

Universidade Federal do Rio de Janeiro

Escola Politécnica

Departamento de Eletrônica e de Computação

**Desreverberação de sinais de voz baseada em critério
perceptual**

Autor:

José María Fernández del Vallado

Orientador:

Prof. Sergio Lima Netto, Ph. D.

Orientador:

Prof. Amaro Azevedo de Lima, Ph. D.

Examinador:

Prof. Eduardo Antônio Barros da Silva, Ph. D.

Examinador:

Prof. Thiago de Moura Prego, D. Sc.

DEL

Julho de 2012

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Escola Politécnica - Departamento de Eletrônica e de Computação

Centro de Tecnologia, bloco H, sala H-217, Cidade Universitária

Rio de Janeiro - RJ CEP 21949-900

Este exemplar é de propriedade da Universidade Federal do Rio de Janeiro, que poderá incluí-lo em base de dados, armazenar em computador, microfilmear ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es) e do(s) orientador(es).

DEDICATÓRIA

Dedico este projeto a todas as pessoas que tornaram possíveis meus estudos no Brasil; todas aquelas pessoas que mais quero: Aos meus pais, aos meus avós, aos meus irmãos, aos meus tios e primos, e a todos meus amigos, por ser todos eles pessoas das que me sento muito orgulhoso e por me ajudar sempre. Especialmente dedico este projeto a Paula, por ser a minha companheira e melhor amiga, e seguir me querendo tanto, ainda a distância.

AGRADECIMENTO

Meus sinceros agradecimentos:

- ao Professor Sergio Lima Netto, por todo o seu interesse e o tempo dedicado, o ensino efetuado, e a quantidade de coisas que com ele aprendi.
- ao Professor Amaro Azevedo de Lima, por me ajudar sempre com o projeto, ser a guia do meu trabalho e por todo o tempo que me dedicou.
- aos alunos e professores da UFRJ que sempre me ajudaram no laboratório e amenizaram o tempo brincando comigo.
- a meu amigo Raúl, por me ajudar sempre e ser a melhor companhia nesses momentos duros da aprendizagem.

José María Fernández del Vallado

RESUMO

A reverberação é um efeito acústico que ocorre nos cenários fechados, onde o tempo de recebimento de dois sons é pequeno. Nos últimos anos, múltiplos estudos tentaram avaliar a percepção deste efeito da forma mais aproximada possível. Uma das tentativas veio de mãos de professores desta Universidade, que desenvolveram o sistema QAreverb baseado na métrica de Allen, e melhoraram a mesma.

Este trabalho se inicia com uma nova métrica de avaliação da reverberação a partir do sistema QAreverb, aumentando ainda mais a correlação da medida final com a avaliação subjetiva dos sinais de voz com reverberação.

Estes sinais de voz provêm de duas bases de dados diferentes, a base de dados MARDY, que contém 32 sinais reverberantes, e a base de dados NBP, que contém 204 sinais reverberantes e foi desenvolvida como parte do sistema QAreverb.

Continuando com o teste do sistema QAreverb modificado, novas funções de mapeamento entre a medida bruta e a medida ajustada para a escala MOS foram desenvolvidas, melhorando em quase todos os casos a correlação da medida mapeada com a avaliação subjetiva para a base de dados NBP, com a devida validação para a base de dados MARDY.

Em segundo lugar, a partir desta nova métrica, aperfeiçoamos o algoritmo de desreverberação de Wu e Wang, aumentando sua eficácia em termos da nova medida de qualidade desenvolvida anteriormente neste projeto.

Palavras-Chave: Reverberação, QAreverb, MARDY, NBP, Filtragem Inversa, Desreverberação.

ABSTRACT

Reverberation is an acoustic effect associated to closed environments, where listeners receive two sounds in a short time interval. Multiple studies were performed during the last years trying to assess reverberation perception as closely as possible. One of those studies came from UFRJ's researchers, who developed the QAreverb system based on Allen's metric, and improved it.

This paper begins with the development of a new evaluation metric based on the QAreverb algorithm, which improves even more the correlation between the final measure and the subjective measure of reverberant speech signals.

Continuing with the assess of these new metric, some experiments were made, consisting on the development of new mapping functions that adjusted the obtained measure to the MOS-scale measure.

In all these experiments were consider two different signal databases: the MARDY database, which contains 32 reverberated signals, and the New Brazilian Portuguese database, which contains 204 reverberated signals and was fully developed for the QAreverb system.

Finally, we studied the performance of the Wu-Wang two-stage dereverberation algorithm based on these new metric, and we improved sound quality of the received signals.

Key-words: Reverberation, QAreverb, MARDY, NBP, Inverse-Filtering, Dereverberation.

SIGLAS

UFRJ - Universidade Federal do Rio de Janeiro

DRER - *Direct to Reverberant Energy Ratio*

EDC - *Energy Decay Curve*

LMS - *Last Mean Squares*

MOS - *Mean Objective Score*

MSE - *Mean Square Error*

NBP - *New Brazilian Portuguese*

PESQ - *Perceptual Evaluation of Speech Quality*

RIR - *Room Impulse Response*

RLS - *Recursive Least-Squares*

RSV - *Room Spectral Variance*

RT - *Reverberation Time*

S-M - *Set-Membership*

SRMR - *Standard Room Mean-Square Residual*

Sumário

1	Introdução	1
1.1	Apresentação Inicial	1
1.2	Organização do Texto	2
2	Avaliação Objetiva da Reverberação	4
2.1	Introdução	4
2.2	Conceito de Reverberação	5
2.2.1	Tempo de reverberação	7
2.2.2	Variância espectral da sala	7
2.2.3	Razão das energias direta e reverberante	8
2.3	Bases de Dados de Sinais Reverberantes	8
2.3.1	Base de dados MARDY	8
2.3.2	Nova base de voz reverberante	9
2.4	Avaliação Objetiva de Sinais Reverberantes	10
2.4.1	Algoritmos anteriores	10
2.4.2	QAreverb	11
2.4.3	QAreverb modificado	11
3	Avaliação da Nova Métrica QAreverb	13
3.1	Introdução	13
3.2	Projeto da QAreverb Modificada	13
3.2.1	Otimização de Q_m	14
3.2.2	Mapeamento não-linear	15
3.2.3	Mapeamento do tipo PESQ	19
3.2.4	Mapeamento PESQ modificado	21

3.3	Conclusão	25
4	Aspectos Teóricos da Desreverberação	26
4.1	Introdução	26
4.2	Algoritmo de Desreverberação de Wu e Wang	27
4.2.1	Filtragem inversa no domínio do tempo	28
4.2.2	Filtragem inversa no domínio da frequência	29
4.3	Filtragem de Wiener	30
4.4	Algoritmo LMS	31
4.5	Aplicação Prática da Filtragem de Wiener	32
4.6	Algoritmo Adaptativo Alternativo	34
5	Modificações Propostas para o algoritmo de Wu-Wang	36
5.1	Introdução	36
5.2	Considerações Iniciais	37
5.3	Modificações do algoritmo de Wu-Wang no domínio do tempo	37
5.4	Conclusão	42
6	Conclusão	44
6.1	Propostas para estudos futuros	45
	Bibliografia	45

Lista de Figuras

2.1	Representação esquemática do efeito de reverberação.	5
2.2	Exemplo de RIR associada ao efeito de reverberação.	6
3.1	Exemplos de mapeamentos não-lineares com mínimos locais no intervalo $[1,5]$, correspondendo a funções não monotonicamente crescentes.	17
3.2	Notas subjetivas da avaliação da reverberação para os 204 sinais da base de dados NBP ('x' conectados) e notas Q_{MOS} modificadas para a configuração $\alpha = 1$, $\beta = 0$ e $\gamma = 0,35$ ('•'), com mapeamento de terceira ordem.	18
3.3	Polinômio de terceira ordem com mínimo local no intervalo $[1,5]$, para o caso de $\alpha = 1$, $\beta = 1$ e $\gamma = 0,3$ correspondente a métrica de $Q_{Areverb}$ original	19
3.4	Notas subjetivas da avaliação da reverberação para os 32 sinais da base de dados MARDY ('x' conectados) e notas Q_{MOS} para o <i>setup</i> $(1,0.5,0.3)$ ('•') usando o mapeamento de terceira ordem	20
3.5	Função de mapeamento PESQ para os coeficientes $c = -3,4986$ e $d = -2,5047$	21
3.6	Notas subjetivas da avaliação da reverberação para os 204 sinais da base de dados NBP ('x' conectados) e notas Q_{MOS} para o <i>setup</i> $(0.5,0,0.15)$ ('•') usando o mapeamento do tipo PESQ.	22
3.7	Notas subjetivas da avaliação da reverberação para os 32 sinais da base de dados MARDY ('x' conectados) e notas Q_{MOS} para o <i>setup</i> $(0.5,0,0.15)$ ('•') usando o mapeamento do tipo PESQ.	22
3.8	Função de mapeamento do tipo PESQ modificado para o caso de $\alpha = 1$, $\beta = 0$, $\gamma = 0.3$, e os coeficientes de mapeamento $c = -5.6503$, $d = -2.3826$, $e = 1.9358$ e $f = 2.5971$	23

3.9	Notas subjetivas da avaliação da reverberação para os 204 sinais da base de dados NBP ('x' conectados) e notas Q_{MOS} para o setup (1,0,0.3) ('•') usando o mapeamento do tipo PESQ modificado. . . .	23
3.10	Notas subjetivas da avaliação da reverberação para os 32 sinais da base de dados MARDY ('x' conectados) e notas Q_{MOS} para o setup (1,0,0.3) ('•') usando o mapeamento do tipo PESQ modificado. . . .	24
4.1	Diagrama de blocos do algoritmo de desreverberação de Wu-Wang. . .	27
4.2	Esquema da filtragem inversa para o algoritmo de Wu e Wang.	28
4.3	Diagrama de blocos do filtro de Wiener.	30
4.4	Representação em diagrama de blocos do problema de equalização do canal.	33
4.5	Convergência do sinal de saída do filtro adaptativo no problema de filtragem inversa.	33
4.6	Convergência do sinal de erro na saída do filtro adaptativo no problema de filtragem inversa para <i>a</i>) escala lineal, e <i>b</i>) escala logarítmica.	34
5.1	Curtose média do algoritmo de Wu e Wang no domínio do tempo para blocos do mesmo tamanho que o tamanho da janela, usando: (a) o algoritmo LMS; (b) o algoritmo SM com $\Gamma = 2.58$ e ponto de parada 10^{-3}	40
5.2	Curtose média do algoritmo original usando o passo $\mu = 3 \times 10^{-9}$, tamanho do filtro $M = 1024$, tamanho de bloco $N = 0.032x F_s$ e $N_i = 500$ iterações.	41

Lista de Tabelas

2.1	Correlação estatística de Q com as notas subjetivas da base de dados NBP para o caso de $\gamma=0$ (Allen) e $\gamma=0,3$ sem/com mapeamento não-linear.	12
2.2	Correlação estatística de Q com as notas subjetivas da base de dados MARDY para o caso de $\gamma=0$ (Allen) e $\gamma=0,3$ sem/com mapeamento não-linear.	12
3.1	Correlação estatística de Q_m com as notas subjetivas da NBP em função do passo utilizado na busca dos valores ótimos de α , β e γ . . .	15
3.2	Configurações com melhores correlações para ambas as bases NBP e MARDY.	19
5.1	Tabela comparativa das notas Q_{MOS} obtidas para o algoritmo de Wu e Wang no domínio do tempo e no domínio do tempo em blocos para um sinal de fala. Os resultados são comparados à saída de cada uma das etapas do algoritmo cego	38
5.2	Tabela comparativa das notas Q_{MOS} obtidas usando o algoritmo de Wu e Wang original e as notas obtidas com as configurações ótimas dos mapeamentos PESQ e PESQ modificado, estudados no capítulo 3, implementados para o algoritmo de duas etapas no domínio do tempo. Os resultados são comparados na saída de cada uma das etapas do algoritmo cego, para o passo ótimo de $\mu = 10$	39
5.3	Avaliação do algoritmo de Wu e Wang, nas versões original e modificada no domínio da frequência com algoritmo SM.	42

5.4	Avaliação do algoritmo de Wu e Wang, nas versões original e modificada no domínio da frequência (comprimento uniformizado dos blocos) com algoritmo SM.	42
-----	---	----

Capítulo 1

Introdução

1.1 Apresentação Inicial

O tema do trabalho é o efeito de reverberação de sinais de voz, tanto o aspecto da medição da sua percepção quanto o de desenvolvimento de novos algoritmos de desreverberação.

O projeto pretende analisar alguns critérios objetivos existentes na literatura, que são utilizados na quantificação da percepção da reverberação em sinais de voz, aperfeiçoando-os. A partir do resultado desse estudo, pretende-se estudar novos algoritmos de desreverberação usando este tipo de critério como medida de desempenho do algoritmo, de modo a otimizar o seu desempenho num domínio perceptual do efeito em si.

Novos métodos de quantificação do efeito da reverberação, no contexto da sua percepção, permitem o desenvolvimento de novos algoritmos (e aperfeiçoamento de algoritmos anteriores) de desreverberação. Neste sentido, os novos algoritmos responderiam a um critério perceptual envolvendo diferentes aspectos do efeito de reverberação. Mais especificamente, veremos que o efeito em si pode ser caracterizado por aspectos como o próprio tempo de reverberação (RT, do inglês *reverberation time*), a variância espectral da sala (RSV, do inglês *room spectral variance*) e a razão das energias direta e reverberante (DRER, do inglês *direct-to-reverberant energy ratio*) [?]. Estas medidas todas podem ser combinadas numa única medida de qualidade perceptual, QAreverb, a qual pode ser usada como medida de desempenho (perceptual) de um algoritmo de desreverberação [?].

Em suma, o objetivo central deste trabalho é o desenvolvimento e aperfeiçoamento de algoritmos de desreverberação baseados em medidas quantitativas da percepção da reverberação em sinais de voz.

