729.1	0,041	0,182	0,029	0,202	Equivalente
Cod. de voz speex	0,011	0,698	0,040	0,697	Equivalente
Cod. de áudio MP3	0,805	0,546	0,845	0,509	Melhor

4.3 Mapeamento das notas

É muito comum que métodos objetivos de avaliação de qualidade utilizem algum tipo de mapeamento das notas de saída dos seus modelos matemáticos/psicoacústicos para uma escala correspondente ao julgamento humano; como exemplo, os métodos padronizados PEAQ e PESQ utilizam um mapeamento desse tipo como último estágio na geração de suas notas.

Nesta parte do trabalho pretende-se testar o desempenho do *DWT-watermarking* Modificado após o mapeamento de sua nota de saída (**PCEWM**—PCEW Modificada) para uma escala correspondente ao julgamento subjetivo (MOS/ODG). Busca-se com isso conhecer melhor o desempenho potencial do método contra testes subjetivos que viessem a ser realizados para sua aferição em condições reais. Espera-se com o mapeamento obter medidas de correlação (*corr*) e diferença absoluta (*dist*) melhores.

4.3.1 Metodologia de mapeamento

A fim de realizar o mapeamento de notas, é necessário definir dois conjuntos distintos de pares de notas. Cada par é formado por uma nota de referência (ODG/PEAQ para áudio e MOS/WPESQ para voz) e uma nota a ser mapeada (PCEWM).

O primeiro conjunto, chamado conjunto de treinamento, tem como função realizar o ajuste dos parâmetros de mapeamento. Esse ajuste é feito através do posicionamento dos pontos formados pelos pares de notas num plano X–Y, para em seguida buscar uma curva que esteja o mais próxima possível dos pontos traçados segundo algum critério; utilizou-se aqui a minimização do erro quadrático médio.

O segundo conjunto de notas, chamado conjunto de teste, avalia a eficácia do mapeamento. Essa avaliação é feita através dos mesmos critérios utilizados para medir o desempenho do PCEWM, i.e., as métricas *corr* e *dist*.

Para o mapeamento, decidiu-se utilizar curvas que fossem definidas com um

número pequeno, porém suficiente, de parâmetros. Assim, depois de terem sido realizados testes informais com diversas curvas, optou-se por duas curvas distintas: uma polinomial do quarto grau, que pode ser definida com 5 parâmetros; e uma exponencial do tipo $y = a + be^{cx}$, que pode ser definida com 3 parâmetros (a, b e c).

Adotaram-se duas estratégias distintas para a separação dos pares de notas entre os conjuntos. Na primeira estratégia, chamada estratégia de divisão dos sinais, dividiram-se os dois conjuntos de maneira que cada um deles ficasse com a metade dos sinais em todas as degradações. Na segunda estratégia, chamada estratégia de divisão das degradações, dividiram-se os dois conjuntos de maneira que cada um deles ficasse com a metade das degradações em todos os sinais. Se, por exemplo, foram utilizados 20 sinais de voz degradados com 7 frequências distintas de filtragem passa-baixas:

- na primeira estratégia teremos 70 sinais para cada conjunto (10 sinais de voz distintos, e 7 níveis de degradação para cada sinal);
- na segunda estratégia teremos o primeiro conjunto—de treinamento—com 60 sinais (20 sinais de voz distintos, e níveis 2, 4 e 6 de degradação para cada sinal), e o segundo conjunto—de teste—com os 80 sinais restantes (20 sinais de voz distintos, e níveis 1, 3, 5 e 7 de degradação para cada sinal).

Temos assim um total de 4 resultados para cada um dos testes, a saber:

1. Mapeamento polinomial e divisão dos sinais
2. Mapeamento exponencial e divisão dos sinais
3. Mapeamento polinomial e divisão das degradações
4. Mapeamento exponencial e divisão das degradações

Vale ressaltar que existe uma “competição” entre o teste com mapeamento polinomial e com mapeamento exponencial, na qual se busca descobrir qual dos dois produz melhores resultados. Por outro lado, a estratégia de divisão dos sinais busca investigar a eficácia do mapeamento na avaliação de sinais não utilizados no treinamento, enquanto na divisão das degradações busca-se investigar a eficácia do mapeamento na avaliação de degradações não utilizadas no treinamento; não há competição entre as estratégias de divisão.

A fim de melhorar o resultado do mapeamento, em boa parte dos conjuntos de pares de notas utilizados para a otimização foram acrescentados pares de notas que correspondiam aos limites do mapeamento, ou seja, notas máximas e mínimas PCEWM e suas notas correspondentes MOS ou ODG. A maior parte das métricas de avaliação melhorou ou obteve o mesmo desempenho com esse acréscimo, à exceção

de um número bem reduzido de casos em que se observaram resultados levemente inferiores.

Nas tabelas a seguir, serão apresentados na linha 1 os resultados obtidos anteriormente (com a base de sinais completa e sem nenhum mapeamento), nas linhas 2 e 5 os resultados sem mapeamento dos conjuntos de teste das divisões; e nas demais linhas os resultados obtidos com os mapeamentos polinomial e exponencial para cada uma das divisões.

Foram apresentados os resultados sem mapeamento dos conjuntos de teste de cada uma divisões para garantir que uma melhora ou piora observada no resultado do mapeamento não fosse oriunda do próprio sub-conjunto de teste da base de notas. Assim, a melhor comparação a se fazer é entre a linha 2 com as linhas 3 e 4, e a linha 5 com as linhas 6 e 7. Vale também comparar o desempenho dos mapeamentos polinomial e exponencial.

A fim de permitir que os resultados deste trabalho possam ser reproduzidos, os parâmetros do mapeamento encontrados foram apresentados no Apêndice B, nas Tabelas B.1 e B.2.

Todas as curvas resultantes dos mapeamentos realizados se encontram apresentadas no Apêndice C. Analisando-se cada uma das curvas obtidas e comparando-as com os resultados apresentados nas tabelas a seguir, será constatado que quanto mais não-linear for a distribuição dos pontos, maior será a melhora nos resultados com o mapeamento. Naturalmente, casos descorrelacionados não devem ser significativamente melhorados.

4.3.2 Resultados

Voz – Ruído

A Tab. 4.4 apresenta os resultados obtidos com o mapeamento na avaliação de sinais de voz contaminados com ruído.

Tabela 4.4: Resultados do mapeamento para sinais de voz contaminados com ruído.

Linha	Conjunto Avaliado	corr	dist
1	Base completa sem mapeamento	0,973	0,205
2	Conj. de teste da div. dos sinais sem map.	0,972	0,204
3	Map. polinomial – div. dos sinais	0,977	0,185
4	Map. exponencial – div. dos sinais	0,971	0,208
5	Conj. de teste da div. das degradações sem map.	0,969	0,217
6	Map. polinomial – div. das degradações	0,972	0,188
7	Map. exponencial – div. das degradações	0,965	0,228

Nas Figs. C.1, C.2, C.3 e C.4 encontram-se apresentados os gráficos relativos às linhas 3, 4, 6 e 7 da Tab. 4.4, respectivamente.

O *DWT-watermarking* Modificado já apresentava bom resultado na avaliação de sinais de voz contaminado com ruído. Esperava-se, dessa forma, uma melhora muito pequena com o mapeamento. Foi o que ocorreu, com uma redução entre 5 e 10% na medida *dist* para o mapeamento polinomial. Com o mapeamento exponencial, não houve mudança.

Foram obtidos resultados similares nas estratégias de divisão dos sinais e de divisão das degradações, demonstrando que os mapeamentos realizados são robustos o suficiente para obter bom desempenho tanto na avaliação de sinais desconhecidos quanto na avaliação de degradações desconhecidas.

Os resultados finais obtidos indicam boa adequação do *DWT-watermarking* Modificado para a avaliação de qualidade de sinais de voz contaminados com ruído branco gaussiano (modelando ruído de fundo de banda larga).

Voz – Perda de Pacotes

A Tab. 4.5 apresenta os resultados obtidos com o mapeamento na avaliação de sinais de voz degradados através de perda de pacotes.

Tabela 4.5: Resultados do mapeamento para sinais de voz degradados por perda de pacotes.

Linha	Conjunto Avaliado	corr	dist
1	Base completa sem mapeamento	0,727	0,161
2	Conj. de teste da div. dos sinais sem map.	0,711	0,167
3	Map. polinomial – div. dos sinais	0,751	0,129
4	Map. exponencial – div. dos sinais	0,752	0,139
5	Conj. de teste da div. das degradações sem map.	0,775	0,205
6	Map. polinomial – div. das degradações	0,807	0,172
7	Map. exponencial – div. das degradações	0,819	0,155

Nas Figs. C.5, C.6, C.7 e C.8 encontram-se apresentados os gráficos relativos às linhas 3, 4, 6 e 7 da Tab. 4.5, respectivamente.

Conforme pode ser observado na linha 1 da Tab. 4.5, apesar de o *DWT-watermarking* Modificado não ter apresentado uma correlação muito elevada na avaliação de qualidade de sinais degradados por perda de pacotes, ele apresentava uma medida *dist* bastante reduzida, indicando precisão nas notas, i.e., que a diferença média entre a nota obtida e a nota real era pequena.

Além disso, como o MOS do WPESQ apresentava comportamento aproximadamente exponencial, conforme pode ser observado nas Figs. 2.7 e 3.6, esperava-se

que com o mapeamento fosse possível melhorar significativamente o desempenho do *DWT-watermarking* Modificado.

De fato, a melhora verificada com o mapeamento foi muito significativa. Foi possível obter uma redução de 15 a 25% no desvio médio das notas (medida *dist*), e um aumento de aproximadamente 5,5% na correlação entre as notas.

Apesar da elevação nas medidas de correlação, elas ainda não podem ser consideradas altas. Por outro lado, as excelentes medidas de desvio médio indicam que as avaliações realizadas com o mapeamento são acuradas. Com isso, fica difícil assegurar a adequação ou não-adequação do método para a avaliação de qualidade de sinais de voz degradados por perda de pacotes. Entretanto, as excelentes medidas *dist* parecem habilitar o método para a avaliação desse tipo de degradação, esperando-se que as medidas de correlação possam ser elevadas com algum dos estudos que serão propostos no Cap. 5.

Voz – Filtragem Passa-Baixas

A Tab. 4.6 apresenta os resultados obtidos com o mapeamento na avaliação de sinais de voz degradados por filtragem passa-baixas.

Tabela 4.6: Resultados do mapeamento para sinais de voz degradados por filtragem passa-baixas.

Linha	Conjunto Avaliado	corr	dist
1	Base completa sem mapeamento	0,848	0,221
2	Conj. de teste da div. dos sinais sem map.	0,838	0,212
3	Map. polinomial – div. dos sinais	0,842	0,207
4	Map. exponencial – div. dos sinais	0,844	0,207
5	Conj. de teste da div. das degradações sem map.	0,849	0,219
6	Map. polinomial – div. das degradações	0,850	0,220
7	Map. exponencial – div. das degradações	0,850	0,217

Nas Figs. C.9, C.10, C.11 e C.12 encontram-se apresentados os gráficos relativos às linhas 3, 4, 6 e 7 da Tab. 4.6, respectivamente.

Não foi possível observar nenhuma melhoria significativa com o mapeamento; já era esperado, já que a curva obtida para o MOS do WPESQ, que pode ser observada na Fig. 3.7, já apresentava comportamento bastante linear.

Como também não se obteve melhoria com o *DWT-watermarking* Modificado em relação ao *DWT-watermarking* Original (ver Tab. 4.3), o resultado final foi equivalente ao obtido no Cap. 3. Assim, não há nenhuma análise dos resultados ou discussão adicional à que foi realizada na Subseção 3.3.4.

Com esses resultados, apesar de não ter sido possível obter uma correlação acima de 0,9 (alta, em termos práticos), o desempenho do *DWT-watermarking* Modificado

na avaliação de sinais degradados por filtragem passa-baixas indica que o método é indicado para a avaliação desse tipo de degradação.

Testes informais indicaram que utilizando outros parâmetros (e.g. diferentes wavelets-mãe) era possível obter resultados melhores, o que corrobora a afirmação de que o método é adequado para a avaliação de qualidade de sinais submetidos a este tipo de degradação. Não se formalizaram esses testes porque uma escolha de parâmetros otimizada para cada uma das degradações foge à proposta desse trabalho, que foi obter um único conjunto de parâmetros para todas as degradações², numa tentativa de obter um método unificado.

Voz – Codificação G.729.1 e Speex

As Tabelas 4.7 e 4.8 apresentam os resultados obtidos com o mapeamento na avaliação de sinais de voz degradados por codificação.

