

CANCELAMENTO DE ECO ELÉTRICO EM REDES TELEFÔNICAS E  
ACÚSTICO EM SISTEMAS DE TELECONFERÊNCIA

Bruno Catarino Bispo

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO  
DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA  
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS  
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE  
EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Aprovada por:

---

Prof. Sergio Lima Netto, Ph.D.

---

Prof. Paulo Sergio Ramirez Diniz, Ph.D.

---

Prof. José Antonio Apolinário Junior, D.Sc.

RIO DE JANEIRO, RJ - BRASIL  
AGOSTO DE 2008

BISPO, BRUNO CATARINO

Cancelamento de Eco Elétrico em Redes  
Telefônicas e Acústico em Sistemas de  
Teleconferência [Rio de Janeiro] 2008

XV, 107 p., 29,7 cm (COPPE/UFRJ,  
M.Sc., Engenharia Elétrica, 2008)

Dissertação - Universidade Federal do  
Rio de Janeiro, COPPE

1.Cancelamento de Eco 2.Cancelamento  
de Eco Elétrico 3.Cancelamento de Eco  
Acústico 4.Filtros Adaptativos

I.COPPE/UFRJ      II.Título (série)

## Agradecimentos

Inicialmente, agradeço e dedico este trabalho aos meus pais Nelson e Maria Alice, por todo esforço realizado durante suas vidas para me proporcionar uma educação digna, pelo enorme amor e carinho, por sempre me apoiarem, incentivando a prosseguir com meus estudos, e à minha irmã Caroline, pelo seu companheirismo e exemplo de dedicação.

Agradeço e dedico à minha madrinha Alzira, ao meu padrinho Abílio e aos meus avós José Bernardo e Ernestina, por toda ajuda prestada aos meus pais durante a minha educação.

Dedico e agradeço à minha esposa Isabel que eu amo tanto, pelo carinho e dedicação em todos os momentos, principalmente nos de maior tensão, e pela compreensão da minha ausência em alguns momentos.

Agradeço ao meu orientador Prof. Sergio Lima Netto, pela sua paciência, pelos conselhos e pela experiência transmitida ao longo do desenvolvimento deste trabalho. Ao Prof. Luiz Wagner Pereira Biscainho pela amizade e pelo apoio. Aos demais professores pelos conhecimentos passados.

Aos amigos do Laboratório de Processamento de Sinais que contribuíram direta ou indiretamente, tanto técnica quanto moralmente, para a realização desta tese: Amaro Azevedo de Lima, Fábio Pacheco Freeland, Vagner Luiz Latsch, Tadeu Nagashima Ferreira, Rafael de Almeida Jesus, Fabiano Tondello Castoldi e Paulo Antonio Andrade Esquef. Aos amigos José Alberto Bassi Junior e Nelson Louza Júnior, pela amizade durante todos esses anos.

Por fim, agradeço a Deus, por tornar 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.)

## CANCELAMENTO DE ECO ELÉTRICO EM REDES TELEFÔNICAS E ACÚSTICO EM SISTEMAS DE TELECONFERÊNCIA

Bruno Catarino Bispo

Agosto/2008

Orientador: Sergio Lima Netto

Programa: Engenharia Elétrica

Sistemas de cancelamento de eco constituem uma importante ferramenta em redes telefônicas e sistemas de teleconferência. Ao falar, o usuário recebe de volta uma reprodução da própria voz que pode ser percebida como um incômodo, devendo ser eliminada ou atenuada.

Neste trabalho é apresentado o desenvolvimento de canceladores de eco elétrico, acústico mono-canal e multi-canal. São utilizados os algoritmos adaptativos clássicos existentes na literatura, assim como os algoritmos de detecção de conversação cruzada. É demonstrada a necessidade da derivação de algoritmos adaptativos robustos a distúrbios de curta duração. Neste sentido, uma alteração é proposta em uma solução atual.

No caso de eco acústico multi-canal, os inúmeros acoplamentos acústicos fazem necessário o decorrelacionamento dos sinais presentes nos canais. Neste contexto, a solução atual adicionando uma função não-linear dos sinais a si mesmos é avaliada. A influência dos mais modernos codificadores de sinais de voz banda larga (até 8 kHz) é estudada e proposta como solução. Por fim, resultados de simulações são apresentadas nos três ambientes em questão.

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

## ECHO CANCELLATION IN TELEPHONE NETWORK AND TELECONFERENCE SYSTEMS

Bruno Catarino Bispo

August/2008

Advisor: Sergio Lima Netto

Department: Electrical Engineering

Echo cancellation systems constitute an important issue in telephone network and teleconference systems. After talking, the user receives back his own voice. This effect can be annoying and should be suppressed or attenuated.

In this work it is presented the development of network, mono-channel and multi-channel acoustic echo cancellers. The adaptive algorithms in the literature are used as well as the double talk detection algorithms. It is demonstrated the necessity of deriving adaptive algorithms robust to short time disturbances. In this context, it is proposed an alteration in a current solution.

In the case of multi-channel acoustic echo, the several acoustics coupling make necessary the descorrelation of the signals present in the channels. The influence of the most current wideband (up to 8 kHz) speech codecs is analyzed and proposed as solution. Finally, simulation results are presented in the three environments considered.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Conceitos Básicos de Cancelamento de Eco</b>	<b>3</b>
2.1	Introdução . . . . .	3
2.2	Eco Elétrico . . . . .	4
2.3	Eco Acústico . . . . .	5
2.4	Cancelamento de Eco . . . . .	6
2.5	Filtragem Adaptativa . . . . .	7
2.5.1	Filtro de Wiener . . . . .	7
2.6	Algoritmos Adaptativos . . . . .	10
2.6.1	Algoritmo <i>Least Mean Square</i> . . . . .	10
2.6.2	Algoritmo <i>Normalized LMS</i> . . . . .	11
2.6.3	Algoritmo <i>Proportionate NLMS</i> . . . . .	11
2.6.4	Algoritmo <i>Normalized Data-Reusing LMS</i> . . . . .	12
2.6.5	Algoritmo <i>Binormalized Data-Reusing LMS</i> . . . . .	13
2.6.6	Algoritmo <i>Proportionate BNDR-LMS</i> . . . . .	13
2.6.7	Complexidade Computacional . . . . .	14
2.7	Conclusões . . . . .	17
<b>3</b>	<b>Cancelamento Robusto de Eco</b>	<b>18</b>
3.1	Introdução . . . . .	18
3.2	Descrição do Problema . . . . .	19
3.3	Detector de Conversação Cruzada . . . . .	21
3.3.1	Algoritmo de Geigel . . . . .	21
3.3.2	Algoritmos baseados em correlação cruzada . . . . .	22

3.4	Algoritmos Robustos . . . . .	25
3.4.1	Algoritmos Robustos Tipo 1 . . . . .	26
3.4.2	Algoritmos Robustos Tipo 2 . . . . .	30
3.5	Conclusões . . . . .	33
<b>4</b>	<b>Cancelamento de Eco Elétrico</b>	<b>35</b>
4.1	Introdução . . . . .	35
4.2	Percursos de Eco . . . . .	36
4.3	Sinais <i>Composite Source Signal</i> . . . . .	39
4.3.1	Sinal Sonoro Limitado em Banda . . . . .	39
4.3.2	Sinal Surdo . . . . .	41
4.3.3	Silêncio . . . . .	43
4.4	Método de Análise Qualitativa . . . . .	43
4.4.1	Métrica Utilizada . . . . .	43
4.4.2	Dispositivo de Medição . . . . .	44
4.5	Simulações . . . . .	45
4.5.1	Teste de Convergência . . . . .	45
4.5.1.1	Descrição e Objetivos . . . . .	45
4.5.1.2	Resultados . . . . .	47
4.5.2	Teste de Re-Convergência . . . . .	51
4.5.2.1	Descrição e Objetivos . . . . .	51
4.5.2.2	Resultados . . . . .	52
4.5.3	Teste de Conversação Cruzada . . . . .	56
4.5.3.1	Descrição e Objetivos . . . . .	56
4.5.3.2	Resultados . . . . .	56
4.6	Conclusões . . . . .	60
<b>5</b>	<b>Cancelamento de Eco Acústico Mono-Canal</b>	<b>62</b>
5.1	Introdução . . . . .	62
5.2	Configuração dos Experimentos . . . . .	63
5.3	Resultados . . . . .	65
5.3.1	Convergência . . . . .	65
5.3.2	Conversação Cruzada . . . . .	69

5.4	Conclusões . . . . .	71
<b>6</b>	<b>Cancelamento de Eco Acústico Multi-Canal</b>	<b>73</b>
6.1	Introdução . . . . .	73
6.2	Problema da Não-Unicidade da Solução . . . . .	74
6.3	Métodos de Análise Qualitativa . . . . .	78
6.3.1	Desalinhamento . . . . .	78
6.3.2	Função Coerência . . . . .	78
6.3.3	W-PESQ . . . . .	79
6.4	Análises Preliminares . . . . .	80
6.4.1	Configuração . . . . .	80
6.4.2	Resultados . . . . .	82
6.5	Algoritmos Adaptativos Multi-Canal . . . . .	93
6.5.1	Configuração dos Experimentos . . . . .	93
6.5.2	Convergência . . . . .	94
6.5.3	Conversação Cruzada . . . . .	97
6.6	Conclusões . . . . .	98
<b>7</b>	<b>Conclusões</b>	<b>100</b>

# Lista de Figuras

2.1	Versão simplificada de um sistema de longa distância. . . . .	4
2.2	Sistema de longa distância com supressores de eco. . . . .	5
2.3	Diagrama ilustrativo do eco acústico. . . . .	6
2.4	Cancelamento de eco elétrico por filtro adaptativo. . . . .	6
2.5	Cancelamento de eco acústico por filtro adaptativo. . . . .	7
2.6	Representação em diagrama de blocos do filtro de Wiener. . . . .	8
3.1	Cancelamento de eco acústico. . . . .	20
3.2	Funções $\psi(z)$ do algoritmo tipo 1 (linha sólida) e tipo 2 (linha tracejada) para $k_0 = 1$ . . . . .	32
4.1	Modelamento de percurso de eco. . . . .	35
4.2	Resposta ao impulso do modelo de percurso de eco 1 com $ERL = 6$ dB e $\delta = 0$ . . . . .	37
4.3	Resposta ao impulso do modelo de percurso de eco 5 com $ERL = 6$ dB e $\delta = 0$ . . . . .	38
4.4	Resposta ao impulso do modelo de percurso de eco 6 com $ERL = 6$ dB e $\delta = 0$ . . . . .	38
4.5	Modelo do sinal CSS [1]. . . . .	39
4.6	Trecho sonoro limitado em banda do sinal CSS para conversação simples. . . . .	40
4.7	Trecho sonoro limitado em banda do sinal CSS para conversação cruzada. . . . .	40
4.8	Resposta em frequência do filtro utilizado para limitar em banda o sinal surdo. . . . .	42
4.9	Resposta em frequência do filtro passa-faixa. . . . .	44

4.10	Esquema para teste de convergência [1]. . . . .	46
4.11	Atenuação necessária ao longo do tempo [1]. . . . .	46
4.12	Atenuação necessária após $(10 + t_d + t_0)$ s [1]. . . . .	47
4.13	Velocidade de convergência dos algoritmos NLMS convencional (linha tracejada) e robusto tipo 1 (linha contínua). . . . .	48
4.14	Velocidade de convergência dos algoritmos robustos tipo 1 NLMS (linha tracejada), PNLMS (linha contínua) e BNDR-LMS (linha ponto-traço). . . . .	49
4.15	Velocidade de convergência dos algoritmos robustos tipo 1 BNDR-LMS (linha tracejada) e P-BNDR-LMS (linha contínua). . . . .	49
4.16	Velocidade de convergência dos algoritmos NLMS robusto tipo 1 (linha contínua) e robusto tipo 2 (linha tracejada). . . . .	50
4.17	Velocidade de convergência dos algoritmos P-BNDR-LMS robusto tipo 1 (linha contínua) e robusto tipo 2 (linha tracejada). . . . .	50
4.18	Esquema para teste de re-convergência [1]. . . . .	51
4.19	Atenuação necessária ao longo do tempo [1]. . . . .	52
4.20	Velocidade de re-convergência dos algoritmos NLMS convencional (linha tracejada) e robusto tipo 1 (linha contínua). . . . .	53
4.21	Velocidade de re-convergência dos algoritmos robustos tipo 1 NLMS (linha tracejada), PNLMS (linha contínua) e BNDR-LMS (linha ponto-traço). . . . .	53
4.22	Velocidade de re-convergência dos algoritmos robustos tipo 1 PNLMS (linha ponto-traço), BNDR-LMS (linha contínua) e P-BNDR-LMS (linha tracejada). . . . .	54
4.23	Velocidade de re-convergência dos algoritmos NLMS robusto tipo 1 (linha contínua) e robusto tipo 2 (linha tracejada). . . . .	55
4.24	Velocidade de re-convergência dos algoritmos P-BNDR-LMS robusto tipo 1 (linha contínua) e robusto tipo 2 (linha tracejada). . . . .	55
4.25	Esquema para teste de conversação cruzada [1]. . . . .	56
4.26	Desempenho em situação de conversação cruzada do algoritmo NLMS convencional (linha tracejada) e robusto tipo 1 (linha contínua). . . . .	57

4.27	Desempenho em situação de conversação cruzada dos algoritmos robustos tipo 1 NLMS (linha tracejada), PNLMS (linha contínua) e BNDR-LMS (linha ponto-traço).	58
4.28	Desempenho em situação de conversação cruzada dos algoritmos robustos tipo 1 BNDR-LMS (linha tracejada) e P-BNDR-LMS (linha contínua).	58
4.29	Desempenho em situação de conversação cruzada do algoritmo NLMS robusto tipo 1 (linha contínua) e robusto tipo 2 (linha tracejada).	59
4.30	Desempenho em situação de conversação cruzada do algoritmo P-BNDR-LMS robusto tipo 1 (linha contínua) e robusto tipo 2 (linha tracejada).	60
5.1	Experimento 1: Estimativa da resposta ao impulso $\mathbf{h}$ da sala de teleconferência.	64
5.2	Erro médio quadrático dos algoritmos LMS, NLMS, PNLMS e NDR-LMS.	65
5.3	Erro médio quadrático dos algoritmos NDR-LMS, BNDR-LMS e P-BNDR-LMS.	66
5.4	Erro médio quadrático dos algoritmos BNDR-LMS, SM-BNDR-LMS e P-BNDR-LMS.	67
5.5	Erro médio quadrático do SM-BNDR-LMS com os valores de $\gamma_0$ mostrados na Tabela 5.2.	68
5.6	Comportamento do erro médio quadrático durante interrupção na adaptação dos coeficientes.	70
5.7	Comportamento do erro médio quadrático durante pausa na adaptação.	70
6.1	Cancelamento de eco acústico multi-canal.	74
6.2	Respostas ao impulso: (a) $\mathbf{g}_1$ e (b) $\mathbf{g}_2$ da sala de transmissão; (c) $\mathbf{h}_1$ e (d) $\mathbf{h}_2$ da sala de recepção.	81
6.3	Módulo da coerência para diferentes níveis de decorrelação: (a) $\alpha = 0$ , (b) $\alpha = 0,2$ , (c)0,4 e (d)0,6.	82

6.4	Desalinhamento utilizando diferentes níveis de decorrelação: (a) $\alpha = 0,2$ , (b)0,4 e (c)0,6. A linha tracejada refere-se ao caso sem decorrelação. . . . .	83
6.5	Relação entre a média do módulo da coerência e o MOS individual dos canais (linha contínua) e entre a média do módulo da coerência e o MOS conjunto dos canais (linha tracejada); o ponto em destaque indicado por * é o caso sem decorrelação. . . . .	84
6.6	Relação entre a média do desalinhamento e o MOS individual dos canais (linha contínua) e entre a média do desalinhamento e o MOS conjunto dos canais (linha tracejada); o ponto em destaque indicado por * é o caso sem decorrelação. . . . .	85
6.7	Módulo da coerência com o codificador Speex com diferentes taxas de codificação: (a)32, (b)26, (c)20 e (d)12 kbps. . . . .	86
6.8	Desalinhamento com o codificador Speex para diferentes taxas de codificação: (a)32, (b)26, (c)20 e (d)12 kbps. A linha tracejada refere-se ao caso sem decorrelação. . . . .	87
6.9	Módulo da coerência com o codificador G.729.1 para diferentes taxas de codificação: (a)32, (b)26, (c)20 e (d)12 kbps. . . . .	87
6.10	Desalinhamento com o codificador G.729.1 para diferentes taxas de codificação: (a)32, (b)26, (c)20 e (d)12 kbps. A linha tracejada refere-se ao caso sem decorrelação. . . . .	88
6.11	Módulo da coerência com a codificação híbrida para diferentes taxas de codificação: (a)32, (b)26, (c)20 e (d)12 kbps. . . . .	89
6.12	Desalinhamento com a codificação híbrida para diferentes taxas de codificação: (a)32, (b)26, (c)20 e (d)12. A linha tracejada refere-se ao caso sem decorrelação. . . . .	90
6.13	Relação entre a média do módulo da coerência e o MOS estéreo dos canais para os diferentes métodos de decorrelação: função retificadora de meia onda (linha contínua), codificador Speex (círculos), codificador G.729.1 (quadrados) e codificação híbrida (triângulos). . . . .	91

6.14	Relação entre a média do desalinhamento e o MOS estéreo dos canais para os diferentes métodos de decorrelação: função retificadora de meia onda (linha contínua), codificador Speex (círculos), codificador G.729.1 (quadrados) e codificação híbrida (triângulos). . . . .	92
6.15	Erro médio quadrático dos algoritmos LMS, NLMS, PNLMS e NDR-LMS. . . . .	94
6.16	Erro médio quadrático dos algoritmos NDR-LMS, BNDR-LMS e P-BNDR-LMS. . . . .	95
6.17	Erro médio quadrático dos algoritmos BNDR-LMS, SM-BNDR-LMS e P-BNDR-LMS. . . . .	96
6.18	Erro médio quadrático do SM-BNDR-LMS com os valores de $\gamma_0$ mostrados na Tabela 6.6. . . . .	96
6.19	Comportamento do erro médio quadrático durante interrupção na adaptação dos coeficientes no caso multi-canal. . . . .	98

# Lista de Tabelas

2.1	Complexidade computacional convencional por atualização dos algoritmos adaptativos. . . . .	14
2.2	Complexidade computacional prática por atualização dos algoritmos adaptativos. . . . .	16
2.3	Algoritmo P-BNDR-LMS Rápido . . . . .	16
3.1	Algoritmo NLMS Robusto Tipo 1 . . . . .	30
3.2	Algoritmo NLMS Robusto Tipo 2. . . . .	33
4.1	Fatores de escalamento $K_i$ e $ERLs$ mínimos [1]. . . . .	37
4.2	Filtro ideal e filtro utilizado para limitar em banda o sinal surdo sem alteração de energia. . . . .	42
4.3	Parâmetros dos algoritmos adaptativos. . . . .	45
5.1	Parâmetros dos algoritmos adaptativos. . . . .	65
5.2	Valores de $\gamma_0$ utilizados no algoritmo SM-BNDR-LMSa e seus respectivos percentuais de atualização. . . . .	69
6.1	Escala MOS para os testes ACR e DCR . . . . .	79
6.2	Resultados da descorrelação para a função retificadora de meia onda com diferentes valores de $\alpha$ . . . . .	90
6.3	Resultados da descorrelação utilizando o codificador Speex com diferentes taxas de codificação. . . . .	91
6.4	Resultados da descorrelação utilizando o codificador G.729.1 com diferentes taxas de codificação. . . . .	91
6.5	Resultados da descorrelação utilizando codificação híbrida com diferentes taxas de codificação. . . . .	92

6.6	Valores de $\gamma_0$ utilizados no algoritmo SM-BNDR-LMS e seus respectivos percentuais de atualização. . . . .	97
-----	--	----

# Capítulo 1

## Introdução

As pessoas geralmente conversam na presença de ecos, que são produzidos pelas reflexões dos sinais de voz nas janelas, objetos e paredes. Normalmente, o atraso entre o sinal de voz e os sinais de eco é pequeno, na ordem de alguns poucos milissegundos, o que faz com que a existência de eco não seja percebida ou que o seu efeito soe mais como uma reverberação.

A maioria das pessoas prefere conversar na presença de uma certa quantidade de reverberação a conversar em um ambiente completamente anecóico, e a quantidade desejável de reverberação depende do ambiente de interesse. Por exemplo, em uma sala de concerto deseja-se uma grande quantidade de reverberação enquanto em uma sala de reunião deseja-se uma pequena quantidade.

No entanto, em conversas telefônicas, seja por telefonia padrão, telefonia móvel ou VoIP, o atraso pode alcançar centenas de milissegundos devido às várias etapas de processamento do sinal de voz: passagem por cabos ou pelo ar, equipamentos de conversão analógico/digital e enlaces via satélites em ligações internacionais. Nestes casos, o eco é percebido como uma perturbação incômoda e deve ser reduzido, ou de preferência, eliminado.

Se a conexão telefônica ocorrer entre dois *handsets*, o único tipo de eco encontrado é o eco elétrico ou de rede. Em conexões locais este tipo de eco não é problema pois os níveis de eco são baixos e os atrasos muito curtos. Porém, em conexões a longa distância, o atraso não é insignificante sendo o eco claramente identificado. A principal fonte de eco elétrico é um circuito denominado híbrida.

Quando a conversação telefone ocorre entre telefones *hands-free* ou entre duas

salas de teleconferência, eco é gerado pelo acoplamento acústico entre a caixa de som e o microfone. Este tipo de eco é denominado eco acústico e, em comparação ao eco elétrico, é mais difícil de ser eliminado pois a resposta ao impulso do percurso do eco é, geralmente, bem maior que no caso elétrico e pode ser alterada a qualquer momento devido à, por exemplo, uma pessoa em movimento.

Algumas novas aplicações exigem uma comunicação mais realista onde faz-se necessário a utilização de sons multi-canal como, por exemplo, estéreo. Nestas aplicações multi-canal existem diversos acoplamentos acústicos entre caixas de som e microfones, e os ecos gerados por estes acoplamentos devem ser cancelados. Entretanto, complicações devido a estes diversos acoplamentos, que não ocorrem no caso mono-canal, fazem este caso difícil de ser contornado.

Esta dissertação tem como objetivo o desenvolvimento de canceladores de eco elétrico para redes telefônicas e acústico para sistemas de teleconferência mono e multi-canal. Apesar de diferirem na natureza e na dificuldade, o cancelamento destes três tipos de ecos é realizado por meio da mesma ferramenta: filtros adaptativos. Os conceitos básicos de cancelamento de eco são descritos no Capítulo 2, onde os algoritmos adaptativos utilizados nesta dissertação são derivados.

O Capítulo 3 apresenta a dificuldade de cancelar o eco em situações de conversação cruzada, onde os usuários falam simultaneamente, sendo necessária a utilização de algoritmos detectores de conversação cruzada e tornar os algoritmos adaptativos robustos a distúrbios de curta duração. Para o último caso, são apresentadas duas soluções existentes na literatura e proposta uma alteração em uma delas.

Neste contexto, o Capítulo 4 descreve o desenvolvimento de um cancelador de eco elétrico de acordo com a atual norma em vigor. O Capítulo 5 apresenta resultados de simulações dos algoritmos adaptativos em um ambiente de cancelamento de eco acústico mono-canal.

O Capítulo 6 aborda o cancelamento de eco acústico multi-canal, explicando a necessidade de descorrelacionar os canais devido aos inúmeros acoplamentos acústicos. Para este problema, são avaliadas uma solução atual encontrada na literatura e a influência dos codificadores de voz. Por fim, o Capítulo 7 conclui esta dissertação resumindo os assuntos abordados e as diferentes soluções existentes, enfatizando os resultados obtidos nos desenvolvimentos dos canceladores de eco.

# Capítulo 2

## Conceitos Básicos de Cancelamento de Eco

### 2.1 Introdução

Apesar das pessoas geralmente conversarem na presença de ecos, estes podem não ser percebidos ou serem confundidos com reverberação dependendo do atraso entre o sinal de voz e os sinais de eco. No entanto, em conversas telefônicas a longas distâncias ou em sistemas de teleconferência, devido ao grande atraso, o eco é claramente identificado tornando-se indesejado.

As Seções 2.2 e 2.3 apresentam, respectivamente, os motivos pelos quais eco elétrico em centrais telefônicas e acústico em sistemas de teleconferência são gerados. A Seção 2.4 apresenta a atual solução para o problema de cancelamento de eco, a utilização de filtros adaptativos, que consiste em sintetizar uma cópia do sinal de eco e subtraí-la do sinal a ser transmitido.

Tal ferramenta, descrita na Seção 2.5, faz uso de algoritmos recursivos de otimização para atualizar os seus coeficientes. A Seção 2.6 descreve brevemente os algoritmos adaptativos clássicos que serão utilizados ao longo desta dissertação e suas respectivas complexidades computacionais. Por fim, a Seção 2.7 resume o presente capítulo.

## 2.2 Eco Elétrico

O eco elétrico na linha telefônica pode ser entendido como uma ligação simplificada entre dois usuários, como representado na Figura 2.1.

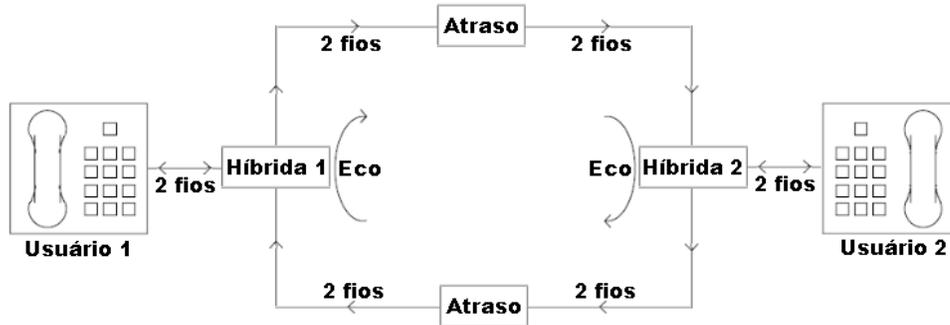


Figura 2.1: Versão simplificada de um sistema de longa distância.

Nesta figura, as ligações entre os usuários e as suas híbridas representam as ligações realizadas por meio de dois fios entre os respectivos telefones e centrais, e nesta parte do circuito ambas as direções de comunicação se propagam no mesmo par de fios. Entre as centrais, as ligações ocorrem por meio de quatro fios e são representadas pelas duas conexões *half-duplex* entre a híbrida 1 e a híbrida 2, sendo o termo híbrida o nome dado ao circuito telefônico que é responsável pela conversão de dois para quatro fios.

O papel do circuito híbrido 1 é conseguir que o sinal proveniente do usuário 2 seja todo dirigido para o segmento de dois fios do usuário 1, não permitindo o retorno de qualquer energia desse sinal. Na prática não se consegue este objetivo porque a variação no comprimento do *loop* do usuário, as características individuais dos aparelhos e a impedância da linha causam a separação imperfeita entre os sinais de emissão e de recepção, existindo sempre o retorno de alguma energia do sinal transmitido. Por conseguinte, o usuário 2 recebe inevitavelmente uma cópia atrasada e atenuada do seu sinal de fala.

A importância do sinal de eco depende do atraso, sendo necessário o seu controle quando o atraso ultrapassa 25 ms [2]. Quanto maior o atraso do sinal retornado mais incômodo se torna ouvi-lo e, conseqüentemente, maior atenuação deve sofrer. Quando não houver atenuação elevada do sinal retornado, a perturbação pode ser tal que a conversação torna-se inviável.

Ainda hoje, os sistemas de comunicações intercontinentais utilizam supressores de eco como mostrado na Figura 2.2, segundo [3]. Estes dispositivos são interruptores comandados por voz que abrem o circuito de emissão quando existe apenas sinal de recepção. O seu comportamento é no entanto deficiente por não haver supressão do eco em períodos de conversação cruzada (quando os locutores falam simultaneamente) onde, neste caso, é suspenso o funcionamento do supressor de eco por meio de um sinal de  $2100 \pm 15$  Hz a um nível de  $-12 \pm 6$  dB até que um silêncio de  $300 \pm 100$  ms seja detectado. Além disso, tal dispositivo não apresenta boa eficiência quando o atraso é acima de 100 ms [4], como ocorre em chamadas intercontinentais via satélite, onde o atraso envolvido em uma transmissão para um satélite e o retorno para a Terra é por volta de 250 ms.

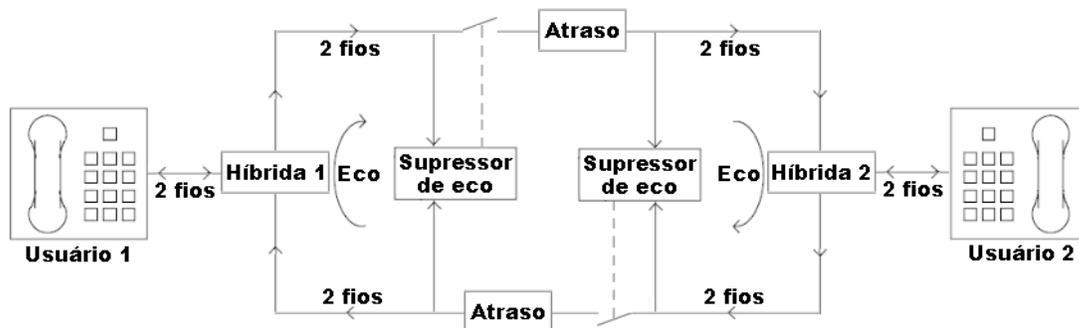


Figura 2.2: Sistema de longa distância com supressores de eco.

## 2.3 Eco Acústico

O eco acústico é o resultado do acoplamento acústico entre o alto-falante e o microfone no outro extremo da linha, como pode ser visto na Figura 2.3. Isto significa que ao falar, o usuário recebe de volta uma reprodução da própria fala que é percebida como uma perturbação incômoda. A popularização dos sistemas de teleconferência e a diminuição do tamanho dos telefones celulares tornam cada vez mais importante o cancelamento desta forma de eco.

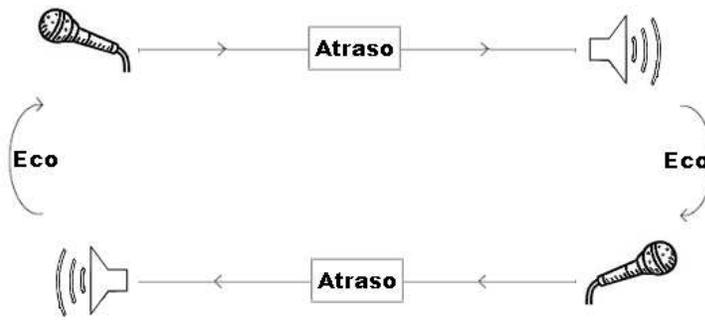


Figura 2.3: Diagrama ilustrativo do eco acústico.

## 2.4 Cancelamento de Eco

Atualmente o cancelamento de eco, seja eco elétrico ou acústico, baseia-se em sistemas adaptativos, que consistem em sintetizar uma cópia do eco e subtraí-la do sinal retornado [4, 5, 6], como pode ser visto nas Figuras 2.4 e 2.5. Na Figura 2.4, o sinal emitido do usuário 1 para o usuário 2 é usado como sinal de referência para o filtro adaptativo à direita, por meio do qual se produz uma estimativa do sinal de eco. A mesma idéia pode ser transportada para o caso de cancelamento de eco acústico, como mostrado na Figura 2.5.

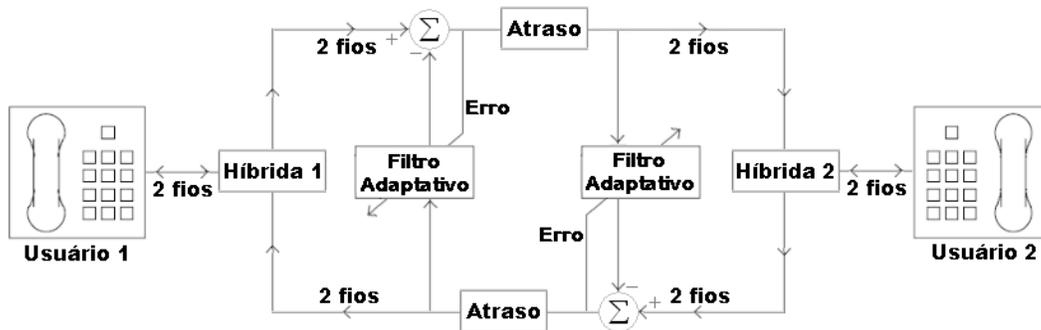


Figura 2.4: Cancelamento de eco elétrico por filtro adaptativo.

Em ambos os casos, o filtro adaptativo ajusta-se continuamente para acomodar as variações do caminho de eco. Em um caso ideal, após a convergência, a função de transferência do filtro adaptativo seria idêntica a do caminho percorrido pelo sinal de eco.

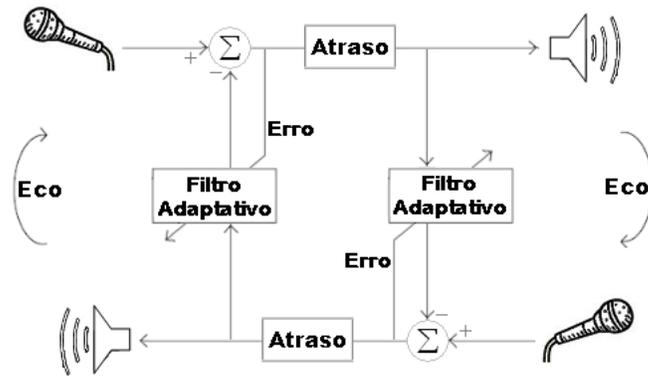


Figura 2.5: Cancelamento de eco acústico por filtro adaptativo.

## 2.5 Filtragem Adaptativa

A filtragem adaptativa constitui atualmente uma ferramenta imprescindível no processamento estatístico de sinais. Sempre que haja necessidade de processar resultados de ambientes com estatísticas desconhecidas ou variantes no tempo, a utilização de um filtro adaptativo representa uma solução atrativa quando confrontada com os resultados permitidos pela utilização dos métodos convencionais. Atualmente, esses tipos de sistemas são aplicados em campos bem diversos como o das comunicações, controle, radar, sonar e engenharia biomédica.

### 2.5.1 Filtro de Wiener

Considere o digrama de blocos da Figura 2.6. O filtro linear discreto é caracterizado pela resposta ao impulso  $w_0, w_1, w_2, \dots$ , e possui como entrada  $x(0), x(1), x(2), \dots$ . Em algum tempo discreto  $n$ , o filtro produz uma saída  $y(n)$ . Esta saída é utilizada para estimar um sinal desejado denotado por  $d(n)$ . Sendo a entrada do filtro e o sinal desejado realizações de processos estocásticos, a estimativa é acompanhada por um erro com características estatísticas próprias. Este erro é denotado por  $e(n)$  e definido como a diferença entre o sinal desejado  $d(n)$  e a saída do filtro  $y(n)$ . O objetivo é eliminar ou reduzir o erro seguindo algum critério estatístico.