Para atingirmos esta meta, o projeto se inicia com um estudo das técnicas atuais de avaliação de qualidade de sinais de voz sob o efeito de reverberação. Em particular, consideramos três passos principais a serem seguidos a partir da técnica QAreverb desenvolvida na referências [1][5]:

(i) Novas formas de combinar as medidas de RT, RSV e DRER para compor a nota final QAreverb serão investigadas.

(ii) Além disto, novos métodos de mapeamento entre a medida bruta e a medida ajustada para a escala MOS serão considerados, procurando aumentar a correlação da medida final com a avaliação subjetiva dos sinais de voz com reverberação. Com estes estudos, pretende-se desenvolver uma nova versão da medida QAreverb com melhor desempenho de estimação da qualidade perceptual.

(iii) A partir desta nova métrica (ou mesmo da anterior, para efeito de comparação), estudaremos o desempenho de algum(ns) algoritmo(s) de desreverberação no contexto perceptual.

(iv) A partir dos resultados do item anterior, procuraremos aperfeiçoar os algoritmos de desreverberação, ou mesmo desenvolver novos algoritmos, usando a métrica perceptual como critério de desempenho.

1.2 Organização do Texto

O capítulo 2 inclui uma introdução ao conceito de reverberação e aos componentes necessários para lhe caracterizar e avaliar. Nesse capítulo são apresentadas a métrica de avaliação QAreverb [?] assim como a sua modificação proposta. Também são apresentadas as bases de dados usadas para a avaliação desta métrica.

O capítulo 3 apresenta o análise da nova métrica e mostra os resultados obtidos. Novos mapeamentos são utilizados para o incremento da correlação entre a nota objetiva e subjetiva.

O capítulo 4 inclui uma apresentação teórica do conceito de desreverberação, do algoritmo de Wu e Wang de duas etapas, e dos principais métodos adaptativos

que são usados no capítulo 5 para melhorar o desempenho deste algoritmo.

No capítulo 5 apresenta propostas as modificações do algoritmo de Wu e Wang para melhorar o seu desempenho. Resultados experimentais são incluídos aqui validando os resultados por nós obtidos.

O capítulo 6 é o capítulo de conclusão, no qual serão resumidas as principais contribuições deste projeto.

Capítulo 2

Avaliação Objetiva da Reverberação

2.1 Introdução

O capítulo que segue apresenta os conceitos teóricos relativos a uma nova métrica de avaliação do efeito de reverberação baseada no algoritmo de Allen que foi posteriormente melhorado, conforme descrito em [1], aumentando o resultado de correlação entre a nota subjetiva proporcionada por ouvintes e a nota objetiva obtida com a métrica.

O capítulo começa com a apresentação do conceito de reverberação, assim como dos principais parâmetros associados a sua medição usados pela métrica proposta (seção 2.2).

Também são apresentados neste capítulo as duas bases de dados usadas para a avaliação da métrica, a base de dados MARDY, e a base de dados NBP (seção 2.3).

Fecha este capítulo a apresentação dos algoritmos de avaliação anteriores a esta nova métrica, o sistema QAreverb original e por último a modificação da métrica QAreverb (2.4).

2.2 Conceito de Reverberação

A reverberação é um efeito acústico que ocorre quando várias cópias de um sinal de som, com diferentes retardos e níveis de intensidade, são percebidas em conjunto. Estas cópias normalmente são devidas às reflexões do sinal num cenário de propagação fechado, conforme ilustrado na Figura 2.1. A reverberação, em geral, ajuda a situar o ouvinte num ambiente acústico. No entanto, uma grande quantidade de reverberação pode atrapalhar a inteligibilidade do sinal de voz.

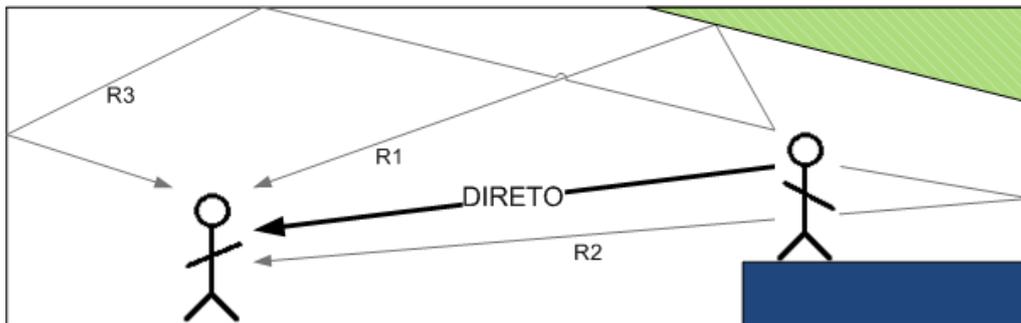


Figura 2.1: Representação esquemática do efeito de reverberação.

Diferentemente do que ocorre quando um sinal tem eco (onde o atraso é grande em relação à percepção de nosso ouvido), a reverberação ocorre quando a diferença dos instantes de recebimento dos sons é inferior a 0,1 s. Como nosso ouvido não percebe dois sons produzidos em tempo inferior a 1/17 de segundo entre eles (alguns autores adotam 1/15), se essas reflexões ocorrem num tempo inferior a isso, não se percebe um novo som mas sim uma continuação do som inicial que vai perdendo “força”.

O efeito de reverberação é normalmente modelado como a convolução de um sinal de áudio $s(n)$ com a resposta impulsiva da sala (*room impulse response*, RIR), $h(t)$, gerando o som com reverberação

$$s_r(n) = \sum_{k=0}^{N-1} h(k)s(n-k), \quad (2.1)$$

onde N é o comprimento de $h(t)$. Na prática, há que considerar duas porções diferentes da RIR: As reflexões iniciais e as reflexões tardias (ver Figura 2.2).

Reflexões iniciais: Esta parcela da RIR é composta por vários impulsos com amplitudes que seguem um decaimento exponencial e contém a maior parte da

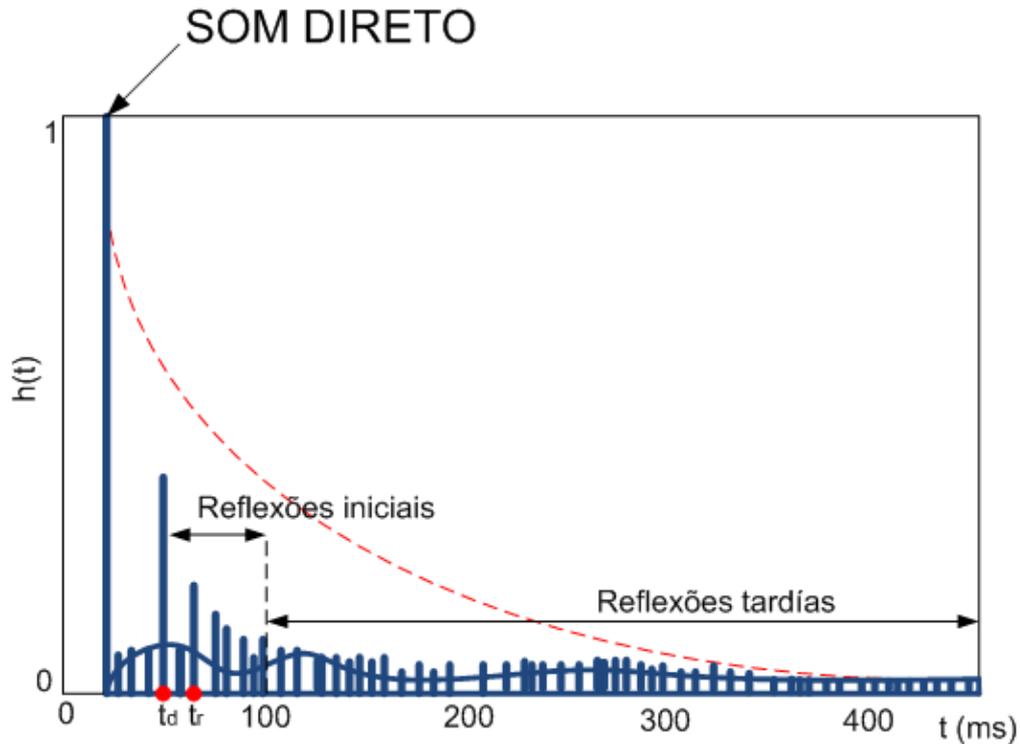


Figura 2.2: Exemplo de RIR associada ao efeito de reverberação.

energia da RIR. O primeiro impulso da figura refere-se ao componente do som direto, definindo o instante temporal t_d . O segundo impulso é a primeira reflexão do som, no instante temporal correspondente à t_r . Esta etapa da reverberação está associada ao efeito também chamado de “coloração”.

Reflexões tardias: Esta parcela inclui os demais componentes da RIR que apresentam uma natureza difusa sem componentes de impulsos isolados significativos.

Há muitas medidas associadas ao efeito de reverberação, sendo que três delas parecem ser as mais importantes para a percepção e, por este motivo, foram usadas pelo sistema QAreverb, que será apresentado na Seção 2.4.2. Estas três medidas são: o tempo de reverberação (*reverberation time*, RT), a variância espectral da sala (*room spectral variance*, RSV) e a razão das energias direta e reverberante (*direct-to-reverberant energy ratio*, DRER). Estas medidas são detalhadas a seguir.

2.2.1 Tempo de reverberação

O tempo de reverberação é uma medida que quantifica o tempo que a energia acústica emitida pelas fontes do som permanece numa sala. O valor do tempo de reverberação varia conforme muitas variáveis de acordo com a geometria da sala e os materiais que compõem os revestimentos do teto e das paredes. Uma estimativa confiável do RT pode ser usada para avaliar as características acústicas de uma sala ou para conceber um esquema de desreverberação próprio para um sistema de áudio particular.

Mais especificamente, o tempo de reverberação T_{60} é definido como o período do tempo requerido pela pressão sonora para decair 60 dB após o sinal de excitação ser desativado [1]. Em locais abertos ou com pouca reverberação o tempo de reverberação é muito próximo de zero; já em locais como catedrais o tempo pode chegar a mais de 20 segundos. Em geral uma reverberação ambiente dura entre 0,2 e 0,3 segundo.

A estimativa do T_{60} utilizando excitação de ruído requer um conjunto de medições, mas Schroeder mostrou que esta estimativa pode ser computada diretamente de uma única RIR medida. Posteriormente, verificou-se que o T_{60} de uma RIR, $h(t)$, dada pode ser estimado com base na curva normalizada de energia de decaimento (EDC) definida como:

$$EDC(t) = 10 \log_{10} \left(\frac{\int_t^{\infty} h^2(\tau) d\tau}{\int_0^{\infty} h^2(\tau) d\tau} \right) \text{ [dB]}, \quad (2.2)$$

onde o denominador garante um valor máximo de EDC de 0 dB quando $t = 0$. Na escala dB, a EDC pode ser aproximada por uma função de primeira ordem e o T_{60} estimado pode ser definido como o intervalo do tempo para que a aproximação de primeira ordem da EDC caia de 0 a -60 dB. Para se obter esta estimativa de primeira ordem, considera-se o intervalo em que a EDC varia de -5 dB até um ponto no qual se considera o sinal significativamente afetado pelo ruído.

2.2.2 Variância espectral da sala

A variância espectral da sala (*room spectral variance*, RSV) é definida como

$$RSV = E_f \{ (I(f) - E_f \{ I(f) \})^2 \}, \quad (2.3)$$

onde $E_f\{\cdot\}$ é o operador média no domínio da frequência e $I(f)$ é o nível de intensidade acústica relativa, definido como

$$I(f) = 10 \log_{10} \left[\frac{H^2(f)}{E_f \{H^2(f)\}} \right] [\text{dB}], \quad (2.4)$$

sendo $H(f)$ a transformada de Fourier de $h(t)$.

2.2.3 Razão das energias direta e reverberante

Esta característica é definida como a relação entre o nível de energia direta E_d (dentro de um intervalo pequeno ao redor de t_d) e o nível de energia reverberante E_r de $h(t)$, de modo que

$$R = \frac{E_d}{E_r} = \frac{\int_{(t_d-1)ms}^{(t_d+1.5)ms} h^2(t) dt}{\int_{(t_d+1.5)ms}^{\infty} h^2(t) dt}. \quad (2.5)$$

Para reduzir a influência do ruído no cômputo de R , é desejável considerar apenas os componentes do sinal 20 dB acima do nível do ruído em $h(t)$ e parar a acumulação de energia no mesmo ponto de parada usado pelo algoritmo de estimativa do T_{60} .

A relação das energias direta e reverberante proporciona ao ouvinte alguma sensação de localização da fonte. Como veremos mais adiante, a grande novidade do sistema QAreverb foi incluir a relação das energias direta a reverberante na quantificação do efeito de reverberação. Esta medida foi incorporada para contemplar situações em que a distância fonte-microfone é menor que uma dada distância crítica, nas quais o RSV não varia de forma significativa.

2.3 Bases de Dados de Sinais Reverberantes

No desenvolvimento do sistema QAreverb foram utilizadas duas bases de dados diferentes: a base MARDY e a base *new Brazilian-Portuguese (NBP)*, descritas a seguir.

2.3.1 Base de dados MARDY

A base de dados MARDY inclui 16 sinais com reverberação naturalmente degradados, gravados diretamente num auditório, assim como as suas respectivas

versões desreverberadas usando um algoritmo de atraso-e-soma, totalizando 32 sinais de som com frequência de amostragem $F_s = 16$ kHz.

A base de dados MARDY considera dois locutores diferentes (1 homem e 1 mulher), 4 valores de distância fonte-microfone ($d = 1, 2, 3$ e 4 m), e dois tipos de painéis de parede (refletivo e absorvivo), que correspondem a um T_{60} estimado de 447 ms e 291 ms, respectivamente.

A base de dados MARDY foi a primeira base de dados criada para avaliar a reverberação dos sinais do som, contendo um número relativamente reduzido de sinais, todos gravados numa mesma sala. Para uma melhor avaliação do efeito, foi preciso o desenvolvimento de uma base de dados maior e mais geral, conforme descrito no próximo item.