Tabela 4.7: Resultados do mapeamento para sinais de voz degradados pelo codificador G.729.1.

Linha	Conjunto Avaliado	corr	dist
1	Base completa sem mapeamento	0,029	0,202
2	Conj. de teste da div. dos sinais sem map.	0,087	0,181
3	Map. polinomial – div. dos sinais	0,128	0,179
4	Map. exponencial – div. dos sinais	0,087	0,181
5	Conj. de teste da div. das degradações sem map.	0,065	0,201
6	Map. polinomial – div. das degradações	-0,053	0,201
7	Map. exponencial – div. das degradações	-0,065	0,201

Tabela 4.8: Resultados do mapeamento para sinais de voz degradados pelo codificador Speex.

Linha	Conjunto avaliado	corr	dist
1	Base completa sem mapeamento	0,040	0,697
2	Conj. de teste da div. dos sinais sem map.	0,106	0,684
3	Map. polinomial – div. dos sinais	0,036	0,691
4	Map. exponencial – div. dos sinais	0,107	0,684
5	Conj. de teste da div. das degradações sem map.	0,009	0,700
6	Map. polinomial – div. das degradações	0,164	0,680
7	Map. exponencial – div. das degradações	0,008	0,700

Nas Figs. C.13, C.14, C.15 e C.16 encontram-se apresentados os gráficos relativos às linhas 3, 4, 6 e 7 da Tab. 4.7, respectivamente. Nas Figs. C.17, C.18, C.19 e C.20

²Os únicos parâmetros que se permitiu serem diferentes para as diferentes degradações foram os parâmetros de mapeamento.

encontram-se apresentados os gráficos relativos às linhas 3, 4, 6 e 7 da Tab. 4.8, respectivamente.

Estes resultados são apresentados apenas por completude, pois já se havia constatado que o *DWT-watermarking* Modificado não é adequado para a avaliação de qualidade de sinais degradados por codificadores de voz.

Áudio – Codificação MP3

A Tab. 4.9 apresenta os resultados obtidos com o mapeamento na avaliação de sinais de áudio degradados por codificação MP3.

Tabela 4.9: Resultados do mapeamento para sinais de áudio degradados por codificação MP3.

Linha	Conjunto avaliado	corr	dist
1	Base completa sem mapeamento	0,845	0,509
2	Conj. de teste da div. dos sinais sem map.	0,847	0,516
3	Map. polinomial – div. dos sinais	0,866	0,455
4	Map. exponencial – div. dos sinais	0,856	0,497
5	Conj. de teste da div. das degradações sem map.	0,858	0,485
6	Map. polinomial – div. das degradações	0,870	0,448
7	Map. exponencial – div. das degradações	0,859	0,484

Nas Figs. C.21, C.22, C.23 e C.24 encontram-se apresentados os gráficos relativos às linhas 3, 4, 6 e 7 da Tab. 4.9, respectivamente.

Com o mapeamento obteve-se um aumento de cerca de 2,5% na correlação, além de uma redução de cerca de 13% no desvio médio das notas (medida dist).

Entretanto, vale ressaltar que o desvio médio das notas obtido (em torno de 0,45) é bastante superior aos valores encontrados nas avaliações dos sinais de voz (em torno de ou inferiores a 0,2). Essa diferença pode ser explicada pela maior variabilidade da base de sinais de áudio.

Os resultados obtidos com a divisão dos sinais e com a divisão das degradações foram muito próximos, indicando robustez do método. Esse resultado é bastante importante, pois, ao contrário dos sinais de voz, os sinais de áudio utilizados são muito diferentes entre si.

Mesmo com esse desvio médio (ou distância média—dist) mais elevado, a correlação em torno de 0,87 aponta que o *DWT-watermarking* Modificado com mapeamento é indicado para a avaliação de qualidade de sinais de áudio que sofreram codificação perceptiva.

4.4 Conclusão

Os objetivos deste capítulo foram alcançados, já que tanto as modificações propostas para o *DWT-watermarking* como o mapeamento das notas foram bem sucedidos.

A melhora obtida com as modificações propostas indica que algumas fraquezas do método original foram corretamente identificadas e reduzidas.

As duas divisões distintas dos sinais de teste e treinamento utilizadas no mapeamento permitiram identificar a robustez dos mapeamentos utilizados.

Foi constatado de forma definitiva a inadequação do método para a avaliação de sinais submetidos à codificação de voz, assim como obtiveram-se indícios fortes da adequação do método para as demais degradações estudadas.

Capítulo 5

Conclusões

Quando este trabalho foi iniciado, o único método de avaliação de qualidade através de marcas d'água na literatura utilizava a DWT. Ele parecia promissor, porém era possível identificar: problemas na metodologia de testes empregada; fragilidades na escolha de parâmetros; utilização de mapeamentos setorizados *ad hoc* das métricas propostas para a escala MOS; e metodologia de avaliação de resultados pouco usual.

Este trabalho buscou analisar e desenvolver a idéia original através de um melhor embasamento do método, com: o estudo da DWT por si só para avaliação de qualidade de sinais de áudio e voz; a reformulação das metodologias de testes e de avaliação de resultados; a sistematização da escolha de parâmetros; a proposta de um novo mapeamento para as notas de saída; e o aperfeiçoamento do método proposto em si.

5.1 Resultados

No Cap. 2 foi analisada a possibilidade de realizar a avaliação de qualidade de sinais de áudio e voz através da apreciação dos maiores coeficientes da DWT. Foi utilizada uma metodologia de avaliação dos coeficientes inspirada nos métodos de avaliação de qualidade utilizando marcas d'água que seriam estudados nos capítulos seguintes. Além disso, nesse capítulo foram definidas as métricas de avaliação de resultados utilizadas durante o restante do trabalho. Os testes realizados indicaram os tipos de degradações para os quais seria possível encontrar melhor desempenho dos métodos de avaliação com base na DWT. Os resultados obtidos indicaram a impossibilidade de avaliar sinais submetidos à codificação de voz (confirmada nos capítulos seguintes). Para os demais testes, os resultados variaram de medianos a muito bons.

No Cap. 3 foi testado o método *DWT-watermarking* em sua forma original, obtendo-se resultados próximos dos do Cap. 2. Foram definidos os parâmetros necessários restantes (boa parte já tinha sido definida no Cap. 2), e foram analisadas

as fraquezas do *DWT-watermarking* da literatura. Criticou-se a metodologia de testes que foi originalmente apresentada com ele, e com a metodologia empregada na dissertação foi possível identificar a viabilidade de se utilizar o *DWT-watermarking* para a avaliação de qualidade de sinais de áudio e voz sujeitos a determinados tipos de degradação.

No Cap. 4 foram propostas alterações ao *DWT-watermarking* a fim de corrigir fraquezas que poderiam estar levando a uma degradação dos resultados. Além disso, realizou-se o mapeamento das notas para uma escala diretamente comparável com a de testes subjetivos. As iniciativas foram muito bem sucedidas, e obteve-se uma melhora significativa para parte das degradações estudadas (tanto com as alterações quanto com o mapeamento propostos). Os resultados finais obtidos—exceto para os sinais degradados por codificadores de voz—se mostraram ao menos satisfatórios, apontando para a continuação das investigações. Vale destacar que o desempenho foi excelente na avaliação de sinais degradados por adição de ruído.

5.2 Contribuições deste trabalho

Destacam-se como contribuições deste trabalho:

O estudo da DWT como domínio para avaliação de qualidade de sinais com referência.

Permitiu identificar a adequação do domínio da DWT para a avaliação de qualidade de sinais de áudio e voz submetidos a alguns tipos de degradação. Além disso, viabilizou a otimização de parâmetros para o *DWT-watermarking*.

A definição da métrica *dist* para aferição do espalhamento dos resultados, como complemento à medida de correlação.

A iniciativa tornou a avaliação dos resultados mais justa e permitiu que métodos com correlações equivalentes pudessem ter seu desempenho diferenciado.

A preparação da base com os sinais de teste e respectivas degradações.

Através da escolha de sinais e degradações controladas foi possível testar todos os métodos analisados na dissertação de forma justa e sistemática. O número de sinais utilizados permite que no futuro sejam feitos testes subjetivos a fim de comprovar os resultados obtidos com o PEAQ/WPESQ.

A estratégia para busca do melhor conjunto de parâmetros a ser utilizado no *DWT-watermarking*.

A metodologia de otimização de parâmetros foi bem sucedida, e permite que trabalhos futuros investiguem conjuntos ótimos de parâmetros para o *DWT-watermarking*. Além disso, a metodologia desenvolvida pode ser estendida para outros possíveis métodos de avaliação de qualidade com base em marcas d'água em domínios diferentes da DWT.

O esquema de otimização do fator Q_s .

Com o esquema proposto foi possível obter o fator Q_s ótimo de forma rápida. Essa maior rapidez permitiu que um conjunto maior de parâmetros pudesse ser testado em um mesmo tempo de processamento computacional.

A mudanças propostas ao *DWT-watermarking* visando ao seu aperfeiçoamento.

As modificações propostas foram testadas e se mostraram bem sucedidas. A complexidade adicionada ao método com essas modificações é desprezível.

A estratégia de mapeamento das notas do *DWT-watermarking* Modificado para ODG/MOS.

Tanto a estratégia de divisão dos conjuntos de treinamento e teste quanto os tipos de curvas testados no mapeamento se mostraram adequados. Os resultados demonstraram que a iniciativa do mapeamento foi bem sucedida.

5.3 Trabalhos futuros

Como possíveis propostas de trabalhos futuros, destacam-se:

Avaliação somente com marca: pesquisa de um método de avaliação de qualidade baseado em marcas d'água em que não seja necessário o envio de nenhuma informação adicional a não ser a marca que foi inserida. Houve tentativas para se retirar do *DWT-watermarking* a necessidade de conhecer o fator de quantização (Q_s) para realizar a avaliação de qualidade, até então não bem sucedidas. Não era possível estimar o Q_s com precisão, e qualquer erro no fator, mesmo que mínimo, impedia o funcionamento do método. Com o pré-conhecimento da marca inserida, e excluída a necessidade de enviar fator Q_s , o método de avaliação de qualidade resultante seria inteiramente sem referência.

Desempenho em outros ruídos aditivos: o excelente resultado obtido na avaliação de sinais de voz contaminados por ruído branco gaussiano qualifica o método para investigações com outros tipos de ruído.

Comparação do método com avaliação humana: a realização de testes subjetivos com indivíduos permitiria melhor avaliação do desempenho do método, pois isolaria os erros inerentes aos métodos objetivos utilizados na comparação.

Substituição da DWT: a utilização de outra transformada como núcleo do método poderia trazer resultados ainda melhores. Transformadas para domínios que tenham sido estudados do ponto de vista da psicoacústica são boas candidatas à investigação.

Especialização do método para determinados tipos de degradação: a otimização dos parâmetros para determinado tipo de degradação pode permitir que o *DWT-watermarking* alcance um desempenho muito melhor. É importante notar que no Cap. 2, por coincidência, foram escolhidos como parâmetros globais (propagados para os capítulos seguintes), os mesmos parâmetros ótimos encontrados para a avaliação de ruído. O método produziu resultados excelentes na avaliação de ruído (o melhor desempenho dentre as degradações estudadas), o que sugere que a otimização de parâmetros pode gerar resultados muito expressivos.

Referências Bibliográficas

- [1] MÖLLER, S., RAAKE, A. “Telephone speech quality prediction: Towards network planning and monitoring models for modern network scenarios”, *Speech Communication*, v. 38, pp. 47–75, Maio 2001.
- [2] PEEBLES, JR., P. Z. *Probability, Random Variables, and Random Signal Principles*. Nova Iorque, EUA, McGraw-Hill, Inc., 1993.
- [3] DE JESUS, R. A. “Análise e Implementação de um Codificador de Áudio MPEG-1 Layer III, Conforme a Norma ISO 11172-3”. . Monografia de conclusão do curso de engenharia eletrônica e de computação, DEL/Poli/UFRJ, Rio de Janeiro, Brasil, Abril 2007.
- [4] Moore, B. C. J. (Ed.). *An Introduction to the Psychology of Hearing*. 5a. ed. Londres, Reino Unido, Academic Press, 2004. ISBN: 0-12-505628-1.
- [5] Leipp, E. (Ed.). *Acoustique et Musique*. 4a. ed. Paris, França, Masson, 1984. ISBN: 2225801967.
- [6] FLETCHER, H., MUNSON, W. A. “Loudness, its definition, measurement, and calculation”, *Journal of the Acoustical Society of America*, v. 5, pp. 82–108, Out. 1933.
- [7] ROBINSON, D. W., DADSON, R. S. “A re-determination of the equal-loudness relations for pure tones”, *British Journal of Applied Physics*, v. 7, pp. 166–181, 1956.
- [8] ZWICKER, E., FASTL, H. *Psychoacoustics, Facts and Models*. Nova Iorque, EUA, Springer Verlag, 1990.
- [9] ITU-T REC. P.800. “Methods for Subjective Determination of Transmission Quality”. . International Telecommunication Union, Genebra, Suíça, 1996.
- [10] ITU-T REC. BS.1116. “Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems”. . International Telecommunication Union, Genebra, Suíça, 1997.

