

UNIVERSIDADE ESTADUAL DE CAMPINAS - UNICAMP
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO - FEEC
DEPARTAMENTO DE COMUNICAÇÃO - DECOM
PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA
ÁREA: TELECOMUNICAÇÕES E TELEMÁTICA

Cancelamento de Eco em Telefonia IP

Autor:

Euler Mendes Rachid

Orientador:

Prof. Dr. Luís Geraldo Pedroso Meloni

*Dissertação apresentada à Coordenação do
Programa de Pós-Graduação em Engenharia
Elétrica da Universidade Estadual de Campinas
como parte dos requisitos exigidos para obtenção do
grau de **Mestre em Engenharia Elétrica***

Banca Examinadora:

Prof. Dr. João Marcos Travassos Romano

Prof. Dr. Luís Geraldo Pedroso Meloni

Prof. Dr. Paulo Cardieri

Prof. Dr. Paulo Mello Marshall

Campinas - São Paulo

Outubro - 2004

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

R114c Rachid, Euler Mendes
Cancelamento de eco em telefonia IP / Euler Mendes
Rachid. --Campinas, SP: [s.n.], 2004.

Orientador: Luís Geraldo Pedroso Meloni
Dissertação (Mestrado) - Universidade Estadual de
Campinas, Faculdade de Engenharia Elétrica e de
Computação.

1. Redes de telefone. 2. Telefone – Ruído. 3. Ruído
elétrico. 4. Redes de telefone – Controle de ruído. 5. Filtros
adaptativos. 6. Garantia de qualidade. 7. Comunicação
interpessoal. 8. Processamento de sinais – Técnicas digitais.
9. Internet (Redes de computação). I. Meloni, Luís Geraldo
Pedroso. II. Universidade Estadual de Campinas. Faculdade
de Engenharia Elétrica e de Computação. III. Título.

Resumo

Este trabalho teve como objetivo o estudo do cancelamento do eco em telefonia IP, onde foi proposto um novo método mais efetivo que o atualmente utilizado nos *gateways*, que leva em consideração a variação do atraso da rede IP. No início é apresentado o conceito relacionado a filtros adaptativos. Posteriormente, é apresentado um estudo sobre o cancelamento do eco na rede de telefonia pública comutada, explicando o funcionamento do atenuador, do supressor de eco e do cancelador de eco. O controle do eco é tratado como uma técnica empregada em VoIP para melhoria da qualidade do sinal de fala. A análise do desempenho do cancelador de eco tradicional é testada através de modelos de caminhos de ecos apresentados na recomendação G.168 do ITU-T. Como métrica é utilizada uma medida baseada no valor da melhoria da perda de retorno de eco (ERLE) e do tempo de convergência do cancelador. Resultados de simulações do método proposto são apresentados e discutidos no final.

Abstract

The objective of this thesis is the study about echo cancellation in a IP telephony, providing a new technique better than the actual one employed in gateways. This technique takes account the fact that the delay in an IP network is variable. This work presents an overview on adaptive filter subject, following an study about echo cancellation in a PSTN network, it explains the functionality of some devices such as attenuators, echo suppressors and echo cancellers. The relation between echo control and the measured voice quality is presented. The echo path models presented in ITU-T G.168 recommendation are used to test the conventional echo canceller performance. For this purpose the ERLE (Echo Return Loss Enhancement) and the convergency time are measured. Finally, simulation results for the new cancellation method are presented and discussed in details.

Apoio Financeiro

Esse trabalho se tornou possível graças a bolsa de mestrado financiada pela Fundação da Universidade Estadual de Campinas (FUNCAMP) em convênio com a Ericsson Telecomunicações.

Agradecimentos

Ao orientador Meloni que possibilitou a realização deste trabalho.

A minha namorada, Alessandra, que me acompanhou durante a elaboração deste trabalho, sempre me incentivando e me dando forças para concluí-lo.

Aos meus pais, Rachid e Natália, e minhas irmãs, Adriana e Cristiane, pelo apoio e incentivo.

Aos amigos do laboratório, Fábio, Glauco, Helder, Jaqueline, Leonardo, Lívio, Lucas, Márzio, Paulo e Sérgio, que me apoiaram e me auxiliaram, sempre dispostos a discutir temas relacionados a tese. Amigos que tornaram o ambiente de trabalho mais alegre e agradável.

Ao Vicente por me auxiliar na utilização do editor de Latex e fornecer um modelo de formato de tese.

A banca examinadora que acrescentou muitas contribuições pertinentes à tese.

A todas outras pessoas que contribuíram diretamente e indiretamente neste trabalho.

Sumário

Lista de Figuras	vi
Lista de Tabelas	viii
1 Introdução	1
1.1 Motivação	1
1.2 Organização	2
2 Filtros Adaptativos	4
2.1 Introdução	4
2.2 Princípios de um Sistema Adaptativo	5
2.2.1 Aplicações que utilizam Filtros Adaptativos	6
2.2.2 Estruturas de Filtros Adaptativos	6
2.2.3 Algoritmos de Otimização	8
2.2.4 Função de Custo	9
2.2.5 Definição do sinal de erro	10
2.3 Filtro de Wiener	10
2.4 Algoritmo LMS (Least Mean Square)	12
2.5 Algoritmo NLMS (Normalized LMS)	13
2.6 Algoritmo RLS (Recursive Least Square)	14
3 Eco na rede telefônica	17
3.1 A Origem do Eco	17
3.2 O Efeito do Eco	19
3.3 Métodos de Controle do Eco	20
3.3.1 Atenuador	20
3.3.2 Supressor de Eco	21
3.3.3 Bobina Híbrida com Autobalanceamento de Impedância	22
3.3.4 Cancelador de Eco	23

4	Qualidade de Serviço em VoIP	25
4.1	Uma Visão Geral de VoIP	25
4.2	Qualidade de Serviço	27
4.2.1	Latência	27
4.2.2	<i>Jitter</i>	28
4.2.3	Codificadores	29
4.2.4	Perda de Pacotes	31
4.2.5	Técnicas de Qualidade de Serviço	31
4.2.6	Métodos utilizados para atingir a Qualidade de Serviço	32
5	Modelos de caminho de eco	37
5.1	Técnicas de Medida Utilizadas para Obter o Modelo do Caminho de Eco	37
5.2	Características do Caminho de Eco Medidas na América do Norte	39
5.3	Modelos Utilizados na Simulação	39
6	Simulações do Cancelador de Eco	52
6.1	Simulações do Cancelador de Eco Convencional	53
6.2	Cancelamento de Eco na rede IP	71
7	Conclusões	77
	Referências Bibliográficas	79

Lista de Figuras

2.1	Diagrama de um filtro adaptativo genérico	6
2.2	Diagrama de blocos de uma estrutura transversal	7
2.3	Diagrama de blocos de uma estrutura FIR em treliça	7
3.1	Diagrama simplificado de uma ligação em um circuito a quatro fios	18
3.2	Diagrama de uma Rede Telefônica Básica	18
3.3	Gráfico da atenuação versus atraso do eco - reproduzido de [1]	20
3.4	Diagrama de um supressor de eco	21
3.5	Diagrama de dois Supressores de Eco em Tandem	22
3.6	Diagrama esquemático do Cancelador de Eco	24
4.1	Representação de uma transmissão VoIP	26
5.1	Método de medição do caminho de eco	38
5.2	Desvio e média dos espectros de 88 chamadas - reproduzido de [2]	40
5.3	Resposta impulsiva do modelo de caminho de eco 1	42
5.4	Resposta em frequência do modelo de caminho de eco 1	42
5.5	Resposta impulsiva do modelo de caminho de eco 2	43
5.6	Resposta em frequência do modelo de caminho de eco 2	44
5.7	Resposta impulsiva do modelo de caminho de eco 3	45
5.8	Resposta em frequência do modelo de caminho de eco 3	45
5.9	Resposta impulsiva do modelo de caminho de eco 4	46
5.10	Resposta em frequência do modelo de caminho de eco 4	47
5.11	Resposta impulsiva do modelo de caminho de eco 5	48
5.12	Resposta em frequência do modelo de caminho de eco 5	48
5.13	Resposta impulsiva do modelo de caminho de eco 6	49
5.14	Resposta em frequência do modelo de caminho de eco 6	50
5.15	Resposta impulsiva do modelo de caminho de eco 7	51
5.16	Resposta em frequência do modelo de caminho de eco 7	51
6.1	Gráfico de convergência para o modelo de caminho de eco 1	55
6.2	Comparação entre o modelo de eco estimado e o modelo 1	56

6.3	Gráfico da média dos erros quadráticos para o modelo de caminho de eco 1	56
6.4	Gráfico de convergência para o modelo de caminho de eco 2	57
6.5	Gráfico da média dos erros quadráticos para o modelo de caminho de eco 2	58
6.6	Comparação entre o modelo de eco estimado e o modelo 2	58
6.7	Gráfico de convergência para o modelo de caminho de eco 3	59
6.8	Gráfico da média dos erros quadráticos para o modelo de caminho de eco 3	60
6.9	Comparação entre o modelo de eco estimado e o modelo 3	60
6.10	Gráfico de convergência para o modelo de caminho de eco 4	61
6.11	Média dos erros quadráticos para o modelo de caminho de eco 4	62
6.12	Comparação entre o modelo de eco estimado e o modelo 4	62
6.13	Gráfico de convergência para o modelo de caminho de eco 4	63
6.14	Média dos erros quadráticos para o modelo de caminho de eco 4	64
6.15	Comparação entre o modelo de eco estimado e o modelo 4	64
6.16	Gráfico de convergência para o modelo de caminho de eco 5	65
6.17	Média dos erros quadráticos para o modelo de caminho de eco 5	66
6.18	Comparação entre o modelo de eco estimado e o modelo 5	66
6.19	Gráfico de convergência para o modelo de caminho de eco 6	67
6.20	Média dos erros quadráticos para o modelo de caminho de eco 6	68
6.21	Comparação entre o modelo de eco estimado e o modelo 6	68
6.22	Gráfico de convergência para o modelo de caminho de eco 7	69
6.23	Média dos erros quadráticos para o modelo de caminho de eco 7	70
6.24	Comparação entre o modelo de eco estimado e o modelo 7	70
6.25	Processo envolvido no cancelamento de eco na rede IP	72
6.26	Sinais utilizados na simulação	74
6.27	Valor do atraso inicial em cada interação da simulação	75
6.28	Gráfico de convergência do cancelador de eco na rede IP	75

Lista de Tabelas

4.1	Valores MOS dos codificadores.	30
5.1	Valores das constantes dos modelos de caminho de eco.	41
5.2	Resposta impulsiva do modelo de caminho de eco 1, $m_1(k)$	41
5.3	Resposta impulsiva do modelo de caminho de eco 2, $m_2(k)$	43
5.4	Resposta impulsiva do modelo de caminho de eco 3, $m_3(k)$	44
5.5	Resposta impulsiva do modelo de caminho de eco 4, $m_4(k)$	46
5.6	Resposta impulsiva do modelo de caminho de eco 5, $m_5(k)$	47
5.7	Resposta impulsiva do modelo de caminho de eco 6, $m_6(k)$	49
5.8	Resposta impulsiva do modelo de caminho de eco 7, $m_7(k)$	50

Capítulo 1

Introdução

Essa dissertação de mestrado apresenta os conceitos importantes relacionados com o cancelamento de eco. Aborda desde a sua origem até a sua aplicação em telefonia IP. Posteriormente, é apresentado um novo modelo de cancelador de eco mais eficaz que os existentes para aplicações de voz sobre IP.

1.1 Motivação

A tendência atual nas telecomunicações é a integração entre a rede de circuitos comutados e a rede de pacotes. Isto possibilitará o uso de aplicações integradas de multimídia interativa, como conferências virtuais utilizando áudio e vídeo. Entretanto, para atingir um nível de serviço adequado é necessário o uso de técnicas de qualidade de serviço (QoS) que garantem a largura de faixa adequada e um atraso mínimo no tempo de transmissão dos dados.

Como a maioria das redes de dados existentes utilizam o protocolo internet (IP), surgiu a possibilidade de utilizar esta rede para transmitir voz, que é chamado de VoIP. Espera-se que no futuro a atual rede telefônica seja substituída por esta nova tecnologia. Grandes empresas de telefonia estão investindo alto no desenvolvimento de soluções em VoIP que irão, inicialmente, funcionar em conjunto com a rede telefônica tradicional.

Um dos maiores problemas encontrados na telefonia IP foi o surgimento do eco¹, devido ao atraso de transmissão significativamente maior na rede IP em relação à rede de telefonia de comutação de circuitos. Este causa uma grande degradação na qualidade da ligação telefônica. Distúrbios deste tipo induzem interrupções extras na conversação que incomodam o usuário, o que motivou o estudo de métodos para controlá-lo. O eco na telefonia é um assunto que já vem sendo tratado há décadas,

¹Existem dois tipos de eco, o eco elétrico, que se refere ao citado no texto, e o eco acústico que é causado por reflexões no ambiente. Este último surgiu na telefonia com a utilização do viva voz.

desde o surgimento da telefonia a longa distância. Várias técnicas para eliminar o eco foram concebidas, sendo o cancelador de eco a mais eficiente.

O eco na telefonia é causado por reflexões que ocorrem devido ao descasamento de impedâncias entre circuitos internos da rede de comutação. A percepção do eco pelo usuário é dependente do atraso de transmissão da rede, um fator que tem um impacto significativo no controle do eco. Em uma chamada telefônica, o tempo gasto para escutá-lo é constante durante toda a ligação. Entretanto, em uma chamada VoIP, o atraso do eco varia dificultando a convergência do algoritmo adaptativo utilizado no cancelador de eco e, conseqüentemente, a redução do eco não será efetiva.

O conceito sobre cancelador de eco na telefonia de circuitos comutados foi desenvolvido há anos atrás, e existem muitas publicações neste tema. Entretanto é difícil encontrar artigos sobre o cancelamento de eco na telefonia IP. Esta tecnologia é recente e ainda está em desenvolvimento.

O objetivo dessa dissertação é fornecer um estudo detalhado sobre o cancelamento de eco empregado na rede IP, englobando todo o assunto relacionado ao cancelador de eco. Além disto, é proposto um novo método de cancelamento mais efetivo que leva em consideração a variação do atraso da rede IP.

1.2 Organização

A seguir, está apresentado uma descrição do conteúdo dos seguintes capítulos que compõem essa dissertação.

O Capítulo 2 apresenta um resumo dos conceitos associados aos filtros adaptativos. No início, está apresentada uma introdução com algumas definições básicas e estruturas de filtros. Citam-se os principais algoritmos de minimização e algumas funções objetivas. Por fim, apresenta-se a teoria do filtro de Wiener e os algoritmos LMS, NLMS e RLS.

O Capítulo 3 apresenta um estudo sobre o cancelamento do eco na telefonia de circuitos comutados. Este capítulo descreve a origem do eco, um estudo do efeito do eco em uma chamada telefônica e métodos de controle do eco. Explica o funcionamento do atenuador, do supressor de eco e do cancelador de eco na rede de telefonia pública comutada (PSTN). Por ser o assunto principal desta dissertação, o último é descrito com mais detalhes.

O Capítulo 4 apresenta os principais conceitos relacionados à qualidade de serviço e diversas técnicas empregadas em VoIP para melhoria da qualidade, incluindo o uso do cancelamento de eco no *gateway* da rede IP.

O Capítulo 5 apresenta os modelos utilizados para simular o caminho de eco em uma rede telefônica. Os modelos apresentados foram utilizados nas simulações do cancelador de eco para avaliar o seu desempenho.

O Capítulo 6 apresenta os resultados da simulação do cancelador de eco, utilizando os modelos apresentados no Capítulo 5. Posteriormente, é apresentado uma simulação do cancelador de eco proposto, que pode ser empregado em um *gateway* para atenuar o eco provindo da rede IP. Esta solução em conjunto com o método utilizado comercialmente, que cancela o lado da rede PSTN, forma um cancelador de eco completo, sem a necessidade de um outro cancelador localizado em um *gateway* remoto.

O Capítulo 7 apresenta um comentário sobre os resultados obtidos e as conclusões gerais deste trabalho.

Capítulo 2

Filtros Adaptativos

O desenvolvimento da teoria de filtros adaptativos foi fundamental para solucionar o efeito do eco na telefonia. Esse capítulo aborda os conceitos associados a filtros adaptativos. No início deste capítulo, será apresentada uma introdução com algumas definições básicas, exemplos de aplicações que utilizam filtros adaptativos e algumas estruturas básicas de filtros. Posteriormente, serão citados alguns algoritmos de otimização e algumas funções custo. Este capítulo termina apresentando a teoria do filtro de Wiener, esta teoria levou ao desenvolvimento dos algoritmos LMS, NLMS e RLS que são revistos no final.

2.1 Introdução

A área de processamento de sinais tem se desenvolvido significativamente nos últimos anos. Projetar um circuito digital tem muitas vantagens sobre o circuito analógico. Os sistemas baseados no processamento digital de sinais são melhores devido à segurança, menor tamanho físico, flexibilidade e precisão [3]; sendo que as duas últimas características são as principais responsáveis pelo rápido crescimento no campo do processamento de sinais adaptativos [4].

Um filtro digital é um tipo de filtro que processa um sinal discreto no tempo representado no formato digital, ou seja, a amplitude do sinal é representada por uma seqüência de bits. Os parâmetros internos e a estrutura de filtros invariantes no tempo são fixos, e se o filtro é linear o sinal de saída é uma função linear do sinal de entrada e/ou do sinal de saída. A filtragem linear é, provavelmente, o tipo mais comum e mais importante entre outros processamentos de sinais [4]. Após definidas as especificações do filtro, o projeto de um filtro linear invariante no tempo deve envolver três etapas: modelagem de uma função de transferência racional que atenda as especificações do projeto, a escolha da estrutura para o filtro, e a escolha da forma de implementação do algoritmo [3].

Um filtro adaptativo é utilizado quando o ambiente externo é inicialmente

desconhecido ou variante no tempo. Desta forma, a adaptação é conveniente para ajustar os parâmetros do filtro, ou quando o ambiente externo varia lentamente [5]. Este tipo de filtro é não linear, pois seus parâmetros são dependentes do sinal de entrada e estão mudando de acordo com a evolução do sistema. Conseqüentemente, a homogeneidade e condições de aditividade não são satisfeitas. Entretanto, se for analisado em um determinado instante de tempo, congelando-se os parâmetros do filtro, o filtro adaptativo é considerado instantaneamente linear [3].

Como os parâmetros do filtro adaptativo mudam para atender aos critérios de desempenho, os filtros adaptativos são variantes no tempo. Desta forma, pode-se considerar um filtro adaptativo como um filtro que executa o passo de aproximação em tempo real [3]. No projeto de filtros não adaptativos é necessário uma identificação completa do sinal de entrada e do sinal de referência para que este atenda às especificações. Quando o processo não está bem definido, os sinais de entrada e de referência não estarão bem definidos. Assim, surge a necessidade de modelar os sinais e, posteriormente, projetar o filtro. Este procedimento pode ter um custo elevado e ser de difícil implementação. Para contornar este problema, emprega-se um filtro adaptativo que executa uma atualização em tempo real de seus parâmetros através de um algoritmo simples, utilizando apenas as informações disponíveis [3].

O algoritmo adaptativo, que tem a função de determinar os valores dos coeficientes do filtro, precisa de um sinal para guiá-lo. Este sinal é chamado de *sinal de referência* ou *sinal desejado*, cuja escolha depende da aplicação.

Devido ao fato dos filtros adaptativos serem sistemas não lineares, analisar seu comportamento é mais complicado do que no caso dos filtros não adaptativos. Por outro lado, como os filtros adaptativos são filtros auto projetáveis seu projeto pode ser considerado menos complexo do ponto de vista prático.

2.2 Princípios de um Sistema Adaptativo

Um diagrama geral de um filtro adaptativo é apresentado na Figura 2.1. Neste diagrama, k representa o número da interação, $x(k)$ o sinal de entrada, $y(k)$ o sinal de saída do filtro adaptativo, e $d(k)$ o sinal desejado. O sinal de erro, $e(k)$, é calculado como a diferença entre o sinal desejado e o sinal de saída, $e(k) = d(k) - y(k)$. Este sinal de erro junto com o sinal de entrada são utilizados para atualizar os coeficientes do filtro, segundo um critério de otimização.

Para uma completa especificação de um sistema adaptativo, é necessária a definição da aplicação, a escolha da estrutura do filtro e a escolha do algoritmo adaptativo [3].

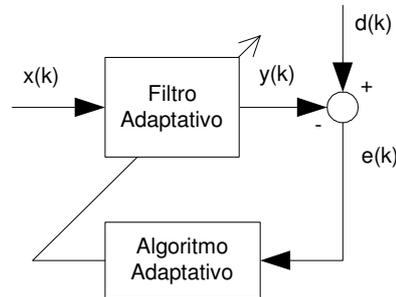


Figura 2.1 Diagrama de um filtro adaptativo genérico

2.2.1 Aplicações que utilizam Filtros Adaptativos

Os sistemas adaptativos podem ser classificados de acordo com a sua aplicação, diferenciando uns dos outros pelas conexões externas do filtro [5]. Podem ser classificados como: identificador, modelador inverso, preditor e cancelador de interferência [6]. Alguns exemplos de aplicação são: identificação de sistemas, equalização de canais, codificação preditiva, cancelamento de eco, melhoria da relação sinal ruído, cancelamento de ruído, análise espectral, e controle [6] [7] [8].

2.2.2 Estruturas de Filtros Adaptativos

A escolha da estrutura para o processo de filtragem tem uma grande influência na operação do algoritmo como um todo [6]. Da estrutura do filtro depende a complexidade computacional do processo e, também, o número de interações necessárias para o algoritmo adaptativo convergir [3]. A resposta ao impulso de um filtro linear determina o tamanho da memória do filtro [6]. Existem duas classes de filtros, os filtros FIR (Finite-duration Impulse Response), de memória finita, e os filtros IIR (Infinite-duration Impulse Response), de memória infinita.

O tipo mais utilizado para filtros adaptativos é o filtro FIR discreto no tempo [5]. São representados pela seguinte equação de diferenças:

$$y(k) = \sum_{i=0}^N w_i x(k-i) \quad (2.1)$$

cuja função de transferência é dado por:

$$H(z) = \sum_{i=0}^N w_i z^{-i} \quad (2.2)$$

A expressão (2.1) define um modelo chamado de média móvel (moving average), cuja função de transferência (2.2) só possui zeros.

Existem várias estruturas que podem implementar um filtro FIR, a mais utilizada é a estrutura na forma direta, ou transversal [5]. Esta estrutura é, simplesmente, a realização direta da função de transferência (2.2). Um diagrama desta estrutura está apresentado na Figura 2.2. Outra estrutura que pode ser utilizada é a estrutura em treliça. Esta é representada por constantes multiplicativas, denominadas de coeficientes de reflexão. A Figura 2.3 apresenta um diagrama desta estrutura na forma direta. Ambas estruturas podem ser implementadas na forma direta, em cascata ou em paralelo [5].

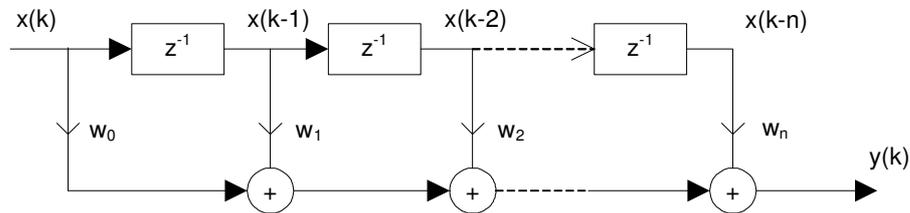


Figura 2.2 Diagrama de blocos de uma estrutura transversal

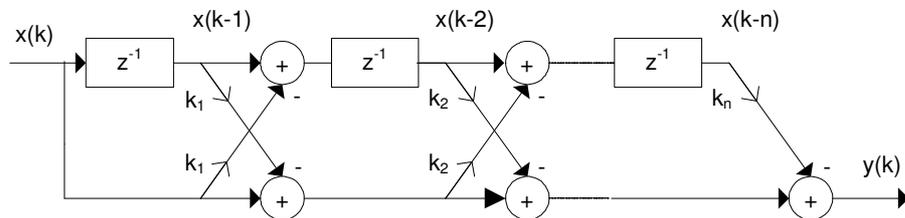


Figura 2.3 Diagrama de blocos de uma estrutura FIR em treliça

Os filtros IIR utilizam estruturas recursivas. A estrutura mais utilizada é a forma canônica direta, devido à sua simples implementação e análise [9]. A característica que distingue um filtro IIR de um filtro FIR é a inclusão de um caminho de realimentação [6]. Um tipo de filtro IIR é o só pólos representado pela seguinte equação de diferenças:

$$y(k) = x(k) + \sum_{j=1}^M p_j y(k-j) \quad (2.3)$$

cuja função de transferência é dado por:

$$H(z) = \frac{1}{1 + \sum_{j=1}^M p_j z^{-j}} \quad (2.4)$$

A expressão (2.3) define um modelo chamado de autoregressivo (autoregressive). Um problema relacionado a esta forma de implementação é o monitoramento da estabilidade do pólo.

O algoritmo adaptativo tem a função de atualizar os coeficientes do filtro. Um algoritmo adaptativo é composto de três itens básicos: definição do algoritmo de minimização, definição da função custo e definição do sinal de erro [3].

2.2.3 Algoritmos de Otimização

A definição do algoritmo de minimização interfere na velocidade de convergência e na complexidade computacional do processo adaptativo. Os métodos de otimização mais utilizados no campo de processamento adaptativo de sinal são o método de Newton, o método Quasi-Newton, e o método do Gradiente [3].

2.2.3.1 Método de Newton

Esse método minimiza uma função custo através de uma aproximação de segunda ordem [3], usando a seguinte fórmula de atualização dos parâmetros:

$$\theta(k+1) = \theta(k) - \mu H_\theta^{-1} \{F[e(k)]\} \nabla_\theta \{F[e(k)]\} \quad (2.5)$$

onde μ é um fator que controla o tamanho do passo do algoritmo, $F[e(k)]$ é a função custo, $H_\theta^{-1} \{F[e(k)]\}$ é a inversa da matriz Hessiana da função custo, e $\nabla_\theta \{F[e(k)]\}$ é o gradiente da função custo com relação aos coeficientes do filtro adaptativo.

2.2.3.2 Método Quasi-Newton

Nesse método é uma derivação do anterior, onde a inversa da matriz Hessiana é estimada através de um cálculo recursivo [3]. A função custo é dada por:

$$\theta(k+1) = \theta(k) - \mu P(k) \nabla_\theta \{F[e(k)]\} \quad (2.6)$$

onde $P(k)$ é uma estimativa de $H_\theta^{-1} \{F[e(k)]\}$, tal que

$$\lim_{k \rightarrow \infty} P(k) = H_\theta^{-1} \{F[e(k)]\} \quad (2.7)$$

Geralmente, utiliza-se o lema da inversão da matriz para estimar a inversa da Hessiana. Da mesma forma, o vetor gradiente também pode ser estimado por um método computacional.