O problema acima descrito possui uma solução simplificada assumindo a utilização de um filtro FIR (do inglês *Finite Impulse Response*) com  $N - 1$  atrasos. Sendo assim o vetor de coeficientes do filtro  $\mathbf{w}$  será da seguinte maneira:

$$\mathbf{w} = [w_0 \ w_1 \ w_2 \ \dots \ w_{N-1}]^T. \quad (2.1)$$

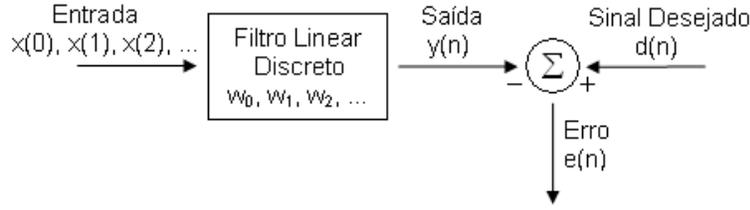


Figura 2.6: Representação em diagrama de blocos do filtro de Wiener.

Considerando o vetor entrada  $\mathbf{x}(n)$  formado pelo valor atual e valores passados

$$\mathbf{x}(n) = [x(n) \ x(n-1) \ x(n-2) \ \cdots \ x(n-N+1)]^T, \quad (2.2)$$

a saída do filtro será o produto interno entre vetor de entrada  $\mathbf{x}(n)$  e o vetor de coeficientes do filtro  $\mathbf{w}$ , dada por:

$$y(n) = \mathbf{x}^T(n)\mathbf{w} = \mathbf{w}^T\mathbf{x}(n). \quad (2.3)$$

O erro de estimativa, denotado por  $e(n)$ , é definido como a diferença entre o sinal desejado  $d(n)$  e a atual saída do filtro  $y(n)$

$$e(n) = d(n) - y(n), \quad (2.4)$$

onde substituindo  $y(n)$  pela equação (2.3), tem-se

$$e(n) = d(n) - \mathbf{x}^T(n)\mathbf{w} = d(n) - \mathbf{w}^T\mathbf{x}(n). \quad (2.5)$$

Define-se, então, a função custo como o valor médio quadrático do erro de estimativa

$$\begin{aligned} E[e^2(n)] &= E[d^2(n) - 2d(n)\mathbf{x}^T(n)\mathbf{w} + \mathbf{w}^T\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{w}] \\ &= E[d^2(n)] + \mathbf{w}^T E[\mathbf{x}(n)\mathbf{x}^T(n)]\mathbf{w} - 2E[d(n)\mathbf{x}^T(n)]\mathbf{w}. \end{aligned} \quad (2.6)$$

Seja  $\mathbf{R}$  a matriz de autocorrelação dos componentes do vetor de entrada

$$\mathbf{R} = E[\mathbf{x}(n)\mathbf{x}^T(n)], \quad (2.7)$$

na forma expandida tem-se:

$$\mathbf{R} = \begin{bmatrix} r_{xx}(0) & r_{xx}(1) & \cdots & r_{xx}(N-1) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}(N-1) & r_{xx}(N-2) & \cdots & r_{xx}(0) \end{bmatrix}. \quad (2.8)$$

De maneira análoga, define-se  $\mathbf{p}$  como o vetor de correlação cruzada entre as componentes do vetor de entrada  $\mathbf{x}(n)$  e o sinal desejado  $d(n)$

$$\mathbf{p} = E[d(n)\mathbf{x}(n)], \quad (2.9)$$

na forma expandida, tem-se

$$\mathbf{p} = [r_{dx}(0) \ r_{dx}(1) \ \dots \ r_{dx}(N-1)]^T. \quad (2.10)$$

Assim, o valor médio quadrático do erro de estimativa pode ser escrito como

$$E[e^2(n)] = E[d^2(n)] + \mathbf{w}^T \mathbf{R} \mathbf{w} - 2\mathbf{p}^T \mathbf{w}. \quad (2.11)$$

Esta função é representada por um parabolóide multidimensional com concavidade orientada no sentido positivo e, conseqüentemente, é caracterizada por um único mínimo. Para determinar o valor mínimo da função custo, calcula-se o seu gradiente

$$\nabla_{\mathbf{w}} E[e^2(n)] = \frac{\partial E[e^2(n)]}{\partial \mathbf{w}} = 2\mathbf{R} \mathbf{w} - 2\mathbf{p}. \quad (2.12)$$

O vetor de coeficientes ótimo  $\mathbf{w}_o$  será aquele para o qual o gradiente se anula, resultando em

$$\mathbf{w}_o = \mathbf{R}^{-1} \mathbf{p}, \quad (2.13)$$

assumindo que a matriz de autocorrelação  $\mathbf{R}$  é não-singular.

O filtro de Wiener determinado pela equação (2.13) é inadequado para situações em que não exista estacionariedade, tornando inviável o conhecimento *a priori* da estatística dos dados  $x(n)$  e  $d(n)$  a serem processados. Nestes casos, o filtro de Wiener poderá não ser realizável ou não ter desempenho ótimo. Uma aproximação para tal situação pode ser feita ocorrendo, primeiramente, uma estimativa da estatística do sinal a ser processado e, em seguida, sua aplicação no cálculo do filtro. Porém, este processo não atende as necessidades computacionais de aplicações em tempo real.

## 2.6 Algoritmos Adaptativos

Os filtros adaptativos são dispositivos que possuem comportamento satisfatório em ambientes onde não haja conhecimento *a priori* das estatísticas dos sinais de entrada. Tais filtros ajustam seus coeficientes de maneira automática baseando-se em algum algoritmo recursivo de otimização, com o intuito de minimizar em tempo real uma função custo. Começa-se com um conjunto de coeficientes iniciais que demonstra desconhecimento do ambiente. Se este for estacionário, os coeficientes do filtro convergem, a cada iteração, para a solução ótima do filtro de Wiener. Em um ambiente não estacionário, o algoritmo possibilita ao filtro acompanhar as variações estatísticas ao longo do tempo, desde que estas sejam suficientemente lentas.

Esta dissertação restringe-se a utilizar algoritmos adaptativos que possuam complexidade computacional suficientemente baixa devido às altas ordens de filtro que são necessárias no ambiente em estudo, de maneira que os algoritmos possam ser implementados em um sistema em tempo real. Apesar da possibilidade de utilização de algoritmos adaptativos em sub-bandas e wavelets, aqui optou-se por algoritmos no domínio do tempo.

### 2.6.1 Algoritmo *Least Mean Square*

O algoritmo LMS é um algoritmo baseado no método do gradiente descendente, que utiliza o valor instantâneo de  $e^2(n)$  como estimativa da função custo calculada por  $E[e^2(n)]$

$$\nabla_{\mathbf{w}} E[e^2(n)] \approx \nabla_{\mathbf{w}} e^2(n) = \frac{\partial e^2(n)}{\partial \mathbf{w}} = 2e(n) \frac{\partial e(n)}{\partial \mathbf{w}}. \quad (2.14)$$

Sendo  $e(n) = d(n) - y(n)$  e  $y(n) = \mathbf{w}^T \mathbf{x}(n)$ , então

$$\frac{\partial e(n)}{\partial \mathbf{w}} = \frac{\partial [d(n) - \mathbf{w}^T \mathbf{x}(n)]}{\partial \mathbf{w}} = -\mathbf{x}(n). \quad (2.15)$$

Logo,

$$\nabla_{\mathbf{w}} e^2(n) = -2e(n)\mathbf{x}(n). \quad (2.16)$$

A equação de atualização dos coeficientes de acordo com o método do gradiente descendente resulta em

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{x}(n), \quad (2.17)$$

onde a constante 2 foi incorporada ao fator  $\mu$  que controla o tamanho do passo do algoritmo, e que deve ser escolhido de maneira que garanta a convergência do algoritmo.

### 2.6.2 Algoritmo *Normalized LMS*

O algoritmo NLMS pode ser derivado por diferentes pontos de vista. A sua interpretação mais simples é como uma implementação especial do algoritmo LMS que leva em consideração as variações do nível do sinal de entrada do filtro. Isto é feito utilizando um tamanho do passo variável normalizado pela energia do sinal de entrada do filtro

$$\mu(n) = \frac{\mu}{\mathbf{x}^T(n)\mathbf{x}(n)}, \quad (2.18)$$

o que garante que o erro *a posteriori*

$$\epsilon(n) = d(n) - \mathbf{w}^T(n+1)\mathbf{x}(n) \quad (2.19)$$

seja minimizado em magnitude e chega-se à equação de atualização

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{\mathbf{x}^T(n)\mathbf{x}(n) + \gamma} e(n)\mathbf{x}(n), \quad (2.20)$$

onde um parâmetro de regularização  $\gamma$  é introduzido para evitar valores muito elevados para o passo de adaptação quando  $\mathbf{x}^T(n)\mathbf{x}(n)$  torna-se pequeno.

### 2.6.3 Algoritmo *Proportionate NLMS*

O algoritmo PNLMS aplica um tamanho de passo adaptativo e individual para cada coeficiente do filtro [7]. Os tamanhos dos passos são calculados por meio da última estimativa dos coeficientes do filtro de tal maneira que os coeficientes grandes recebam altos tamanhos de passo, o que faz com que os coeficientes ativos sejam ajustados mais rápidos que os coeficientes não ativos (pequenos ou zeros). O algoritmo PNLMS adapta os coeficientes como

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n) + \gamma} \mathbf{G}(n)\mathbf{x}(n)e(n), \quad (2.21)$$

onde

$$\mathbf{G}(n) = \text{diag} \{g_0(n)g_1(n) \cdots g_{N-1}(n)\}, \quad (2.22)$$

$$g_l(n) = \frac{\gamma_l(n)}{\frac{1}{N} \sum_{i=0}^{N-1} \gamma_i(n-1)}, l = 0, 1, \dots, N-1, \quad (2.23)$$

$$\gamma_l(n) = \max \{\gamma_{\min}, |w_l(n)|\}, l = 0, 1, \dots, N-1, \quad (2.24)$$

$$\gamma_{\min} = \varrho \max \{\delta, |w_0(n)|, \dots, |w_{N-1}(n)|\}. \quad (2.25)$$

Os parâmetros  $\delta$  e  $\varrho$  são números positivos com valores típicos  $\delta = 0.01$  e  $\varrho = \frac{5}{N}$  [8]. O termo  $\gamma_{\min}$  protege  $w_l(n)$  de não ser adaptado quando este é muito menor que o maior coeficiente e  $\delta$  regula a atualização quando todos os coeficientes são inicializados com zeros.

## 2.6.4 Algoritmo *Normalized Data-Reusing* LMS

Existem situações onde é possível reutilizar dados antigos para melhorar a convergência dos filtros adaptativos. Os algoritmos *data-reusing* são considerados uma alternativa para aumentar a velocidade de convergência dos algoritmos adaptativos em situações onde o sinal de entrada é correlacionado como, por exemplo, sinais de voz [6]. O preço a ser pago neste caso é o aumento do desajuste do algoritmo, onde a relação entre o aumento na velocidade de convergência e o aumento no desajuste pode ser controlada pelo tamanho do passo do algoritmo.

Sendo  $L$  o número de dados antigos a ser utilizado, o NDR-LMS atualiza os coeficientes do filtro como [9]

$$\mathbf{w}_{i+1}(n) = \mathbf{w}_i(n) + \frac{\mu}{\mathbf{x}^T(n-i)\mathbf{x}(n-i)} e_i(n) \mathbf{x}(n-i) \quad (2.26)$$

onde  $i = 0, 1, \dots, L$  e

$$\mathbf{w}_0(n) = \mathbf{w}_L(n-1) \quad (2.27)$$

$$e_i(n) = d(n) - \mathbf{w}_i^T(n) \mathbf{x}(n) \quad (2.28)$$

$$\mathbf{w}_0(n+1) = \mathbf{w}_L(n) \quad (2.29)$$

### 2.6.5 Algoritmo *Binormalized Data-Reusing* LMS

O BNDR-LMS faz reuso de um par de dado antigo, além de combinar projeções ortogonais de duas direções de gradientes consecutivos e normalização, para obter uma convergência mais rápida quando comparada aos outros algoritmos do tipo LMS [10]. Os coeficientes são atualizados como

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \left[ \frac{\lambda_1}{2} \mathbf{x}(n) + \frac{\lambda_2}{2} \mathbf{x}(n-1) \right], \quad (2.30)$$

onde  $\mu$  é o fator que controla o tamanho do passo e

$$\frac{\lambda_1}{2} = \frac{e(n) \|\mathbf{x}(n-1)\|^2 - \epsilon(n-1) \mathbf{x}^T(n-1) \mathbf{x}(n)}{\|\mathbf{x}(n)\|^2 \|\mathbf{x}(n-1)\|^2 - [\mathbf{x}^T(n) \mathbf{x}(n-1)]^2}, \quad (2.31)$$

$$\frac{\lambda_2}{2} = \frac{\epsilon(n-1) \|\mathbf{x}(n)\|^2 - e(n) \mathbf{x}^T(n-1) \mathbf{x}(n)}{\|\mathbf{x}(n)\|^2 \|\mathbf{x}(n-1)\|^2 - [\mathbf{x}^T(n) \mathbf{x}(n-1)]^2} \quad (2.32)$$

e  $\epsilon(n-1)$  é o erro *a posteriori* na iteração  $n-1$

$$\epsilon(n-1) = d(n-1) - \mathbf{w}^T(n) \mathbf{x}(n-1). \quad (2.33)$$

O algoritmo BNDR-LMS, assim como o algoritmo NLMS, faz parte da família do algoritmo *affine projection* (APA) [6], que generaliza a idéia de reutilização de  $L$  pares de dados para aumentar a convergência dos algoritmos adaptativos em situações em que o sinal de entrada é correlacionado. Enquanto o algoritmo NLMS é o caso particular do APA para  $L = 0$ , o algoritmo BNDR-LMS é uma forma fechada do APA para  $L = 1$ .

### 2.6.6 Algoritmo *Proportionate* BNDR-LMS

Utilizando o mesmo princípio apresentado na Subseção 2.6.3, o algoritmo P-BNDR-LMS [11] junta as idéias de reutilização de dados antigos e de tamanhos de passo individuais para cada coeficiente do filtro adaptativo. Assim, o P-BNDR-LMS atualiza os coeficientes do filtro como

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \mathbf{G}(n) \left[ \frac{\lambda_1}{2} \mathbf{x}(n) + \frac{\lambda_2}{2} \mathbf{x}(n-1) \right], \quad (2.34)$$

onde  $\mu$  é o fator que controla o tamanho do passo,  $\mathbf{G}(n)$  é definido na equação (2.22),

$$\frac{\lambda_1}{2} = \frac{e(n)\mathbf{x}^T(n-1)\mathbf{G}(n)\mathbf{x}(n-1) - \epsilon(n-1)\mathbf{x}^T(n-1)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)\mathbf{x}^T(n-1)\mathbf{G}(n)\mathbf{x}(n-1) - [\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n-1)]^2} \quad (2.35)$$

e

$$\frac{\lambda_2}{2} = \frac{\epsilon(n-1)\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n) - e(n)\mathbf{x}^T(n-1)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)\mathbf{x}^T(n-1)\mathbf{G}(n)\mathbf{x}(n-1) - [\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n-1)]^2}. \quad (2.36)$$

## 2.6.7 Complexidade Computacional

Uma importante característica dos algoritmos adaptativos, juntamente com a velocidade de convergência, é a complexidade computacional. Define-se por complexidade computacional a quantidade de operações aritméticas que o algoritmo realiza a cada iteração. Esta característica ganha ainda mais importância quando se trata de aplicações em tempo real.

Idealmente, deseja-se que o algoritmo tenha uma rápida velocidade de convergência e uma baixa complexidade computacional. Entretanto, estes dois requisitos costumam ser contraditórios. Geralmente, para acelerar a convergência dos algoritmos adaptativos é necessário aumentar a sua complexidade computacional. Esta relação velocidade de convergência *versus* complexidade computacional deve ser analisada. A Tabela 2.1 mostra a complexidade por atualização dos algoritmos adaptativos derivados na Seção 2.6.

Tabela 2.1: Complexidade computacional convencional por atualização dos algoritmos adaptativos.

Algoritmo	Multiplicações	Adições	Divisões	Subtrações
LMS	$N + 1$	$N$	0	0
NLMS	$2N + 1$	$2N$	1	0
PNLMS	$3N + 1$	$3N$	$N + 2$	0
NDR-LMS	$5N + 2$	$5N$	2	1
BNDR-LMS	$6N + 8$	$5N$	1	4
P-BNDR-LMS	$8N + 9$	$7N$	$N + 2$	4

Entretanto, os algoritmos que envolvem o cálculo do produto interno dos sinais de entrada do filtro podem ter sua complexidade reduzida efetuando esta medida de maneira recursiva. Sendo a entrada do filtro adaptativo na interação  $n$

$$\mathbf{x}(n) = [x(n) \ x(n-1) \ x(n-2) \ \cdots \ x(n-N+1)]^T, \quad (2.37)$$

os produtos internos  $\mathbf{x}^T(n)\mathbf{x}(n)$  e  $\mathbf{x}^T(n)\mathbf{x}(n-1)$  podem ser calculados recursivamente como

$$\mathbf{x}^T(n)\mathbf{x}(n) = \|\mathbf{x}(n)\|^2 = \|\mathbf{x}(n-1)\|^2 + x^2(n) - x^2(n-N) \quad (2.38)$$

e

$$\begin{aligned} \mathbf{x}^T(n)\mathbf{x}(n-1) &= \mathbf{x}^T(n-1)\mathbf{x}(n-2) + x(n)x(n-1) \\ &\quad - x(n-N)x(n-N-1), \end{aligned} \quad (2.39)$$

onde  $\|\mathbf{x}(n-1)\|^2$  e  $\mathbf{x}^T(n-1)\mathbf{x}(n-2)$  estão disponíveis pois foram calculados na iteração passada. Assim, operações que formalmente exigiriam  $N$  multiplicações e  $N$  somas têm suas complexidades computacionais reduzidas para 2 multiplicações, 1 soma e 1 subtração.

Esta recursão não pode ser utilizada no caso de algoritmos do tipo *proportionate* devido à inclusão da matriz  $\mathbf{G}$ . Como ela é atualizada a cada iteração, o valor atual de  $\mathbf{x}^T\mathbf{G}\mathbf{x}$  não possui nenhuma ligação direta com o valor da iteração passada.

Porém, uma versão “rápida” do algoritmo P-BNDR-LMS pode ser derivada desconsiderando a utilização da matriz  $\mathbf{G}$  na normalização do algoritmo, permitindo, assim, a utilização do cálculo recursivo para os cálculos dos produtos internos dos sinais de entrada do filtro adaptativo. Além disso, o cálculo do erro *a posteriori*

$$\epsilon(n-1) = d(n-1) - \mathbf{w}^T(n)\mathbf{x}(n-1) \quad (2.40)$$

pode ser aproximado por [12]

$$\epsilon(n-1) \approx (1 - \mu)e(n-1), \quad (2.41)$$

substituindo uma operação de  $N$  multiplicações,  $N$  somas e 1 subtração por apenas 1 multiplicação e 1 subtração. O algoritmo BNDR-LMS também faz uso desta aproximação. A Tabela 2.2 mostra a complexidade computacional prática por atualização dos algoritmos adaptativos, enquanto a Tabela 2.6.7 apresenta o pseudocódigo da versão rápida do P-BNDR-LMS.

Tabela 2.2: Complexidade computacional prática por atualização dos algoritmos adaptativos.

Algoritmo	Multiplicações	Adições	Divisões	Subtrações
LMS	$N + 1$	$N$	0	0
NLMS	$N + 3$	$N + 1$	1	1
PNLMS	$3N + 1$	$3N$	$N + 2$	0
NDR-LMS	$3N + 4$	$3N + 1$	2	2
BNDR-LMS	$3N + 12$	$3N + 2$	1	6
P-BNDR-LMS	$4N + 14$	$3N + 2$	$N + 2$	6

Tabela 2.3: Algoritmo P-BNDR-LMS Rápido

---


$$e(n) = d(n) - \mathbf{w}^T(n)\mathbf{x}(n)$$

$$\epsilon(n-1) = (1 - \mu)e(n-1)$$

$$r_{22}(n) = r_{11}(n-1)$$

$$r_{11}(n) = r_{11}(n-1) + x^2(n) - x^2(n-N)$$

$$r_{12}(n) = r_{12}(n-1) + x(n)x(n-1) - x(n-N)x(n-N-1)$$

$$den = r_{11}(n)r_{22}(n) - r_{12}^2(n) + \gamma$$

$$\frac{\lambda_1}{2} = e(n)r_{22} - \epsilon(n-1)r_{12}$$

$$\frac{\lambda_2}{2} = \epsilon(n-1)r_{11} - e(n)r_{12}$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{den}\mathbf{G}(n) \left[ \frac{\lambda_1}{2}\mathbf{x}(n) + \frac{\lambda_2}{2}\mathbf{x}(n-1) \right]$$


---

## 2.7 Conclusões

Este capítulo abordou a existência de eco em sistemas de comunicação, diferenciando as suas naturezas elétrica e acústica, e a necessidade de dispositivos que atenuem este tipo de perturbação. Explicou-se a necessidade de dispositivos mais eficientes que os supressores de eco, devido ao seu não funcionamento em situações de conversação cruzada, quando os locutores falam ao mesmo tempo.

Como solução para o problema de cancelamento de eco faz-se uso de filtros adaptativos, dispositivos que possuem comportamento satisfatório em ambientes onde não haja conhecimento *a priori* das estatísticas dos sinais de entrada. Objetiva-se sintetizar uma cópia do eco e subtraí-la do sinal contaminado adequadamente, que será o sinal efetivamente transmitido.

Discutiu-se os princípios dos filtros adaptativos direcionando à derivação dos algoritmos adaptativos que serão abordados. Comparou-se a complexidade computacional dos algoritmos adaptativos, uma importante característica para implementações em tempo real, focando-se em soluções práticas para redução das mesmas.

# Capítulo 3

## Cancelamento Robusto de Eco

### 3.1 Introdução

Cancelamento de eco, como visto anteriormente, baseia-se em filtros adaptativos que consistem em sintetizar uma cópia do eco e subtraí-la do sinal a ser transmitido. Tais filtros ajustam seus coeficientes com o intuito de minimizar a energia da componente de eco. Porém, se o sinal a ser transmitido conter, além do eco, algum outro sinal energeticamente relevante proveniente do ambiente externo, este último agirá como um ruído descorrelacionado para o algoritmo adaptativo, prejudicando o processo de estimação dos seus coeficientes e fazendo com que o algoritmo possa divergir ou permitir uma passagem excessiva de eco.

A Seção 3.2 exemplifica o problema da influência externa no processo de estimação do filtro adaptativo em um ambiente de cancelamento de eco acústico mono-canal, o qual é matematicamente detalhado. Como uma primeira solução para proteger o filtro destas influências externas, a Seção 3.3 apresenta o conceito de algoritmos de detecção de conversação cruzada e os principais algoritmos existentes na literatura.

Entretanto, estes algoritmos são passíveis de falhas que são suficientes para influenciarem o comportamento dos algoritmos adaptativos causando divergência. Por isso, a Seção 3.4 descreve o conceito de algoritmos adaptativos robustos a distúrbios de curta duração, onde uma atual derivação encontrada na literatura é apresentada e é proposta uma alteração para uma segunda derivação. Por fim, a Seção 3.5 resume o capítulo enfatizando as principais contribuições.

## 3.2 Descrição do Problema

Em um sistema de teleconferência, a fala do locutor distante  $x(n)$ , ao ser reproduzida pelas caixas de som, pode retornar ao microfone percorrendo diversos percursos. Entre estes percursos, estão incluídos o caminho direto assim como infinitas reflexões, sendo que, em todos os casos, ocorre uma atenuação do sinal. Em geral, esta atenuação se torna mais brusca conforme o aumento do percurso, fazendo com que componentes de grande atraso possuam baixa energia. Assim, pode-se simplificar o problema em questão, considerando apenas a existência de um número finito de reflexões.

Com isso, o eco captado pelo microfone pode ser modelado como o resultado da passagem do sinal de fala do locutor distante  $x(n)$  por um filtro FIR de resposta ao impulso

$$\mathbf{h} = [h_0 \ h_1 \ \dots \ h_{N-1}]^T. \quad (3.1)$$

Então, o eco acústico é somado à fala do locutor próximo  $v(n)$  gerando o sinal contaminado

$$d(n) = \mathbf{h}^T \mathbf{x}(n) + v(n), \quad (3.2)$$

que é transmitido ao locutor distante, onde

$$\mathbf{x}(n) = [x(n) \ x(n-1) \ \dots \ x(n-N+1)]^T. \quad (3.3)$$

Atualmente, o cancelamento de eco acústico baseia-se em sistemas adaptativos como representado na Figura 3.1 [4, 5, 6]. Nestes sistemas, modela-se o percurso do eco  $\mathbf{h}$  por meio do filtro adaptativo  $\mathbf{w}$ , obtendo-se assim uma estimativa do eco

$$y(n) = \mathbf{w}^T \mathbf{x}(n), \quad (3.4)$$

onde

$$\mathbf{w} = [w_0 \ w_1 \ \dots \ w_{N-1}]^T. \quad (3.5)$$

Esta estimativa do eco é subtraída do sinal contaminado, gerando-se o sinal de erro

$$\begin{aligned} e(n) &= d(n) - y(n) \\ &= (\mathbf{h}^T - \mathbf{w}^T) \mathbf{x}(n) + v(n), \end{aligned} \quad (3.6)$$

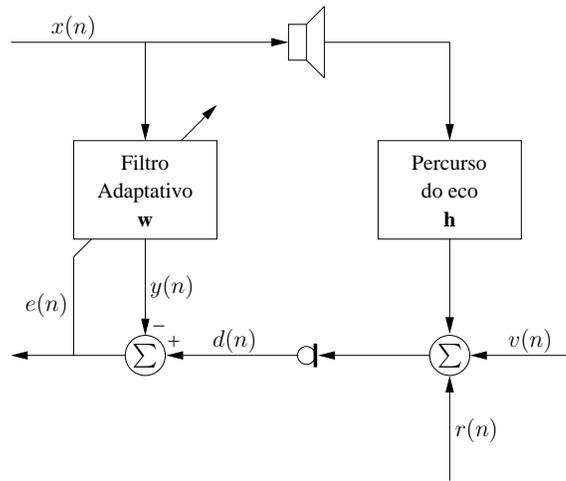


Figura 3.1: Cancelamento de eco acústico.

que é efetivamente o sinal transmitido ao locutor distante.

No caso ideal,  $\mathbf{w} = \mathbf{h}$  e conseqüentemente  $e(n) = v(n)$ , sendo todo o sinal de eco cancelado. Entretanto, a aproximação de  $\mathbf{h}$  por um filtro  $\mathbf{w}$  de comprimento e precisão finitos impede que o percurso de eco seja perfeitamente modelado.

Na prática, se durante a adaptação dos coeficientes  $\mathbf{w}$  do filtro a fala do locutor próximo  $v(n)$  for zero e o ruído do locutor próximo  $r(n)$  for insignificante, o filtro adaptativo  $\mathbf{w}$  consegue convergir para uma boa estimativa do percurso do eco  $\mathbf{h}$ , cancelando com sucesso o eco.

Entretanto, quando  $v(n)$  e  $x(n)$  são simultaneamente diferentes de zero, caracterizando uma situação de conversação cruzada (DT, do inglês *double talk*), a fala do locutor próximo  $v(n)$  age como um ruído descorrelacionado para o algoritmo adaptativo ao aumentar bruscamente a amplitude do sinal desejado  $d(n)$  e do erro  $e(n)$ . Isto pode fazer com que o filtro adaptativo divirja ou permita uma passagem excessiva de eco não-cancelado até o locutor distante. A solução comum para este problema é diminuir ou parar completamente a adaptação do filtro quando a presença da fala do locutor próximo for detectada. Esta é a regra do detector de conversação cruzada (DTD, do inglês *double talk detector*).

Entretanto, na prática, a detecção de conversação cruzada não ocorre perfeitamente. Os algoritmos de detecção de DT costumam apresentar alguns tipos de falhas, como demorar algum tempo para detectar uma situação de DT ou mesmo não detectá-las completamente. Apesar destes problemas não serem temporalmente

longos, são suficientes para influenciarem o comportamento do algoritmo adaptativo causando divergência.

Outro problema possível de ser encontrado é a falsa detecção de DT quando ocorre uma variação do percurso do eco, principalmente no caso de eco acústico. Este é um caso em que o filtro adaptativo precisa corrigir as mudanças no percurso do eco, sendo desejado que a adaptação não seja desnecessariamente parada por uma falsa detecção de DT. Além disso, o ruído do locutor próximo  $r(n)$  não deve ser detectado como DT.

Diante destes problemas, faz-se necessário o desenvolvimento de dois tipos de algoritmos adaptativos robustos a distúrbios de curta duração devido às falhas de detecção de DT. A intenção é fazer com que os algoritmos sejam poucos sensíveis aos erros de detecção de DT e, conseqüentemente, não permitam uma passagem excessiva de eco.

### 3.3 Detector de Conversação Cruzada

A base para a maioria dos algoritmos de detecção de DT envolve o cálculo de uma variável de detecção através dos dados disponíveis, como a fala do locutor próximo, fala do locutor distante e o sinal de erro, e a comparação da variável de detecção com um limiar constante. Dependendo se a variável de detecção estiver acima ou abaixo do limiar, uma situação de DT é declarada ou não.

Existem vários algoritmos diferentes na literatura sobre como calcular uma variável de detecção para um DTD. Aqui, daremos importância a alguns algoritmos conhecidos, os quais são baseados em comparação de energia e correlação cruzada.

#### 3.3.1 Algoritmo de Geigel

O algoritmo de Geigel compara a amplitude do sinal recebido  $d(n)$  com a máxima amplitude das  $N$  amostras mais recentes da fala do locutor distante  $x(n)$ , onde  $N$  é o comprimento do filtro adaptativo. O percurso do eco geralmente atenua a fala do locutor distante  $x(n)$ , fazendo com que a amplitude do sinal recebido  $d(n)$  contendo apenas o eco seja menor que a do sinal recebido  $d(n)$  contendo tanto o eco quanto a fala do locutor próximo  $v(n)$ . O algoritmo de Geigel calcula a sua variável

de detecção  $\xi_1$  como

$$\xi_1 = \frac{|d(n)|}{\max \{|x(n)|, \dots, |x(n - N + 1)|\}}. \quad (3.7)$$

Se  $\xi_1$  for maior que o limiar  $T$ , DT é declarado. Caso contrário, não é. A escolha de  $T$  precisa ser feita com cuidado pois afetará drasticamente o desempenho do DTD. Uma constante de tempo  $t_{hold}$  denominada *hangover time* é especificada de maneira que se DT é declarado, a adaptação dos coeficientes do filtro adaptativo é interrompida por este período de tempo além do fim da situação da DT.

### 3.3.2 Algoritmos baseados em correlação cruzada

O DTD baseado em correlação cruzada entre  $x(n)$  e  $e(n)$  proposto em [13] faz uso do princípio da ortogonalidade. Quando o filtro adaptativo converge para a sua solução ótima, o princípio da ortogonalidade entre a fala do locutor distante  $x(n)$  e o erro  $e(n)$  é satisfeito. Numa situação de DT, o sinal recebido  $d(n)$  e o erro ficam maiores devido à presença da fala do locutor próximo  $v(n)$ . Entretanto, como  $v(n)$  é descorrelacionado com  $x(n)$ , o princípio da ortogonalidade continua valendo. Por outro lado, quando ocorre uma variação no percurso do eco, o princípio da ortogonalidade pode não ser satisfeito. Dado o vetor de correlação cruzada entre  $\mathbf{x}(n)$  e  $e(n)$

$$\mathbf{c}_{xe} = [c_{xe,0} c_{xe,1} \dots c_{xe,N-1}]^T, \quad (3.8)$$

onde

$$c_{xe,i} = \frac{E \{x(n-i)e(n)\}}{\sqrt{E \{x^2(n-i)\} E \{e^2(n)\}}}, \quad (3.9)$$

a variável de detecção é definida como

$$\xi_2 = \frac{1}{N} \sum_{i=0}^{N-1} |c_{xe,i}|. \quad (3.10)$$

Quando  $\xi_2 \leq T$ , o filtro adaptativo terá convergido. Caso contrário, não terá convergido ou o percurso do eco terá se modificado. Em geral, este tipo de algoritmo não detecta uma condição de DT explicitamente. Ao invés, ele decide quando o filtro adaptativo convergiu ou não. Quando o filtro tiver convergido, a adaptação será parada para proteger o filtro de interferências causadas por DT. Se

o filtro adaptativo não tiver convergido ou o percurso do eco tiver sido alterado, o filtro continuará se adaptando.

Entretanto, percebeu-se que a mesma idéia utilizando vetor de correlação cruzada entre  $\mathbf{x}(n)$  e  $d(n)$  apresentava um resultado mais robusto e confiável [14]. Assim sendo, o vetor de correlação cruzada é definido como

$$\mathbf{c}_{xd} = [c_{xd,0} c_{xd,1} \cdots c_{xd,N-1}]^T, \quad (3.11)$$

onde

$$c_{xd,i} = \frac{E \{x(n-i)d(n)\}}{\sqrt{E \{x^2(n-i)\} E \{d^2(n)\}}}. \quad (3.12)$$

Neste caso, a variável de detecção é definida como

$$\xi_3 = \|\mathbf{c}_{xd}\|_\infty = \max_i |c_{xd,i}|, \quad i = 0, 1, \dots, N-1. \quad (3.13)$$

Mais uma vez, a idéia é comparar  $\xi_3$  com um limiar  $T$ . Se  $\xi_3 \geq T$ , então não há DT; se  $\xi_3 < T$ , então há DT.

Porém, nenhum desses vetores de correlação cruzada é bem normalizado. No geral, apenas podemos dizer que  $\xi_i \leq 1$ , não podendo inferir nada sobre o seu valor quando  $v = 0$ . A quantidade de correlação irá depender da estatística dos sinais e do percurso do eco. Como conseqüência, o melhor valor de  $T$  poderá variar muito de um experimento para outro.

Sendo assim, a referência [14] propôs um vetor de correlação cruzada normalizado, onde  $\xi_4 = 1$  quando  $v = 0$ . Supondo  $v = 0$

$$\sigma_d^2 = \mathbf{h}^T \mathbf{R}_{xx} \mathbf{h}, \quad (3.14)$$

onde  $\mathbf{R}_{xx} = E[\mathbf{x}(n)\mathbf{x}^T(n)]$ . Sendo  $d(n) = \mathbf{h}^T \mathbf{x}(n)$ , tem-se que

$$\mathbf{r}_{xd} = \mathbf{R}_{xx} \mathbf{h} \quad (3.15)$$

e a equação (3.14) pode ser re-escrita como

$$\sigma_d^2 = \mathbf{r}_{xd}^T \mathbf{R}_{xx}^{-1} \mathbf{r}_{xd}. \quad (3.16)$$

Em geral, para  $v \neq 0$

$$\sigma_d^2 = \mathbf{r}_{xd}^T \mathbf{R}_{xx}^{-1} \mathbf{r}_{xd} + \sigma_v^2. \quad (3.17)$$

Dividindo a equação (3.16) por  $\sigma_v^2$  e tirando-se a sua raiz quadrada, obtém-se a nova variável de decisão

$$\xi_4 = \sqrt{\mathbf{r}_{xd}^T (\sigma_v^2 \mathbf{R}_{xx})^{-1} \mathbf{r}_{xd}} \quad (3.18)$$

$$= \|\mathbf{c}_{xd}\|_2, \quad (3.19)$$

onde

$$\mathbf{c}_{xd} = (\sigma_v^2 \mathbf{R}_{xx})^{-\frac{1}{2}} \mathbf{r}_{xd} \quad (3.20)$$

é chamado de vetor de correlação normalizado entre  $\mathbf{x}$  e  $d$ .

