USP

Campus de São Carlos

# APLICAÇÃO DE REDES NEURAIS ARTIFICIAIS E FILTRO DE KALMAN PARA REDUÇÃO DE RUÍDOS EM SINAIS DE VOZ

Autor: Antonio Marcos Selmini

Orientador: Prof. Dr Marcelo Basílio Joaquim

UNIVERSIDADE DE SÃO PAULO

ESCOLA DE ENGENHARIA DE SÃO CARLOS

# APLICAÇÃO DE REDES NEURAIS ARTIFICIAIS E FILTRO DE KALMAN PARA REDUÇÃO DE RUÍDOS EM SINAIS DE VOZ

ANTONIO MARCOS SELMINI

Serviço de Pós-Graduação EESC/USP
EXEMPLAR REVISADO
Data de entrada no Serviço: 13 / 08 / 01
Ass.: Gui Ehrme

Dissertação apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica.

EESC - USP
Biblioteca
S. CARLOS

ORIENTADOR: Prof. Dr. Marcelo Basílio Joaquim

DEDALUS - Acervo - EESC
31100036747

São Carlos
2001

Class. TESE-EESC
Cutt. 5276
Tombo T0181/01

nt 1199246

Ficha catalográfica preparada pela Seção de Tratamento
da Informação do Serviço de Biblioteca – EESC/USP

S468a

Selmini, Antonio Marcos
Aplicação de redes neurais artificiais e filtro de Kalman para redução de ruídos em sinais de voz / Antonio Marcos Selmini. -- São Carlos, 2001.

Dissertação (Mestrado) -- Escola de Engenharia de São Carlos-Universidade de São Paulo, 2001.
Área: Engenharia Elétrica.
Orientador: Prof. Dr. Marcelo Basílio Joaquim.

1. Filtro de Kalman. 2. Filtro de Kalman estendido. 3. Redes neurais artificiais. 4. Redes RBF. I. Título.

**FOLHA DE JULGAMENTO**

Candidato: Tecnólogo **ANTONIO MARCOS SELMINI**

Dissertação defendida e julgada em 19-06-2001 perante a Comissão Julgadora:

Prof. Dr. **MARCELO BASÍLIO JOAQUIM (Orientador)** — Aprovado
(Escola de Engenharia de São Carlos/USP)

Prof. Tit. **JOSÉ CARLOS PEREIRA** — Aprovado
(Escola de Engenharia de São Carlos/USP)

Prof. Tit. **FABIO VIOLARO** — Aprovado
(Universidade Estadual de Campinas - UNICAMP)

Profa. Associada **VILMA ALVES DE OLIVEIRA**
Coordenadora do Programa de Pós-Graduação
em Engenharia Elétrica

**JOSÉ CARLOS A. CINTRA**
Presidente da Comissão de Pós-Graduação da EESC

*Aos meus pais e meus avós*
*que sempre acreditaram em mim.*

# AGRADECIMENTOS

Ao Prof. Dr. Marcelo Basílio Joaquim, pela orientação e principalmente pela paciência.

Às Profas. Vilma e Célia dos Colégios Objetivo de Itajobi e Catanduva respectivamente, pelo grande incentivo e ajuda que foram essenciais para que eu pudesse chegar até este ponto.

À Cristina da Dataflash Informática e ao Clóvis do Colégio Fênix que permitiram meu afastamento semanal do trabalho para que este projeto de pesquisa pudesse ser realizado.

Às amigas Sandra, Fátima, Giovana e Kátia pelas caronas para São Carlos.

Ao meu grande amigo Armando, obrigado!

À Profa. Zélia do Colégio Fênix pelas dicas e correção deste trabalho.

# SUMÁRIO

**LISTA DE FIGURAS** ..... **viii**

**LISTA DE TABELAS** ..... **xi**

**LISTA DE SIGLAS E ABREVIATURAS** ..... **xii**

**LISTA DE SÍMBOLOS** ..... **xiii**

**RESUMO** ..... **xvi**

**ABSTRACT** ..... **xvii**

**1 – INTRODUÇÃO** ..... **1**

1.1 – Motivação ..... 1

1.2 – Objetivos ..... 2

1.3 – Organização do trabalho ..... 3

**2 – MODELO FISIOLÓGICO DE PRODUÇÃO DE VOZ E PREDIÇÃO LINEAR** ..... **6**

2.1 – Introdução ..... 6

2.2 – Modelo Fisiológico de Produção de Voz ..... 7

2.3 – Representação de Sinais de Voz ..... 11

2.4 – Modelo de Produção de Voz ..... 11

2.5 – Predição Linear de Sinais de Voz ..... 17

2.6 – Método da Autocorrelação ..... 25

2.7 – Método da Covariância ..... 27

2.8 – Considerações Finais ..... 29

**3 – ENRIQUECIMENTO DE SINAIS DE VOZ ..... 30**

3.1 – Introdução ..... 30

3.2 – Classificação dos Métodos de Enriquecimento de Sinais de Voz ..... 33

3.2.1 – Subtração Espectral ..... 34

3.3 – Filtros ótimos ..... 51

3.3.1 – Filtro de Wiener ..... 53

3.4 – Considerações Finais ..... 57

**4 – FILTRO DE KALMAN ..... 59**

4.1 – Introdução ..... 59

4.2 – Formulação do Filtro de Kalman ..... 62

4.3 – Modelo Recursivo do Filtro de Kalman ..... 70

4.4 – Forma Alternativa do Filtro de Kalman ..... 74

4.5 – Divergência do Filtro de Kalman ..... 79

4.6 – Filtro de Kalman Estendido ..... 80

4.7 – Considerações Finais ..... 87

**5 – REDES NEURAIS ARTIFICIAIS ..... 88**

5.1 – Introdução ..... 88

5.2 – Histórico ..... 89

5.3 – Neurônios Biológicos e Artificiais ..... 91

5.4 – Principais Arquiteturas de Redes Neurais ..... 95

5.4.1 – Redes com Camada Única ..... 95

5.4.2 – Redes de Múltiplas Camadas ..... 95

5.4.3 – Redes Recorrentes ..... 97

5.5 – Aprendizagem em Redes Neurais Artificiais ..... 98

5.5.1 – Aprendizagem Supervisionada ..... 98

5.5.2 – Aprendizagem Não-Supervisionada ..... 99

5.5.3 – Aprendizagem por Reforço ..... 100

5.6 – Redes com Múltiplas Camadas ..... 101

5.6.1 – Algoritmo de Retropropagação ..... 102

5.6.2 – Resumo do Algoritmo de Retropropagação ..... 106

5.7 – Redes de Função de Base Radial ("*Radial Basis Function – RBF*") ..... 107

5.7.1 – Treinamento das Redes RBF ..... 109

5.8 - Comparação entre Redes RBF e Perceptrons de Múltiplas Camadas ..... 112

5.9 – Considerações Finais ..... 113

**6 – FILTRO DUPLO DE KALMAN ESTENDIDO ..... 114**

6.1 – Introdução ..... 114

6.2 – Descrição do Algoritmo ..... 115

6.3 – Alterações no Algoritmo ..... 118

6.4 – Estimação dos Estados ..... 119

6.5 – Estimação dos Pesos ..... 120

6.6 – Implementação e Avaliação do Algoritmo ..... 121

6.6.1 – Processamento Utilizando Ruído Branco Gaussiano ..... 122

6.6.2 – Processamento Utilizando Ruído Branco Não Estacionário ..... 127

6.6.3 – Pré-Filtragem Utilizando o Filtro de Média ..... 130

6.7 – Considerações Finais ..... 133

**7 – CONCLUSÕES ..... 135**

**BIBLIOGRAFIA ..... 139**

# LISTA DE FIGURAS

FIGURA 2.1 – Aparelho fonador humano ..... 8

FIGURA 2.2 – Diagrama em blocos do processo de produção da voz ..... 9

FIGURA 2.3 – Etapas do processo de produção da voz ..... 10

FIGURA 2.4a – Forma de onda para um som sonoro ..... 12

FIGURA 2.4b – Forma de onda para um som surdo ..... 12

FIGURA 2.5 – Modelo linear de produção da voz ..... 14

FIGURA 2.6 – Modelo geral discreto no tempo para produção de sinais de voz ..... 18

FIGURA 2.7 – Preditor linear de ordem $M$ ..... 19

FIGURA 3.1 – Fontes de degradação empregada no enriquecimento de sinais de voz ..... 32

FIGURA 3.2 – Diagrama em blocos do sistema proposto por BOLL ..... 40

FIGURA 4.1 – Ciclo do filtro de Kalman discreto ..... 71

FIGURA 4.2 – Ciclo de operação do filtro de Kalman discreto ..... 73

FIGURA 4.3 – Ciclo de operação do filtro de Kalman discreto usando formas alternativas ..... 78

FIGURA 4.4 – Ciclo de operação do filtro de Kalman estendido ..... 86

FIGURA 5.1 – Modelo anatômico simplificado de um neurônio biológico ..... 92

FIGURA 5.2 – Modelo de um neurônio artificial ..... 93

FIGURA 5.3 – Modelo não linear de um neurônio artificial. ..... 94

FIGURA 5.4 – Rede neural com apenas uma camada de neurônios ..... 95

FIGURA 5.5 – Rede neural com múltiplas camadas ..... 96

FIGURA 5.6 – Rede neural recorrente sem neurônios escondidos e sem laços de auto-realimentação ..... 97

FIGURA 5.7 – Rede neural recorrente com neurônios ocultos e laços de auto-realimentação ..... 98

FIGURA 5.8 – Modelo de aprendizagem por reforço ..... 101

FIGURA 5.9 – Rede neural com duas camadas ocultas e uma camada de saída ..... 102

FIGURA 5.10 – Rede de função de base radial ..... 108

FIGURA 6.1 – Diagrama em blocos do algoritmo proposto por WAN & NELSON (1998b) ..... 117