2.2.3.3 Método do Gradiente

Esse tipo de algoritmo minimiza a função custo seguindo a direção contrária ao vetor gradiente da função [3]. A equação de atualização dos parâmetros do filtro é dada por:

$$\theta(k+1) = \theta(k) - \mu \nabla_{\theta} \{F[e(k)]\} \quad (2.8)$$

O método do Gradiente é o mais fácil de implementar, pois é necessário apenas o cálculo do gradiente da função custo. Por outro lado, o método de Newton converge mais rapidamente, precisando de um menor número de interações para aproximar do ponto ótimo. Já o método Quasi-Newton tem um bom compromisso entre ambos. Entretanto, o algoritmo Quasi-Newton está sujeito a problemas de instabilidade devido a sua forma recursiva de estimar a inversa da matriz Hessiana [3].

2.2.4 Função de Custo

A função custo é uma função que depende dos sinais de entrada $x(k)$, de referência $d(k)$, e de saída $y(k)$. Esta deve satisfazer as propriedades de não negatividade e a de optimalidade. Ou seja:

$$F[x(k), d(k), y(k)] \geq 0, \forall x(k), d(k), y(k) \quad (2.9)$$

$$F[x(k), d(k), y(k)] = 0 \quad (2.10)$$

Em um processo adaptativo, o algoritmo adaptativo tenta minimizar a função objetiva de tal forma que $y(k)$ aproxime de $d(k)$, e como consequência, $\theta(k)$ converge para θ_o , onde θ_o é o conjunto de coeficientes ótimos que minimizam a função custo [3]. Como o sinal de erro é uma função dos sinais $x(k)$, $d(k)$ e $y(k)$, a função custo pode ser representada como uma função direta do sinal de erro. Ou seja, $F = F[e(k)] = F[e(x(k), y(k), d(k))]$ [3].

A complexidade do vetor gradiente e o cálculo da matriz Hessiana dependem da escolha da função custo. Uma lista das funções objetivas mais utilizadas na implementação de algoritmos adaptativos está apresentada abaixo [3]:

- Mean-Square Error (MSE):

$$F[e(k)] = E[|e(k)|^2] \quad (2.11)$$

- Least Squares (LS):

$$F[e(k)] = \frac{1}{k+1} \sum_{i=0}^k |e(k-i)|^2 \quad (2.12)$$

- Weighted Least Squares (WLS):

$$F[e(k)] = \sum_{i=0}^k \lambda^i |e(k-i)|^2 \quad (2.13)$$

onde λ é uma constante menor que 1;

- Instantaneous Square Value (ISV):

$$F[e(k)] = |e(k)|^2 \quad (2.14)$$

Como a função MSE utiliza um operador esperança, cuja estatística necessita de uma infinita quantidade de informação para ser medida, é impossível ser implementada na prática. As outras funções objetivas tendem a aproximar desta função ideal. Elas diferem entre si em complexidade de implementação e nas características de convergência [3].

2.2.5 Definição do sinal de erro

A escolha do sinal de erro afeta várias características do algoritmo, tais como, a complexidade computacional, a velocidade de convergência, robustez, a ocorrência de polarização e soluções múltiplas [3].

2.3 Filtro de Wiener

A solução de Wiener representa a solução mínima do erro médio quadrático (MSE) de um filtro linear discreto no tempo [6]. Essa solução depende da matriz de auto-correlação do sinal de entrada e da correlação cruzada entre o sinal de entrada e o sinal de referência.

Como citado anteriormente, o filtro FIR na forma direta, ou transversal, é o mais utilizado para realizar um filtro adaptativo [3] [5]. Este possui uma relação simples e analiticamente calculável entre os seus parâmetros e sua função de transferência [5], facilitando o cálculo da solução ótima do filtro adaptativo. Sua saída é dada pela seguinte expressão:

$$y(k) = \sum_{i=0}^N w_i(k)x(k-i) = \mathbf{w}^T(k)\mathbf{x}(k) \quad (2.15)$$

onde $\mathbf{x}(k) = [x(k) \ x(k-1) \ \dots \ x(k-N)]^T$ e $\mathbf{w}(k) = [w_0(k) \ w_1(k) \ \dots \ w_N(k)]^T$ são, respectivamente, vetores de entrada e dos parâmetros do filtro. A ordem do filtro adaptativo é representada por N . Como o sinal de saída $y(k)$ é uma combinação

linear dos coeficientes do filtro, existe apenas uma única solução ótima para o erro médio quadrático (Mean Square Error - MSE) [3].

Utilizando a função custo do erro médio quadrático (2.11) e a equação (2.15), obtém-se:

$$E[e^2(k)] = \xi(k)$$

Considerando o sinal de erro como:

$$e(k) = d(k) - y(k) = d(k) - \mathbf{w}^T(k)\mathbf{x}(k) \quad (2.16)$$

Obtém-se:

$$\begin{aligned} \xi(k) &= E[d^2(k) - 2d(k)y(k) + y^2(k)] \\ \xi(k) &= E[d^2(k) - 2d(k)\mathbf{w}^T(k)\mathbf{x}(k) + \mathbf{w}^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}(k)] \\ \xi(k) &= E[d^2(k)] - 2E[d(k)\mathbf{w}^T(k)\mathbf{x}(k)] + E[\mathbf{w}^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}(k)] \end{aligned}$$

Para um filtro com os coeficientes fixos, a função MSE é dada por:

$$\begin{aligned} \xi(k) &= E[d^2(k)] - 2\mathbf{w}^T E[d(k)\mathbf{x}(k)] + \mathbf{w}^T E[\mathbf{x}(k)\mathbf{x}^T(k)]\mathbf{w} \\ \xi(k) &= E[d^2(k)] - 2\mathbf{w}^T \mathbf{p} + \mathbf{w}^T \mathbf{R} \mathbf{w} \end{aligned} \quad (2.17)$$

onde $\mathbf{p} = E[d(k)\mathbf{x}(k)]$ é o vetor de correlação cruzada entre o sinal desejado e o sinal de entrada, e $\mathbf{R} = E[\mathbf{x}(k)\mathbf{x}^T(k)]$ é a matriz de auto-correlação do sinal de entrada. Pode-se observar na expressão (2.17) que a função objetiva é uma função quadrática com relação aos coeficientes do filtro. Logo, a minimização desta função leva a uma solução ótima única, obtendo um vetor dos coeficientes ótimos, \mathbf{w}_o .

O vetor gradiente da função custo (2.17), com relação aos parâmetros do filtro, é dado por:

$$\mathbf{g}_w = \frac{\partial \xi}{\partial \mathbf{w}} = \left[\frac{\partial \xi}{\partial w_0} \frac{\partial \xi}{\partial w_1} \dots \frac{\partial \xi}{\partial w_N} \right] = -2\mathbf{p} + 2\mathbf{R}\mathbf{w} \quad (2.18)$$

Igualando o gradiente a zero e assumindo que \mathbf{R} é uma matriz não singular, obtém-se a solução de Wiener:

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p} \quad (2.19)$$

Na prática, uma estimativa precisa de \mathbf{R} e \mathbf{p} não é possível de se calcular. Quando o sinal de entrada e o sinal desejado são ergódicos¹, é possível utilizar o tempo médio (time averages) para estimar \mathbf{R} e \mathbf{p} , o que é, implicitamente, feito pela maioria dos algoritmos adaptativos [3].

¹Um processo $x(n)$ é ergódico se todas as suas propriedades estatísticas se podem determinar a partir de uma única realização. Para tal, o processo deverá ser estacionário, podendo-se, então, substituir médias de conjuntos por médias temporais.

2.4 Algoritmo LMS (Least Mean Square)

Utilizando um algoritmo baseado no método do gradiente, (2.8), pode-se aproximar da solução de Wiener (2.19) da seguinte forma [3]:

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \hat{\mathbf{g}}_{\mathbf{w}}(k) = \mathbf{w}(k) + 2\mu(\hat{\mathbf{p}}(k) - \hat{\mathbf{R}}(k)\mathbf{w}(k)) \quad (2.20)$$

para $k = 0, 1, 2, \dots$, onde $\hat{\mathbf{g}}_{\mathbf{w}}(k)$ representa uma estimativa do vetor gradiente da função custo com relação aos coeficientes do filtro.

A matriz de auto-correlação \mathbf{R} e o vetor de correlação cruzada \mathbf{p} , são estimados pelos seus valores instantâneos da seguinte forma:

$$\hat{\mathbf{R}}(k) = \mathbf{x}(k)\mathbf{x}^T(k) \quad (2.21)$$

$$\hat{\mathbf{p}}(k) = d(k)\mathbf{x}(k) \quad (2.22)$$

A estimativa do vetor gradiente fica:

$$\begin{aligned} \hat{\mathbf{g}}_{\mathbf{w}}(k) &= -2d(k)\mathbf{x}(k) + 2\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}(k) = 2\mathbf{x}(k)(-d(k) + \mathbf{x}^T(k)\mathbf{w}(k)) \\ \hat{\mathbf{g}}_{\mathbf{w}}(k) &= -2e(k)\mathbf{x}(k) \end{aligned} \quad (2.23)$$

Estimar \mathbf{R} e \mathbf{p} por (2.21) e (2.22) é equivalente a substituir a função custo MSE (2.11) por ISV (2.14). Isto pode ser demonstrado derivando-se a função objetiva ISV com relação aos coeficientes do filtro:

$$\frac{\partial e^2}{\partial \mathbf{w}} = \left[2e(k) \frac{\partial e}{\partial \mathbf{w}_0} \quad 2e(k) \frac{\partial e}{\partial \mathbf{w}_1} \quad \dots \quad 2e(k) \frac{\partial e}{\partial \mathbf{w}_N} \right]^T = -2e(k)\mathbf{x}(k) = \hat{\mathbf{g}}_{\mathbf{w}}(k) \quad (2.24)$$

que é o mesmo resultado obtido em (2.23).

A equação de atualização dos parâmetros do filtro torna-se:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + 2\mu e(k)\mathbf{x}(k) \quad (2.25)$$

onde μ representa o passo de busca. Este deve ser escolhido em uma faixa que garanta a convergência do algoritmo.

O algoritmo LMS fica da seguinte forma:

$$\begin{aligned} \mathbf{x}(0) &= \mathbf{w}(0) = [0 \ 0 \ \dots \ 0]^T \\ \text{Para } k &\geq 0, \text{ executar:} \\ e(k) &= d(k) - \mathbf{x}^T(k)\mathbf{w}(k) \\ \mathbf{w}(k+1) &= \mathbf{w}(k) + 2\mu e(k)\mathbf{x}(k) \end{aligned}$$

2.5 Algoritmo NLMS (Normalized LMS)

O algoritmo LMS normalizado converge mais rápido que o algoritmo LMS, por utilizar um fator de convergência variável [3]. A equação de atualização é expressa como:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + 2\mu_k e(k)\mathbf{x}(k) = \mathbf{w}(k) + \Delta\mathbf{w}'(k) \quad (2.26)$$

onde μ_k é o fator de convergência. Este deve ser escolhido para o algoritmo convergir rapidamente.

O erro quadrático instantâneo é dado por:

$$e^2(k) = d^2(k) + \mathbf{w}^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}(k) - 2d(k)\mathbf{w}^T(k)\mathbf{x}(k) \quad (2.27)$$

Alterando o vetor dos coeficientes de tal forma que:

$$\mathbf{w}'(k) = \mathbf{w}(k) + \Delta\mathbf{w}'(k)$$

O erro quadrático (2.27) torna-se:

$$\begin{aligned} e'^2(k) &= e^2(k) + 2\Delta\mathbf{w}'^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}(k) + \Delta\mathbf{w}'^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\Delta\mathbf{w}'(k) \\ &\quad - 2d(k)\Delta\mathbf{w}'^T(k)\mathbf{x}(k) \end{aligned}$$

Como:

$$\Delta e^2(k) \triangleq e'^2(k) - e^2(k)$$

Obtém-se:

$$\begin{aligned} \Delta e^2(k) &= 2\Delta\mathbf{w}'^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}(k) + \Delta\mathbf{w}'^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\Delta\mathbf{w}'(k) \\ &\quad - 2d(k)\Delta\mathbf{w}'^T(k)\mathbf{x}(k) \end{aligned} \quad (2.28)$$

Considerando que:

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

A expressão (2.28) pode ser escrita da seguinte forma:

$$\Delta e^2(k) = -2\Delta\mathbf{w}'^T(k)\mathbf{x}(k)e(k) + \Delta\mathbf{w}'^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\Delta\mathbf{w}'(k) \quad (2.29)$$

Como:

$$\Delta\mathbf{w}'(k) = 2\mu_k e(k)\mathbf{x}(k) \quad (2.30)$$

Substituindo (2.30) em (2.29), obtém-se:

$$\Delta e^2(k) = -4\mu_k e^2(k)\mathbf{x}^T(k)\mathbf{x}(k) + 4\mu_k^2 e^2(k)[\mathbf{x}(k)\mathbf{x}^T(k)]^2 \quad (2.31)$$

O valor de μ_k tal que $\frac{\partial \Delta e^2(k)}{\partial \mu_k} = 0$ é dado por:

$$\mu_k = \frac{1}{2\mathbf{x}^T(k)\mathbf{x}(k)} \quad (2.32)$$

Substituindo (2.32) em (2.31), obtém-se $\Delta e^2(k) = -e^2(k)$, que corresponde ao seu ponto mínimo.

Assim sendo, a equação de atualização (2.26) torna-se:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \frac{e(k)\mathbf{x}(k)}{\mathbf{x}^T(k)\mathbf{x}(k)} \quad (2.33)$$

Um fator de convergência fixo μ_n é introduzido na fórmula (2.33) para controlar o desajuste, já que a fórmula foi baseada em valores instantâneos dos erros quadráticos e não do MSE. Um parâmetro γ também é introduzido para prevenir um valor muito elevado para passo de atualização, quando $\mathbf{x}^T(k)\mathbf{x}(k)$ torna-se muito pequeno. Finalmente, a fórmula de atualização dos coeficientes do filtro adaptativo fica da seguinte forma:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \frac{\mu_n}{\gamma + \mathbf{x}^T(k)\mathbf{x}(k)} e(k)\mathbf{x}(k) \quad (2.34)$$

O algoritmo NLMS está apresentado abaixo [3].

$$\mathbf{x}(0) = \mathbf{w}(0) = [0 \ 0 \ \dots \ 0]^T$$

Escolher um valor para μ_n de tal forma que: $0 < \mu_n \leq 2$

Escolher um valor pequeno para γ .

Para, $k \geq 0$, executar:

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

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \frac{\mu_n}{\gamma + \mathbf{x}^T(k)\mathbf{x}(k)} e(k)\mathbf{x}(k)$$

2.6 Algoritmo RLS (Recursive Least Square)

O algoritmo RLS baseia-se em um cálculo na forma recursiva para encontrar a solução que minimiza a soma dos erros quadráticos. Este algoritmo é conhecido por sua rápida convergência, independente do espalhamento dos autovalores da matriz de auto-correlação do sinal de entrada. Tem um excelente desempenho quando empregado em sistemas variantes no tempo. Porém, possui uma maior complexidade computacional e sofre problemas de estabilidade, que não ocorre com o algoritmo LMS [10] [11].

Dado um filtro FIR realizado na forma direta com o sinal de entrada dado por $\mathbf{x}(k) = [x(k) \ x(k-1) \ \dots \ x(k-N)]^T$, onde N é a ordem do filtro, e os coeficientes do

filtro dado por $\mathbf{w}(k) = [w_0(k) w_1(k) \dots w_N(k)]^T$, como anteriormente. Os coeficientes são adaptados para minimizar a função custo, WLS, que no caso dos algoritmos *least-squares* é uma função determinística, dada por:

$$\begin{aligned}\xi^d(k) &= \sum_{i=0}^k \lambda^{k-i} e^2(i) \\ &= \sum_{i=0}^k \lambda^{k-i} [d(i) - \mathbf{x}^T(i) \mathbf{w}(k)]^2\end{aligned}\quad (2.35)$$

onde $e(i)$ é o erro de saída no instante i . O parâmetro λ é um fator de peso exponencial que deve ser escolhido na faixa $0 \ll \lambda \leq 1$. É também chamado fator de esquecimento, desde que a informação passada tem um peso que diminui exponencialmente o seu efeito na atualização dos coeficientes do filtro.

Diferenciando $\xi^d(k)$ em relação ao $\mathbf{w}(k)$, tem-se que:

$$\frac{\partial \xi^d(k)}{\partial \mathbf{w}(k)} = -2 \sum_{i=0}^k \lambda^{k-i} \mathbf{x}(i) [d(i) - \mathbf{x}^T(i) \mathbf{w}(k)] \quad (2.36)$$

Igualando a expressão (2.36) a zero, para encontrar o vetor ótimo $\mathbf{w}(k)$ que minimiza a função custo, tem-se que:

$$\sum_{i=0}^k \lambda^{k-i} \mathbf{x}(i) \mathbf{x}^T(i) \mathbf{w}(k) - \sum_{i=0}^k \lambda^{k-i} \mathbf{x}(i) d(i) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (2.37)$$

Isolando $\mathbf{w}(k)$ da expressão (2.37), obtém-se o vetor dos coeficientes ótimos:

$$\begin{aligned}\mathbf{w}(k) &= \left[\sum_{i=0}^k \lambda^{k-i} \mathbf{x}(i) \mathbf{x}^T(i) \right]^{-1} \sum_{i=0}^k \lambda^{k-i} \mathbf{x}(i) d(i) \\ &= \mathbf{R}_D^{-1}(k) \mathbf{p}_D(k)\end{aligned}\quad (2.38)$$

onde $\mathbf{R}_D(k)$ e $\mathbf{p}_D(k)$ são chamados matriz de correlação determinística do sinal de entrada e vetor de correlação cruzada determinístico entre os sinais de entrada e desejado, respectivamente.

Na equação (2.38), assume-se que a matriz $\mathbf{R}_D(k)$ é não singular. Caso contrário, deve ser utilizada uma matriz inversa generalizada para obter uma solução para $\mathbf{w}(k)$ que minimiza $\xi^d(k)$ [12].

No algoritmo RLS convencional, é utilizado o lema da inversão de matriz para o cálculo da matriz inversa. Desta forma, a complexidade computacional é menor,

comparada com o cálculo direto da matriz inversa. Assim sendo, a inversa da matriz de correlação determinística é calculada da seguinte forma:

$$\mathbf{S}_D(k) = \mathbf{R}_D^{-1}(k) = \frac{1}{\lambda} \left[\mathbf{S}_D(k-1) - \frac{\mathbf{S}_D(k-1)\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{S}_D(k-1)}{\lambda + \mathbf{x}^T(k)\mathbf{S}_D(k-1)\mathbf{x}(k)} \right] \quad (2.39)$$

O algoritmo RLS convencional está apresentado abaixo.

$$\mathbf{S}_D(-1) = \delta \mathbf{I}$$

onde δ pode ser a inversa da potência do sinal de entrada estimado

$$\mathbf{p}_D(-1) = \mathbf{x}(-1) = [0 \ 0 \ \dots \ 0]^T$$

Para $k \geq 0$, executar:

$$\mathbf{S}_D(k) = \frac{1}{\lambda} \left[\mathbf{S}_D(k-1) - \frac{\mathbf{S}_D(k-1)\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{S}_D(k-1)}{\lambda + \mathbf{x}^T(k)\mathbf{S}_D(k-1)\mathbf{x}(k)} \right]$$

$$\mathbf{p}_D(k) = \lambda \mathbf{p}_D(k-1) + d(k)\mathbf{x}(k)$$

$$\mathbf{w}(k) = \mathbf{S}_D(k)\mathbf{p}_D(k)$$

Se necessário, calcule:

$$y(k) = \mathbf{w}^T(k)\mathbf{x}(k)$$

$$e(k) = d(k) - y(k)$$

Capítulo 3

Eco na rede telefônica

O atraso nas ligações telefônicas passou a causar problemas durante os anos de 1920 a 1930, quando surgiram os primeiros circuitos a longa distância. Nesta época, circuitos de distância maior que 800 Km (aproximadamente 500 milhas) eram considerados circuitos de longa distância [13]. Este atraso se tornou suficientemente grande para causar ecos que incomodavam o assinante. Isto motivou a pesquisa de métodos de controle de eco na rede de telefonia pública comutada (PSTN). Foi então que surgiram os atenuadores, bobinas híbridas com autobalanceamento de impedância, os supressores de eco e, posteriormente, os canceladores de eco.

3.1 A Origem do Eco

A Figura 3.1 representa, de forma simplificada, o circuito envolvido em uma conversa entre duas pessoas através de uma ligação telefônica. Neste diagrama, cada linha representa um par de fios. Os assinantes estão conectados por um circuito de quatro fios durante todo o percurso da ligação. Como indicado na figura por P_1 e P_2 , existem perdas na linha que devem ser compensadas pelos ganhos G_1 e G_2 . As perdas não são concentradas em um ponto como demonstra a figura, mas estão distribuídas em todo o circuito. Ocorrem também atrasos nas linhas, A_1 e A_2 , que não podem ser compensados. Este atraso é dependente do tipo de circuito utilizado. Por exemplo, a transmissão pode ser feita por satélite, fibras ópticas, fios de cobre ou ondas de rádio. O atraso não pode ser menor que o tempo equivalente à distância entre os assinantes dividido pela velocidade da luz. No caso de uma ligação local, onde poucos equipamentos estão envolvidos, o atraso total envolvido na ligação será bem próximo deste valor mínimo [13].

O atraso interfere na conversação de tal forma que, levando em consideração a Figura 3.1, quando o assinante A terminar de falar, este só poderá escutar a resposta do assinante B após um tempo equivalente a soma dos atrasos envolvidos, ou seja, após $t = A_1 + A_2$.

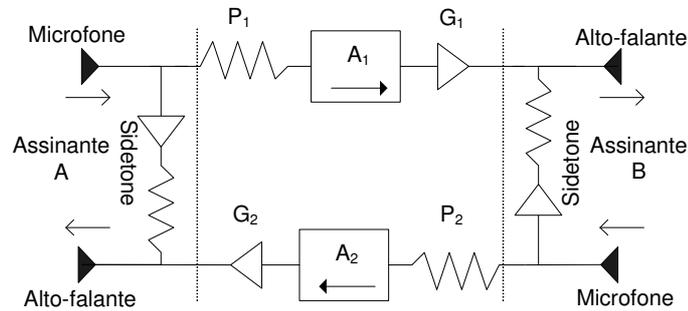


Figura 3.1 Diagrama simplificado de uma ligação em um circuito a quatro fios

Circuitos de longa distância necessitam de um ganho elevado, por este motivo, é utilizado circuitos de quadro fios que permitem o uso deste ganho sem aparecer ruídos na linha. Em circuitos curtos, como a linha que liga o assinante à central e o entroncamento local, é mais viável economicamente o uso de circuitos a dois fios [13]. Como existe um número elevado deste tipo de circuito, ao reduzir um par de fios se economiza muito cobre. Porém, é necessário ligar os dois tipos de circuitos e, assim, surge um novo problema. Para não ocorrer reflexão do sinal no fio, é necessário um casamento perfeito de impedâncias entre os dois meios. Como representado na Figura 3.2, a conexão entre os circuitos de dois fios e os circuitos de quatro fios é realizada através de bobinas híbridas, H_1 , H_2 , H_3 e H_4 . Como o casamento entre as impedâncias da linha e da rede, R_1 , R_2 , R_3 e R_4 , nunca é perfeito, uma parcela do sinal transmitido retorna da híbrida causando o eco, representado na Figura 3.2 por e_1 e e_2 .

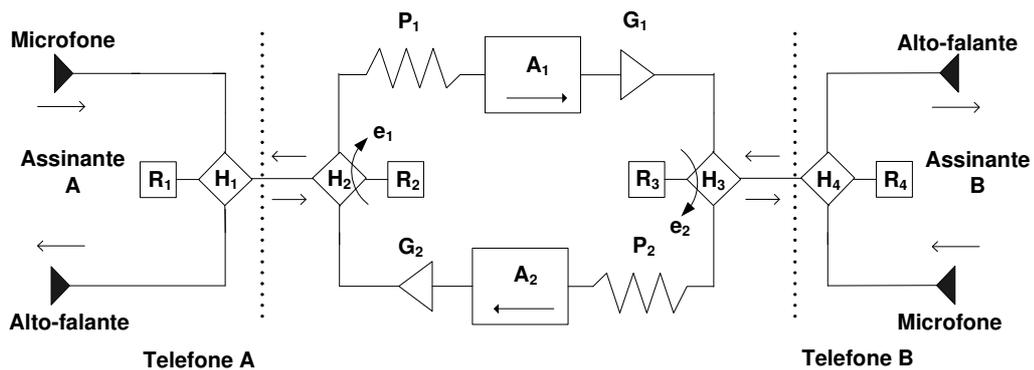


Figura 3.2 Diagrama de uma Rede Telefônica Básica

Em um sistema de transmissão, qualquer descasamento de impedâncias causará

uma reflexão da energia de volta à fonte e poderá causar eco. No caso do circuito a quatro fios, como os amplificadores são dispositivos que transmitem o sinal unidirecionalmente, o eco não consegue se propagar ao longo da linha, ficando confinado em pequenas seções do sistema. Estas reflexões não interferem no sinal que chega ao assinante. Entretanto, no circuito a dois fios, em qualquer ponto que ocorra o descasamento da impedância, o eco pode retornar na linha com pouca atenuação. Na literatura é bastante comum considerar que a reflexão é causada apenas na bobina híbrida. Isto é permitido quando o sistema está em repouso. Entretanto, para problemas envolvendo os transitórios na rede, deve-se considerar os pontos que ocorrem reflexões no circuito a dois fios [13].

Na Figura 3.2 está ilustrada uma situação típica de uma ligação telefônica entre dois assinantes. Os aparelhos telefônicos possuem híbridas, H_1 e H_4 , que convertem uma transmissão a quatro fios, um par do microfone e um par do alto-falante, em uma transmissão a dois fios, para conectar o telefone na linha do assinante. Na central, outras híbridas, H_2 e H_3 , convertem novamente de dois para quatro fios, para transmissão em longa distância. Em uma conversa entre dois assinantes A e B , a fala retorna de cada uma das híbridas. O sinal que retorna no próprio telefone do assinante, devido à híbrida interna, é chamado de *sidetone* [13]. Este tipo de eco possui um atraso nulo, ocorre quase que instantaneamente com a fala. Não causa perda na qualidade da ligação. No caso do assinante A falando e o assinante B escutando, a fala retorna tanto na híbrida H_2 quanto na híbrida H_3 . Como o tempo de propagação na linha do assinante é curto, a primeira é mascarada pelo *sidetone*. Enquanto a segunda, causa o chamado *eco do falante* [13]. Este pode possuir um atraso significativo, que leva o assinante A a escutar o eco de sua voz. O sinal refletido em H_3 pode refletir novamente em H_1 e H_2 causando o chamado *eco do ouvinte* [13], que é escutado pelo assinante B após ele escutar a fala original do A . Esse processo é repetido até que a perda no caminho circulante reduz o eco de tal forma que o assinante não possa mais escutá-lo. Em um circuito telefônico típico, apenas o *eco do falante* é controlado, uma vez que se este for eliminado, o *eco do ouvinte* também será eliminado.