2.3.2 Nova base de voz reverberante

A base de dados NBP foi baseada em 4 sinais anecoicos pronunciados por dois falantes (1 homem e 1 mulher) usando uma frequência de amostragem $F_s = 48$ kHz. Cada sinal é composto de duas frases em português brasileiro, separadas em média por aproximadamente 1,7 s, sendo a duração média de 8,4 s para todas as frases da base de dados. O efeito de reverberação foi imposto nos sinais anecoicos usando três enfoques diferentes:

- Reverberação artificial: O efeito de reverberação é emulado por 6 RIRs geradas artificialmente, produzindo um total de 24 sinais. Nestas RIRs, as reflexões iniciais foram modeladas pelo método das imagens, com uma distância fonte-microfone fixa $d = 1,8$ m numa sala virtual de dimensões (Comprimento x Largura x Altura) = $4 \text{ m} \times 3 \text{ m} \times 3 \text{ m}$. As RIRs artificiais foram criadas com tempos de reverberação de 200, 300 e 400 ms. Com o método modificado de Gardner, foram emulados os tempos de reverberação maiores, com valores de 500, 600 e 700 ms.
- Reverberação natural: Neste contexto, foram gerados 68 sinais reverberados a partir de 17 RIRs diferentes e dos 4 sinais anecoicos. As RIRs foram obtidas de 4 salas diferentes e diferentes distancias fonte-microfone d .
- Reverberação real: Nesta parte, foram gerados 108 sinais com reverberação a

partir da reprodução e gravação dos 4 sinais anecoicos em 7 salas com diferentes características de reverberação. Em cada sala foram consideradas 5 diferentes distâncias fonte-microfone d , a menos da sala menor, na que só foram utilizadas 3 distâncias.

Desta maneira, obtivemos o número total de 204 sinais (24 artificiais, 68 naturais, 108 reais e os 4 anecoicos). Testes subjetivos de avaliação da qualidade dos sinais com reverberação foram feitos usando a escala MOS 1–5 (*mean objective score*), usando 30 ouvintes por sinal.

2.4 Avaliação Objetiva de Sinais Reverberantes

2.4.1 Algoritmos anteriores

Existem muitas métricas para avaliar o efeito de reverberação, sendo que a métrica de interesse neste projeto é o algoritmo de Allen, pois é o algoritmo em que se baseia a métrica QAreverb.

Berkley e Allen [?] identificaram duas variáveis físicas para a qualidade de um som reverberado: O tempo de reverberação e a distância fonte-ouvinte.

Como foi explicado na seção 2.2, a resposta impulsiva da sala é composta de 3 partes: O sinal original, as reflexões iniciais, e as reflexões tardias. As reflexões iniciais distorcem o espectro do sinal do som, caracterizando uma resposta em frequência não plana. Este efeito também se conhece como coloração.

Considerando que a coloração pode ser caracterizada pela variância da resposta em frequência da sala, Allen (Allen, 1982), propôs uma medida P da reverberação percebida combinando o tempo de reverberação e a RSV da forma

$$P = P_0 - T_{60}.RSV, \quad (2.6)$$

onde P_0 define o valor máximo da escala de P .

Outras tentativas de quantificar a qualidade de um sinal de voz incluem: ITU W-PESQ (7 kHz), ITU P.563 (4 kHz), *Reverberation Decay Curve* R_{DT} (4kHz) e *Speech-to-Reverberant Modulation energy Ratio* SRMR (4 ou 8 kHz), sendo estas últimas duas desenvolvidas especificamente para avaliar o efeito de reverberação.

2.4.2 QAreverb

Apesar de ignorada no trabalho pioneiro de Allen, a razão E_d/E_r das energias direta e reverberante proporciona uma chave fundamental para avaliar a inteligibilidade do som num espaço fechado, pois entrega ao ouvinte alguma sensação da distância e localização da fonte. Experimentos práticos também indicaram a importância da razão das energias direta e reverberante na percepção subjetiva da reverberação. De fato, ainda que a RSV seja fortemente relacionada a $DREER$ quando $R \geq 1$, as duas medidas são dissociadas quando a distância da fonte ao microfone é suficientemente grande, o que corresponde a $R < 1$.

Portanto, uma nova medida Q para avaliar a qualidade da reverberação foi criada, combinando as características T_{60} , RSV e incorporando a relação de energias $DREER$ à medida de Allen, de forma que:

$$Q = -\frac{T_{60} \cdot RSV}{(E_d/E_r)^\gamma}, \quad (2.7)$$

com o expoente γ determinado empiricamente numa etapa de treinamento do sistema, onde o caso particular de $\gamma = 0$ corresponde à medida original de Allen.

Usando a base de dados NBP, a máxima correlação obtida entre a nota subjetiva (MOS) e a nota objetiva Q do sistema QAreverb foi de $\rho = 91\%$ para $\gamma=0,3$, utilizando o mapeamento de terceira ordem descrito na seção (3.2.2.1), aumentando assim o resultado de $\rho = 85\%$ obtido pela métrica de Allen, como detalhado na tabela 2.1.

A base de dados MARDY foi utilizada na fase experimental do sistema QAreverb para validar os resultados obtidos com a base de dados NBP, pois a base MARDY contém sinais que não foram utilizadas no desenvolvimento do sistema. Na tabela (2.2) se pode observar como para os valores de $\gamma=0$ e $\gamma=0.3$, utilizando o mapeamento de terceira ordem com os mesmos valores de $x_1=0.0017$, $x_2=0.0598$, $x_3=0.7014$, e $x_4=4.5387$ ajustados para a base de dados NBP, também se obtêm valores altos de correlação.

2.4.3 QAreverb modificado

Com o objetivo de determinar a importância de cada medida (T_{60} , RSV e E_d/Er) e aumentar ainda mais a correlação entre a nota objetiva e a nota subjetiva

Tabela 2.1: Correlação estatística de Q com as notas subjetivas da base de dados NBP para o caso de $\gamma=0$ (Allen) e $\gamma=0,3$ sem/com mapeamento não-linear.

Algoritmo	Melhor correlação (ρ)[%]
Q($\gamma=0$) Allen	85/85
Q($\gamma=0.3$) Sistema QAreverb original	81/91

Tabela 2.2: Correlação estatística de Q com as notas subjetivas da base de dados MARDY para o caso de $\gamma=0$ (Allen) e $\gamma=0,3$ sem/com mapeamento não-linear.

Algoritmo	Melhor correlação (ρ)[%]
Q($\gamma=0$) Allen	91/92
Q($\gamma=0.3$) Sistema QAreverb original	95/95

MOS, consideramos uma nova medida de qualidade Q_m , incluindo expoentes α , β e γ a cada uma das medidas, de modo que

$$Q_m = -\frac{T_{60}^\alpha \cdot RSV^\beta}{(E_d/E_r)^\gamma}. \quad (2.8)$$

Um programa em Matlab foi criado para se buscar os melhores valores de α , β e γ , sempre no intervalo $[0,1]$, de modo a maximizar a correlação estatística de Q_m com os valores de MOS obtidos para as bases NBP.

Numa fase posterior de análise, avaliamos estes resultados para a base de dados MARDY, da mesma forma que foi feito para o sistema QAreverb original.

O estudo efetuado com esta métrica modificada, assim como os resultados do análise são descritos no capítulo 3.

Capítulo 3

Avaliação da Nova Métrica QAreverb

3.1 Introdução

No capítulo 2 foram apresentados os conceitos teóricos associados à nova métrica de avaliação da reverberação proposta (Q_m), a partir de uma modificação da métrica do sistema QAreverb.

Neste capítulo são mostrados os resultados desta nova métrica para diferentes casos, o primeiro deles usando a métrica diretamente (seção 3.2.1), seguido de um estudo do desempenho da métrica usando diferentes tipos de mapeamento para a escala MOS (seções 3.2.2, 3.2.3 e 3.2.4).

Por último, fecha este capítulo a seção de conclusão (3.3) onde são resumidos os principais resultados e as conclusões obtidas no estudo efetuado.

3.2 Projeto da QAreverb Modificada

A modificação proposta para o sistema QAreverb consistiu na determinação do conjunto dos valores de α , β e γ que maximizam a correlação de Q_m na equação (2.8) com os valores de MOS para os sinais da base NBP. O procedimento de busca dos valores ótimos foi realizado em Matlab.

Num primeiro procedimento fizemos a análise da correlação diretamente, sem nenhum tipo de mapeamento; posteriormente, testamos diferentes tipos de mapea-

mento para aumentar ainda mais a correlação. O procedimento foi o mesmo para cada tipo de experimento, e foi composto pelas seguintes fases:

- Geração dos coeficientes α , β e γ a partir das definições dos respectivos intervalos e passos de busca.
- Obtenção dos melhores resultados de correlação de Q_m com os valores de MOS da NBP, usando os valores de α , β e γ definidos no passo anterior.
- Refinamento da procura: Para os melhores valores de α , β e γ obtidos no passo anterior, reduzir o intervalo e o passo de busca e realizar nova procura de valores ótimos.
- Teste de diferentes tipos de mapeamento para os valores de α , β e γ obtidos após a etapa de refinamento. No caso especial do mapeamento de terceira ordem, foi feita a comprovação de que a função polinomial resultante era sempre crescente.
- Avaliação destes resultados com a base de dados MARDY.

Os itens a seguir mostram os resultados obtidos para cada tipo de experimento.

3.2.1 Otimização de Q_m

Num primeiro passo, estabelecemos os valores de α , β e γ no intervalo $[0,1]$ usando um passo inicial de 0,1 para cada um destes parâmetros. Após se encontrar o conjunto α , β e γ ótimo nesta primeira busca refinamos o passo e fizemos nova busca em torno deste conjunto ótimo.

A Tabela 3.1 mostra os resultados obtidos com os diferentes passos, referentes ao intervalo $[0,1]$ completo.

Os resultados desta tabela indicam uma correlação máxima de **90,12%** obtida com $\alpha = 0,62-0,66$, $\beta = 0$ e $\gamma = 0,17$. Para este experimento, o refinamento com passo de 0,01 já foi suficiente para gerar a configuração ótima.

Comparando este resultado com o resultado do sistema QAreverb sem mapeamento, ($\rho = 81\%$), a melhora deste primeiro experimento é de aproximadamente

Tabela 3.1: Correlação estatística de Q_m com as notas subjetivas da NBP em função do passo utilizado na busca dos valores ótimos de α , β e γ .

Passo	Melhor correlação	Setup ótimo
0,1	89,89%	$\alpha = 0,7, \beta = 0, \gamma = 0.2$
0,01	90,12%	$\alpha = 0,62 - 0,66, \beta = 0, \gamma = 0,17$
0,001	90,12%	$\alpha = 0,62 - 0,66, \beta = 0, \gamma = 0,17$

10 %, uma grande melhora pois sem nenhum tipo de mapeamento o resultado da correlação já alcançou o melhor resultado do sistema original, obtido com o mapeamento de terceira ordem.

O resultado de $\beta = 0$, que elimina a influência da componente RSV ao compor a nota Q , talvez possa ser explicado pela uso da medida E_d/E_r no sistema modificado, que parece ser redundante com a medida RSV , tornado esta última desnecessária.

A avaliação do sistema utilizando a base de dados MARDY também foi satisfatória, pois para os mesmos valores de α , β e γ obtidos na análise anterior (tabela 3.1), obtivemos um valor de correlação de $\rho = 95\%$, a mesma correlação obtida pelo sistema QAreverb original para a base de dados MARDY (Tabela 2.2).

Este primeiro experimento foi muito satisfatório já que nos proporcionou altas expectativas de poder aumentar a correlação máxima de 91% obtida pelo sistema QAreverb para o mapeamento de terceira ordem (tabela 2.1), pois, como era de se esperar, a correlação para a nova métrica de QAreverb aumentaria mapeando os resultados para a escala MOS.

3.2.2 Mapeamento não-linear

Com o objetivo de melhorar a correlação ainda mais, foram estudados diversos mapeamentos não-lineares do valor de Q_m para a escala MOS de 1 a 5.

3.2.2.1 Mapeamento polinomial de terceira ordem

O mapeamento polinomial de terceira ordem foi usado no sistema QAreverb original para mapear os valores de Q para a escala MOS de 1 a 5. No sistema original, este mapeamento produz um valor de correlação maior do que o obtido

sem mapeamento.

O mapeamento polinomial de terceira ordem é caracterizado por

$$Q_{map} = x_1 Q_m^3 + x_2 Q_m^2 + x_3 Q_m + x_4, \quad (3.1)$$

com os valores de x_1 , x_2 , x_3 e x_4 determinados de forma a maximizar a correlação de Q_{map} com as avaliações subjetivas da base NBP.

Na prática, os ouvintes podem dar diferentes notas quando são considerados diferentes níveis de reverberação no teste subjetivo. Desta forma, o mapeamento não-linear pode ser acompanhado de um ajuste linear da forma

$$Q_{MOS} = a Q_{map} + b, \quad (3.2)$$

onde a e b são determinados empiricamente. Este segundo mapeamento reduz o erro quadrático médio (MSE) entre a nota objetiva e a nota subjetiva, sem mudar o fator de correlação associado.

3.2.2.2 Projeto de QAreverb com mapeamento de terceira ordem

Inicialmente, para fazer este mapeamento no Matlab, consideramos todos os possíveis valores de α , β e γ e dos coeficientes x_1 , x_2 , x_3 e x_4 . Infelizmente, porém, esta busca se mostrou extremamente dispendiosa em termos computacionais, particularmente quando consideramos pequenos passos para cada variável. Numa segunda proposta, consideramos a busca dos valores ótimos de x_1 , x_2 , x_3 e x_4 apenas para os valores de α , β e γ obtidos no processo de otimização anterior, sem mapeamento.