FIGURA 6.2 – Forma de onda da palavra "elétrica" ..... 123

FIGURA 6.3 – Ruído branco gerado para contaminar o sinal da figura 6.2 ..... 123

FIGURA 6.4 – Sinal de voz contaminado por ruído branco gaussiano ..... 124

FIGURA 6.5 – Sinal de voz processado pelo FDKE ..... 124

FIGURA 6.6 – Espectrograma do sinal original ..... 125

FIGURA 6.7 – Espectrograma do sinal corrompido por ruído branco gaussiano ..... 126

FIGURA 6.8 – Espectrograma do sinal filtrado ..... 126

FIGURA 6.9 – Ruído branco não estacionário ..... 127

FIGURA 6.10 – Sinal de voz contaminado com ruído branco não estacionário ..... 128

FIGURA 6.11 – Resultado da filtragem de um sinal contaminado com ruído branco não estacionário ..... 128

FIGURA 6.12 – Espectrograma do sinal de voz corrompido
com ruído branco não estacionário ..... 129

FIGURA 6.13 – Espectrograma do sinal resultante do processamento
de um sinal corrompido com ruído branco não estacionário ..... 129

FIGURA 6.14 – Sinal de voz contaminado por ruído branco
gaussiano antes da pré-filtragem ..... 131

FIGURA 6.15 – Sinal resultante da pré-filtragem ..... 131

FIGURA 6.16 – Resultado da filtragem do sinal pré-filtrado ..... 132

FIGURA 6.17 - Espectrograma do sinal de voz corrompido
com ruído branco gaussiano ..... 133

FIGURA 6.18 – Espectro do sinal resultante do processamento pelo
FDKE, sendo que o pré processamento foi realizado pelo
filtro de média ..... 133

# LISTA DE TABELAS

TABELA 2.1 – Modelo somente com pólos usando o método da autocorrelação ..... 27

TABELA 2.2 – Modelo somente com pólos usando o método da covariância ..... 29

TABELA 3.1 – Resultados dos algoritmos $H_1$, $H_2$ e $H_3$ para os sinais SP1 e SP2 degradados por ruído branco ..... 47

TABELA 3.2 – Resultados dos algoritmos $H_1$, $H_2$ e $H_3$ para os sinais SP1 e SP2 degradados por ruído de automóvel ..... 48

TABELA 3.3 – Equações para o filtro de Wiener ..... 57

TABELA 4.1 – Equações de propagação das estimativas para o filtro de Kalman discreto ..... 72

TABELA 4.2 – Equações de atualização das estimativas para o filtro de Kalman discreto ..... 72

TABELA 4.3 – Equações de propagação das estimativas para o filtro de Kalman estendido ..... 87

TABELA 4.2 – Equações de atualização das estimativas para o filtro de Kalman estendido ..... 87

TABELA 5.1 – Algoritmo *k*-médias ..... 111

# LISTA DE ABREVIATURAS E SIGLAS

| | |
|---|---|
| RBF | - Radial Basis Function (Função de Base Radial) |
| MLP | - Multilayer Perceptrons (Perceptrons Multicamadas) |
| AR | - Auto-regressivo |
| ARMA | -Auto-regressivo média móvel |
| MA | - Média Móvel |
| SNR | - Signal Noise Rate |
| TF | - Transformada de Fourier |
| HFT | - Hands Free Telephone |
| NSEP | - Nova Subtração Espectral de Potências |
| SNR-Seg | - Signal Noise Rate Segmental |
| SEA | - Subtração Espectral em Amplitudes |
| SEP | - Subtração Espectral de Potências |
| SEM | - Subtração Espectral Modificada |
| RNA | - Redes Neurais Artificiais |
| EKG | - Eletrocardiograma |
| EEG | - Eletroencefalograma |
| EMG | - Eletromiograma |
| ENG | - Eletroneurograma |
| FDKE | - Filtro Duplo de Kalman Estendido |
| RAM | - Randon Access Memory |

# LISTA DE SÍMBOLOS

| | |
|---|---|
| $G$ | - modelo glotal |
| $V$ | - modelo do trato vocal |
| $L$ | - modelo de radiação dos lábios |
| $f$ | - freqüência formantes |
| $e$ | - sinal de erro |
| $a$ | - coeficientes de predição linear |
| $z^{-1}$ | - unidade de atraso |
| $x$ | - sinal de voz livre de ruídos |
| $X(z)$ | - transformada $z$ do sinal de voz livre de ruídos |
| $\hat{x}(k)$ | - estimativa do sinal de voz |
| $A(k-1)$ | - matriz de transição de estados variante no tempo |
| $r$ | - autocorrelação do sinal de voz |
| $w_R$ | - janela de análise |
| $\tilde{x}$ | - sinal de voz resultante da aplicação de uma janela de análise |
| $c$ | - covariância do sinal de voz |
| $y$ | - sinal de voz corrompido por ruído aditivo |
| $v$ | - ruído aditivo |
| $Y(e^{j\omega})$ | - espectro de amplitude para o sinal corrompido por ruído aditivo |
| $X(e^{j\omega})$ | - espectro de amplitude para o sinal de voz livre de ruídos |
| $V(e^{j\omega})$ | - espectro de amplitude para o ruído aditivo |
| $\hat{V}(e^{j\omega})$ | - estimativa do espectro de amplitude para o ruído aditivo |
| $\theta_V(e^{j\omega})$ | - fase do ruído |
| $\theta_Y(e^{j\omega})$ | - fase do sinal de voz corrompido por ruído |
| $H(e^{j\omega})$ | - filtro de subtração espectral |

| | |
|---|---|
| $M(e^{j\omega})$ | - espectro de amplitude para o valor médio do sinal corrompido por ruído |
| $\lvert X(e^{j\omega})\rvert$ | - magnitude do espectro do sinal de voz original |
| $\lvert X(e^{j\omega})\rvert^2$ | - espectro de potências para o sinal de voz original |
| $\lvert Y(e^{j\omega})\rvert^2$ | - espectro de potências para o sinal de voz corrompido por ruído aditivo |
| $\lvert V(e^{j\omega})\rvert^2$ | - espectro de potências para o ruído aditivo |
| $\Gamma(.)$ | - função gama |
| $I_0$ | - função de Bessel modificada do tipo zero |
| $I_1$ | - função de Bessel modificada do tipo um |
| $D(e^{j\omega})$ | - fator relativo de degradação |
| $w$ | - ruído de estado |
| $Q$ | - matriz de covariância para o ruído de estado |
| $R$ | - matriz de covariância para o ruído aditivo |
| $C$ | - matriz de medidas |
| $P^-$ | - matriz de covariância do erro de estimação antes da anexação de uma nova medida |
| $P$ | - matriz de covariância do erro de estimação depois da anexação de uma nova medida |
| $\hat{x}^-(k)$ | - estimativa do sinal de voz antes da anexação de uma nova medida |
| $K$ | - Ganho de Kalman |
| $I$ | - matriz identidade |
| $C^T$ | - transposta da matriz de medidas |
| $f(.)$ | - função não-linear que relaciona o estado $x$ no instante de tempo $k$ e $k-1$ |
| $h(.)$ | - função não-linear que representa as conexões entre o vetor de estados $x$ e a medida de ruído $v$ |
| $F$ | - linearização da rede neural |
| $H$ | - linearização da matriz de medidas |

| | |
|---|---|
| $u_j$ | - saída para o neurônio *j* |
| $o_j$ | - saída desejada para o neurônio *j* |
| $\varphi(.)$ | - função de ativação para a rede neural |
| $t_j$ | - campo local para o neurônio *j* |
| $\omega_{ji}$ | - pesos sinápticos entre o neurônio *i* e o neurônio *j* |
| $\delta_i$ | - gradiente local para o neurônio *i* |
| $c$ | - centro da função de base radial |
| $r$ | - raio da função de base radial |
| $\eta$ | - taxa de aprendizagem |
| $\alpha$ | - taxa de momento |

# RESUMO

Selmini, A. M. (2001). *Aplicação de Redes Neurais Artificiais e Filtro de Kalman para Redução de Ruídos em Sinais de Voz*. São Carlos, 2001. Dissertação (mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo.

A filtragem, na sua forma mais geral, tem estado presente na vida do homem há muito tempo. Com o surgimento de novas tecnologias (surgimento da eletricidade e a sua evolução) e o desenvolvimento da computação, as técnicas de filtragem passaram a ser utilizadas largamente na engenharia para a filtragem (separação) de sinais elétricos. Normalmente, os sistemas de comunicação (telefonia móvel e fixa, sinais recebidos de satélites e outros sistemas) contêm sinais indesejáveis responsáveis pela degradação do sinal original. Dentro deste contexto, este projeto de pesquisa apresenta um estudo do algoritmo Filtro Duplo de Kalman Estendido, onde um filtro de Kalman e duas redes neurais são empregados para a redução de ruídos em sinais de voz. O algoritmo estudado foi aplicado ao processamento de um sinal corrompido por dois tipos de ruídos diferentes: ruído branco gaussiano e ruído branco não estacionário, conseguindo-se bons resultados. Uma melhora sensível do sinal filtrado pode ser conseguida com técnicas de pré-filtragem do sinal. Neste trabalho foi utilizado o filtro de médias para a pré-filtragem, obtendo um sinal filtrado com ruído musical de baixa intensidade.

**Palavras-chave:** Filtro de Kalman, Filtro de Kalman Estendido, Redes Neurais Artificiais, Redes RBF.

# ABSTRACT

Selmini, A. M. (2001). *Application of Artificial Neural Networks and Kalman Filtering for Reduction of Noise in Speech Signals*. São Carlos, 2001. Dissertação (mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo.

Filtering in it's most general kind has been present in men's life for a long time. With the appearance of new technologies (appearance of electricity and it's evolution) and the development of the computer science, the filtering techniques started to be widely used in engineering to the filtering (separation) of electric signals. Normally the communication systems (fixed and mobile telephony, signals sent from satellites and other systems) bring undesired results responsible for the degradation of the original signal. Within this context, this research project shows a study of the algorithm Dual Extended Kalman Filtering, in which a Kalman filter and two neural networks are used for the reduction of noise in speech signals. The algorithm studied was applied to the processing of a signal corrupted by two types of different noises: gaussian white noise and non stationary white noise obtaining good results. A significant improvement of the filtered noise can be obtained with techniques of pre-filtering of the signal. In this research the average filter for a pre-filtering was used, obtaining a filtered signal with musical noise of low intensity.