Substituindo as equações (3.15) e (3.17) em (3.18), tem-se que a variável de decisão é igual a

$$\xi_4 = \frac{\sqrt{\mathbf{h}^T \mathbf{R}_{xx} \mathbf{h}}}{\sqrt{\mathbf{h}^T \mathbf{R}_{xx} \mathbf{h} + \sigma_v^2}}. \quad (3.21)$$

Sendo assim, deduz-se da equação (3.21) que para  $v = 0$ ,  $\xi_4 = 1$  e para  $v \neq 0$ ,  $\xi_4 < 1$ . Note também que  $\xi_4$  é insensível a variações no percurso do eco quando  $v = 0$ .

O vetor de correlação normalizado também pode ser obtido utilizando, ao invés da correlação cruzada entre o sinal de voz do locutor distante  $x(n)$  e a saída do filtro adaptativo  $e(n)$ , a correlação cruzada entre o sinal capturado pelo microfone  $d(n)$  e a saída do filtro  $e(n)$  [15]

$$\begin{aligned} r_{ed} &= E[e(n)d(n)] \\ &= E[(v(n) + \mathbf{h}^T \mathbf{x}(n) - \mathbf{w}^T \mathbf{x}(n)) (v(n) + \mathbf{x}^T(n) \mathbf{h})] \\ &= E[(\mathbf{h}^T \mathbf{x}(n) - \mathbf{w}^T \mathbf{x}(n)) \mathbf{x}^T(n) \mathbf{h} + v^2(n)] \\ &= (\mathbf{h}^T - \mathbf{w}^T) \mathbf{R}_{xx} \mathbf{h} + \sigma_v^2. \end{aligned} \quad (3.22)$$

Neste caso, a variável de detecção de conversação cruzada é definida como [15]

$$\xi_5 = 1 - \frac{r_{ed}}{\sigma_d^2}. \quad (3.23)$$

Substituindo as equações (3.14) e (3.22) em (3.23), obtém-se

$$\xi_5 = 1 - \frac{(\mathbf{h}^T - \mathbf{w}^T) \mathbf{R}_{xx} \mathbf{w} + \sigma_v^2}{\mathbf{h}^T \mathbf{R}_{xx} \mathbf{h} + \sigma_v^2} \quad (3.24)$$

$$\xi_5 = \frac{\mathbf{w}^T \mathbf{R}_{xx} \mathbf{h}}{\mathbf{h}^T \mathbf{R}_{xx} \mathbf{h} + \sigma_v^2} \quad (3.25)$$

de onde nota-se que, para  $v = 0$ ,  $\xi_5 \approx 1$  e para  $v \neq 0$ ,  $\xi_5 > 0$ , satisfazendo as condições necessárias de um detector de conversação cruzada ótimo.

Percebe-se que as equações (3.21) e (3.24), além da raiz quadrada, diferem apenas no numerador, onde  $\xi_4$  utiliza a resposta ao impulso do percurso de eco  $\mathbf{h}$  e  $\xi_5$  utiliza os coeficientes do filtro adaptativo  $\mathbf{w}$ . Entretanto, na implementação prática de  $\xi_4$  substitui-se  $\mathbf{h}$  por  $\mathbf{w}$  por não haver conhecimento *a priori* da resposta ao impulso do percurso do eco. Além disso, as estatísticas reais dos sinais não estão disponíveis na prática. Como consequência, a variável de decisão é calculada por meio de estimativas destas variáveis, que no caso de  $\xi_5$  resultam em [15]

$$r_{ed}(n) = \lambda_{dtd} r_{ed}(n-1) + (1 - \lambda_{dtd}) e(n)d(n) \quad (3.26)$$

$$\sigma_d^2(n) = \lambda_{dtd} \sigma_d^2(n-1) + (1 - \lambda_{dtd}) d(n)d(n), \quad (3.27)$$

onde  $e(n)$  é a saída do cancelador de eco e  $d(n)$  é o sinal capturado pelo microfone no instante de tempo  $n$ , e  $\lambda_{dtd}$  é um fator de esquecimento menor que 1.

### 3.4 Algoritmos Robustos

Geralmente, no contexto de identificação de sistemas, o sinal de erro  $e(n)$  é considerado ter média zero, branco e gaussiano. Apesar destas suposições serem muito convenientes e extremamente úteis na prática, algoritmos adaptativos otimizados nestas condições são muito sensíveis a variações destas suposições.

Um exemplo é o caso de cancelamento de eco combinado com um DTD. Algumas vezes, o DTD falha ao detectar o início e o fim de uma situação de DT e como resultado, um alto nível de sinal de voz na saída do cancelador de eco distorce o processo de adaptação dos coeficientes do filtro. A taxa de ocorrência destes distúrbios de alta intensidade depende da eficiência do DTD e da intensidade de DT. Desejando-se que o algoritmo seja insensível a estes erros de detecção, robustez contra erros de detecção de DT devem ser incorporados à estimação do algoritmo. Por robustez, entende-se insensibilidade a pequenos distúrbios.

Os algoritmos adaptativos apresentados podem ser feitos robustos a distúrbios de curta duração modificando a função custo as quais são baseados. Entretanto, em geral, estas modificações diminuem a velocidade de convergência. O grande problema de cancelamento de eco robusto é desenvolver um algoritmo que não divirja ou divirja lentamente em situações de conversação cruzada, e que seja capaz de rastrear rapidamente variações no percurso de eco quando elas ocorrerem. Estes dois requerimentos são contraditórios. A solução para este problema é a combinação de um DTD com métodos estatísticos robustos utilizando uma variável de escalamento adaptativa  $s$ .

### 3.4.1 Algoritmos Robustos Tipo 1

Algoritmos robustos podem ser obtidos alterando a função custo para

$$J = E \left\{ \rho \left[ \frac{|e(n)|}{s} \right] \right\}, \quad (3.28)$$

onde  $s$  é um fator de escalamento e  $\rho(z)$  é uma função com derivada monotonicamente não-decrescente que deve ser escolhida apropriadamente de maneira que

$\lim_{|e(n)| \rightarrow \infty} \left| \nabla_{\mathbf{w}} \rho \left[ \frac{|e(n)|}{s} \right] \right| < \infty$  [16]. Em [8] propõe-se

$$\rho(z) = \begin{cases} \frac{|z|^2}{2}, & \text{se } |z| \leq k_0 \\ k_0 |z| - \frac{k_0^2}{2}, & \text{se } |z| > k_0 \end{cases} \quad (3.29)$$

a qual é denominada de função Huber e  $k_0$  é uma constante arbitrária.

O novo algoritmo NLMS pode ser obtido pelo método de Newton

$$w(n+1) = w(n) - \mu [\nabla_{\mathbf{w}}^2 J]^{-1} \nabla_{\mathbf{w}} J, \quad (3.30)$$

onde, devido à escolha efetuada na equação (3.29),

$$\nabla_{\mathbf{w}} J = E \left\{ -\text{sign}[e(n)] \psi \left[ \frac{|e(n)|}{s} \right] \frac{\mathbf{x}(n)}{s} \right\} \quad (3.31)$$

e  $\psi(|z|) = \frac{d\rho(|z|)}{d|z|}$  é um decisor

$$\psi \left[ \frac{|e(n)|}{s} \right] = \min \left\{ \frac{|e(n)|}{s}, k_0 \right\}. \quad (3.32)$$

Assumindo-se que o vetor  $\mathbf{x}(n)$  é independente do erro  $e(n)$ , o que é verdade até a convergência, a matriz hessiana é calculada como

$$\begin{aligned}
\nabla_{\mathbf{w}}^2 J &= \nabla_{\mathbf{w}} \left\{ E \left[ -\text{sign}[e(n)] \psi \left[ \frac{|e(n)|}{s} \right] \frac{\mathbf{x}(n)}{s} \right] \right\}^T \\
&= E \left\{ - \left[ \nabla_{\mathbf{w}} \{ \text{sign}[e(n)] \} \psi \left[ \frac{|e(n)|}{s} \right] \right. \right. \\
&\quad \left. \left. + \text{sign}[e(n)] \nabla_{\mathbf{w}} \left\{ \psi \left[ \frac{|e(n)|}{s} \right] \right\} \right] \frac{\mathbf{x}(n)}{s} \right\} \\
&= E \left\{ - \left[ -2\delta[e(n)] \psi \left[ \frac{|e(n)|}{s} \right] s \right. \right. \\
&\quad \left. \left. - \text{sign}^2[e(n)] \psi' \left[ \frac{|e(n)|}{s} \right] \frac{\mathbf{x}(n)}{s} \right] \frac{\mathbf{x}^T(n)}{s} \right\}, \tag{3.33}
\end{aligned}$$

onde  $\delta(\cdot)$  é o delta de Dirac. Desconsiderando o caso não provável  $e(n) = 0$ ,

$$\begin{aligned}
\nabla_{\mathbf{w}}^2 J &= E \left\{ - \left[ \mathbf{0}^T - \text{sign}^2[e(n)] \psi' \left[ \frac{|e(n)|}{s} \right] \frac{\mathbf{x}(n)}{s} \right] \frac{\mathbf{x}^T(n)}{s} \right\} \\
&= E \left\{ \psi' \left[ \frac{|e(n)|}{s} \right] \frac{\mathbf{x}(n) \mathbf{x}^T(n)}{s^2} \right\} \\
&= \frac{c}{s^2} \mathbf{R}_{xx}, \tag{3.34}
\end{aligned}$$

o qual, objetivando reduzir a complexidade computacional, é aproximado como uma matriz diagonal

$$\nabla_{\mathbf{w}}^2 J \approx c \frac{\mathbf{x}^T(n) \mathbf{x}(n)}{s^2} \mathbf{I}. \tag{3.35}$$

Com isso, o novo algoritmo NLMS atualiza os coeficientes do filtro segundo

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu \mathbf{x}(n)}{\mathbf{x}^T(n) \mathbf{x}(n) + \gamma} \psi \left[ \frac{|e(n)|}{s} \right] \text{sign}[e(n)] s, \tag{3.36}$$

onde a constante  $c$  está incluída no tamanho do passo  $\mu$ .

O fator de escalamento  $s$  deve ser robusto a distúrbios de curta duração, devido à conversação cruzada, e ser capaz de perseguir mudanças de longa duração no sinal de erro devido à mudanças do percurso de eco. Para satisfazer este requisito, o fator de escalamento é definido como [8]

$$s(n+1) = \lambda_s s(n) + \frac{1 - \lambda_s}{\beta} \psi \left[ \frac{|e(n)|}{s(n)} \right] s(n), \tag{3.37}$$

onde  $\beta$  é uma constante de normalização escolhida de modo que para um processo gaussiano  $z$  de variância unitária e média zero

$$E[\psi(z) - \beta] = 0, \quad (3.38)$$

resultando em

$$\beta = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \psi(z) e^{-\frac{1}{2}z^2} dz \quad (3.39)$$

$$= \sqrt{\frac{2}{\pi}} \left( 1 - e^{-\frac{1}{2}k_0^2} \right) + k_0 \operatorname{erfc} \left( \frac{k_0}{\sqrt{2}} \right), \quad (3.40)$$

onde

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt. \quad (3.41)$$

Assim, a atual estimativa  $s(n)$  é governada aproximadamente pelas últimas  $\frac{1}{1-\lambda_s}$  amostras de  $e(n)$ , onde  $\lambda_s$  é uma constante de fator de esquecimento ligeiramente menor do que 1. Logo, quando o algoritmo ainda não tiver convergido, estando o nível do sinal de erro alto,  $s(n)$  será grande. Quando o algoritmo tiver convergido, estando o nível do sinal de erro baixo,  $s(n)$  será pequena. Então, a variável de escalamento  $s(n)$  terá seu comportamento ligado ao nível de convergência do algoritmo adaptativo por meio do sinal de erro e deve ser inicializada com um alto valor.

Quando o algoritmo ainda não tiver convergido e, conseqüentemente,  $s(n)$  estiver grande, o decisor estará na sua parte linear e o algoritmo robusto se comportará como o algoritmo NLMS convencional. Quando o decisor determinar que o algoritmo convergiu, o que dependerá do valor de  $e(n)$  e de  $s(n)$ , o algoritmo se comportará de maneira diferente do NLMS convencional, o que chamaremos aqui de modo robusto, diminuindo o passo a ser dado na direção da adaptação. Neste caso, o sentido da adaptação é determinado por  $\operatorname{sign}[e(n)]$ . Uma das conseqüências da utilização deste método é a diminuição do desajuste do algoritmo adaptativo, como será demonstrado posteriormente.

Se uma alteração repentina no percurso de eco ocorrer, o algoritmo não a perseguirá imediatamente. Entretanto, como o fator de escalamento  $s(n)$  começará a aumentar devido ao aumento do nível do sinal de erro, o decisor voltará a cair

na sua parte linear aumentando, assim, a velocidade de convergência do algoritmo robusto.

No caso de ocorrer uma falha da detecção de uma situação de DT, apesar do nível do sinal de erro aumentar bruscamente, como a variável de escalamento  $s(n)$  acompanhará o aumento no nível do sinal do erro lentamente devido ao seu modo de atualização, fazendo com que o algoritmo continue por mais tempo em modo robusto. Idealmente, isto dará tempo para o DTD agir e diminuirá a taxa de divergência em intervalos de DT menores que  $\frac{1}{1-\lambda_s}$  amostras.

Assim como nos DTD, é interessante incluir um *hangover time* para controle do fator de escalamento. Quando uma situação DT é detectada, a adaptação de  $s(n)$  deve ser interrompida durante algum período de tempo, o qual deve ser preferencialmente igual ao do DTD. Porém, até ser detectada, DT já pode ter resultado em altos valores para o fator de escalamento. Para garantir que qualquer distúrbio em  $s(n)$  decaia lentamente quando uma situação de DT for detectada, permite-se que, neste caso, o fator de escalamento decaia até um valor mínimo  $s_{min}$ .

A relação entre robustez e velocidade de convergência do algoritmo adaptativo é determinada pela velocidade de rastreamento da variável de escalamento  $s(n)$ , a qual é controlada por um único parâmetro  $\lambda_s$ . Porém, este parâmetro também influencia na velocidade de convergência do algoritmo adaptativo. Em uma situação de convergência, para uma mesma constante  $k_0$ , quanto mais rápido  $s(n)$  acompanhar a diminuição do sinal de erro, mais rapidamente o algoritmo entrará em modo robusto.

O algoritmo NLMS robusto é descrito na Tabela 3.2. Analogamente, os outros algoritmos adaptativos citados na Seção 2.6 podem ser modificados gerando-se uma variante robusta.

Tabela 3.1: Algoritmo NLMS Robusto Tipo 1

---

**Filtro**

$$e(n) = d(n) - \mathbf{w}^T(n)\mathbf{x}(n)$$

$$\psi \left[ \frac{|e(n)|}{s(n)} \right] = \min \left\{ \frac{|e(n)|}{s(n)}, k_0 \right\}$$

**DTD**

se conversação cruzada:

$$\mu = 0$$

$$s(n+1) = \lambda_s s(n) + (1 - \lambda_s) s_{min}$$

caso contrário:

$$\mu = \mu_0$$

$$s(n+1) = \lambda_s s(n) + \frac{1 - \lambda_s}{\beta} \psi \left[ \frac{|e(n)|}{s(n)} \right] s(n)$$

**Algoritmo Adaptativo**

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu \mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{x}(n) + \gamma} \psi \left[ \frac{|e(n)|}{s(n)} \right] \text{sign}[e(n)]s(n)$$


---

### 3.4.2 Algoritmos Robustos Tipo 2

Da mesma maneira como descrito na Subseção 3.4.1, a referência [17] propõe uma mudança na função custo da forma

$$J = E \left\{ \rho \left[ \frac{e(n)}{s} \right] \right\}, \quad (3.42)$$

de maneira que

$$\rho(z) = \ln[\cosh(z)]. \quad (3.43)$$

Devido a esta escolha, o gradiente de  $J$  é

$$\nabla_{\mathbf{w}} J = -\frac{\mathbf{x}(n)}{s} \psi \left[ \frac{e(n)}{s} \right], \quad (3.44)$$

onde

$$\psi(z) = \frac{d\rho(z)}{dz} \quad (3.45)$$

$$= \tanh(z). \quad (3.46)$$

Logo, a segunda derivada de  $J$  é

$$\nabla_{\mathbf{w}}^2 J = \frac{\mathbf{x}(n)\mathbf{x}^T(n)}{s^2} \psi' \left[ \frac{e(n)}{s} \right], \quad (3.47)$$

onde

$$\psi'(z) = \operatorname{sech}^2(z). \quad (3.48)$$

Porém, aqui é proposta uma pequena variação em  $\psi(z)$  para poder haver uma comparação com a utilizada na Subseção 3.4.1

$$\psi(z) = k_0 \tanh(z). \quad (3.49)$$

A inclusão da constante  $k_0$  permite controlar o tamanho do passo do algoritmo em seu modo robusto além de, neste caso, igualar o valor da função  $\psi(z)$  deste algoritmo com a do algoritmo tipo 1. Porém, da maneira como é inserida, a constante  $k_0$  será sempre um ganho dado ao tamanho do passo do algoritmo.

A função  $\psi(z)$  neste caso possui, ao contrário da utilizada na Subseção 3.4.1, um comportamento suave como pode ser visto na Figura 3.2. Além disto, ela estará sempre limitada superiormente pela função  $\psi(z)$  do algoritmo robusto tipo 1 o que, provavelmente, implicará em uma menor velocidade de convergência do algoritmo e em uma maior robustez a distúrbios de curta duração.

Novamente, deriva-se o novo algoritmo NLMS pelo método de Newton resultando na atualização dos coeficientes do filtro

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu \mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{x}(n) \psi' \left[ \frac{e(n)}{s(n)} \right] + \gamma} \psi \left[ \frac{e(n)}{s} \right] s(n). \quad (3.50)$$

Para evitar divergência do algoritmo robusto, impede-se que  $\psi'(z)$  seja menor que 0,5. Na prática, calcula-se  $\psi'(z)$  de acordo com a equação (3.48) limitando-o inferiormente a este valor. A inclusão deste fator na atualização dos coeficientes deve, provavelmente, aumentar a velocidade de convergência do algoritmo devido ao seu baixo valor.

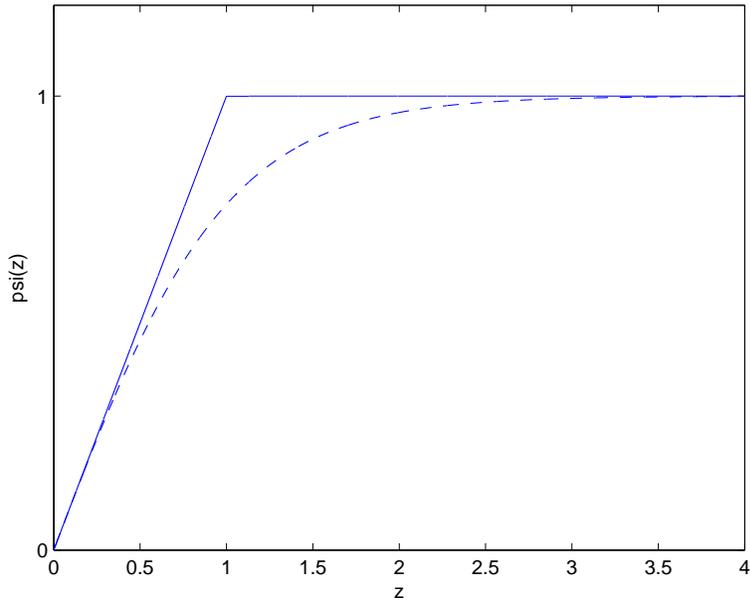


Figura 3.2: Funções  $\psi(z)$  do algoritmo tipo 1 (linha sólida) e tipo 2 (linha tracejada) para  $k_0 = 1$ .

Utilizando os mesmos critérios para a determinação da equação (3.37), o fator de escalamento é estimado como

$$s(n+1) = \lambda_s s(n) + (1 - \lambda_s) \frac{s(n)}{\psi' \left[ \frac{e(n)}{s(n)} \right]} \left| \psi \left[ \frac{e(n)}{s(n)} \right] \right|. \quad (3.51)$$

O algoritmo NLMS robusto tipo 2 é descrito na Tabela 3.2. Analogamente, os outros algoritmos adaptativos citados na Seção 2.6 podem ser derivados de maneira robusta.

Tabela 3.2: Algoritmo NLMS Robusto Tipo 2.

---

**Filtro**

$$e(n) = d(n) - \mathbf{w}^T(n)\mathbf{x}(n)$$

$$\psi \left[ \frac{e(n)}{s(n)} \right] = k_0 \tanh \left[ \frac{e(n)}{s(n)} \right]$$

$$\psi' \left[ \frac{e(n)}{s(n)} \right] = \operatorname{sech}^2 \left[ \frac{e(n)}{s(n)} \right]$$

se  $\psi' \left[ \frac{e(n)}{s(n)} \right] < 0.5$ :  $\psi' \left[ \frac{e(n)}{s(n)} \right] = 0.5$

**DTD**

se conversação cruzada:

$$\mu = 0$$

$$s(n+1) = \lambda_s s(n) + (1 - \lambda_s) s_{min}$$

caso contrário:

$$\mu = \mu_0$$

$$s(n+1) = \lambda_s s(n) + \frac{1 - \lambda_s}{\psi' \left[ \frac{e(n)}{s(n)} \right]} \left| \psi \left[ \frac{e(n)}{s(n)} \right] \right| s(n)$$

**Algoritmo Adaptativo**

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu \mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{x}(n)\psi' \left[ \frac{e(n)}{s(n)} \right] + \gamma} \psi \left[ \frac{e(n)}{s(n)} \right]$$


---

## 3.5 Conclusões

Este capítulo discutiu o problema de conversação cruzada em dispositivos de cancelamento de eco devido à utilização de filtros adaptativos. Notou-se necessário um detector de conversação cruzada encarregado de diminuir ou parar completamente a adaptação do filtro quando a presença da fala do locutor próximo for detectada para evitar que o algoritmo divirja.

Foram apresentados os princípios dos algoritmos de detecção de conversação cruzada e os algoritmos existentes para tal função. Entretanto, na prática a detecção de conversação cruzada não ocorre perfeitamente. Os algoritmos de DT costumam apresentar alguns tipos de falhas, como demorar algum tempo para detectar uma situação de DT ou mesmo não detectá-las completamente. Estes problemas são suficientes para influenciarem o comportamento do algoritmo adaptativo causando divergência.

Por isso, fez-se necessário o desenvolvimento de algoritmos adaptativos robustos a distúrbios de curta duração devido às falhas de detecção de DT. A intenção é fazer com que os algoritmos sejam poucos sensíveis aos erros de detecção de DT e, conseqüentemente, não permitam uma passagem excessiva de eco. Uma atual solução encontrada na literatura, denominada nesta dissertação de tipo 1, é apresentada. A inclusão de uma constante na função decisora de uma segunda solução, denominada tipo 2, é proposta para que possa haver um controle no tamanho do passo do algoritmo em seu modo robusto e uma comparação mais justa com a solução tipo 1.

# Capítulo 4

## Cancelamento de Eco Elétrico

### 4.1 Introdução

Neste capítulo será discutido o desenvolvimento de um cancelador de eco elétrico. Para tal, será utilizada como base a recomendação ITU-T G.168 [1], a qual resume brevemente o tema em questão e define o desempenho mínimo que um cancelador de eco elétrico deve apresentar nos três ambientes de interesse: convergência, re-convergência e situação de conversação cruzada.

Segundo [1], um esquema de simulação realístico de cancelamento de eco elétrico é ilustrado na Figura 4.1, onde  $g(k)$  é um dos possíveis percursos de eco encontrados em [1], os quais serão detalhados na Seção 4.2.

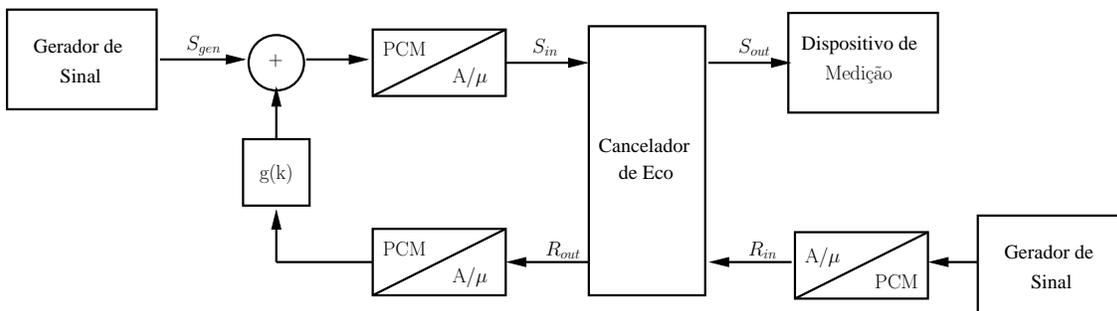


Figura 4.1: Modelamento de percurso de eco.

Percebe-se que uma das características deste modelo é a presença de várias codificações/decodificações entre PCM linear e PCM  $\mu/A$ -law, resultando em uma inserção de ruído de codificação, visto que a codificação/decodificação PCM linear  $\rightarrow$  PCM  $\mu/A$ -law  $\rightarrow$  PCM linear é uma operação com perdas. Neste trabalho, optou-se

pela codificação  $\mu$ -law.

Os sinais utilizados nas simulações do sistema de cancelamento de eco elétrico serão os sinais CSS (*Composite Source Signal*) definidos em [1] e anteriormente em [18]. Esses sinais artificiais são feitos de tal maneira que possuem características semelhantes aos sinais de voz reais, contendo trechos de sinal sonoro, surdo e silêncio, como será descrito na Seção 4.3.

A Seção 4.4 descreve o método RMS utilizado para medir a quantidade de energia dos sinais CSS e o dispositivo de medição de energia de curtos trechos de sinais, que envolve, além do método RMS, um filtro passa-banda e um filtro exponencial de 1 pólo.

Por fim, a Seção 4.5 apresenta os testes elaborados para validar o desempenho de canceladores de eco elétrico em situações de convergência, re-convergência e conversação cruzada.

## 4.2 Percursos de Eco

De acordo com [1], os percursos de eco podem ser simulados como filtros digitais lineares com resposta ao impulso  $g(n)$  de duração  $t_d$ . Para possibilitar vários possíveis atrasos ( $\delta$ ) e diferentes atenuações ( $ERL$ ),  $g(n)$  é definido como

$$g(n) = (10^{-\frac{ERL}{20}} K_i) m_i(n - \delta), \quad (4.1)$$

ou seja, uma versão atrasada e atenuada de qualquer uma das seqüências  $m_i(n)$ , onde  $i = 1, 2, \dots, 8$ . As seqüências  $m_i(n)$  representam percursos de eco com diferentes características de dispersão e duração.

As seqüências utilizadas serão  $m_1(n)$ ,  $m_5(n)$  e  $m_6(n)$  e para cada uma dessas seqüências, os valores de  $K_i$  e  $ERL$  mínimo são fixados como mostra a Tabela 4.1. Segundo [1],  $m_5(n)$  e  $m_6(n)$  são modelos de percursos de eco medidos na América do Norte. Já percursos de eco medidos na França apresentam características equivalentes a  $m_1(n)$ ,  $m_5(n)$  e  $m_6(n)$ .

Tabela 4.1: Fatores de escalamento  $K_i$  e  $ERL$ s mínimos [1].

Percurso de Eco	Fator de Escalamento $K_i$	$ERL$ mínimo para CSS
1	$1.39 \times 10^{-5}$	6 dB
5	$9.33 \times 10^{-6}$	6 dB
6	$1.51 \times 10^{-5}$	6 dB

Nas simulações de cancelamento de eco elétrico utilizaram-se, em todos os casos,  $\delta = 0$  e o  $ERL$  mínimo, ou seja, quando a atenuação natural é a menor possível. Assim, o cancelador de eco é exposto à situação mais crítica, onde o sinal de eco possui alta energia e, conseqüentemente, precisa sofrer uma grande atenuação. Sendo assim, as respostas ao impulso dos percursos de eco utilizados são mostradas nas Figuras 4.2, 4.3 e 4.4.

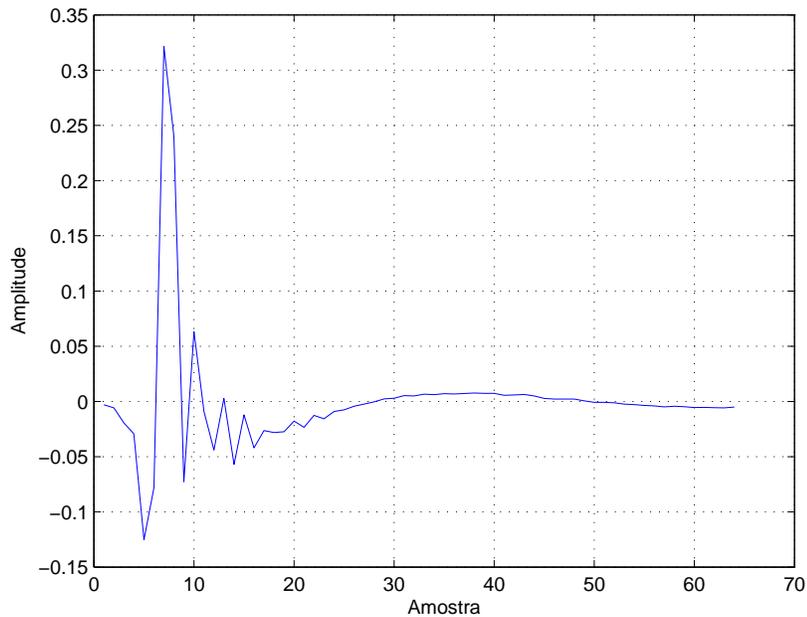


Figura 4.2: Resposta ao impulso do modelo de percurso de eco 1 com  $ERL = 6$  dB e  $\delta = 0$ .

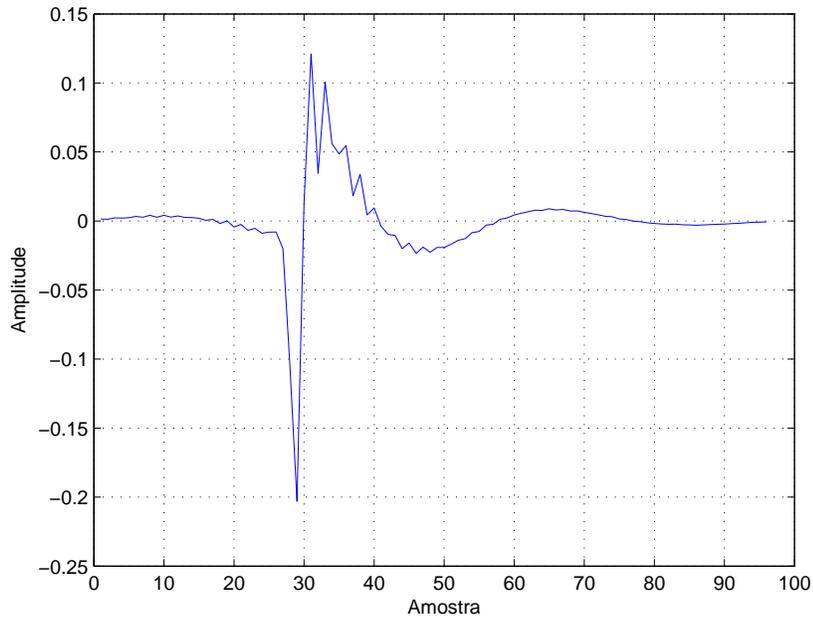


Figura 4.3: Resposta ao impulso do modelo de percurso de eco 5 com  $ERL = 6$  dB e  $\delta = 0$ .

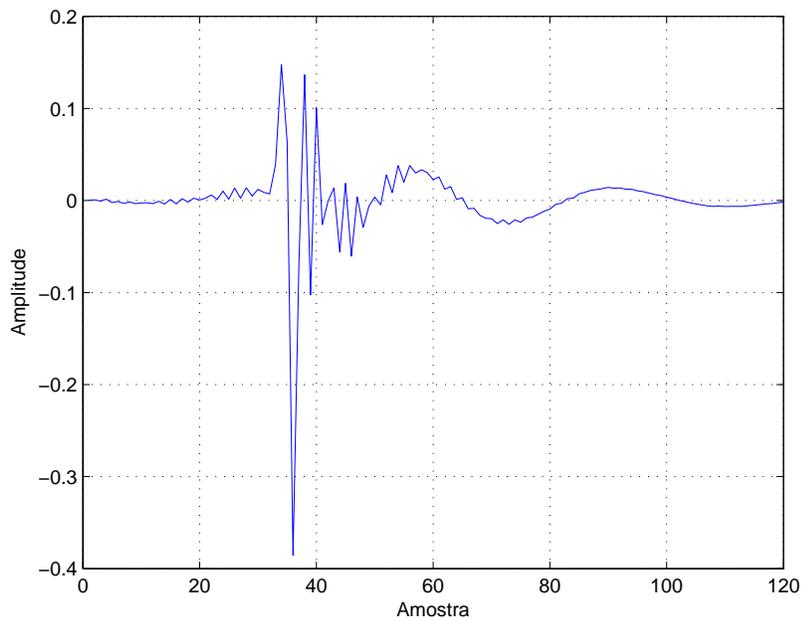


Figura 4.4: Resposta ao impulso do modelo de percurso de eco 6 com  $ERL = 6$  dB e  $\delta = 0$ .

### 4.3 Sinais *Composite Source Signal*

Os sinais CSS são definidos em [1, 18] para os casos de conversação simples e cruzada, onde o último é utilizado como sinal de voz do locutor próximo. Os sinais CSS apresentam características semelhantes aos sinais de voz reais, sendo constituídos por três trechos em ordem: sinal sonoro, sinal surdo e silêncio; onde a duração das seqüências depende do tipo do sinal CSS. No final, cada seqüência é repetida com a amplitude invertida.

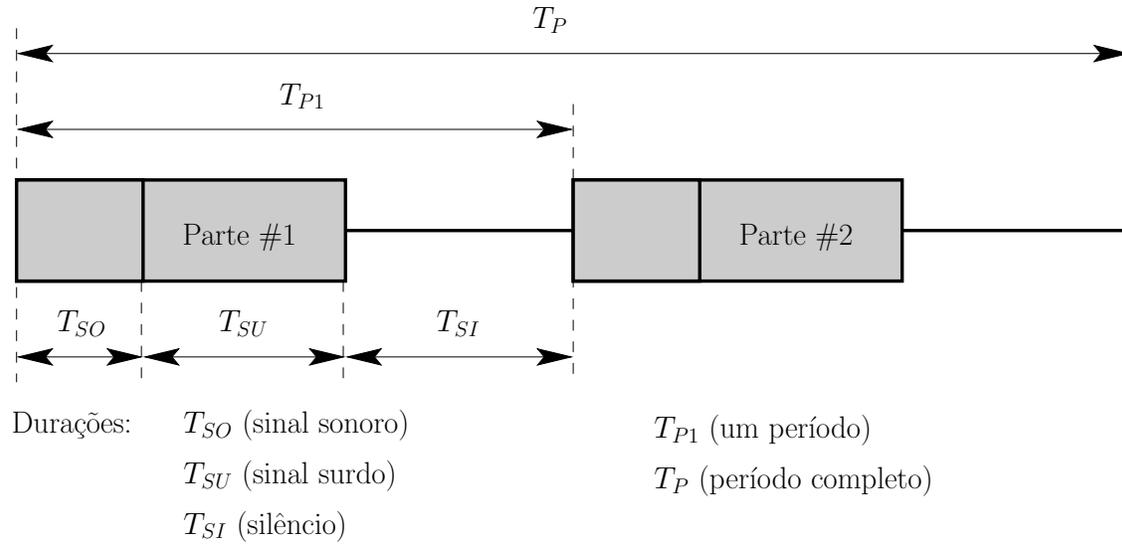


Figura 4.5: Modelo do sinal CSS [1].