O resultado desta busca foi uma correlação de 92,8% obtida para $\alpha \approx 0,6$, $\beta = 0$, $\gamma = 0,2$ e $x_1 = -3.7220$, $x_2 = -6.8956$, $x_3 = -0.7568$ e $x_4 = 4.3850$. A fim de minimizar o erro quadrático médio entre a nota objetiva e subjetiva, fizemos o mapeamento linear do tipo $Q_{MOS} = a Q_m + b$ com os coeficientes $a = 1.0000$ e $b = -5.9636 \times 10^{-10}$. Estes valores indicam que o ajuste de escala foi praticamente desprezível.

Infelizmente, como comprovamos numa etapa posterior (seção ??), o mapeamento não-linear encontrado possuía mínimos locais no intervalo de interesse 1–5 da escala MOS, o que o tornava não monotonicamente crescente, conforme ilustrado na Figura 3.1. Isto nos obrigou a descartar este resultado inicial.

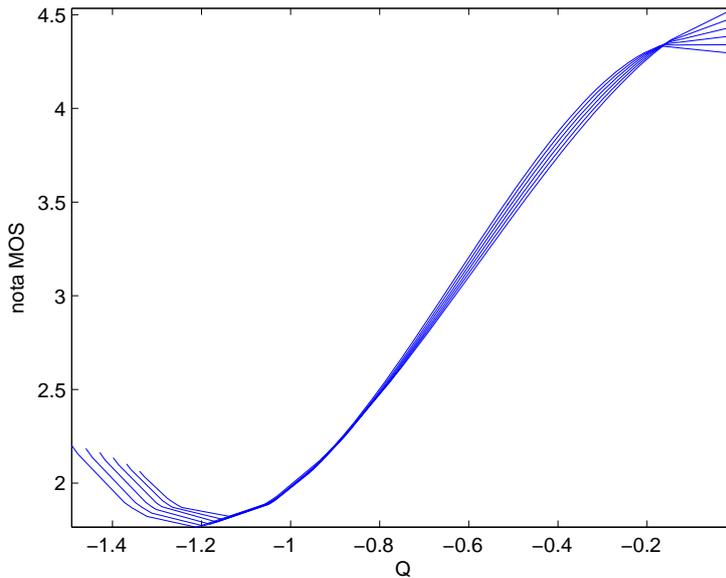


Figura 3.1: Exemplos de mapeamentos não-lineares com mínimos locais no intervalo $[1,5]$, correspondendo a funções não monotonicamente crescentes.

Para resolver o problema de mapeamentos não monotonicamente crescentes, buscamos outros valores de α , β e γ que proporcionaram uma boa correlação. O melhor resultado para estes novos *setups* foi $\alpha = 1$, $\beta = 0$, $\gamma = 0.3-0.35$, $x_1 = -0.1489$, $x_2 = 1.1089$, $x_3 = 3.8250$ e $x_4 = 4.5559$, caracterizando uma correlação de **92,7%**. Para este *setup*, obtivemos um mapeamento linear com $a = 1.0000$ e $b = 4.8161 \times 10^{-9}$, o que indica, novamente, um ajuste mínimo para a escala MOS.

Nesta nova configuração, o valor de α muda de 0,6 para 1, dando um maior peso ao componente T_{60} na composição da nota QAreverb modificada. Comparando a correlação aqui obtida com a do sistema QAreverb original, como dado na tabela 2.1, vemos que a métrica modificada resulta num acréscimo de quase o 2%, eliminando o efeito da *RSV*.

Os resultados do sistema QAreverb modificado com mapeamento de terceira ordem para a base NBP são mostrados na figura 3.2.

A avaliação do resultado para a base MARDY, no entanto, obteve um valor de correlação $\rho = 91\%$ menor do que o correspondente ao sistema QAreverb original (95%). Porém, devemos ressaltar que o sistema QAreverb original usava um mapeamento não-linear não monotonicamente crescente, conforme visto na figura 3.3, tornando um pouco injusta esta comparação direta entre os dois sistemas.

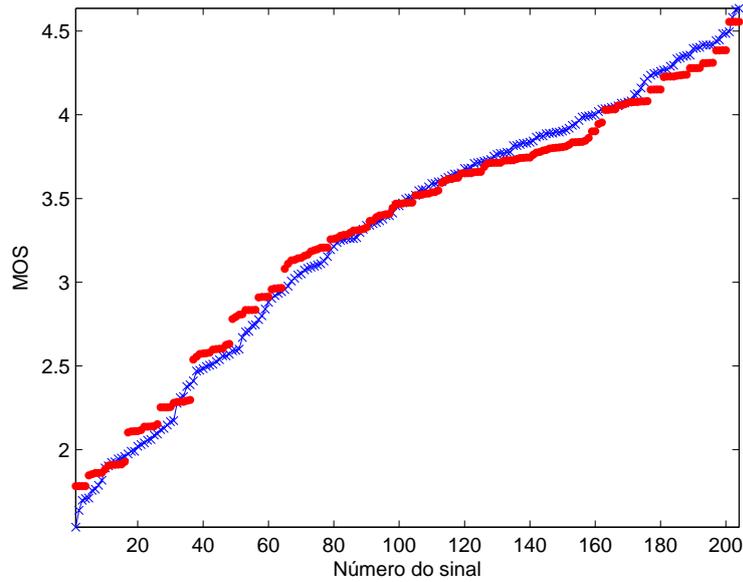


Figura 3.2: Notas subjetivas da avaliação da reverberação para os 204 sinais da base de dados NBP ('x' conectados) e notas Q_{MOS} modificadas para a configuração $\alpha = 1$, $\beta = 0$ e $\gamma = 0,35$ ('•'), com mapeamento de terceira ordem.

3.2.2.3 Busca de resultados sub-ótimos da base NBP para melhorar a correlação da base MARDY com a métrica Q_m

Prosseguindo com o experimento da nova métrica de $Q_{Areverb}$, o passo seguinte foi procurar diferentes configurações de α , β e γ que proporcionaram um bom resultado de correlação para ambas as bases NBP e MARDY.

Na tabela 3.2 são mostrados estes resultados sub-ótimos com os seus respectivos *setups* e coeficientes do mapeamento de terceira ordem x_1, x_2, x_3, x_4 . Na tabela são mostrados apenas os resultados em que as correlações para ambas as bases eram superiores às obtidas pelo sistema original.

Os resultados do sistema $Q_{Areverb}$ com mapeamento de terceira ordem para a base MARDY são mostrados na figura (3.4).

Numa etapa posterior, ignoramos os mapeamentos de terceira ordem que possuísem derivada nula no intervalo de 1 a 5, de modo a obter uma função monotonicamente crescente neste intervalo. Infelizmente, nesta etapa de avaliação tivemos que desconsiderar todos os resultados obtidos anteriormente, pois todos eles tinham raiz da derivada do polinômio no intervalo 1–5. Além disso, esta nova busca

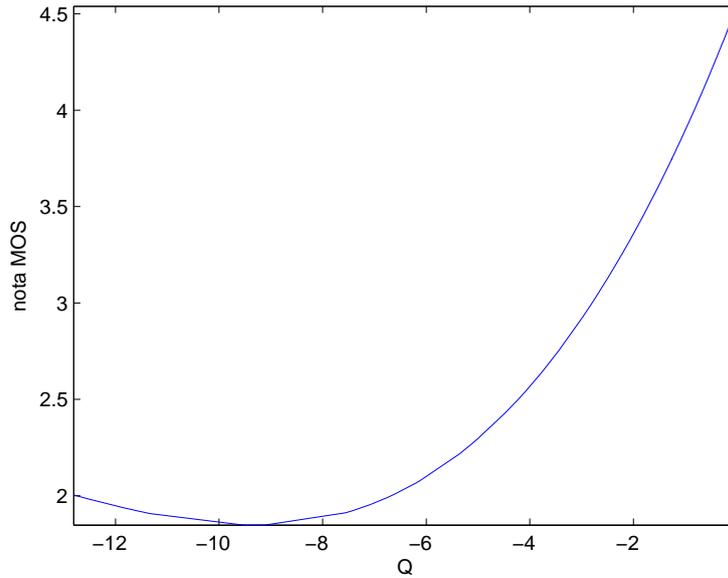


Figura 3.3: Polinômio de terceira ordem com mínimo local no intervalo $[1,5]$, para o caso de $\alpha = 1$, $\beta = 1$ e $\gamma = 0,3$ correspondente a métrica de QAreverb original

Tabela 3.2: Configurações com melhores correlações para ambas as bases NBP e MARDY.

<i>setup</i>	$\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$	$\rho[\%]$ NBP	$\rho[\%]$ MARDY
$\alpha = 1, \beta = 0.2, \gamma = 0.3$	(-0.1791,0.0991,2.2927,4.5197)	92.5	95.32
$\alpha = 0.9, \beta = 0.2, \gamma = 0.3$	(-0.1803,0.0806,2.2931,4.5808)	92.5	95.31
$\alpha = 1, \beta = 0.3, \gamma = 0.3$	(-0.0698,0.1974,2.0303,4.5324)	92.4	95.3
$\alpha = 1, \beta = 0.4, \gamma = 0.3$	(-0.0232,0.2167,1.7802,4.5420)	92.3	95.28
$\alpha = 0.9, \beta = 0.3, \gamma = 0.3$	(-0.0722,0.1786,2.0248,4.5920)	92.3	95.27
$\alpha = 1, \beta = 0.5, \gamma = 0.3$	(-0.0048,0.1990,1.5476,4.5484)	92.2	95.25

proporcionou somente dois *setups* válidos, ambos com correlações inferiores a 82%.

Assim, para melhorar o desempenho do sistema QAreverb modificado, vamos considerar novos tipos de mapeamentos.

3.2.3 Mapeamento do tipo PESQ

Nesta seção, consideramos o mapeamento usado na métrica PESQ que tem a seguinte forma:

$$Q_{map} = 1 + \frac{4}{1 + \exp[(c \cdot Q_m) + d]}, \quad (3.3)$$

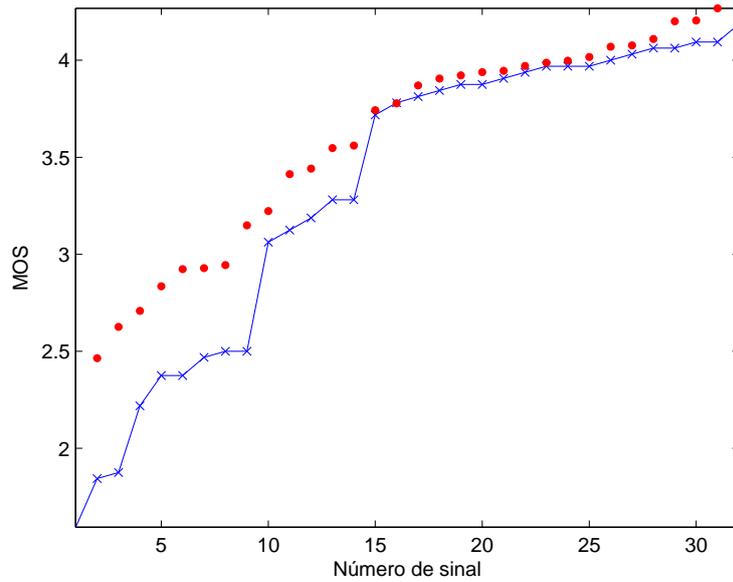


Figura 3.4: Notas subjetivas da avaliação da reverberação para os 32 sinais da base de dados MARDY ('x' conectados) e notas Q_{MOS} para o *setup* (1,0.5,0.3) ('•') usando o mapeamento de terceira ordem

com c e d determinados empiricamente de modo a maximizar a correlação da nova métrica mapeada com as notas subjetivas da base NBP.

Usando este mapeamento, obtivemos uma correlação máxima de $\rho = 91,7\%$, para $\alpha = 0.4 - 0.5$, $\beta = 0$, $\gamma = 0.13 - 0.16$, $c = -3.4410$ e $d = -2.4450$. Esta configuração se assemelha à obtida na otimização de Q_m com mapeamento polinomial, novamente com $\beta = 0$ em que o efeito da *RSV* é ignorado. O mapeamento resultante é mostrado na figura 3.5 e as notas objetivas obtidas para a base NBP são ilustradas na figura 3.6.

Avaliando o resultado para a base MARDY, conforme visto na figura 3.7, obtivemos uma correlação de $\rho = 95\%$ para a configuração descrita acima. Isto indica um resultado satisfatório para a base NBP e equivalente ao obtido anteriormente para a base MARDY.

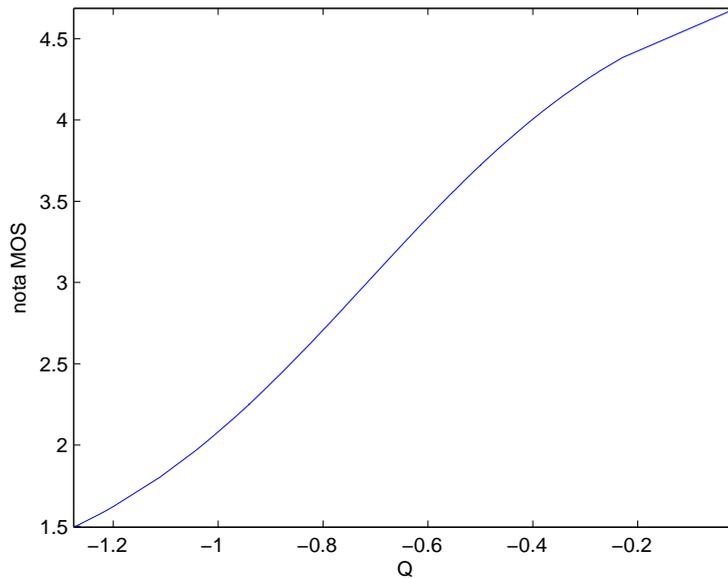


Figura 3.5: Função de mapeamento PESQ para os coeficientes $c = -3,4986$ e $d = -2,5047$.