- [11] ITU-T REC. BS.1534. “Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems”. . International Telecommunication Union, Genebra, Suíça 2001.
- [12] KARJALAINEN, M. “A New Auditory Model for the Evaluation of Sound Quality of Audio Systems”. In: *Proc. of ICASSP’85 – IEEE International Conference on Acoustics, Speech, and Signal Processing*, v. 10, pp. 608–611, Tampa, EUA, Mar. 1985.
- [13] ITU-R REC. BS.1387. “Method for Objective Measurements of Perceived Audio Quality”. . International Telecommunication Union, Genebra, Suíça, 1998.
- [14] THIEDE, T., TREURNIET, W. C., BITTO, R., et al. “PEAQ—The ITU Standard for Objective Measurement of Perceived Audio Quality”, *Journal of the Audio Engineering Society*, v. 48, n. 1/2, pp. 3–29, Jan./Fev. 2000.
- [15] KABAL, P. *An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality*. Relatório técnico, McGill University, TSP Lab, Department of Electrical and Computer Engineering, Maio 2002.
- [16] ITU-T REC. P.862. “Perceptual Evaluation of Speech Quality (PESQ): Objective Method for End-to-end Speech Quality Assessment of Narrow Band Telephone Networks and Speech Codecs”. . International Telecommunication Union, Genebra, Suíça, 2005.
- [17] ANDERSON, J. “Methods for Measuring Perceptual Speech Quality”, . Agilent *White Paper*, Disponível em <http://cp.literature.agilent.com/litweb/pdf/5988-2352EN.pdf> (acessado em Fev. 2007), 2001.
- [18] RIX, A., BEERENDS, J. G., HOLLIER, M., et al. “PESQ - The New ITU Standard for End-to-End Speech Quality Assessment”, *109th Audio Engineering Society Convention*, Los Angeles, EUA, Set. 2000. Preprint 5360.
- [19] PSYTECHNICS LIMITED. “PESQ: An Introduction”, . Psytechnics *White Paper*, Disponível em <http://psytechnics.com/site/sections/downloads/download.php?fileID=64> (Acessado em Fev. 2007), Set. 2001.
- [20] RIX, A., BEERENDS, J. G., HOLLIER, M., et al. “Perceptual Evaluation of Speech Quality (PESQ) - A new method for speech quality assessment

- of telephone networks and codecs”. In: *Proc. of ICASSP'01 – IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 749–754, Salt Lake City, EUA, Maio 2001.
- [21] RIX, A., HOLLIER, M., HEKSTRA, A., et al. “Perceptual Evaluation of Speech Quality (PESQ), The New ITU Standard for End-to-End Speech Quality Assessment, Part I – Time-Delay Compensation”, *Journal of the Audio Engineering Society*, v. 50, n. 10, pp. 755–764, Out. 2002.
- [22] BEERENDS, J. G., HEKSTRA, A., RIX, A., et al. “Perceptual Evaluation of Speech Quality (PESQ), The New ITU Standard for End-to-End Speech Quality Assessment, Part II – Psychoacoustic Model”, *Journal of the Audio Engineering Society*, v. 50, n. 10, pp. 765–778, Out. 2002.
- [23] RIX, A. “Comparison between subjective listening quality and P.862 PESQ scores”, . Psytechnics *White Paper*, Disponível em <http://psytechnics.com/site/sections/downloads/download.php?fileID=16> (acessado em Fev. 2007), Set. 2003.
- [24] TU, R., ZHAO, J. “A novel semi-fragile audio watermarking scheme”. In: *Proc. of HAVE 2003 – The 2nd IEEE Internatioal Workshop on Haptic, Audio and Visual Environments and Their Applications*, pp. 89–94, Ottawa, Canadá, Set. 2003.
- [25] CAI, R., ZHAO, J. “Audio quality measurement by using digital watermarking”. In: *Proc. of Canadian Conference on Electrical and Computer Engineering*, v. 2, pp. 1159–1162, Ottawa, Canadá, Maio 2004.
- [26] CAI, L., ZHAO, J. “Speech Quality Evaluation: A New Application of Digital Watermarking”. In: *Proc. of IMTC'05 – IEEE Conference on Instrumentation and Measurement Technology*, v. 1, pp. 726–731, Maio 2005.
- [27] CAI, L., TU, R., ZHAO, J., et al. “Speech Quality Evaluation: A New Application of Digital Watermarking”, *IEEE Transactions on Instrumentation and Measurement*, v. 56, n. 1, pp. 45–55, Fev. 2007.
- [28] BRANDENBURG, K. “Evaluation of Quality for Audio Encoding at Low Bit Rates”, *82nd Audio Engineering Society Convention*, Fev. 1987. Preprint 2433.
- [29] BURRUS, C. S., GOPINATH, R. A., GUO, H. *Introduction to Wavelets and Wavelet Transforms: A Primer*. Upper Saddle River, EUA, Prentice-Hall, 1998. ISBN: 0-13-489600-9. With additional material and programs by Jan E. Odegard and Ivan W. Selesnick.

- [30] DAUBECHIES, I. *Ten lectures on Wavelets*. Filadelfia, EUA, Society for Industrial and Applied Mathematics, 1992.
- [31] DAUBECHIES, I. “The Wavelet Transform, Time-Frequency Localization and Signal Analysis”, *IEEE Transactions on Information Theory*, v. 36, n. 5, pp. 961–1005, Set. 1990.
- [32] Kahrs, M., Brandenburg, K. (Eds.). *Applications of digital signal processing to audio and acoustics*. Norwell, EUA, Kluwer, 1998. ISBN: 0-7923-8130-0.
- [33] P.SUP23: ITU-T CODED-SPEECH DATABASE. Disponível em <http://www.itu.int/rec/T-REC-P.Sup23/en>. Acessado em out. 2008.
- [34] SPEEX: A FREE CODEC FOR FREE SPEECH. Disponível em <http://www.speex.org/>. Acessado em out. 2008.
- [35] KABAL, P. Conjunto de rotinas “AFsp”, Disponível em <http://www-mmsp.ece.mcgill.ca/Documents/Software/Packages/AFsp/AFsp.html>. Acessado em out. 2008.
- [36] BENEDETTO, F., GIUNTA, G., NERI, A., et al. “Digital Audio Watermarking for QoS Assessment of MP3 Music Signals”. In: *Proc. of 14th European Signal Processing Conference - EUSIPCO*, Florença, Italia, Set 2006.
- [37] ARYA, A. K., LATHER, J. S., DEWAN, L. “A Tool for Audio Quality Evaluation Under Hostile Environment”, *International Journal of Computer and Information Science and Engineering*, v. 2, n. 4, pp. 240–250, 2008.
- [38] ITU-R REC. BS.1116-1. “Methods for the Subjective Assessment of small Impairments in Audio Systems including Multichannel Sound Systems”. . International Telecommunication Union, Genebra, Suíça, Out., 1997.

Apêndice A

Otimização de Q_s

O pseudocódigo abaixo apresenta a forma como se realiza a otimização de Q_s . Utilizam-se limites superiores e inferiores para Q_s que são inicializados em valores extremos (10000 e 1, respectivamente), assim como se utilizam passos variados de acordo com a diferença encontrada.

Ressalta-se que a otimização é feita analisando-se a nota de avaliação do sinal marcado em primeiro lugar, para em seguida analisar-se a distorção da marca extraída.

As constantes utilizadas no início do código (89, -50, 29 etc.) foram escolhidas de forma arbitrária apenas para acelerar o processo de busca do Q_s ótimo. Elas não garantem a convergência do código, que é garantida através do estabelecimento dos limites máximo e mínimo durante a otimização.

```
QsAnterior = Qs;
if (nota < notaDesejada - 0.5)
    Qs = Qs + 89;
elseif (nota > notaDesejada + 0.5)
    Qs = Qs - 50;
elseif (nota < notaDesejada - 0.1)
    Qs = Qs + 29;
elseif (nota > notaDesejada + 0.1)
    Qs = Qs - 15;
elseif (nota < notaDesejada - 0.05)
    Qs = Qs + 19;
elseif (nota > notaDesejada + 0.05)
    Qs = Qs - 10;
elseif (nota < notaDesejada - 0.01)
    Qs = Qs + 3;
elseif (nota > notaDesejada + 0.01)
```

```

        Qs = Qs-2;
else
    if (distorcao > (distorcaoMaxima))
        Qs = Qs-1;
    endif
endif

if (Qs>QsAnterior)
    QsMin = QsAnterior;
endif
if (Qs<QsAnterior)
    QsMax = QsAnterior;
endif
if (Qs<QsMin)
    Qs = QsMin+1;
    QsMin = Qs;
endif
if (Qs>QsMax)
    Qs = QsMax-1;
    QsMax = Qs;
endif
if (abs(QsMin-QsMax)<=1)
    fimDaOtim = sim;
endif

```

Apêndice B

Parâmetros obtidos para o mapeamento do PCEWM

A seguir serão apresentados os parâmetros obtidos através de otimização para o mapeamento do PCEWM (*DWT-watermarking* Modificado) nas notas MOS (WPESQ) e ODG (PEAQ).

Para o mapeamento, conforme explicado na Subseção 4.3.1, foram utilizados os seguintes modelos:

Polinomial —Utilizando-se um polinômio do quarto grau, são necessários 5 parâmetros (α , β , γ , δ e η) para defini-lo:

$$y = \alpha x^4 + \beta x^3 + \gamma x^2 + \delta x + \eta.$$

Exponencial —Utilizando-se uma curva exponencial, são utilizados 3 parâmetros (a, b e c) para defini-la:

$$y = a + be^{cx}.$$

Nas tabelas a seguir, serão apresentados os parâmetros para todas as degradações nas quais o *DWT-watermarking* Modificado se mostrou adequado para efetuar a avaliação da qualidade¹. Apresentar-se-ão tanto os parâmetros obtidos com a estratégia de divisão dos sinais quanto os obtidos com a divisão das degradações.

¹Foram excluídos os parâmetros obtidos com os codificadores de voz.

Tabela B.1: Parâmetros obtidos para o mapeamento polinomial.

Teste	α	β	γ	δ	η
Voz – Ruído					
Divisão dos sinais	25,28	-53,07	36,16	-5,39	1,04
Divisão das degradações	23,60	-48,98	33,37	-4,96	1,05
Voz – Perda de Pacotes					
Divisão dos sinais	-5,15	18,17	-11,17	2,13	1,00
Divisão das degradações	19,58	-28,23	15,11	-2,53	1,03
Voz – Filtragem passa-baixas					
Divisão dos sinais	-33,97	70,10	-47,47	15,31	1,03
Divisão das degradações	-39,26	80,27	-53,46	16,39	1,06
Áudio – Codificação MP3					
Divisão dos sinais	33,75	-78,68	56,86	-8,23	-3,82
Divisão das degradações	30,13	-69,38	49,11	-6,12	-3,83

Tabela B.2: Parâmetros obtidos para o mapeamento exponencial.

Teste	a	b	c
Voz – Ruído			
Divisão dos sinais	-12,18	12,72	0,236
Divisão das degradações	-5,182	5,780	0,462
Voz – Perda de Pacotes			
Divisão dos sinais	0,822	0,0480	4,481
Divisão das degradações	0,995	0,0134	5,696
Voz – Filtragem passa-baixas			
Divisão dos sinais	0,0015	2,232	0,905
Divisão das degradações	-16,60	18,69	0,153
Áudio – Codificação MP3			
Divisão dos sinais	5,219	-10,59	-0,887
Divisão das degradações	3,179	-8,371	-1,146

Apêndice C

Gráficos do mapeamento do PCEWM

A seguir serão apresentados os gráficos resultantes dos mapeamentos do PCEWM (*DWT-watermarking* Modificado) nas notas MOS (WPESQ) e ODG (PEAQ). Os círculos representam cada uma das notas obtidas para um mesmo arquivo, enquanto as linhas representam a curva que melhor realiza o mapeamento dos eixos, de acordo com o critério de minimização do erro quadrático médio.