**Keywords:** Kalman filter, Extended Kalman filter, Artificial Neural Networks, Networks RBF.

# Capítulo 1

# Introdução

## 1.1. Motivação

A filtragem na sua forma mais geral tem estado presente na vida do homem a muito tempo. Inicialmente a filtragem surgiu com o objetivo de separar componentes de misturas, ou seja, a partir de misturas contendo componentes indesejáveis o homem sentiu necessidade de separar estes componentes para assegurar a sua própria sobrevivência.

Com a evolução do homem, o surgimento de novas tecnologias (surgimento da eletricidade e sua evolução) e o desenvolvimento da computação, as

técnicas de filtragem passaram a ser utilizadas largamente na engenharia para a filtragem (separação) de sinais elétricos. Normalmente, os sistemas de comunicação (telefonia móvel e fixa, sinais recebidos de satélites e outros sistemas) contêm sinais indesejáveis responsáveis pela degradação do sinal original. Dessa forma, vários cientistas começaram a pesquisar técnicas e aperfeiçoar as já existentes para a redução de ruídos em sinais das diversas áreas como citado acima.

Várias técnicas foram desenvolvidas, dentre elas: filtros de freqüência seletiva (filtro de Butterworth, Chebyshev, Elíptico), filtros ótimos (filtros de Wiener e de Kalman) e recentemente, a aplicação de redes neurais artificiais configuradas como filtro para a redução de ruídos. Dentre essas técnicas, tanto a filtragem de Kalman quanto as redes neurais artificiais têm produzido resultados eficientes e satisfatórios quanto a qualidade e a inteligibilidade do sinal resultante.

## 1.2. Objetivos

Com base nesse contexto e principalmente no grande desenvolvimento de sistemas baseados em sinais de voz (compressão, análise e síntese), este trabalho apresenta um estudo e a implementação de uma técnica baseada no filtro de Kalman e em redes neurais artificiais para a redução de ruídos em sinais de voz.

O algoritmo estudado foi desenvolvido pelos pesquisadores americanos WAN & NELSON (1998b), onde dois filtros de Kalman juntamente com uma rede neural são empregados para a filtragem de sinais de voz. No trabalho proposto, será estudado o emprego das redes RBF (*"Radial Basis Function"*) com funções de ativação do tipo gaussiana a este algoritmo. Como as redes RBF juntamente com as

redes MLP (*"Multilayer Perceptrons"*) são consideradas aproximadores universais de funções, teoricamente são redes equivalentes PARK & SANDBERG (1991). Os dois modelos apresentam várias diferenças, cada um com suas vantagens e limitações. Uma das principais vantagens das redes RBF sobre as redes MLP é o tempo de treinamento da rede, que por sua vez é mais rápido.

Para o desenvolvimento e a implementação do algoritmo as seguintes etapas foram realizadas:

1. Estudo do processo de formação da voz, sua propagação e a teoria de predição linear para sinais de voz;
2. Estudo das principais técnicas e algoritmos para filtragem de sinais de voz;
3. Estudo detalhado do filtro de Kalman e sua aplicação em sistemas lineares e não-lineares;
4. Estudo das redes neurais artificiais e suas aplicações em processamento digital de sinais de voz;
5. Implementação do filtro de Kalman e de redes neurais artificiais.

## 1.3. Organização do trabalho

Esta dissertação está dividida em 7 capítulos:

1. Introdução;
2. Modelo Fisiológico de Produção de Voz e Predição Linear;
3. Enriquecimento de Sinais de Voz;
4. Filtro de Kalman;

5. Redes Neurais Artificiais;
6. Filtro Duplo de Kalman Estendido;
7. Conclusões.

No capítulo 2, são apresentados os principais tópicos relacionados ao processamento digital de sinais de voz tais como: modelo fisiológico de produção, representação (formas de onda para os sons sonoros e surdos) e o modelo de predição linear, utilizado para representar o sinal quando da utilização do filtro de Kalman.

O capítulo 3 é responsável pela descrição de duas técnicas de filtragem muito empregadas em processamento digital de sinais de voz: a subtração espectral de potência e o filtro de Wiener.

No capítulo 4, é apresentada a teoria do filtro de Kalman, sua utilização, suas equações, propriedades, características e problemas que possam ser encontrados quando o filtro é utilizado.

O capítulo 5 traz a teoria das Redes Neurais Artificiais. Neste capítulo é apresentada a definição de redes neurais, suas semelhanças com o cérebro humano e as principais técnicas de aprendizagem. Em seguida é apresentado um resumo sobre as redes com múltiplas camadas, com o seu algoritmo de retropropagação e as redes RBF que serão utilizadas neste trabalho.

A descrição do algoritmo implementado é apresentada no capítulo 6. Este capítulo também apresenta os resultados da avaliação do algoritmo proposto. Os testes foram realizados com dois tipos de ruídos: ruído branco gaussiano e ruído branco não estacionário. Em um terceiro teste foi realizado uma pré-filtragem no

sinal para auxiliar na redução do ruído antes de se aplicar o filtro de Kalman e melhorar as estimativas iniciais do sinal.

Finalmente, o Capítulo 7 reúne as conclusões do trabalho desenvolvido e algumas idéias que foram surgindo ao longo do desenvolvimento do projeto e que serão propostas para trabalhos futuros.

# Capítulo 2

# Modelo Fisiológico de Produção de Voz e Predição Linear

## 2.1. Introdução

Este capítulo tem por objetivo apresentar o modelo fisiológico para a produção de sinais de voz, bem como, o modelo matemático que representa a forma de onda desses sinais.

Para aplicações em sistemas para voz, tais como: análise e síntese, enriquecimento para aplicações em sistemas de telefonia celular, reconhecimento e outras aplicações, é necessário entender o processo de formação da voz e técnicas de processamento digital de sinais. O capítulo inicia-se com uma breve introdução à

teoria acústica, lembrando que, o objetivo deste capítulo não é desenvolver um estudo da teoria acústica e sim apresentar uma breve noção sobre a produção de sinais de voz bem como o seu modelo matemático.

O capítulo se encerra com a teoria da predição linear aplicada a sinais de voz para a sua representação quando da utilização do filtro de Kalman. Neste tópico são abordados dois métodos para a determinação dos coeficientes de predição para modelos de sinais de voz: o *método da autocorrelação* e o *método da covariância*.

## 2.2. Modelo Fisiológico de Produção de Voz

Os sinais de voz são formados por uma seqüência de sons que por sua vez representam a informação transmitida RABINER & SCHAFER (1978). Cada língua apresenta regras diferentes para o arranjo desses sons. O estudo dessas regras de linguagem é denominado lingüística e o estudo dos sons é chamado fonética.

A figura 2.1 mostra o esquema do aparelho fonador humano.

A voz é uma onda de pressão acústica que se origina a partir de movimentos voluntários dos órgãos vocais humanos DELLER et al. (1993). O aparelho vocal humano pode ser considerado um tubo acústico, tendo um comprimento aproximado de 17 cm. De acordo com a figura 2.1, o aparelho vocal inicia-se nas pregas vocais, onde o ar vindo dos pulmões é modificado pelo trato vocal, e termina nos lábios por onde o ar é irradiado. Outro componente muito importante durante o processo de fonação é o véu palatino. Este componente desempenha papel importante durante a geração de sons nasalados. Para a geração destes sons o véu palatino é baixado e o trato nasal é acoplado acusticamente ao trato

vocal gerando sons nasais. Da mesma forma, durante o processo de sons não nasais, a cavidade nasal é bloqueada pela ação do véu palatino de modo que nenhum som seja irradiado pelo nariz.

1. cavidade nasal
2. cavidade bucal
3. lábios
4. língua
5. palato duro
6. véu palatino
7. palato mole
8. laringe
9. esôfago
10. traquéia
11. pregas vocais

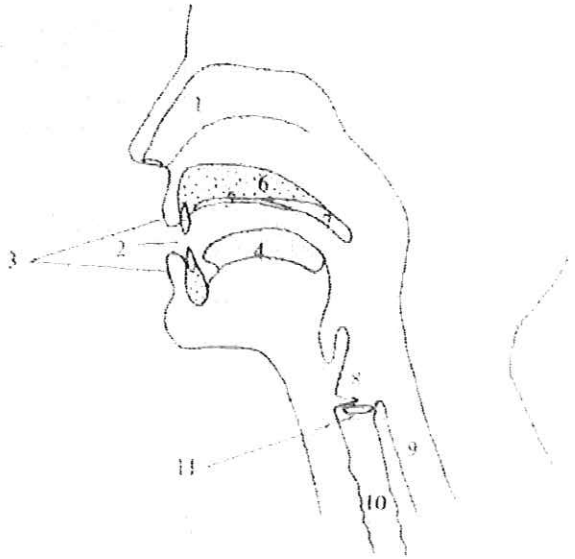


FIGURA 2.1 – Aparelho fonador humano.

A figura 2.2 mostra o diagrama em blocos do processo de produção da voz humana.

No processo de fonação, o ar armazenado nos pulmões é expelido passando pelas pregas vocais, que por sua vez vibram, constituindo dessa forma a fonte de excitação para a produção de sons. No trato vocal este fluxo de ar sofre influência das estruturas ressonantes, produzindo deste modo diferentes sons da fala.

Estes sons gerados pelo sistema vocal, em resposta à excitação das pregas vocais, podem ser classificados em três classes de acordo com o modo de

excitação: *sons sonoros*, *fricativos* e *plosivos* RABINER & SCHAFER (1978) PARSONS (1987).

Os *sons sonoros* são resultantes da passagem do ar vinda dos pulmões e forçados na glote (abertura entre as pregas vocais). Durante a passagem do ar, as pregas vocais vibram (abrem e fecham) de forma oscilatória produzindo pulsos quase regulares responsáveis pela excitação do trato vocal, que por sua vez produz sons periódicos. A taxa de vibração das pregas vocais depende de um conjunto de características do locutor, tais como: sexo, idade, pressão do ar nos pulmões e também do comprimento, espessura e da tensão das pregas vocais JOAQUIM (1988). Quanto maior a tensão nas pregas vocais, maior será a taxa de vibração, que por sua vez é definida como a freqüência fundamental da voz. Exemplos de sons sonoros: [a], [i], [e].