3.2.4 Mapeamento PESQ modificado

Por último, outro mapeamento foi considerado, generalizando o mapeamento do tipo PESQ para a forma

$$Q_{map} = e + \frac{f}{1 + \exp[(c \cdot Q) + d]}, \quad (3.4)$$

com c , d , e e f obtidos empiricamente.

Neste caso, a correlação máxima foi de $\rho = 93\%$ obtida para $\alpha = 1$, $\beta = 0 - 0,3$, $\gamma = 0,3$, $c = -3.1213$, $d = -2.6964$, $e = 1.9483$ e $f = 2.6964$. Com esta modificação do mapeamento PESQ, ilustrado na figura 3.8, obtivemos a maior correlação de todos os mapeamentos feitos e uma configuração muito similar àquela que gerava a maior correlação com mapeamento polinomial.

Para a base MARDY, esta configuração gerou uma correlação de $\rho = 95,5$, melhorando ligeiramente o resultado da métrica original para as duas bases de dados.

Os resultados do sistema QAreverb com mapeamento PESQ modificado para a base NBP e a base MARDY são mostrados nas figuras 3.9 e 3.10, respectivamente.

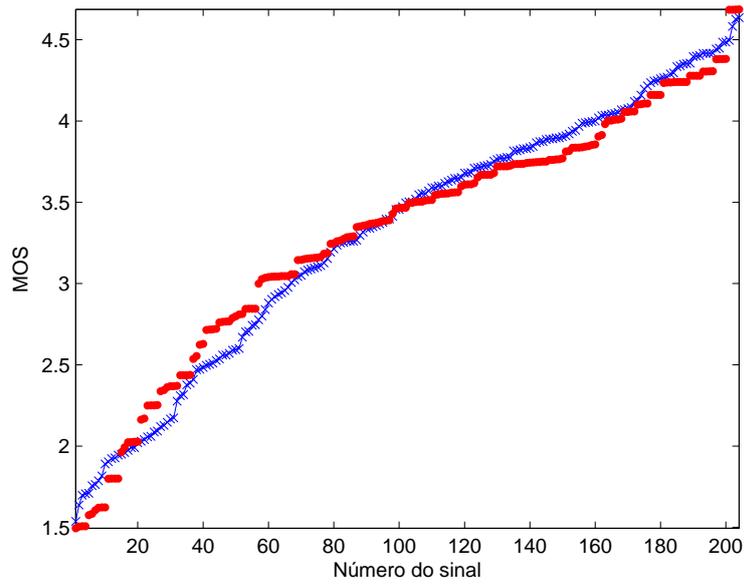


Figura 3.6: Notas subjetivas da avaliação da reverberação para os 204 sinais da base de dados NBP ('x' conectados) e notas Q_{MOS} para o setup (0.5,0,0.15) ('•') usando o mapeamento do tipo PESQ.

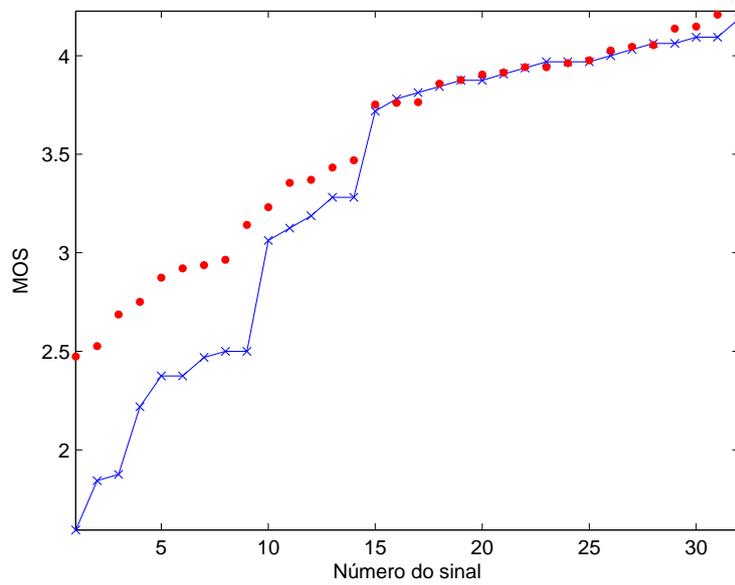


Figura 3.7: Notas subjetivas da avaliação da reverberação para os 32 sinais da base de dados MARDY ('x' conectados) e notas Q_{MOS} para o setup (0.5,0,0.15) ('•') usando o mapeamento do tipo PESQ.

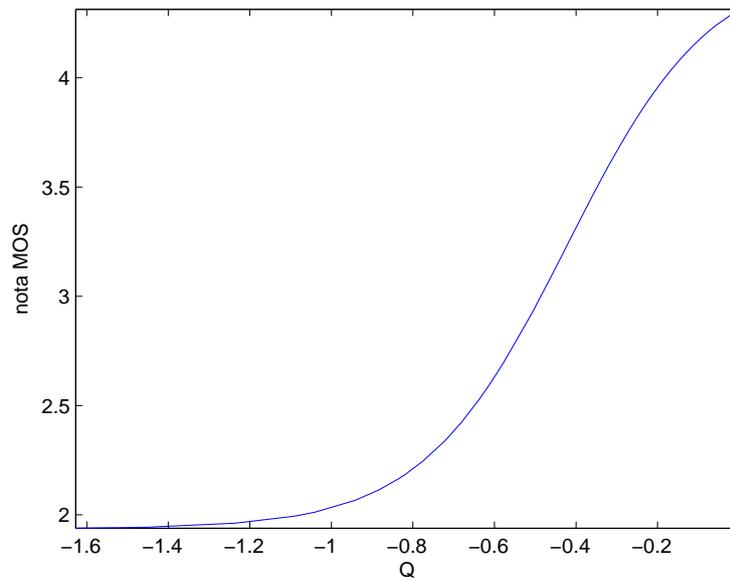


Figura 3.8: Função de mapeamento do tipo PESQ modificado para o caso de $\alpha = 1$, $\beta = 0$, $\gamma = 0.3$, e os coeficientes de mapeamento $c = -5.6503$, $d = -2.3826$, $e = 1.9358$ e $f = 2.5971$.

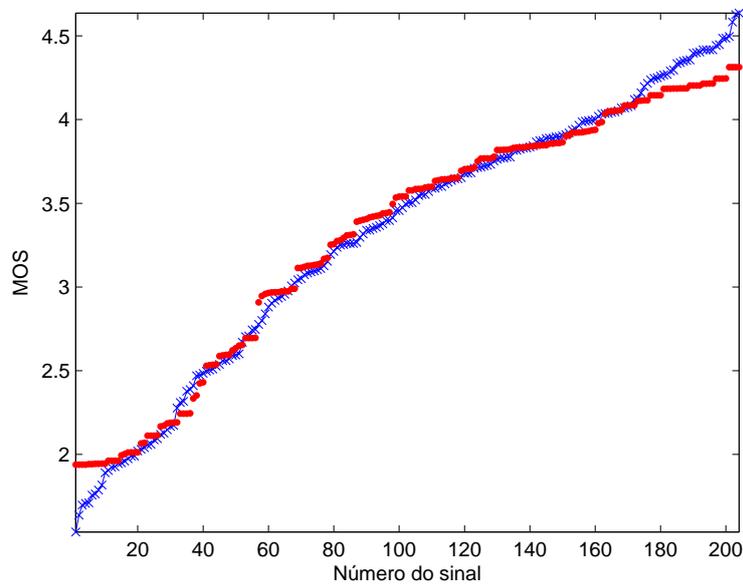


Figura 3.9: Notas subjetivas da avaliação da reverberação para os 204 sinais da base de dados NBP ('x' conectados) e notas Q_{MOS} para o setup (1,0,0.3) ('•') usando o mapeamento do tipo PESQ modificado.

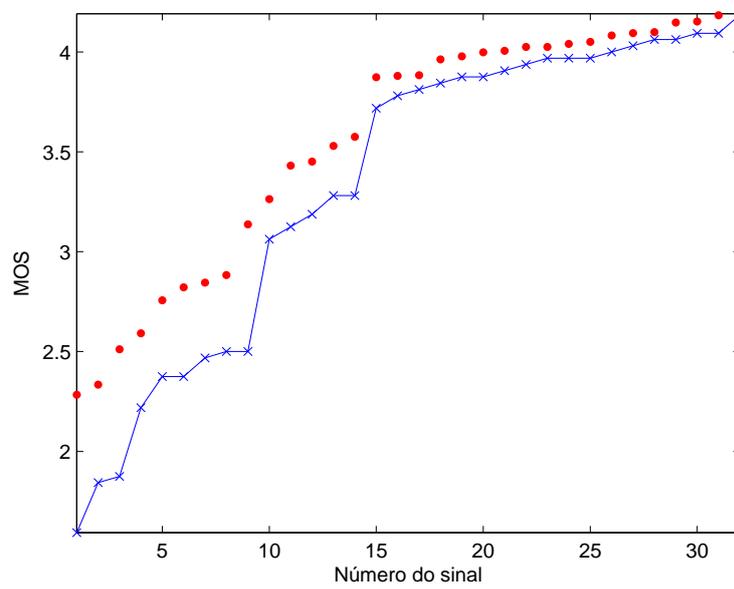


Figura 3.10: Notas subjetivas da avaliação da reverberação para os 32 sinais da base de dados MARDY ('x' conectados) e notas Q_{MOS} para o setup (1,0,0.3) ('•') usando o mapeamento do tipo PESQ modificado.

3.3 Conclusão

Neste capítulo abordamos a tarefa de procurar novas métricas para a melhora da avaliação do efeito de reverberação, incorporando, respectivamente, os expoentes α , β e γ aos componentes T_{60} , RSV , e R do sistema QAreverb, e testando esta métrica modificada para diferentes tipos de mapeamento não-linear.

O sistema modificado foi testado usando duas bases de dados compostas, respectivamente, por 204 e 32 sinais reverberantes. Os resultados indicaram um incremento das correlações das notas objetivas do método modificado com as notas subjetivas de ambas as bases para 93% e 95,5%, respectivamente, em comparação aos valores de 91% e 95% obtidos pelo sistema original.

Em particular, observamos em vários casos, que o expoente β para a RSV ficava nulo, indicando uma ausência deste componente na composição da métrica modificada final.

Capítulo 4

Aspectos Teóricos da Desreverberação

4.1 Introdução

Na primeira parte deste projeto (capítulos 2 e 3) foram analisadas diferentes formas de combinar os componentes da medida Q do sistema QAreverb original, obtendo novas configurações com importantes melhoras no desempenho da métrica resultante.

Esta segunda parte do projeto, que abarca os capítulos 4 e 5, vai-se valer dos resultados obtidos na primeira parte e, a partir da apresentação de novos conceitos relativos aos processos de desreverberação e filtragem adaptativa, melhorar o desempenho do algoritmo Wu e Wang, que também será apresentado neste capítulo.

Neste sentido, o presente capítulo inclui uma apresentação teórico-prática do conceito de desreverberação (seção 4.2) e do algoritmo de Wu e Wang para desreverberação de sinais de fala (seção ??). Após esta introdução, serão apresentados a filtragem de Wiener (seção 4.3) e um estudo de algoritmos adaptativos (seção 4.6) *least means squares* (LMS) e *set membership* (SM), a serem incorporados na algoritmo de Wu e Wang.

4.2 Algoritmo de Desreverberação de Wu e Wang

Como colocado no Capítulo 2.1, a reverberação é um efeito acústico que acontece em qualquer cenário fechado, produzindo uma degradação do sinal que tanto pode tornar o sinal mais agradável para o ouvinte quanto pode, até mesmo, dificultar sua inteligibilidade. Para tentar mitigar este efeito, os algoritmos de desreverberação tentam realizar uma desconvolução cega do sinal sem prévio conhecimento da RIR correspondente.

No algoritmo de Wu e Wang, uma primeira etapa consiste na estimação do filtro inverso para reduzir o efeito da coloração e uma segunda etapa perfaz a subtração espectral para minimizar a influência da reverberação em longo prazo. O diagrama de blocos deste algoritmo está representado na figura 4.1, onde $y(n)$ corresponde ao sinal reverberante, $z(n)$ corresponde ao sinal filtrado inversamente e $x(n)$ ao sinal desreverberado.



Figura 4.1: Diagrama de blocos do algoritmo de desreverberação de Wu-Wang.

Neste trabalho, procuramos modificar o processo de filtragem inversa do algoritmo de Wu e Wang, de modo a otimizar o seu desempenho em termos perceptuais, segundo o sistema QAreverb modificado na primeira parte deste projeto.

De fato, a ideia da filtragem inversa é reconstruir uma estimativa do sinal de fala original, reduzindo o efeito da coloração. Esta etapa, faz a estimação do filtro inverso a partir da maximização da curtose do sinal de resíduos da predição linear (PL), conforme indicado na figura 4.2.

O problema de estimar os coeficientes do filtro inverso pode ser analisado no domínio do tempo e no domínio da frequência. As seguintes subseções apresentam ambos os modelos.

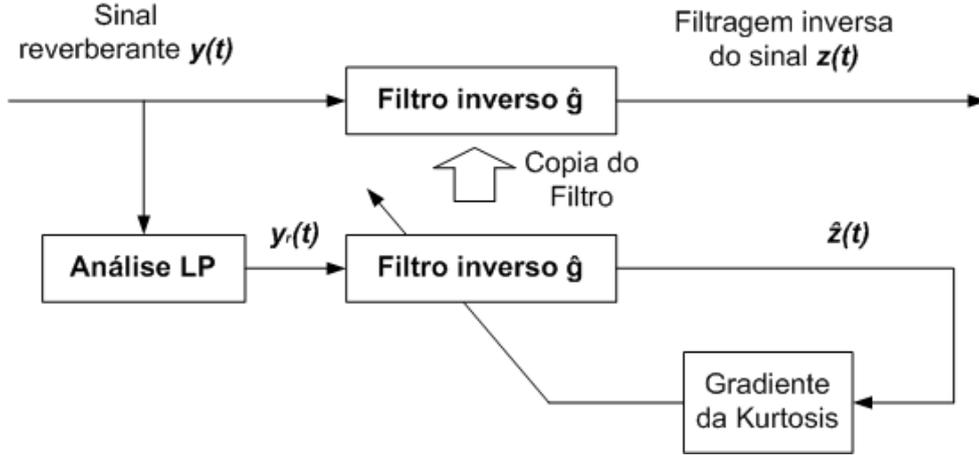


Figura 4.2: Esquema da filtragem inversa para o algoritmo de Wu e Wang.