Para possibilitar melhor comparação, o eixo das abscissas, que contém as notas do *DWT-watermarking* Modificado, foi colocado por completo em todos os gráficos (de 0 a 1), enquanto o eixo das ordenadas (MOS/ODG) foi deixado livre para que se pudesse visualizar a curva de mapeamento encontrada por completo, mesmo nos casos em que a curva obtida foi bastante inadequada devido à descorrelação entre as notas utilizadas para sua construção. Os trechos em que as curvas deixam de ser monotônicas são aqueles que não possuem notas mapeadas, por serem muito altas ou muito baixas; nesses trechos toma-se como mapeamento o valor máximo/mínimo.

Maiores informações sobre o procedimento de otimização podem ser obtidos na Subseção 4.3.1. As métricas de avaliação resultantes de cada mapeamento podem ser obtidas na Subseção 4.3.2.

Com os gráficos é possível efetuar tanto uma comparação entre o mapeamento polinomial e exponencial obtido para cada conjunto de pares de notas, como também uma comparação entre o conjunto de pares de notas obtidos através da divisão dos sinais e o conjunto obtido através da divisão das degradações.

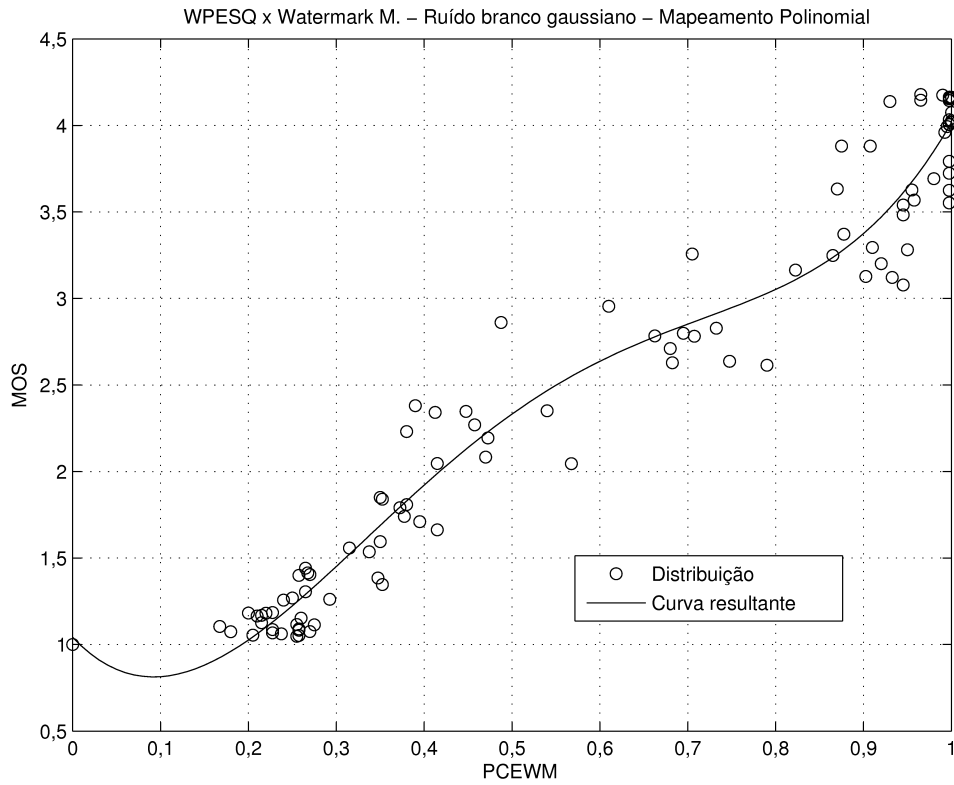


Figura C.1: Mapeamento polinomial e divisão dos sinais — sinais de voz contaminados por ruído branco.

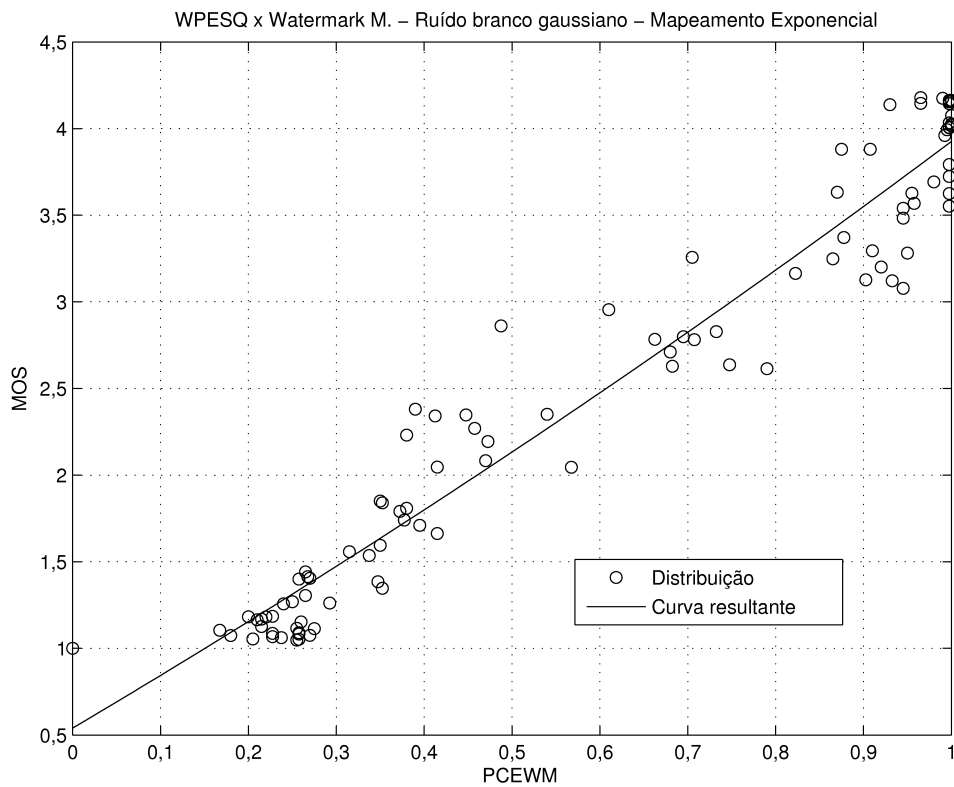


Figura C.2: Mapeamento exponencial e divisão dos sinais — sinais de voz contaminados por ruído branco.

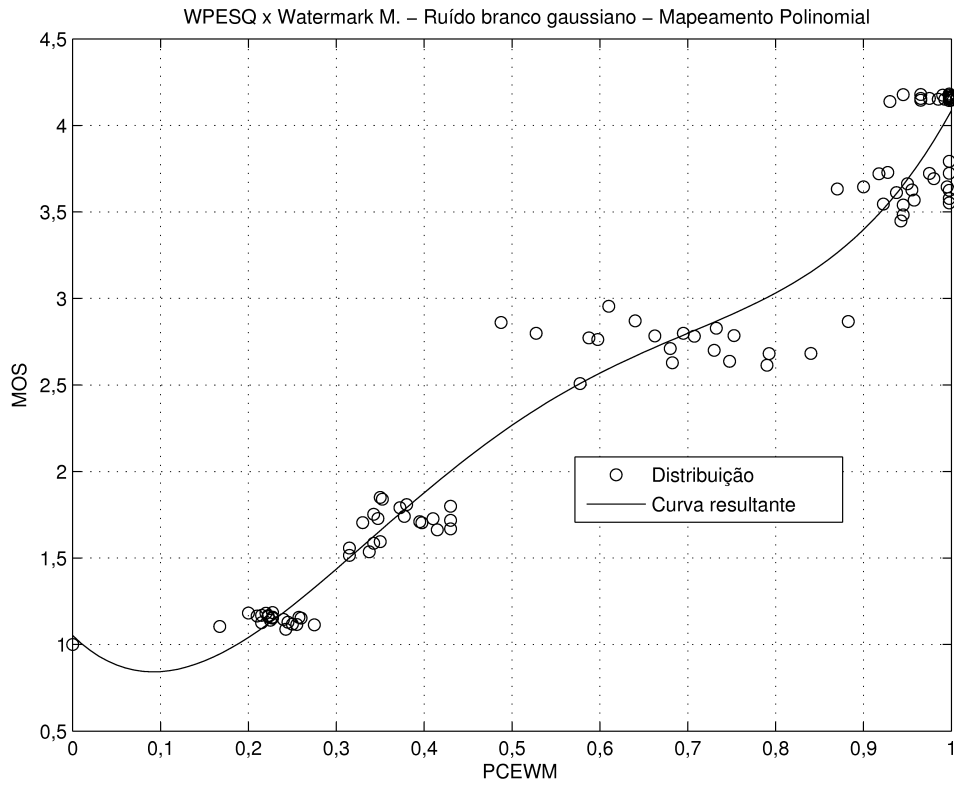


Figura C.3: Mapeamento polinomial e divisão das degradações — sinais de voz contaminados por ruído branco.

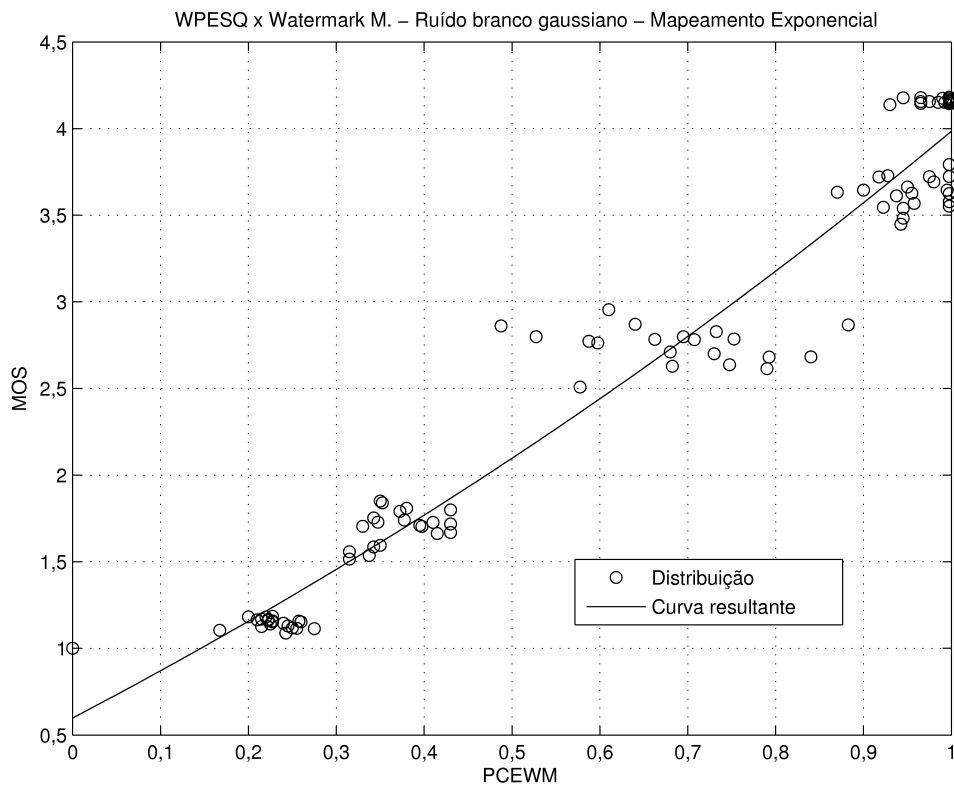


Figura C.4: Mapeamento exponencial e divisão das degradações — sinais de voz contaminados por ruído branco.