3.2 O Efeito do Eco

O eco torna-se um incômodo a medida que o atraso aumenta. Ecos com pouco atraso são mascarados pela fala e são até mesmo agradáveis de se ouvir durante a fala, dando a impressão de que o ouvinte está escutando bem o que está sendo dito e que a linha não está “muda”. Mas quando o atraso aumenta, de tal forma que ao terminar a fala o assinante percebe a própria voz, o eco passa a degradar a qualidade da ligação. Por exemplo, se o tempo de *round-trip delay* é de aproximadamente 250ms, o eco se torna bastante irritante mesmo com um nível bastante baixo [13].

Testes para determinar a tolerância ao eco foram reportados por Clark e Mathes em 1925 [14]. Em 1953, Phillips [1] repetiu este trabalho. O resultado obtido está mostrado na Figura 3.3. O gráfico representa a relação entre o atraso total do eco, *round-trip delay*, e a menor atenuação que o eco deve sofrer, *echo path round-trip loss*, para a ligação atingir uma boa qualidade na opinião de vários ouvintes. A curva representa um valor médio da tolerância ao eco. Aproximadamente 68% dos ouvintes tem uma tolerância ao eco de 5 dB acima ou abaixo da curva, enquanto os 32% restantes estão fora desta faixa. Além disto, levando em consideração a Figura 3.2, e que as perdas na rede são iguais em ambas as direções; neste estudo, também foi determinado que, na prática a perda no caminho de eco é maior que o dobro da perda na rede mais a perda de retorno na híbrida H_3 , ou seja, $P_T > 2 \times P_1 + P_{H_3}$, sendo $P_1 = P_2$.

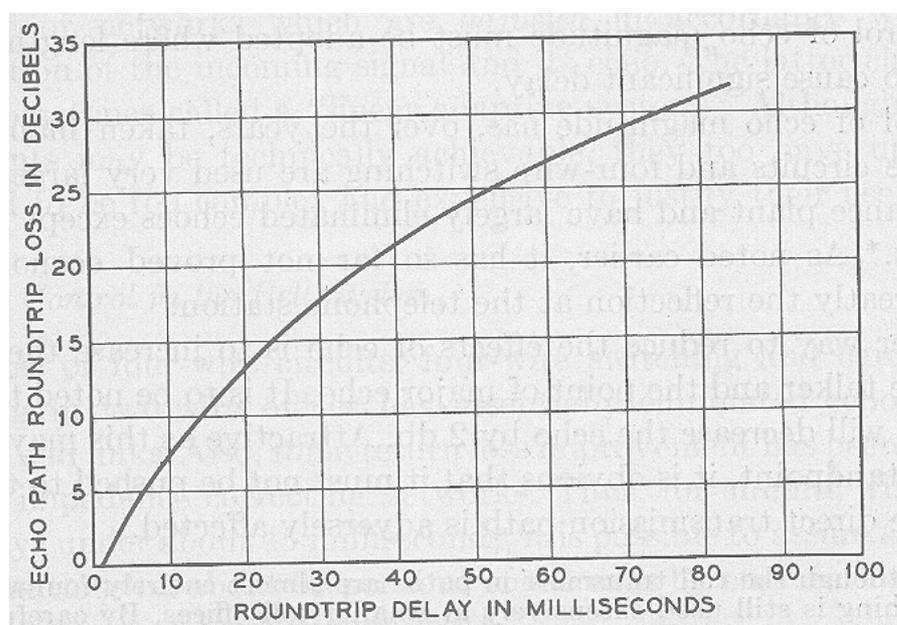


Figura 3.3 Gráfico da atenuação versus atraso do eco - reproduzido de [1]

3.3 Métodos de Controle do Eco

3.3.1 Atenuador

Como diminuir o atraso na rede não é uma forma simples de se eliminar o eco, as soluções encontradas baseiam-se em controlar a amplitude do eco. O eco pode ter

seu sinal atenuado introduzindo uma perda na linha no circuito de quatro fios, tanto na transmissão, quanto na recepção. Observando a Figura 3.2 estas perdas podem ser colocadas no circuito em série com os ganhos G_1 e G_2 . Porém, o sinal de fala também será atenuado. Para cada x dB de atenuação no sinal de fala, o sinal de eco será atenuado de $2x$ dB. Apesar de ser um método simples de se colocar em prática, a perda introduzida no sistema não pode ser elevada, senão o sinal transmitido terá pouca qualidade, atingindo níveis impraticáveis. O atenuador de eco pode ser utilizado para circuitos com um tempo de *round-trip delay* menor que 45ms [13]. Para atrasos maiores, a perda que deve ser introduzida é tanta que atenua sinal de fala além do nível aceitável. Nestes casos são utilizados supressores de eco.

3.3.2 Supressor de Eco

No final dos anos 50, do século passado, surgiu o primeiro supressor de eco. No início, foram utilizados para controlar o eco gerado por circuitos de satélite [15]. O supressor de eco consiste, basicamente, de dois comutadores operados por voz que interrompem o caminho de retorno do eco. A Figura 3.4 ilustra um modelo de um supressor de eco simples. Quando o assinante em A fala, sua voz é transmitida através do circuito a quatro fios no caminho 1, isto faz com que o comutador C_B opere e desabilite o caminho 2 antes que o eco e_B gerado em H_B retorne. Similarmente, uma fala em B causa a operação no comutador C_A que bloqueia o sinal e_A gerado em H_A .

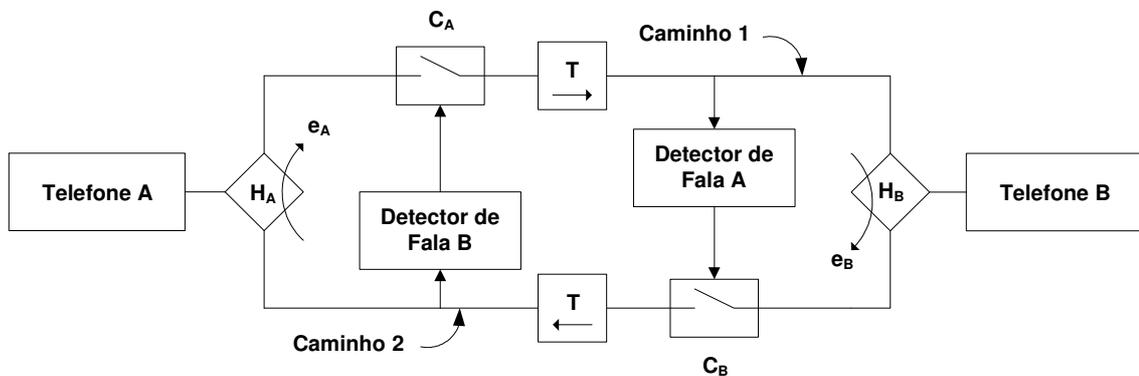


Figura 3.4 Diagrama de um supressor de eco

Na prática, os supressores de eco utilizados são mais sofisticados que o apresentado na Figura 3.4. São empregados em circuitos de 1600 a 2400 Km. Podem reduzir o eco a um nível satisfatório, para um atraso de *round-trip* de aproximadamente 50 a 100ms.

Embora os supressores de eco reduzem o eco causado por problemas de transmissão na rede, eles também causam uma distorção nas primeiras sílabas, um ajuste artificial do volume e eliminam a capacidade de *dupla fala*, reduzindo a habilidade do sistema em permitir uma conversação natural [15]. Esse dispositivo tem a desvantagem de tornar a comunicação unidirecional, não permitindo a transmissão da fala em ambas direções [13]. Outro problema que pode ocorrer é o chamado *lock-out*. Este efeito é o bloqueio total da transmissão. Ocorre quando vários supressores de eco são usados em *tandem*. Como mostrado na Figura 3.5, pode existir um certo atraso T entre dois supressores em uma ligação telefônica. Quando ocorre uma interrupção na conversa, os caminhos nos dois sentidos ficam disponíveis. Neste estado, o assinante em B começa a falar e o comutador C_3 é operado. O sinal leva um tempo T para chegar no comutador C_1 . Antes deste tempo, o assinante em A começa a falar sem saber que o assinante em B está falando, causando a operação do comutador C_2 . Ambos os comutadores interrompem o circuito. Como um assinante não está escutando o outro, até que um deles pare de falar, o circuito poderá ficar interrompido por um período de tempo. A. W. Horton [16] estudou este fenômeno e mostrou como calcular a probabilidade de ocorrência do mesmo.

Vários artigos na literatura especializada descrevem as características, vantagens e desvantagens do supressor de eco [13] [17] [18].

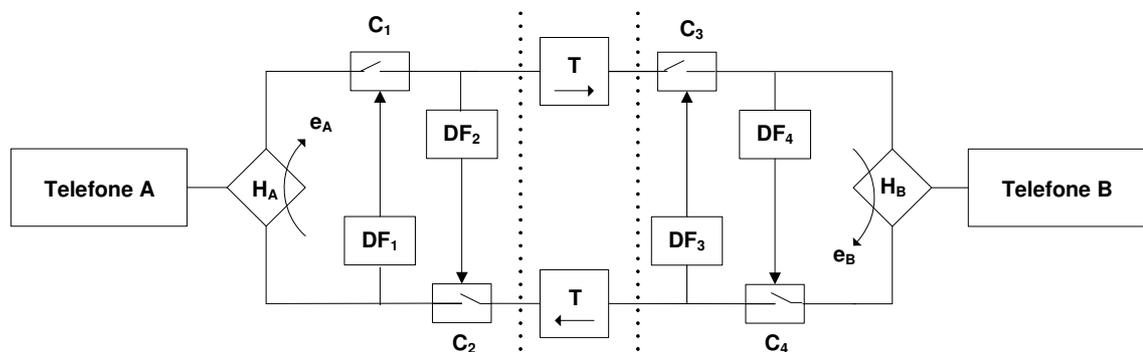


Figura 3.5 Diagrama de dois Supressores de Eco em Tandem

3.3.3 Bobina Híbrida com Autobalanceamento de Impedância

Foram feitos estudos para desenvolver bobinas híbridas com autobalanceamento de impedância. A maioria dos esquemas envolve a medida de corrente e tensão associados com a impedância da linha e introduz correntes ou tensões de

compensação para balanceamento. Na prática este método se mostrou bastante complexo e caro para que justificasse seu uso [13].

3.3.4 Cancelador de Eco

3.3.4.1 Histórico

A idéia do cancelador de eco foi originalmente proposta por *B.F. Logan* e *J.L.Kelly Jr* [19]. As primeiras publicações foram feitas em 1966 através de dois artigos publicados no *Bell Systems Technical Journal* por *Sondhi & Presti* [19], e *Becker & Rudin* [20]. Em 1967, *Sondhi* fez uma nova publicação mais detalhada [21].

Após o desenvolvimento da teoria nos laboratórios AT&T, o primeiro cancelador de eco foi produzido no final da década por COMSAT TeleSystem [15]. Baseado em processos analógicos, o cancelador de eco foi implementado para testar seu desempenho na rede de comunicação via satélite para ligações longa distância intercontinental. Foi comercialmente inviável devido ao imenso tamanho físico e alto custo de produção.

No final da década de 70, com o desenvolvimento da tecnologia de circuitos digitais, foi possível a produção dos primeiros canceladores de eco pela COMSAT [15]. Foram baseados na tecnologia de circuitos digitais utilizando interface analógica com a rede. Entretanto, o intenso processamento necessário para atualizar os coeficientes do filtro e o elevado custo inviabilizavam o seu uso comercial.

No início dos anos 80, com a evolução dos semicondutores, foram desenvolvidos canceladores de eco com interface mais sofisticada e sistemas multicanais baseados na nova tecnologia de processadores digitais de sinais [15]. A microeletrônica avançou em uma taxa elevada e os dispositivos tornaram-se cada vez mais baratos. Este desenvolvimento tecnológico viabilizou o uso do cancelador de eco não somente para circuitos via satélite, mais também para circuitos de longa distância terrestres. Os canceladores de eco desenvolvidos tiveram um desempenho superior ao supressor de eco, melhorando a qualidade das ligações.

Nos últimos 20 anos, a comunidade científica tem tido um grande esforço para desenvolver canceladores de eco que utilizam outras estruturas de filtros mais eficientes, algoritmos adaptativos com convergência mais rápida e com menor esforço computacional. Entretanto, atualmente, os canceladores de eco ainda se baseiam no conceito proposto por *Kelly* e *Logan*, utilizando um filtro com a estrutura transversal e um algoritmo adaptativo simples baseado no método do gradiente (*steepest descent*). O algoritmo do mínimo quadrado médio normalizado (NLMS) é o mais empregado para este fim.

3.3.4.2 Conceito

Como o supressor de eco não tinha um resultado satisfatório, J. L. Flanagan e D. W. Hagelberger sugeriram um dispositivo de laço aberto que consistia de um filtro transversal que aproximava a resposta impulsiva do sistema. O eco foi tratado como uma versão linearmente filtrada do sinal de fala original. Um trem de pulsos era transmitido através do circuito para medir a sua resposta impulsiva e, assim, configurar os coeficientes do filtro que modela o sistema. O eco é cancelado subtraindo o sinal sintetizado pelo filtro do sinal de retorno. Porém, o caminho de eco não é constante. Além da brusca mudança devido à conexão ou desconexão de telefones de extensão durante a conversação, ou transferência de chamadas via PABX, também existem mudanças lentas em ganho e outras flutuações da função de transferência do caminho de eco [21]. Então, para um dispositivo de loop aberto funcionar na prática, é necessário ajustar o filtro transversal durante a conversação. A transmissão contínua de trens de pulsos para tal ajuste seria intolerável para o assinante.

A proposta feita por *Kelly e Logan* contorna essas dificuldades. O sinal de fala é utilizado no lugar do trem de pulsos e um laço de controle continuamente adapta o filtro transversal para as mudanças no caminho. O diagrama do cancelador proposto está representado na Figura 3.6. Pode-se observar que o cancelador é apenas um filtro adaptativo cujos coeficientes estão representados por W . Estes são adaptados de acordo com o sinal resultante da multiplicação do sinal de entrada, $x(t)$, com o sinal de erro, $e(t)$, ponderado pela constante K .

Utilizando o filtro adaptativo, o cancelador de eco sintetiza uma réplica do eco, $y'(t)$, e o subtrai do sinal de retorno da híbrida, $y(t)$, que é composto pelo eco, $ax(t)$, mais um sinal de ruído $r(t)$. O sinal residual é o sinal de erro $e(t)$.

Neste esquema, considera-se que o assinante do outro lado da linha não está falando. Desta forma, o ideal seria um sinal de erro nulo. No caso em que os dois assinantes estão falando, a adaptação dos coeficientes é congelada e o sinal de erro deve ser equivalente ao sinal de fala do assinante remoto.

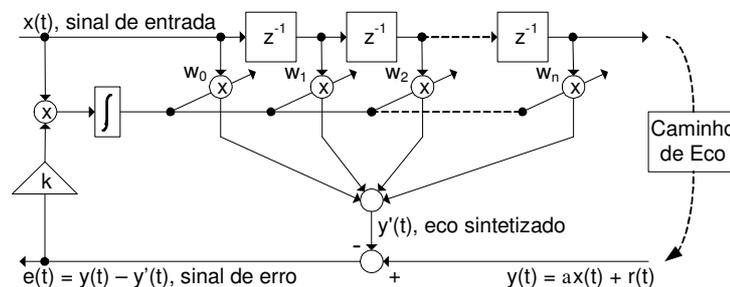


Figura 3.6 Diagrama esquemático do Cancelador de Eco

Capítulo 4

Qualidade de Serviço em VoIP

Na área de redes, a qualidade de serviço (QoS) representa a capacidade da rede em fornecer um serviço diferenciado para um determinado tráfego. Este termo refere-se a um conjunto de parâmetros que garantem o desempenho em relação à qualidade e disponibilidade de serviço, compreendendo atraso máximo, largura de banda e prioridade dos pacotes. Em processamento de sinais digitais, a qualidade de serviço envolve a medida da qualidade do sinal de fala, que depende do codificador utilizado, além de blocos funcionais responsáveis pela melhoria do sinal, como os canceladores de eco, algoritmos de tratamento de pacotes perdidos, *buffer de jitter*, supressores de silêncio e geração de ruído de fundo de conforto. Esse capítulo apresenta os principais conceitos relacionados à qualidade de serviço e diversas técnicas empregadas em voz sobre IP (VoIP).

4.1 Uma Visão Geral de VoIP

A tendência atual é a convergência entre a rede de circuitos comutados e a rede de pacotes. Essa é a maior prioridade para muitos gerentes de redes, fabricante de equipamentos e provedores de serviços. As organizações estão buscando soluções que irão possibilitá-las utilizar o excesso da capacidade das redes para transmissão de voz e dados. É bem provável que depois que a telefonia IP estiver bem consolidada, aplicações integradas de multimídia interativa, como conferências virtuais utilizando áudio e vídeo, passarão a fazer parte do nosso cotidiano.

Resumidamente, VoIP (Voice over Internet Protocol) é a tecnologia que utiliza o protocolo da internet (IP) para transmitir voz sobre uma rede de dados, podendo ser internet ou intranet. A principal vantagem de se utilizar a rede de pacotes para transmitir voz é o aumento da eficiência e diminuição dos custos. O provedor de serviços VoIP pode aumentar rapidamente o número de serviços prestados e utilizar a largura de banda existente de forma eficiente, atendendo melhor a satisfação do cliente [22]. Pode-se disponibilizar serviços para transmissão de voz e de dados

utilizando os mesmos equipamentos da rede. O meio físico é compartilhado com vários usuários, possibilitando que várias ligações ocorram ao mesmo instante, tendo assim, um maior aproveitamento da banda disponível.

A rede PSTN tem a vantagem de garantir a largura de banda necessária e fixa para a transmissão de voz com qualidade telefônica e um atraso mínimo. Porém, é pouco eficiente para tráfego de dados [22]. Sua largura de banda é limitada em 64 kbps e o tráfego de dados necessita de uma taxa de transferência maior. A desvantagem é o desperdício de recursos, devido ao fato da conexão ser dedicada, mesmo que não existam dados para serem transmitidos na linha, esta não pode ser utilizada para outra transmissão enquanto a conexão estiver estabelecida.

A rede IP foi construída para suportar aplicações de dados que não são de tempo real, como transferência de arquivos ou email. Esses são caracterizados pelo tráfego em rajadas, com picos ocasionais de demanda por maior largura de banda e não são sensíveis ao atraso. É uma rede de comutação de pacotes, onde os dados são empacotados e enviados para o destino através de vários roteadores. Não existe o estabelecimento de uma chamada. Os pacotes não precisam seguir o mesmo caminho. Se ocorrer um congestionamento em algum ponto da rede, os roteadores são programados para escolherem uma nova rota para os pacotes.

A rede de pacotes é, geralmente, representada por uma nuvem indicando a incerteza do caminho que os pacotes irão seguir. Para se transmitir o sinal de telefonia nesta rede é necessário que se tenha dispositivos conectados às suas bordas para converter o sinal para o formato utilizado. Na rede IP estes dispositivos são chamados de *gateways*. A Figura 4.1 apresenta um exemplo de uma ligação utilizando a tecnologia VoIP que envolve dois *gateways*.

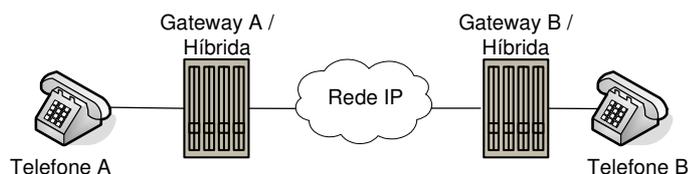


Figura 4.1 Representação de uma transmissão VoIP

Para uma conversação com qualidade, a voz precisa ser transmitida em tempo real. Se uma rede de comutação de pacotes for utilizada, os pacotes perdidos não podem ser retransmitidos. Para uma boa qualidade na transmissão, a fala não pode ser cortada e nem muito atrasada. Além disto, existem também outros fatores como o *jitter*, a perda de pacotes e o eco que devem ser controlados. Para este controle são utilizadas técnicas de qualidade de serviço.

4.2 Qualidade de Serviço

O termo QoS indica a capacidade de uma rede de fornecer um serviço com qualidade para um determinado tráfego de rede. A tecnologia de QoS inclui a utilização de prioridades com largura de banda dedicada, controle de *jitter*, latência e perda de pacotes. Além disto, é importante que a prioridade de um ou mais fluxos de pacotes não impeçam que outros sejam transmitidos.

O emprego de VoIP está relacionado com a qualidade de serviço da rede IP. É possível atingir uma qualidade em VoIP tão boa quanto em uma chamada telefônica na rede PSTN, utilizando-se as técnicas de qualidade de serviço que serão citadas a seguir [23].

4.2.1 Latência

A transmissão de pacotes de dados de voz pela rede IP não ocorre instantaneamente. Latência é o tempo necessário para os pacotes trafegarem do transmissor para o receptor, incluindo o tempo gasto na codificação, empacotamento no transmissor, o tempo gasto para a decodificação e tratamento do *jitter* no receptor. A latência é freqüentemente apresentada como um valor médio.

A conversação, geralmente, envolve uma parada ao redor de 200 ms para mudar de locutor. Se a latência da rede for maior que 150 ms, o fluxo da conversa fica comprometido, podendo as duas pessoas começar a falar ao mesmo tempo, uma interrompendo a outra [23]. O atraso ou a latência entre os falantes é um dos fatores mais crucial na determinação subjetiva da qualidade de uma chamada telefônica.

A recomendação G.114 do ITU-T considera o atraso de transmissão da rede para aplicações de voz. Define três faixas de aceitação para o atraso:

- 0-150 ms, aceitável para a maioria das aplicações.
- 150-400 ms, aceitável desde que o administrador tenha conhecimento do tempo de transmissão e seu impacto na qualidade da transmissão para aplicações do usuário.
- Acima de 400 ms, inaceitável para propósitos gerais, entretanto, em alguns casos excepcionais este limite pode ser excedido.

Essas recomendações são para conexões com um controle de eco adequado, que implica o uso de canceladores de eco.

A voz que trafega pela rede de pacotes é processada em várias etapas. O sinal analógico originado no telefone é digitalizado utilizando a modulação PCM pelo codificador de voz (vocoder). Estas amostras PCM são comprimidas por um algoritmo (codec) e formatadas em pacotes para transmissão através da WAN (wide

area network). No outro lado da rede de pacotes as mesmas funções são realizadas na ordem inversa.

Todas as etapas envolvidas nesse processo acrescentam uma parcela ao atraso total da transmissão. A seguir está detalhado algumas dessas fontes de atraso.

4.2.1.1 Atraso de codificação

É o tempo que o processador digital de sinais (DSP) gasta para comprimir um bloco de amostras PCM. Devido a existência de vários codificadores, esse atraso varia de acordo com o codificador de voz utilizado. Por exemplo, o codificador G.729A analisa um bloco de 10 ms de amostras PCM para posteriormente comprimi-lo.

4.2.1.2 Atraso de *bufferização*

Depois que a voz é empacotada, o quadro é armazenado em um *buffer* para transmissão na rede. Devido a alta prioridade do pacote de voz, este deve esperar somente se já estiver sendo transmitido um pacote de dados ou se estiver outro pacote de voz na fila. Este tempo de espera é chamado de atraso de *bufferização*.

4.2.1.3 Atraso de roteamento na rede IP

Ao transportar voz pela Internet, existe uma informação limitada, e basicamente nenhum controle, sobre o caminho que os pacotes poderiam seguir entre dois *gateways* quaisquer. Na rede IP pode existir atrasos longos, alta perda de pacotes, *jitter* variado e pacotes fora de ordem.

Em uma configuração Intranet, entretanto, é possível obter muita informação e controle sobre o caminho do pacote de voz. Neste caso, uma rede bem definida de roteadores e switches pode fornecer um serviço de ótima qualidade.

Pacotes transportados na rede IP são atrasados ao passar por cada roteador. O atraso no roteador vai depender de sua configuração, desempenho, capacidade e carga. Muitos fatores podem influenciar negativamente no atraso de roteamento, principalmente, um alto volume de pacotes grandes chegando ao mesmo tempo com o tráfego de voz. Entretanto, em uma rede gerenciável é possível priorizar portas de telefonia IP sobre portas de dados genéricos, diminuindo o atraso total da rede.

4.2.2 *Jitter*

Um grande problema enfrentado na transmissão VoIP é a característica aleatória do atraso da rede IP. Em uma conexão típica, os pacotes passam por vários roteadores até atingir o destino. Os pacotes não são transmitidos pelo mesmo caminho, podendo alguns pacotes levar mais tempo que outros. Cada roteador acrescenta um atraso variado, dependendo do número de pacotes que estão no *buffer*.

Além disto, parte da rede pode ficar congestionada. A flutuação do instante da chegada dos pacotes ao redor do valor médio da latência é chamado de *jitter*. Para amenizar esta variação, os pacotes são coletados e armazenados em um *buffer* por um tempo suficiente para permitir que o pacote mais lento chegue no destino em tempo hábil para ser tocado na seqüência correta. Isto causa um atraso adicional.

4.2.3 Codificadores

O uso de codificadores tem várias vantagens, tais como, compressão da taxa de bits, recuperação de pacotes perdidos e detectores de atividade de voz com gerador de ruídos de conforto. Existem diversos tipos, os mais utilizados comercialmente na telefonia IP estão citados abaixo.

G.711 Codifica um fluxo de fala em uma taxa de 64 kbps. Este codec possui uma qualidade denominada de telefônica, equivalente a voz na rede PSTN, e utiliza a mesma largura de banda de um canal de voz da rede de circuitos comutados.

G.723.1 Codifica a uma taxa de 5,3 ou 6,3 kbps, usa a codificação preditiva linear (LPC) e uma busca em dicionários que requer muito processamento computacional.

G.729A Este algoritmo trabalha na taxa de 8 kbps. É empregado por muitos anos como o codificador de fala no mercado de Frame Relay. Para telefonia IP, foi escolhido pelo IMTC (Internacional Multimedia Teleconferencing Consortium) no fórum VoIP, como codificador alternativo para voz sobre IP [24].

Os codificadores comprimem blocos de amostras PCM. Esses blocos variam em comprimento de acordo com o codificador utilizado. Por exemplo, o tamanho de um bloco básico do G.729A é de 10 ms, enquanto que para o G.723.1 é de 30 ms. A taxa de transmissão destes codificadores é menor que a do G.711, gastando menos largura de banda para transmitir a fala. Para a codificação, os *vocoders* precisam armazenar os dados em um *buffer* para executar a segmentação da fala. Um pequeno atraso é acrescentado para o *vocoder* armazenar e executar os cálculos matemáticos. Além disto, ocorre um atraso adicional para realizar o processamento de compressão de fala. Os cálculos são realizados pelo processador, no qual o *vocoder* está executando, por exemplo um processador digital de sinais (DSP).

Os *vocoders* também degradam o sinal dependendo do codificador utilizado. Uma medida subjetiva de qualidade é o MOS (Mean Opinion Score). A tabela 4.1 apresenta os valores MOS atribuído a cada codificador.

Vocoder	Taxa(kbps)	MOS
G.711 (PCM)	64	4,1
G.723 (MP-MLQ)	6,3	3,9
G.723.1 (ACELP)	5,3	3,65
G.729A (CS-ACELP)	8	3,7

Tabela 4.1 Valores MOS dos codificadores.