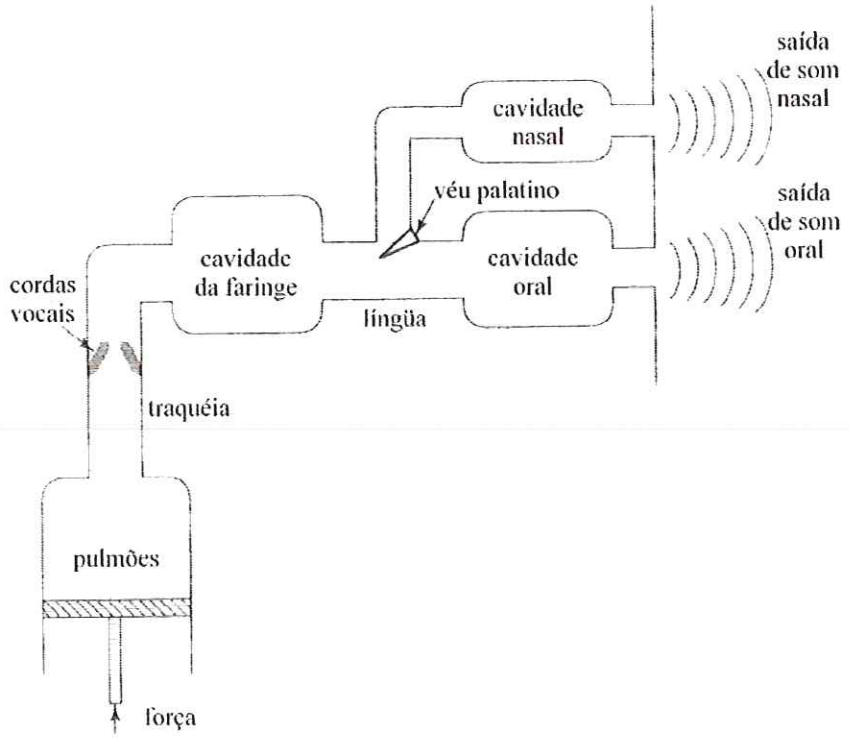



FIGURA 2.2 – Diagrama em blocos do processo de produção da voz DELLER et al. (1993).

Na produção de *sons fricativos*, as pregas vocais mantém-se abertas e em algum ponto do trato vocal forma-se uma constrição. Em seguida, uma quantidade de ar forçada através desta constrição produz uma "turbulência". Esta turbulência por sua vez constitui uma fonte de excitação para o trato vocal, que em resposta a esta excitação, gera sons fricativos. A forma de onda dos sons fricativos é semelhante a um ruído branco. Os sons fricativos podem ser divididos em dois grupos: *fricativos sonoros*, onde ocorre a vibração das pregas vocais e *fricativos surdos*, onde não ocorre a vibração das pregas vocais. Exemplos de sons fricativos sonoros: [z], [v] e fricativos surdos: [s], [f].

Os *sons plosivos* são resultantes de um fechamento completo da região do trato vocal na direção dos lábios. A pressão do ar é aumentada no ponto do fechamento e subitamente liberada. Estes sons são caracterizados por um instante de silêncio seguidos de uma "explosão". Exemplos desses sons são: [p], [b], [t].

De acordo com o exposto até este ponto, a geração de sinais de voz pode ser composta por três partes, conforme indica a figura 2.3. As três partes do processo são: fonte de excitação, trato vocal e radiação.

A fonte de excitação é responsável por excitar (alimentar) o trato vocal que por sua vez atua como um filtro acústico, variante no tempo, que modela o sinal de voz através da ação de seus articuladores. Por último este sinal é irradiado pela boca ou nariz.

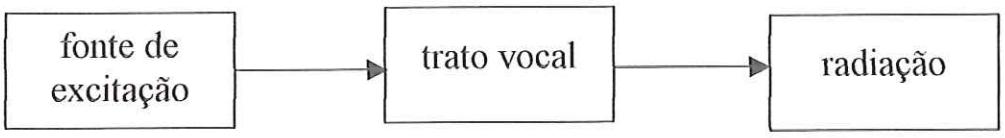



FIGURA 2.3 – Etapas do processo de produção de voz JOAQUIM (1988).

## 2.3. Representação de Sinais de Voz

A figura 2.4 apresenta a forma de onda de dois sons apresentados na seção anterior.

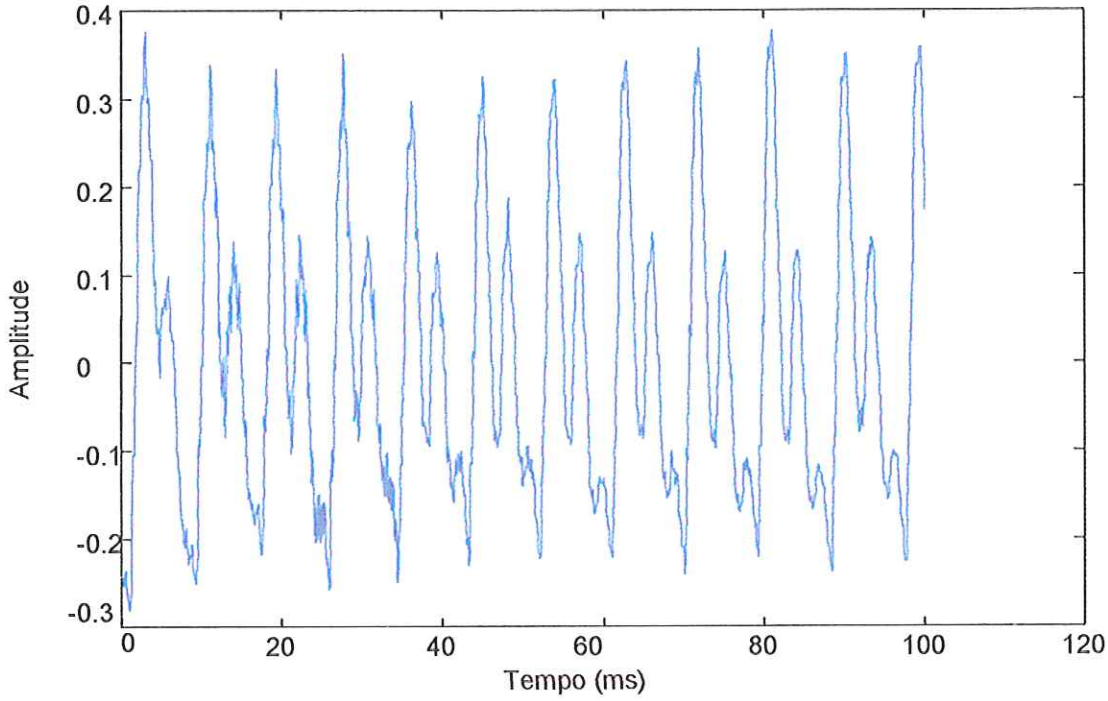
Analisando a figura 2.4, percebe-se claramente que a forma de onda pode ser separada em duas componentes: uma delas apresenta sinal semelhante a um ruído branco (sons surdos) enquanto que a outra componente apresenta sinais periódicos (sons sonoros). Para os sons surdos percebe-se uma variação aleatória da amplitude do sinal em relação ao tempo. Para os sons sonoros (periódicos), o intervalo de tempo de repetição é chamado período de tom ($T$) ou *pitch* e o seu inverso $\frac{1}{T}$ é chamado de freqüência fundamental RABINER & SCHAFER (1978).

As freqüências predominantes, que são ressaltadas pelas cavidades ressonantes na produção de sons sonoros, são chamadas de freqüências formantes, enquanto que a fonte de excitação é denominada de fonte primária.

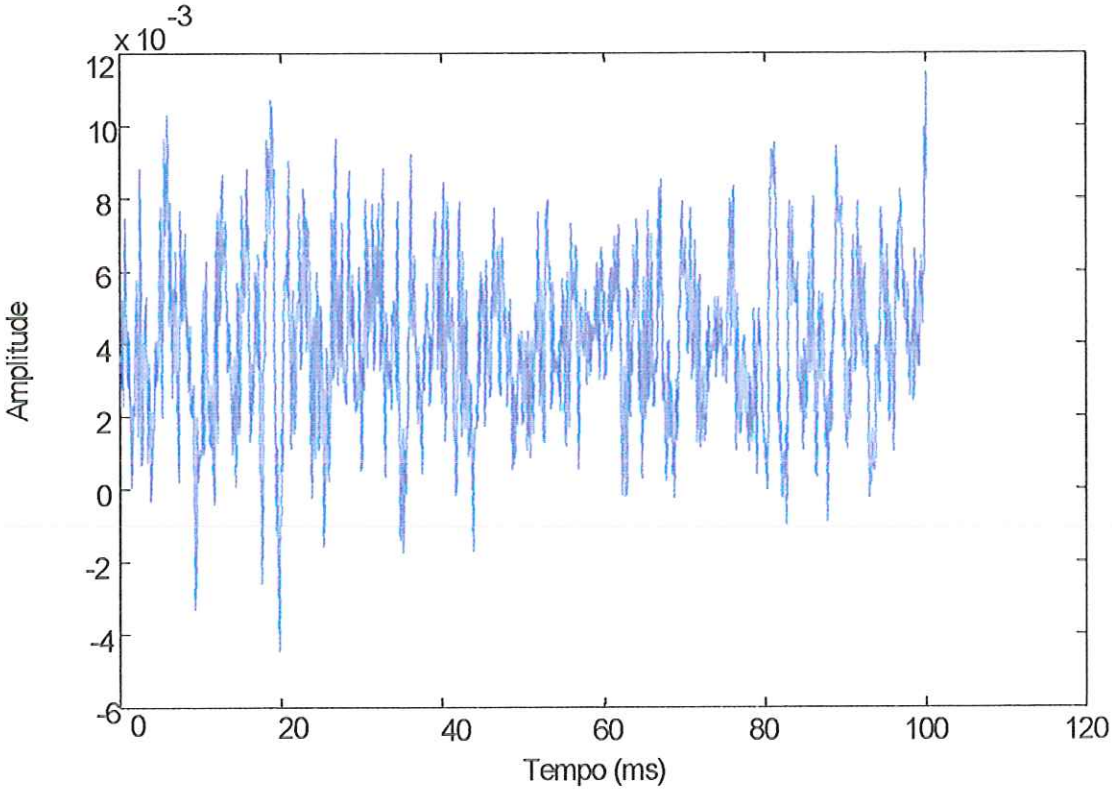
No espectro para sons surdos, a fonte primária de excitação é semelhante a um ruído branco que por sua vez gera um sinal cuja forma de onda é semelhante a um ruído branco (sem periodicidade).