#### 4.3.1 Sinal Sonoro Limitado em Banda

A referência [1] define um vetor com 134 valores em 16 bits referentes ao sinal sonoro limitado em banda entre 200 Hz e 3,6 kHz para o caso de conversação simples. A Figura 4.6 ilustra 10 repetições deste vetor. Com uma taxa de amostragem de 44,1 kHz, os 134 valores representam um intervalo de tempo de 3,04 ms, sendo que a seqüência é repetida 16 vezes para alcançar uma duração  $T_{SO} = 48,62$  ms.

Para o caso de conversação cruzada, um vetor com 229 valores em 16 bits é definido como sinal sonoro. A Figura 4.7 ilustra 10 repetições deste vetor. Utilizando uma taxa de amostragem igual a 44,1 kHz, os 229 valores constituem um intervalo de tempo de 5,19 ms, sendo que a seqüência é repetida 14 vezes para alcançar uma duração  $T_{SO} = 72,69$  ms.

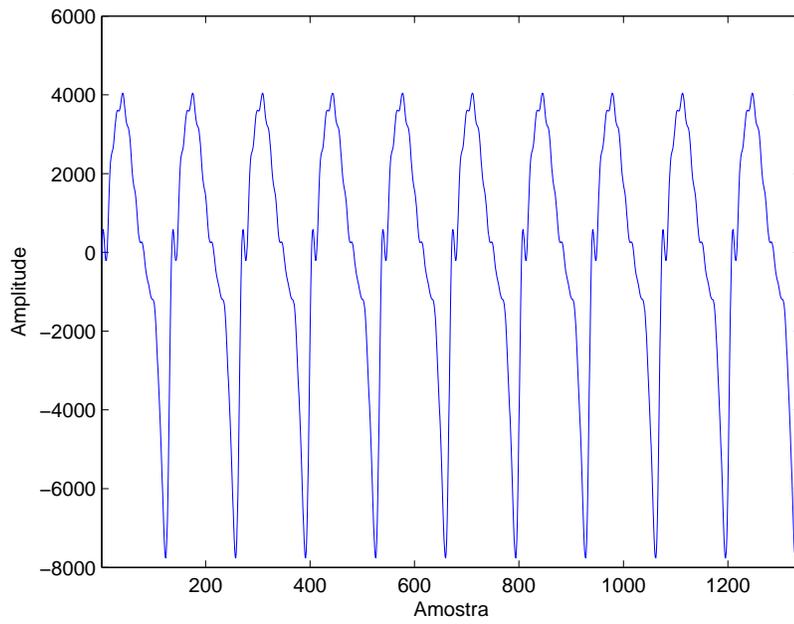


Figura 4.6: Trecho sonoro limitado em banda do sinal CSS para conversação simples.

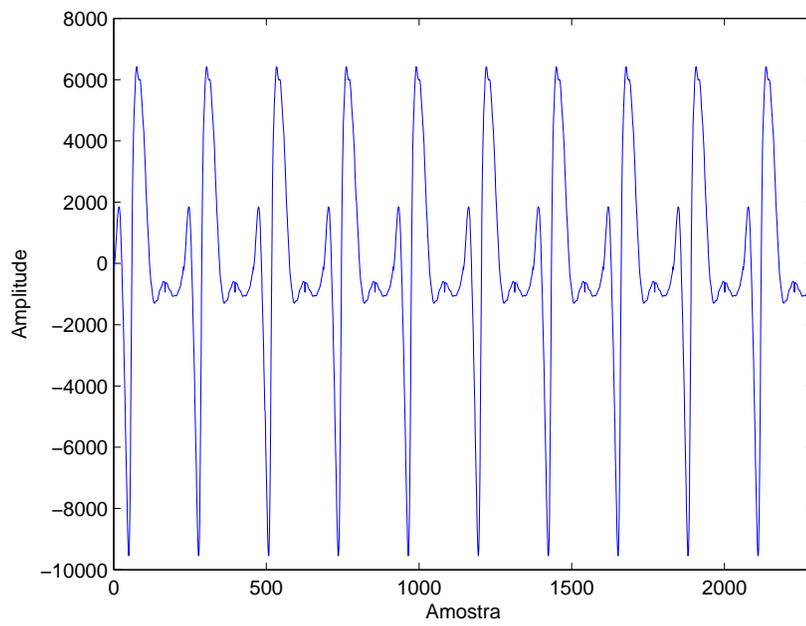


Figura 4.7: Trecho sonoro limitado em banda do sinal CSS para conversação cruzada.

### 4.3.2 Sinal Surdo

O sinal presente após o sinal sonoro artificial é o sinal surdo. Este sinal contém características de sinais ruidosos, sendo gerado por meio de um pseudo ruído no caso de conversação simples. Na sua construção, primeiramente um espectro complexo é produzido no domínio da frequência de acordo com

$$H(n) = W(n)e^{j i_n \pi}, \quad (4.2)$$

onde

$$n = -\frac{M}{2}, \dots, -1, 1, \dots, \frac{M}{2}, \quad (4.3)$$

$$i_n = \{+1, 0\} \text{ aleatório}, \quad (4.4)$$

$$i_n = -i_{-n}. \quad (4.5)$$

O índice  $M$  representa o número de pontos da FFT. Utilizando  $M = 8192$ , existirão 3715 pontos da FFT na faixa entre 0 e 20 kHz devido à resolução de 5,4 Hz (44,1 kHz/8192). O valor de  $W(n)$  é escolhido constante e igual a 305360 para que a energia na banda de 0 e 20 kHz dos sinais sonoros e surdos seja igual.

Então, calcula-se a IFFT do espectro  $H(n)$  gerando o sinal no domínio do tempo

$$S(n) = \frac{1}{M} \sum_{k=-\frac{M}{2}}^{\frac{M}{2}} H(n)e^{j2\pi nk}, \quad k = -\frac{M}{2}, \dots, \frac{M}{2} - 1, \quad (4.6)$$

que é repetida 1,077 vezes para atingir a duração  $T_{SU} = 200$  ms.

A Tabela 4.2 mostra a diferença entre o filtro ideal e o filtro de ordem 8192 utilizado com o objetivo de limitar a banda entre 200 Hz e 3,6 kHz do sinal surdo sem variar a energia. O filtro utilizado possui resposta em frequência exibida na Figura 4.8.

Para o caso de sinais CSS para conversação cruzada, o sinal surdo é definido como ruído branco gaussiano com banda até 20 kHz e duração  $T_{SU} = 200$  ms. O nível do ruído aleatório é determinado de maneira que a energia na banda entre 0 e 20 kHz dos sinais sonoros e surdos seja igual. Com o intuito de limitar a banda entre 200 Hz e 3,6 kHz do sinal surdo sem alterar a energia, o filtro definido na Tabela 4.2 é aplicado.

Tabela 4.2: Filtro ideal e filtro utilizado para limitar em banda o sinal surdo sem alteração de energia.

Filtro Ideal		Filtro Utilizado	
Frequência (Hz)	Módulo (dB)	Frequência (Hz)	Módulo (dB)
50	-25,8	49,4838	-24,9913
100	-12,8	98,9677	-10,1078
200	17,4	200,4096	17,4214
215	17,8	215,2447	17,7528
500	12,2	499,7868	12,2072
1000	7,2	999,5736	7,2188
2850	0	2850,3	-0,0018
3600	-2	3599,9	-2,1859
3660	-20	3661,8	-20,7151
3680	-30	3679,1	-29,0138

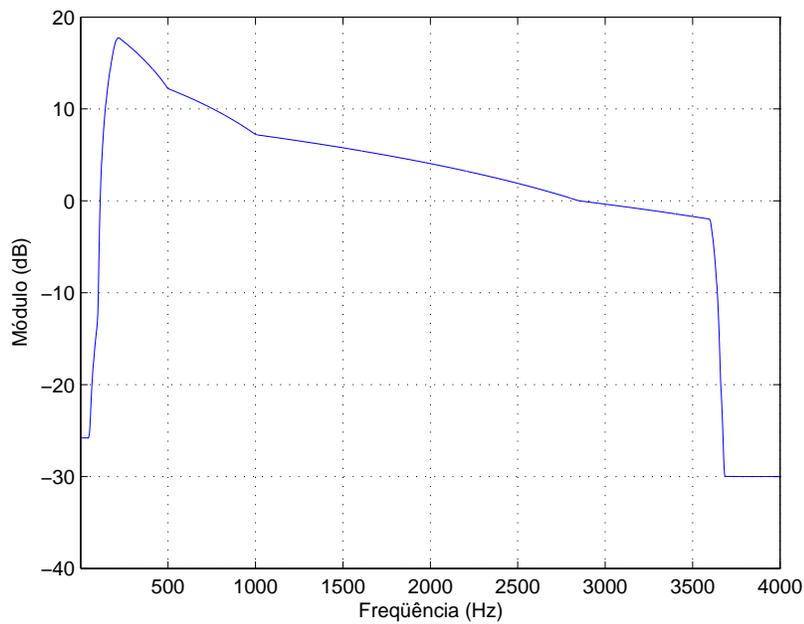


Figura 4.8: Resposta em frequência do filtro utilizado para limitar em banda o sinal surdo.

### 4.3.3 Silêncio

O trecho de silêncio para conversação simples é representado por um vetor de zeros com duração  $T_{SI} = 101,38$  ms, fazendo com que a seqüência sinal sonoro, sinal surdo e silêncio possua  $T_{P1} = 350$  ms de duração. Após isso, o sinal CSS para conversação simples é repetido com amplitude invertida e reamostrado para 8 kHz, a taxa de telefonia fixa, apresentando um tamanho de 5600 amostras e duração  $T_P = 700$  ms.

No caso de sinais CSS para conversação cruzada, o trecho de silêncio é definido como um vetor de zeros com duração  $T_{SU} = 127,31$  ms, resultando em  $T_{P1} = 400$  ms de duração para a seqüência sinal sonoro, sinal surdo e silêncio. O sinal CSS para conversação cruzada é, então, repetido com amplitude invertida e reamostrado para 8 kHz, resultando em um tamanho de 6400 amostras e duração  $T_P = 800$  ms.

## 4.4 Método de Análise Qualitativa

Nesta seção serão apresentados a métrica utilizada para medir a quantidade de energia dos sinais CSS e o dispositivo de medição para curtos trechos de sinais, que envolve um filtro passa-banda e um filtro exponencial de 1 pólo.

### 4.4.1 Métrica Utilizada

Para propósito de testes, a métrica especificada para medição da quantidade de energia dos sinais CSS é um método RMS definido como [1]

$$S(n) = 3,17 + 20\log_{10} \left[ \frac{\sqrt{\frac{2}{k} \sum_{i=n}^{n-k+1} e_i^2}}{8159} \right], \quad (4.7)$$

para codificação  $\mu$ -law, onde  $S(n)$  é o nível de energia do sinal em dBm0,  $n$  é o indexador no tempo e  $k$  é o número de amostras levadas em conta na medição, sendo  $k$  um múltiplo do período do sinal CSS. Nesta dissertação, foi utilizado  $k$  igual ao período do sinal CSS.

Em alguns momentos, a métrica definida será aplicada apenas na parte ativa dos sinais CSS, ou seja, excluindo os trechos de silêncio. Sendo  $L_{Rin}$  o nível de

energia do sinal  $R_{in}$ , o nível de energia da parte ativa de  $R_{in}$  pode ser obtido para os sinais CSS para conversação simples e cruzada, respectivamente, como

$$L_{Rin,act} = L_{Rin} + 1,49, \quad (4.8)$$

$$L_{Rin,act} = L_{Rin} + 1,66. \quad (4.9)$$

#### 4.4.2 Dispositivo de Medição

Nos testes que virão a seguir, será necessário medir o nível de energia de curtos trechos de sinais. Isto é feito utilizando um dispositivo de medição que integra 4 blocos e utiliza os valores dos sinais de interesse codificados em PCM linear.

O primeiro bloco refere-se a um filtro linear passa-faixa de ordem 100 com resposta em frequência mostrada na Figura 4.9. Este filtro é utilizado para atenuar as componentes em frequência fora da banda de interesse e para eliminar o ganho DC devido à codificação A-law [1].

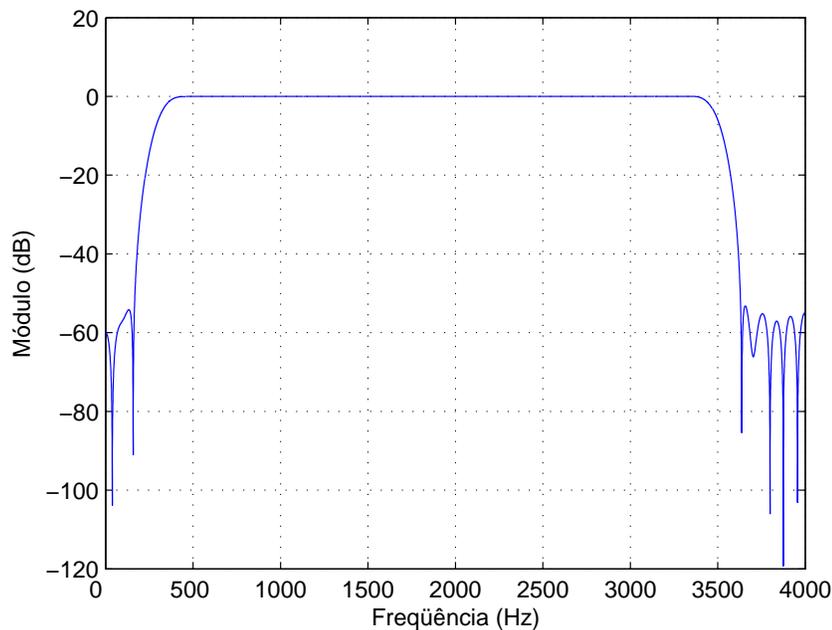


Figura 4.9: Resposta em frequência do filtro passa-faixa.

O segundo bloco eleva ao quadrado as amostras do sinal de saída do filtro linear passa-banda, enquanto o terceiro filtra este novo sinal por um filtro exponencial de 1 pólo com tempo de decaimento de 35 ms, ou seja, um filtro passa-baixa com um único pólo em 0,9964. O quarto bloco refere-se à medição em si e utiliza a métrica descrita na Subseção 4.4.1.

## 4.5 Simulações

Esta seção descreve os principais testes definidos em [1] para a validação de um cancelador de eco elétrico, restringindo-se aos testes de convergência, re-convergência e conversação cruzada. As Seções 4.5.1, 4.5.2 e 4.5.3, além de descreverem as configurações e os desempenhos mínimos exibidos por [1], apresentam os resultados obtidos pelos algoritmos adaptativos clássicos nas situações de convergência, re-convergência e conversação cruzada, respectivamente.

Nas simulações seguintes foram utilizados o esquema de simulação apresentado na Seção 4.1, os percursos de eco definidos na Seção 4.2 e os sinais CSS descritos na Seção 4.3. Os resultados das simulações representam a média de 20 experimentos, sendo cada experimento composto por 40 sinais CSS para conversação simples e 35 para conversação cruzada, e todas as medidas foram obtidas utilizando o método descrito na Seção 4.4.

Os valores dos parâmetros dos algoritmos adaptativos foram estimados empiricamente de maneira que todos apresentassem o mesmo desajuste, tornando mais justa a comparação de velocidade de convergência e re-convergência. Já os parâmetros dos algoritmos robustos foram estimados de maneira que o algoritmo tipo 1 apresentasse uma grande robustez no teste de conversação cruzada. Todos os parâmetros são encontrados na Tabela 4.3.

Tabela 4.3: Parâmetros dos algoritmos adaptativos.

Algoritmo	$\mu$	$\gamma$	$k_0$	$\lambda_s$
NLMS	0,8	$10^5$	0,75	0,9985
PNLMS	0,8	$10^5$	0,75	0,9985
BNDR-LMS	0,8	$10^{15}$	0,75	0,9985
P-BNDR-LMS	0,8	$10^{18}$	0,75	0,9985

### 4.5.1 Teste de Convergência

#### 4.5.1.1 Descrição e Objetivos

Este teste objetiva garantir que o cancelador de eco convirja rapidamente para todas as combinações de níveis de sinais de entrada e de percursos de eco, e

que o nível de retorno do eco seja suficientemente baixo. No início da conexão, a convergência deve ser rápida o suficiente para não ser percebida.

O teste é realizado segundo o esquema exibido na Figura 4.10, onde o filtro é inicializado com zeros e a adaptação começa, no mínimo, 200 ms antes da presença de sinal na entrada do filtro no tempo  $t_0$ .



Figura 4.10: Esquema para teste de convergência [1].

Segundo [1], para todos os valores de  $-30 \text{ dBm0} \leq L_{Rin,act} \leq 0 \text{ dBm0}$ , de  $ERL \geq 6 \text{ dB}$  e do comprimento do filtro  $\Delta \geq t_d$ , a perda  $(L_{Rin,act} - L_{RES})$  deve ser maior ou igual a mostrada na Figura 4.11, onde  $L_{RES}$  é igual a  $(L_{Rin} - ERL)$  menos a atenuação gerada pelo cancelador de eco elétrico. Após  $(10 + t_d + t_0) \text{ s}$ , a perda  $(L_{Rin} - L_{RES})$  deve ser maior que ou igual a mostrada na Figura 4.12.

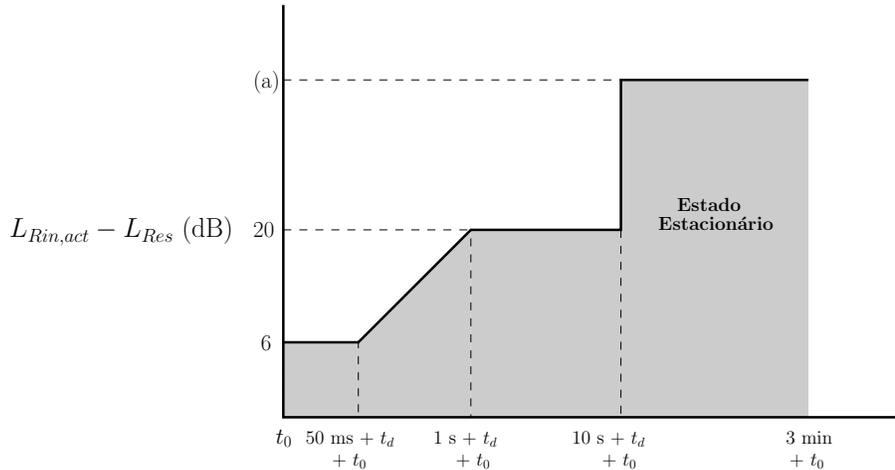


Figura 4.11: Atenuação necessária ao longo do tempo [1].

A variável  $(L_{Rin,act} - L_{RES})$  pode ser substituída por  $(L_{Sin} - L_{Sout} + ERL)$ , onde  $L_{Sin}$  e  $L_{Sout}$  são os níveis de energia de  $S_{in}$  e  $S_{out}$ , respectivamente. Neste teste, o nível do sinal  $R_{in}$ ,  $L_{Rin}$ , é medido como descrito na Subseção 4.4.1, enquanto a medição dos outros sinais é feita como descrito na Subseção 4.4.2. Nota-se que  $(L_{Sin} - L_{Sout} + ERL)$  é a atenuação aplicada pelo cancelador mais a atenuação natural  $ERL$  do percurso de eco.

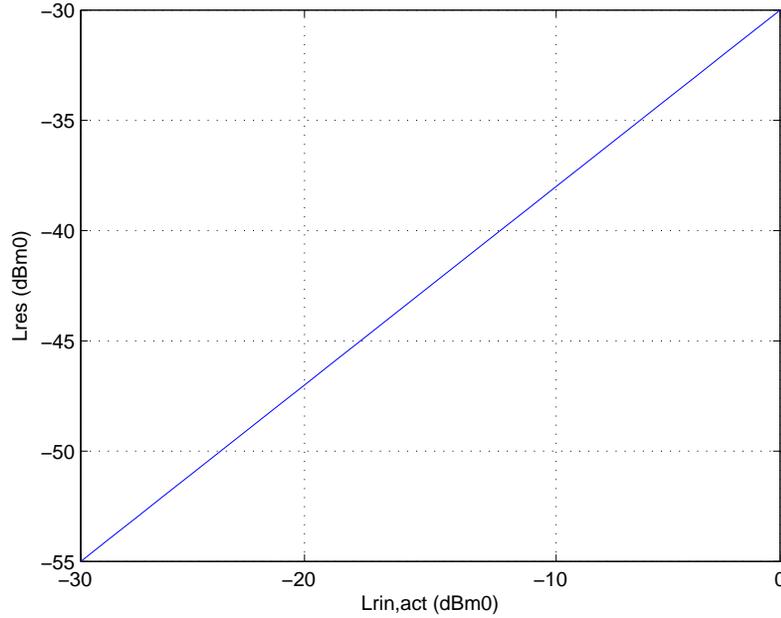


Figura 4.12: Atenuação necessária após  $(10 + t_d + t_0)$  s [1].

#### 4.5.1.2 Resultados

Em todas as simulações realizadas utilizou-se  $L_{Rin,act} = -2,9352$  dBm0, sendo este o valor médio encontrado nos sinais CSS criados como descrito na Seção 4.3 sem aplicar nenhum tipo de ganho aos mesmos,  $ERL = 6$  dB, sendo este o valor mínimo para todos os percursos de eco selecionados, e  $\Delta = 256$  ms. Este valor de  $L_{Rin,act}$  é bem próximo do valor máximo permitido para a energia do sinal de entrada, enquanto este valor de  $ERL$  é a menor atenuação natural permitida. Em conjunto, estes valores fazem com que seja gerado um sinal de eco com alta energia, o que força o cancelador de eco a trabalhar em uma situação crítica onde há a necessidade de aplicar uma grande atenuação ao sinal de eco.

Como indicado na Figura 4.12, com este valor de  $L_{Rin,act}$  o cancelador de eco deve ser capaz de produzir um sinal de saída com  $L_{RES} = -32,446$  dBm0, ou seja, deve ser capaz de aplicar uma atenuação de 29,5108 dB após  $(10 + t_d + t_0)$  s.

Inicialmente, comparamos o comportamento do algoritmo NLMS robusto tipo 1 com o do NLMS convencional. A média do comportamento destes algoritmos nos três percursos em questão é mostrada na Figura 4.13. O início de adaptação do filtro  $t_0$  é indicado com o instante 0 s do gráfico e, como descrito na Seção 4.10, a presença de sinal ocorre após 200 ms.

Percebe-se que ambos os algoritmos satisfazem as condições de convergência e que, como esperado, o algoritmo robusto apresenta uma menor velocidade de convergência e um menor desajuste.

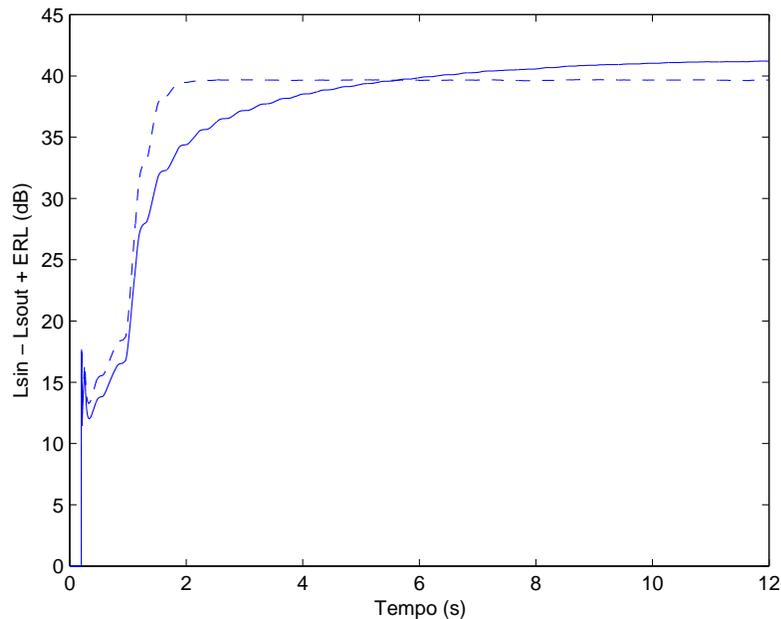


Figura 4.13: Velocidade de convergência dos algoritmos NLMS convencional (linha tracejada) e robusto tipo 1 (linha contínua).

Certificado o funcionamento do algoritmo robusto, a velocidade dos algoritmos robustos NLMS, PNLMS e BNDR-LMS são confrontadas na Figura 4.14. Os algoritmos PNLMS e BNDR-LMS apresentam uma convergência mais rápida que o NLMS, com pequena vantagem para o BNDR-LMS e, conseqüentemente, satisfazem as condições mínimas exigidas.

A comparação entre os algoritmos BNDR-LMS e o P-BNDR-LMS encontra-se na Figura 4.15, de onde conclui-se que o algoritmo P-BNDR-LMS apresentou a maior velocidade de convergência entre todos os algoritmos testados, apesar do melhor início de convergência do BNDR-LMS.

Por fim, interessava-nos confrontar os algoritmos robustos tipo 1 e tipo 2. Para isso, utilizamos nos algoritmos tipo 2 os mesmos valores de  $k_0$  e  $\lambda_s$  para visualizarmos claramente as conseqüências da utilização de uma função custo diferente. A Figura 4.16 mostra o desempenho do algoritmo NLMS robusto tipo 2 comparado com o do NLMS tipo 1.

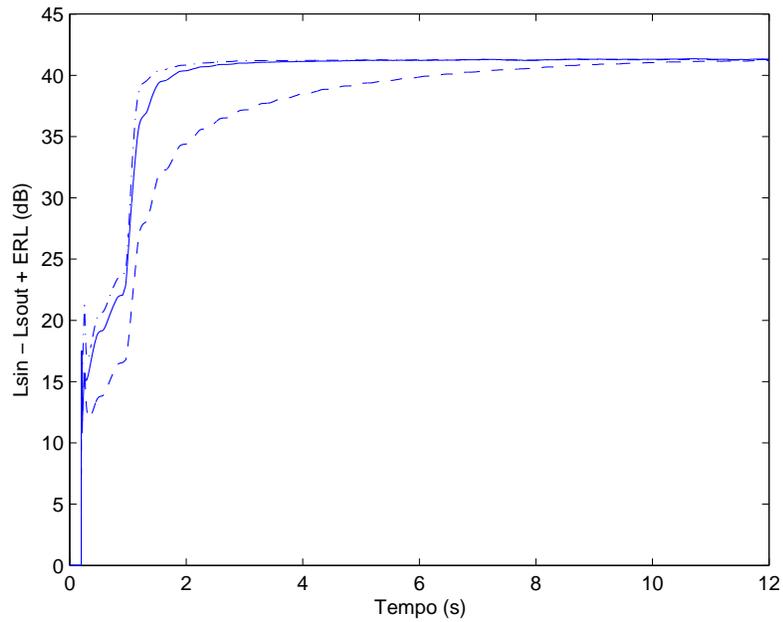


Figura 4.14: Velocidade de convergência dos algoritmos robustos tipo 1 NLMS (linha tracejada), PNLMS (linha contínua) e BNDR-LMS (linha ponto-traço).

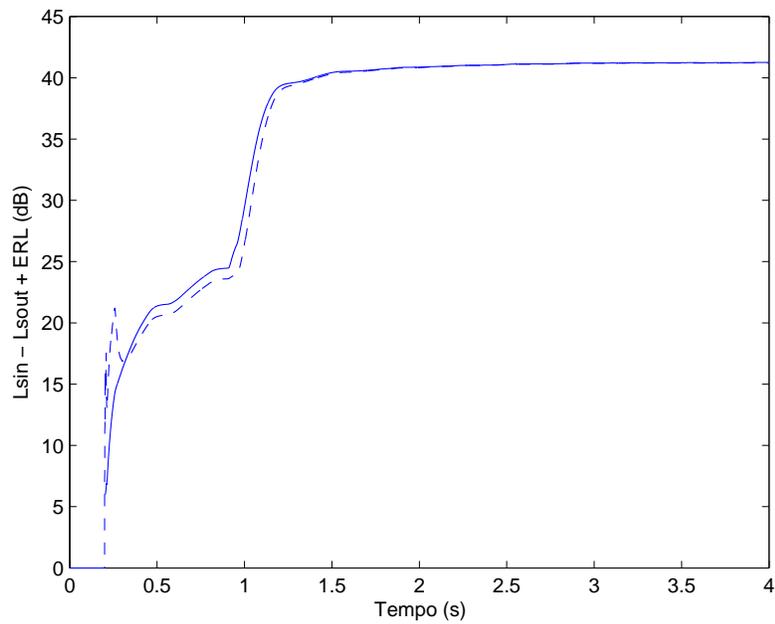


Figura 4.15: Velocidade de convergência dos algoritmos robustos tipo 1 BNDR-LMS (linha tracejada) e P-BNDR-LMS (linha contínua).

Nota-se que o algoritmo tipo 2 apresenta uma velocidade de convergência mais rápida que o algoritmo tipo 1. Entretanto, esta diferença mostrou-se variável dependendo do algoritmo. No algoritmo P-BNDR-LMS a diferença é mínima como

pode ser visto na Figura 4.17, ainda mais se levarmos em consideração que, neste caso, a utilização do algoritmo tipo 2 apresentou uma pequena variação no desajuste.

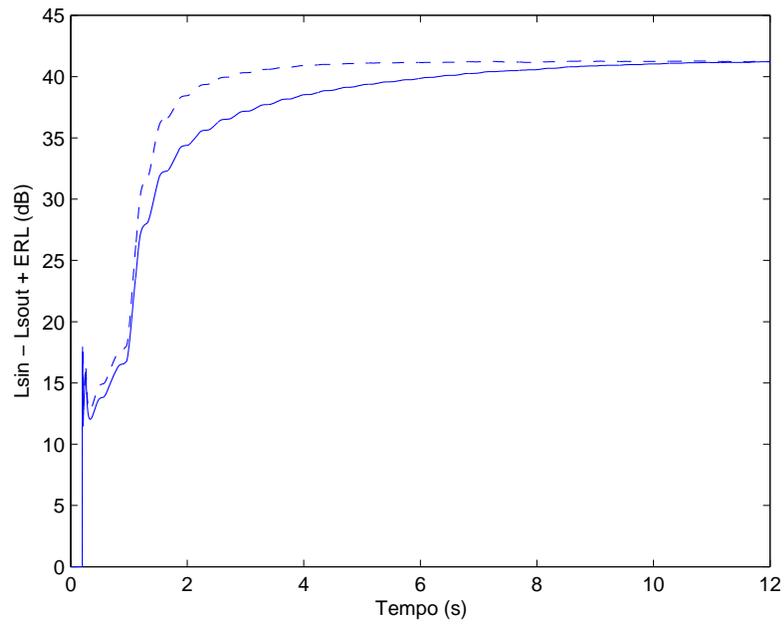


Figura 4.16: Velocidade de convergência dos algoritmos NLMS robusto tipo 1 (linha contínua) e robusto tipo 2 (linha tracejada).

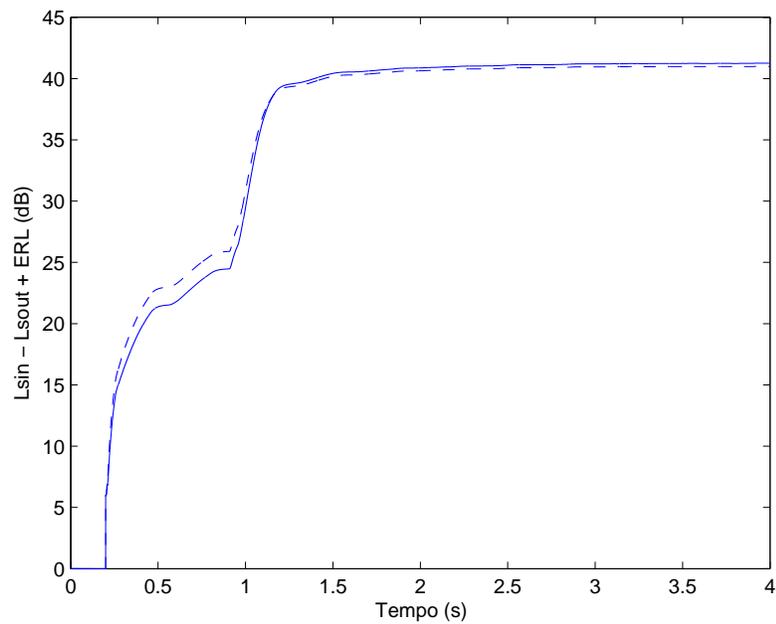


Figura 4.17: Velocidade de convergência dos algoritmos P-BNDR-LMS robusto tipo 1 (linha contínua) e robusto tipo 2 (linha tracejada).

## 4.5.2 Teste de Re-Convergência

### 4.5.2.1 Descrição e Objetivos

Este teste objetiva garantir que o cancelador de eco re-convirja rapidamente para todas as combinações de níveis de sinais de entrada em ambientes após a ocorrência de variações bruscas no percurso do eco. Espera-se que o nível de retorno do eco após uma mudança brusca no percurso de eco seja suficientemente baixo em pouco tempo.

O teste é realizado segundo o esquema exibido na Figura 4.18. O filtro adaptativo deve ter convergido para o percurso de eco inicial no momento em que ocorre a alteração do percurso, no tempo  $t_0$ . As alterações presentes nas simulações são percurso 1→percurso 5, percurso 5→percurso 6 e percurso 6→percurso 1.

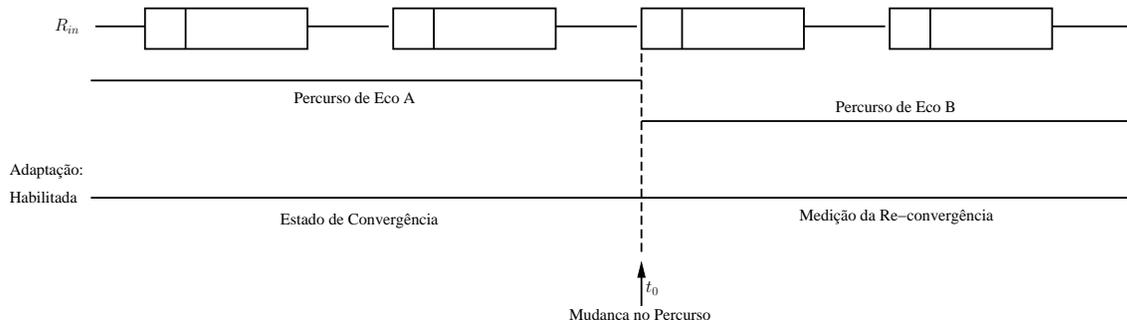


Figura 4.18: Esquema para teste de re-convergência [1].

Segundo [1], para todos os valores de  $-30 \text{ dBm0} \leq L_{Rin,act} \leq 0 \text{ dBm0}$ , de  $ERL \geq 6 \text{ dB}$  e do comprimento do filtro  $\Delta \geq t_d$  para ambos os percursos, a perda ( $L_{Rin,act} - L_{RES}$ ) deve ser maior ou igual à mostrada na Figura 4.19. Após  $(10 + t_d + t_0)$  s, a perda ( $L_{Rin,act} - L_{RES}$ ) deve ser maior que ou igual a exibida na Figura 4.12.