4.2.3.1 Medida da Qualidade do Sinal de Voz

Para medida da qualidade, podem ser usados métodos objetivos e subjetivos. O MOS (Mean Opinion Score) é uma medida subjetiva tradicionalmente empregada, definida na recomendação ITU-T P.800. A pontuação do MOS varia de 1 (ruim) a 5 (excelente). Um sistema com uma pontuação maior ou igual a 4 é considerado um sistema de qualidade telefônica. Como exemplo, o algoritmo de modulação PCM (Pulse Code Modulation) definido no padrão G.711 do ITU-T tem um MOS de 4,1. Além desta medida, existem métodos objetivos que tentam prever o MOS, como por exemplo PAMS, PSQM e PESQ [25].

PAMS (Perceptual Analysis / Measurement System) Mede o sinal de voz em termos do esforço e da qualidade do som escutado pelo ouvinte. O sinal original e o sinal degradado pelo sistema são comparados. A quantidade de diferentes tipos de erros encontrados na versão degradada é analisada e uma pontuação MOS é estimada.

PSQM (Perceptual Speech Quality Measurement) Foi desenvolvido para medir a qualidade objetiva do codec. Está especificado na recomendação P.861 do ITU-T. O PSQM compara um sinal original com uma versão degradada para medir distorções, ruído e fidelidade. Uma versão melhorada foi aprovada pelo ITU-T, PSQM+, que trata de grandes distorções transientes com maior eficácia. Nenhuma das versões reagem adequadamente com filtragem, atraso variável e curtas distorções localizadas.

PESQ (Perceptual Evaluation of Speech Quality) Especificado na recomendação P.862, o PESQ é um método de medida da qualidade da chamada, em que um sinal de voz conhecido é enviado através da conexão de teste e então comparado com o sinal original. Uma pontuação análoga à produzida pelo MOS é obtida. É utilizado para medir distorções no sinal de fala. Entretanto, o PESQ não mede os efeitos do eco, largura de faixa e outras imperfeições, e torna-se menos apurado para taxas de perda de pacotes maiores que 20%.

4.2.4 Perda de Pacotes

Como VoIP é um serviço de tempo real, pacotes perdidos não podem ser retransmitidos. Os pacotes de voz podem conter 20 ou 30 ms de informação de fala, a perda destes pacotes reduzem significativamente a qualidade da chamada.

Da mesma forma que pacotes são atrasados ao passar pelo roteador, existe a possibilidade deles serem perdidos ou descartados em períodos de congestionamento. Os roteadores são projetados para transmitir todos os pacotes que chegam na porta de saída correta, mas um roteador sobrecarregado pode perder uma seqüência inteira de pacotes IP. A perda de pacotes degrada o desempenho de uma chamada na telefonia IP. Embora a maioria dos *gateways* empregam algoritmos de codificação e técnicas para tratar essa possibilidade, uma perda de mais de 5% dos pacotes tem um efeito muito degradante na conversação. Através do planejamento e gerenciamento da rede, as perdas de pacotes podem ser minimizadas e freqüentemente eliminadas [24].

4.2.5 Técnicas de Qualidade de Serviço

A capacidade de transmissão de dados de uma rede é chamada de largura de banda ou largura de faixa. Utilizar uma largura de faixa inadequada causa tanto atraso, quanto perda de pacotes. Como o tráfego na rede IP é irregular, deve-se empregar técnicas para priorizar os pacotes de voz em caso de congestionamento. Algumas destas técnicas são CoS e IntServ.

4.2.5.1 Classe de Serviço (CoS)

É um modo de administrar o tráfego na rede agrupando tipos semelhantes de tráfego e tratando cada tipo como uma classe. Para cada classe, é atribuído um nível de prioridade diferente. Alguns tipos semelhantes de tráfego são e-mail, fluxo de voz, fluxo de vídeo, transferência de arquivos, etc. A classe de serviço não garante um nível de serviço, em termos de largura de banda e tempo de entrega. Utiliza a política de melhor esforço (best-effort). A tecnologia CoS é mais simples para ser administrada e escalonada quando a rede cresce em estrutura e em volume de tráfego. Duas tecnologias de CoS são o Tipo de Serviço (ToS) e o Serviço Diferenciado (DiffServ).

Tipo de Serviço O cabeçalho do pacote IP contém um byte chamado de tipo de serviço. Este é dividido em duas seções, um campo de cinco bits chamado de ToS e um campo precedente, os três bits mais significativos. Esta extensão do campo ToS, permite a camada 3 do cabeçalho IPv4 conter oito valores precedentes para indicar tipo de serviço. Esse byte possui valores que são utilizados pelos roteadores

e *switches* de nível 3 para especificar como a rede deve agir para transmitir cada pacote, levando em consideração o atraso, a confiabilidade, o custo e a taxa de transmissão.

Serviços Diferenciados A tecnologia anteriormente citada não fornece um gerenciamento de tráfego, simplesmente rotulando as prioridades nos pacotes. O campo ToS foi redefinido pelo grupo IETF (Internet Engineering Task Force), recebendo o nome de DiffServ. Seis bits do campo são utilizados para selecionar o comportamento por nó da rede, PHB (Per Hop Behavior), que define como os pacotes são enfileirados nos nós da rede. Descreve um nível particular de serviço em termos de largura de banda, teoria de fila, e decisões de perda ou descarte de pacotes.

4.2.5.2 Serviços de Internet (IntServ)

Outra tecnologia utilizada é a Internet Service (IntServ) que garante QoS, enquanto que os outros serviços utilizam a política de melhor esforço. Implementa um protocolo de sinalização de QoS para reserva de recursos, tal como largura de faixa, para um fluxo unidirecional ou caminho da rede. Para cada chamada, dois fluxos devem ser configurados.

4.2.6 Métodos utilizados para atingir a Qualidade de Serviço

4.2.6.1 Redução do atraso e do *jitter*

Para reduzir o efeito do *jitter*, os pacotes que chegam no receptor são armazenados em um *buffer*. Após capturar uma quantidade ótima de pacotes de dados, o decodificador inicia o processamento dos pacotes a uma taxa constante. Este tempo de espera inicial é chamado de *playout time*. Para manter a latência efetiva tão baixa quanto possível, a espera inicial deve ser mínima. Porém, se o *buffer* esvaziar, alguns pacotes serão perdidos. Geralmente, isto se deve ao congestionamento momentâneo da rede. Para não ocorrer perdas, o número de pacotes armazenados deve ser grande, o que leva a um aumento na latência. Existe um compromisso entre o *playout time* e a redução do atraso que deve ser otimizado.

Foram propostos vários algoritmos para calcular o melhor tempo de espera para armazenar os pacotes. Uma proposta é medir a variação no número de pacotes no *buffer* sobre um período de tempo, e ajustar o *playout time* adaptativamente. Isto funciona melhor em redes ATM, que fornece uma pequena variação do *jitter*.

Para redes IP, existem vários algoritmos. Um algoritmo simples conta o número de pacotes que são descartados por chegarem atrasados e cria uma razão destes

pacotes com os que foram processados com sucesso. Esta razão é utilizada para ajuste do *playout time*.

4.2.6.2 Priorização de pacotes VoIP

Para priorizar o tráfego de voz, as camadas 2 e 3 da pilha de protocolo de rede são utilizadas para indicar a prioridade através das técnicas de QoS. CoS deve ser implementada para Ethernet. O campo ToS do cabeçalho IP deve ser utilizado para indicar prioridade dos pacotes. O DiffServ e o RSVP devem ser implementados no roteador para utilizar os bits ToS e reservar largura de banda. Além do uso de outras técnicas, também é importante que o aplicativo VoIP nos terminais utilize o campo ToS para rotular as prioridades dos pacotes IP enviados na rede.

4.2.6.3 Fragmentação de pacotes em links de baixa velocidade

Um grande pacote de dados pode ocupar um link de baixa velocidade por um tempo significativo, de forma a atrasar alguns pacotes VoIP aumentando o *jitter*. Em um link de 256 kbps, um pacote de 1 Kbyte ocupará o link por 31,25 ms, equivalente ao atraso de um pacote VoIP de 30 ms. Portanto é imprescindível o uso de técnicas de fragmentação e intercalamento de pacotes entre roteadores de links de baixa velocidade. Sendo que um pacote maior é fragmentado em pacotes menores e intercalados por pacotes de voz de maior prioridade.

4.2.6.4 Redução do *overhead* do cabeçalho IP

A largura de banda necessária para transmissão de voz pode ser reduzida se o *overhead* de cabeçalho for reduzido. O RFC2508 do IETF descreve o protocolo CRTP para compressão de cabeçalho de tempo real. Normalmente, o cabeçalho IP possui 44 bytes. Utilizando a compressão, o cabeçalho fica reduzido para 2 ou 4 bytes, diminuindo significativamente a largura de faixa por pacote.

4.2.6.5 Recuperação de pacotes perdidos

Pacotes perdidos podem degradar significativamente o sinal de fala, dependendo da rede e do codificador utilizado. Pelo fato da rede IP atual não garantir a qualidade de serviço, usualmente ocorrerá a perda de pacotes em picos de carga e congestionamento. Para transmissão de dados que não são de tempo real, os pacotes perdidos podem ser retransmitidos, entretanto a perda de pacotes de voz não pode ser tratada da mesma forma.

Alguns esquemas são utilizados por software VoIP para contornar este problema. Um deles é a interpolação de pacotes perdidos. Esse esquema é um método simples que preenche o tempo entre quadros de fala não contínuos. Funciona bem quando

a perda de pacotes é pouco freqüente, mas no caso de perdas em rajadas, um outro método deve ser empregado. Por exemplo, o envio dos quadros de fala do pacote anterior junto com o atual. Este método tem a desvantagem de aumentar a largura de banda e o atraso. Um método híbrido combinando os anteriores pode ser utilizado, gastando uma largura de banda menor para enviar informações redundantes.

Como exemplo, os algoritmos PLC (Packet Loss Concealment) e o PLR (Packet Loss Recovery) são executados nos terminais para recuperar os pacotes perdidos sem precisar retransmiti-los.

4.2.6.6 Redução da utilização do canal

A compressão da fala reduz a largura de faixa utilizada pelos pacotes de voz, é possível obter uma alta taxa de compressão de bits, porém a qualidade do sinal de fala é reduzida. Como exemplo, pode ser utilizado o codificador G.729A que possui uma boa taxa de compressão, 8 kbps, sendo oito vezes menor que a taxa do G.711, 64 kbps.

Pode-se utilizar técnicas que economizam a largura de banda nos períodos de silêncio. Junto com os codificadores, são implementados blocos funcionais com esta finalidade. Por exemplo, o detector de atividade de voz (VAD) e o gerador de ruído de conforto (CNG).

Em uma comunicação VoIP, quando o detector VAD percebe a ausência de fala, pacotes de controle são enviados ao invés de sinais contendo silêncio. Os pacotes de controle possuem menos bits que os pacotes de sinais. No receptor, os pacotes de controle são identificados e o gerador CNG toca um ruído de fundo, ao invés de interromper o sinal de fala. O usuário não percebe mudança no sinal durante a conversa. Esta técnica é bem eficiente, reduzindo em até 50% a utilização do canal em uma transmissão full-duplex, pois na maior parte do tempo apenas um usuário fala enquanto o outro escuta.

4.2.6.7 Gerenciamento da rede IP

O desempenho da rede IP na Internet é altamente variável e freqüentemente inaceitável. Por outro lado, a qualidade de VoIP sobre uma rede gerenciável, Intranet, ou VPN (Virtual Private Network) pode ser muito boa e relativamente quantificável.

No passado, Intranets foram implementadas usando o aluguel de linhas de rede caras por grandes corporações e agências do governo. Mas com o aumento de provedores ISPs (Internet Service Providers) e provedores de comunicação de dados é possível obter serviços de rede privada virtual com garantia de qualidade de serviços.

A habilidade para realocar um *gateway* em um ponto de presença ISP tem se tornado uma forma barata de minimizar a latência na rede IP, eliminando um par de roteadores em cada ponta da rede. Devido a proliferação de provedores de serviço de rede IP e produtos oferecidos, é agora possível criar um backbone da telefonia IP com facilidade.

O custo já não é proibitivo e o atraso nos pacotes, a perda e o *jitter* podem ser gerenciados para obter uma qualidade excelente. Com tal gerenciamento de rede, o atraso pode ser mantido abaixo de 100 ms, a perda de pacote menor que 3% e o *jitter* abaixo de 60 ms, resultando em uma alta qualidade de conversação [24].

4.2.6.8 Utilização de Canceladores de Eco

O aparelho telefônico é projetado para gerar o sinal *sidetone* (ocorre devido a híbrida interna do aparelho) que é o retorno da voz do microfone para o alto-falante. Este ocorre quase que instantaneamente com a fala, dando a impressão que o circuito está funcionando. Porém, o eco é um fator negativo que causa um desconforto e deve ser eliminado [23].

O eco elétrico, que será o assunto desse trabalho sendo citado apenas como eco, surge devido à reflexão do sinal na híbrida na rede de comutação de circuitos. A híbrida localiza-se na central telefônica e é responsável pela conexão do enlace local a dois fios no enlace a longa distância a quatro fios.

Outra fonte de eco está no próprio aparelho telefônico. Devido ao acoplamento acústico ruim entre o microfone e o alto-falante e a reflexão do som na sala do interlocutor, o som do alto-falante retorna pelo microfone ocorrendo o chamado eco acústico.

O eco não é gerado dentro da rede IP. Em uma ligação local na rede de comutação de circuitos, a latência é tão baixa que o eco é mascarado pelo *sidetone*. Isto ocorre para um atraso menor que 50 ms. Para ligações a longa distância, principalmente envolvendo enlaces de satélites geoestacionários que tem um atraso de transmissão em torno de 250 ms, valores maiores são obtidos. O aumento no atraso intensifica o problema do eco. A rede PSTN possui canceladores de eco para eliminar qualquer eco com atraso acima de 50 ms [22]. Se o atraso for menor, o eco não será cancelado pelo cancelador empregado na rede PSTN, de acordo com a recomendação G.165 do ITU [26], que é o padrão original para canceladores de eco. A recomendação G.168 [2], mais recente, padroniza o cancelador de eco em uma forma mais estrita que a anterior. Não sendo cancelado na rede PSTN, o eco se propaga pela rede IP que introduz um atraso muito maior. Para impedir que isto ocorra, deve ser implementado um cancelador de eco no *gateway* IP.

O cancelamento do eco executado no *gateway* pode ser *near-end*, referindo-se ao lado mais próximo da fonte de eco, ou *far-end*, referindo-se ao lado mais afastado da fonte. O cancelador utilizado comercialmente cancela o eco na terminação mais

próxima, cujo caminho de eco é menor.

Uma característica que distingue os canceladores é o tempo de atraso do caminho de eco, *echo path delay*, medido em milissegundos, que o cancelador pode processar. O comprimento padrão é na faixa de 32 ms a 128 ms. O caminho de eco é medido do cancelador de eco para a híbrida mais o retorno da híbrida para o cancelador.

É importante controlar a intensidade do eco. A recomendação G.168 do ITU-T cita que a perda no caminho de eco com o cancelador deve ser de no mínimo de 55 dB.

A fala dupla (*double-talk*) ocorre quando os dois lados transmitem voz ao mesmo tempo. Este fenômeno é muito comum na conversação, dificultando a identificação do eco pelo cancelador. Por este motivo, é implementado um detector de fala dupla que funciona em conjunto com o cancelador. Quando detectado este período, o algoritmo adaptativo do cancelador de eco é congelado, evitando-se assim que o sinal de fala do lado remoto atrapalhe a convergência do algoritmo.

Durante a fala dupla, o sinal de fala do lado remoto é adicionado ao sinal de eco. Desta forma, o sinal resultante da subtração do sinal de retorno da híbrida pelo sinal de eco sintetizado é equivalente ao sinal de fala do lado remoto. Se este valor for utilizado como sinal de erro, os coeficientes do algoritmo adaptativo irão divergir dos coeficientes do modelo.

A maioria dos canceladores de eco atual não estima o atraso inicial do caminho de eco. Se forem configurados para um tempo de 128 ms e o eco ocorrer antes, este será cancelado, porém é gasto um processamento desnecessário. Seria mais eficiente se ocorresse uma estimativa do atraso do eco e o cancelamento processado apenas no instante mais provável do eco aparecer [22].

Os canceladores VoIP existentes comercialmente cancelam apenas o eco da terminação mais próxima. Para um cancelamento efetivo, é necessário a existência de canceladores em todos os *gateways* VoIP da rede. Não há como garantir que todos os equipamentos utilizados na rede IP tenham um cancelador de eco. Se não tiverem, não adianta um alto investimento em um cancelador na rede local, se o eco vem remotamente na rede IP. A idéia proposta nesta dissertação é um cancelador que cancela inclusive o eco proveniente da terminação remota. Faz-se necessário um método para estimar o tempo de atraso total da rede IP e do uso do *playout time*. Este método será detalhado no Capítulo 6.

O sucesso de VoIP depende muito da aceitação dos usuários. Para eles migrarem da rede PSTN para a rede IP, é necessária uma boa relação custo benefício. A qualidade do sinal transmitido deve ser próxima da telefonia pública. Para este fim, os métodos para atingir QoS devem ser empregados em VoIP.

Capítulo 5

Modelos de caminho de eco

A pesquisa científica está a todo tempo buscando modelos matemáticos para simular situações reais. Os simuladores estão ficando cada vez mais sofisticados, representando os fenômenos físicos de uma forma muito fiel. A grande vantagem do emprego de simuladores está na possibilidade de análise de diversas configurações sistemáticas de forma exaustiva; ao passo que se realizada por meio de modelos reais, os custos dos modelos podem ser muito elevados e o tempo de suas implementações demorado. Neste capítulo são apresentados os modelos utilizados para simular o caminho de eco em uma rede telefônica. Os modelos apresentados foram utilizados nas simulações do cancelador de eco para avaliar o seu desempenho. Os resultados serão apresentados no Capítulo 6.

5.1 Técnicas de Medida Utilizadas para Obter o Modelo do Caminho de Eco

A recomendação G.168 do ITU-T [2], no seu apêndice II, cita as técnicas de medida utilizadas para obter a resposta impulsiva do caminho de eco na rede telefônica. Foram medidos vários caminhos de eco na rede norte americana e os resultados analisados. Essa seção apresenta um resumo deste apêndice com alguns comentários.

A Figura 5.1 ilustra o esquema de geração e gravação dos sinais utilizados. Este foi conectado à central telefônica através de uma interface T1. Todo o circuito envolvido na transmissão até a central está na porção do circuito a quatro fios, com canais de transmissão (TX) e recepção (RX) independentes.

Para realizar as medidas, o equipamento deve discar o número do telefone de destino no início do processo. Após estabelecida a conexão, o sinal de teste é enviado para o destino através da interface T1, passando pela central local e propagando pela rede PSTN. Parte do sinal transmitido é refletido na híbrida no *far-end*. Este sinal

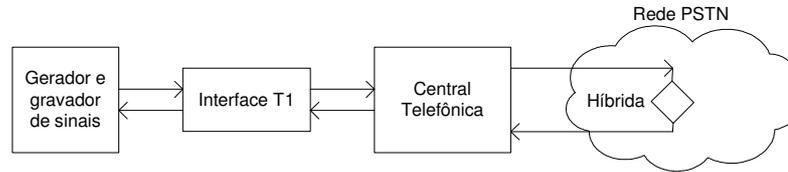


Figura 5.1 Método de medição do caminho de eco

de retorno junto com o sinal enviado é gravado para medida do caminho de eco.

O sinal de teste enviado é constituído de três partes. O primeiro segmento é constituído por um tom de 2100 Hz com a duração de 1,35 ms, utilizado para desabilitar os supressores e canceladores de eco que podem estar ativos no enlace durante as medições. Este tom de desabilitação deve ser de acordo com as recomendações G.164 [27] e G.165 [26] do ITU-T. O segundo segmento consiste de uma pausa de 80 ms para obter a característica do ruído ambiente. Este tempo é menor que a pausa utilizada para ativar o cancelador de eco, $250 \text{ ms} \pm 150 \text{ ms}$ [27] [26]. O terceiro segmento é um ruído branco gaussiano com duração de 5 s, utilizado para identificar a resposta impulsiva do caminho de eco.

A solução para encontrar a resposta impulsiva é a mesma utilizada para identificação de sistemas. Pode-se utilizar o método *Least-Square* (LS) ou o método iterativo do algoritmo *Normalized Least Mean Square* (NLMS).

Duas características importantes do caminho de eco são o tempo de dispersão¹ e sua resposta em frequência. A resposta impulsiva do caminho de eco tem uma duração finita que é referida como tempo de dispersão. Quanto maior a dispersão, maior deve ser o número de coeficientes do filtro adaptativo utilizado no cancelamento do eco. Por exemplo, para uma dispersão de 10 ms, o número de coeficientes do filtro para uma taxa de 8000 Hz é calculado como:

$$f = 8000 \text{ Hz} \Rightarrow T = \frac{1}{8000} = 0,000125 \text{ s}$$

$$N = \frac{0,010}{0,000125} = 80$$

onde N é o número de coeficientes do filtro.

¹Pelo gráfico da resposta impulsiva é possível medir o tempo de dispersão do caminho de eco. Mede-se o intervalo de tempo entre o início da oscilação da resposta impulsiva e o instante em que o estado de estabilidade é atingido.

5.2 Características do Caminho de Eco Medidas na América do Norte

A recomendação G.168 do ITU-T [2] apresenta uma análise de medições de vários caminhos de eco realizadas na América do Norte. Foram realizadas 101 chamadas de longa distância no período de junho de 1998 a abril de 1999. As chamadas foram originadas de Montreal com os seguintes destinos: Arizona, British, Columbia, California, Lousiana, Manitoba, Massachusetts, Michigan, Minnesota, Missouri, Nevada, New York, North Carolina, Ontario, Quebec, Saskatchena, Texas e Wisconsin.

Foi verificado o seguinte resultado:

- O tempo de dispersão dos caminhos de eco medidos não ultrapassou 12 ms. A média foi de 6,02 ms com desvio padrão de 2,26 ms.
- A resposta em frequência dos caminhos de eco se apresentaram relativamente plana na banda de passagem. A maioria apresentou um pico em 250 Hz. Na Figura 5.2, reproduzida da recomendação G.168 do ITU-T [2], foi calculado a média dos espectros das respostas impulsivas medidas e representado pela linha contínua. As linhas tracejadas representam a região de desvio padrão.
- Foram encontrados alguns casos de dupla reflexão, porém não ocorreram casos com maior número de reflexões.

A resposta em frequência praticamente plana na banda de passagem indica que o caminho de eco atenua praticamente pelo mesmo valor todas as frequências do sinal de voz. Outro resultado importante é que as dispersões não ultrapassaram 12 ms. Logo, o filtro de apenas 96 coeficientes será suficiente para envolver toda a resposta impulsiva.

5.3 Modelos Utilizados na Simulação

Os seguintes modelos de caminho de eco, citados no anexo D da recomendação G.168 do ITU-T [2], foram utilizados na simulação do cancelador de eco. O caminho de eco é simulado por um filtro digital com a seguinte resposta impulsiva:

$$g(k) = 10^{ERL_i/20} K_i m_i(k - \delta) \quad (5.1)$$

Onde:

ERL_i = perda de retorno do eco do i -ésimo modelo de caminho de eco

K_i = fator de escala do i -ésimo modelo de caminho de eco

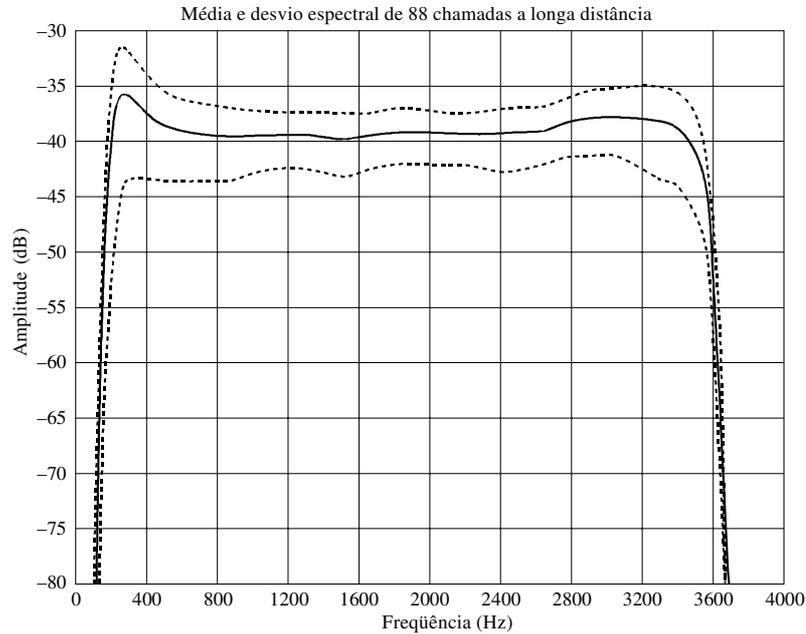


Figura 5.2 Desvio e média dos espectros de 88 chamadas - reproduzido de [2]

m_i = resposta impulsiva do i -ésimo modelo de caminho de eco

δ = atraso do modelo de caminho de eco em número de amostras

Para levar em consideração os vários atrasos, diferentes perdas de caminho de eco, diversas características de dispersão e duração do eco, a função $g(k)$ é escolhida como uma versão atrasada e atenuada das respostas impulsivas do modelo de caminho de eco, m_i .

Os valores de ERL_i e K_i estão apresentados na tabela 5.1. Os valores de m_i estão apresentados nas tabelas² 5.2, 5.3, 5.4, 5.5, 5.6, 5.7 e 5.8.

Os quatro primeiros modelos de caminho de eco, $m_i(k)$ para $i = 1, 2, 3$ e 4 , foram gerados por um simulador de uma rede híbrida. A resposta impulsivas destes estão apresentados nas Figuras 5.3, 5.5, 5.7 e 5.9. Os outros três modelos, $m_i(k)$ para $i = 5, 6$ e 7 , representam modelos digitais reais do caminho de eco medido de redes telefônicas na América do Norte. As respostas impulsivas destes estão apresentados nas Figuras 5.11, 5.13 e 5.15. As suas respostas em frequência estão apresentados nas Figuras 5.4, 5.6, 5.8, 5.10, 5.12, 5.14 e 5.16.

Observa-se pelos gráficos que a maioria dos modelos, exceto o modelo $m_7(k)$, tem uma única reflexão. Os modelos simulados $m_1(k)$, $m_2(k)$, $m_3(k)$ e $m_4(k)$ tem uma

²Os valores das tabelas são lidos da esquerda para a direita.

dispersão de aproximadamente 4 ms, 8 ms, 10 ms e 12 ms, para uma frequência de amostragem de 8 kHz. O modelo $m_5(k)$ tem uma dispersão de aproximadamente 6 ms. Este modelo é o mais comum. O modelo $m_6(k)$ tem uma dispersão maior, de aproximadamente 10 ms. Esta dispersão ocorre devido ao pico no espectro na frequência em torno de 250 Hz. O modelo $m_7(k)$ tem duas reflexões, as duas respostas impulsivas se sobrepõem, com dispersão total de aproximadamente 6 ms. Os três últimos modelos apresentam um atraso inicial de 2 ms, também chamado de atraso puro.

i	$ERL_i(dB)$	K_i
1	7,6	$1,39 \times 10^{-5}$
2	12,2	$1,35 \times 10^{-5}$
3	9,0	$1,52 \times 10^{-5}$
4	8,6	$1,77 \times 10^{-5}$
5	15,5	$9,33 \times 10^{-6}$
6	21,3	$1,51 \times 10^{-5}$
7	19,0	$1,31 \times 10^{-5}$

Tabela 5.1 Valores das constantes dos modelos de caminho de eco.