## 2.4. Modelo de Produção de Voz

Durante muito tempo os cientistas visaram entender e reproduzir o processo de produção da voz, pois este representa o meio básico de comunicação. As primeiras tentativas de construir um modelo semelhante ao nosso trato vocal resultaram em máquinas mecânicas para a produção de voz que com o tempo e novas




(a)




(b)

FIGURA 2.4 – (a) Forma de onda para um som sonoro, (b) Forma de onda para um som surdo.

técnicas levaram a construção de sistemas elétricos. Atualmente os sistemas são baseados em modelos matemáticos DELLER et al. (1993).

O primeiro documento relatando esforços para conseguir reproduzir os sons da voz data de 1779 por C. G. Kratzenstein, na tentativa de produzir e explicar as diferenças entre as vogais. Desde então, vários resultados de pesquisas foram surgindo procurando reproduzir o trato vocal. Durante o século XIX, vários sistemas mecânicos foram apresentados com o objetivo de modelar e sintetizar sinais de voz. O século XX é caracterizado por modelos elétricos, principalmente pelo surgimento dos circuitos e sistemas baseados em computador.

O processo de formulação matemática da produção de voz é uma conseqüência da construção de aparelhos elétricos e mecânicos. Este processo matemático é importante pois fornece a base para aplicações em análise e síntese de sinais de voz, codificação, enriquecimento e reconhecimento. O sistema de produção da voz é um processo complexo descrito por um conjunto de equações diferenciais que descrevem a propagação do ar no trato vocal. Este por sua vez depende de um conjunto de características tais como DELLER et al. (1993):

- natureza e forma variante do sistema vocal;
- cavidade nasal;
- efeitos das paredes do tecido do trato vocal;
- efeitos provocados pela combinação da estrutura subglotal (pulmões e traquéia) com as estruturas ressonantes do trato vocal.

O modelo linear apresentado a seguir foi proposto por Fant na década de 60 FANT (1960). Todo o desenvolvimento das equações e tópicos da teoria acústica são mostrados em FANT (1960) e FLANAGAN (1972).

Como mostrado nas seções anteriores e na figura 2.3, o processo de produção de voz pode ser dividido em três partes. Dessa forma, existem três componentes separadas para modelar: fonte de excitação, trato vocal e radiação, como mostra a figura 2.5.

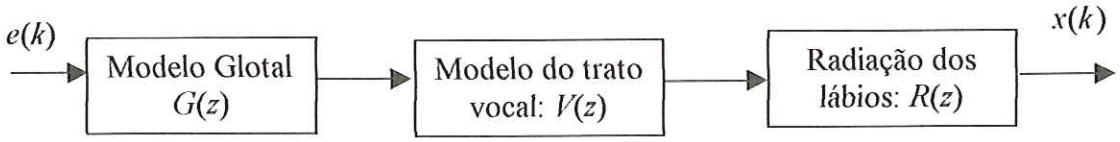



FIGURA 2.5 – Modelo linear de produção da voz FANT (1960), JOAQUIM (1988).

O sinal $e(k)$ apresentado na figura 2.5 é um sinal de excitação que depende das características do sinal de voz. Quando o sinal de voz apresenta características sonoras, o sinal de excitação $e(k)$ assume a forma de pulsos periódicos separados pelo período de tom. Quando o sinal de voz apresentar características não-sonoras, o sinal $e(k)$ assume a forma de um ruído.

O fluxo de ar que passa pelo trato vocal através da glote é modelado por um filtro passa-baixas com uma freqüência de corte de aproximadamente 100 Hz, com uma função de transferência dada por:

$$G(z) = \frac{1}{(1 - e^{-cT} z^{-1})^2} \tag{2.1}$$

sendo $G(z)$ a transformada $z$ do sinal $g(kT)$ amostrado a cada $T$ segundos.

O trato vocal pode ser modelado por um filtro somente com pólos com $n$ freqüências formantes $(f_n)$ que representam as ressonâncias do trato vocal. A equação do filtro somente com pólos é descrita por:

$$V(z) = \frac{1}{\prod_{i=1}^{n} \left\{ 1 - 2e^{-C_i T} \cos(b_i T) z^{-1} + e^{-2C_i T} z^{-2} \right\}} \tag{2.2}$$

a *n-ésima* freqüência pode ser calculada por: $f_n = \frac{b_n}{2\pi}$.

A terceira parte do modelo linear consiste no processo de radiação dos sons pelo lábio através do qual o sinal de voz se torna uma onda de pressão acústica. A equação que descreve o modelo de radiação dos lábios é apresentada abaixo:

$$L(z) = 1 - z^{-1} \tag{2.3}$$

Combinando os três modelos apresentados acima tem-se que:

$$G(z)V(z)L(z) = \frac{\left(1 - z^{-1}\right)}{(1 - e^{-cT} z^{-1})^2 \left[ \prod_{i=1}^{n} \left( 1 - 2e^{-C_i T} \cos(b_i T) z^{-1} + e^{-2C_i T} z^{-2} \right) \right]} \tag{2.4}$$

De acordo com FANT (1960), o numerador da equação pode ser cancelado com o termo $\left(1-e^{cT}z^{-1}\right)^2$ pois $cT$ é geralmente muito menor que 1, fazendo com que o modelo se transforme em um filtro somente com pólos com uma função de transferência dada pela equação:

$$G(z)V(z)L(z)=\frac{1}{(1-e^{-cT}z^{-1})\left[\prod_{i=1}^{n}\left(1-2e^{-c_iT}\cos(b_iT)z^{-1}+e^{-2c_iT}z^{-2}\right)\right]} \tag{2.5}$$

O modelo linear de produção da voz mostrado na figura 2.5 e descrito pela função de transferência dada pela eq. (2.5) pode ser representado matematicamente por:

$$X(z)=\frac{1}{A(z)}E(z) \tag{2.6}$$

sendo que $\frac{1}{A(z)}$ é um filtro linear somente com pólos, variante no tempo definido pela eq. (2.7):

$$\frac{1}{A(z)}=\frac{1}{\sum_{i=0}^{M}a_iz^{-1}}=G(z)V(z)L(z) \tag{2.7}$$

sendo $a_0 = 1$. O filtro $\frac{1}{A(z)}$ representa o comportamento espectral da voz.

Resumindo, tem-se que o sinal de voz pode ser representado por uma única função de transferência somente com pólos que combina os modelos glotal, do trato vocal e da radiação pelos lábios.

## 2.5. Predição Linear de Sinais de Voz

O modelo de predição linear representa uma poderosa ferramenta matemática para a análise de sinais de voz. De acordo com este modelo a amostra de um sinal pode ser expressa em termos de suas amostras passadas, ou seja, como uma combinação linear de seus valores passados. O sinal é modelado por um filtro somente com pólos, denominado filtro preditor linear, cujos coeficientes são obtidos de modo a minimizar o erro quadrático médio entre as amostras originais e as amostras preditas.

A predição linear e suas aplicações têm sido reportadas em vários trabalhos desde 1968 e utilizadas principalmente em engenharia, nas áreas de controle, teoria da informação, identificação de sistemas e estimação de sinais. A sua ampla aplicabilidade deve-se principalmente a sua capacidade para apresentar estimativas precisas utilizando representação do sinal com baixa taxa de bits, sendo ideal para o armazenamento da informação. Outra característica importante desse método é a simplicidade, velocidade e precisão dos cálculos, fazendo com que este método se torne uma das principais técnicas de análise e síntese de sinais de voz JOAQUIM (1988).

A figura 2.6 representa o modelo linear para a produção de sinais de voz descrito nas seções anteriores. De acordo com o diagrama em blocos da figura 2.6, o sistema é excitado por um trem de impulsos para a geração de sinais de voz sonoros e por uma seqüência de ruído branco para a produção de sons surdos.

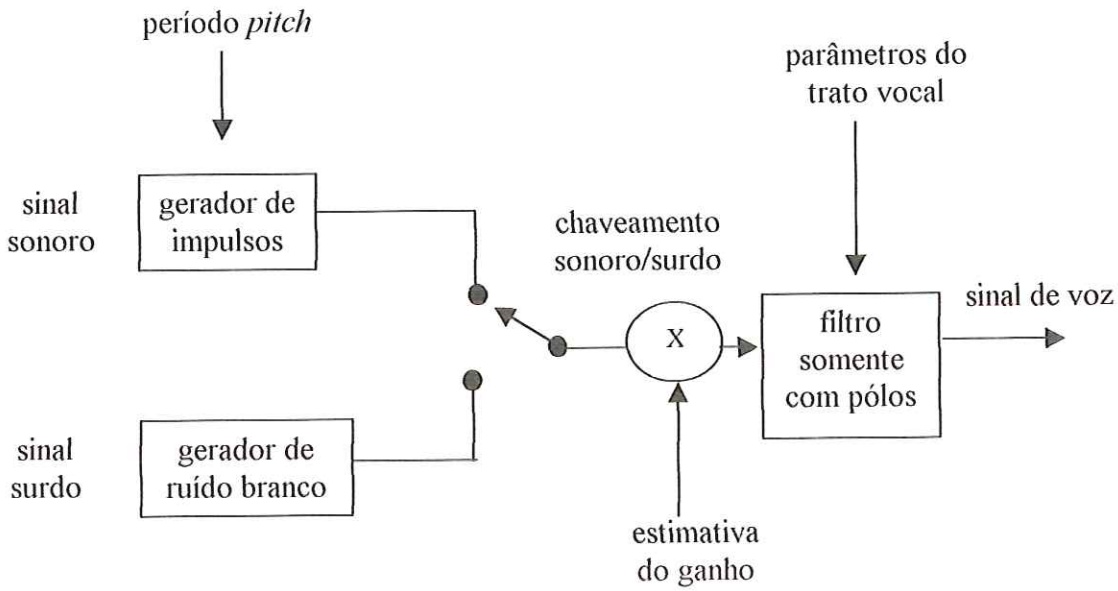



FIGURA 2.6 – Modelo geral discreto no tempo para produção de sinais de voz.