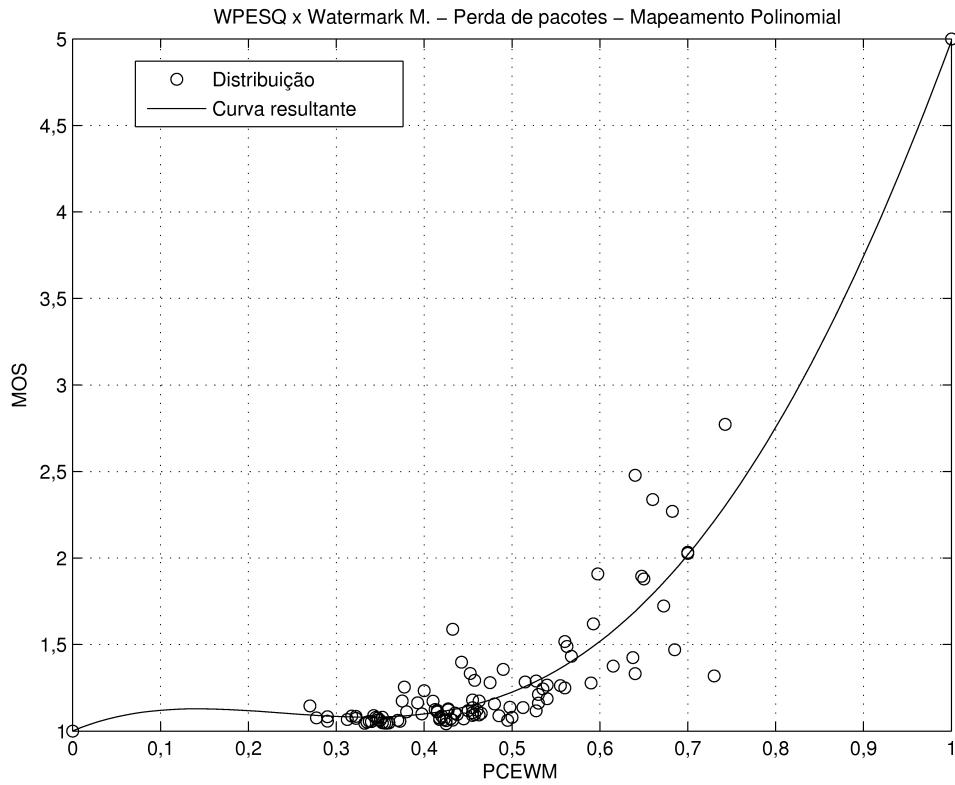


Figura C.5: Mapeamento polinomial e divisão dos sinais — sinais de voz contaminados por perda de pacotes.

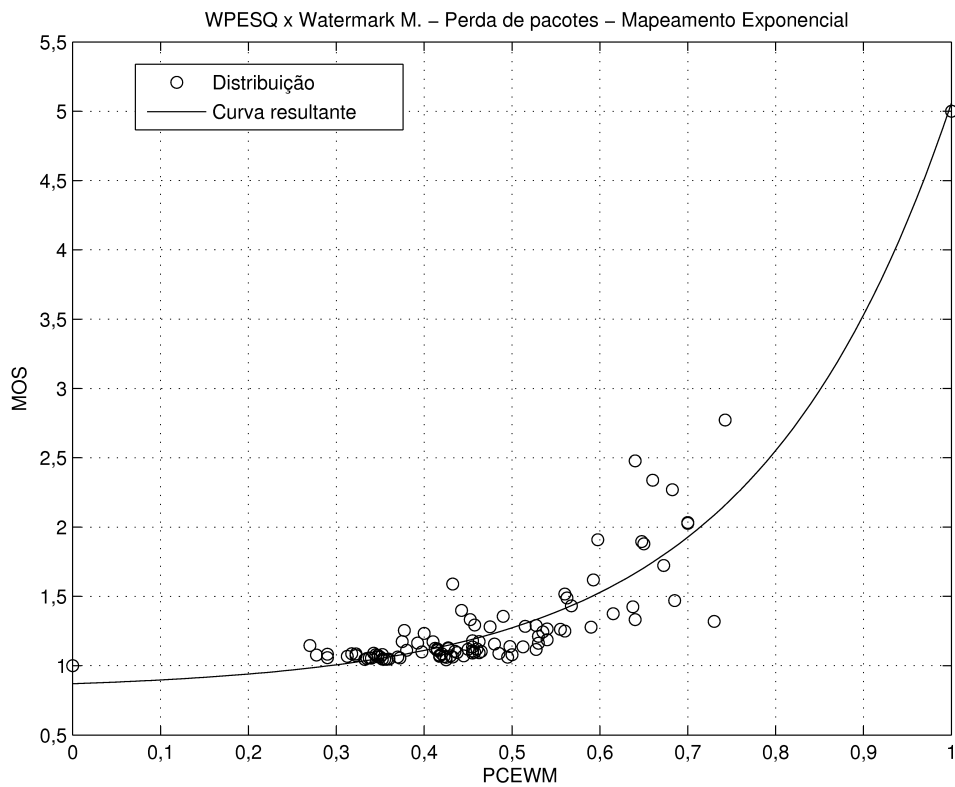


Figura C.6: Mapeamento exponencial e divisão dos sinais — sinais de voz contaminados por perda de pacotes.

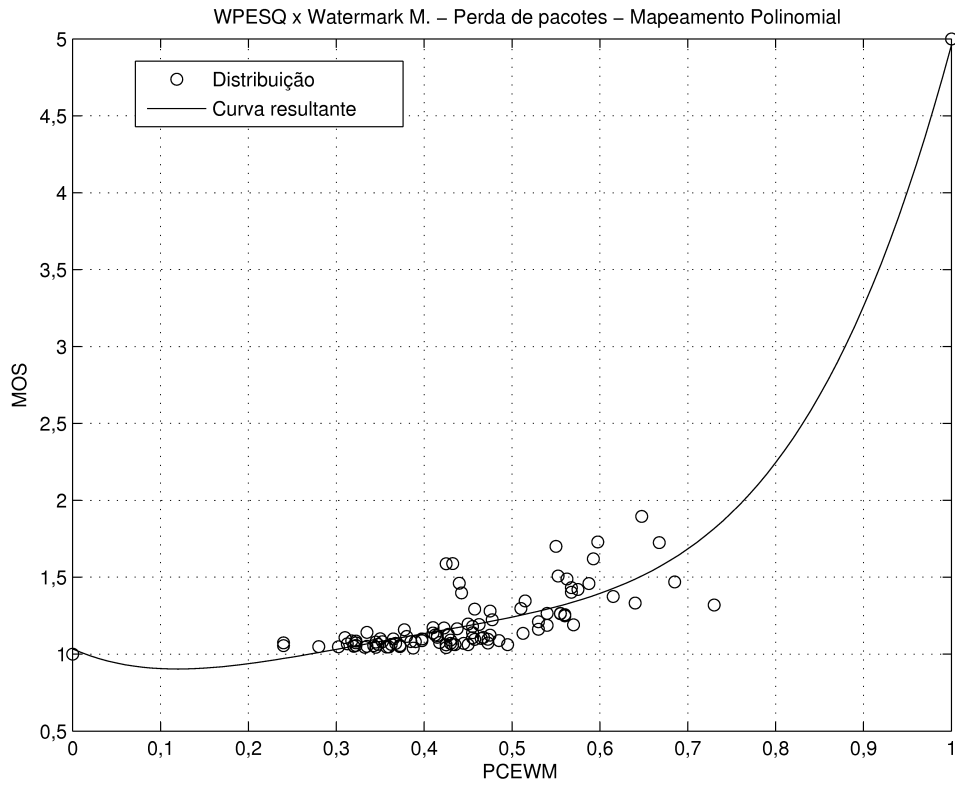


Figura C.7: Mapeamento polinomial e divisão das degradações — sinais de voz contaminados por perda de pacotes.

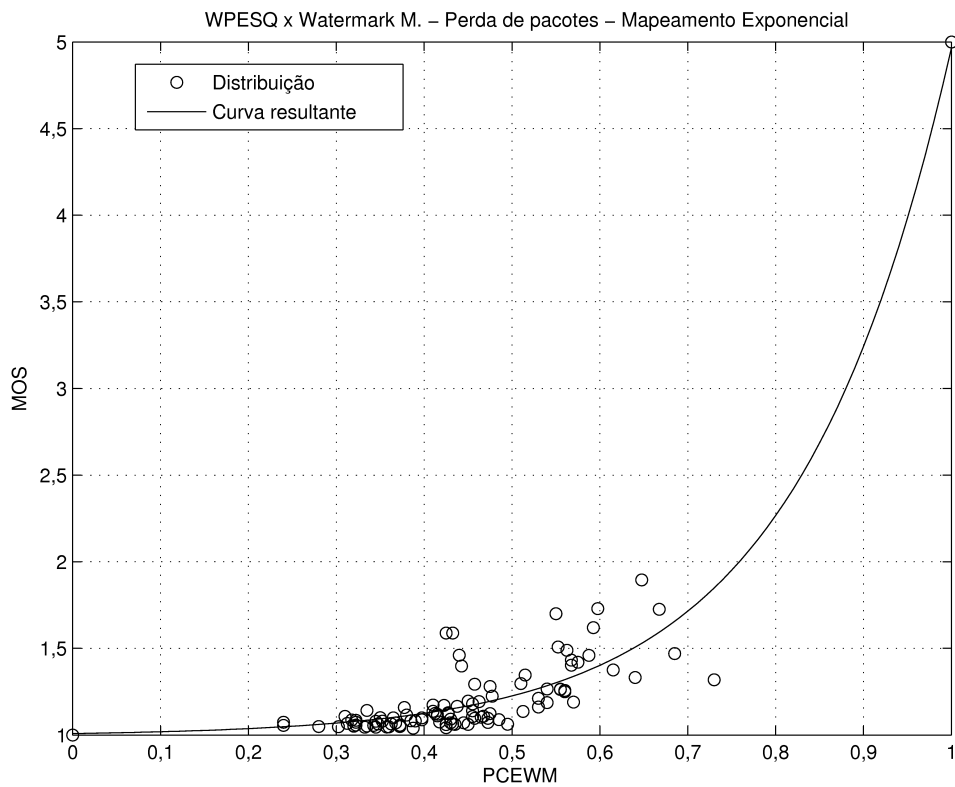


Figura C.8: Mapeamento exponencial e divisão das degradações — sinais de voz contaminados por perda de pacotes.

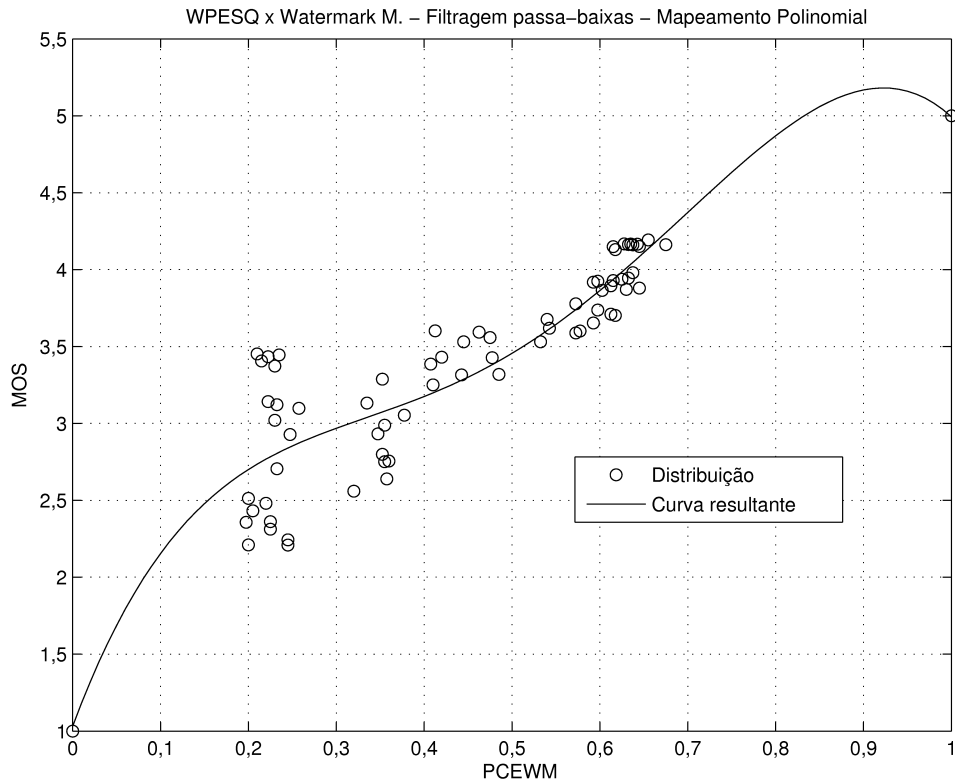


Figura C.9: Mapeamento polinomial e divisão dos sinais — sinais de voz contaminados por filtragem passa-baixas.