-436	-829	-2797	-4208	-17968	-11215	46150	34480	-10427	9049	-1309
-6320	390	-8191	-1751	-6051	-3796	-4055	-3948	-2557	-3372	-1808
-2259	-1300	-1098	-618	-340	-61	323	419	745	716	946
880	1014	976	1033	1091	1053	1042	794	831	899	716
390	313	304	304	73	-119	-109	-176	-359	-407	-512
-580	-704	-618	-685	-791	-772	-820	-839	-724	*	*

Tabela 5.2 Resposta impulsiva do modelo de caminho de eco 1, $m_1(k)$

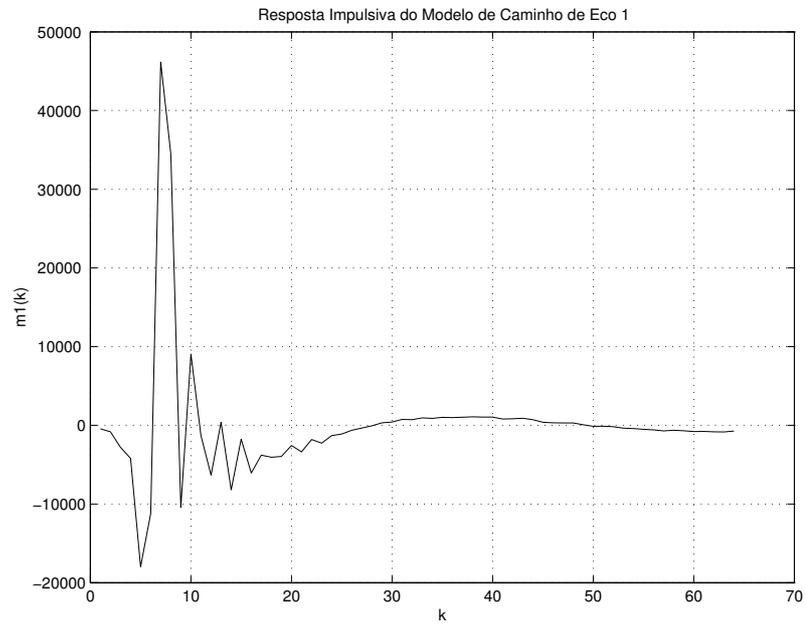


Figura 5.3 Resposta impulsiva do modelo de caminho de eco 1

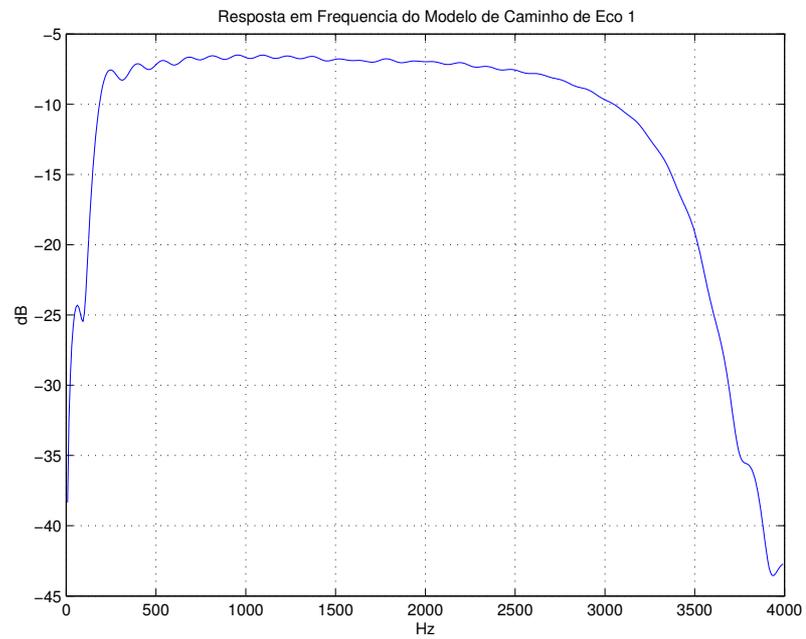


Figura 5.4 Resposta em frequência do modelo de caminho de eco 1

-381	658	1730	-51	-3511	-1418	7660	8861	-8106	-21370	-5307
23064	24020	1020	-12374	-16296	-19524	-7480	13509	17115	13952	13952
97	-9326	-9046	-15208	-9853	-3858	-1979	6029	5616	7214	6820
3935	3919	921	1316	-693	-759	-1517	-2176	-2028	-2654	-1814
-2077	-1468	-1221	-842	-463	-298	-68	64	493	723	789
954	756	839	872	1020	789	822	558	658	476	377
377	262	97	-68	-183	-232	-331	-347	-430	-314	-430
-463	-463	-414	-381	-479	-479	-512	-479	-397	-430	-397
-298	-265	-249	-216	-249	-265	-166	-232	*	*	*

Tabela 5.3 Resposta impulsiva do modelo de caminho de eco 2, $m_2(k)$

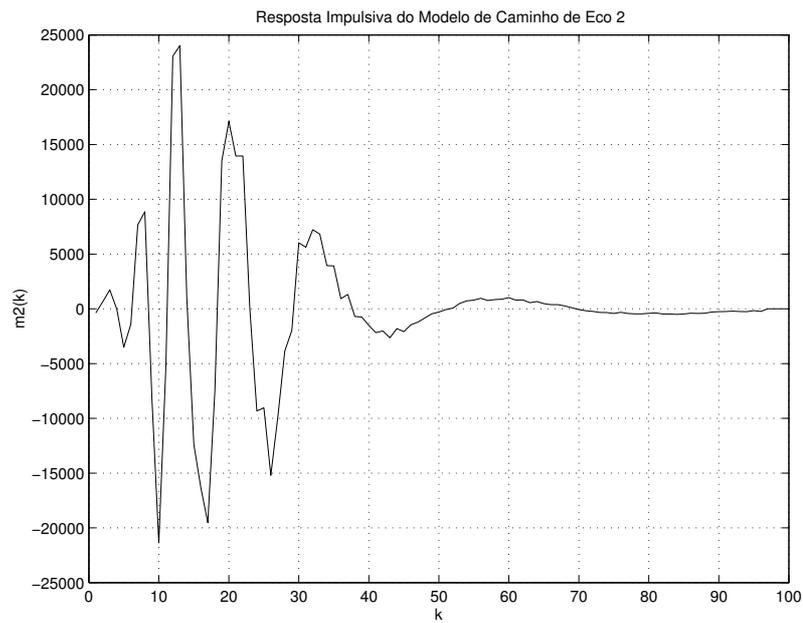


Figura 5.5 Resposta impulsiva do modelo de caminho de eco 2

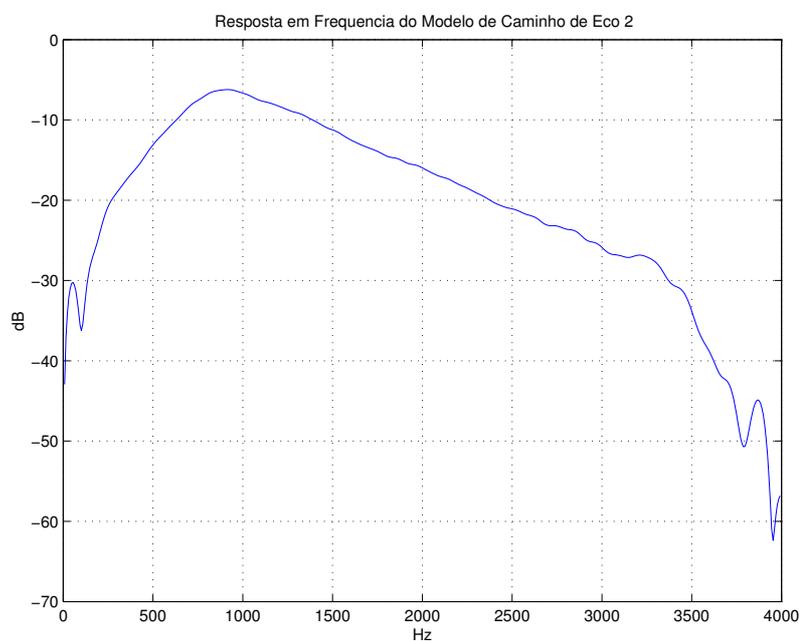


Figura 5.6 Resposta em frequência do modelo de caminho de eco 2

-448	-436	2230	2448	-4178	-7050	5846	18581	2322	-26261	-16249
21637	25649	-2267	-10311	-4693	-12690	-7428	14164	13467	4438	8627
456	-11879	-6352	-5104	-7496	3271	6566	4277	11131	7562	1475
3728	-3525	-7301	-3101	-9269	-6146	-2553	-6272	811	124	788
5147	2172	5387	4598	3535	4004	2311	2150	1017	330	-139
-573	-1100	-1157	-1180	-1455	-1123	-1386	-1123	-1066	-1020	-1100
-1008	-1077	-1088	-917	-917	-963	-814	-871	-734	-642	-562
-356	-379	-345	-230	-233	-333	-356	-390	-310	-265	-368
-310	-310	-390	-482	-459	-482	-551	-573	*	*	*

Tabela 5.4 Resposta impulsiva do modelo de caminho de eco 3, $m_3(k)$

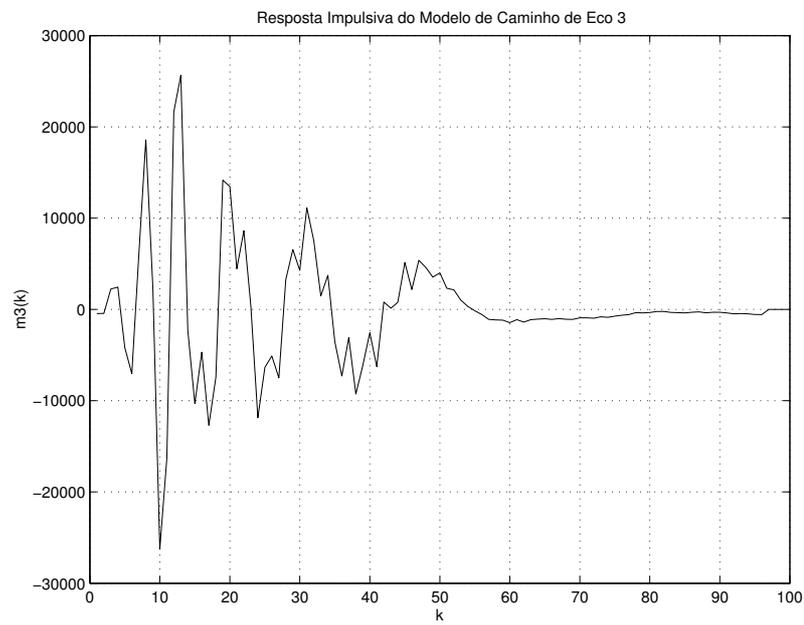


Figura 5.7 Resposta impulsiva do modelo de caminho de eco 3

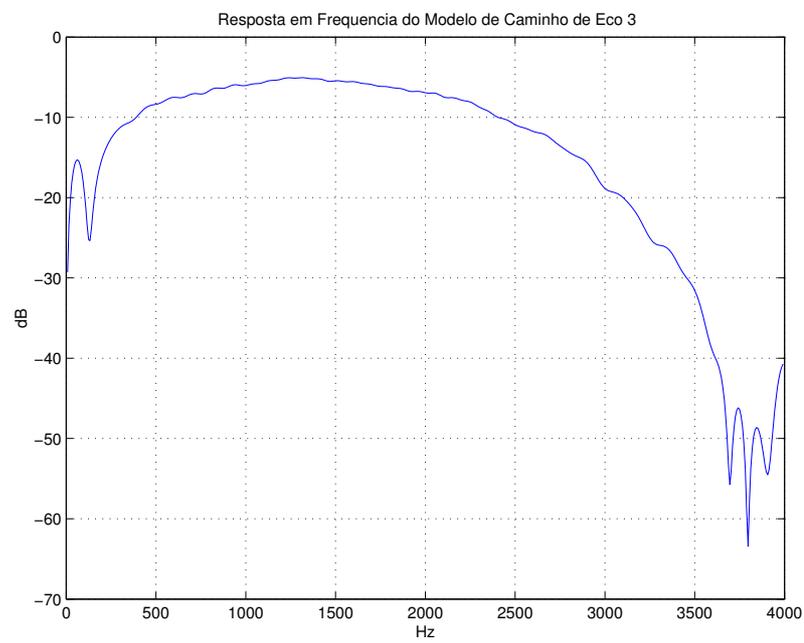


Figura 5.8 Resposta em frequência do modelo de caminho de eco 3

160	312	-241	-415	897	908	-1326	-1499	2405	3347	-3624
-7733	4041	14484	-1477	-21739	-4470	25356	11458	-19696	-11800	5766
789	6633	14624	-6975	-17156	-187	149	1515	14907	4345	-7128
-2757	-10185	-7083	6850	3944	6969	8694	-4068	-3852	-5793	-9371
453	1060	3965	9463	2393	2784	-892	-7366	-3376	-5847	-2399
3011	1537	6623	4205	1602	1592	-4752	-3646	-5207	-5577	-501
-1174	4041	5647	4628	7252	2123	2654	-881	-4113	-3244	-7289
-3830	-4600	-2508	431	-144	4184	2372	4617	3576	2382	2839
-404	539	-1803	-1401	-1705	-2269	-783	-1608	-220	-306	257
615	225	561	8	344	127	-57	182	41	203	-111
95	-79	30	84	-13	-68	-241	-68	-24	19	-57
-24	30	-68	84	-155	-68	19	*	*	*	*

Tabela 5.5 Resposta impulsiva do modelo de caminho de eco 4, $m_4(k)$

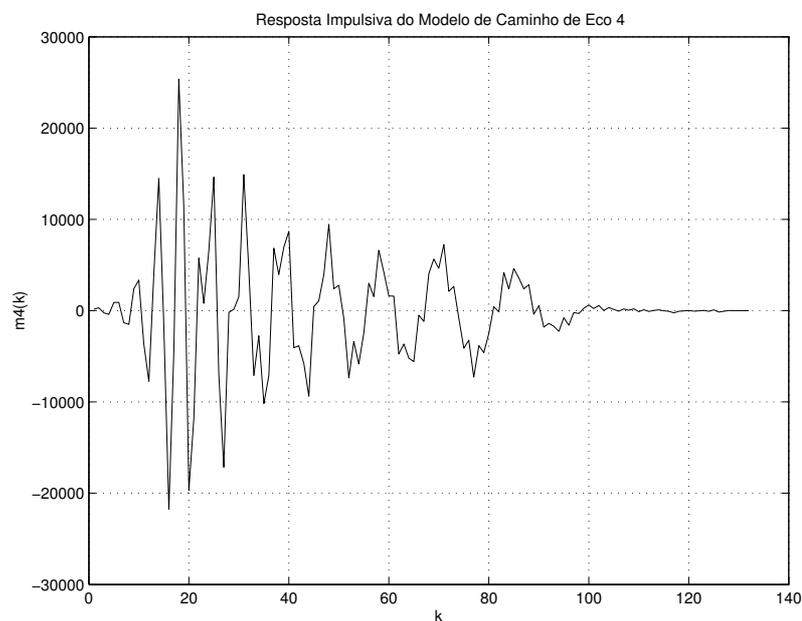


Figura 5.9 Resposta impulsiva do modelo de caminho de eco 4

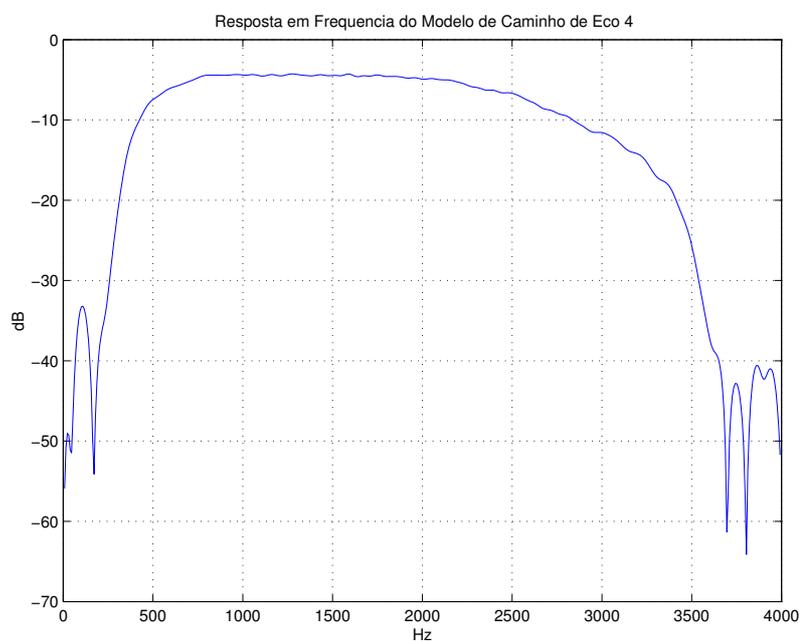


Figura 5.10 Resposta em frequência do modelo de caminho de eco 4

293	268	475	460	517	704	581	879	573	896	604
787	561	538	440	97	265	-385	20	-938	-523	-1438
-1134	-1887	-1727	-1698	-4266	-22548	-43424	2743	25897	7380	21499
11983	10400	11667	3889	7241	925	2018	-821	-2068	-2236	-4283
-3406	-5022	-4039	-4842	-4104	-4089	-3582	-2978	-2734	-1805	-1608
-645	-495	279	471	947	1186	1438	1669	1640	1901	1687
1803	1543	1566	1342	1163	963	733	665	323	221	-14
-107	-279	-379	-468	-513	-473	-588	-612	-652	-616	-566
-515	-485	-404	-344	-290	-202	-180	-123	*	*	*

Tabela 5.6 Resposta impulsiva do modelo de caminho de eco 5, $m5(k)$

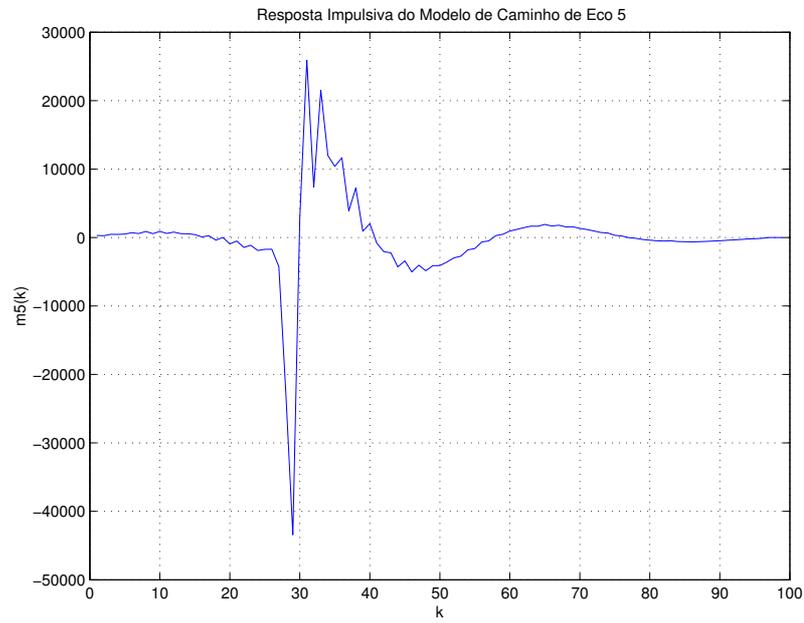


Figura 5.11 Resposta impulsiva do modelo de caminho de eco 5

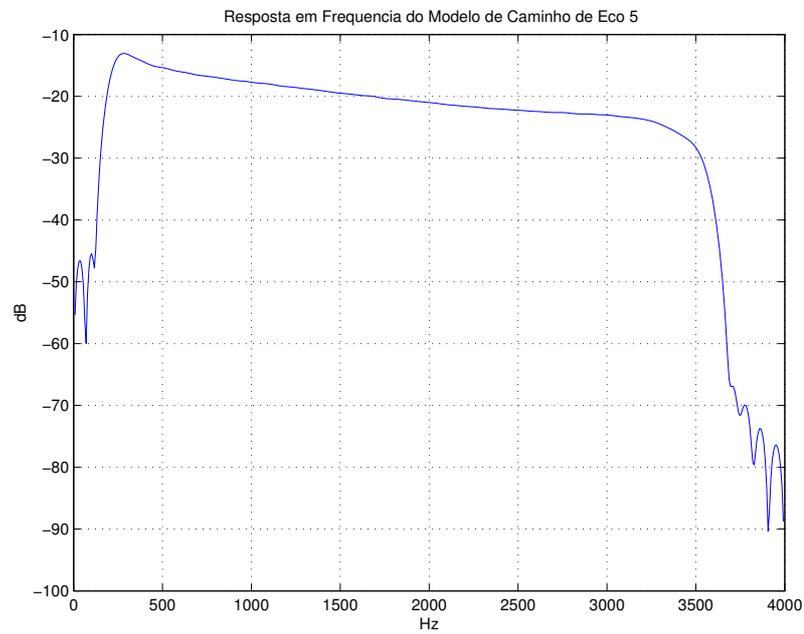


Figura 5.12 Resposta em frequência do modelo de caminho de eco 5

29	109	-83	198	-294	-135	-415	-202	-444	-337	-313
-450	-105	-503	145	-490	267	-231	340	77	343	783
158	1341	195	1798	344	1845	629	1604	1182	940	5163
19522	8421	-50953	-9043	18046	-13553	13336	-3471	-107	1788	-7409
2469	-7994	490	-3860	-837	490	-636	3682	1141	5019	2635
5025	3946	4414	4026	3005	3380	1616	2007	158	388	-1198
-1117	-2134	-2547	-2589	-3310	-2778	-3427	-2779	-3116	-2502	-2399
-1956	-1539	-1239	-570	-377	251	331	964	1177	1449	1564
1724	1871	1767	1802	1630	1632	1379	1271	1063	856	711
482	289	54	-137	-321	-490	-638	-764	-836	-800	-859
-838	-837	-834	-740	-673	-581	-493	-436	-327	-201	*

Tabela 5.7 Resposta impulsiva do modelo de caminho de eco 6, $m6(k)$

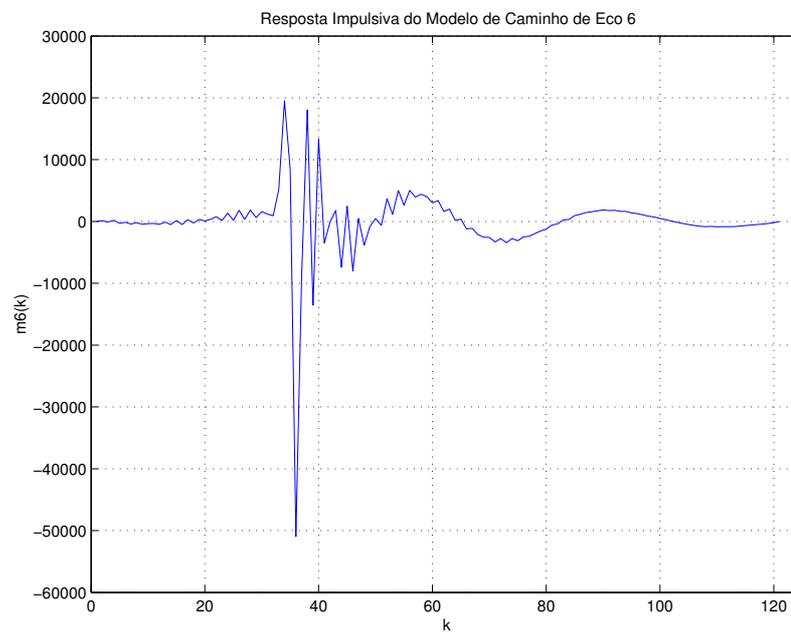


Figura 5.13 Resposta impulsiva do modelo de caminho de eco 6

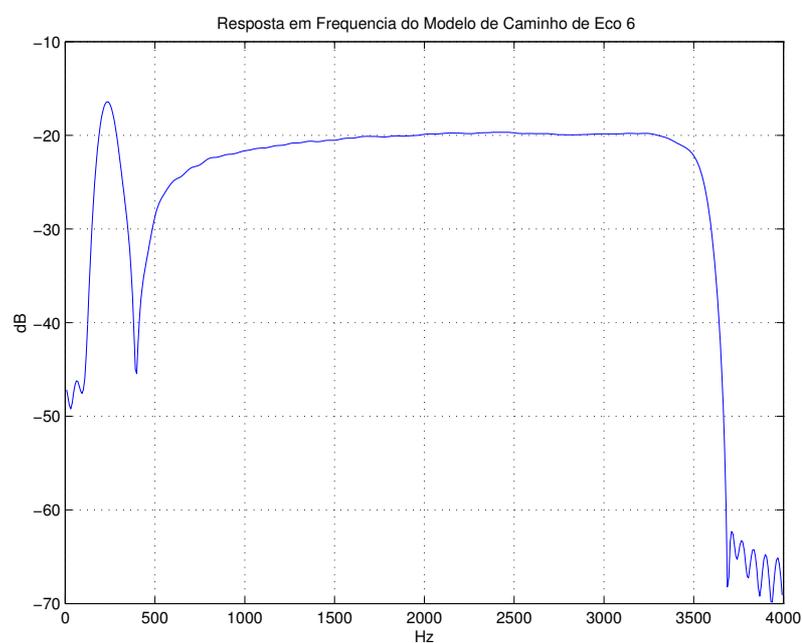


Figura 5.14 Resposta em frequência do modelo de caminho de eco 6

258	-111	337	-319	347	-434	192	-450	-108	-343	-596
-177	-1187	-52	-1781	-147	-1959	-326	-1601	-1389	-13620	-720
33818	-10683	-6742	12489	-9862	8950	-1574	758	3526	-3118	2421
-8966	-4901	11385	18072	-14410	-7473	19836	-16854	-3115	9483	-17799
7399	-4342	-7415	7929	-10726	6239	-2526	-1317	5345	-4565	6868
-2195	3425	1969	-109	3963	-1275	3087	-892	1239	2	-427
596	-1184	551	-1244	141	-743	-415	-372	-769	-183	-785
-270	-659	-377	-523	-325	-245	-255	-60	35	218	149
340	233	365	303	251	230	209	179	*	*	*

Tabela 5.8 Resposta impulsiva do modelo de caminho de eco 7, m7(k)

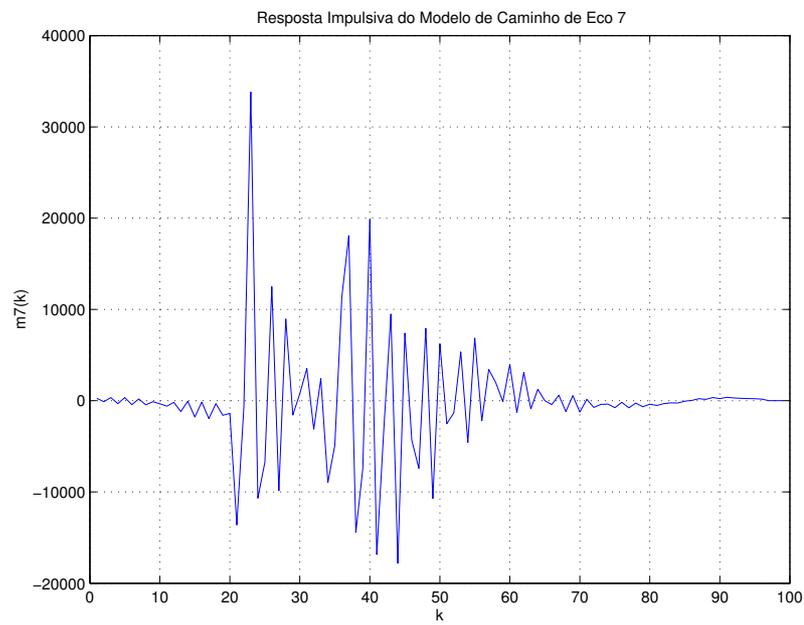


Figura 5.15 Resposta impulsiva do modelo de caminho de eco 7

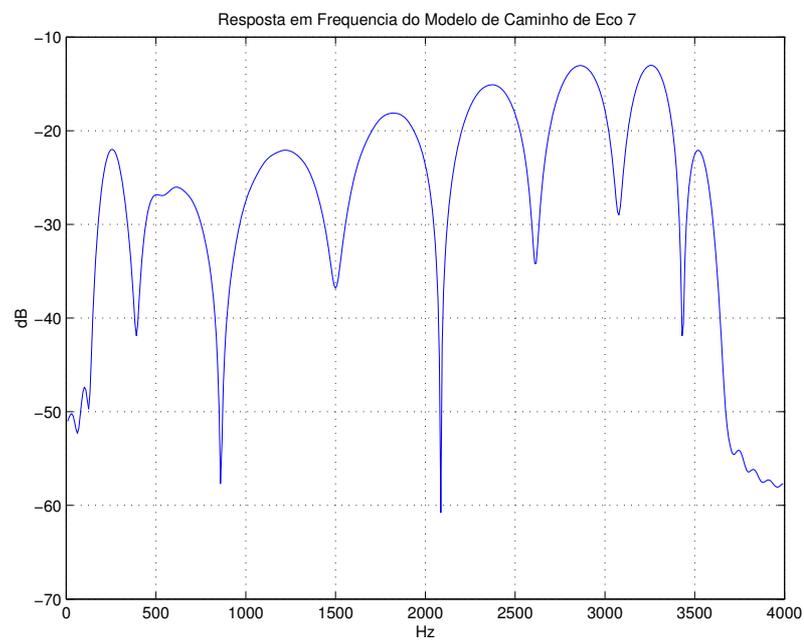


Figura 5.16 Resposta em frequência do modelo de caminho de eco 7

Capítulo 6

Simulações do Cancelador de Eco

Neste capítulo, serão apresentados os resultados da simulação do cancelador de eco, utilizando os modelos apresentados no Capítulo 5. Os princípios do funcionamento deste cancelador convencional, empregado para ligações longa distância, é o mesmo utilizado nos *gateways* para eliminar o eco provindo da rede de telefonia pública.