Como visto em seções anteriores, o modelo de produção linear de sinais de voz é constituído basicamente por três partes combinadas: modelo glotal, modelo do trato vocal e radiação pelos lábios, que são agrupados em um filtro digital recursivo, cujos coeficientes são facilmente obtidos através do método de predição linear. O diagrama em blocos da figura 2.7 mostra o esquema de um filtro digital somente com pólos utilizado na produção de sinais de voz.

A função de transferência para o filtro preditor é dada pela seguinte equação:

$$1 - A(z) = -\sum_{i=1}^{M} a_i z^{-i} \tag{2.8}$$

sendo $a_0 = 1$, $M$ a ordem do filtro (número de coeficientes do filtro preditor linear) e $z^{-1}$ representa a unidade de atraso. A escolha da ordem do filtro depende de um conjunto de características, sendo as principais: a natureza do som e ressonância do trato vocal. Um valor estipulado para $M$ deve estar compreendido entre 8 e 12 JOAQUIM (1988).

As equações que definem o modelo do filtro digital somente com pólos são dadas por:

$$\begin{cases} X(z) = \dfrac{1}{A(z)} E(z) & (2.9) \\ E(z) = A(z) X(z) & (2.10) \end{cases}$$

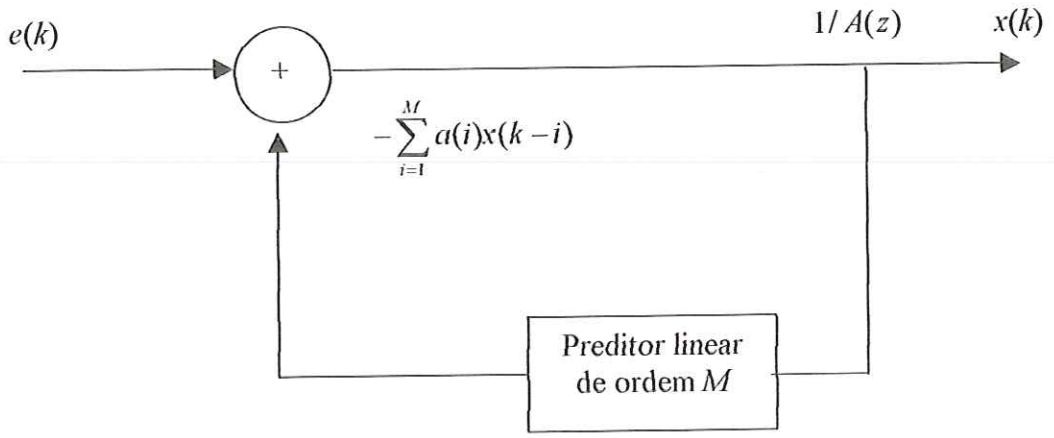



FIGURA 2.7 – Preditor linear de ordem $M$.

A eq. (2.9) é a equação de síntese, enquanto que a eq. (2.10) é a equação de análise. $X(z)$ representa a transformada $z$ do sinal de voz original $x(k)$. A função de transferência $E(z)$ é a transformada $z$ da função de excitação do filtro digital. Além do modelo auto-regressivo (AR) empregado neste trabalho, outros modelos podem ser empregados, dentre eles: auto-regressivo média móvel (ARMA) e o média móvel (MA).

A função de excitação, dependendo do som, pode assumir a forma de impulsos periódicos para a produção de sons sonoros ou ruído branco para a produção de sons surdos. A função de transferência $\frac{1}{A(z)}$ define o filtro somente com pólos e $A(z)$ é um filtro FIR (somente com zeros) representado pela seguinte equação:

$$A(z) = 1 + \sum_{i=1}^{M} a_i z^{-i} \tag{2.11}$$

Substituindo a eq. (2.11) em (2.10) tem-se que:

$$E(z) = \left(1 + \sum_{i=1}^{M} a_i z^{-i}\right) X(z) \tag{2.12}$$

Tomando-se a transformada $z$ inversa da equação acima tem-se que:

$$e(k) = x(k) + \sum_{i=1}^{M} a_i x(k-i) \tag{2.13}$$

Pela análise da eq. (2.13), observa-se que o preditor linear é formado por uma combinação linear dos valores atuais do sinal de voz $x(k)$ e valores passados da amostra $x(k-i)$. A amostra predita do sinal de voz $\hat{x}(k)$, de acordo com o preditor linear de ordem $M$ mostrado na figura 2.7, é dada por:

$$\hat{x}(k) = -\sum_{i=1}^{M} a_i x(k-i) \tag{2.14}$$

O modelo de predição linear descrito pela eq. (2.14) será utilizado para representar o sinal quando da utilização do filtro de Kalman. O filtro utiliza espaços de estados e o modelo de predição utilizado para representar o sinal é dado por:

$$\hat{x}(k) = A(k-1)x(k-1) \tag{2.15}$$

sendo $A$ a matriz de transição de estados representada por ANDERSON & MOORE (1979):

$$A = \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_{M-1} & -a_M \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{2.16}$$

sendo $a_i$ os coeficientes de predição linear determinados de modo a minimizar o sinal de erro conforme eq. (2.13), segundo um determinado critério estabelecido. O método mais comum e o mais utilizado em processamento digital de sinais é o da minimização da soma dos quadrados das amostras de $e(k)$, resultando em equações lineares facilmente tratáveis.

O sinal de excitação $e(k)$ do preditor linear pode ser interpretado como o erro de predição (ou resíduo) entre a amostra $x(k)$ e o sinal de voz predito definido pela eq. (2.14).

$$e(k) = x(k) - \hat{x}(k) \tag{2.17}$$

A expressão para o erro quadrático médio entre o sinal de voz $x(k)$ e o sinal de voz predito $\hat{x}(k)$ é dada por:

$$\varepsilon = \sum_{k=n_0}^{n_1} e^2(k) \tag{2.18}$$

sendo $n_0$ e $n_1$ os limites do somatório.

Substituindo a eq. (2.13) na eq. (2.18) tem-se que:

$$\varepsilon = \sum_{k=n_0}^{n_1} \left\{ \sum_{i=0}^{M} a_i x(k-i) \right\}^2 \tag{2.19}$$

Desenvolvendo a eq. (2.19) obtém-se a seguinte expressão para o erro médio quadrático:

$$\varepsilon = \sum_{k=n_0}^{n_1} \sum_{i=0}^{M} \sum_{j=0}^{M} a_i a_j x(k-i)x(k-j) \tag{2.20}$$

Rearranjando os termos da eq. (2.20) e invertendo a ordem dos somatórios tem-se:

$$\varepsilon = \sum_{i=0}^{M} \sum_{j=0}^{M} a_i a_j \sum_{k=n_0}^{n_1} x(k-i)x(k-j) \tag{2.21}$$

Na eq. (2.21) o termo: $\sum_{k=n_0}^{n_1} x(k-i)x(k-j)$ representa a função de autocorrelação do sinal de voz $x(k)$ e é representado pelo termo $r_{ij}$. Dessa forma, pode-se reescrever a eq. (2.21) da seguinte forma:

$$\varepsilon = \sum_{i=0}^{M} \sum_{j=0}^{M} a_i a_j r_{ij} \tag{2.22}$$

Para minimizar o erro quadrático médio da eq. (2.22), calculam-se as derivadas parciais do erro em relação aos coeficientes $a_j$ e iguala-se o resultado a zero:

$$\frac{\partial \varepsilon}{\partial a_j} = \frac{\partial}{\partial a_j}\left\{\sum_{k=n_0}^{n_1}\left[\sum_{i=0}^{M} a_i x(k-i)\right]^2\right\} = 0$$

$$= 2\sum_{i=0}^{M} a_i \sum_{k=n_0}^{n_1} x(k-i)x(k-j) = 0$$

$$\sum_{i=0}^{M} a_i r_{ij} = 0 \qquad j = 1, 2, 3, \ldots, M \qquad (2.23)$$

A eq. (2.23) pode ser escrita da seguinte forma:

$$\sum_{i=1}^{M} a_i r_{ij} = -r_{0j}\ ; \qquad j = 1,2,3,\ldots,M \qquad (2.24)$$

Resolvendo o conjunto de $M$ equações lineares dado pela eq. (2.24) determina-se o conjunto de coeficientes $(a_i)$ do filtro preditor linear. Esta equação também pode ser representada na forma matricial:

$$\mathbf{R_x a} = \mathbf{r_j} \qquad (2.25)$$

ou

$$\begin{bmatrix} r_{11} & r_{21} & r_{31} & \cdots & r_{M1} \\ r_{12} & r_{22} & r_{32} & \cdots & r_{M2} \\ r_{13} & r_{23} & r_{33} & \cdots & r_{M3} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{1M} & r_{2M} & r_{3M} & \cdots & r_{MM} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_M \end{bmatrix} = -\begin{bmatrix} r_{01} \\ r_{02} \\ r_{03} \\ \vdots \\ r_{0M} \end{bmatrix} \tag{2.26}$$

sendo $\mathbf{R}_x$ a matriz de autocorrelação do sinal. $\mathbf{R}_x$ é uma matriz quadrada de ordem $M$, hermetiana e Toeplitz. Os coeficientes $a_i$ são dispostos em um vetor e representam os coeficientes do filtro digital. Resolvendo o conjunto de $M$ equações lineares apresentado acima determinam-se os $M$ coeficientes $a_i$ do filtro preditor, lembrando que, na solução das equações são necessárias as amostras do sinal de voz num dado intervalo para a determinação da matriz de autocorrelação.