Novamente, a variável ( $L_{Rin,act} - L_{RES}$ ) pode ser substituída por ( $L_{Sin} - L_{Sout} + ERL$ ), onde  $L_{Sin}$  e  $L_{Sout}$  são os níveis de energia de  $S_{in}$  e  $S_{out}$ , respectivamente. O nível do sinal  $R_{in}$ ,  $L_{Rin}$ , é medido como descrito na Subseção 4.4.1, enquanto a medição dos outros sinais é feita de acordo com a Subseção 4.4.2.

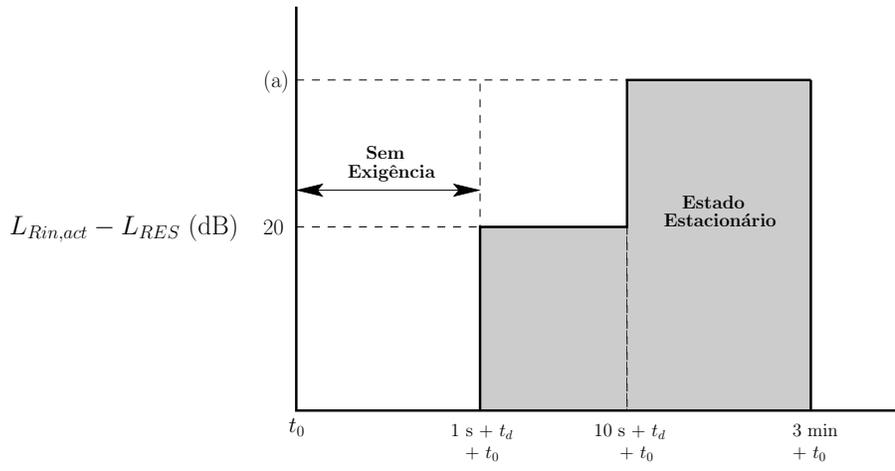


Figura 4.19: Atenuação necessária ao longo do tempo [1].

#### 4.5.2.2 Resultados

Assim como na Subseção 4.5.1, em todas as simulações realizadas foram utilizadas  $L_{Rin,act} = -2,9352$  dBm0, sendo este o valor médio encontrado nos sinais CSS criados como descrito na Seção 4.3 sem aplicar nenhum tipo de ganho aos mesmos,  $ERL = 6$  dB, sendo este o valor mínimo para todos os percursos de eco selecionados, e  $\Delta = 256$  ms. Com este valor de  $L_{Rin,act}$  o cancelador de eco deve ser capaz de produzir um sinal de saída com  $L_{RES} = -32,446$  dBm0, ou seja, deve ser capaz de aplicar uma atenuação de 29,5108 dB após  $(10 + t_d + t_0)$  s.

Mais uma vez, iniciamos comparando o comportamento do algoritmo NLMS robusto tipo 1 com o do NLMS convencional. A média do comportamento destes algoritmos nos três percursos em questão é mostrada na Figura 4.20. A alteração no percurso de eco ocorre no instante  $t_0 = 0$  s do gráfico.

Percebe-se que apenas o algoritmo NLMS convencional satisfaz as exigências de re-convergência. O algoritmo NLMS robusto tipo 1 não satisfaz a condição em  $(1 + t_d + t_0)$  s e está no limite do exigido em  $(10 + t_d + t_0)$  s. Vale lembrar que uma menor velocidade de re-convergência do algoritmo robusto já era esperado devido ao modo como é derivado, como explicado na Subseção 3.4.1.

Confrontando agora os algoritmo robustos NLMS, PNLMS e BNDR-LMS, percebe-se pela Figura 4.21 que nenhum deles satisfaz a exigência em  $(1 + t_d + t_0)$  s. Entretanto, os algoritmos PNLMS e BNDR-LMS apresentam uma re-convergência bem superior à do NLMS, tendo o BNDR-LMS o melhor resultado, apesar do PNLMS ser superior nos instantes iniciais.

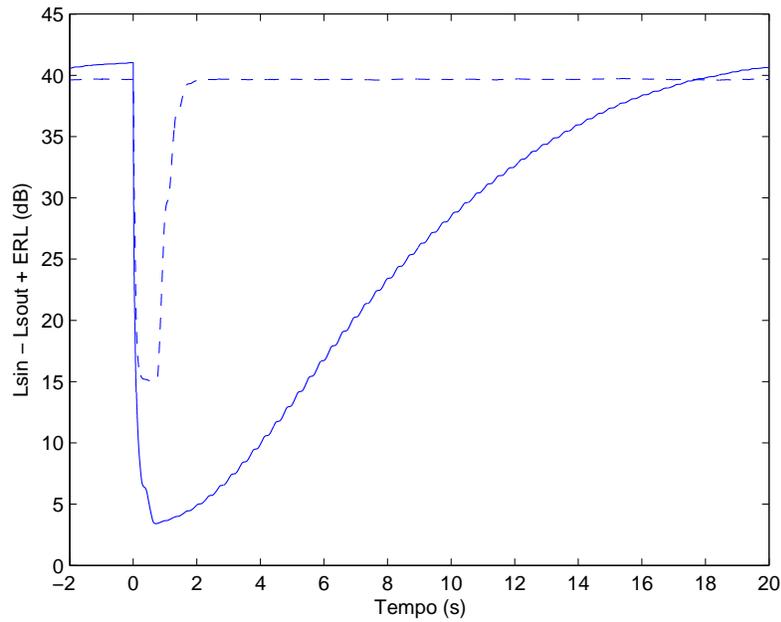


Figura 4.20: Velocidade de re-convergência dos algoritmos NLMS convencional (linha tracejada) e robusto tipo 1 (linha contínua).

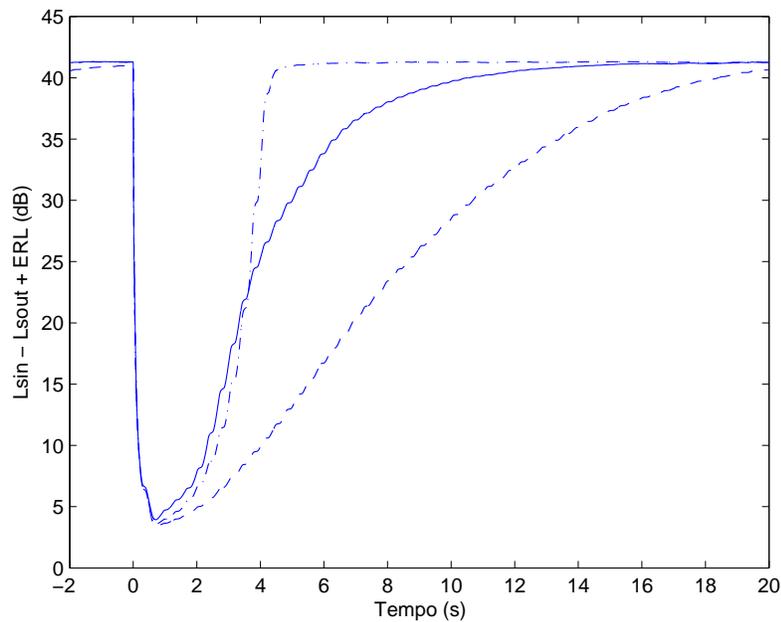


Figura 4.21: Velocidade de re-convergência dos algoritmos robustos tipo 1 NLMS (linha tracejada), PNLMS (linha contínua) e BNDR-LMS (linha ponto-traço).

A Figura 4.22 compara os algoritmos PNLMS, BNDR-LMS e P-BNDR-LMS, de onde conclui-se que o algoritmo P-BNDR-LMS apresentou uma velocidade de re-convergência superior à de todos os algoritmos testados. Porém, ainda não suficiente

para satisfazer a exigência de 20 dB de atenuação no instante  $(1 + t_d + t_0)$  s, onde  $t_d$  é a duração do percurso de eco. Este é o preço a ser pago para fazer o filtro adaptativo robusto em situações de conversação cruzada, como será comprovado na Subseção 4.5.3.

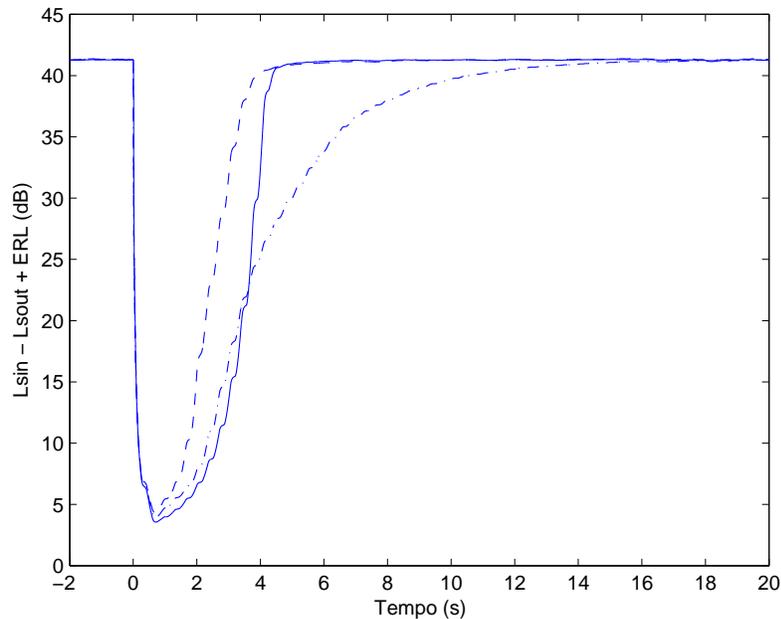


Figura 4.22: Velocidade de re-convergência dos algoritmos robustos tipo 1 PNLMS (linha ponto-traço), BNDR-LMS (linha contínua) e P-BNDR-LMS (linha tracejada).

Defrontando os algoritmos robustos NLMS tipo 1 e tipo 2 na Figura 4.23, percebe-se que, novamente, o algoritmo tipo 2 apresenta uma re-convergência mais rápida. O mesmo resultado foi obtido nos demais algoritmos, como exemplificado para o P-BNDR-LMS na Figura 4.24. Entretanto, ainda não sendo suficiente para satisfazer a primeira exigência de [1].

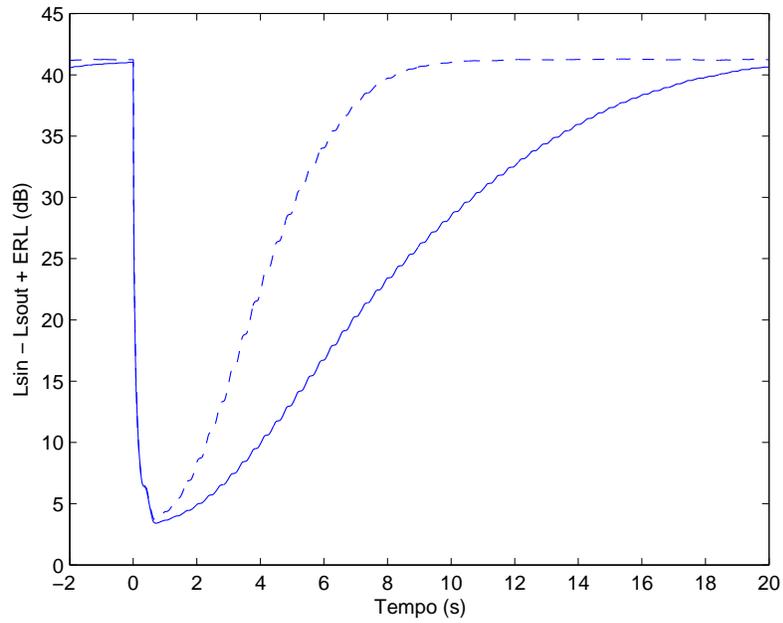


Figura 4.23: Velocidade de re-convergência dos algoritmos NLMS robusto tipo 1 (linha contínua) e robusto tipo 2 (linha tracejada).

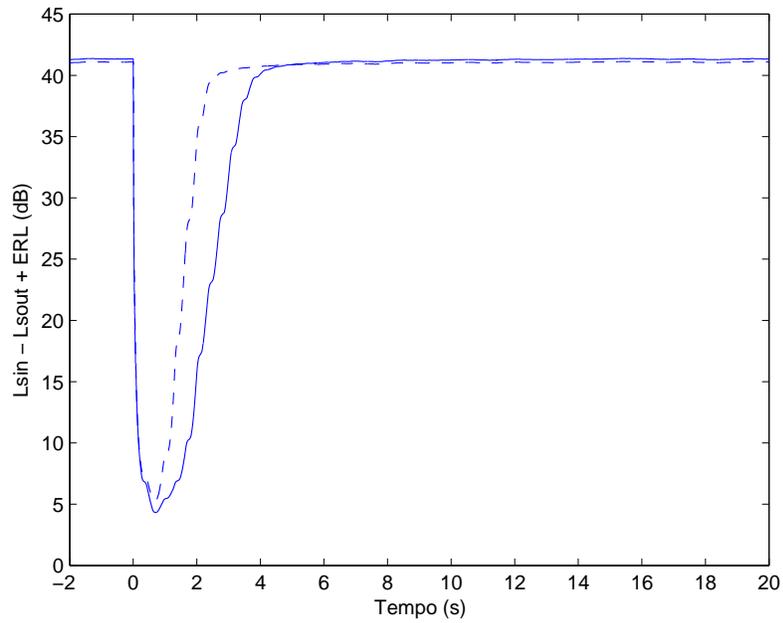


Figura 4.24: Velocidade de re-convergência dos algoritmos P-BNDR-LMS robusto tipo 1 (linha contínua) e robusto tipo 2 (linha tracejada).

## 4.5.3 Teste de Conversação Cruzada

### 4.5.3.1 Descrição e Objetivos

Este teste objetiva garantir que o cancelador de eco apresente um bom desempenho em situações de conversação cruzada, não diverjindo e evitando excessiva redução da atenuação aplicada ao sinal de eco. Segundo [1], o teste é realizado segundo o esquema exibido na Figura 4.25.

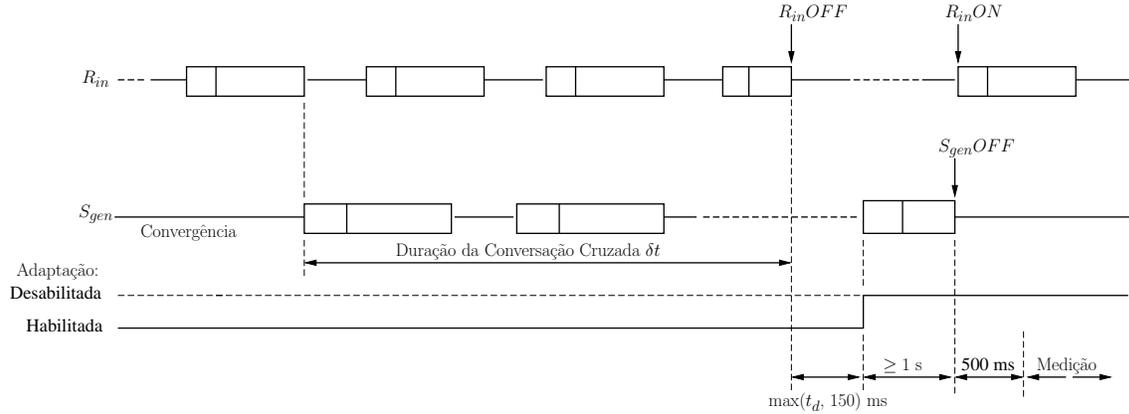


Figura 4.25: Esquema para teste de conversação cruzada [1].

Neste experimento, porém, a medição será feita durante todo o tempo, mesmo antes da conversação cruzada. Assim, a atenuação é calculada como  $(L_{Sin-S_{gen}} - L_{Sout-S_{gen}} + ERL)$ .

De acordo com [1], para todos valores de  $-30 \text{ dBm0} \leq L_{Rin,act} \leq 0 \text{ dBm0}$ , de  $L_{S_{gen}} \geq L_{Rin}$ , de  $ERL \geq 6 \text{ dB}$  e do comprimento do filtro  $\Delta \geq t_d$ , o nível de retorno de eco não deve cair mais que 10 dB em relação à atenuação prevista na Figura 4.12.

### 4.5.3.2 Resultados

Quanto maior  $L_{S_{gen}}$  mais fácil é para detectar situações de DT. Por isso, para manter o mesmo princípio usado anteriormente de expor o cancelador de eco às situações mais críticas, em nossas simulações foi usado  $L_{S_{gen}} \cong L_{Rin}$ . Como DTD utilizamos o algoritmo de Geigel considerando uma atenuação de 3 dB ( $T = \sqrt{2}$ ) e  $t_{hold} = 40 \text{ ms}$ .

Objetivando verificar a eficiência do algoritmo robusto, confrontamos na Figura 4.26 o algoritmo NLMS convencional e o algoritmo NLMS robusto tipo 1, ambos fazendo uso do Geigel DTD.

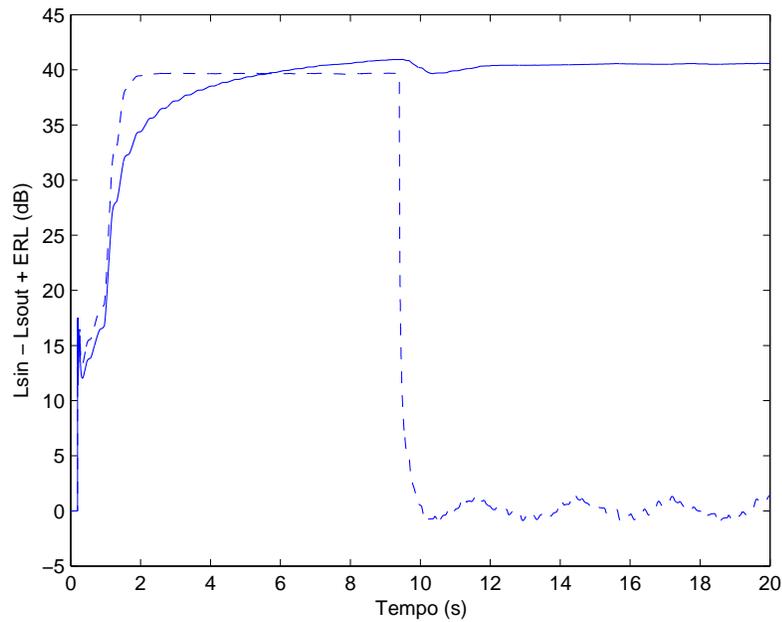


Figura 4.26: Desempenho em situação de conversação cruzada do algoritmo NLMS convencional (linha tracejada) e robusto tipo 1 (linha contínua).

Percebe-se, neste caso, a grande vantagem do algoritmo robusto em relação aos algoritmos convencionais. As falhas de detecção de situações de DT pelo algoritmo de Geigel resultam em um aumento excessivo do desajuste do NLMS convencional, enquanto o algoritmo NLMS robusto tipo 1 praticamente não é afetado.

A Figura 4.27 exhibe o comportamento dos algoritmos robustos tipo 1 NLMS, PNLMS e BNDR-LMS nas mesmas situações. Novamente, os algoritmos são levemente afetados pela situação de DT, não tendo suas convergências decaindo 10 dB em nenhum momento. Do mesmo modo, a Figura 4.28 exhibe o comportamento do P-BNDR-LMS confrontado com o BNDR-LMS.

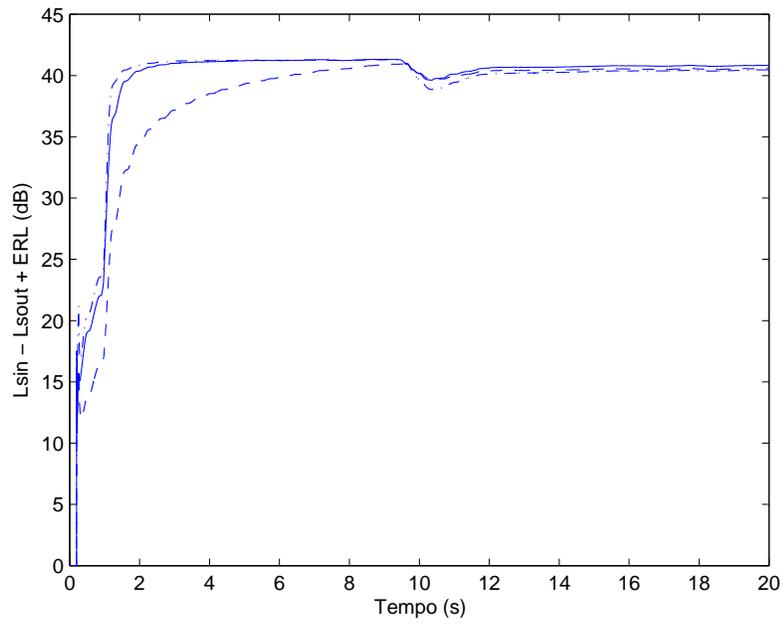


Figura 4.27: Desempenho em situação de conversação cruzada dos algoritmos robustos tipo 1 NLMS (linha tracejada), PNLMS (linha contínua) e BNDR-LMS (linha ponto-traço).

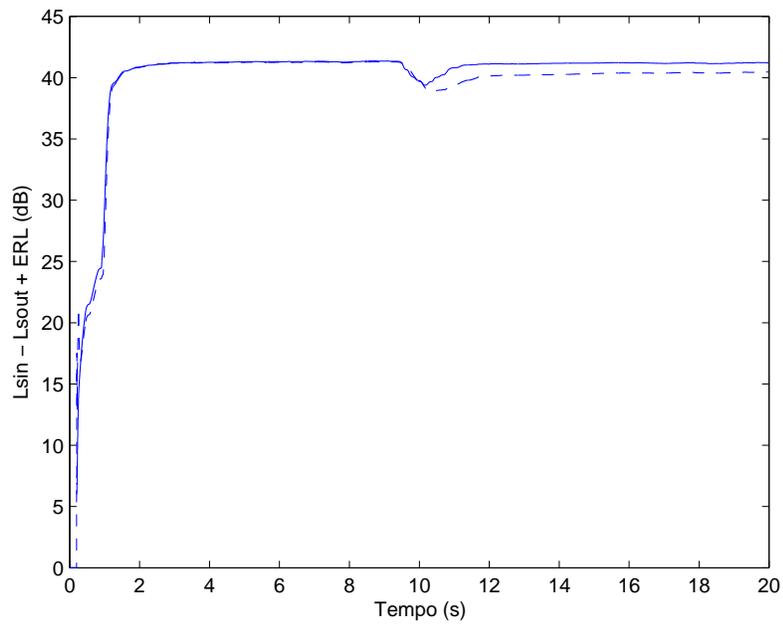


Figura 4.28: Desempenho em situação de conversação cruzada dos algoritmos robustos tipo 1 BNDR-LMS (linha tracejada) e P-BNDR-LMS (linha contínua).

Por fim, confrontamos o algoritmo NLMS robusto tipo 2 com o anteriormente utilizado. Como mostrado nas Figuras 4.29 e 4.30, apesar de apresentar um bom comportamento, não variando demasiadamente a atenuação aplicada ao sinal de eco, a variação apresentada pelo algoritmo robusto tipo 2 foi maior que a do algoritmo robusto tipo 1. Isto pode ser explicado pelo fato que, geralmente, algoritmos com uma maior velocidade de convergência tendem a ser mais sensíveis aos erros do DTD.

As simulações mostraram que os algoritmos robustos são muito menos sensíveis a falhas de detecção de DT que os algoritmos convencionais, tendo suas convergências levemente alteradas. Entretanto, o custo desta eficiência em situação de DT é uma menor capacidade de re-convergência como demonstrado nos testes da Subseção 4.5.2. Porém, na média a utilização de algoritmos robustos mostrou ser a melhor opção.

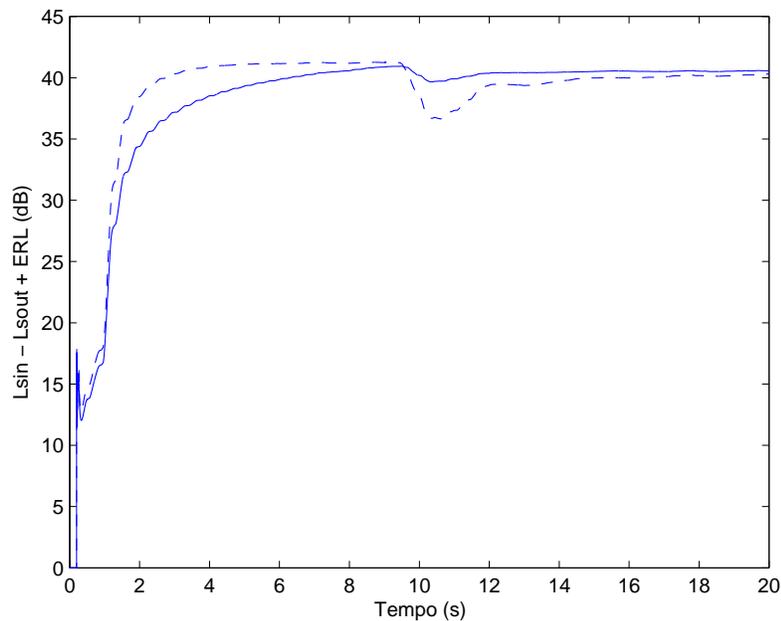


Figura 4.29: Desempenho em situação de conversação cruzada do algoritmo NLMS robusto tipo 1 (linha contínua) e robusto tipo 2 (linha tracejada).

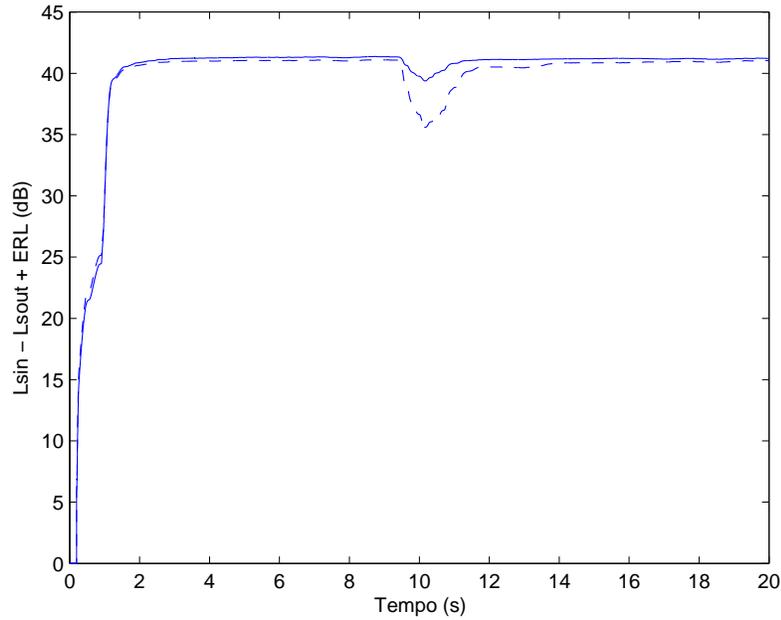


Figura 4.30: Desempenho em situação de conversação cruzada do algoritmo P-BNDR-LMS robusto tipo 1 (linha contínua) e robusto tipo 2 (linha tracejada).

## 4.6 Conclusões

Este capítulo descreveu o sistema de simulação para o desenvolvimento de um cancelador de eco elétrico definido na recomendação ITU-T G.168 [1]. Foram detalhados os percursos de eco que apresentam características semelhantes a percursos reais assim como os sinais CSS que simulam as características dos sinais de voz reais.

Os algoritmos adaptativos foram testados nas situações de convergência, re-convergência e conversação cruzada. Os resultados obtidos foram comparados com as exigências mínimas existentes na recomendação. As exigências não foram satisfeitas apenas no caso de nível de re-convergência após aproximadamente 1 s depois da alteração do percurso de eco utilizando algoritmos robustos, e no caso de situação de conversação cruzada utilizando algoritmos convencionais. Entretanto, a utilização de algoritmos robustos mostrou-se claramente necessária devido ao grande aumento do desajuste encontrado no segundo caso.

Entre os algoritmos adaptativos utilizados, o algoritmo P-BNDR-LMS foi o que apresentou o melhor desempenho destacando-se a sua capacidade de convergência. Porém, este algoritmo possui a maior complexidade computacional entre todos. Com um menor custo, o algoritmo BNDR-LMS apresentou um desempenho similar ao P-BNDR-LMS, distanciando-se apenas nos instantes iniciais no caso de re-convergência.

Comparando os algoritmos robustos tipo 1, encontrado na literatura, e tipo 2, proposto nesta dissertação, concluiu-se que, com os mesmos parâmetros, o algoritmo tipo 2 apresenta uma maior velocidade de convergência e re-convergência, e por conseqüência, apresenta uma maior sensibilidade aos erros de detecção de conversação cruzada. Isto deve-se a uma maior influência da inclusão da variável  $\psi'(|z|)$  em comparação à alteração de  $\psi(|z|)$  ocorrida no algoritmo robusto tipo 2.

# Capítulo 5

## Cancelamento de Eco Acústico

### Mono-Canal

#### 5.1 Introdução

Em um sistema de teleconferência, o eco acústico é o resultado do acoplamento entre o alto-falante e o microfone no outro extremo da conexão. Isto significa que, ao falar, o usuário recebe de volta uma reprodução da própria fala que pode ser percebida como uma perturbação incômoda. A popularização dos sistemas de teleconferência faz com que cada vez seja mais importante o cancelamento desta forma de eco.

Este capítulo apresenta simulações de cancelamento de eco acústico mono-canal em uma sala de teleconferência utilizando os algoritmo adaptativos descritos no Capítulo 2. A Seção 5.2 descreve as configurações dos experimentos, incluindo a criação de uma base de dados contendo sinais de eco gerados em tempo real e o método de análise do desempenho utilizado na avaliação dos algoritmos, assim como os procedimentos realizados para a especificação dos parâmetros dos algoritmos.

Na Seção 5.3 são apresentados os resultados obtidos nas situações de convergência e de conversação cruzada, onde nesta última situação é discutido um problema encontrado devido às características da sala de teleconferência onde ocorreram as gravações. A Seção 5.4 resume o presente capítulo enfatizando as principais contribuições.

## 5.2 Configuração dos Experimentos

Como descrito no Capítulo 3, o cancelamento de eco acústico mono-canal utiliza a fala do locutor distante  $x(n)$  para modelar o percurso de eco  $\mathbf{h}$  por meio do filtro adaptativo  $\mathbf{w}$ , obtendo uma estimativa do eco  $y(n)$  a qual é subtraída do sinal contaminado  $d(n)$ , gerando o sinal de erro

$$\begin{aligned} e(n) &= d(n) - y(n) \\ &= (\mathbf{h}^T - \mathbf{w}^T) \mathbf{x}(n), \end{aligned} \quad (5.1)$$

que é efetivamente o sinal transmitido ao locutor distante na ausência de sinal de voz do locutor próximo.

Para as simulações de cancelamento de eco acústico foram realizadas gravações de sinais de eco em tempo real. Para isso, foram utilizados sinais originalmente gravados com uma taxa de amostragem de 16 kHz formados por concatenações de frases foneticamente balanceadas [19]. A geração dos sinais de eco foi feita reproduzindo estes sinais e gravando, em tempo real, o eco capturado por um microfone. Este procedimento foi realizado em uma sala de teleconferência sem preparo acústico. Para facilitar as avaliações, o sinal de fala do locutor próximo  $s(n)$  foi sempre zero.

Como índice de desempenho dos algoritmos adaptativos foi utilizado o erro médio quadrático (MSE) do erro residual definido por

$$\text{MSE} = \frac{P_e}{P_d} = \frac{\text{LPF}[e^2(n)]}{\text{LPF}[d^2(n)]}, \quad (5.2)$$

onde LPF é um filtro passa-baixas com um simples pólo em 0,9996. Em todas as simulações subseqüentes foram utilizados 150 trechos de 20 s.

Com o objetivo de ter uma idéia do comprimento do filtro necessário no sistema de cancelamento de eco acústico, foi feita a estimativa da resposta ao impulso  $\mathbf{h}$  da sala de teleconferência, a qual é mostrada na Figura 5.1.

Por esta figura, observa-se que a primeira reflexão ocorre com um atraso de aproximadamente 250 amostras e que as reflexões mais significativas concentram-se nas primeiras 2000 amostras. Para uma melhor estimativa da ordem do filtro, os algoritmos adaptativos descritos na Seção 2.6 foram testados com diferentes comprimentos de filtro ( $N = 1024, 2048, 4096, 8192$ ). Os resultados encontrados indicaram que alguns algoritmos adaptativos, como LMS, NLMS e NDR-LMS, apresentavam

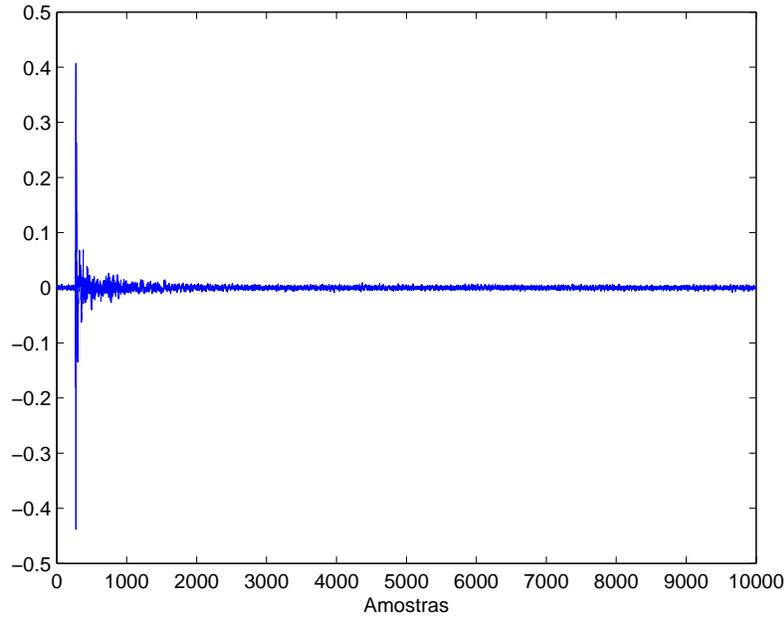


Figura 5.1: Experimento 1: Estimativa da resposta ao impulso  $\mathbf{h}$  da sala de teleconferência.

uma maior atenuação com  $N = 8192$ , enquanto outros algoritmos com  $N = 4096$ . Como em todos os casos a diferença de desempenho dos algoritmos com este valores de comprimento de filtro foi pequena, de aproximadamente 1 dB, decidiu-se utilizar nas simulações subsequentes  $N = 4096$  por este valor resultar em uma menor complexidade computacional.

Diferentemente do Capítulo 4, os parâmetros dos algoritmos adaptativos foram estimados, por meio de uma busca extensiva, de forma que o MSE médio resultante fosse minimizado. Este procedimento foi feito devido à grande dificuldade de igualar o desajuste dos algoritmos, onde seria necessário um número extremamente alto de simulações, que se tornariam exaustivas devido à elevada ordem dos filtros adaptativos que faz-se necessária. Além disso, alguns algoritmos necessitariam de um tempo muito elevado para alcançar o estado estacionário, como poderá ser observado na Subseção 5.3.1.

Além disso, devido às características do ambiente de simulação, que serão discutidas na Subseção 5.3.2, a utilização de algoritmos robustos mostrou-se ineficiente, não sendo aplicada. Assim, os parâmetros dos algoritmos adaptativos para a ordem do filtro  $N = 4096$  são mostrados na Tabela 5.1.

Tabela 5.1: Parâmetros dos algoritmos adaptativos.