Além destes resultados, será apresentado um método de utilização deste cancelador para atenuar o eco provindo da rede IP. Para que nenhum dos usuários em uma comunicação escute o eco, é necessário que se tenha um cancelador de eco em cada *gateway* IP, evitando que o eco se propague na rede. Se por acaso um dos *gateways* não fornecer este bloco funcional, o usuário no lado oposto a este *gateway* poderá perceber um eco de sua própria voz. Assim sendo, da forma que o eco é atualmente tratado, são necessários dois canceladores de eco trabalhando em conjunto. Pode-se gastar recursos investindo em uma solução ineficiente, totalmente dependente de um outro equipamento que não pertence à mesma rede que está utilizando o cancelador. Este capítulo apresentará resultados de simulação de um cancelador de eco que pode ser empregado em um *gateway* para atenuar o eco provindo do *far-end*, no outro lado da rede IP. Esta solução em conjunto com o método utilizado comercialmente, que cancela o *near-end*, forma um cancelador de eco completo, sem a necessidade de um outro cancelador localizado em um *gateway* remoto.

Foi utilizado o aplicativo Matlab versão 6.1 para realizar toda a simulação que será apresentada neste capítulo.

6.1 Simulações do Cancelador de Eco Convencional

Em todas as simulações realizadas foi utilizado o mesmo cancelador de eco. Este é composto por um filtro adaptativo cuja estrutura utilizada é a transversal FIR e o algoritmo adaptativo utilizado é o NLMS para atualização dos parâmetros do filtro.

A estrutura FIR é muito utilizada devido a questões de estabilidade e pouca complexidade, as propriedades de convergência do algoritmo são bem conhecidas. A desvantagem desta estrutura é a sua grande dependência da duração da resposta impulsiva do caminho de eco. Quanto maior a duração do eco, maior deve ser o número de coeficientes do filtro para cancelá-lo e menor a velocidade de convergência. Se a duração do eco variar em torno de 10 a 60 ms, com uma taxa de 8 KHz, o número de coeficientes necessários será de 80 a 480.

O algoritmo NLMS, apresentado na seção 2.5, é o mais utilizado devido a sua facilidade de implementação e boa estabilidade. A desvantagem é a sua dependência à auto-correlação do sinal de referência, o tempo de convergência aumenta para sinais muito correlacionados como o sinal de voz.

O valor da constante de passo do algoritmo é muito importante na determinação da velocidade de convergência, estabilidade e erro residual após a convergência. Um passo largo de atualização leva à uma rápida convergência, porém o erro residual torna-se elevado e as chances de ocorrer instabilidade aumentam [28]. O valor da constante de passo depende da potência do sinal de entrada. Para contornar este problema, o termo de atualização é normalizado pela potência do sinal de entrada. Esta é a diferença entre o algoritmo LMS e o algoritmo NLMS.

Os modelos da resposta impulsiva do caminho de eco foram utilizados para gerar o sinal de eco, como se ocorresse na híbrida. Foi utilizado um sinal aleatório como sinal de entrada. Este sinal foi filtrado por um filtro FIR cujos coeficientes foram calculados de acordo com a expressão (5.1). O sinal de saída corresponde ao sinal retornado da híbrida, ou seja, o próprio eco.

O princípio do cancelador de eco é estimar a característica do caminho de eco, utilizando um filtro adaptativo, e posteriormente utilizar esta estimativa para gerar uma réplica do eco. Esta réplica é, então, subtraída do sinal de retorno da híbrida. O filtro adaptativo é utilizado devido ao fato do caminho de eco ser desconhecido e variante no tempo [28].

Seguindo esse princípio, foram realizadas várias simulações com cada um dos modelos citados no Capítulo 5. Em cada simulação, o mesmo sinal de entrada aleatório citado anteriormente foi utilizado como sinal de entrada no algoritmo adaptativo, e como sinal desejado foi utilizado o sinal de eco gerado. O filtro adaptativo é composto por 128 coeficientes, equivalendo a 16 ms de atraso do caminho de eco, conforme demonstrado na expressão (6.1), também chamado de

cauda do eco ou memória do filtro.

$$Atraso = \frac{1}{8000} \cdot 128 = 0,016s, \quad (6.1)$$

considerando que a frequência de amostragem é de 8 kHz.

A recomendação G.165 [26] cita os requerimentos básicos para os canceladores de eco usados na telefonia. Nesta padronização, está especificado que o tempo de convergência deve ser menor que 500 ms e que este estado é atingido quando o nível de potência do sinal de retorno fica 27 dB abaixo do sinal de referência. O tempo de convergência foi utilizado para analisar o desempenho do algoritmo para os diversos modelos de eco.

O desempenho do cancelador de eco é, usualmente, dado em termos da quantidade de atenuação do sinal de eco introduzido pelo cancelador de eco, ERLE (Echo Return Loss Enhancement). Não é considerada a atenuação causada por nenhum processamento não linear que possa existir após o cancelador. Este valor é obtido da seguinte forma:

$$ERLE = 10 \cdot \log_{10} \left(\frac{E[e^2(n)]}{E[(e(n) - \hat{e}(n))^2]} \right) \quad (dB) \quad (6.2)$$

onde:

$e(n)$ representa o sinal de eco, ou retorno da híbrida;

$\hat{e}(n)$ representa a estimativa do sinal de eco.

As recomendações G.165 [26] e G.168 [2] citam o uso de um processamento não linear após o cancelador de eco, embora não exija que uma aplicação o utilize. Este funciona eliminando o eco residual. Opera em conjunto com um detector de dupla fala; quando ocorre a fala nas duas direções, o detector desabilita o processamento não linear e congela os coeficientes do filtro adaptativo. Esta é uma forma de assegurar que o algoritmo adaptativo não fique instável e que a fala transmitida junto com o sinal de eco não seja atenuada. O detector baseia-se no nível de potência do sinal de retorno e no nível de potência do sinal enviado para decidir quando ocorre a dupla fala. O processador não linear funciona bloqueando sinais de baixo nível de potência e deixando passar sinais de alto nível de potência, reduzindo o eco a um nível aceitável.

As simulações realizadas e demonstradas nessa seção não utilizam o processador não linear e nem um detector de dupla fala. Assume-se que ocorre apenas a fala em uma direção. Desta forma, o ideal seria que o ERLE tendesse para o infinito rapidamente. Entretanto, existem fatores que limitam este resultado:

- Os filtros adaptativos possuem uma precisão finita dos coeficientes e dos registradores utilizados no cálculo.
- O número de coeficientes é finito impedindo que todo o caminho de eco seja modelado.
- O valor da constante de passo para atualização dos parâmetros que determina a velocidade de convergência, estabilidade e erro residual.
- Fatores externos como não linearidades no caminho de eco introduzidas pelos diversos componentes nos circuitos de transmissão e pela híbrida.

A Figura 6.1 mostra a convergência do cancelador de eco para o modelo de caminho de eco 1, utilizando como parâmetro o ERLE. Foram utilizados 128 coeficientes no filtro adaptativo. Considerando que a convergência é atingida quando o ERLE fica acima de 27 dB, como citado anteriormente, verifica-se que após 680 interações o algoritmo converge, representando um tempo de convergência de 85 ms, para a frequência de amostragem de 8 kHz. Através da Figura 6.2, verifica-se que os coeficientes estimados na simulação foram idênticos aos utilizados no modelo, por isto ocorreu a sobreposição total da curva azul pela vermelha. O elevado valor do ERLE apresentado na Figura 6.1, sem atingir o nível estacionário do erro residual (Figura 6.3), ocorreu devido a alta precisão utilizada na simulação e o pequeno passo de atualização dos coeficiente aplicado ao algoritmo adaptativo. Como a simulação foi realizada no Matlab, ambiente em que as variáveis do programa possuem uma precisão de 64 bits em ponto flutuante, a convergência dos coeficientes continua encontrando os valores fracionários dos coeficientes do modelo.

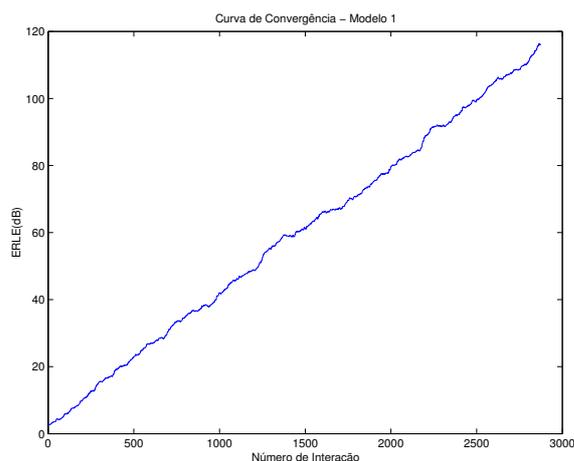


Figura 6.1 Gráfico de convergência para o modelo de caminho de eco 1

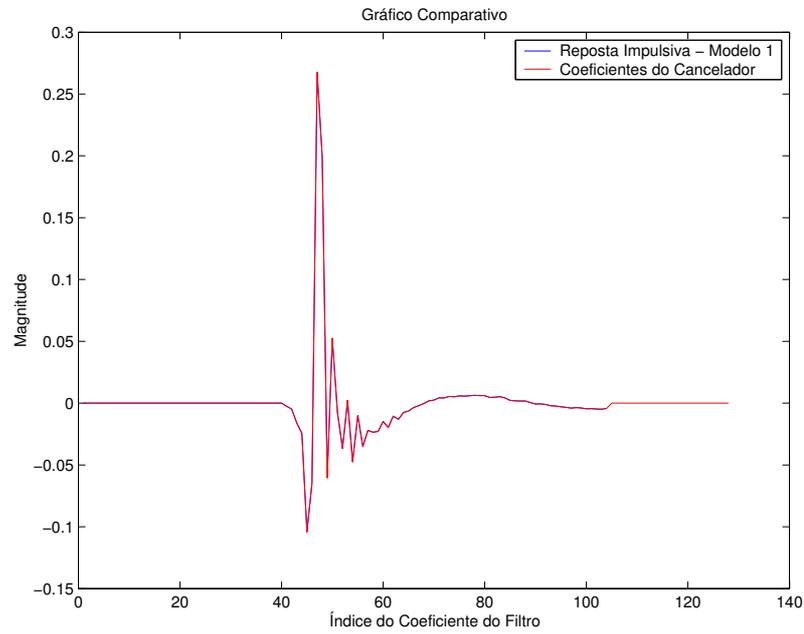


Figura 6.2 Comparação entre o modelo de eco estimado e o modelo 1

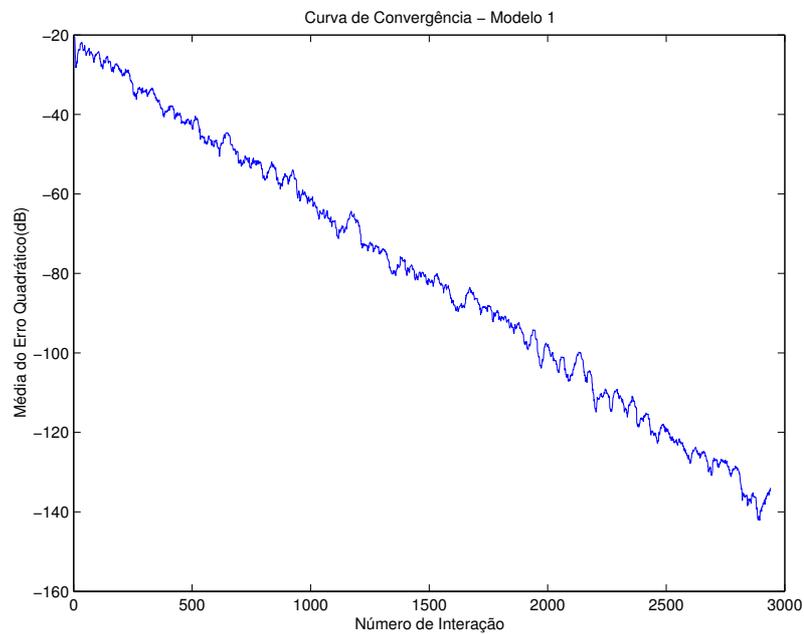


Figura 6.3 Gráfico da média dos erros quadráticos para o modelo de caminho de eco 1

Para o modelo do caminho de eco 2, foi realizada a mesma simulação. A Figura 6.4 mostra a convergência do cancelador de eco para o modelo de caminho de eco 2, utilizando como parâmetro o ERLE. Utilizando-se a mesma análise anterior, após 680 interações o algoritmo converge, representando um tempo de convergência de 85 ms. Pode-se notar que o algoritmo se estabiliza em torno de 35 dB após 1000 interações. Na Figura 6.5, observa-se um erro residual em torno de -60 dB. Na Figura 6.6 pode-se verificar que os coeficientes estimados na simulação estão bem próximos do ideal. O número de coeficientes do filtro adaptativo é um pouco menor que o modelo, porém foi suficiente para modelar os principais componentes oscilatórios da resposta impulsiva levando a um pequeno erro residual. Foram utilizados 128 coeficientes no filtro adaptativo da simulação, enquanto o modelo utiliza 140 coeficientes. Observa-se também um atraso inicial de 40 amostras, representando 5 ms a 8 kHz.

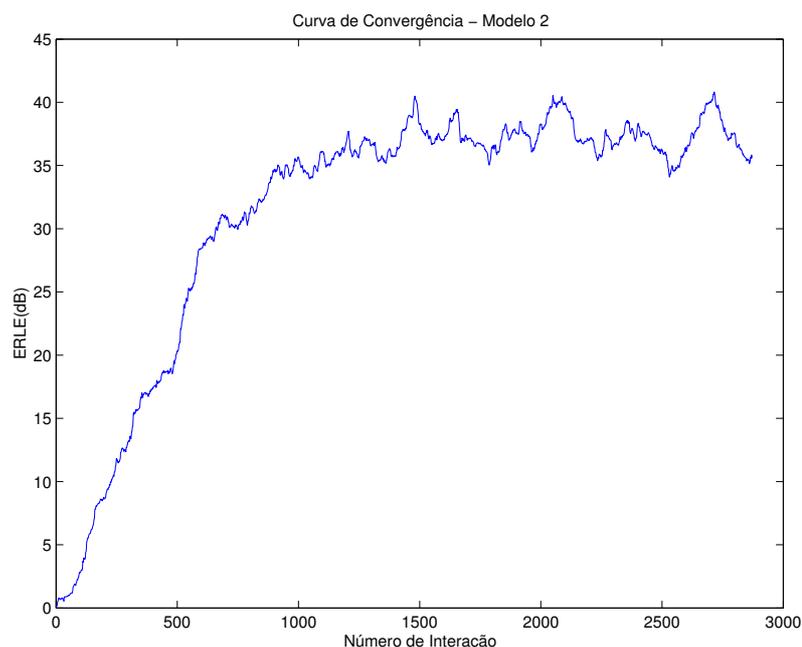


Figura 6.4 Gráfico de convergência para o modelo de caminho de eco 2

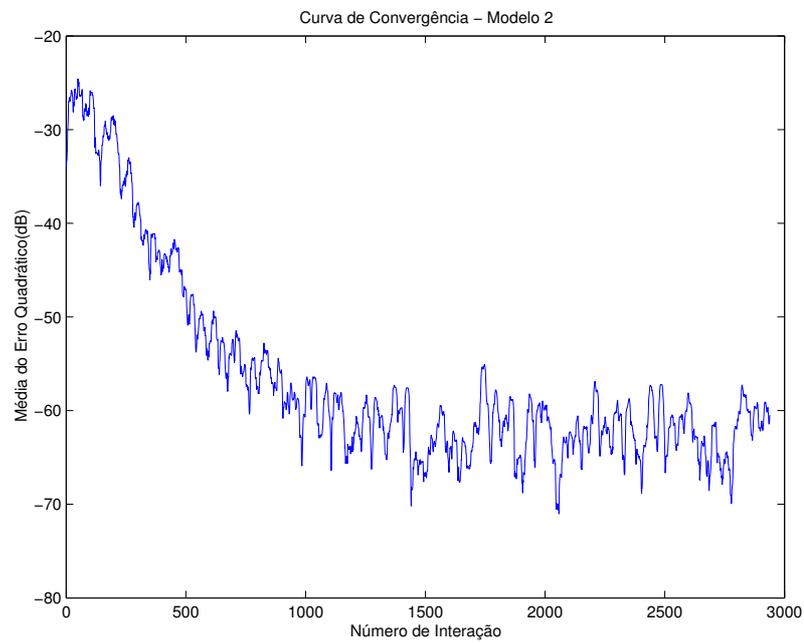


Figura 6.5 Gráfico da média dos erros quadráticos para o modelo de caminho de eco 2

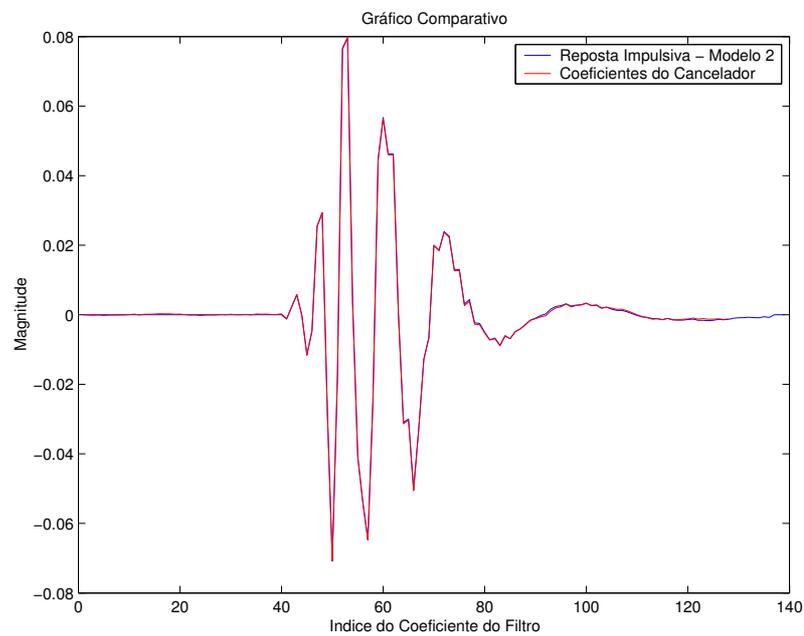


Figura 6.6 Comparação entre o modelo de eco estimado e o modelo 2

A Figura 6.7 mostra a convergência do cancelador de eco para o modelo de caminho de eco 3, utilizando como parâmetro o ERLE. O algoritmo converge após 755 interações, representando um tempo de convergência de 94 ms. Pode-se notar que o algoritmo se estabiliza em torno de 30 dB após 1000 interações. Na Figura 6.8, observa-se um erro residual em torno de -50 dB. Através da Figura 6.9, pode-se verificar que os coeficientes estimados na simulação estão bem próximos do ideal. O número de coeficientes do filtro adaptativo é um pouco menor que o modelo, porém foi suficiente para modelar os principais componentes oscilatórios da resposta impulsiva levando a um pequeno erro residual. Foram utilizados 128 coeficientes no filtro adaptativo da simulação, enquanto o modelo utiliza 140 coeficientes. Observa-se também um atraso inicial de 40 amostras, representando 5 ms a 8 kHz.

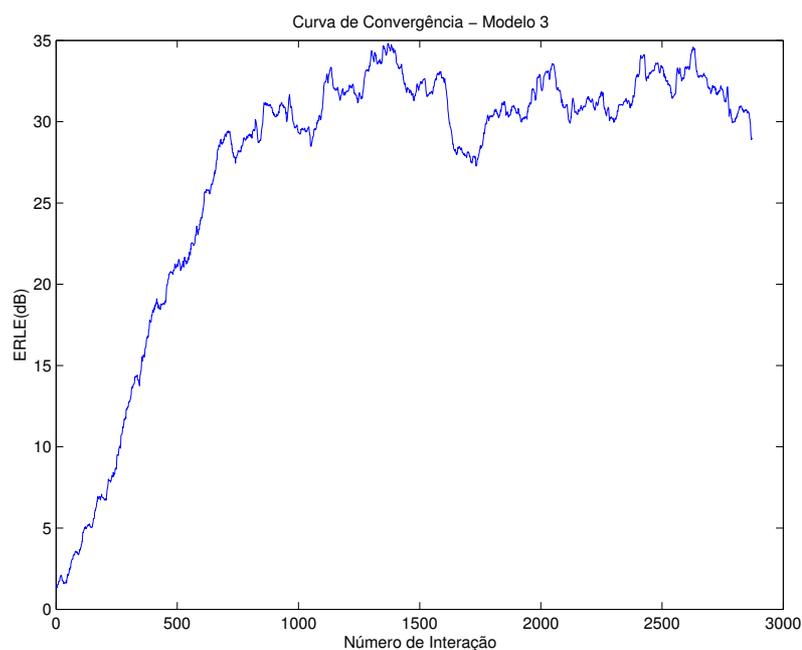


Figura 6.7 Gráfico de convergência para o modelo de caminho de eco 3

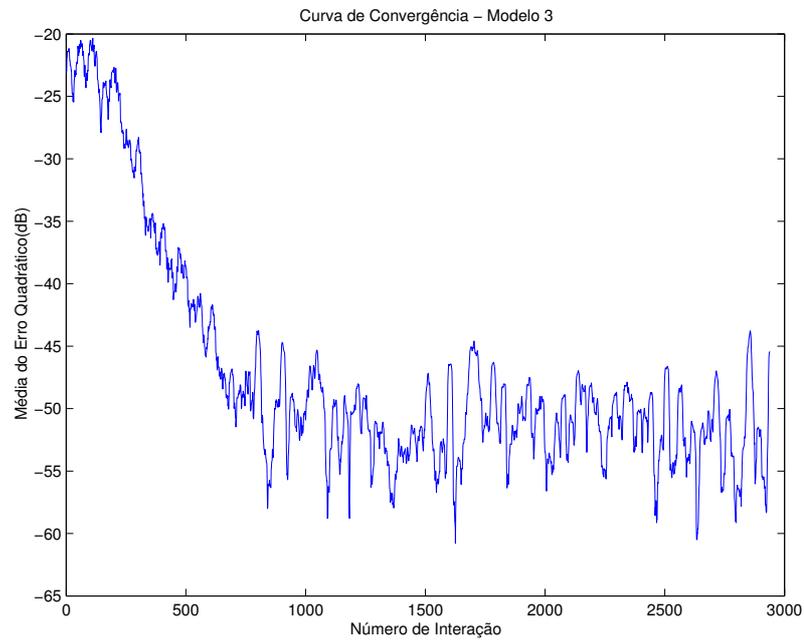


Figura 6.8 Gráfico da média dos erros quadráticos para o modelo de caminho de eco 3

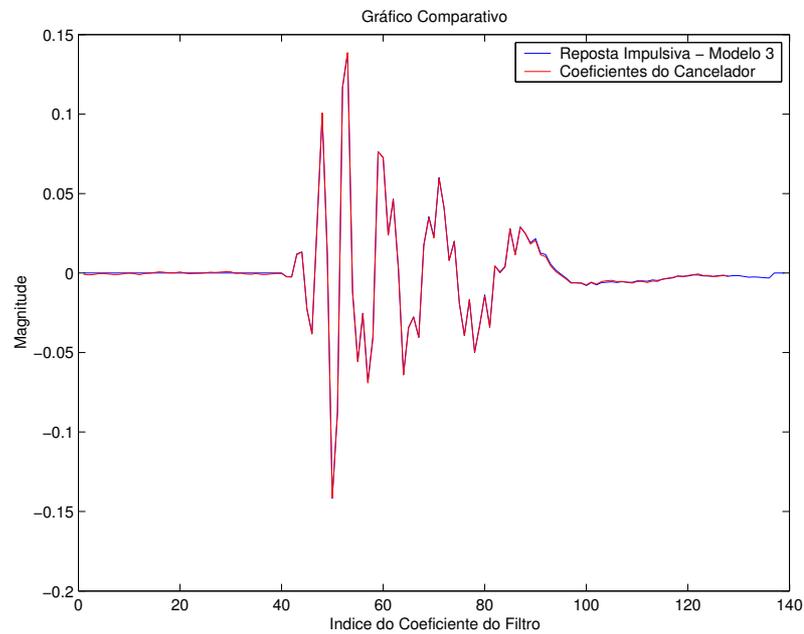


Figura 6.9 Comparação entre o modelo de eco estimado e o modelo 3

A Figura 6.10 mostra a convergência do cancelador de eco para o modelo de caminho de eco 4, utilizando como parâmetro o ERLE. Utilizando-se a mesma métrica anterior, o algoritmo converge, porém não atingi o limiar esperado. O valor máximo do ERLE atingido é menor que 27 dB. Pode-se notar que o algoritmo estabiliza-se em torno de 20 dB após 1000 interações. Na Figura 6.11, observa-se um erro residual em torno de -40 dB. Na Figura 6.12, pode-se verificar que os coeficientes estimados na simulação tentam atingir o ideal. Neste caso, o número de coeficientes do filtro adaptativo é menor que o modelo do caminho de eco 4, não sendo suficiente para modelar toda a resposta impulsiva, deixando de estimar uma parte significativa, levando a um grande erro residual. Foram utilizados 128 coeficientes no filtro adaptativo da simulação, enquanto o modelo utiliza 173 coeficientes. Observa-se também um atraso inicial de 40 amostras, representando 5 ms a 8 kHz.