4.2.1 Filtragem inversa no domínio do tempo

Para apresentar este estudo seguiremos o esquema da figura 4.2.

Sendo $\hat{g} = [g(1), g(2), \dots, g(L)]$ o filtro inverso de ordem L , o sinal do som filtrado inversamente é

$$z(t) = \hat{g}\hat{y}(t), \quad (4.1)$$

onde $\hat{y} = [y(t - L + 1), \dots, y(t - 1), y(t)]^T$ e y o sinal de fala com reverberação. No algoritmo original, este sinal tem uma frequência de amostragem de 16 kHz.

Para criar a melhor estimaco possvel do sinal devemos calcular o resduo y_r LP do sinal reverberante, que tem uma maior curtose do que o sinal original, tal que

$$\tilde{z}(t) = \hat{g}\hat{y}_r(t). \quad (4.2)$$

onde $\hat{y}_r = [y_r(t - L + 1), \dots, y_r(t - 1), y_r(t)]^T$.

O filtro inverso timo \hat{g}  obtido maximizando a curtose de $\tilde{z}(t)$, definida por

$$f(t) = \frac{4(E[\tilde{z}^2(t)]\tilde{z}^3(t) - E[\tilde{z}^4(t)]\tilde{z}(t))}{E^3[\tilde{z}^2(t)]}. \quad (4.3)$$

Esta otimizao pode ser feita mediante um algoritmo adaptativo, neste caso o algoritmo LMS (que ser apresentado na seo 4.4) para blocos de tamanho L do sinal. Usando uma busca baseada no gradiente da funo objetivo, o problema de otimizao pode ser formulado no domnio do tempo da seguinte maneira:

$$\hat{g}(t + 1) = \hat{g}(t) + \mu f(t)\hat{y}_r(t), \quad (4.4)$$

onde μ  a taxa de aprendizagem para cada passo.

4.2.2 Filtragem inversa no domínio da frequência

A formulação no domínio do tempo pode demorar muito em convergir, como comprovamos experimentalmente mais adiante. Desta forma, também é possível formular o processo da filtragem inversa no domínio da frequência, usando uma estrutura em blocos que melhora a convergência do processo adaptativo. Deste modo, o sinal é processado bloco a bloco usando a transformada rápida de Fourier (FFT).

O filtro inverso no domínio da frequência é descrito por

$$G'(n+1) = G(n) + \frac{\mu}{M} \sum_{m=1}^M F(m)Y_r^*(m), \quad (4.5)$$

$$G(n+1) = \frac{G'(n+1)}{G'(n+1)}, \quad (4.6)$$

onde $F(m)$ e $Y_r(m)$ são, respectivamente, as FFTs dos sinais $f(t)$ e $\hat{y}_r(t)$ para o bloco m , e o índice (*) denota o operador conjugado complexo. $G(n)$ é a FFT de \hat{g} na iteração número n , e M é o número de blocos.

Por fim, o sinal filtrado inversamente $z(t)$ é obtido pela convolução do som reverberante com o filtro inverso. Neste algoritmo foi escolhido um passo de $\mu = 3 \times 10^{-9}$ e um total de 500 iterações.

A implementação do algoritmo de Wu e Wang, fornecida pelos próprios autores, não está coerente com o artigo que descreve o algoritmo [?] original. De fato, na implementação fornecida, o algoritmo usa repetidamente a informação de todos os blocos para atualizar o filtro. Conforme indicado em [?], para ser coerente com o artigo [?], o sinal reverberante $z(n)$ deve ser dividido em M janelas de N amostras cada com 0% de sobreposição, onde $N = N_{ms} \times \frac{F_s}{1000}$, N_{ms} é o tamanho da janela em milissegundos e F_s a frequência de amostragem.

Os resíduos LP são calculados para cada janela usando um filtro LP com L coeficientes, gerando assim o resíduo $y_r(n; m)$, onde n é o índice de amostragem para a janela m . Esta implementação está coerente com o artigo original, em que se usa a informação de apenas um bloco para atualizar o filtro em cada iteração.

4.3 Filtragem de Wiener

O processo de filtragem inversa utilizado no algoritmo de desreverberação de Wu e Wang é um caso particular da filtragem de Wiener que é estudada nesta seção. O objetivo do filtro de Wiener é reduzir o ruído que tem um sinal, comparando esse sinal com uma estimativa do mesmo, sem ruído. Na filtragem de Wiener são assumidos os seguintes pontos:

- O sinal e o ruído são processos estacionários lineares, com espectro conhecido, ou autocorrelação e correlação cruzada conhecidas.
- O filtro deve ser fisicamente realizável.

Em particular, o filtro de interesse é a aplicação da filtragem de Wiener para o caso particular de respostas ao impulso finitas (FIR, do inglês *finite-duration impulse response*). Na figura 4.3 temos o diagrama de blocos do filtro FIR de Wiener para sequências em tempo discreto. Neste processo, o sinal de entrada $x(n)$ passa pelo filtro de Wiener $w(n)$ e o resultado $y(n)$ é comparado com um sinal de referência $d(n)$ gerando o erro da filtragem $e(n)$.

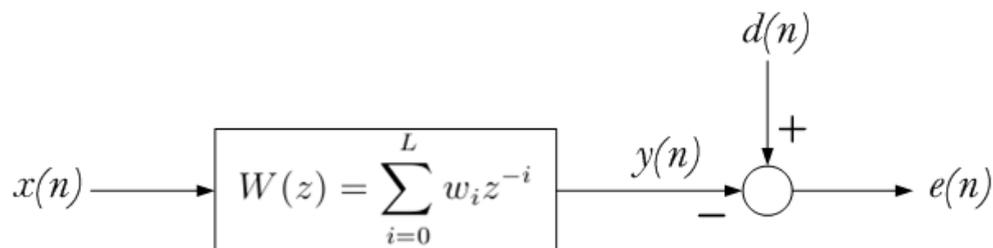


Figura 4.3: Diagrama de blocos do filtro de Wiener.

Na filtragem de Wiener, o objetivo é determinar os coeficientes $\mathbf{w} = [w_0 \ w_1 \ \dots \ w_L]^T$ de um filtro digital com relação entrada-saída descrita por

$$y(n) = w_0 x(n) + w_1 x(n-1) + \dots + w_L x(n-L), \quad (4.7)$$

onde L é a ordem do filtro, de modo que $y(n)$ seja a mais parecida possível a um sinal de referência $d(n)$ dado. Para avaliar o grau de similaridade entre $y(n)$ e $d(n)$, usa-se o erro quadrático médio (MSE, do inglês *mean squared error*) entre estes

sinais, tal que

$$\begin{aligned}
\xi &= E \{e^2(n)\} \\
&= E \{(d(n) - y(n))^2\} \\
&= E \{(d(n) - w_0x(n) - w_1x(n-1) - \dots - w_Lx(n-L))^2\}. \quad (4.8)
\end{aligned}$$

Para minimizar esta medida, igualamos o gradiente de ξ a zero,

$$\nabla_{\mathbf{w}}\xi = 2\mathbf{R}\mathbf{w}^* - 2\mathbf{p} = \mathbf{0}, \quad (4.9)$$

onde, assumindo estacionaridade do sinal de entrada,

$$\mathbf{R} = \begin{bmatrix} E\{x^2(n)\} & E\{x(n)x(n-1)\} & \dots & E\{x(n)x(n-L+1)\} \\ E\{x(n-1)x(n)\} & E\{x^2(n)\} & \dots & E\{x(n)x(n-L+2)\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{x(n-L+1)x(n)\} & E\{x(n-L+2)x(n)\} & \dots & E\{x^2(n)\} \end{bmatrix} \quad (4.10)$$

$$\mathbf{p} = \left[E\{x(n)x(n-1)\} \quad E\{x(n)x(n-2)\} \quad \dots \quad E\{x(n)x(n-L)\} \right]^T. \quad (4.11)$$

Com isto, a partir da expressão (4.9), é possível determinar os coeficientes do filtro ótimo pela relação

$$\mathbf{w}^* = \mathbf{R}^{-1}\mathbf{p}. \quad (4.12)$$

4.4 Algoritmo LMS

O algoritmo de Wiener descrito acima não é eficiente em termos computacionais para altas ordens L do filtro devido à inversão de matriz requerida na equação (4.12), que é uma operação de complexidade $O[L^2]$ ¹. Para evitar a inversão de matriz, Widrow e Hopf encontraram uma solução simples e elegante, originando o algoritmo LMS (no inglês, *least mean squares*). A idéia proposta foi fazer um algoritmo iterativo (neste contexto também chamado de adaptativo) para o vetor de coeficientes \mathbf{w} acompanhando o sentido oposto do gradiente da função objetivo ξ

¹Em geral, a complexidade da inversão de uma matriz $L \times L$ é da ordem de $O[L^3]$. Porém a estrutura regular da matriz \mathbf{R} , chamada de matriz Toeplitz (por ter coeficientes constantes em longo de todas suas diagonais principal e secundárias) permite algoritmos eficientes de inversão com complexidade $O[L^2]$

em função de \mathbf{w} , isto é

$$\begin{aligned}
\mathbf{w}(n+1) &= \mathbf{w}(n) - \nabla_{\mathbf{w}}\xi \\
&= \mathbf{w}(n) - \nabla_{\mathbf{w}}E\{e^2(n)\} \\
&= \mathbf{w}(n) - E\{\nabla_{\mathbf{w}}e^2(n)\} \\
&= \mathbf{w}(n) - 2E\{e(n)\nabla_{\mathbf{w}}e(n)\} \\
&= \mathbf{w}(n) - 2E\{e(n)\nabla_{\mathbf{w}}(d(n) - y(n))\} \\
&= \mathbf{w}(n) + 2E\{e(n)\nabla_{\mathbf{w}}y(n)\}.
\end{aligned} \tag{4.13}$$

Com isto, usando a relação (4.7), tem-se

$$\mathbf{w}(n+1) = \mathbf{w}(n) + 2E\{e(n)\mathbf{x}(n)\}, \tag{4.14}$$

onde $\mathbf{x}(n) = [x(n) \ x(n-1) \ \dots \ x(n-L)]^T$. O grande artifício, porém, do algoritmo LMS foi a substituição, na equação (4.14), do operador valor esperado $E\{\cdot\}$ por uma constante de ganho μ para controlar o passo de atualização do algoritmo resultante, fazendo com que o processo iterativo se tornasse extremamente simples e robusto:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{x}(n). \tag{4.15}$$

Este algoritmo é mais eficiente computacionalmente que o algoritmo de Wiener quando o número de coeficientes a inverter na matriz de correlações é alto, mas também precisa de ajustar o passo μ de modo a garantir um bom processo de convergência.

4.5 Aplicação Prática da Filtragem de Wiener

A equalização do canal é um caso particular da filtragem de Wiener, conforme descrito na figura 4.4.

Nesta configuração, o filtro adaptativo é usado para obter o modelo inverso de um sistema desconhecido. É por isto que esta configuração também é chamada filtragem inversa (que é o nome que aparece nos algoritmos de desreverberação). Numa situação ideal, o modelo inverso representa a função de transferência inversa da que caracteriza o sistema desconhecido. A versão atrasada do sistema desconhecido representa a resposta desejada para o filtro adaptativo.

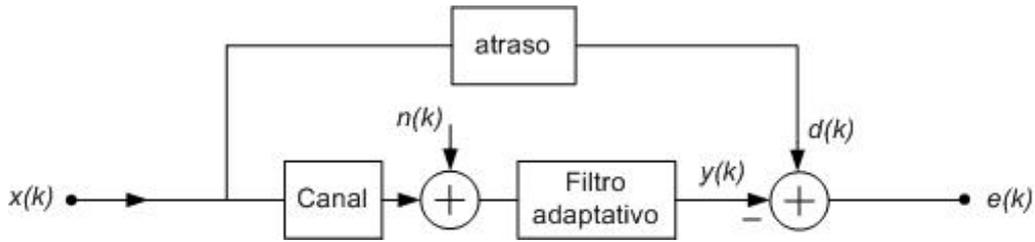


Figura 4.4: Representação em diagrama de blocos do problema de equalização do canal.

Por exemplo, para um esquema de uma sinusoidal com 1000 amostras: $x(n) = \text{sen}(n)$; com $n = 1 : 1000$ e um canal descrito pela função de transferência

$$H(z) = \frac{1}{(z^{-1} - 0.3)(z^{-1} + 0.2)}, \quad (4.16)$$

a convergência do sinal de saída do filtro adaptativo usando o algoritmo LMS, em relação ao sinal de saída desejada, é mostrada na Figura 4.5, e a função de erro $e(n)$ correspondente é indicada na Figura 4.6.