Algoritmo	$\mu$	$\gamma$
LMS	$3.7 \times 10^{-12}$	—
NLMS	1	$10^{10}$
PNLMS	1	$10^{10}$
NDR-LMS	1,7	$10^{10}$
BNDR-LMS	1,2	$10^{18}$
P-BNDR-LMS	0,9	$10^{17}$

## 5.3 Resultados

### 5.3.1 Convergência

As Figuras 5.2 e 5.3 representam a convergência média de cada algoritmo adaptativo, cujos principais parâmetros indicados na Tabela 5.1 foram determinados empiricamente minimizando o MSE resultante de cada algoritmo.

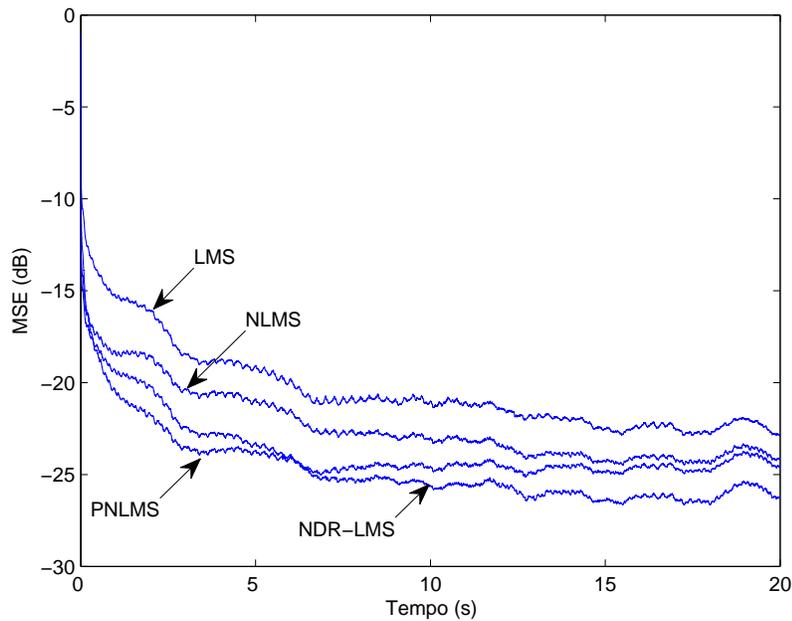


Figura 5.2: Erro médio quadrático dos algoritmos LMS, NLMS, PNLMS e NDR-LMS.

Como esperado, o algoritmo LMS resultou em uma convergência mais lenta dentre todos os algoritmos. A técnica de proporcionalidade aumenta a velocidade de

convergência do algoritmo NLMS de uma maneira mais eficiente que a reutilização de dados do algoritmo NDR-LMS que, por outro lado, apresenta uma convergência final mais significativa que o PNLMS.

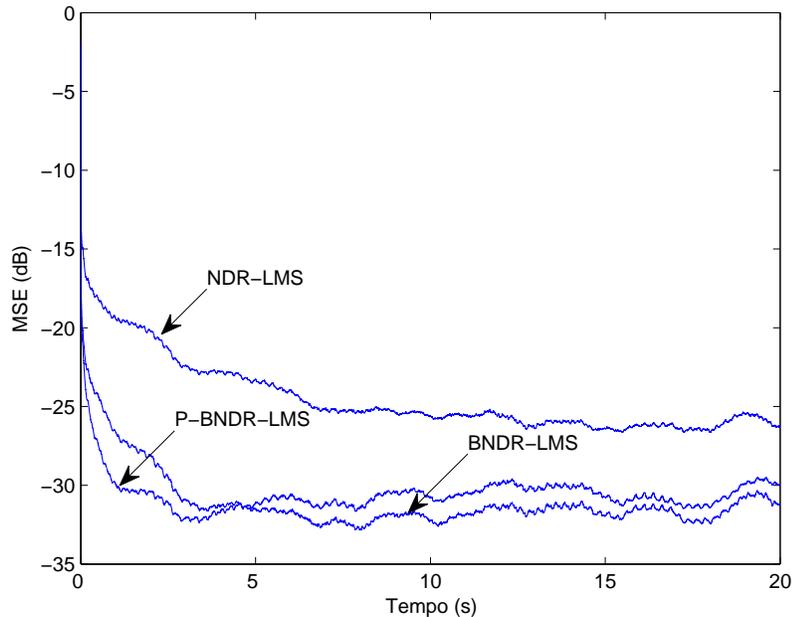


Figura 5.3: Erro médio quadrático dos algoritmos NDR-LMS, BNDR-LMS e P-BNDR-LMS.

Sem oferecer um grande aumento na complexidade computacional, o algoritmo BNDR-LMS apresenta um desempenho superior ao PNLMS e ao NDR-LMS, tanto em velocidade de convergência quanto em atenuação final. Apesar da incorporação da técnica de proporcionalidade resultar em um ganho na velocidade de convergência, a perda na atenuação final juntamente com o aumento de complexidade computacional não justifica a sua utilização.

Os desempenhos dos algoritmos adaptativos em um ambiente de cancelamento de eco acústico são influenciados pelo fato de que os sinais de voz são não-estacionários e correlacionados. A primeira característica requer rápida convergência dos algoritmos que, geralmente, está diretamente relacionada ao aumento do desajuste. Já a segunda característica é explorada pelos algoritmos que fazem reuso de dados. Isto explica o motivo pelo qual algoritmos adaptativos com maior desajuste, como BNDR-LMS e P-BNDR-LMS, tiveram desempenhos superiores aos algoritmos com baixo desajuste como o LMS.

Além disso, a variação da resposta ao impulso da sala de teleconferência durante as gravações, que será demonstrada na Subseção 5.3.2, ressalta ainda mais a importância de algoritmos adaptativos com alta velocidade de convergência. Outra importante consequência desta característica é que alguns algoritmos adaptativos não alcançaram o estado estacionário mesmo após 20 s de simulação, como pode ser notado nas Figuras 5.2 e 5.3.

Devido a não utilização de algoritmos robustos por causa da característica do ambiente citada acima, utilizou-se o sistema de filtragem *set-membership*, que considera o parâmetro  $\mathbf{w}(n)$  estimado aceitável se o erro  $e(n)$  associado for menor que um limite específico  $\gamma_0$ . Uma idéia semelhante é utilizada pelos algoritmos robustos, onde ocorre a diminuição do passo do algoritmo se a razão do erro pelo fator de escalamento  $s(n)$  for menor que a constante  $k_0$ . Diferentemente, neste caso o mecanismo *set-membership* não atualiza os coeficientes do filtro. A Figura 5.4 exibe o desempenho do algoritmo SM-BNDR-LMS com o limite do erro  $\gamma_0 = 1$ .

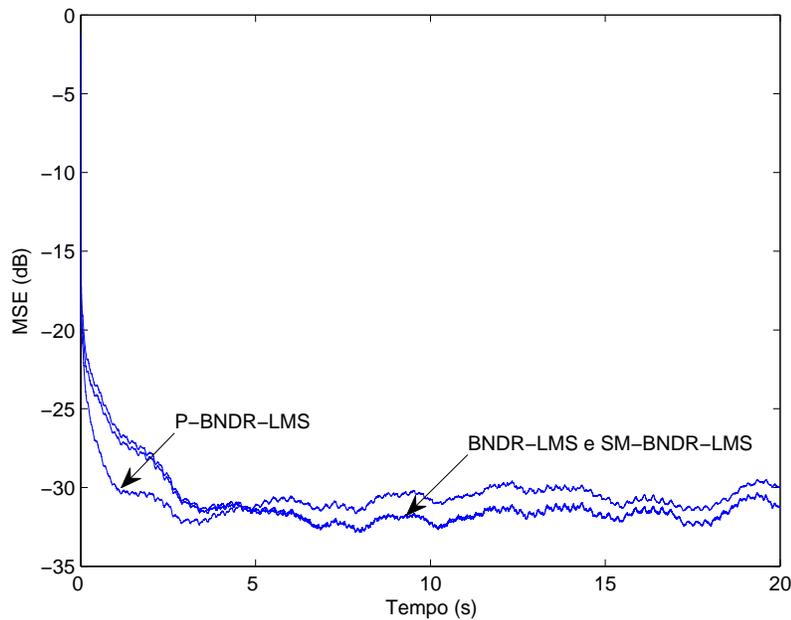


Figura 5.4: Erro médio quadrático dos algoritmos BNDR-LMS, SM-BNDR-LMS e P-BNDR-LMS.

Como demonstrado em [20], a utilização do mecanismo *set-membership* juntamente com o algoritmo BNDR-LMS [21] é uma alternativa viável para reduzir a complexidade sem afetar significativamente o desempenho em sistemas de cancelamento de eco. Isto é obtido variando-se o limite específico  $\gamma_0$  do erro como exibido na Figura 5.5, onde são utilizados os valores de  $\gamma_0$  exibidos na Tabela 5.2 com as suas respectivas complexidades computacionais.

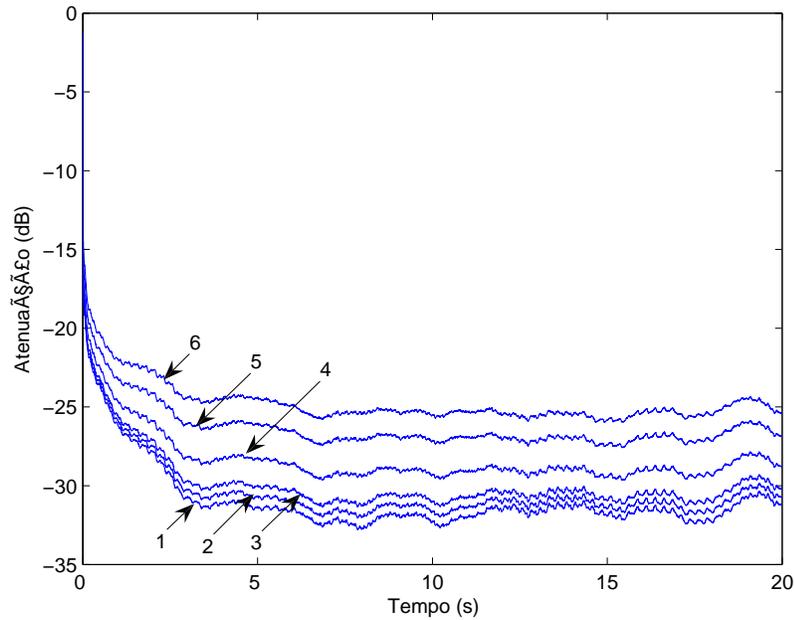


Figura 5.5: Erro médio quadrático do SM-BNDR-LMS com os valores de  $\gamma_0$  mostrados na Tabela 5.2.

Com a utilização do mecanismo *set-membership* consegue-se reduzir a complexidade computacional média do algoritmo BNDR-LMS em até 75%, apresentando uma perda na atenuação final menor que 2 dB. Em um caso mais extremo, uma redução em até 31% da complexidade resulta em um custo computacional semelhante ao do NLMS porém com um desempenho consideravelmente superior.

Tabela 5.2: Valores de  $\gamma_0$  utilizados no algoritmo SM-BNDR-LMSa e seus respectivos percentuais de atualização.

MSE	$\gamma_0$	%
1	1	97,93
2	10	84,77
3	20	75,80
4	50	59,05
5	100	42,62
6	150	31,60

### 5.3.2 Conversação Cruzada

Durante os testes em situações de conversação cruzada, um comportamento similar apresentado por todos os algoritmos foi identificado. Espera-se que, durante uma interrupção na adaptação dos coeficientes do filtro, o MSE mantenha-se no mesmo nível onde encontrava-se antes da interrupção. No entanto, percebeu-se nas simulações que sempre havia uma piora significativa do MSE durante o período de conversação cruzada, e que isto não estava sendo causado por erros de detecção do DTD mas, sim, por simplesmente haver uma interrupção na adaptação dos coeficientes do filtro. Este comportamento é ilustrado na Figura 5.6 onde os algoritmos NLMS e BNDR-LMS têm sua adaptação interrompida sem a inclusão do sinal de voz do locutor próximo  $s(n)$ .

Um comportamento similar aos apresentados pelos algoritmos é obtido fazendo os coeficientes da resposta ao impulso da sala de teleconferência variarem lentamente no tempo. Durante a convolução entre o sinal do locutor distante  $\mathbf{x}(n)$  e a resposta ao impulso  $\mathbf{h}$  da sala de teleconferência, ao adicionar-se ruído branco a  $\mathbf{h}$  a cada iteração, por meio da função *randn* do Matlab dividida por uma constante da ordem de  $10^4$ , observou-se que o mesmo comportamento era encontrado. A Figura 5.7 ilustra esta situação utilizando uma razão eco-ruído de 40 dB.

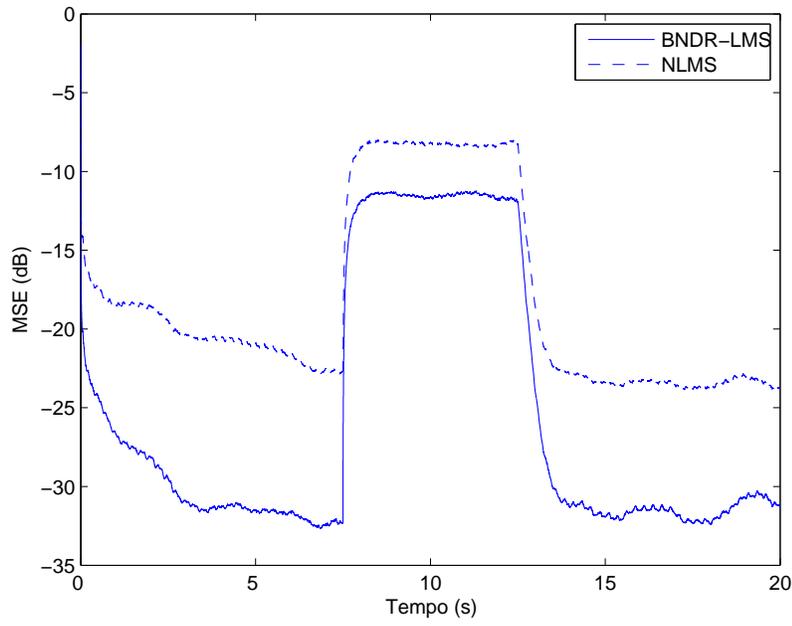


Figura 5.6: Comportamento do erro médio quadrático durante interrupção na adaptação dos coeficientes.

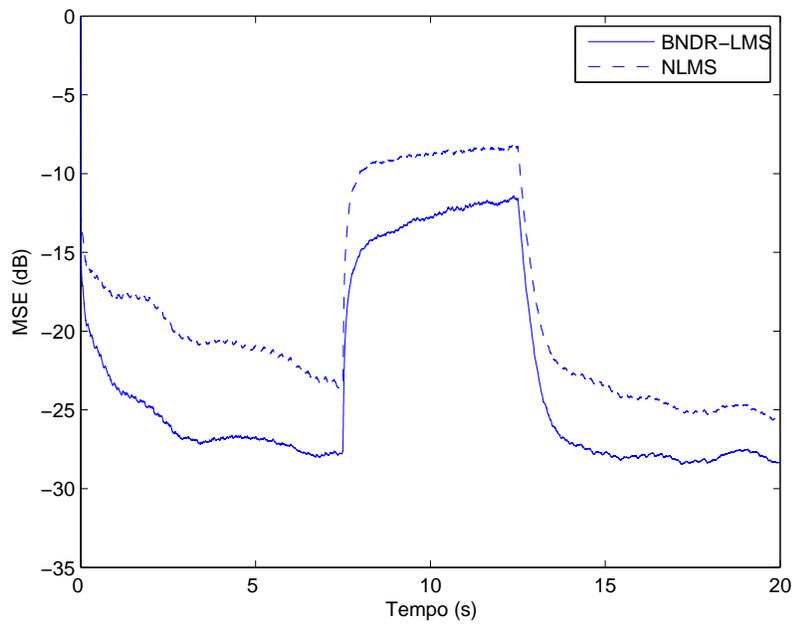


Figura 5.7: Comportamento do erro médio quadrático durante pausa na adaptação.

Na prática, este comportamento pode ser justificado por a sala de teleconferência não possuir preparo acústico e ser limitada, em parte, por divisórias, sendo sujeita a interferências externas. Esta característica torna ineficiente a utilização de DTD uma vez que, ao proteger o filtro adaptativo interrompendo a adaptação dos coeficientes em uma situação de conversação cruzada, o filtro deixa de acompanhar as variações do percurso do eco resultando em um aumento excessivo do eco no sinal a ser transmitido.

Uma outra consequência é o comprometimento da utilização de algoritmos robustos justificada pelo fato de que, em um ambiente com esta característica, os algoritmos adaptativos encontram-se sempre em uma situação de re-convergência, rastreando as variações do percurso de eco mesmo que sejam lentas. Como exemplificado no Capítulo 4, os algoritmos adaptativos robustos apresentam como característica uma redução na velocidade de re-convergência por basearem-se pela variável de escalamento  $s(n)$ , a qual é responsável por rastrear de maneira lenta  $e(n)$ . Assim, os algoritmos robustos não conseguem obter um bom desempenho, atenuando o eco de uma maneira menos eficiente que os algoritmos convencionais.

Ainda assim, a utilização de algoritmos robustos não evitaria o aumento da energia do eco a ser transmitido uma vez que isto não ocorreria devido às falhas de detecção de conversação cruzada e, sim, devido à interrupção da atualização dos coeficientes do filtro que aconteceria mesmo utilizando algoritmos adaptativos robustos.

## 5.4 Conclusões

Este capítulo apresentou simulações de cancelamento de eco acústico mono-canal utilizando sinais gravados em tempo real. Uma estimativa da resposta ao impulso da sala de teleconferência foi realizada para definir a ordem do sistema a ser utilizada. Os parâmetros de cada algoritmo adaptativo foram estimados por meio de uma busca extensiva objetivando minimizar o MSE resultante em uma situação de convergência num intervalo de tempo de 20 s.

Os resultados mostraram que, apesar de possuir um maior desajuste, o algoritmo adaptativo BNDR-LMS obteve o melhor desempenho entre os algoritmos

avaliados. Isto é justificado por este algoritmo reutilizar dados para aumentar a velocidade de convergência em situações onde o sinal de entrada é correlacionado, sendo esta uma das principais características dos sinais de voz. Além disso, a demonstrada variação da resposta ao impulso da sala de teleconferência durante as gravações, que fez com que os algoritmos estivessem sempre em uma situação de re-convergência, ressaltou ainda mais a necessidade de algoritmos adaptativos com alta velocidade de convergência.

Esta característica da sala de teleconferência ainda resultou em uma transmissão excessiva de eco em situações de conversação cruzada por haver, neste caso, uma interrupção na adaptação dos coeficientes do filtro, fazendo com que estes não acompanhassem as variações do percurso do eco. Com isso, a utilização de algoritmos robustos tornou-se injustificável além de ineficiente, por estes caracterizarem-se por terem uma baixa velocidade de re-convergência devido ao modo como são derivados.

Por fim, a utilização do sistema de filtragem *set-membership* mostrou-se eficiente ao reduzir a complexidade computacional média do algoritmo BNDR-LMS em quase 25% sem afetar de maneira drástica o seu desempenho.

# Capítulo 6

## Cancelamento de Eco Acústico

### Multi-Canal

#### 6.1 Introdução

Com o aumento do uso dos sistemas de teleconferência, houve uma necessidade de desenvolvimento de sistemas mais sofisticados e realistas. Uma das mudanças implementadas foi o realismo espacial, desejado em comunicações de áudio e vídeo entre grupos de pessoas. Para isso, foi necessário levar em consideração áudio multi-canal, no qual envolve no mínimo 2 canais, isto é, som estéreo.

Este capítulo aborda o tema cancelamento de eco acústico multi-canal em uma sala de teleconferência utilizando os algoritmos adaptativos descritos no Capítulo 2. A Seção 6.2 detalha matematicamente o assunto enfatizando o problema da não-unicidade da solução ótima que, diferentemente do caso mono-canal, não implica no total alinhamento entre a resposta ao impulso  $\mathbf{w}$  do filtro adaptativo e a resposta ao impulso  $\mathbf{h}$  da sala de teleconferência, ou seja, em  $\mathbf{w} = \mathbf{h}$ . Ainda nesta seção, uma solução encontrada na literatura para este problema é apresentada e é proposta a utilização de codificadores de voz como método de descorrelação. Os métodos têm seus desempenhos avaliados na Seção 6.4.

Na Seção 6.5 são apresentados os resultados obtidos em um ambiente de cancelamento de eco acústico multi-canal nas situações de convergência e de conversação cruzada. A Seção 6.6 resume o presente capítulo enfatizando as principais contribuições.

## 6.2 Problema da Não-Unicidade da Solução

Um diagrama esquemático da aplicação de cancelamento de eco acústico estéreo em teleconferência é mostrado na Figura 6.1. No lado direito encontra-se a sala de transmissão contendo dois microfones, os quais são responsáveis pela captura de sinais da fala do locutor por meio de dois percursos acústicos que são caracterizados pelas respostas ao impulso  $\mathbf{g}_1$  e  $\mathbf{g}_2$ . Esses sinais estéreos são, então, transmitidos para as caixas de som na sala de recepção a esquerda, que são acopladas a um dos microfones por meio dos percursos indicados pelas respostas ao impulso  $\mathbf{h}_1$  e  $\mathbf{h}_2$ , produzindo o sinal de saída  $d(n)$ . O mesmo ocorre para o outro microfone na sala de recepção, mas por simplicidade, apenas o cancelamento de eco para um dos microfones será discutido, sendo o outro similar. Se nada for feito, o sinal  $d(n)$  contendo eco capturado pelo microfone chegará às caixas de som na sala de transmissão.

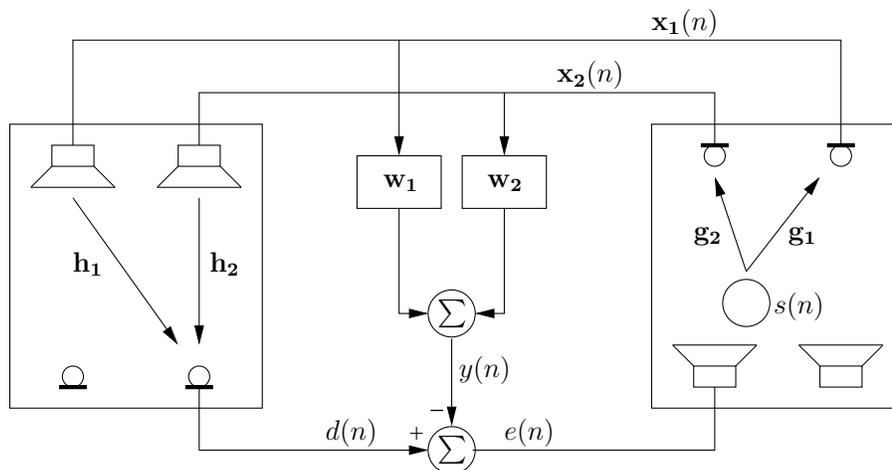


Figura 6.1: Cancelamento de eco acústico multi-canal.

Assim como no caso mono-canal, o cancelamento de eco acústico estéreo é realizado por meio de uso de filtros adaptativos FIR que modelam a resposta ao impulso do percurso do eco. A Figura 6.1 generaliza essa idéia usando dois filtros adaptativos  $\mathbf{w}_1$  e  $\mathbf{w}_2$  para modelar os dois percursos de eco  $\mathbf{h}_1$  e  $\mathbf{h}_2$  na sala de recepção [22]. Utilizando os sinais  $\mathbf{x}_1(n)$  e  $\mathbf{x}_2(n)$ , os filtros produzem a estimativa  $y(n)$  que é subtraída do sinal de eco  $d(n)$ , formando o sinal de erro  $e(n)$ , o qual se deseja que seja o menor possível na ausência de um locutor próximo.

O eco  $d(n)$  e o sinal de erro  $e(n)$  podem ser expressos, respectivamente, como

$$d(n) = \mathbf{h}_1^T \mathbf{x}_1(n) + \mathbf{h}_2^T \mathbf{x}_2(n), \quad (6.1)$$

$$e(n) = d(n) - \mathbf{w}_1^T \mathbf{x}_1(n) - \mathbf{w}_2^T \mathbf{x}_2(n), \quad (6.2)$$

onde  $\mathbf{w}_1$  e  $\mathbf{w}_2$  são os vetores de tamanho  $N$  dos coeficientes dos filtros adaptativos e

$$\mathbf{x}_1(n) = [x_1(n) \ x_1(n-1) \ x_1(n-2) \ \cdots \ x_1(n-N+1)]^T, \quad (6.3)$$

$$\mathbf{x}_2(n) = [x_2(n) \ x_2(n-1) \ x_2(n-2) \ \cdots \ x_2(n-N+1)]^T. \quad (6.4)$$

Assim, o sinal de erro pode ser escrito de maneira mais compacta

$$e(n) = d(n) - \mathbf{w}^T \mathbf{x}(n), \quad (6.5)$$

onde  $\mathbf{w} = [\mathbf{w}_1^T | \mathbf{w}_2^T]^T$  é a concatenação de  $\mathbf{w}_1$  e  $\mathbf{w}_2$ , assim como,  $\mathbf{x}(n) = [\mathbf{x}_1^T(n) | \mathbf{x}_2^T(n)]^T$ .

Dessa maneira, o sinal  $d(n)$  pode ser escrito como

$$d(n) = \mathbf{h}_1^T \mathbf{x}_1(n) + \mathbf{h}_2^T \mathbf{x}_2(n) = \mathbf{h}^T \mathbf{x}(n). \quad (6.6)$$

Em termos de  $\mathbf{h}$ , a equação (6.5) pode ser re-escrita como

$$e(n) = \mathbf{h}^T \mathbf{x}(n) - \mathbf{w}^T \mathbf{x}(n) = (\mathbf{h}^T - \mathbf{w}^T) \mathbf{x}(n) = \tilde{\mathbf{w}}^T \mathbf{x}(n). \quad (6.7)$$

Para que o eco seja totalmente cancelado, deve-se ter

$$\tilde{\mathbf{w}}_1^T * \mathbf{x}_1(n) + \tilde{\mathbf{w}}_2^T * \mathbf{x}_2(n) = 0. \quad (6.8)$$

Para o caso de apenas um locutor como exibido na Figura 6.1

$$[\tilde{\mathbf{w}}_1^T * \mathbf{g}_1 + \tilde{\mathbf{w}}_2^T * \mathbf{g}_2] * \mathbf{s}(n) = 0, \quad (6.9)$$

onde  $s(n)$  é o sinal gerado pelo locutor na sala de transmissão. No domínio da frequência, a equação (6.9) torna-se

$$[\tilde{W}_1(e^{j\omega})G_1(e^{j\omega}) + \tilde{W}_2(e^{j\omega})G_2(e^{j\omega})]S(e^{j\omega}) = 0, \quad (6.10)$$

onde as transformadas discretas de Fourier das funções no tempo são denotadas pelas respectivas letras maiúsculas. Considere primeiramente o caso mono-canal, ou seja,  $G_2(e^{j\omega}) = 0$ . Neste caso, desconsiderando os zeros de  $G_1(e^{j\omega})S(e^{j\omega})$ , a equação (6.9) ocorre quando  $\tilde{W}_1(e^{j\omega}) = 0$ . Assim, o completo alinhamento ( $\mathbf{w} = \mathbf{h}$ )

é alcançado garantindo que  $G_1(e^{jw})S(e^{jw})$  não seja zero em nenhuma frequência. No caso estéreo, mesmo que  $S(e^{jw})$  não possua zeros na região de frequências de interesse, o melhor que pode ser obtido é

$$\tilde{W}_1(e^{jw})G_1(e^{jw}) + \tilde{W}_2(e^{jw})G_2(e^{jw}) = 0. \quad (6.11)$$

Para esta equação existem infinitas soluções não implicando em  $\tilde{W}_1(e^{jw}) = \tilde{W}_2(e^{jw}) = 0$ , e conseqüentemente em  $\mathbf{w} = \mathbf{h}$ , que é a condição de completo alinhamento. O problema de canceladores de eco estéreo é aparente na equação (6.10): mesmo que as respostas ao impulso da sala de recepção  $\mathbf{h}_1$  e  $\mathbf{h}_2$  sejam fixadas, qualquer variação em  $G_1(e^{jw})$  e  $G_2(e^{jw})$  requer ajustes de  $\tilde{W}_1(e^{jw})$  e  $\tilde{W}_2(e^{jw})$ , exceto na condição improvável  $\tilde{W}_1(e^{jw}) = \tilde{W}_2(e^{jw}) = 0$ . Então, o algoritmo adaptativo não deve apenas rastrear as variações na sala de recepção mas também as variações na sala de transmissão. As últimas variações são particularmente difíceis de serem rastreadas; se um locutor pára de falar e um outro começa a falar em um lugar diferente, as respostas ao impulso  $\mathbf{g}_1$  e  $\mathbf{g}_2$  mudam bruscamente. Com isso, o grande desafio é desenvolver um algoritmo que convirja independentemente das variações na sala de transmissão, como no caso de eco mono-canal.

Entretanto, através da equação (6.8) percebe-se que  $e(n) = 0$  implica em  $\mathbf{w} = \mathbf{h}$  somente se não houver uma relação linear entre os sinais  $x_1(n)$  e  $x_2(n)$ . Então, o problema é encontrar um método para descorrelacionar  $x_1(n)$  e  $x_2(n)$  sem afetar a qualidade e a percepção estéreo dos mesmos.

Algumas técnicas para descorrelacionar os sinais  $x_1(n)$  e  $x_2(n)$  foram desenvolvidas nos últimos anos. A primeira idéia foi adicionar ruído aleatório independente em cada canal [23]

$$x'_i(n) = x_i(n) + v_i(n), \quad i = 1, 2, \quad (6.12)$$

onde  $v_1(n)$  e  $v_2(n)$  são dois ruídos brancos independentes. Uma segunda idéia foi modular cada canal com ruído aleatório independente [24]

$$x'_i(n) = [1 + \varepsilon_i(n)]x_i(n), \quad i = 1, 2, \quad (6.13)$$

onde

$$\varepsilon_i(n) = \alpha\varepsilon_i(n-1) + (1-\alpha)v_i(n) \quad (6.14)$$

e  $v_1(n)$  e  $v_2(n)$  são dois ruídos brancos independentes novamente. Entretanto, nenhuma dessas técnicas apresentou resultados satisfatórios pois degradavam muito a qualidade dos sinais. Tentando evitar esta degradação, [25] teve a idéia de adicionar a cada canal algo parecido com o próprio sinal. Assim, surgiu a técnica de acrescentar aos sinais uma função não-linear dos próprios sinais

$$x'_i(n) = x_i(n) + \alpha f[x_i(n)], \quad i = 1, 2, \quad (6.15)$$

onde a função  $f$  deve ser não-linear para evitar uma relação linear entre  $x'_1(n)$  e  $x'_2(n)$ . Dentre várias funções testadas, a que apresentou a melhor relação decorrelação/qualidade foi a função retificadora de meia onda [25, 26]

$$x'_i(n) = x_i(n) + \alpha \left( \frac{x_i(n) + |x_i(n)|}{2} \right), \quad i = 1, 2. \quad (6.16)$$

Uma versão melhorada desta técnica utiliza retificadores de meia onda positivo e negativo em cada canal respectivamente,

$$\begin{aligned} x'_1(n) &= x_1(n) + \alpha \left( \frac{x_1(n) + |x_1(n)|}{2} \right), \\ x'_2(n) &= x_2(n) + \alpha \left( \frac{x_2(n) - |x_2(n)|}{2} \right). \end{aligned} \quad (6.17)$$

A referência [26] mostra que, para sinais codificados em PCM, a percepção estéreo não é alterada por este método mesmo para  $\alpha = 0,4$  e que a qualidade dos sinais presentes nos canais não é facilmente percebida devido à natureza do sinal de voz e o efeito de mascaramento psicoacústico. Isto é explicado pelo fato que a periodicidade do sinal permanece inalterada e que durante períodos de silêncios nenhuma distorção é inserida. Além disso, a estrutura harmônica dos sinais acabam mascarando as distorções harmônicas inseridas.

Algoritmos que utilizam efetivamente o modelo psicoacústico humano para que distorções inseridas por eles não sejam percebidas são os codificadores de voz. A compressão com perdas realizadas pelos codificadores é feita de modo que, após a decodificação, o sinal resultante seja perceptivelmente parecido com o original. Nesta dissertação, a influência dos mais atuais codificadores de voz para sinais de banda larga (8 kHz) na decorrelação dos canais de um ambiente de teleconferência estéreo é estudada.

Os codificadores de voz a serem analisados serão o G.729.1 [27] e o Speex [28]. O G.729.1 é um codificador padronizado pela ITU para sinais de voz com taxa de

amostragem de 16 kHz. O Speex é um codificador com código livre apto a utilizar sinais de voz com taxas de amostragem até 32 kHz.

## 6.3 Métodos de Análise Qualitativa

### 6.3.1 Desalinhamento

Em um ambiente de cancelamento de eco acústico multi-canal é importante conhecer a diferença entre a solução do filtro adaptativo  $\mathbf{w} = [\mathbf{w}_1^T | \mathbf{w}_2^T]^T$  e a resposta ao impulso  $\mathbf{h} = [\mathbf{h}_1^T | \mathbf{h}_2^T]^T$  da sala de recepção devido às características do problema em questão descrito na Seção 6.2. Esta diferença, geralmente, é quantizada por

$$\epsilon = \frac{\|\mathbf{h} - \mathbf{w}\|}{\|\mathbf{h}\|}, \quad (6.18)$$

a qual é denominada desalinhamento normalizado e é utilizada como indicador da eficiência de métodos de descorrelação. No caso multi-canal, é possível ter um bom cancelamento de eco mesmo quando o desalinhamento for grande, ou seja, quando  $\mathbf{w}$  convergir para uma das soluções da equação (6.10) que não seja  $\mathbf{w} = \mathbf{h}$ . Entretanto, neste caso, o sistema de cancelamento de eco será muito sensível a variações em  $\mathbf{g}_1$  e  $\mathbf{g}_2$ , o que prejudicará o desempenho geral do sistema. O objetivo principal do cancelamento de eco acústico multi-canal é evitar este problema.

### 6.3.2 Função Coerência

Outra informação necessária para o caso multi-canal é quanto os sinais presentes nos dois canais são correlacionados. A referência [25] estabelece uma ligação entre o condicionamento da matriz de auto-correlação dos sinais de entrada  $x_1$  e  $x_2$  e a função coerência

$$\gamma(f) = \frac{S_{x_1x_2}(f)}{\sqrt{S_{x_1x_1}(f)S_{x_2x_2}(f)}}, \quad (6.19)$$

onde  $S_{xx}(f)$  é a densidade espectral de potência, definida como

$$S_{x_1x_2}(f) = \sum_{k=-\infty}^{\infty} E[x_1(n)x_2(n-k)] e^{-j2\pi fk} \quad (6.20)$$

$$= \sum_{k=-\infty}^{\infty} R_{x_1x_2}(k) e^{-j2\pi fk}. \quad (6.21)$$

Na prática, a função coerência é utilizada para medir a correlação cruzada entre os sinais de entrada e também como indicador da eficiência de métodos de descorrelação.

### 6.3.3 W-PESQ

Apesar de haver a necessidade de descorrelacionar os sinais presentes nos canais de um sistema de teleconferência multi-canal, o método de descorrelação não deve afetar significativamente a qualidade dos sinais. Por isso, uma medida importante a ser considerada é a degradação gerada pelo método de descorrelação.