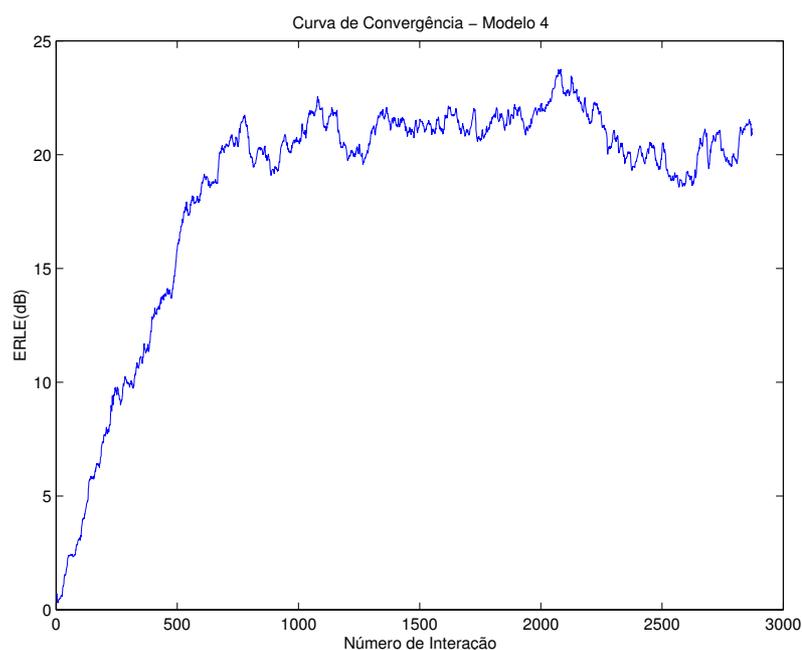


Figura 6.10 Gráfico de convergência para o modelo de caminho de eco 4

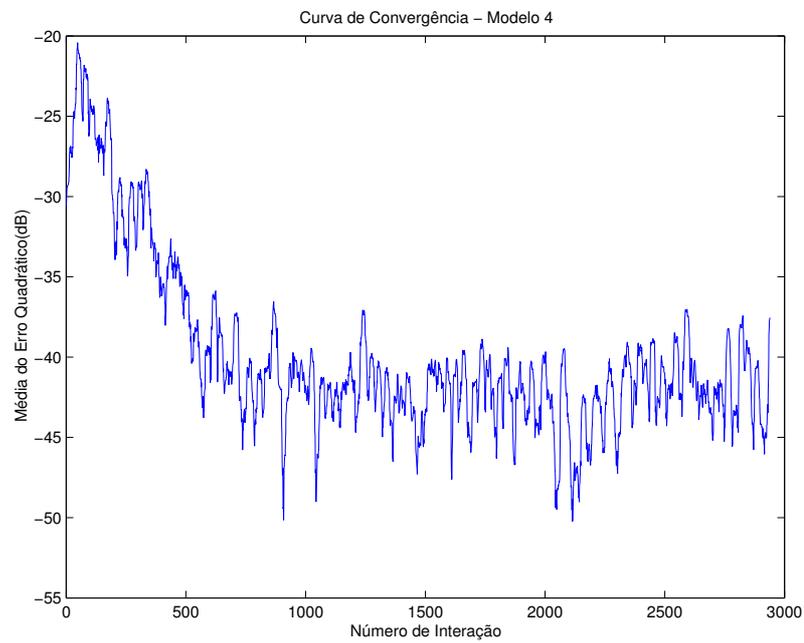


Figura 6.11 Média dos erros quadráticos para o modelo de caminho de eco 4

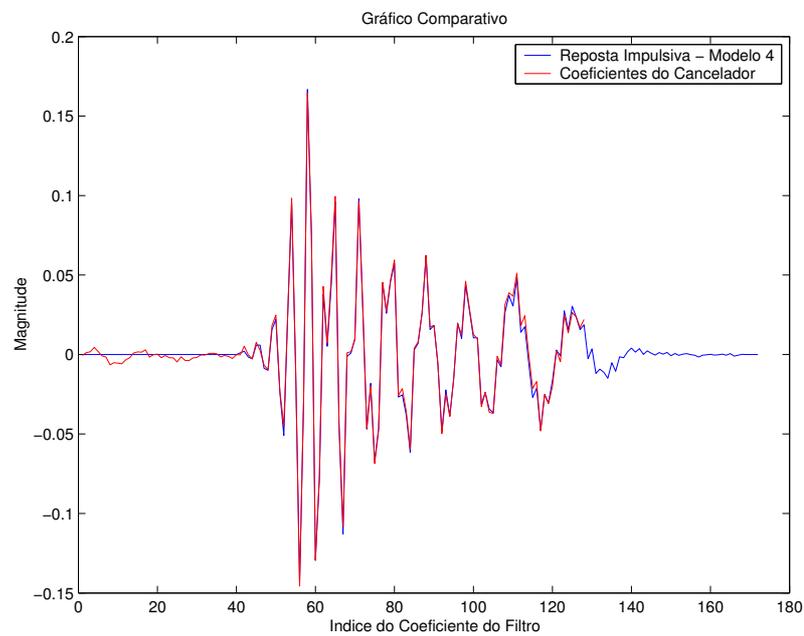
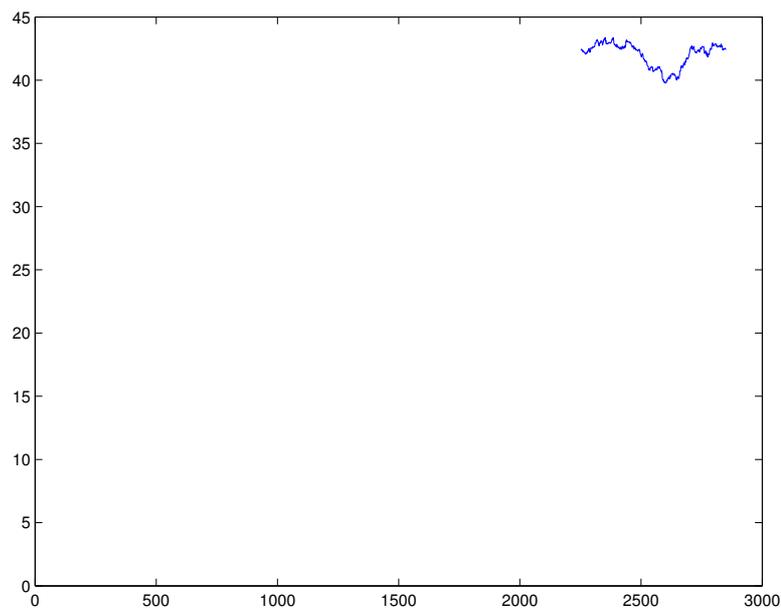


Figura 6.12 Comparação entre o modelo de eco estimado e o modelo 4

Como o cancelador de eco não atingiu o limiar esperado para o modelo 4, foi realizado uma nova simulação utilizando um número maior de coeficientes para o filtro adaptativo. Foram utilizados 150 coeficientes. A Figura 6.13 mostra a convergência do cancelador de eco para o modelo de caminho de eco 4, utilizando como parâmetro o ERLE. O algoritmo converge após 805 interações, correspondendo a um tempo de convergência de 100 ms, e estabiliza em torno de 40 dB após 1200 interações. Na Figura 6.14, observa-se um erro residual em torno de -60 dB. Na Figura 6.15, pode-se verificar que os coeficientes estimados na simulação atingem valor ideal. Com o aumento do número de coeficientes do filtro adaptativo, o cancelador de eco obteve o desempenho desejado.



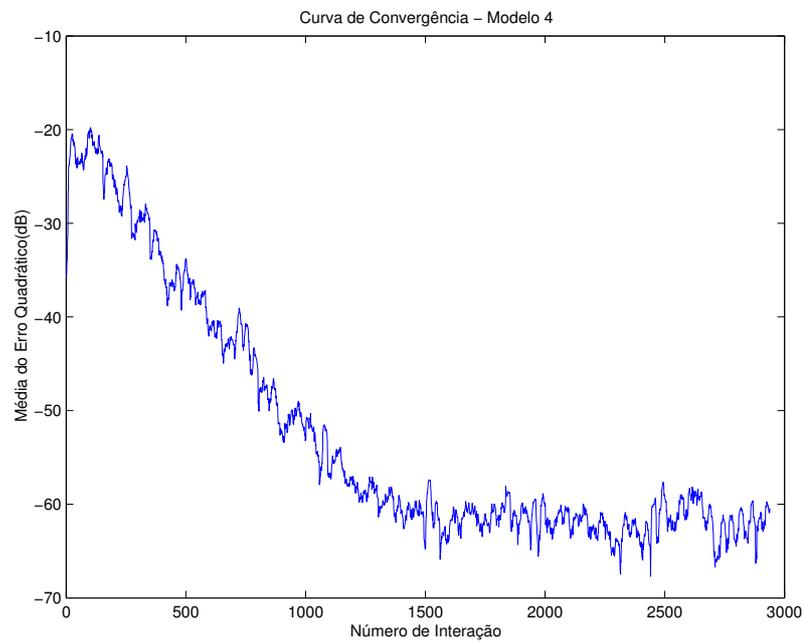


Figura 6.14 Média dos erros quadráticos para o modelo de caminho de eco 4

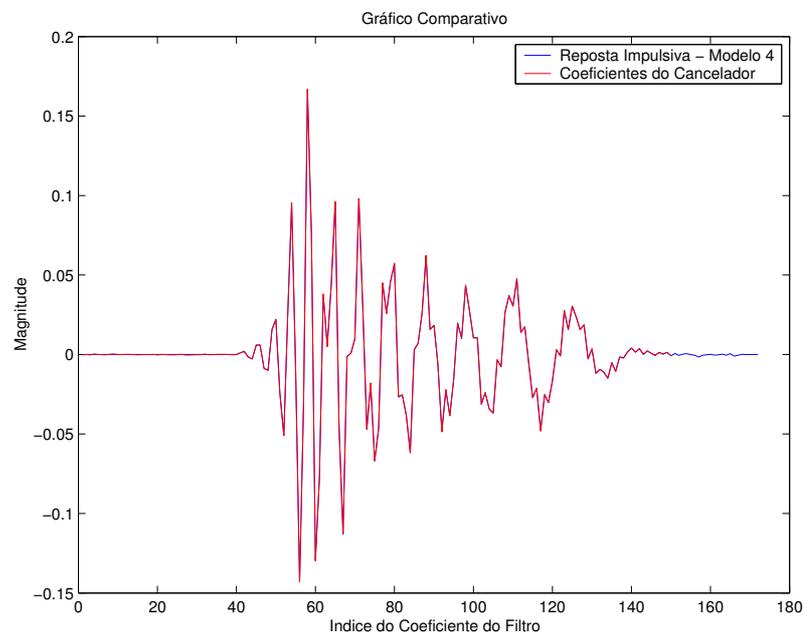


Figura 6.15 Comparação entre o modelo de eco estimado e o modelo 4

A Figura 6.16 mostra a convergência do cancelador de eco para o modelo de caminho de eco 5, utilizando como parâmetro o ERLE. O algoritmo converge após 710 interações, representando um tempo de convergência de 88 ms. Pode-se notar que o algoritmo estabiliza-se em torno de 33 dB após 1000 interações. Na Figura 6.17, observa-se um erro residual em torno de -65 dB. Na Figura 6.18, pode-se verificar que os coeficientes estimados na simulação estão bem próximos do ideal. O número de coeficientes do filtro adaptativo é um pouco menor que o modelo, porém foi suficiente para modelar os principais componentes oscilatórios da resposta impulsiva levando a um pequeno erro residual. Foram utilizados 128 coeficientes no filtro adaptativo da simulação, enquanto o modelo utiliza 140 coeficientes. Observa-se também um atraso inicial de 60 amostras, representando 7,5 ms a 8 kHz.

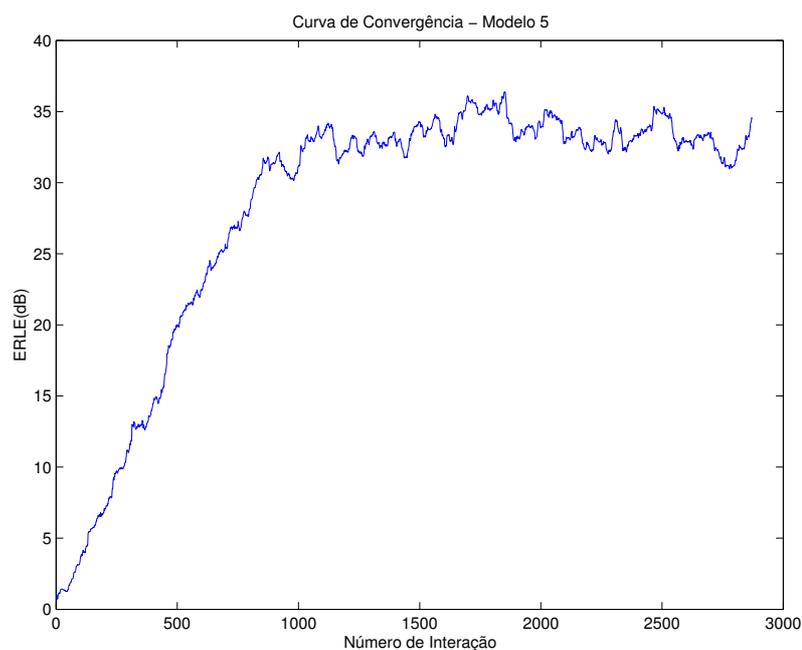


Figura 6.16 Gráfico de convergência para o modelo de caminho de eco 5

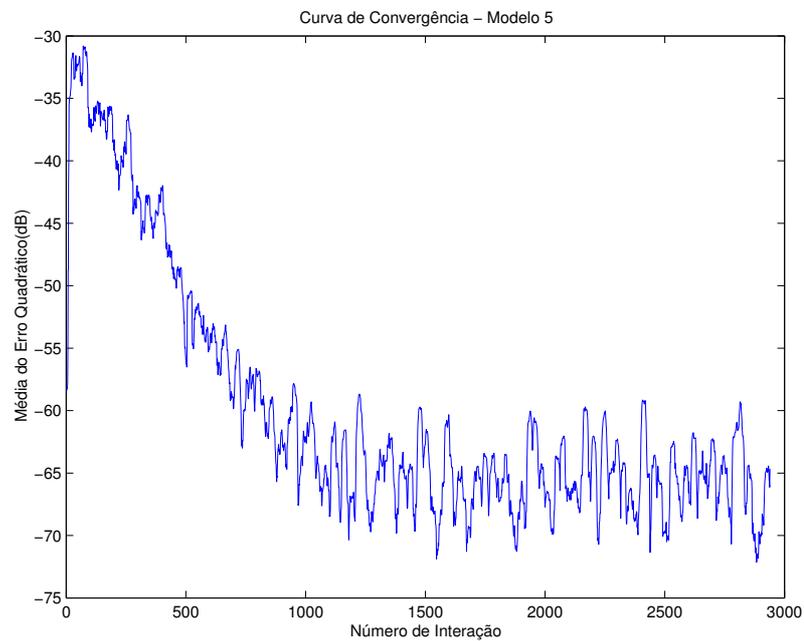


Figura 6.17 Média dos erros quadráticos para o modelo de caminho de eco 5

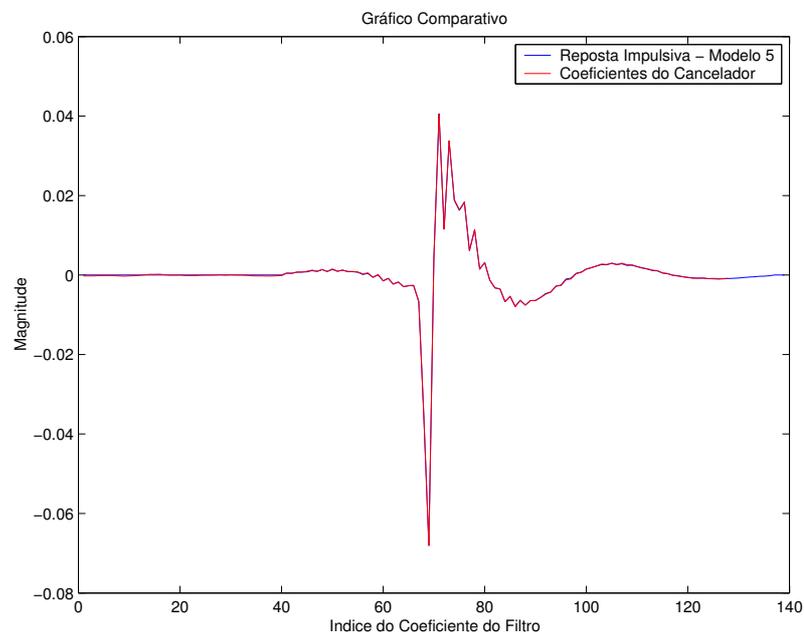


Figura 6.18 Comparação entre o modelo de eco estimado e o modelo 5

A Figura 6.19 mostra a convergência do cancelador de eco para o modelo de caminho de eco 6, utilizando como parâmetro o ERLE. Utilizando-se a mesma métrica anterior, o algoritmo converge, porém não atingi o limiar esperado. O valor máximo do ERLE atingido é menor que 27 dB. Pode-se notar que o algoritmo se estabiliza em torno de 17 dB após 1000 interações. Na Figura 6.20, observa-se um erro residual em torno de -50 dB. Na Figura 6.21, pode-se verificar que os coeficientes estimados na simulação tentam atingir o ideal. Neste caso, o número de coeficientes do filtro adaptativo é menor que o modelo do caminho de eco 6, não sendo suficiente para modelar toda a resposta impulsiva, deixando de estimar uma parte significativa, levando a um grande erro residual. Foram utilizados 128 coeficientes no filtro adaptativo da simulação, enquanto o modelo utiliza 160 coeficientes. Observa-se também um atraso inicial de 60 amostras, representando 7,5 ms a 8 kHz.

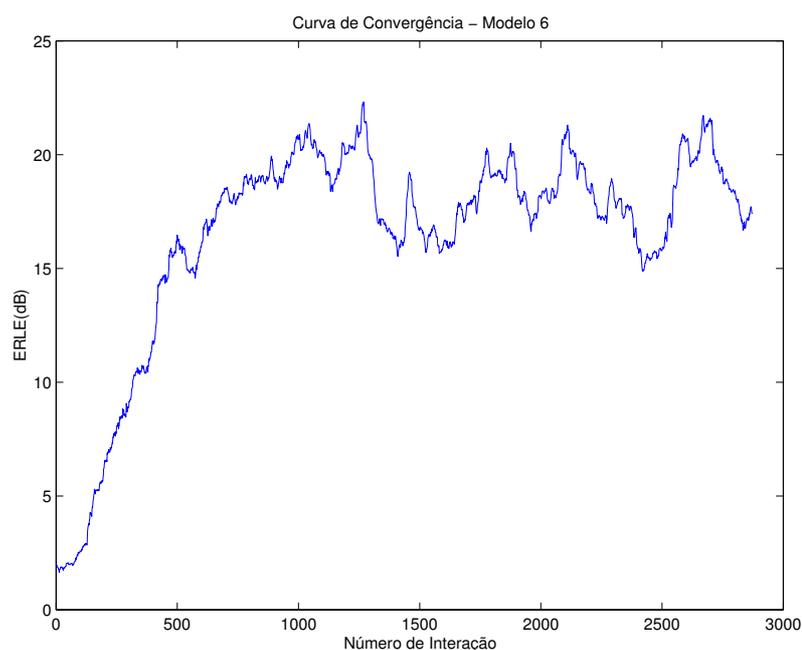


Figura 6.19 Gráfico de convergência para o modelo de caminho de eco 6

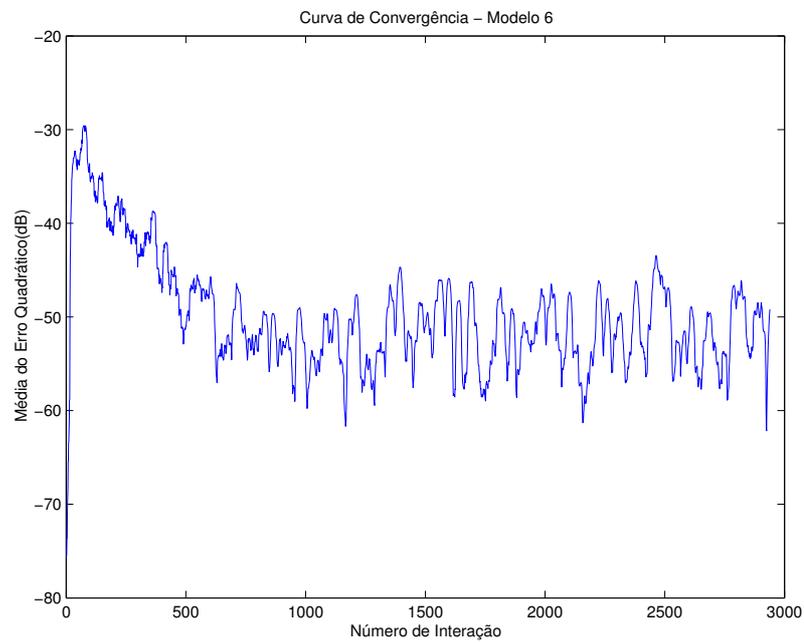


Figura 6.20 Média dos erros quadráticos para o modelo de caminho de eco 6

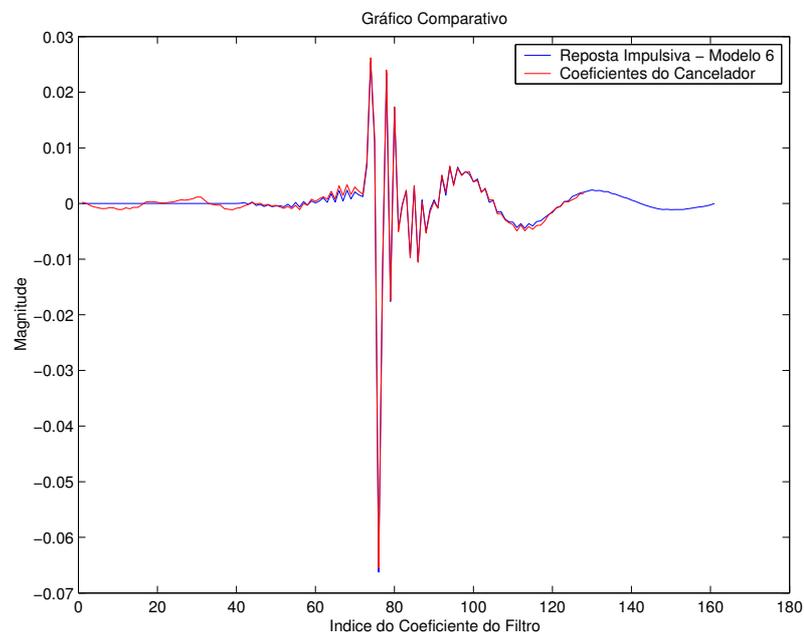


Figura 6.21 Comparação entre o modelo de eco estimado e o modelo 6

A Figura 6.22 mostra a convergência do cancelador de eco para o modelo de caminho de eco 7, utilizando como parâmetro o ERLE. O algoritmo converge após 870 interações, representando um tempo de convergência de 109 ms. Pode-se notar que o algoritmo estabiliza-se em torno de 35 dB após 1200 interações. Na Figura 6.23, observa-se um erro residual em torno de -70 dB. Observando a Figura 6.24 pode-se verificar que os coeficientes estimados na simulação estão bem próximos do ideal. O número de coeficientes do filtro adaptativo é um pouco menor que o modelo, porém foi suficiente para modelar os principais componentes oscilatórios da resposta impulsiva, levando a um pequeno erro residual. Foram utilizados 128 coeficientes no filtro adaptativo da simulação, enquanto o modelo utiliza 140 coeficientes. Observa-se também um atraso inicial de 60 amostras, representando 7,5 ms a 8 kHz.