Tendo as amostras do sinal de voz, pode-se definir inicialmente dois métodos para a determinação dos coeficientes do filtro preditor linear: *método da autocorrelação* e *método de covariância*.

## 2.6. Método da Autocorrelação

Neste método supõem-se que o sinal $x(k)$ é conhecido apenas dentro de um intervalo finito de tamanho $N$ $[0; N-1]$ e que é aproximado por um modelo somente com pólos, como descrito nas seções anteriores. Os coeficientes $a_i$ são determinados para minimizar o erro quadrático médio.

É assumido que o valor do sinal é $x(k) = 0$ para $k < 0$ ou $k > N-1$. Este método é conhecido como *método da autocorrelação* e um novo sinal $\tilde{x}(k)$ é gerado

pela aplicação de uma janela de análise ao sinal $x(k)$. Estas janelas de análise podem ser do tipo Hamming, Hanning, Bartlett, Blackman, Kaiser e retangular. Cada um dos tipos de janelas apresentam características diferentes para manipular os dados a serem analisados.

$$\tilde{x}(k) = x(k) w_R(k) \tag{2.27}$$

sendo $w_R(k)$ uma janela de análise como especificado acima.

A aplicação de uma janela de análise assegura que o modelo será estável, ou seja, os pólos da função de transferência estarão dentro do círculo de raio unitário e a estrutura Toeplitz da eq. (2.24) também será assegurada HAYES (1996).

Dessa forma, o cálculo da função de autocorrelação na eq. (2.19), dentro do intervalo de comprimento $N$ $[0; N-1]$, é realizado utilizando-se o sinal $\tilde{x}(k)$ ao invés de $x(k)$:

$$r_{ij} = \sum_{m=k}^{N-1} \tilde{x}(m)\tilde{x}(m-k) \tag{2.28}$$

As equações para o modelo somente com pólos usando o método da autocorrelação estão resumidas na tabela 2.1. Para a resolução do conjunto de $M$ equações lineares é utilizada a recursão de Levinson-Durbin.

TABELA 2.1 – Modelo somente com pólos usando o método da autocorrelação.

| | |
|---|---|
| **Equações normais** | $\sum_{i=1}^{M} a_i r_{ij} = -r_{0j}$ , para $j = 1,2,3,\ldots,M$ <br> $r_{0j} = \sum_{m=k}^{N-1} \tilde{x}(m)\tilde{x}(m-k)$ |
| **Erro mínimo** | $\varepsilon = r_{ij}(0) + \sum_{i=1}^{M} a_i r_{0j}$ |

## 2.7. Método da Covariância

Como apresentado na seção anterior, o método da autocorrelação considera um intervalo finito $[0, N-1]$ em um sinal e força o sinal a zero fora deste intervalo aplicando uma janela de análise.

Nesta seção é apresentado outro método, conhecido como *método da covariância*. Neste método, nenhuma restrição é considerada para os dados fora do intervalo especificado e nenhuma janela de análise é aplicada ao sinal. Para este método, a estrutura Toeplitz da equação normal é perdida HAYES (1996). Como para o método da autocorrelação, este método é utilizado para determinar os coeficientes $a_i$ do filtro preditor linear que minimizam a soma dos quadrados do erro $e(k)$.

Considerando o mesmo intervalo do sinal apresentado na seção anterior $[0, N-1]$, o erro quadrático pode ser minimizado para os valores de $k$ no intervalo $[p, N]$. A expressão para o erro é dada por:

$$\varepsilon^{C} = \sum_{k=0}^{N-1} |e(k)|^2 \tag{2.28}$$

Dessa forma, não é necessário aplicar janelas nos dados ou fazer qualquer restrição para os dados fora do intervalo especificado. O cálculo dos coeficientes do filtro somente com pólos é dado pela seguinte equação na forma matricial:

$$\begin{bmatrix} c(1,1) & c(1,2) & c(1,3) & \cdots & c(1,M) \\ c(2,1) & c(2,2) & c(2,3) & \cdots & c(2,M) \\ c(3,1) & c(3,2) & c(3,3) & \cdots & c(3,M) \\ \vdots & \vdots & \vdots & & \vdots \\ c(M,1) & c(M,2) & c(M,3) & \cdots & c(M,M) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_M \end{bmatrix} = \begin{bmatrix} c(1,0) \\ c(2,0) \\ c(3,0) \\ \vdots \\ c(M,0) \end{bmatrix} \tag{2.29}$$

e o erro mínimo é dado por:

$$\{\varepsilon^{C}\}_{\min} = c_x(0,0) + \sum_{k=1}^{M} a(k) c_x(0,j) \qquad j = 1,2,3,\ldots,M \tag{2.29}$$

A tabela 2.2 mostra de forma resumida as equações para a modelagem de um sinal, usando um modelo somente com pólos, pelo método da covariância.

Como citado na seção anterior, a recursão de Levinson-Durbin é utilizada para calcular os coeficientes do modelo somente com pólos. A determinação dos

coeficientes utilizando o método da covariância pode ser realizada pela decomposição de Cholesky.

TABELA 2.2 – Modelo somente com pólos usando o método da covariância.

| | |
|---|---|
| **Equações normais** | $\sum_{m=1}^{M} a(m)c(i,j) = c(\mathrm{m},0)$, para $m = 1,2,3,\ldots,M$ <br> $c(i,j) = \sum_{k=0}^{M-1} x(k-\mathrm{i})x(k-j), \quad k,i \geq 0$ |
| **Erro mínimo** | $\varepsilon^{C} = c(0) + \sum_{k=1}^{M} a(k)c(0,j)$, para $j = 1,2,3,\ldots,M$ |

## 2.8. Considerações Finais

Neste capítulo foi apresentado de forma resumida o processo de formação de sinais de voz, bem como, a sua representação.

A teoria da predição linear também foi abordada. Esta representa uma ferramenta indispensável para a análise de sinais. Também foi apresentado o modelo de predição linear comumente empregado na literatura e que será o modelo adotado neste trabalho para a representação de sinais para aplicações com o filtro de Kalman.

O capítulo se encerra com dois métodos clássicos utilizados para determinar os coeficientes do preditor linear: *método da autocorrelação* e *método da covariância*.

# Capítulo 3

# Enriquecimento de Sinais de Voz

## 3.1. Introdução

Em muitas aplicações de processamento de sinais de voz, a presença de ruído aditivo pode degradar a qualidade e a inteligibilidade dos sinais. Basicamente, no processo de comunicação entre dois sistemas e/ou duas pessoas, dependendo do ambiente, pode haver uma perda na qualidade da informação transmitida. A qualidade de um sinal de voz pode ser degradada na conversão da informação (microfone), na transmissão (canais de comunicação) ou na reprodução (alto-falante).

Com base nestas informações, muitas pesquisas têm sido realizadas nessa área e conseqüentemente muitos algoritmos de enriquecimento de sinais de voz têm

surgido com o objetivo de eliminar ou mesmo reduzir a intensidade do ruído aditivo, melhorar a qualidade do sinal transmitido e até mesmo eliminar interferências dos meios de comunicação DELLER et al. (1993). No processo de redução de ruídos, a relação sinal ruído (SNR – *signal noise rate*) deve ser levada em consideração, ou seja, deverá haver uma melhora no sinal de voz filtrado em relação ao sinal contaminado.

A figura 3.1 mostra de forma geral as possíveis fontes de distorção de um sinal de voz no processo de comunicação e as aplicações onde um sinal enriquecido (livre de ruído) faz-se necessário.

De acordo com a figura 3.1, alguns algoritmos de enriquecimento de sinais de voz visam a redução de ruídos introduzidos na fonte (origem do sinal). Na fonte, a distorção pode ser causada tanto por ruídos aditivos (ruídos de fundo) quanto por ruídos provocados pelos próprios locutores. Além das distorções na fonte, o sinal de voz também pode ser corrompido durante o processo de codificação.

Em ambos os casos, os algoritmos de enriquecimento de sinais são aplicados aos sinais corrompidos melhorando dessa forma sua qualidade. A figura 3.1 também destaca algumas das várias aplicações em que um sinal de voz livre de ruído pode ser empregado. Uma das situações é quando o sinal percorre um canal de comunicação, podendo ser corrompido tanto na fonte quanto no momento da transmissão, e este sinal é aplicado a um algoritmo de enriquecimento para que quando o sinal atinja seu alvo (ouvido humano) possa ser inteligível. Outra aplicação destacada é quando um sinal emitido por uma fonte, e conseqüentemente corrompido nesta por um ruído aditivo, atinja seu alvo (ouvido humano) sem ser transmitido por um canal de comunicação. Neste caso, um algoritmo de enriquecimento também

pode ser aplicado ao sinal. Outra aplicação muito empregada é nos sistemas de reconhecimento de sinais de voz onde algoritmos de enriquecimento também podem ser aplicados aos sinais antes mesmo de atingir os sistemas de reconhecimento.