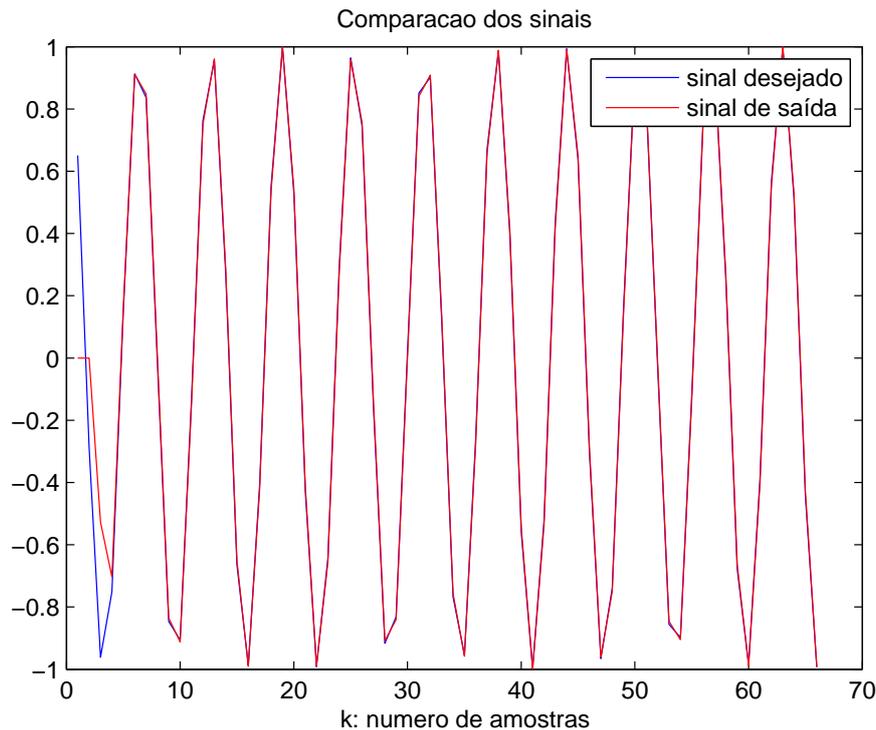


Figura 4.5: Convergência do sinal de saída do filtro adaptativo no problema de filtragem inversa.

Na figura 4.5 podemos ver como o sinal de saída $y(n)$ converge quase totalmente para o sinal de entrada $x(n)$ devidamente atrasado de um determinado

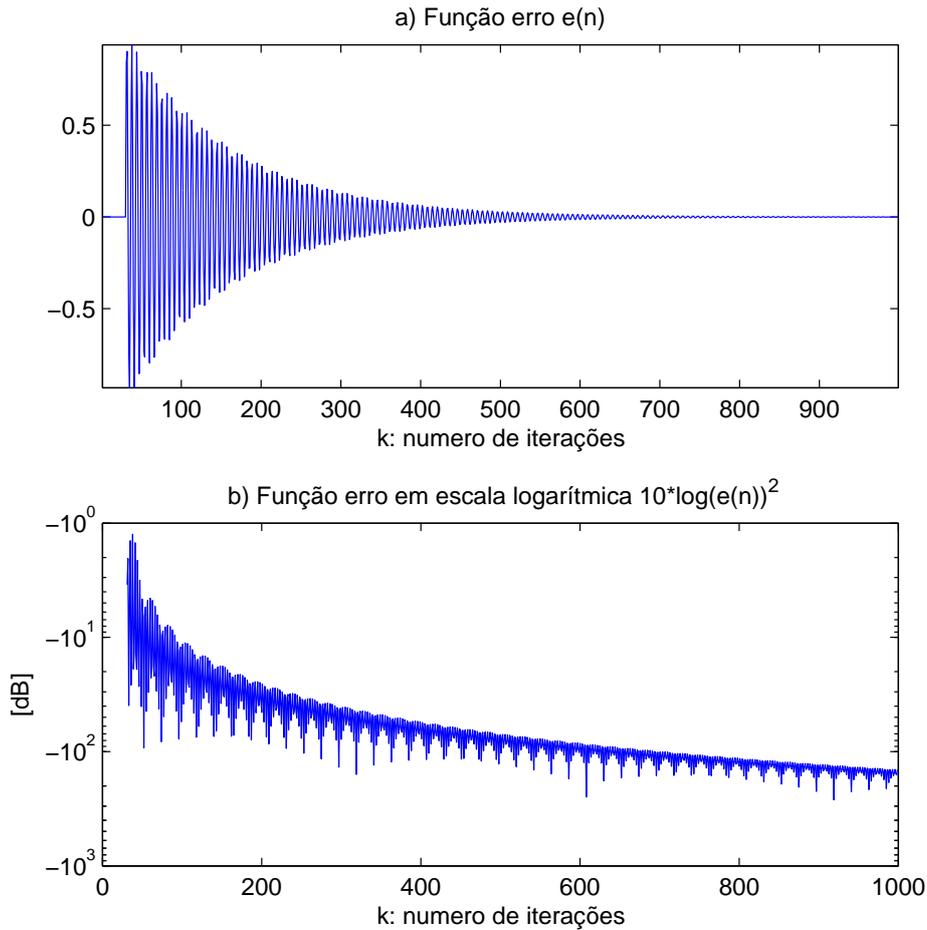


Figura 4.6: Convergência do sinal de erro na saída do filtro adaptativo no problema de filtragem inversa para *a)* escala lineal, e *b)* escala logarítmica.

número de amostras. Além disto, na figura 4.6 vemos, ainda nas primeiras iterações do algoritmo, um erro inicial significativo que rapidamente converge para zero após a devida equalização da função de transferência $H(z)$.

4.6 Algoritmo Adaptativo Alternativo

A técnica *set membership* (SM) é em verdade uma pequena modificação do algoritmo LMS com o intuito de reduzir o número de vezes que atualizamos os coeficientes $\mathbf{w}(n)$ do filtro adaptativo. A ideia básica do contexto SM é realizar um teste em cada iteração, comprovando se o valor absoluto do erro é maior que uma cota determinada Γ . Se é assim, então atualizamos os coeficientes $\mathbf{w}(n)$ do

filtro adaptativo, seguindo o esquema LMS. Caso contrário os coeficientes $\mathbf{w}(n+1)$ permanecem inalterados.

A proposta SM apresentada em [?] é dada por:

$$\mu(n) = \begin{cases} 1 - \frac{\Gamma}{|e(k)|} & , \text{se } |e(k)| \geq \Gamma \\ 0 & , \text{caso contrario} \end{cases} \quad (4.17)$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n) \mathbf{x}(n) \quad (4.18)$$

Na prática, um valor muito alto de Γ dificulta a convergência do algoritmo SM e um valor muito baixo torna este algoritmo similar ao LMS. Sendo assim, a escolha do limiar Γ adequado é feita empiricamente para cada aplicação.

De modo geral, o algoritmo SM é mais rápido e mais eficiente computacionalmente que o LMS. A única desvantagem dele é que precisa de uma boa cota Γ , pois caso contrário convergirá a um valor com mais erro que o desejado.

Capítulo 5

Modificações Propostas para o algoritmo de Wu-Wang

5.1 Introdução

O capítulo 4 foi uma introdução teórico-prática dos conceitos da filtragem inversa (em particular do algoritmo de desconvolução de Wu e Wang), assim como de alguns conceitos de filtragem adaptativa.

No presente capítulo introduzimos algumas modificações ao algoritmo original de Wu e Wang para melhorar o seu desempenho. Estas modificações começam pela implementação do algoritmo no domínio do tempo, com o que verificamos uma velocidade de convergência menor do que a do algoritmo no domínio da frequência. Após este experimento, consideramos o uso do algoritmo SM, em substituição ao LMS, o que melhorou muito tanto a velocidade de convergência como a nota MOS resultantes. Em seguida, fizemos a mesma modificação para o algoritmo em blocos apresentado na seção ??, melhorando ainda mais a qualidade do sinal desreverberado e a nota Q_{MOS} .

Em todos estes casos, o cálculo da nota $Q_{Areverb}$ da métrica original foi comparada com a nota obtida pelas novas métricas estudadas no capítulo 3, observando-se uma ligeira melhora neste sentido.

5.2 Considerações Iniciais

O objetivo deste capítulo é introduzir algumas modificações ao algoritmo de Wu e Wang, resumido no capítulo ??, para melhorar seu desempenho, de acordo com as seguintes medidas:

- **Eficiência:** Avaliada em termos de velocidade de convergência ou, equivalentemente, em número de iterações necessárias para convergência do algoritmo adaptativo.
- Q_{MOS} : Já considerando o sistema modificado estudado no capítulo 3. capítulo 3

Todas as modificações foram testadas para um único sinal.

Para a avaliação da nota Q mapeada à escala MOS de 1 a 5, além da métrica original de Q_{Areverb} usando o mapeamento de terceira ordem, apresentada na seção 2.4.2, fizemos uso das novas métricas de avaliação analisadas em 3.2. Mais especificamente o mapeamento dos tipos PESQ e PESQ modificado.

A seção a seguir apresenta as modificações propostas para o algoritmo de duas etapas de Wu e Wang.

5.3 Modificações do algoritmo de Wu-Wang no domínio do tempo

O algoritmo original de Wu e Wang foi desenvolvido no domínio da frequência, pois, como foi assinalado em [?], no domínio do tempo os coeficientes do filtro inverso podem demorar muito em convergir, ou às vezes nem convergir. De todas as maneiras, decidimos implementar este algoritmo no domínio do tempo para comprovar se, fazendo alguma modificação, podíamos melhorar algum dos indicadores apresentados na seção 5.2.

A primeira tentativa foi implementar este algoritmo amostra a amostra. Efetivamente, experimentos feitos demonstraram que o tempo de convergência que precisava o algoritmo no domínio do tempo era muito superior do que o tempo necessário para convergir usando o algoritmo no domínio da frequência. A partir destes resul-

tados, fizemos uma nova modificação, testando o sinal dividido em blocos, do mesmo modo que foi feito no algoritmo original de Wu e Wang.

Segundo o estudado em ??, o sinal reverberante $z(n)$ deve ser dividido em M janelas de N amostras com 0% de superposição, sendo $N = N_{ms} \times \frac{F_s}{1000}$, onde N_{ms} é o tamanho da janela (em ms) e F_s a frequência de amostragem. Para a implementação do algoritmo foi usado um tamanho do filtro $M = 1024$, um tamanho do bloco $N = 0.032F_s$ e uma frequência de amostragem de $F_s = 48$ kHz.

Fazendo assim, o tempo necessário para realizar a filtragem inversa foi muito inferior ao tempo que precisava o algoritmo filtrando amostra a amostra, e similar ao tempo requerido pelo algoritmo original no domínio da frequência. Devido ao tempo limitado para o desenvolvimento deste projeto, estas duas versões do algoritmo (amostra a amostra e dividindo em blocos) para ver se os resultados obtidos eram o suficientemente similares.

Mediante testes subjetivos feitos por 4 pessoas do laboratório determinamos que a qualidade acústica dos sinais não proporcionava nenhuma diferença significativa entre os sinais. Em seguida, comparamos a nota obtida pela métrica de QAreverb para cada uma das etapas do algoritmo de Wu e Wang, conforme indicado na tabela 5.1, resultando, mais uma vez, em números bastante similares para todos os efeitos práticos.

Tabela 5.1: Tabela comparativa das notas Q_{MOS} obtidas para o algoritmo de Wu e Wang no domínio do tempo e no domínio do tempo em blocos para um sinal de fala. Os resultados são comparados à saída de cada uma das etapas do algoritmo cego

	Domínio do tempo	Domínio do tempo em blocos
Saída 1ª etapa	2.1972	2.1975
Saída 2ª etapa	2.8205	2.8194

De modo geral, observamos que a nota obtida pelo algoritmo anterior era muito ruim quando comparada com a nota obtida com o algoritmo original ($Q_{MOS} = 3.88$), o que nos fez considerar outra modificação do algoritmo no domínio do tempo. Esta configuração consistiu em considerar blocos do mesmo tamanho que o tamanho da janela, e comparar assim somente partes do sinal de forma independente. Neste

caso, para a configuração do algoritmo foi usado um tamanho de janela de $M = N = 1024$ e um passo ótimo do algoritmo adaptativo $\mu = 10$. A execução deste algoritmo gerou resultados ligeiramente superiores ao anterior, mas ainda inferior à nota obtida pelo algoritmo original.

Numa nova tentativa, substituímos o algoritmo LMS pelo SM. Nesta nova análise vamos considerar ainda os mapeamentos da nota QAreverb que melhoram a correlação entre a nota objetiva e subjetiva da métrica original quando aplicando configurações ótimas de α , β e γ , conforme descrito na seção ???. O resultado é sintetizado na tabela 5.2.

Tabela 5.2: Tabela comparativa das notas Q_{MOS} obtidas usando o algoritmo de Wu e Wang original e as notas obtidas com as configurações ótimas dos mapeamentos PESQ e PESQ modificado, estudados no capítulo 3, implementados para o algoritmo de duas etapas no domínio do tempo. Os resultados são comparados na saída de cada uma das etapas do algoritmo cego, para o passo ótimo de $\mu = 10$

	Q_{map} 1ª etapa	Q_{map} 2ª etapa	Tempo duração(s)
Alg. Orig.	3.13	3.88	112.91
LMS (Map.3)	2.26	3.08	133.52
SM (Map.3)	2.20	2.81	1.15
LMS (PESQ)	1.74	3.75	132.64
SM (PESQ)	1.73	3.50	1.13
LMS (PESQ2)	1.95	3.85	134.46
SM (PESQ2)	1.95	3.57	1.09

Por essa tabela, podemos ver como o uso dos novos mapeamentos proporcionam uma melhora da nota Q_{map} quando comparada com o mapeamento de terceira ordem da métrica de QAreverb original (seção 2.4.2), sendo a melhor das notas aquela relativa à configuração ótima para o mapeamento PESQ modificado (seção 3.2.4), seguida da configuração ótima para o mapeamento PESQ original (seção 3.2.3). Também se pode observar que o tempo de duração do processamento quando usamos o algoritmo SM é muito menor, requerendo um tempo de duração da ordem de mais de 100 vezes menor.