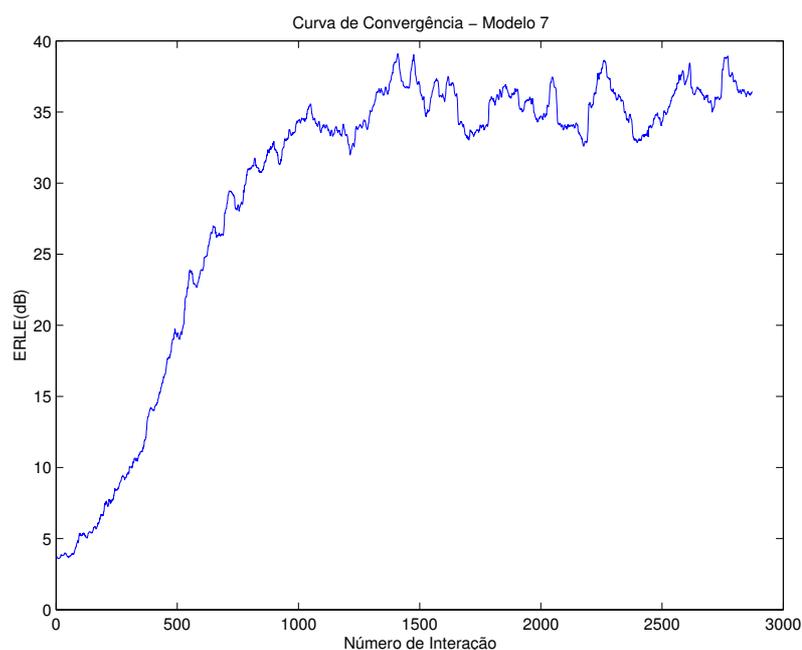


Figura 6.22 Gráfico de convergência para o modelo de caminho de eco 7

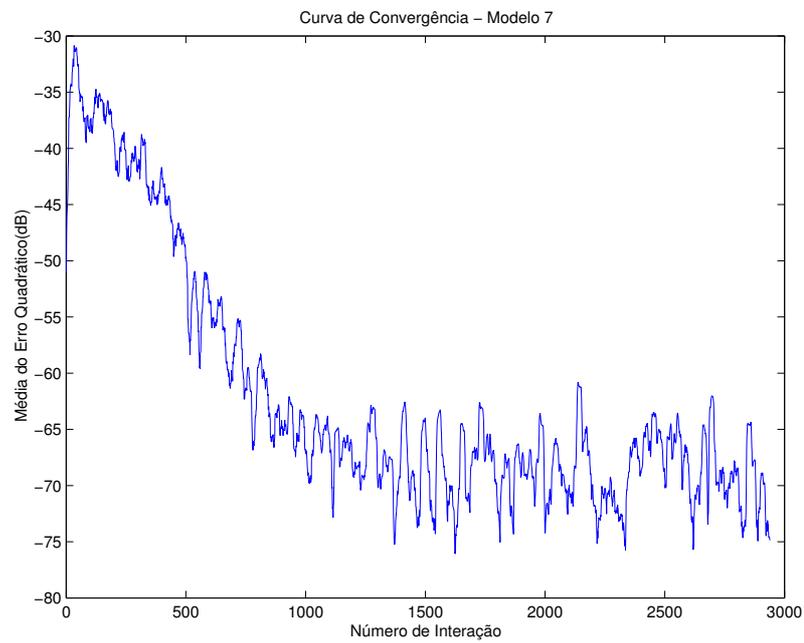


Figura 6.23 Média dos erros quadráticos para o modelo de caminho de eco 7

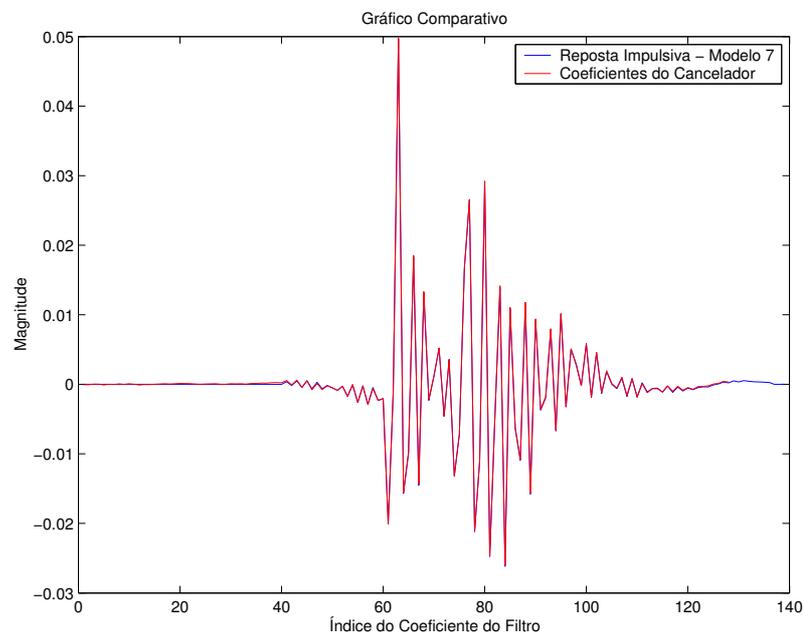


Figura 6.24 Comparação entre o modelo de eco estimado e o modelo 7

O cancelador de eco obteve um resultado satisfatório. O algoritmo adaptativo empregado conseguiu encontrar uma boa estimativa do caminho de eco para os modelos utilizados. Desta forma, é possível sintetizar o sinal de eco e subtraí-lo do sinal de retorno da híbrida, resultando em um sinal de retorno de eco desprezível.

Conforme visto na simulação dos modelos 4 e 6, mesmo que o número de coeficientes utilizados no filtro adaptativo seja insuficiente para envolver toda a resposta impulsiva do caminho de eco, o algoritmo adaptativo convergiu. Os coeficientes estimados ficaram próximos dos ideais. Apenas uma pequena cauda da resposta impulsiva do caminho de eco não foi modelado, devido ao número inferior de coeficientes do filtro adaptativo em relação ao número de coeficientes do modelo. Mesmo para um caminho de eco com dupla reflexão como o modelo 7, o cancelador obteve um bom desempenho.

Para demonstrar o desempenho do cancelador para o modelo 4, foi aumentado o número de coeficientes do filtro adaptativo e executado uma nova simulação. Conforme apresentado nas figuras 6.13 e 6.15, o algoritmo convergiu no limiar esperado e uma boa estimativa dos coeficientes foi obtida.

Em uma aplicação real, deve-se realizar vários testes com diferentes números de coeficientes até obter um bom desempenho. O aumento do número de coeficientes acarreta em uma maior carga computacional e maior gasto de memória. Na prática, são utilizados algoritmos que identificam o atraso inicial do caminho de eco, atraso puro, e realizam os cálculos a partir desta janela de tempo. Com isto é possível reduzir o número de coeficientes do filtro. Outro método também utilizado é o uso de canceladores em cascata, geralmente três, sendo que cada um fica responsável por cancelar o eco em faixas diferentes de tempo [22] [29] [30].

6.2 Cancelamento de Eco na rede IP

A rede IP não é uma fonte geradora de eco. O eco surge na parte analógica da rede de telefonia pública. Porém, como a rede IP possui uma latência maior, o efeito do eco fica mais evidente ao passar pela rede IP.

A rede de telefonia possui uma latência pequena. Geralmente, o eco retorna para o falante em um intervalo de tempo tão curto, que parece estar ocorrendo instantaneamente, não precisando ser atenuado. O intervalo entre o instante em que a fala é transmitida e o instante em que o eco chega no transmissor é chamado de atraso.

A rede IP é uma rede de roteamento de pacotes, originalmente não concebida para o tráfego de sinais de tempo real, tais como voz, devido ao comportamento aleatório de seu tráfego, ocorrendo com muita frequência variações da taxa de transmissão. Não possui reserva de banda e todo o recurso é compartilhado por várias aplicações. Com o surgimento do IPv6, muitas melhorias utilizando as técnicas de qualidade de

serviço deverão tornar VoIP mais eficiente. Fatores como o *jitter*, a latência, perda de pacotes além do eco influenciam na qualidade da transmissão de voz.

Devido ao fato da latência na rede IP ser bem maior do que na rede de telefonia, um sinal de eco proveniente da rede PSTN que possui um atraso pequeno, que não causaria um incômodo ao ouvinte, ao passar pela rede IP, torna-se um incômodo para ele. Atualmente, o atraso acrescentado é na ordem de décimos de segundo. O uso do cancelador de eco no *gateway* é importante para evitar que o eco da rede PSTN entre na rede IP.

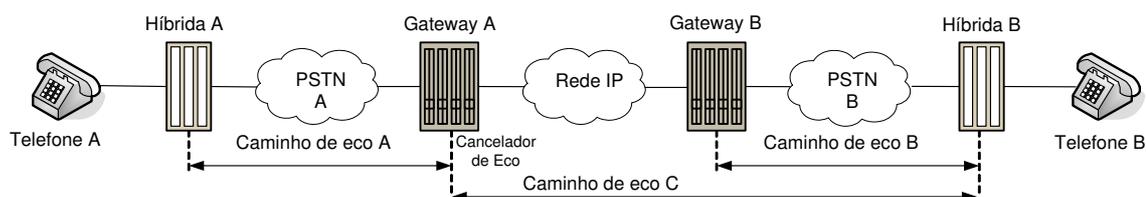


Figura 6.25 Processo envolvido no cancelamento de eco na rede IP

Considerando-se a situação apresentada na Figura 6.25, um aparelho telefônico *A* conectado à rede *PSTN A*, se comunica com um aparelho *B* conectado a outra rede *PSTN B*. As redes estão conectadas com a rede IP através de dois *gateways*. O *gateway A* que possui uma conexão com a rede *PSTN A* e o *gateway B* que se conecta com a rede *PSTN B*. O *gateway A* possui um cancelador de eco impedindo que o eco se propague para a rede IP. O *gateway B* não possui nenhum cancelador de eco. Após estabelecida a comunicação entre os aparelhos, um assinante fala em *B*, sua voz se propaga na rede *PSTN B* até o *gateway B*. No *gateway* o sinal de voz que pode estar no formato PCM linear é codificado por um *codec* de fala, podendo ser o G.711, e empacotado em vários pacotes IPs. Estes trafegam pela rede IP até o *gateway A*. Os pacotes chegam desordenados e alguns podem se perder. Estes pacotes são armazenados em um *buffer de jitter* para serem reordenados e processados após um tempo denominado *playout time*. Existem algoritmos que conseguem interpolar alguns pacotes perdidos, estes podem ser utilizados para melhorar a qualidade do sinal. Após um intervalo de tempo, os pacotes são decodificados, transformados em amostras PCM e comprimidos utilizando-se a lei-*A* ou lei- μ . Posteriormente, são enviados pela rede *PSTN A* até o aparelho *A*. Porém, na rede telefônica, o sinal que trafega em um circuito a quatro fios se reflete na híbrida que se localiza na placa do assinante na central telefônica. Após alguns milésimos de segundos, este sinal está de volta ao *gateway A*. Como o cancelador de eco está em perfeito funcionamento, este sinal não será empacotado de volta para a rede IP. Para este cancelamento ocorrer, o cancelador deve armazenar os sinais que estão sendo enviados para a rede PSTN por um certo tempo, equivalente ao atraso do caminho de eco da rede *PSTN*

A. O número de coeficientes utilizados no filtro adaptativo do cancelador é calculado a partir deste valor. Se tudo estiver operando bem, o sinal de eco será subtraído pelo sinal estimado pelo filtro do cancelador. Neste caso, o eco proveniente do *caminho de eco A* está sendo cancelado. O sinal resultante é identificado como um ruído pelo codificador de voz do *gateway A*. Sendo assim, serão gerados pacotes de controle, identificando que o assinante *A* não está falando, e enviados para o *gateway B* que gera um sinal de ruído ambiente para o assinante *B*. Este processo é executado pelos codificadores de voz que possuem as funções VAD e CNG, citados no capítulo 4.

Na situação considerada anteriormente, o assinante *B* estava falando enquanto o assinante *A* estava apenas ouvindo. Em uma situação inversa, em que o *A* está falando para o *B*, o processo só se difere em um ponto. Como o *gateway B* não possui um cancelador de eco, o sinal que retorna da híbrida na rede *PSTN B* é transmitido para o *gateway A* como se fosse a fala do assinante *B*. O assinante *A* vai escutar uma replica de sua fala após alguns décimos de segundo. Neste caso, todo o esforço para implementar um cancelador de eco no *gateway A* serviu para o assinante da rede *PSTN B* não escutar o eco de sua voz. Enquanto que o assinante da rede *PSTN A* não está sendo favorecido.

A proposta é implementar um cancelador de eco que trate do eco que chega no *gateway* pela rede IP. Utilizando como exemplo a situação anterior, um cancelador desta forma localizado no *gateway A* deverá ter a mesma estimativa do *caminho de eco B* que teria se estivesse no *gateway B*. A única diferença é que o atraso inicial, ou atraso puro, será muito maior e irá variar com o tempo. Devido à latência e à variação da taxa de transmissão da rede IP. Estes fatores são controlados pelo *playout time*. O cancelador vai utilizar este valor de tempo como estimativa do atraso inicial. Assim, o cancelador estima o *caminho de eco C* indicado na Figura 6.25.

Como na rede IP o atraso dos pacotes são diferentes uns dos outros e muda a todo o instante, o receptor precisa armazenar uma quantidade de pacotes antes de processá-los. A variação em relação ao tempo médio gasto para o pacote ser transmitido do destinatário para o destino é chamado de *jitter*. A quantidade de pacotes armazenados influencia no atraso total da comunicação. Quanto maior o *jitter*, maior deverá ser o número de pacotes armazenados para que não ocorra cortes no sinal de fala e, conseqüentemente, maior será o atraso acrescentado.

Em uma transmissão com tratamento de *jitter*, os pacotes que chegam no receptor são armazenados na memória e ordenados de acordo com seu índice. No início, o algoritmo espera a chegada de uma certa quantidade de pacotes para posteriormente processá-los. Este tempo de espera é denominado *playout time*. Ao término deste período, inicia a decodificação dos pacotes a uma taxa constante. Se algum pacote não chegar antes do instante certo para ser tocado, este será considerado perdido e o próximo será processado.

Existem algoritmos que estimam o *playout time*, tentando minimizá-lo mantendo

a taxa aceitável de perda de pacotes. Os algoritmos de controle de *jitter* utilizam uma estimativa inicial baseada no tempo gasto por um pacote para ir e voltar da origem para o destino. Uma nova estimativa do *playout time* é realizada em períodos de silêncio da conversação. Assim a taxa continua constante enquanto tiver atividade de voz para processamento.

A simulação realizada considera que ocorre uma variação aleatória do *playout time* nos intervalos de silêncio. Assume-se que o valor do tempo de atraso estimado está disponível para o cancelador de eco. A Figura 6.26 mostra os sinais de entrada e de eco utilizado na simulação. O intervalo de silêncio apresentado no gráfico corresponde a 187 ms. A Figura 6.27 apresenta os valores do *playout time* utilizado para cada interação da simulação, sendo que o valor inicial foi de 100 ms. Após cada intervalo de silêncio, um novo valor do *playout time* é utilizado. Este valor foi gerado por uma variável aleatória que é acrescentada ao valor anterior, representando o comportamento aleatório do *jitter*. A Figura 6.28 mostra as curvas de convergência utilizando o valor ERLE como parâmetro para dois métodos diferentes de cancelamento.

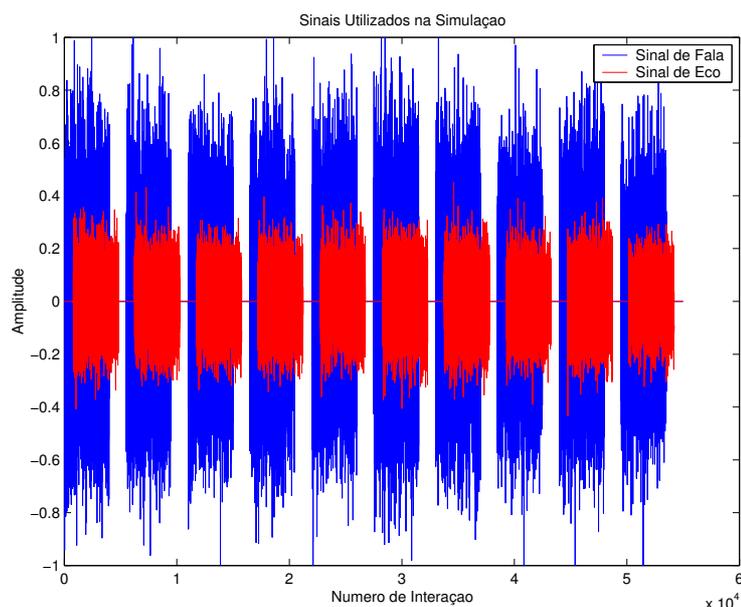


Figura 6.26 Sinais utilizados na simulação

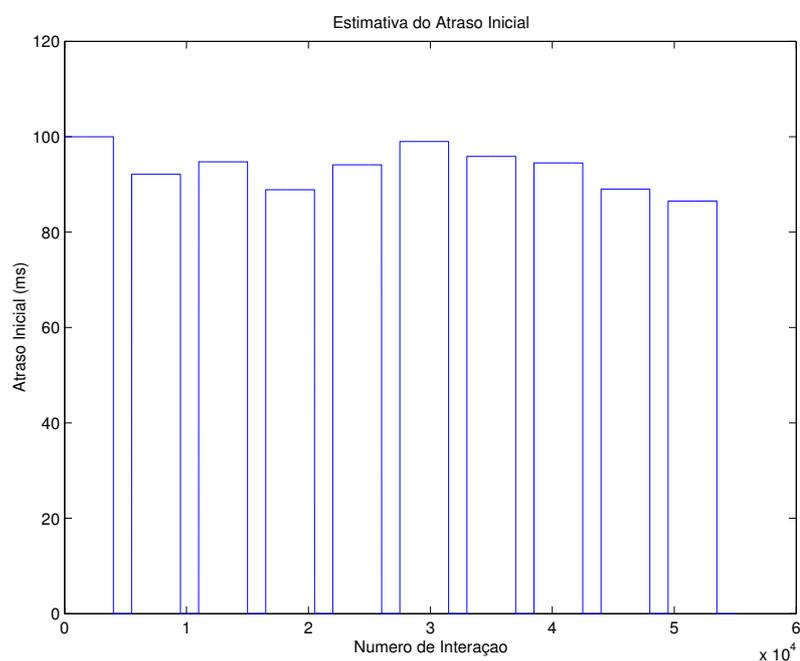


Figura 6.27 Valor do atraso inicial em cada interação da simulação

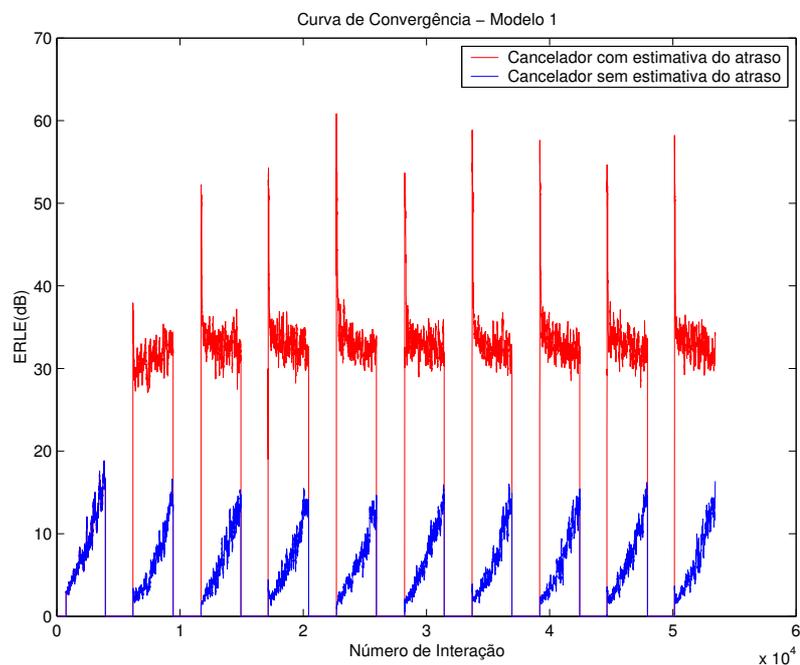


Figura 6.28 Gráfico de convergência do cancelador de eco na rede IP

Na Figura 6.28, a curva vermelha representa a convergência de um cancelador que utiliza o valor do *playout time* para estimar o atraso inicial do caminho de eco. Assim, quando ocorre a mudança de atraso, o algoritmo não demora para convergir novamente. Os coeficientes obtidos no período de fala anterior, são utilizados novamente, sendo que apenas é acrescentado um atraso na resposta do filtro. A curva azul representa o desempenho de um cancelador que não utiliza a estimativa do atraso. Sempre que ocorre uma mudança no atraso puro do caminho de eco, o algoritmo precisa convergir novamente, buscando novos coeficientes ótimos. Esta é a razão da queda no valor do ERLE logo que o atraso muda. Pode-se observar também que no início, os dois algoritmos têm a mesma resposta. Ambos são inicializados com um atraso de 100 ms, o que justifica a sobreposição inicial da curva azul à vermelha.

Em uma rede IP, que possui atraso variado, para o cancelador de eco modelar o caminho de eco que passa por esta rede, é preciso estimar a variação do atraso da rede. Como estimativa, pode ser utilizado o tempo de *playout*. Na simulação apresentada, pode-se observar que o cancelador utilizando este artefato obteve um desempenho melhor, mantendo a convergência após a mudança do atraso da rede. Enquanto que o cancelador sem estimar a variação do atraso não consegue convergir. A elucidação deste fato é uma das contribuições mais importantes deste trabalho.

Capítulo 7

Conclusões

Essa dissertação apresentou um estudo do cancelamento do eco aplicado tanto na telefonia IP quanto na rede de telefonia comutada. Como os canceladores de eco baseiam-se na teoria de filtros adaptativos, para desenvolvimento do tema, foi apresentado no Capítulo 2 os conceitos básicos da teoria de filtros adaptativos. Entretanto, o objetivo desse trabalho não é propor uma nova estrutura de filtros, nem muito menos um novo algoritmo para adaptação dos coeficientes do filtro. Portanto, foi utilizado apenas o filtro FIR, com o algoritmo adaptativo NLMS tradicionalmente utilizado. No Capítulo 6, foi simulado o seu desempenho utilizando os modelos do caminho de eco da recomendação G.168 do ITU-T [2]. Foi verificado um desempenho satisfatório com uma rápida convergência para os filtros que possuem o número de coeficientes suficiente para envolver toda a resposta impulsiva. Para os casos em que a quantidade dos coeficientes foi menor que o necessário, foi verificado um erro residual que pode ser elevado o suficiente para impedir a convergência do algoritmo para o nível aceitável de ERLE estipulado na recomendação G.168.

Na simulação realizada na seção 6.1, foram utilizados os modelos da rede PSTN e a mesma estrutura do cancelador de eco desta rede. O funcionamento do cancelador tradicionalmente utilizado nas centrais telefônicas para ligações a longa distância é semelhante ao apresentado. Entretanto a importância dessa simulação é apresentada na seção 6.2, quando se justifica o uso deste cancelador no *gateway* IP para impedir que o eco mesmo com um pequeno atraso deve ser eliminado antes de entrar na rede IP. Foi verificado que devido à esta rede de dados possuir um elevado atraso de transmissão, o eco torna-se um incômodo ao ouvinte do outro lado da rede. Finalmente, observa-se que o cancelador utilizado na rede IP é o mesmo empregado na rede PSTN.

Além deste estudo realizado, na seção 6.2 foi proposto um novo método de cancelador IP que cancela o eco nas duas direções, impedindo que o eco se propague tanto da rede PSTN para a rede IP, quanto da rede IP para a rede PSTN. Este cancelador utiliza uma estimativa da variação do atraso da rede IP, como parâmetro

de entrada do algoritmo adaptativo para aproximar os coeficientes do filtro ótimo mais rapidamente. Foi mostrado na Figura 6.28 que o cancelador proposto conseguiu convergir mesmo com as variações do atraso da rede IP, enquanto que o convencional não obteve um bom resultado.

A motivação para utilizar o novo método proposto é desenvolver um cancelador de eco IP independente do *gateway* do destino. Assim, obtém-se um investimento em um cancelador de eco para benefício da qualidade da própria rede. Além disto, o novo método apresenta o mesmo gasto computacional do cancelador IP empregado atualmente, pois o filtro possui o mesmo número de coeficientes, a mesma estrutura e o mesmo algoritmo adaptativo; porém exige um cancelador de eco para cada sentido de transmissão da fala.

Referências Bibliográficas

- [1] PHILLIPS, G. M. Echo and its effects on the telephone user. *Bell Laboratories Record*, v. 32, p. 281–4, August 1954.
- [2] ITU-T. Recommendation G.168 - digital network echo cancellers. Telecommunication standardization section of ITU, International Telecommunication Union, April 2000.
- [3] DINIZ, P. S. R. *Adaptive filtering - algorithms and practical implementation*. Kluwer Academic Publishers.
- [4] BELLANGER, M. *Adaptive digital filters and signal analysis*. Marcel Dekker, 1987.
- [5] HONIG, M. L. & MESSERCHMITT, D. G. *Adaptive filters: Structures, algorithms, and applications*. Kluwer Academic Publisher, 1984.
- [6] HAYKIN, S. *Adaptive filter theory*. 4th edition. ed. Prentice Hall, 2002.
- [7] WIDROW, B. & STEARNS, S. D. *Adaptive signal processing*. Prentice Hall, 1985.
- [8] JOHNSON, D. H. & DUDGEON, D. E. *Array signal processing*. Prentice Hall, 1993.
- [9] ANTONIOU, A. *Digital filters: Analysis, design, and applications*. McGraw Hill, 1992.
- [10] ARDALAN, S. H. Floating-point analysis of recursive least-squares and least-mean squares adaptive filters. *IEEE Trans. on Circuits and Systems*, v. CAS-33, p. 1192–208, December 1986.
- [11] CIOFFI, J. M. Limited precision effects in adaptive filtering. *IEEE Trans. on Circuits and Systems*, v. CAS-34, p. 821–33, July 1987.
- [12] GOODWIN, G. C. & PAYNE, R. L. *Dynamic system identification: Experiments design and data analysis*. Academic Press, 1977.

- [13] EMILING, J. W. & MITCHELL, D. The effects of time delay and echoes on telephone conversations. *Bell System Technical Journal*, p. 2869–91, November 1963.
- [14] CLARK, A. B. & MATHES, R. C. *Echo suppressor for long telephone circuits*, June 1925. Diss. mestrado.
- [15] TELLABS. Echo cancellation. Disponível na internet via Web Proforum Tutorials. <http://www.iec.org>, January 2004.
- [16] HORTON Jr., W. A. The occurrence and effects of lockout occasioned by two echo suppressors. *B.S.T.J.*, v. 17, p. 258–80, April 1938.
- [17] BRADY, P. T. & HELDER, G. K. Echo supressor design in telephone communications. *B. S. T. J.*, v. 42, p. 2893–917, November 1963.
- [18] RIESZ, R. R. & KLEMMER, E. T. Subjective evaluation of delay and echo supressors in telephone communications. *B. S. T. J.*, v. 42, p. 2919–41, November 1963.
- [19] SONDHI, M. M. & PRESTI, A. J. A self-adaptive echo canceller. *Bell System Technical Journal*, v. 45, n. 10, p. 1851–4, December 1966.
- [20] BECKER, F. K. & RUDIN, H. R. Application of automatic transversal filters to the problem of echo suppression. *Bell System Technical Journal*, v. 45, n. 10, p. 1847–50, December 1966.
- [21] SONDHI, M. M. An adaptive echo canceller. *Bell System Technical Journal*, v. 46, n. 3, p. 497–511, March 1967.
- [22] CORMACK, B. M. Eliminating echo critical for voip. Whitepaper, Disponível na internet via <http://www.commsdesign.com>, June 2001.
- [23] INTEL. Overcoming barriers to high-quality voice over IP deployments. Whitepaper, Disponível na internet via <http://www.intel.com>.
- [24] MOCKINGBIRD NETWORKS. Transporting voice over IP - the issues of quality, echo and latency. Whitepaper, Disponível na internet via <http://itpapers.zdnet.com>.
- [25] GL Communication Inc. Voice quality on next generation networks demands testing tools. Whitepaper, Disponível na internet via <http://whitepapers.zdnet.co.uk>.

-
- [26] ITU-T. Recommendation G.165 - echo cancellers. Telecommunication standardization section of ITU, International Telecommunication Union, March 1993.
- [27] ITU-T. Recommendation G.164. Telecommunication standardization section of ITU, International Telecommunication Union.
- [28] MURANO, K.; UNAGAMI, S.; AMANO, F. Echo cancelation and applications. Technical report, IEEE Communitations Magazine, January 1990.
- [29] MAPLETREE NETWORKS. Echo cancelation for voice over IP. Technical report, www.mapletree.com, 2001.
- [30] IEEE. IEEE standard equipment requirements and measurements techniques for analog transmission parameters for telecommunications. Std 743-1995, Transmission Systems and Access Committee of the IEEE Communications Society, December 1995.