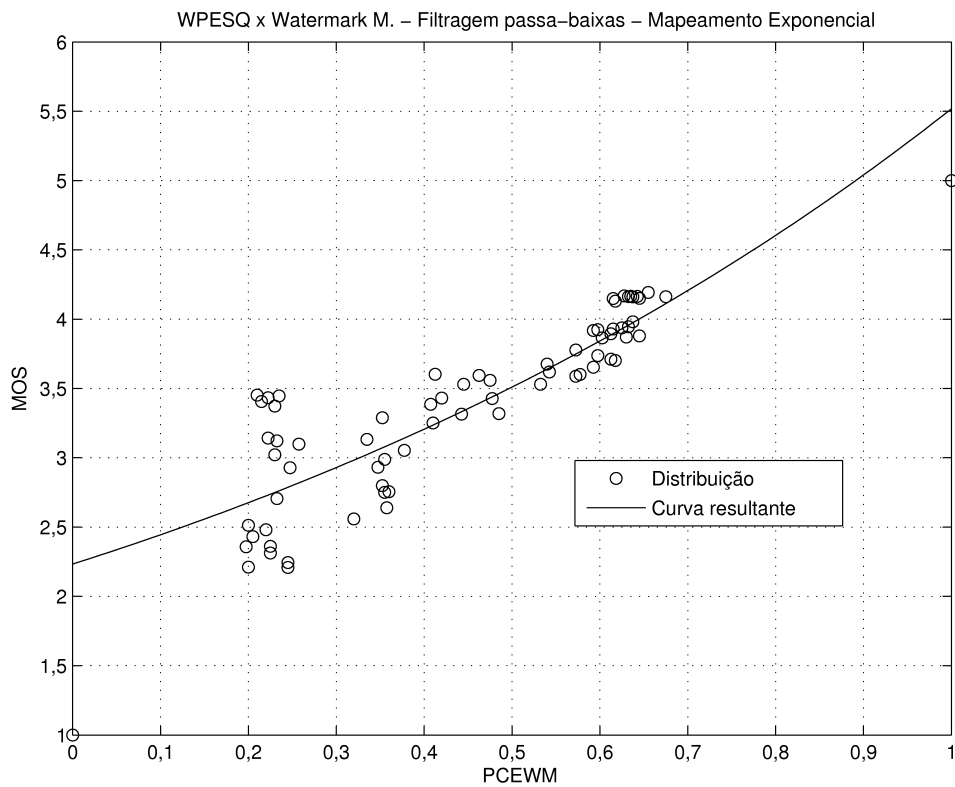


Figura C.10: Mapeamento exponencial e divisão dos sinais — sinais de voz contaminados por filtragem passa-baixas.

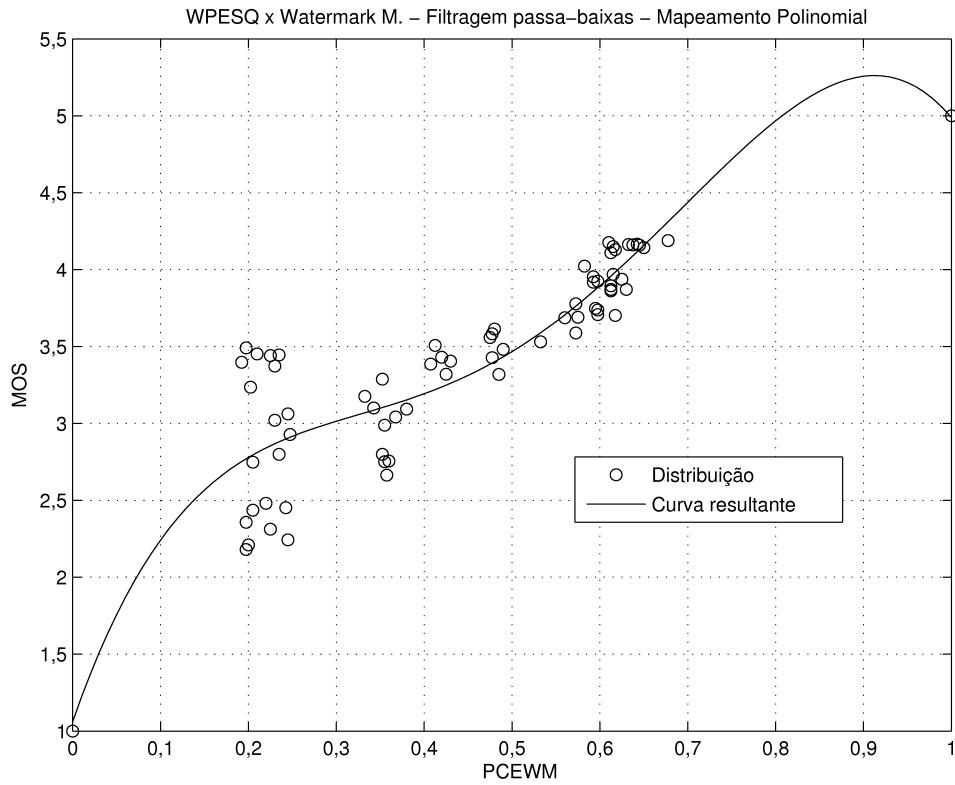


Figura C.11: Mapeamento polinomial e divisão das degradações — sinais de voz contaminados por filtragem passa-baixas.

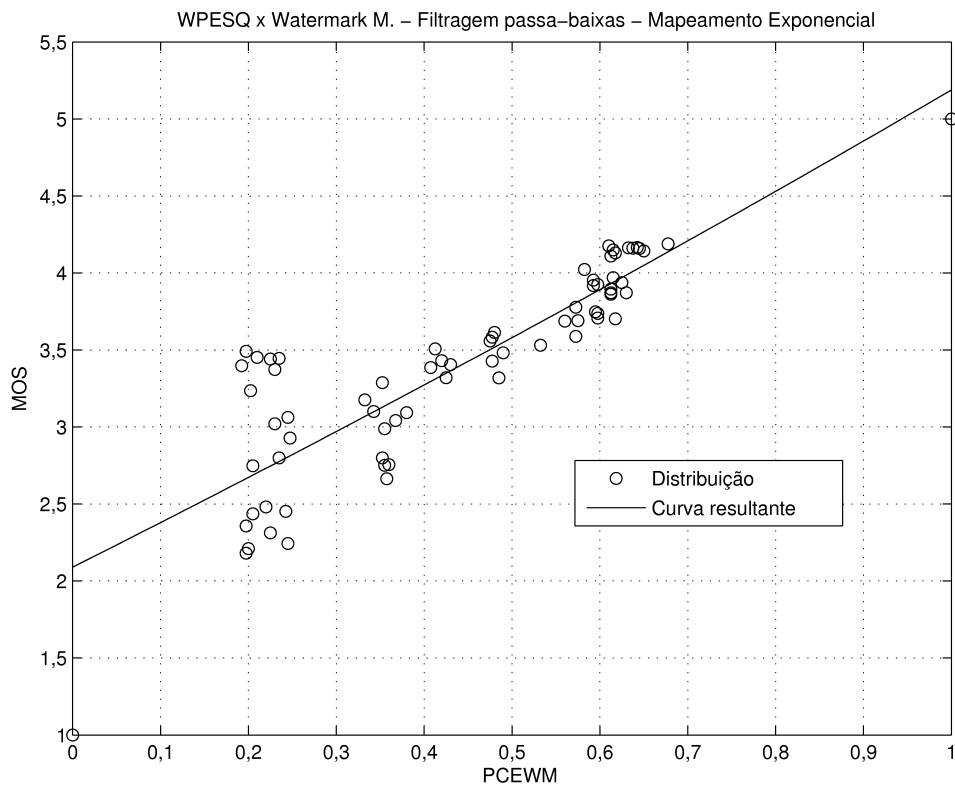


Figura C.12: Mapeamento exponencial e divisão das degradações — sinais de voz contaminados por filtragem passa-baixas.

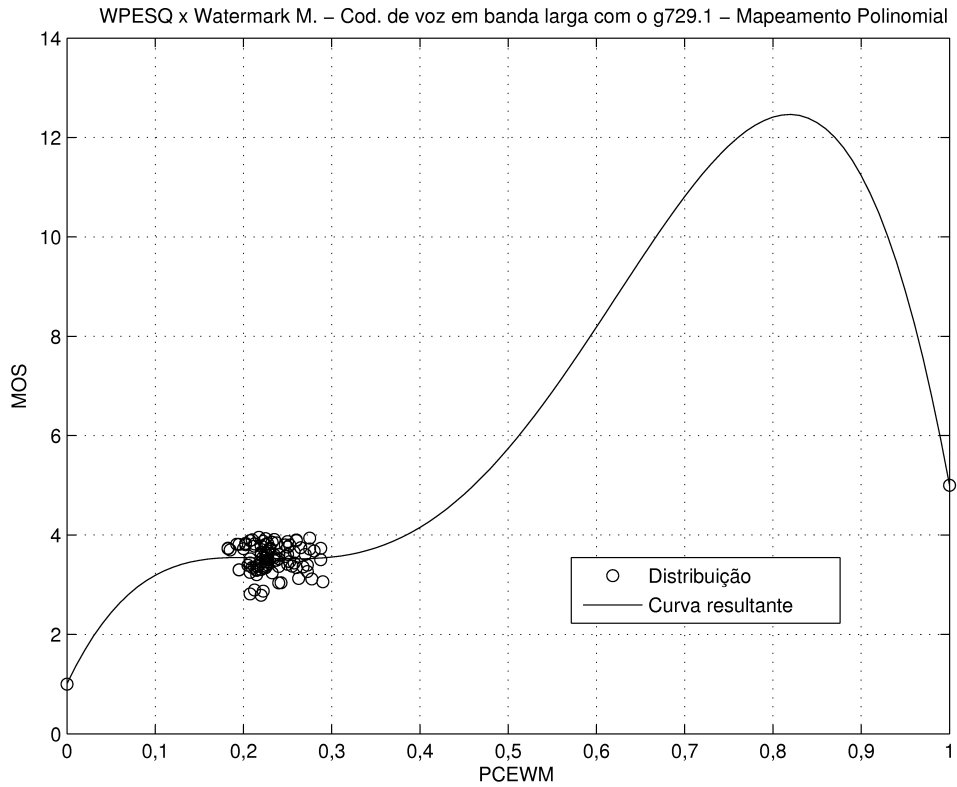


Figura C.13: Mapeamento polinomial e divisão dos sinais — sinais de voz contaminados pelo codificador G.729.1.

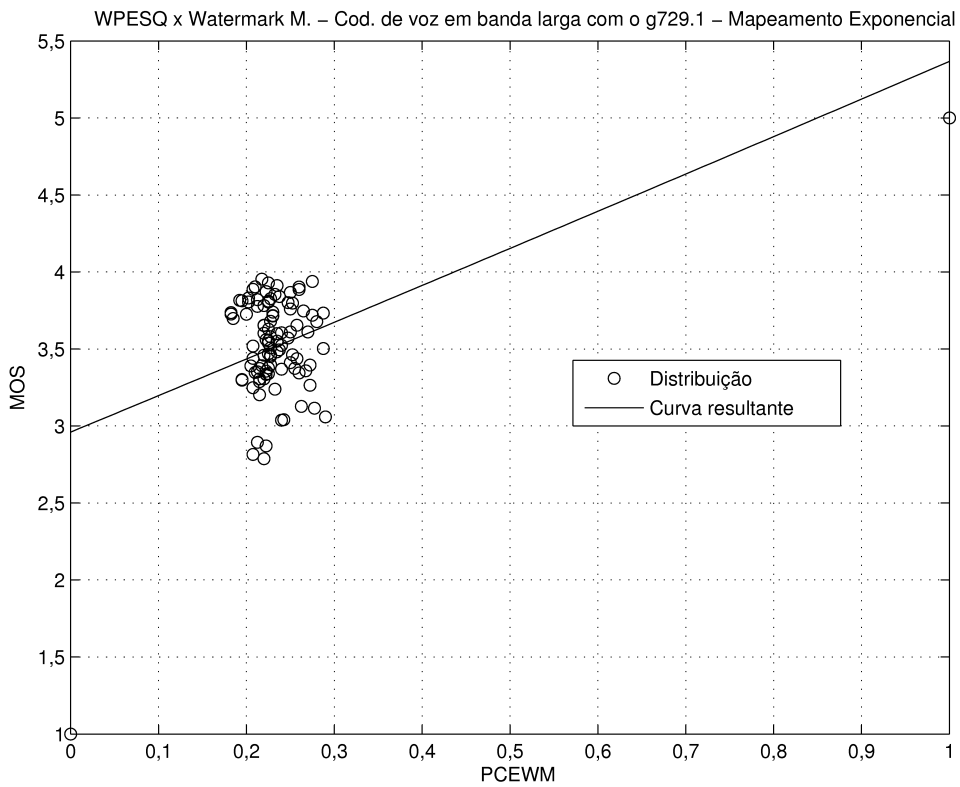


Figura C.14: Mapeamento exponencial e divisão dos sinais — sinais de voz contaminados pelo codificador G.729.1.