A forma mais eficiente para medir a qualidade de sinais de voz é a realização de testes subjetivos, onde os sinais são expostos à pessoas e estas são questionadas sobre a qualidade dos mesmos. Estes testes subjetivos são padronizados em [29], a qual descreve cinco tipos de testes subjetivos e as condições de como a nota média de qualidade MOS (do inglês *Mean Opinion Score*) de cada sinal deve ser determinada.

As categorias mais usuais de testes subjetivos são a ACR (do inglês *Absolute Category Rating*) e a DCR (do inglês *Degradation Category Rating*). Na ACR, o avaliador é apresentado a um simples sinal, o sinal em teste, o qual deve avaliá-lo em uma escala de qualidade de 5 pontos de acordo com a Tabela 6.1. Na DCR, dois sinais são apresentados ao avaliador. Um deles é identificado como sinal de referência e o outro é o sinal em teste. O avaliador deve quantificar o nível percebido de degradação no sinal em teste comparando-o com o sinal de referência em uma escala de degradação de 5 pontos.

Tabela 6.1: Escala MOS para os testes ACR e DCR

Escala MOS	ACR	DCR
5	Excelente	Inaudível
4	Boa	Audível mas não incômoda
3	Regular	Ligeiramente incômoda
2	Pobre	Incômoda
1	Ruim	Muito incômoda

Este tipo de avaliação precisa de um grande número de pessoas e envolve um procedimento custoso em relação ao tempo. Então, métodos objetivos de ava-

liação de qualidade, que não dependem de avaliadores, começaram a ser desenvolvidos [30, 31]. Atualmente, esta área encontra-se muito desenvolvida sendo o PESQ [32] (do inglês *Perceptual Evaluation of Speech Quality*) o estado da arte em avaliação objetiva de qualidade de sinais de voz, projetado para sinais de voz com taxa de amostragem de 8 kHz. Uma versão banda larga deste algoritmo, para sinais com largura de banda de 50 até 7000 Hz, denominada W-PESQ [33] (do inglês *Wideband-PESQ*) encontra-se hoje disponível.

O algoritmo W-PESQ é um algoritmo de avaliação de sinais de voz com referência que apresenta uma correlação de 0.9355 com o MOS subjetivo [32]. Ele avalia o sinal de teste comparando-o com um sinal de referência que, usualmente, não apresenta degradação. Este algoritmo será utilizado para avaliar a quantidade de degradação que o método de descorrelação insere nos sinais. Vale ressaltar que a nota máxima dada pelo algoritmo W-PESQ, quando os sinais de referência e em teste são idênticos, é igual a 4,644.

## 6.4 Análises Preliminares

### 6.4.1 Configuração

Durante as simulações de um ambiente de cancelamento de eco acústico multi-canal foram utilizadas as respostas ao impulso mostradas na Figura 6.2.

As respostas ao impulso  $\mathbf{g}_1$  e  $\mathbf{g}_2$  da sala de transmissão são versões reamostradas para 16 kHz de taxa de amostragem das respostas ao impulso encontradas em [34]. Ruído é concatenado à cauda das respostas para extendê-las até o tamanho de 4000 amostras. Estas respostas foram obtidas realizando a medição a diferentes distâncias da caixa de som em uma sala de reunião acusticamente preparada e, originalmente, foram gravadas com uma taxa de amostragem de 32 kHz.

As respostas ao impulso  $\mathbf{h}_1$  e  $\mathbf{h}_2$  da sala de recepção são respostas ao impulso gravadas de um ambiente sem preparo acústico, sendo que  $\mathbf{h}_1$  foi obtida estando a caixa de som e microfone alinhados enquanto em  $\mathbf{h}_2$  desalinhados.

A referência [18] fornece sinais gravados em vários idiomas com uma taxa de amostragem de 16 kHz para validação do algoritmo W-PESQ. Cada idioma possui 47 sinais, sendo todos os sinais gravados por 4 locutores distintos. Cada sinal é

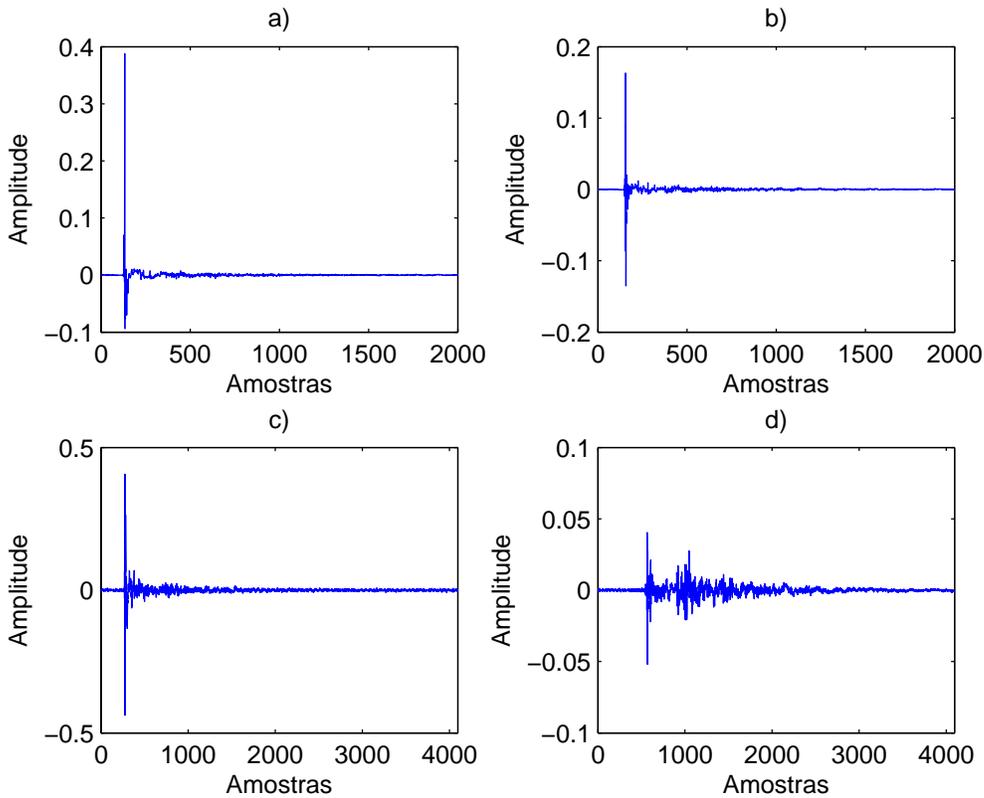


Figura 6.2: Respostas ao impulso: (a) $g_1$  e (b) $g_2$  da sala de transmissão; (c) $h_1$  e (d) $h_2$  da sala de recepção.

composto por duas frases separadas por um trecho de silêncio. Existem também silêncios no início e no fim de cada sinal de maneira que o sinal possua 8 s de duração.

Nas simulações de avaliação do desempenho dos métodos de decorrelação foram utilizados sinais do idioma inglês. Porém, os sinais foram concatenados de maneira que possuam no mínimo 16 s de duração e o excesso de silêncio foi retirado. Assim, nas simulações subsequentes foram utilizados 48 sinais de voz.

Com o objetivo de representar um caso real, utilizou-se o algoritmo BNDR-LMS com ordem do cada subfiltro  $N = 3000$ , para que este não fosse capaz de modelar toda a resposta ao impulso da sala de teleconferência. Os outros parâmetros utilizados foram  $\mu = 0,8$  e  $\gamma = 10^{15}$ . Devido aos diferentes ganhos aplicados aos sinais pelos métodos de decorrelação, nos cálculos da medida de desalinhamento os sinais decorrelacionados tiveram suas energias equiparadas com os sinais originais.

## 6.4.2 Resultados

O método de decorrelação utilizando a função retificadora de meia onda existente na literatura foi avaliado verificando o comportamento das três medidas de análise descritas na Subseção 6.3 perante a variação de  $\alpha$ . Como ilustração da quantidade de decorrelação causada pelo método, a Figura 6.3 compara a quantidade de correlação existente entre os sinais presentes nos canais quando nenhum tipo de algoritmo de decorrelação é aplicado e quando a função retificadora de meia onda é utilizada com  $\alpha = 0,2, 0,4$  e  $0,6$ .

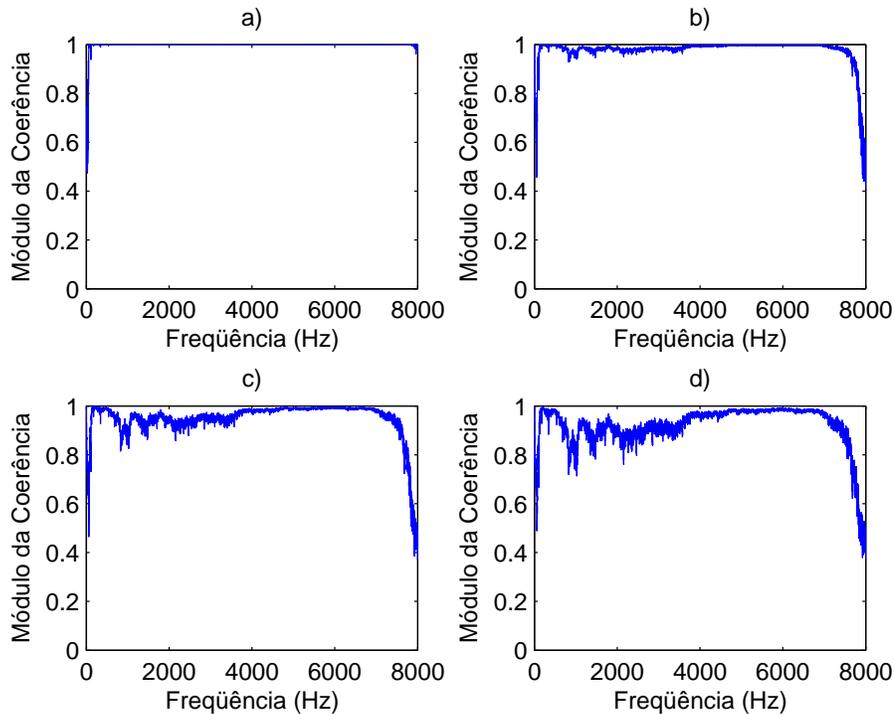


Figura 6.3: Módulo da coerência para diferentes níveis de decorrelação: (a)  $\alpha = 0$ , (b)  $\alpha = 0,2$ , (c)  $0,4$  e (d)  $0,6$ .

Nota-se que, como esperado, quanto mais não-linearidade é inserida nos canais maior a decorrelação entres eles. Além disso, apesar de haver uma decorrelação em toda faixa do espectro, esta é enfatizada na faixa de 500 à 4000 Hz por esta ser a faixa de frequência onde concentram-se os sinais de voz.

Da mesma maneira, a diminuição da quantidade de desalinhamento gerada pelo decorrelação dos canais devido à inserção de não-linearidade é comprovada pela Figura 6.4. Nela, verifica-se a eficiência do método de decorrelação em aproximar

a resposta ao impulso  $\mathbf{w}$  do filtro adaptativo à resposta ao impulso  $\mathbf{h}$  da sala de teleconferência. No entanto, os níveis de desalinhamento ainda ficam distantes do caso mono-canal, onde consegue-se obter desalinhamento de  $-20$  dB. Mas ainda assim, os resultados obtidos são superiores ao caso onde não é utilizado método de decorrelação dos canais.

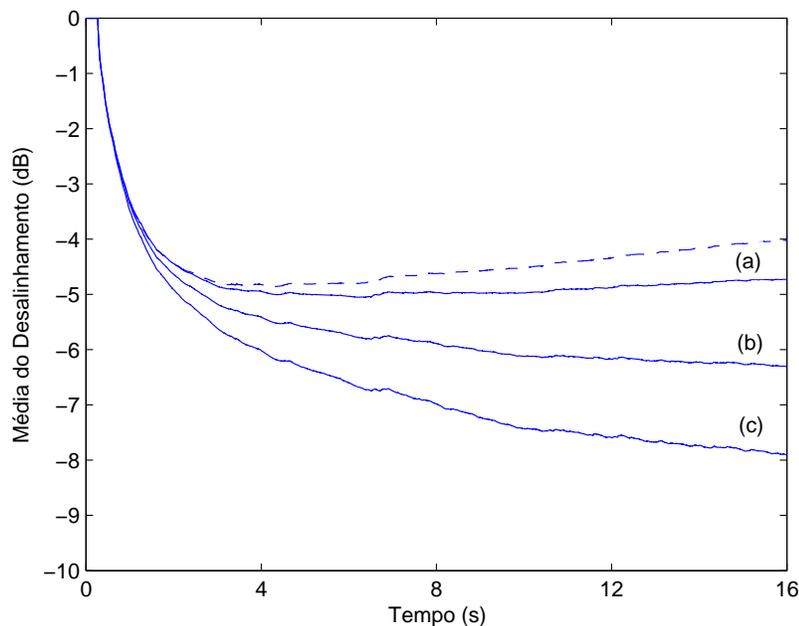


Figura 6.4: Desalinhamento utilizando diferentes níveis de decorrelação: (a)  $\alpha = 0,2$ , (b)  $0,4$  e (c)  $0,6$ . A linha tracejada refere-se ao caso sem decorrelação.

Entretanto, a decorrelação dos canais e, conseqüentemente, a melhora do desempenho do sistema é obtida sacrificando a qualidade dos sinais de voz. Com o intuito de avaliar a qualidade do sinal estéreo após o método de decorrelação, foram realizados dois tipos de experimentos utilizando o algoritmo W-PESQ. Em um primeiro caso, avaliou-se a qualidade dos sinais presentes em cada canal antes e após o método de decorrelação.

Porém, este tipo de avaliação não considera o efeito dos dois canais em conjunto, formando o sinal estéreo. Supondo o ouvinte posicionado simetricamente entre as caixas de som do sistema de teleconferência, pode-se considerar que o som final escutado pelo ouvinte é a soma dos sinais presentes em cada canal. Portanto, em um segundo experimento avaliou-se a qualidade da soma dos sinais originalmente presentes nos canais e após a decorrelação. A relação correlação *versus* qualidade

é mostrada para os casos de avaliação mono (linha contínua) e estéreo (linha tracejada) na Figura 6.5, enquanto a relação desalinhamento *versus* qualidade é exibida na Figura 6.6.

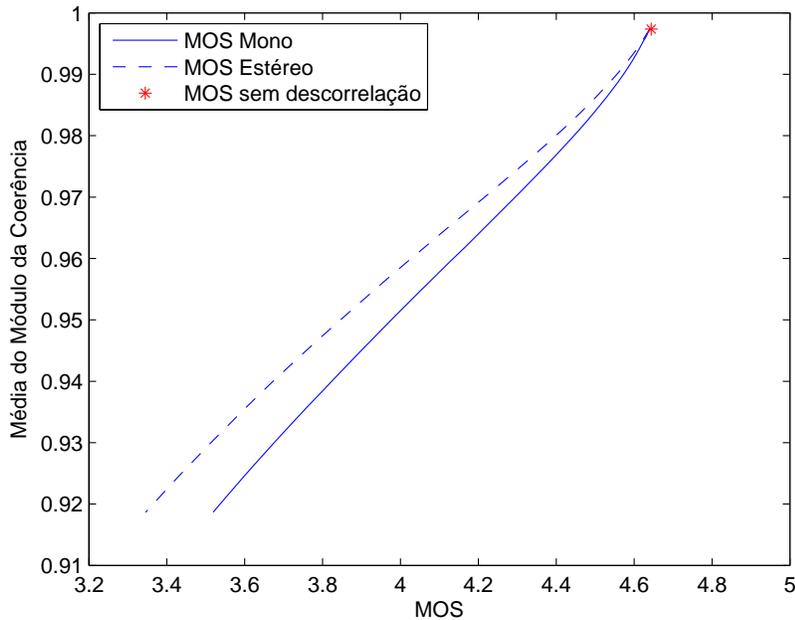


Figura 6.5: Relação entre a média do módulo da coerência e o MOS individual dos canais (linha contínua) e entre a média do módulo da coerência e o MOS conjunto dos canais (linha tracejada); o ponto em destaque indicado por \* é o caso sem descorrelação.

Por estas figuras, nota-se que no segundo experimento, onde tenta-se levar em conta a característica estéreo do sistema, a degradação da qualidade é menos sensível às intensas inserções de não-linearidade do que no caso onde os canais são avaliados individualmente. Uma provável explicação para este fato é que, devido à definição das respostas ao impulso  $\mathbf{g}_1$  e  $\mathbf{g}_2$  da sala de transmissão presente na Subseção 6.4.1, o sinal presente no canal 1 possui uma maior intensidade que o do canal 2. Isto faz que ao somar os canais, tanto a degradação do sinal encontrado no canal 2 quanto o próprio sinal sejam mascarados pela alta intensidade do sinal presente no canal 1.

Da mesma maneira, os codificadores de voz também tiveram seus desempenhos como método de descorrelação avaliados. O nível de descorrelação foi controlado pela taxa de codificação, a qual foi utilizada nos valores de 32, 26, 20 e 12 kbps.

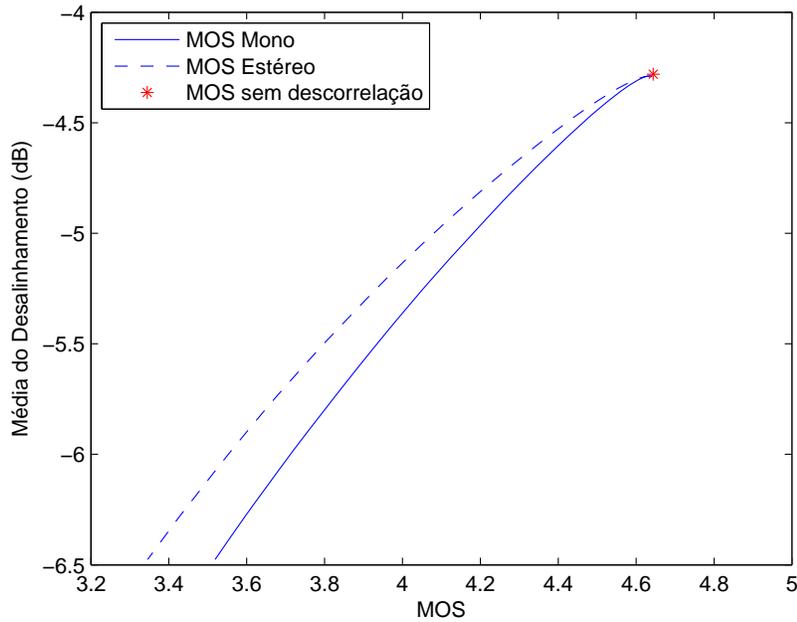


Figura 6.6: Relação entre a média do desalinhamento e o MOS individual dos canais (linha contínua) e entre a média do desalinhamento e o MOS conjunto dos canais (linha tracejada); o ponto em destaque indicado por \* é o caso sem descorrelação.

Para o codificador de voz Speex, a quantidade de correlação existente entre os sinais presentes nos canais para as taxas de codificação utilizadas é ilustrada na Figura 6.7.

Percebe-se que quanto menor a taxa de codificação, e conseqüentemente uma maior quantidade de degradação é inserida no sinal, menor é a correlação entre os canais. Porém, diferentemente do caso do método de descorrelação anterior onde a descorrelação concentrava-se nas baixas freqüências (até 4 kHz), o nível de descorrelação obtido pelo codificador Speex apresenta uma relação direta com a freqüência: quanto maior a faixa de freqüência maior a descorrelação inserida.

Uma explicação para este fato é que componentes de sinal nas freqüências mais altas são mais difíceis de serem percebidas pelo ouvido humano. O codificador de voz faz uso dessa informação para comprimir sinais de voz retirando, principalmente, informações contidas nesta faixa de espectro. Para estes níveis de correlação, o desalinhamento final do sistema é mostrado na Figura 6.8.

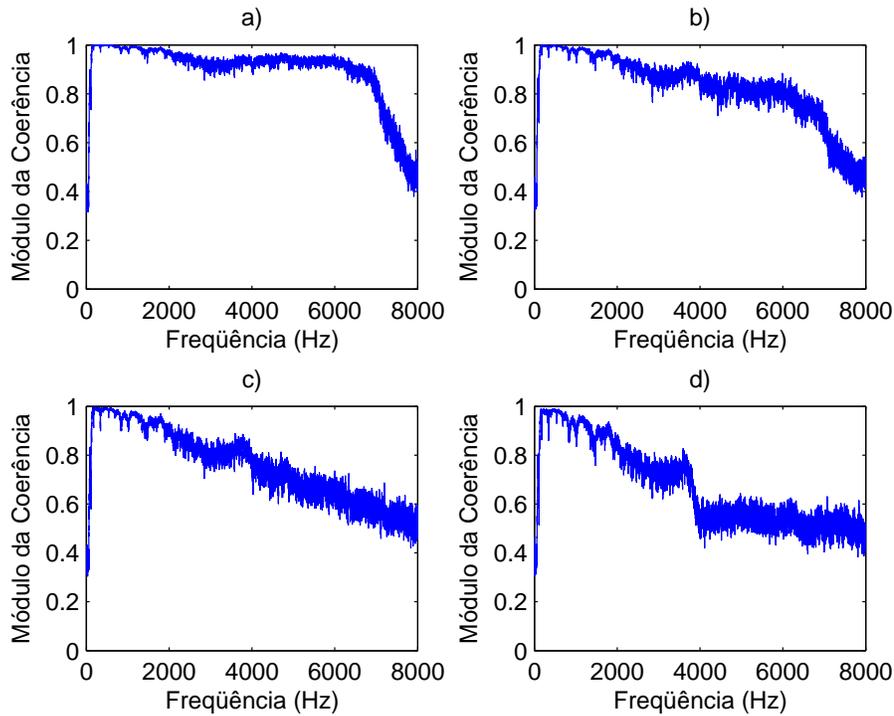


Figura 6.7: Módulo da coerência com o codificador Speex com diferentes taxas de codificação: (a)32, (b)26, (c)20 e (d)12 kbps.

O mesmo processo é repetido para o codificador de voz G.729.1. A Figura 6.9 ilustra a quantidade de correlação entre os canais para as diferentes taxas de codificação. Pelo mesmo motivo ocorrido com o Speex, a quantidade de descorrelação inserida concentra-se nas altas frequências. Entretanto, neste caso, o aumento do nível de descorrelação em relação ao espectro de frequência não é tão comportado como o obtido pelo Speex. Já o desalinhamento resultante utilizando o codificador de voz G.729.1 como método de descorrelação é mostrado na Figura 6.10.

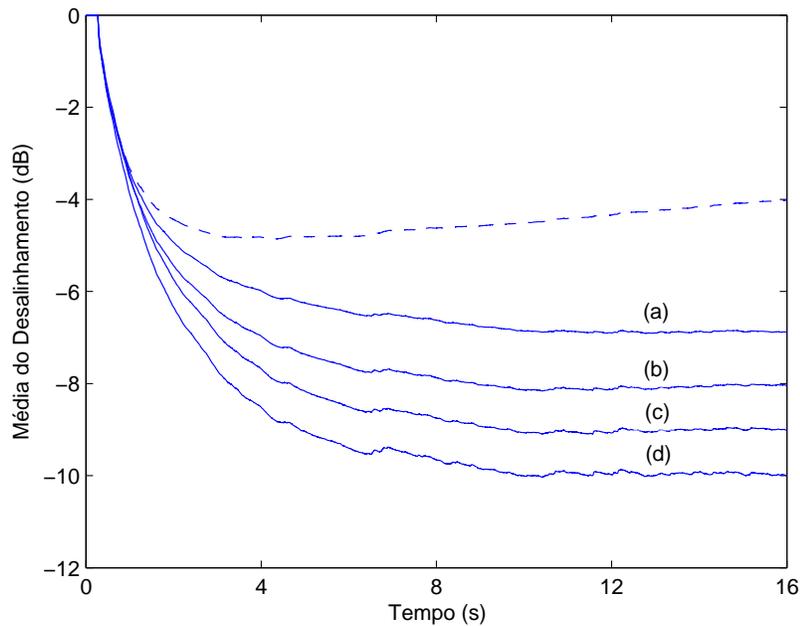


Figura 6.8: Desalinhamento com o codificador Speex para diferentes taxas de codificação: (a)32, (b)26, (c)20 e (d)12 kbps. A linha tracejada refere-se ao caso sem decorrelação.

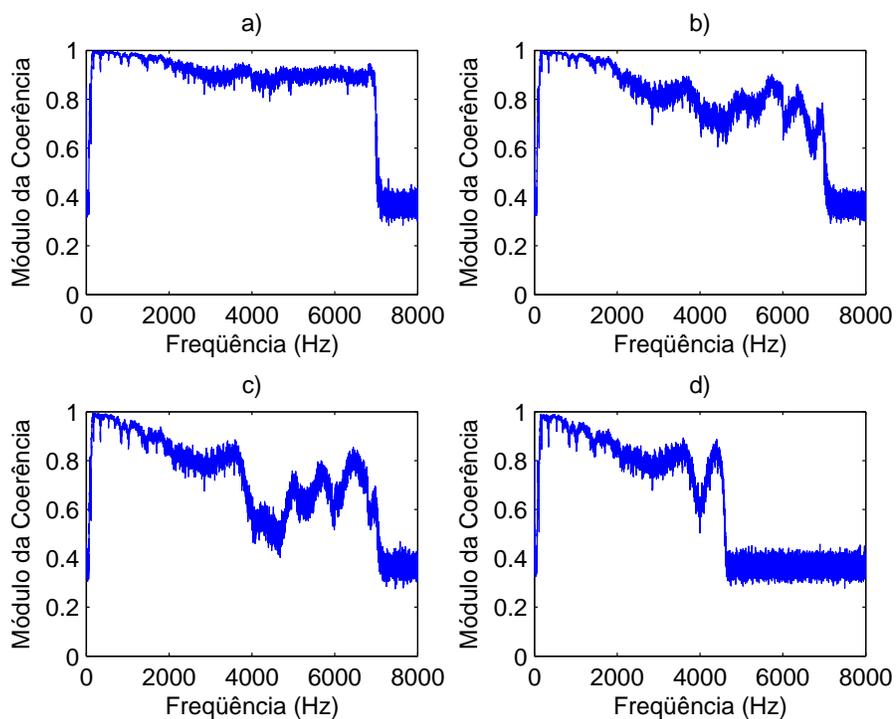


Figura 6.9: Módulo da coerência com o codificador G.729.1 para diferentes taxas de codificação: (a)32, (b)26, (c)20 e (d)12 kbps.

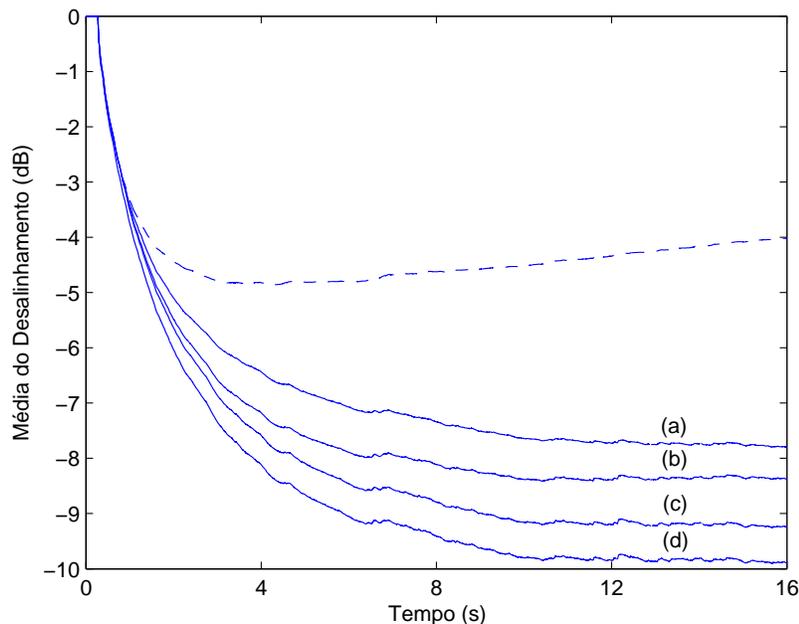


Figura 6.10: Desalinhamento com o codificador G.729.1 para diferentes taxas de codificação: (a)32, (b)26, (c)20 e (d)12 kbps. A linha tracejada refere-se ao caso sem descorrelação.

Como alternativa para melhorar o desempenho geral do sistema, cada canal foi codificado por um dos codificadores de voz. A idéia foi utilizar as diferentes características dos codificadores para aumentar a descorrelação entre os canais. Os níveis de correlação existentes entre os canais utilizando o codificador Speex no canal 1 e o codificador G.729.1 no canal 2, ambos com taxas de amostragem de 32, 26, 20 e 12 kbps, é exibida na Figura 6.11. Já a Figura 6.12 mostra o desalinhamento obtido utilizando a codificação híbrida como método de descorrelação.

Concluindo a análise da eficiência dos codificadores de voz como métodos de descorrelação dos canais de um ambiente de teleconferência estéreo, a Figura 6.13 exhibe a relação MOS estéreo *versus* correlação média entre os canais para os três esquemas de codificação propostos. Da mesma maneira, a Figura 6.14 exhibe a relação MOS estéreo *versus* desalinhamento médio. Nos dois casos, os resultados obtidos pelos codificadores de voz são comparados com os resultados obtidos com o método de adição de não-linearidade encontrado na literatura.

Por estas figuras, conclui-se que os codificadores de voz descorrelacionam os canais de maneira mais eficiente que o método atual existente na literatura,

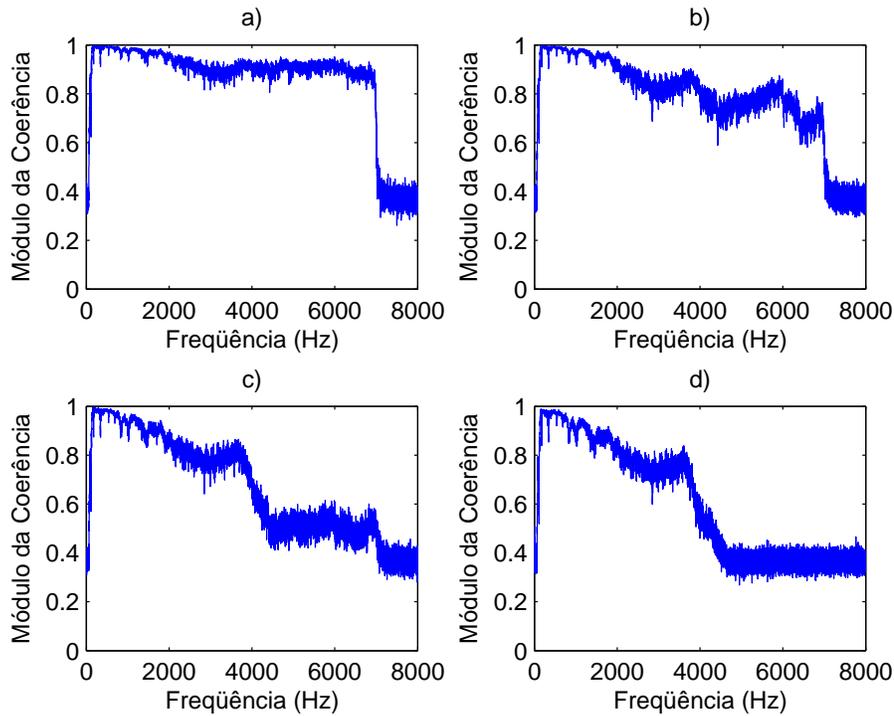


Figura 6.11: Módulo da coerência com a codificação híbrida para diferentes taxas de codificação: (a)32, (b)26, (c)20 e (d)12 kbps.

resultando em maior descorrelação e qualidade dos sinais. Entre os dois codificadores utilizados, de uma maneira geral, o codificador Speex apresentou um desempenho superior ao G.729.1, principalmente com taxa de codificação de 12 kbps.

Exceto pelo caso de taxa de amostragem de 32 kbps onde o Speex com 26 kbps apresenta um melhor desempenho, a utilização de um codificador em cada canal resultou em um menor desalinhamento sem alterar significativamente o MOS estéreo. Por exemplo, o processo de codificação híbrido com taxa de codificação de 26 kbps apresenta uma maior qualidade dos sinais que o uso dos codificadores com taxa de 20 kbps, mantendo o mesmo nível de desalinhamento. Além disso, o método de codificação híbrido mostrou-se a melhor opção para os casos onde aplica-se uma maior descorrelação.

Por fim, as Tabelas 6.2, 6.3, 6.4 e 6.5 resumem a avaliação dos métodos de descorrelação avaliados.

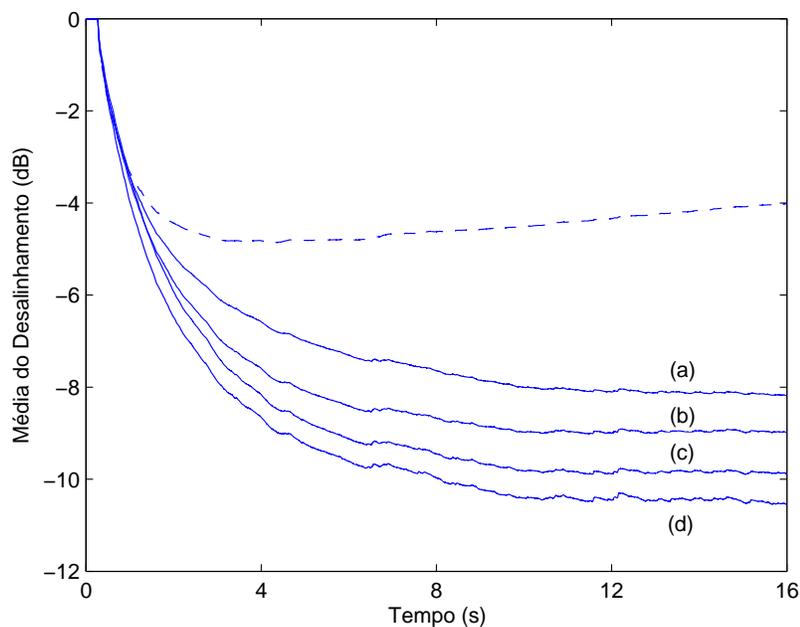


Figura 6.12: Desalinhamento com a codificação híbrida para diferentes taxas de codificação: (a)32, (b)26, (c)20 e (d)12. A linha tracejada refere-se ao caso sem descorrelação.

Tabela 6.2: Resultados da descorrelação para a função retificadora de meia onda com diferentes valores de  $\alpha$ .