Como se estudou na seção ??, a eficiência do algoritmo SM depende da escolha

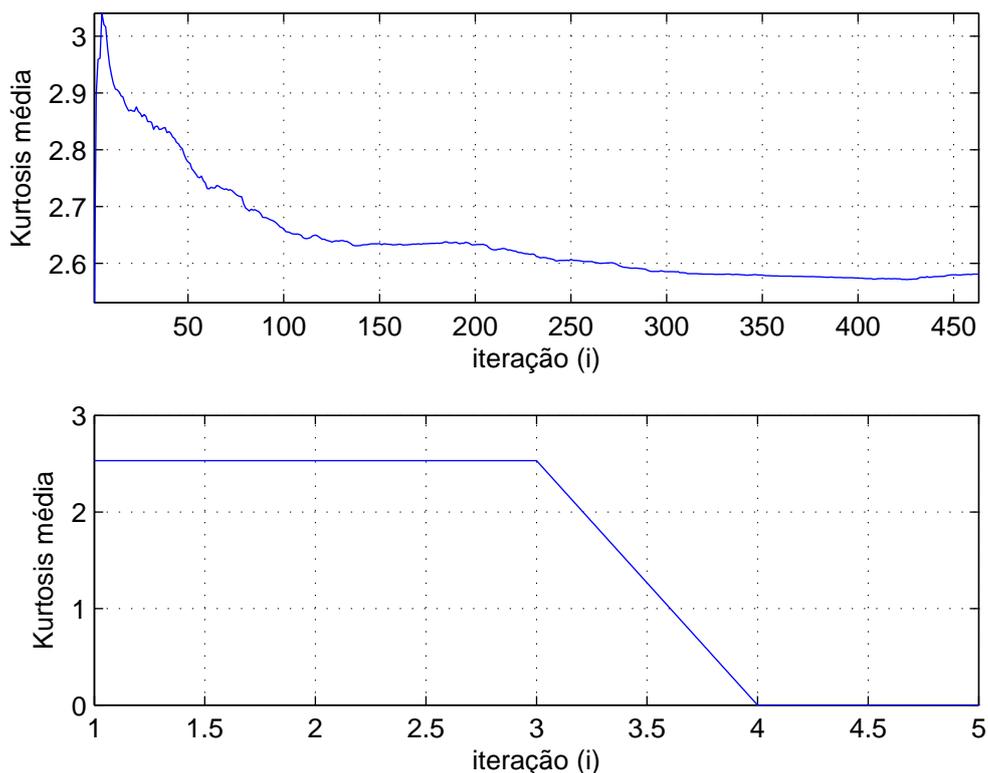


Figura 5.1: Curtose média do algoritmo de Wu e Wang no domínio do tempo para blocos do mesmo tamanho que o tamanho da janela, usando: (a) o algoritmo LMS; (b) o algoritmo SM com $\Gamma = 2.58$ e ponto de parada 10^{-3} .

de uma boa cota do erro para não limitar o erro mínimo a uma cota sub-ótima. No presente problema, a situação é a mesma substituindo a figura-de-mérito MSE pela curtose. Nesse sentido, tentamos estabelecer a melhor cota possível $\Gamma = 2.58$ pelo valor mínimo observado para a curtose do algoritmo LMS (figura 5.1), resultando num processo para o algoritmo SM conforme visto na figura 5.1, onde se observa a necessidade de apenas 3 iterações para se atingir o valor mínimo.

Infelizmente, porém, nenhuma das notas obtidas nestes testes com o algoritmo SM ($Q_{map} = 3.75$ para o mapeamento PESQ e $Q_{map} = 3.85$ para o mapeamento PESQ modificado) superaram o nota obtida pelo algoritmo de Wu-Wang original ($Q_{map} = 3.88$). Aproveitando os resultados obtidos nesta seção, onde se determinou que o algoritmo SM reduz o tempo de convergência na ordem de mais de 100 vezes, e que as configurações ótimas dos mapeamentos PESQ proporcionaram um melhoramento da nota Q_{map} quando comparados com a configuração ótima do

mapeamento de terceira ordem para a métrica original, a próxima análise foi o uso do domínio da frequência.

Neste caso, porém, o limiar do algoritmo SM foi escolhido como $\Gamma = 8.0433$, gerando o processo de convergência mostrado na figura 5.2.

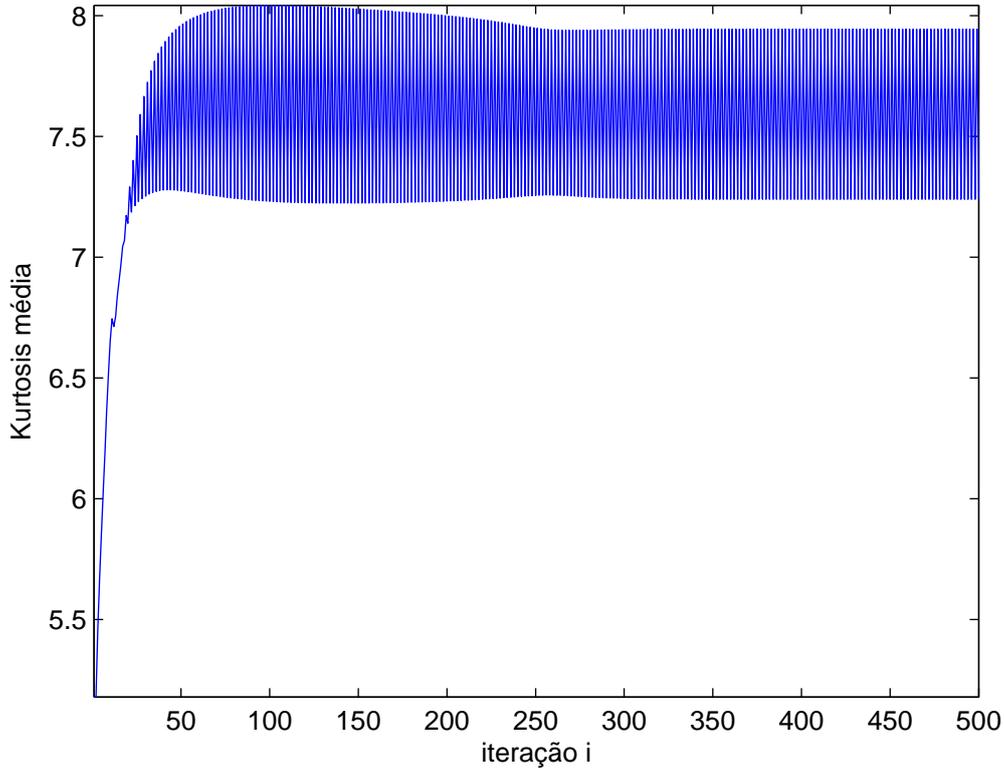


Figura 5.2: Curtose média do algoritmo original usando o passo $\mu = 3 \times 10^{-9}$, tamanho do filtro $M = 1024$, tamanho de bloco $N = 0.032x F_s$ e $N_i = 500$ iterações.

A tabela 5.3 mostra os resultados obtidos na análise de desempenho do algoritmo SM no domínio da frequência. Nela podemos ver como os tempos de duração para o algoritmo SM novamente são muito inferiores aos requeridos pelo algoritmo LMS, mas, novamente as notas ficam um pouco pior. Também se pode observar como os mapeamentos PESQ e PESQ modificado novamente ganham do mapeamento de terceira ordem.

Por fim, consideramos o uso de filtros com mesmo comprimento que o bloco de sinal, ainda no domínio da frequência. Neste caso, a escolha da cota limite para o algoritmo SM considerava o valor mínimo obtido e não o seu valor final. Com isto, com esta modificação, não somente conseguimos diminuir o tempo de duração,

Tabela 5.3: Avaliação do algoritmo de Wu e Wang, nas versões original e modificada no domínio da frequência com algoritmo SM.

	Q_{map} 1ª etapa	Q_{map} 2ª etapa	Tempo duração(s)
Map.3	3.13	3.88	112.91
Map.3 (SM)	1.86	3.20	0.78
PESQ	3.26	3.99	115.65
PESQ (SM)	1.77	3.61	0.76
PESQ Modif.	3.27	4.04	116.75
PESQ Modif. (SM)	1.95	3.71	0.67

mas também aumentamos a nota Q_{MOS} resultante, conforme indicado na tabela 5.4. Novamente a melhor nota para o algoritmo SM corresponde ao mapeamento PESQ modificado, que produz um aumento na nota do 6.5% quando comparado com o mapeamento original. O segundo melhor resultado corresponde ao mapeamento PESQ, que produz um aumento do 6.2%, e por último o mapeamento de terceira ordem, que melhorou em quase 1%.

Tabela 5.4: Avaliação do algoritmo de Wu e Wang, nas versões original e modificada no domínio da frequência (comprimento uniformizado dos blocos) com algoritmo SM.

	Q_{map} 1ª etapa	Q_{map} 2ª etapa	Tempo duração(s)
Alg. Orig.	3.13	3.88	112.91
LMS (Map.3)	2.68	3.70	52.71
SM (Map.3)	2.20	3.91	0.59
LMS (PESQ)	3.01	3.90	52.7
SM (PESQ)	2.68	4.12	0.63
LMS (PESQ2)	2.92	3.97	53.97
SM (PESQ2)	1.95	4.13	0.65

5.4 Conclusão

Neste capítulo foram testadas diferentes modificações do algoritmo de Wu e Wang que melhoraram o desempenho computacional dele, assim como a nota final Q_{MOS} do processo de desreverberação. Verificamos que o algoritmo no domínio do

tempo efetivamente era pior do que o algoritmo no domínio da frequência, baseando esta comparação em indicadores de eficiência e nota MOS.

Por outro lado, melhoramos o desempenho do algoritmo original no domínio da frequência, dividindo o sinal em blocos do mesmo tamanho que o tamanho da janela e modificando o algoritmo adaptativo LMS pelo algoritmo SM, reduzindo o tempo de duração da etapa de filtragem inversa numa ordem de mais de 100 vezes.

Também obtivemos outro resultado muito importante: Utilizando as novas métricas de avaliação desenvolvidas anteriormente neste projeto constatamos um aumento da nota com o uso do algoritmo SM. No caso, a melhor destas notas correspondeu a um aumento de 6.5% respeito ao algoritmo original, quando utilizando o algoritmo SM.

Capítulo 6

Conclusão

Neste projeto tivemos dois objetivos principais:

- O primeiro deles foi a busca de novas versões da métrica QAreverb para avaliação subjetiva do efeito de reverberação, incorporando expoentes α , β e γ aos três componentes do sistema (T_{60} , RSV , e R , respectivamente). O sistema foi testado usando duas bases de dados, uma delas composta por 204 sinais reverberantes e a outra composta por 32 sinais reverberantes. No melhor caso (já considerando novos mapeamentos), obtivemos valores de correlação de 93% e 95.5%, respectivamente, para cada uma das bases de dados, enquanto que a métrica QAreverb original obtinha correlações iguais a 91% e 95% para estas mesmas bases.
- O segundo objetivo foi testar diferentes modificações do algoritmo de duas etapas de Wu e Wang que melhoraram o desempenho computacional dele, assim como a nota final Q_{MOS} do processo de desreverberação. Melhoramos o desempenho do algoritmo original no domínio da frequência, dividindo o sinal em blocos do mesmo tamanho que o tamanho da janela e modificando o algoritmo adaptativo LMS pelo algoritmo *set-membership* (SM), reduzindo o tempo de duração da etapa de filtragem inversa numa ordem de mais de 100 vezes. Com esta mesma substituição, conseguimos aumentar a nota perceptual em cerca de 6.5% usando a nova métrica QAreverb com o mapeamento do tipo PESQ modificado.

Por tudo isto, consideramos satisfatórios os resultados aqui alcançados.

6.1 Propostas para estudos futuros

Para novos projetos que dariam continuidade ao aqui apresentado, se sugerem as seguintes propostas:

- Testar diferentes tipos de mapeamentos à escala MOS de 1 a 5 utilizando funções de mapeamento sempre crescentes, por exemplo o mapeamento logarítmico, e comprovar se é possível melhorar ainda mais a correlação entre a nota objetiva e subjetiva.
- Por outro lado, se recomenda repetir a análise feito no capítulo 5 para a base de dados NBP toda, pois todo o estudo feito aqui considerou um único sinal devido a que o tempo para fazer este projeto foi muito limitado. Estes experimentos mais amplos seriam fundamentais para validar de forma concreta os resultados aqui obtidos para um único sinal.
- Também se deveria implementar o algoritmo RLS para a etapa de filtragem inversa e comparar o resultado com os resultados obtidos na análise de capítulo 5.
- Por último, se propõe uma outra modificação do algoritmo de Wu e Wang que seria a troca da figura de mérito curtose por uma outra medida de qualidade, como por exemplo a entropia. Novamente, seria interessante investigar a ocorrência de algum ganho computacional ou mesmo na qualidade percebida do sinal desreverberado.

Referências Bibliográficas

- [1] T. de M. Prego, “Acerca da Reverberação de Sinais de Voz: Quantificação Perceptual e Aperfeiçoamento de Algoritmos de Desreverberação,” D.Sc. Thesis, COPPE/UFRJ, Junho de 2012.
- [2] T. de M. Prego, A. A. de Lima, S. L. Netto, B. Lee, A. Said, R. W. Schafer e T. Kalker, “A blind algorithm for reverberation-time estimation using subband decomposition of speech signals,” *Journal of Acoustical Society of America*, vol. 131, no. 4, pp. 2811-2816, Abril de 2012.
- [3] A. A. de Lima, T. de M. Prego, S. L. Netto, B. Lee, A. Said, R. W. Schafer, T. Kalker e M. Fozunbal, “On the quality-assessment of reverberated speech,” *Speech Communication*, vol. 54, no. 3, pp. 393-401, Março de 2012.
- [4] T. de M. Prego, A. A. de Lima e S. L. Netto, “Perceptual analysis of higher-order statistics in estimating reverberation,” *Proc. International Symposium on Communications, Control and Signal Processing*, Rome, Italy, Maio de 2012.
- [5] A. A. de Lima, T. de M. Prego, S. L. Netto, B. Lee, A. Said, R. W. Schafer, T. Kalker e M. Fozunbal, “Feature analysis for quality assessment of reverberated speech,” *Proc. IEEE International Workshop on Multimedia Signal Processing*, Rio de Janeiro, Brazil, Outubro de 2009.
- [6] M. Wu e D. Wang, “A Two-Stage algorithm for one-microphone reverberant speech enhancement,” *IEEE Trans. on Audio, Speech and Lang. Proc.*, vol. 14, Maio de 2006.