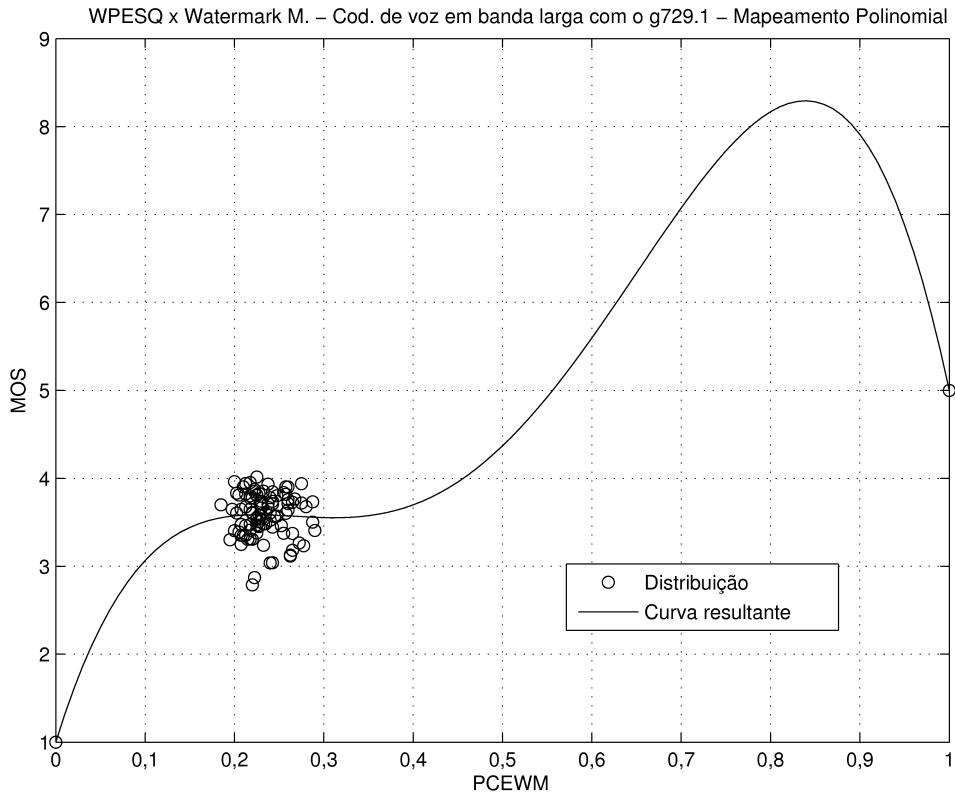


Figura C.15: Mapeamento polinomial e divisão das degradações — sinais de voz contaminados pelo codificador G.729.1.

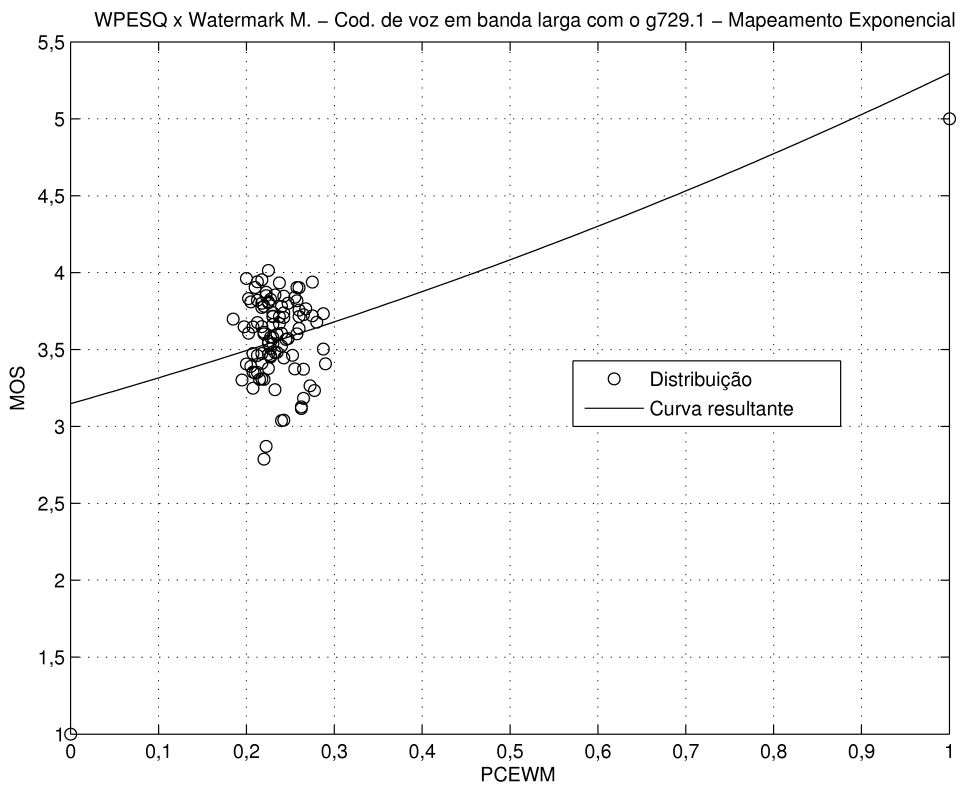


Figura C.16: Mapeamento exponencial e divisão das degradações — sinais de voz contaminados pelo codificador G.729.1.

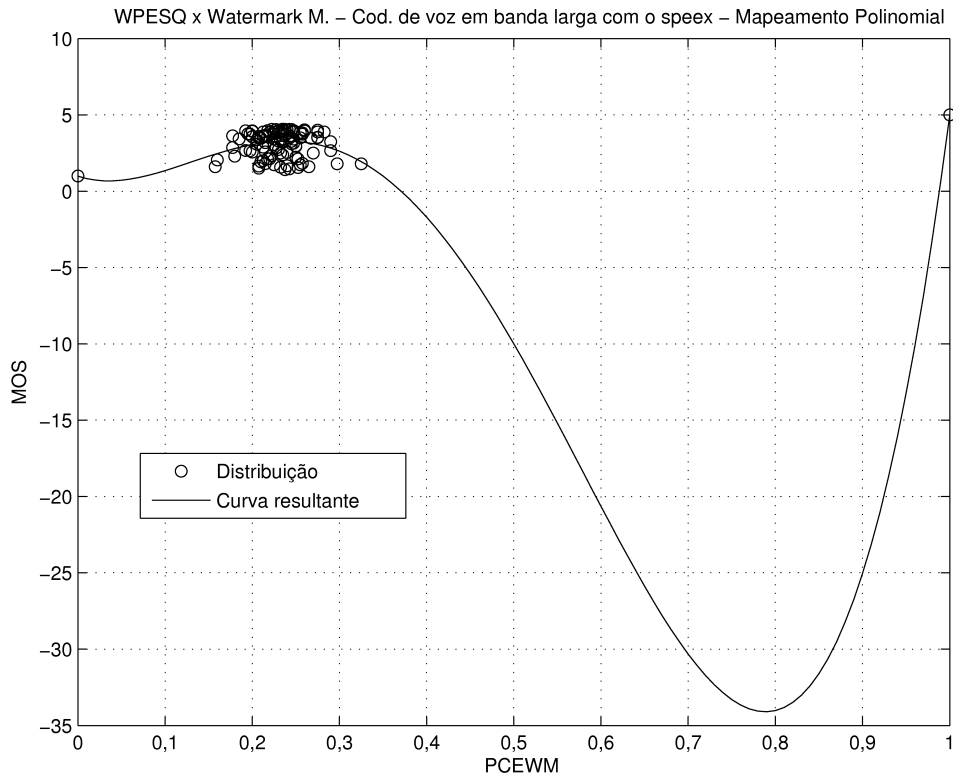


Figura C.17: Mapeamento polinomial e divisão dos sinais — sinais de voz contaminados pelo codificador speex.

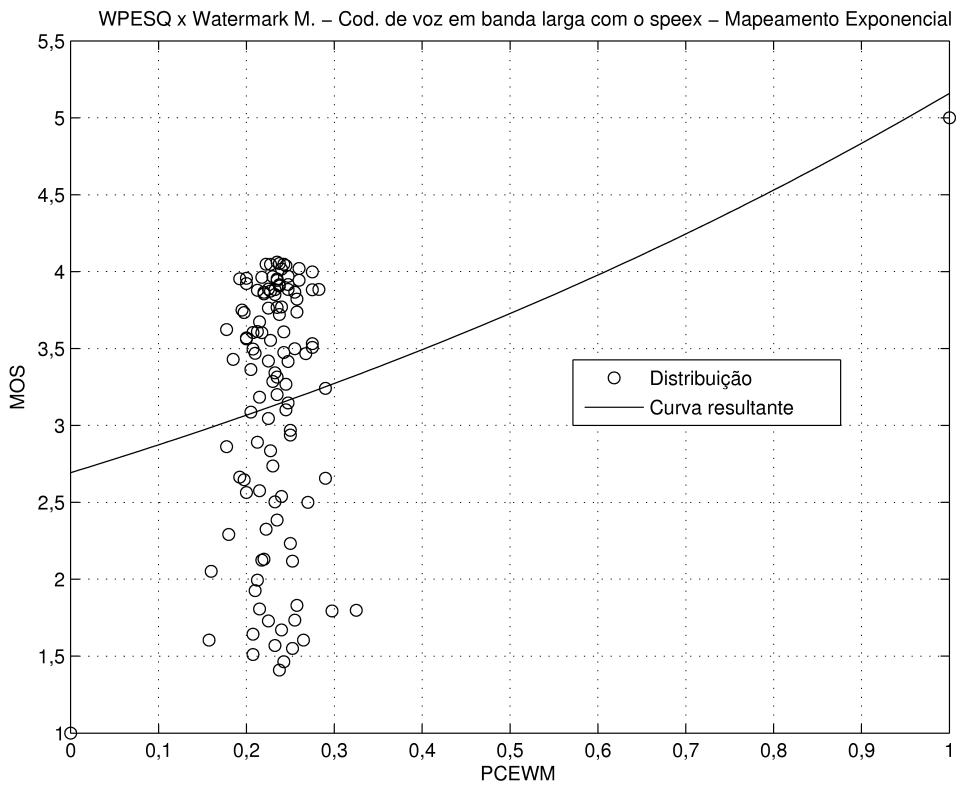


Figura C.18: Mapeamento exponencial e divisão dos sinais — sinais de voz contaminados pelo codificador speex.

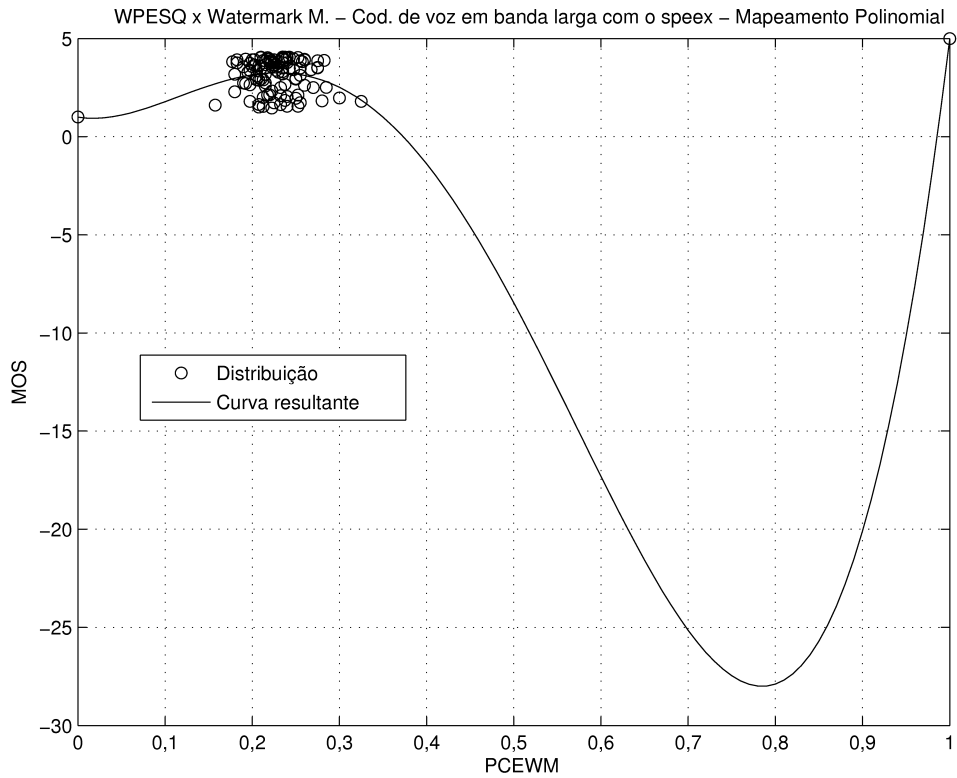


Figura C.19: Mapeamento polinomial e divisão das degradações — sinais de voz contaminados pelo codificador speex.