$\alpha$	MOS mono	MOS estéreo	Correlação	Desalinhamento (dB)
0,2	4,3860	4,3246	0,9759	-4,6278
0,25	4,2799	4,1983	0,9691	-4,8143
0,3	4,1682	4,0648	0,9620	-5,0253
0,35	4,0518	3,9315	0,9548	-5,2537
0,4	3,9378	3,8013	0,9475	-5,4934
0,45	3,8262	3,6761	0,9402	-5,7393
0,5	3,7178	3,5591	0,9329	-5,9871
0,55	3,6152	3,4486	0,9257	-6,2334
0,6	3,5186	3,3454	0,9187	-6,4757

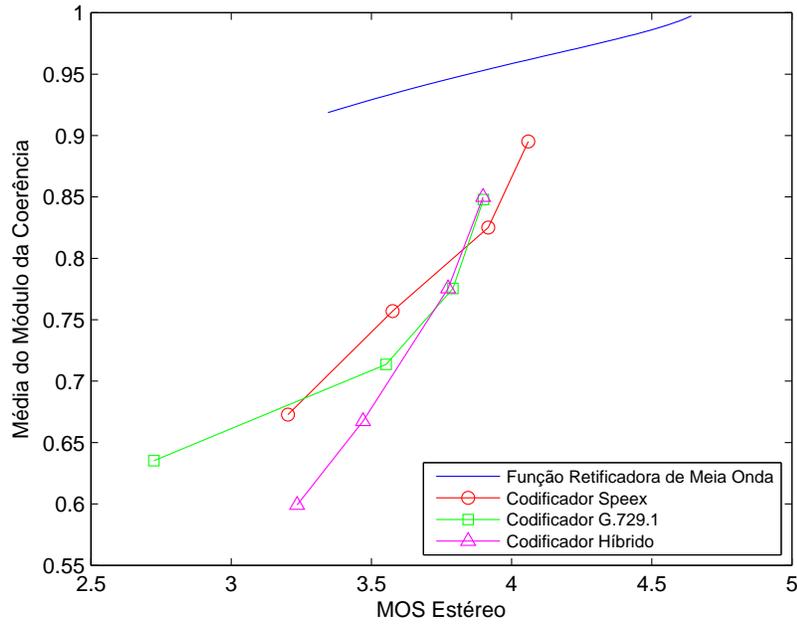


Figura 6.13: Relação entre a média do módulo da coerência e o MOS estéreo dos canais para os diferentes métodos de descorrelação: função retificadora de meia onda (linha contínua), codificador Speex (círculos), codificador G.729.1 (quadrados) e codificação híbrida (triângulos).

Tabela 6.3: Resultados da descorrelação utilizando o codificador Speex com diferentes taxas de codificação.

Taxa (kbps)	MOS mono	MOS estéreo	Correlação	Desalinhamento (dB)
32	4,1199	4,0602	0,8952	-6,0811
26	3,9817	3,9174	0,8251	-7,0992
20	3,6633	3,5760	0,7570	-7,8403
12	3,3091	3,2029	0,6727	-8,6624

Tabela 6.4: Resultados da descorrelação utilizando o codificador G.729.1 com diferentes taxas de codificação.

Taxa (kbps)	MOS mono	MOS estéreo	Correlação	Desalinhamento (dB)
32	3,9327	3,8999	0,8479	-6,6855
26	3,8423	3,7910	0,7755	-7,3063
20	3,6295	3,5525	0,7137	-7,8883
12	2,7652	2,7242	0,6352	-8,4405

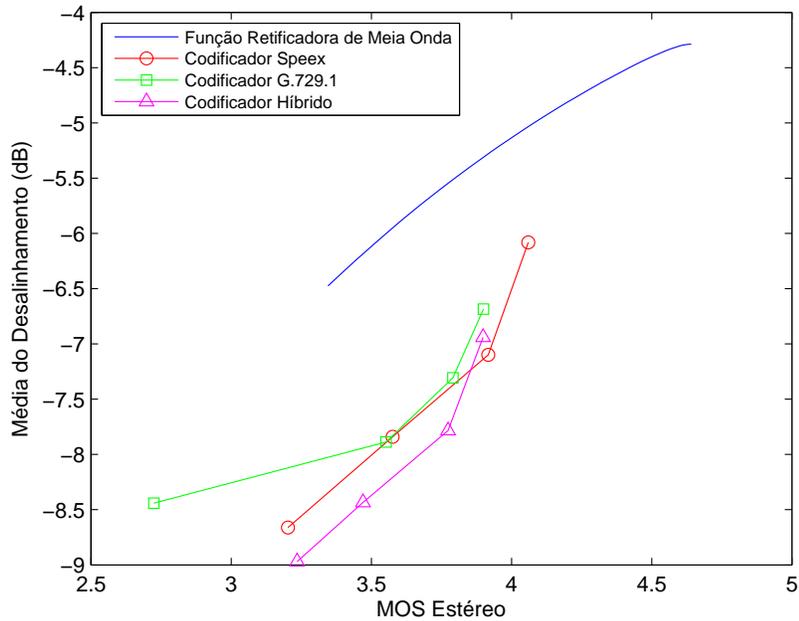


Figura 6.14: Relação entre a média do desalinhamento e o MOS estéreo dos canais para os diferentes métodos de descorrelação: função retificadora de meia onda (linha contínua), codificador Speex (círculos), codificador G.729.1 (quadrados) e codificação híbrida (triângulos).

Tabela 6.5: Resultados da descorrelação utilizando codificação híbrida com diferentes taxas de codificação.

Taxa (kbps)	MOS mono	MOS estéreo	Correlação	Desalinhamento (dB)
32	3,9863	3,8987	0,8499	-6,9403
26	3,8730	3,7739	0,7754	-7,7841
20	3,5895	3,4703	0,6673	-8,4352
12	2,9625	3,2350	0,5991	-8,9704

## 6.5 Algoritmos Adaptativos Multi-Canal

A derivação dos algoritmos adaptativos clássicos para o caso multi-canal é muito semelhante ao caso mono-canal. Sendo  $\mathbf{w} = [\mathbf{w}_1^T | \mathbf{w}_2^T]^T$  a concatenação de  $\mathbf{w}_1$  e  $\mathbf{w}_2$ , assim como,  $\mathbf{x}(n) = [\mathbf{x}_1^T(n) | \mathbf{x}_2^T(n)]^T$ , o algoritmo NLMS multi-canal atualiza os coeficientes do filtro segundo

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{\mathbf{x}_1^T(n)\mathbf{x}_1(n) + \mathbf{x}_2^T(n)\mathbf{x}_2(n)} e(n)\mathbf{x}(n) \quad (6.22)$$

onde  $\mu$  é o tamanho do passo do algoritmo. Do mesmo modo, os outros algoritmos adaptativos podem ser derivados para o caso multi-canal.

### 6.5.1 Configuração dos Experimentos

Assim como no caso mono-canal, como índice de desempenho dos algoritmos adaptativos foi utilizado o erro médio quadrático (MSE) do erro residual definido por

$$\text{MSE} = \frac{P_e}{P_d} = \frac{\text{LPF}[e^2(n)]}{\text{LPF}[d^2(n)]}, \quad (6.23)$$

onde LPF é um filtro passa-baixas com um simples pólo em 0,9996. Em todas as simulações subseqüentes foram utilizados 150 trechos de 20 s.

Os sinais utilizados como fonte sonora para a geração de eco multi-canal foram, do mesmo modo que o caso mono-canal, formados por concatenações de frases foneticamente balanceadas [19] originalmente gravadas com uma taxa de amostragem de 16 kHz. Estes sinais foram reproduzidos por uma caixa de som e simultaneamente gravados, em tempo real, por dois microfones posicionados simetricamente em relação à caixa de som. Estes sinais gravados referem-se aos sinais presentes nos dois canais de comunicação.

A geração dos sinais de eco multi-canal foi feita reproduzindo estes sinais de maneira estéreo e gravando simultaneamente, em tempo real, o eco capturado por um microfone. Este procedimento foi realizado em uma sala de teleconferência sem preparo acústico. Para facilitar as avaliações, o sinal de fala do locutor próximo foi sempre zero.

Nas simulações de cancelamento de eco multi-canal foram utilizados os parâmetros otimizados no caso mono-canal. Neste caso, cada subfiltro  $\mathbf{h}_1$  e  $\mathbf{h}_2$  possui comprimento igual a 4096, resultando em  $N = 8192$ .

### 6.5.2 Convergência

As Figuras 6.15 e 6.16 representam a convergência média para  $N = 8192$  de cada algoritmo adaptativo, cujos principais parâmetros foram determinados empiricamente minimizando o MSE resultante de cada algoritmo no caso mono-canal.

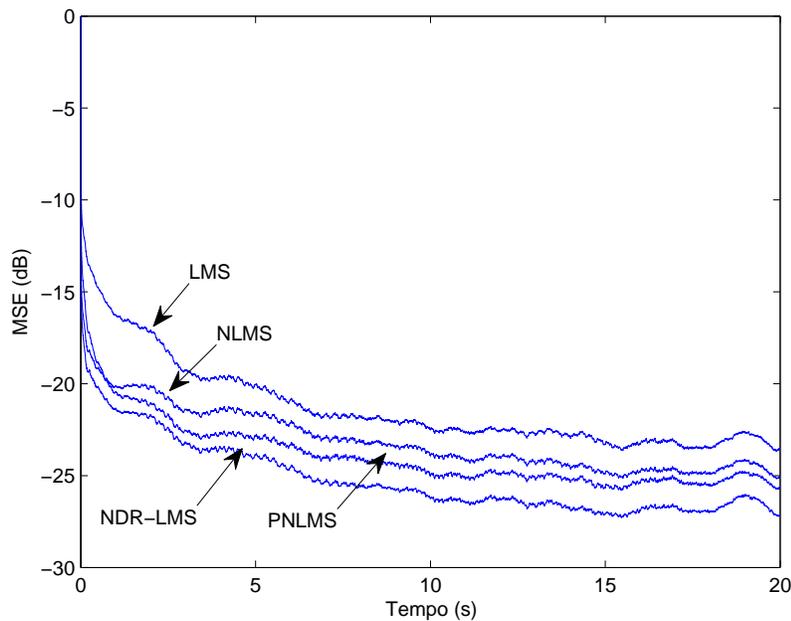


Figura 6.15: Erro médio quadrático dos algoritmos LMS, NLMS, PNLMS e NDR-LMS.

Os algoritmos adaptativos apresentaram um comportamento similar ao caso mono-canal, com destaque para os algoritmos BNDR-LMS. O bom desempenho dos algoritmos com alto desajuste é explicado pela variação da resposta ao impulso da sala de teleconferência, que será demonstrada mais adiante, a qual requer algoritmos com rápida velocidade de convergência. Tal característica da sala torna ineficiente a utilização de algoritmos robustos.

Repetindo o caso mono-canal, utilizou-se o mecanismo *set-membership* juntamente ao algoritmo BNDR-LMS. A Figura 6.17 compara o desempenho do algoritmo

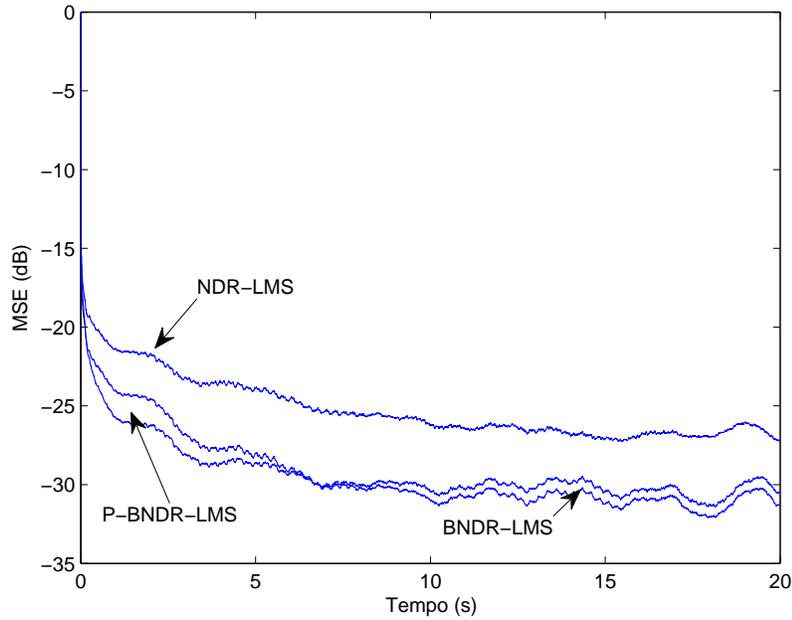


Figura 6.16: Erro médio quadrático dos algoritmos NDR-LMS, BNDR-LMS e P-BNDR-LMS.

SM-BNDR-LMS utilizando  $\gamma_0 = 10^{-1}$  com os algoritmos BNDR-LMS e P-BNDR-LMS. Para este valor de  $\gamma_0$ , o algoritmo SM-BNDR-LMS apresenta um desempenho bem semelhante ao BNDR-LMS e em alguns momentos, até superior. Com os valores de  $\gamma_0$  exibidos na Tabela 6.6 juntamente com os seus respectivos percentuais de atualização, o algoritmo SM-BNDR-LMS apresenta os desempenhos mostrados na Figura 6.18.

Nota-se que, a utilização da filtragem *set-membership* consegue reduzir a complexidade do algoritmo BNDR-LMS da mesma maneira como ocorrido no caso mono-canal. Porém, para os mesmo valores de  $\gamma_0$  utilizados no caso mono-canal, o SM-BNDR-LMS apresentou uma maior redução na complexidade computacional e, por conseqüência, no desempenho do sistema. Isto deve-se ao fato de que o sinal de eco multi-canal possuir, em geral, um nível de energia mais baixo que o mono-canal, por ser proveniente de uma fonte sonora que é atenuada duas vezes em seqüência.

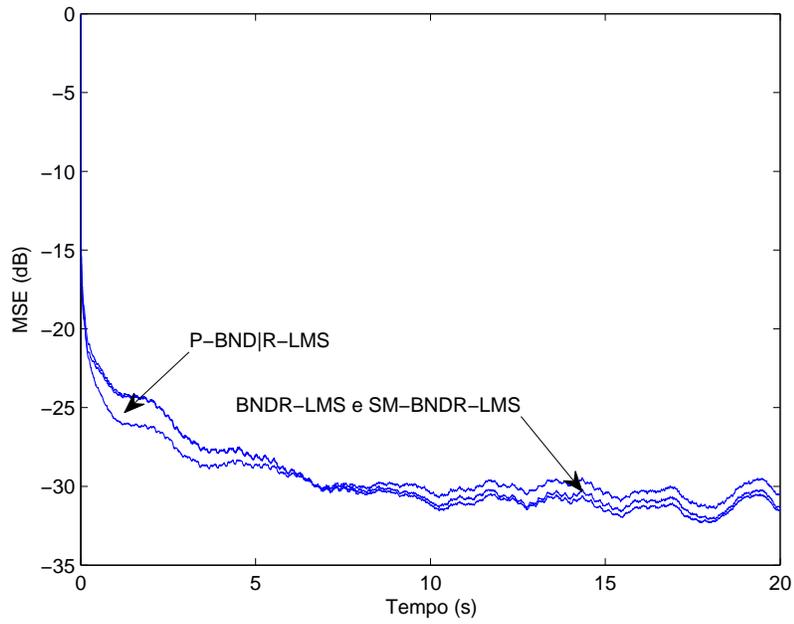


Figura 6.17: Erro médio quadrático dos algoritmos BNDR-LMS, SM-BNDR-LMS e P-BNDR-LMS.

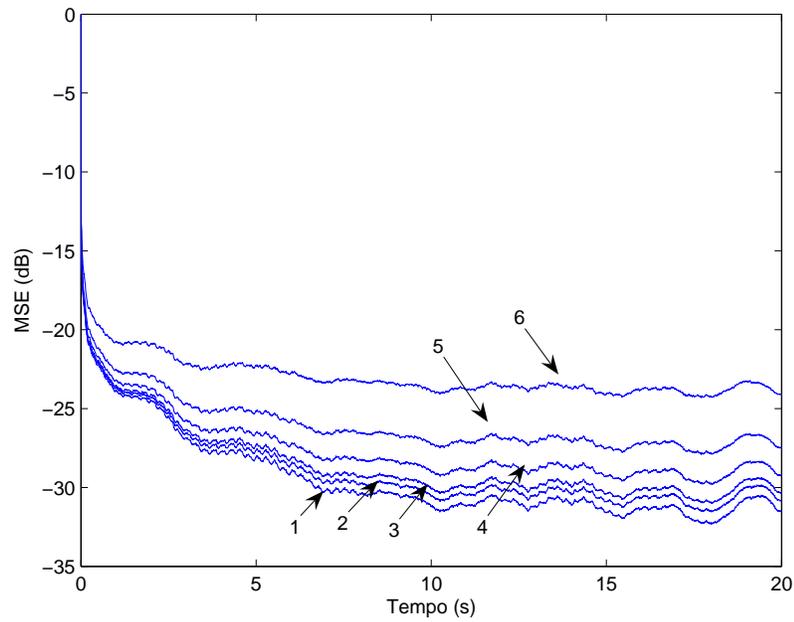


Figura 6.18: Erro médio quadrático do SM-BNDR-LMS com os valores de  $\gamma_0$  mostrados na Tabela 6.6.

Tabela 6.6: Valores de  $\gamma_0$  utilizados no algoritmo SM-BNDR-LMS e seus respectivos percentuais de atualização.

MSE	$\gamma_0$	%
1	$10^{-1}$	98,34
2	3	84,33
3	5	78,05
4	10	68,49
5	20	57,46
6	50	38,83

### 6.5.3 Conversação Cruzada

O mesmo comportamento encontrado em situação de conversação cruzada no caso de cancelamento de eco acústico mono-canal ocorreu no caso multi-canal. Durante a interrupção da adaptação dos coeficientes do filtro adaptativo devido à detecção de conversação cruzada, um aumento excessivo do MSE era obtido. Porém, este aumento ocorria mesmo não havendo a presença de sinal de voz do locutor próximo, sendo causado apenas pela pausa na adaptação dos coeficientes do filtro. Este comportamento utilizando os algoritmos NLMS e BNDR-LMS é ilustrado na Figura 6.19.

A este comportamento é dada a mesma explicação do caso mono-canal: variância da resposta ao impulso da sala de teleconferência. Apesar da utilização de métodos de decorrelação melhorar o desempenho final do sistema como demonstrado na Seção 6.4, eles não são capazes de decorrelacionar completamente os canais e, conseqüentemente, fazer com que o sistema seja insensível a variações das respostas ao impulso da sala de transmissão. Assim, esta variação pode estar ocorrendo tanto na sala de transmissão quanto na sala de recepção por ambas não possuírem preparo acústico, sendo sensíveis a interferências externas.

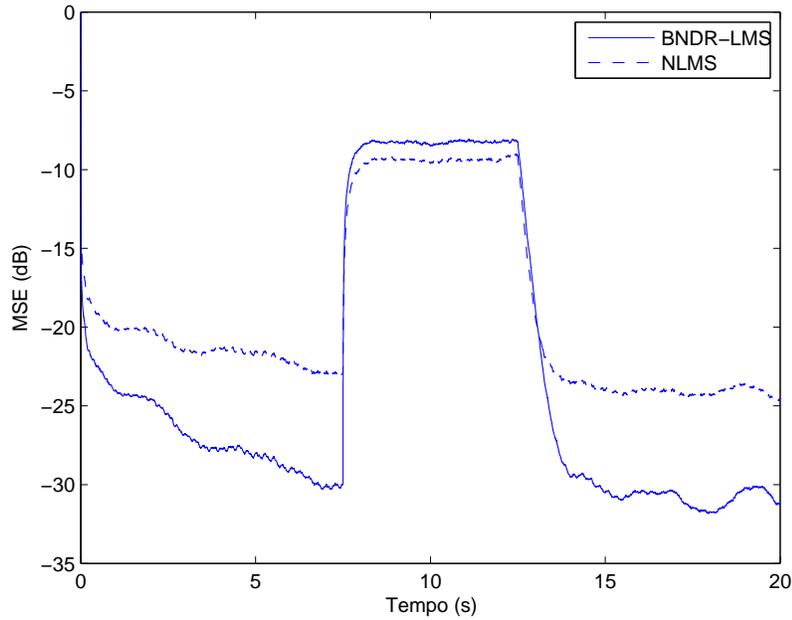


Figura 6.19: Comportamento do erro médio quadrático durante interrupção na adaptação dos coeficientes no caso multi-canal.

## 6.6 Conclusões

Este Capítulo abordou o assunto cancelamento de eco acústico multi-canal descrevendo matematicamente e enfatizando o problema da não-unicidade da solução ótima que, diferentemente do caso mono-canal, não implica no total alinhamento entre a resposta ao impulso  $\mathbf{w}$  do filtro adaptativo e a resposta ao impulso  $\mathbf{h}$  da sala de teleconferência.

Uma solução atual encontrada na literatura, o acréscimo de uma função não-linear dos próprios sinais a si mesmos, foi apresentada e testada sobre diferentes aspectos. Tal solução é capaz de proporcionar uma melhora no desempenho final do sistema sem degradar demasiadamente a qualidade dos sinais, dependendo da escolha do parâmetro  $\alpha$ . Porém, o nível de desalinhamento ainda encontra-se distante do obtido no caso de cancelamento de eco acústico mono-canal.

Uma solução utilizando codificadores de voz para sinais de banda larga foi proposta como método de decorrelação dos canais. Simulações mostram que ambos os codificadores utilizados, G.729.1 e Speex, apresentaram resultados superiores à primeira solução, obtendo um maior nível de decorrelação entre os canais do sistema

e uma maior qualidade dos sinais após o processo de decorrelação. Além disso, um método de codificação híbrido utilizando um codificador em cada canal mostrou-se a melhor solução para obter um baixo nível de desalinhamento.

Foi exemplificada a derivação da versão multi-canal dos algoritmos adaptativos clássicos utilizados nesta dissertação. Simulações em um ambiente de cancelamento de eco acústico multi-canal utilizando sinais gravados em tempo real foram apresentadas.

Os algoritmos apresentaram desempenhos similares ao caso mono-canal, destaque aos algoritmos BNDR-LMS apesar destes possuírem alto desajuste. Mais uma vez, este fato é explicado pela variação da resposta ao impulso da sala de teleconferência que, no caso multi-canal, devido ao problema da não-unicidade da solução pode estar ocorrendo tanto na sala de transmissão quanto na de recepção. Este problema ressalta a importância de algoritmos adaptativos com rápida velocidade de convergência, tornando ineficiente a utilização de algoritmos robustos.

A utilização do sistema de filtragem *set-membership* juntamente com o algoritmo BNDR-LMS mostrou-se novamente eficiente ao reduzir a complexidade computacional média do algoritmo BNDR-LMS em quase 25% sem afetar de maneira drástica o desempenho do sistema.

# Capítulo 7

## Conclusões

Esta dissertação abordou o problema da existência de ecos em sistemas de comunicação. O Capítulo 2 classificou os tipos de eco de acordo com as suas naturezas elétrica ou acústica, e mostrou a necessidade de dispositivos que atenuem este tipo de perturbação. A solução atual para o problema de cancelamento de eco é a utilização de filtros adaptativos, dispositivos que possuem comportamento satisfatório em ambientes onde não haja conhecimento *a priori* das estatísticas dos sinais de entrada.

Tais filtros ajustam seus parâmetros de maneira automática baseando-se em algum algoritmo recursivo de otimização, com o intuito de minimizar em tempo real uma função custo. Os algoritmos adaptativos clássicos foram brevemente derivados e tiveram as suas complexidade computacionais comparadas.

O Capítulo 3 discutiu o problema de conversação cruzada em dispositivos de cancelamento de eco devido a utilização de filtros adaptativos. Provou-se necessário um detector de conversação cruzada (DT) encarregado de parar completamente a adaptação do filtro, quando a presença da fala do locutor próximo for detectada, para evitar que o algoritmo divirja.

Foram apresentados os princípios dos algoritmos de detecção de conversação cruzada e os principais algoritmos existentes na literatura para tal função. Porém, na prática, a detecção de conversação cruzada não ocorre perfeitamente. Os algoritmos de DT costumam apresentar alguns tipos de falhas como demorar algum tempo para detectar uma situação de DT ou mesmo não detectá-las completamente. Estes problemas são suficientes para influenciarem o comportamento do algoritmo

adaptativo causando divergência.

Por isso, fez-se necessário o desenvolvimento de algoritmos adaptativos robustos a distúrbios de curta duração. A intenção foi fazer com que os algoritmos sejam poucos sensíveis aos erros de detecção de DT e, conseqüentemente, não permitam uma passagem excessiva de eco. Uma atual solução encontrada na literatura, denominada nesta dissertação de tipo 1, foi apresentada. Uma alteração foi proposta para uma segunda solução, denominada tipo 2, a qual foi comparada com a solução tipo 1.

No Capítulo 4 canceladores de eco elétrico foram desenvolvidos de acordo com as características exigidas em [1] utilizando vários algoritmos adaptativos. Os algoritmos foram confrontados em situações de convergência, re-convergência e conversação cruzada, sempre verificando se os desempenhos supriam as exigências mínimas da atual norma em vigor. Mostrou-se que a utilização de algoritmos robustos faz-se, realmente, necessária para que não haja uma passagem excessiva de eco em situações de conversação cruzada. No entanto, a utilização destes tipos de algoritmo sacrifica a velocidade de convergência e, principalmente, de re-convergência, fazendo com que, neste último caso, uma das exigências de [1] não seja alcançada.

Comparando os algoritmos robustos tipo 1 e tipo 2, concluiu-se que, com a alteração proposta e utilizando os mesmos valores para os parâmetros, o algoritmo robusto tipo 2 apresentou uma maior velocidade de convergência e de re-convergência. No entanto, este algoritmo mostrou-se um pouco mais sensível aos erros de detecção de DT resultando em uma maior redução no MSE durante estes períodos. Em relação aos algoritmos adaptativos, o algoritmos P-BNDR-LMS, por juntar as técnicas de proporcionalidade e reutilização de dados, apresentou o melhor desempenho durante as simulações, seguido de perto pelo algoritmo BNDR-LMS. Porém, o algoritmo BNDR-LMS apresenta uma menor complexidade computacional que o algoritmo P-BNDR-LMS.

Já o Capítulo 5 apresentou simulações em um ambiente de cancelamento de eco acústico mono-canal utilizando sinais gravados em tempo real. Os resultados mostraram que, apesar de possuir um maior desajuste, o algoritmo adaptativo BNDR-LMS obteve o melhor desempenho entre os algoritmos avaliados. Isto é justificado por este algoritmo reutilizar dados para aumentar a velocidade de con-

vergência em situações onde o sinal de entrada é correlacionado, sendo esta uma das principais características dos sinais de voz. Além disso, a demonstrada variação da resposta ao impulso da sala de teleconferência durante as gravações, que fez com que os algoritmos estivessem sempre em uma situação de re-convergência, ressaltou ainda mais a necessidade de algoritmos adaptativos com alta velocidade de convergência.

A variação da resposta ao impulso da sala de teleconferência foi identificada durante os testes em situações de conversação cruzada, onde a interrupção da atualização dos coeficientes do filtro resultava em um aumento excessivo do MSE, mesmo na ausência de sinal de voz do locutor próximo. Esta característica pode ser explicada pelo fato que a sala de teleconferência não era acusticamente preparada e fez com que a utilização de algoritmos robustos tornasse injustificável além de ineficiente, por estes caracterizarem-se por terem uma baixa velocidade de re-convergência devido ao modo como são derivados. Por fim, a utilização do sistema de filtragem *set-membership* mostrou-se eficiente ao reduzir a complexidade computacional média do algoritmo BNDR-LMS em quase 25% sem afetar de maneira drástica o seu desempenho.

O Capítulo 6 abordou o assunto cancelamento de eco acústico multi-canal descrevendo matematicamente e enfatizando o problema da não-unicidade da solução ótima que, diferentemente do caso mono-canal, não implica no total alinhamento entre a resposta ao impulso  $\mathbf{w}$  do filtro adaptativo e a resposta ao impulso  $\mathbf{h}$  da sala de teleconferência.

A solução atual encontrada na literatura, o acréscimo de uma função não-linear dos próprios sinais a si mesmos, foi apresentada e testada sobre diferentes aspectos. Tal solução é capaz de proporcionar uma melhora no desempenho final do sistema sem degradar demasiadamente a qualidade dos sinais, dependendo da escolha do parâmetro  $\alpha$ . Porém, o nível de desalinhamento ainda encontra-se distante do obtido no caso de cancelamento de eco acústico mono-canal.

Uma solução utilizando codificadores de voz para sinais de banda larga foi proposta como método de descorrelação dos canais. Simulações mostram que ambos os codificadores utilizados, G.729.1 e Speex, apresentaram resultados superiores à primeira solução, obtendo um maior nível de descorrelação entre os canais do sistema e uma maior qualidade dos sinais após o processo de descorrelação. Além disso, um

método de codificação híbrido utilizando um codificador em cada canal mostrou-se a melhor solução para obter um baixo nível de desalinhamento.

Nas simulações de cancelamento de eco acústico multi-canal, os algoritmos apresentaram desempenhos similares ao caso mono-canal, destaque aos algoritmos BNDR-LMS apesar destes possuírem alto desajuste. Mais uma vez, este fato é explicado pela variação da resposta ao impulso da sala de teleconferência que, no caso multi-canal, devido ao problema da não-unicidade da solução pode estar ocorrendo tanto na sala de transmissão quanto na de recepção. Este problema ressalta a importância de algoritmos adaptativos com rápida velocidade de convergência, tornando ineficiente a utilização de algoritmos robustos. Mais uma vez, a filtragem *set-membership* mostrou-se eficiente ao reduzir a complexidade computacional média do algoritmo BNDR-LMS sem afetar demasiadamente o desempenho do sistema.

As contribuições desta tese mostram que apesar desta área de pesquisa estar solidificada, ainda existem temas que podem ser aprofundados. Assim, para trabalhos futuros são apresentadas as seguintes propostas:

- desenvolvimento de algoritmos adaptativos robustos que possuam uma velocidade de re-convergência mais rápida;
- verificação dos resultados de cancelamento de eco acústico em salas com preparo acústico;
- avaliação de outras combinações para métodos de decorrelação: utilização de codificadores de voz juntamente com a função retificadora de meia onda, utilização de um método em cada canal, utilização de codificadores de voz em cascata, utilização de codificadores de voz e áudio em cada canal;
- desenvolvimento de novas técnicas para decorrelacionar os canais.

# Referências Bibliográficas

- [1] ITU-T Rec. G.168, “Digital network echo cancellers”, International Telecommunications Union, Geneva, Switzerland, 2004.
- [2] ITU-T Rec. G.131, “Talker echo and its control”, International Telecommunications Union, Geneva, Switzerland, 2003.
- [3] ITU-T Rec. G.164, “Echo suppressors”, International Telecommunications Union, Geneva, Switzerland, 1988.
- [4] WIDROW, B., STEARNS, S. D., *Adaptive Signal Processing*. Englewood Cliffs, NJ, Prentice Hall, 1985.
- [5] HAYKIN, S., *Adaptive Filter Theory*. 3 ed. Upper Saddle River, New Jersey, Prentice Hall, 1996.
- [6] DINIZ, P. S. R., *Adaptive Filtering: Algorithms and Practical Implementation*. 2 ed. Norwell, MA, Kluwer Academic Publishers, 2002.
- [7] DUTTWEILER, D. L., “Proportionate normalized least mean square adaptation in echo cancellers”, *IEEE Transactions on Speech and Audio Processing*, v. 8, pp. 508–518, September 2000.
- [8] GÄNSLER, T., GAY, S. L., SONDHI, M. M., *et al.*, “Double-talk robust fast converging algorithms for network echo cancellation”, *IEEE Transactions on Speech and Audio Processing*, v. 8, pp. 656–663, November 2000.
- [9] CAMPOS, M. L. R. D., DINIZ, P. S. R., APOLINÁRIO JR., J. A., “On normalized data-resusing and affine-projections algorithms”. In: *6th IEEE International Conference on Electronics, Circuits and Systems*, pp. 843–846, 1999.

- [10] APOLINÁRIO JR., J. A., CAMPOS, M. L. R. D., DINIZ, P. S. R., “Convergence analysis of the binormalized data-reusing LMS algorithm”, *IEEE Transactions on Signal Processing*, v. 48, n. 11, pp. 3235–3242, November 2000.
- [11] GÄNSLER, T., BENESTY, J., GAY, S. L., *et al.*, “A robust proportionate affine projection algorithm for network echo cancellation”. In: *Proceedings of IEEE Conference on Acoustics, Speech, and Signal Processing*, pp. 793–796, June 2000.
- [12] GAY, S. L., TAVATHIA, S., “The fast affine projection algorithm”. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 3023–3026, 1995.
- [13] YE, H., WU, B. X., “A new double-talk detection algorithm based on orthogonality theorem”, *IEEE Transactions on Communication*, v. 39, pp. 1542–1545, November 1991.
- [14] BENESTY, J., MORGAN, D. R., CHO, J. H., “A new class of doubletalk detectors based on cross-correlation”, *IEEE Transactions on Speech and Audio Processing*, v. 8, pp. 168–172, March 2000.
- [15] IQBAL, M. A., STOKES, J. W., GRANT, S. L., “Normalized double-talk detection based on microphone and aec error cross-correlation”. In: *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 360–363, July 2007.
- [16] HUBER, P. J., *Robust Statistics*. New York, Wiley, 1981.
- [17] BENESTY, J., GÄNSLER, T., “A robust fast recursive least squares adaptive algorithm”, *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, v. 6, pp. 3785–3788, May 2001.
- [18] ITU-T Rec. P.501, “Test signals for use in telephony”, International Telecommunications Union, Geneva, Switzerland, 1996.
- [19] ALCAIM, A., SOLEWICZ, J. A., MORAES, J. A. D., “Frequência de ocorrência dos fonemas e listas de frases foneticamente balanceadas no português falado no Rio de Janeiro”, Dezembro 2002.

- [20] BISPO, B. C., NETTO, S. L., “Comparação de algoritmos adaptativos para cancelamento de eco acústico”, Setembro 2007.
- [21] DINIZ, P. S. R., WERNER, S., “Set-membership binormalized data-reusing LMS algorithm”, *IEEE Transactions on Signal Processing*, v. 51, n. 1, pp. 124–134, January 2003.
- [22] SONDHI, M. M., REYNOLDS, R., MORGAN, D. R., “Acoustic echo cancellation for stereophonic teleconferencing”. In: *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, May 1991.
- [23] SONDHI, M. M., MORGAN, D. R., “Stereophonic acoustic echo cancellation - an overview of the fundamental problem”, *IEEE Signal Processing Letters*, v. 2, pp. 148–151, August 1995.
- [24] SHIMAUCHI, S., MAKINO, S., “Stereo projection echo canceller with true echo path estimation”. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 3059–3062, 1995.
- [25] BENESTY, J., MORGAN, D. R., SONDHI, M. M., “A better understanding an a improved solution to the specific problems of stereophonic acoustic echo cancellation”, *IEEE Transactions on Speech and Audio Processing*, v. 6, n. 2, pp. 156–165, March 1998.
- [26] MORGAN, D. R., HALL, J. L., BENESTY, J., “Investigation of several types of nonlinearities for use in stereo acoustic echo cancellation”, *IEEE Transactions on Speech and Audio Processing*, v. 9, n. 6, pp. 686–696, September 2001.
- [27] ITU-T Rec. G.729.1, “G.729-based embedded variable bit-rate coder: an 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729”, International Telecommunications Union, Geneva, Switzerland, 2006.
- [28] “The Speex Codec Manual Version 1.2 Beta”, <http://www.speex.org>, (accessed in July 2008).
- [29] ITU-T Rec. P.800, “Methods for subjective determination of transmission quality”, International Telecommunications Union, Geneva, Switzerland, 1996.

- [30] LIMA, A. A. D., FREELAND, F. P., JESUS, R. A. D., *et al.*, “On the Quality Assessment of Sound Signals”. In: *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 416–419, May 2008.
- [31] LIMA, A. A. D., FREELAND, F. P., ESQUEF, P. A. A., *et al.*, “Reverberation assessment in audioband speech signals for telepresence systems”. In: *Proceedings of International Conference on Signal Processing and Multimedia Applications*, July 2008.
- [32] 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 Telecommunications Union, Geneva, Switzerland, 2005.
- [33] ITU-T Rec. P.862.2, “Wideband extension to recommendation P.862 for the assessment of wideband telephone networks and speech codecs”, International Telecommunications Union, Geneva, Switzerland, 2005.
- [34] ITU-T Rec. G.191, “Software tools for speech and audio coding standardization”, International Telecommunications Union, Geneva, Switzerland, 2005.