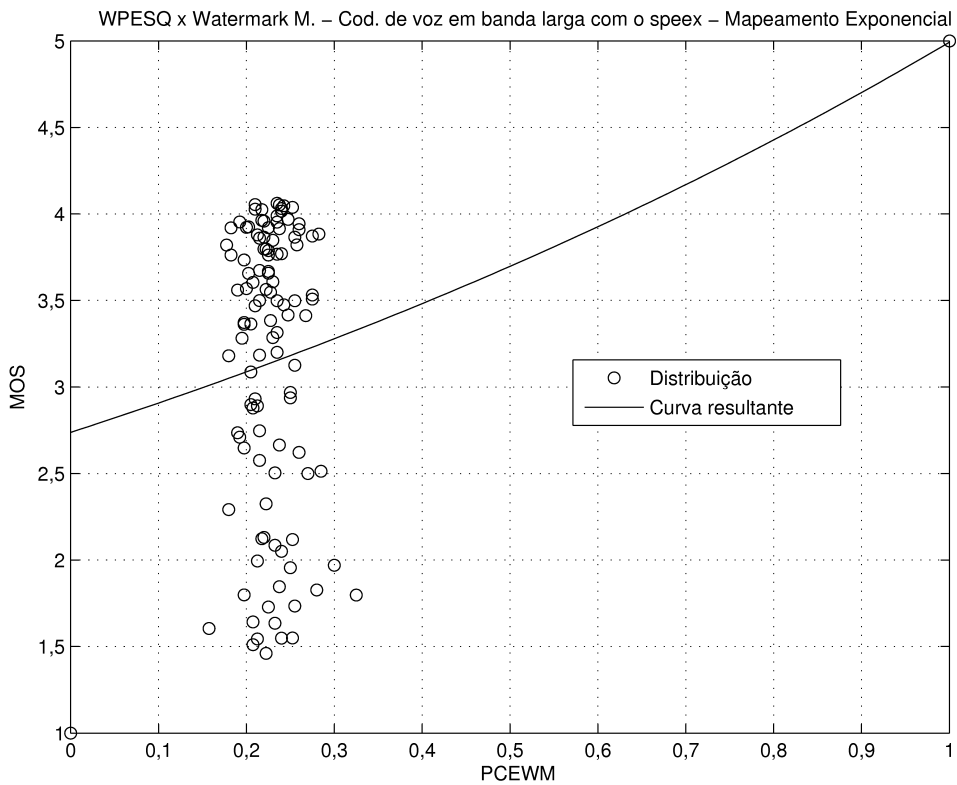


Figura C.20: Mapeamento exponencial e divisão das degradações — sinais de voz contaminados pelo codificador speex.

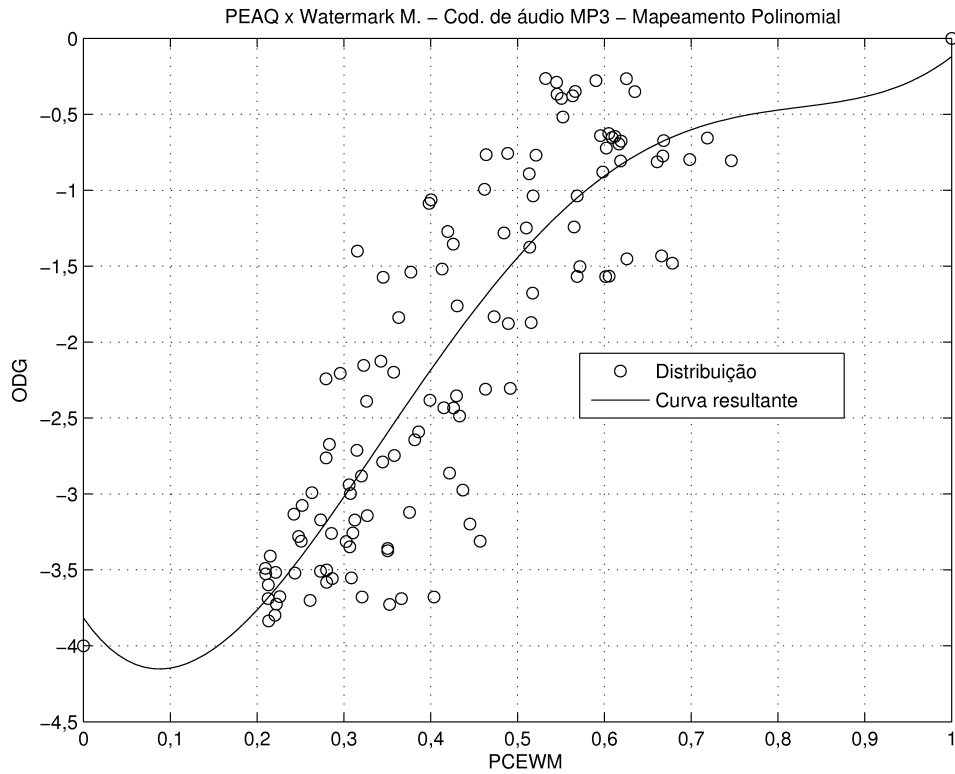


Figura C.21: Mapeamento polinomial e divisão dos sinais — sinais de áudio degradados por codificação MP3.

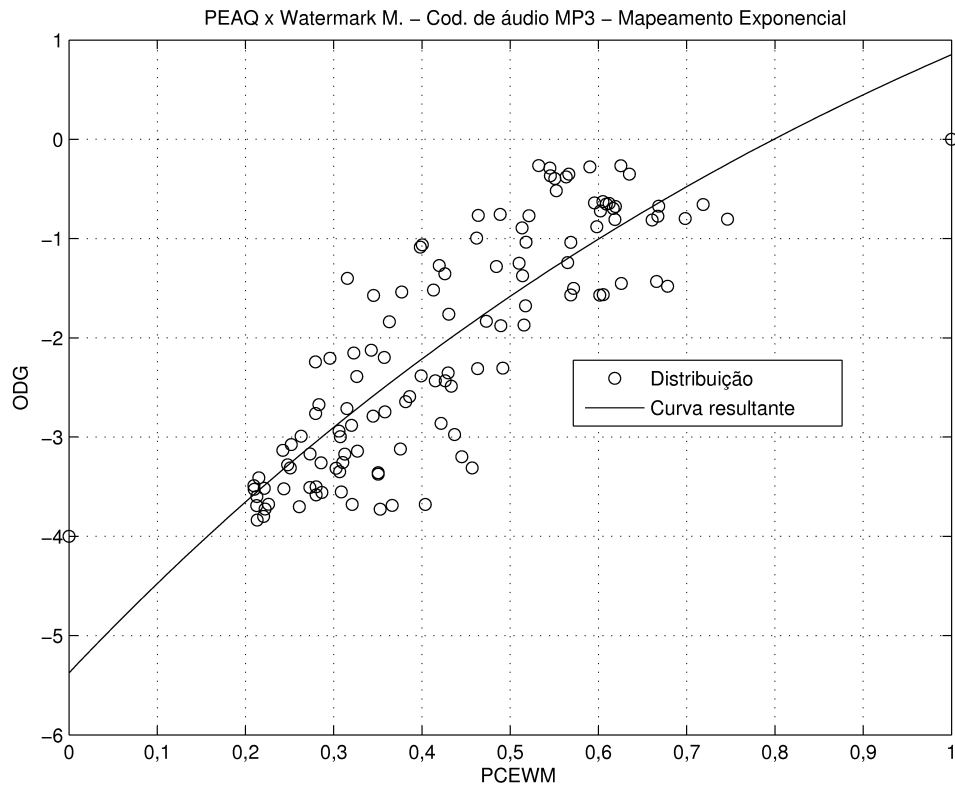


Figura C.22: Mapeamento exponencial e divisão dos sinais — sinais de áudio degradados por codificação MP3.

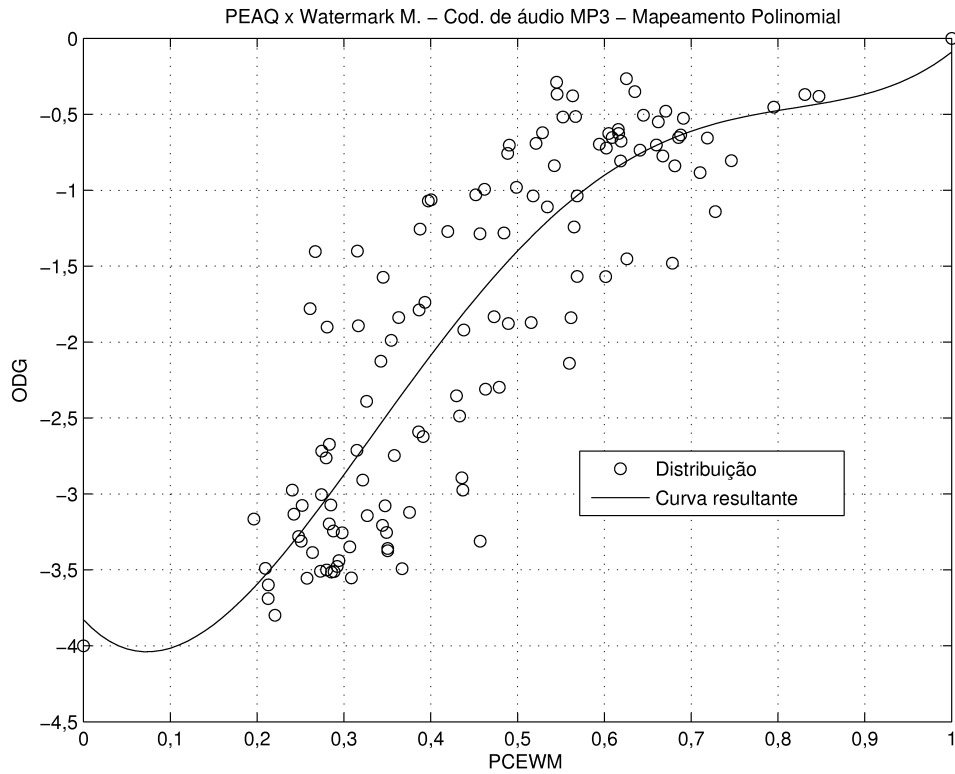


Figura C.23: Mapeamento polinomial e divisão das degradações — sinais de áudio degradados por codificação MP3.

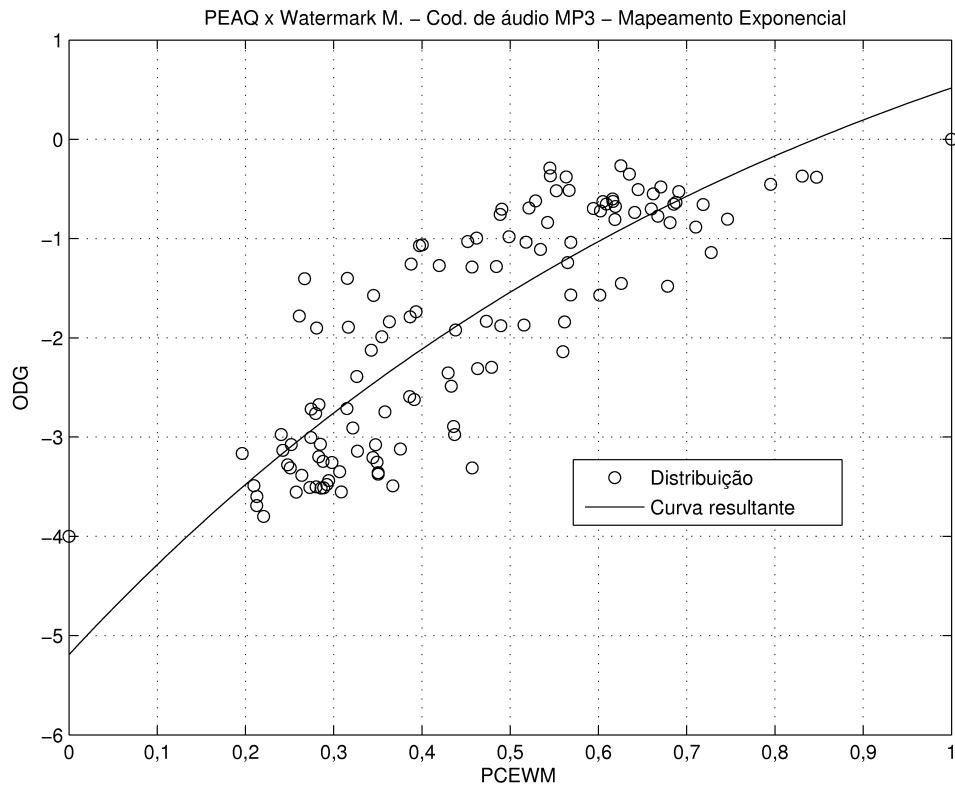


Figura C.24: Mapeamento exponencial e divisão das degradações — sinais de áudio degradados por codificação MP3.