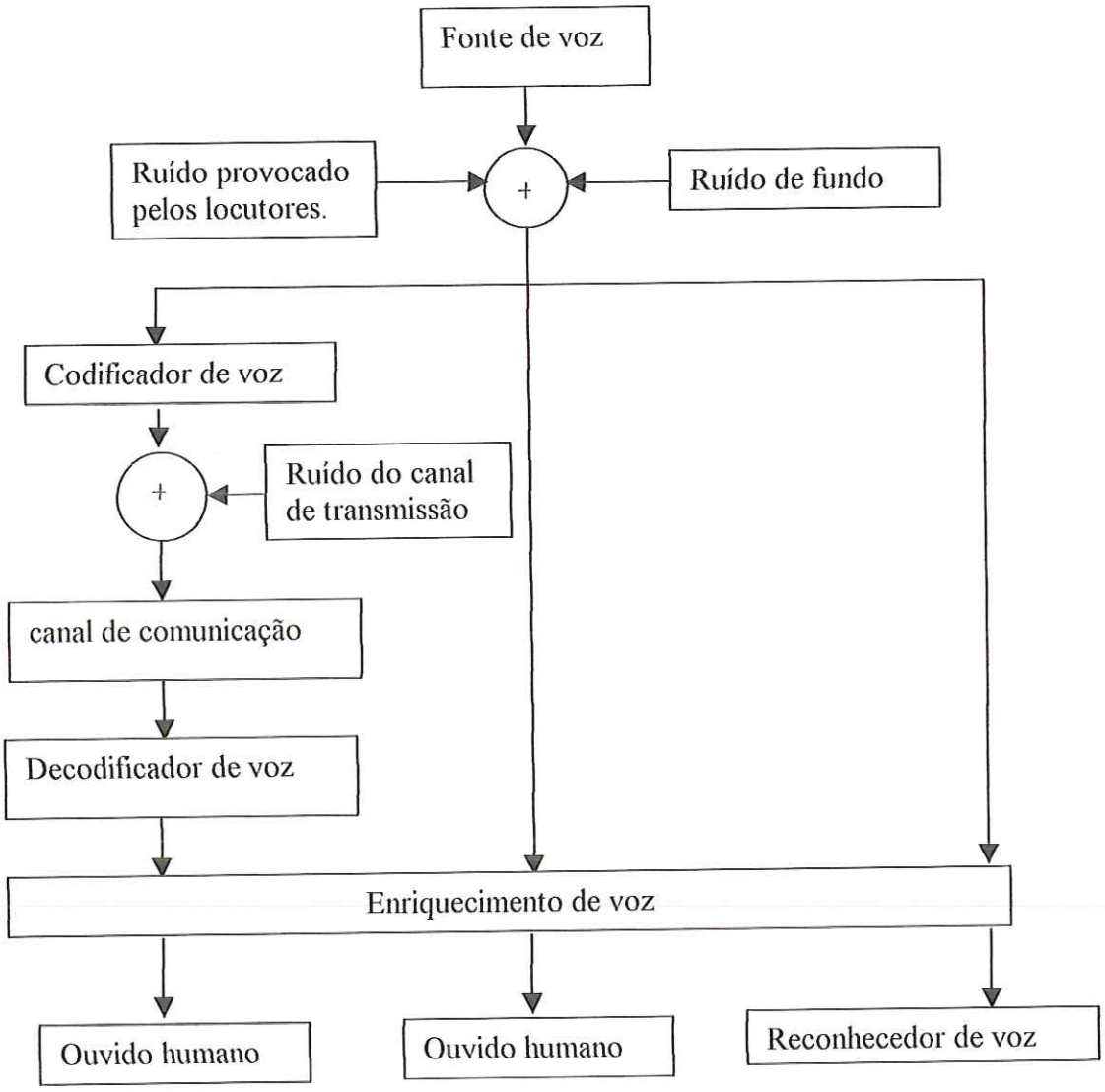



FIGURA 3.1 – Fontes de degradação empregadas no enriquecimento de sinais de voz.

## 3.2. Classificação dos Métodos de Enriquecimento de Sinais de Voz

Segundo DELLER et al. (1993), existem várias classificações para os métodos de enriquecimento de sinais de voz. Uma das maneiras diz respeito ao modo pelo qual o sinal é modelado. Algumas técnicas são baseadas em modelos estocásticos, onde um critério matemático é necessário, enquanto que outros métodos são baseados em aspectos perceptuais, onde aspectos importantes da percepção humana são melhorados.

Existe também a classificação em relação ao canal (meio de transmissão). Pode-se ter aplicações para um canal simples e aplicações para um canal duplo (ou multicanal). Para aplicações envolvendo um canal simples considere a existência de um único microfone. Sinais de voz em telefone também são considerados canais simples.

Além dessas classificações apresentadas acima, existem outros métodos de enriquecimento. Essas classes apresentam um conjunto de hipóteses, vantagens e limitações. Dentre esses métodos destacam-se: técnicas espectrais a curto intervalo de tempo (subtração espectral), filtros ótimos (filtros de Wiener e Kalman) e aplicações de redes neurais artificiais.

Neste capítulo serão abordadas a subtração espectral de potência e a teoria do filtro de Wiener aplicada a filtragem de sinais de voz. O filtro de Kalman e as redes neurais artificiais serão abordados nos capítulos 4 e 5 respectivamente.

Para a aplicação dos métodos citados acima, algumas hipóteses serão consideradas:

- a distorção do sinal original é provocada por ruído aditivo;

- o sinal de voz e o ruído são descorrelacionados.

### 3.2.1. Subtração Espectral

A técnica de subtração espectral considera o enriquecimento de sinais de voz no domínio da freqüência. É uma técnica muito empregada por ser uma das primeiras técnicas aplicadas a sinais de voz e uma das mais fáceis para implementação em computador. Nesta técnica, o enriquecimento do sinal de voz corrompido é realizado através de uma subtração espectral entre o sinal de voz corrompido e a estimativa do ruído.

O sinal de voz nesta técnica é modelado como um processo aleatório, tendo ruído aditivo, que por sua vez pode ser considerado em curtos intervalos de tempo. As estatísticas de segunda ordem do ruído podem ser estimadas durante os instantes de silêncio do sinal.

O sinal de voz corrompido por ruído aditivo é dado pela eq. (3.1).

$$y(k) = x(k) + v(k) \tag{3.1}$$

sendo que $x(k)$ representa o sinal de voz original, $v(k)$ representa o ruído aditivo e $y(k)$ o sinal de voz corrompido por ruído.

Para a aplicação dos métodos de subtração espectral, a transformada de Fourier (TF) é aplicada na eq. (3.1) e obtém-se:

$$Y(e^{j\omega}) = X(e^{j\omega}) + V(e^{j\omega}) \tag{3.2}$$

sendo $Y(e^{j\omega})$ o espectro de amplitude para o processo $y(k)$. Similarmente $X(e^{j\omega})$ e $V(e^{j\omega})$ representam os espectros de amplitudes para os processos $x$ e $v$. Pela análise da eq. (3.2) nota-se que tendo o espectro de amplitude do sinal disponível $y(k)$ e tendo uma estimativa do espectro de amplitude do ruído $\hat{V}(e^{j\omega})$, determina-se a estimativa do espectro de amplitude do sinal livre de ruído $X(e^{j\omega})$, de acordo com a eq. (3.3):

$$\hat{X}(e^{j\omega}) = Y(e^{j\omega}) - \hat{V}(e^{j\omega}) \tag{3.3}$$

BOLL (1979) desenvolveu uma técnica baseada na subtração espectral. Nesta técnica, o espectro de amplitude do sinal de voz original é obtido subtraindo-se o espectro de amplitude do ruído do espectro de amplitude do sinal de voz ruidoso, como apresentado na eq. (3.3). No erro resultante da subtração, ou seja, na diferença entre o sinal de voz original e o sinal de voz resultante da subtração, são aplicados quatro métodos objetivando a redução do erro.

No algoritmo empregado por BOLL, o sinal de voz corrompido por ruído é dado pela eq. (3.1). Neste sinal é aplicado diretamente a transformada de Fourier (TF) e obtém-se o sinal dado pela eq. (3.2)

Para a determinação do filtro de subtração espectral $H(e^{j\omega})$, BOLL substituiu a magnitude do ruído $|V(e^{j\omega})|$ pelo seu valor médio $M(e^{j\omega})$, determinado durante atividades não-sonoras. A fase do ruído $\theta_v(e^{j\omega})$ também foi substituída pela fase $\theta_y(e^{j\omega})$ de $Y(e^{j\omega})$, resultando dessa forma no estimador de subtração espectral:

$$\hat{X}(e^{j\omega}) = \left[|Y(e^{j\omega})| - M(e^{j\omega})\right] e^{j\theta_y(e^{j\omega})} \tag{3.4}$$

ou

$$\hat{X}(e^{j\omega}) = H(e^{j\omega}) Y(e^{j\omega}) \tag{3.5}$$

sendo $H(e^{j\omega})$ definido pela seguinte relação $1 - \dfrac{M(e^{j\omega})}{|Y(e^{j\omega})|}$ que representa o filtro de subtração espectral. O valor $M(e^{j\omega})$ é o valor médio do ruído determinado durante os instantes de silêncio e $|X(e^{j\omega})|$ é a magnitude do espectro do sinal de voz original.

O erro espectral resultante do estimador espectral é dado pela eq. (3.6).

$$\varepsilon(e^{j\omega}) = \hat{X}(e^{j\omega}) - X(e^{j\omega}) \;=\; V(e^{j\omega}) - M(e^{j\omega}) e^{j\theta_y(e^{j\omega})} \tag{3.6}$$

BOLL desenvolveu algumas modificações que são empregadas para tentar reduzir o valor do erro resultante do estimador espectral. Dentre essas modificações pode-se destacar a magnitude média, redução do erro residual, retificação de meia onda e atenuação do sinal durante os trechos de silêncio.

- **Magnitude média:** de acordo com a eq. (3.6), nota-se que o erro espectral é a diferença entre o espectro do ruído e o seu valor médio. Tendo em vista a redução do erro de estimação, BOLL sugeriu substituir o valor da magnitude do sinal ruidoso $|Y(e^{j\omega})|$ na eq. (3.4) pelo seu valor médio $\overline{|Y(e^{j\omega})|}$, levando dessa forma a uma redução do erro de estimação.

$$\overline{|Y(e^{j\omega})|} = \frac{1}{M}\sum_{i=0}^{M-1}|Y_i(e^{j\omega})| \tag{3.7}$$

sendo $Y_i(e^{j\omega})$ a transformada da *i-ésima* janela $y(k)$. Desse modo, o estimador espectral utilizando a média do espectro do ruído é dado por:

$$\hat{X}_M(e^{j\omega}) = \left[\left|\overline{Y(e^{j\omega})}\right| - M(e^{j\omega})\right]e^{j\theta_y(e^{j\omega})} \tag{3.8}$$

sendo $X_M(e^{j\omega})$ é o valor do sinal de voz determinado a partir da subtração espectral entre o valor médio do espectro do sinal corrompido

por ruído $\overline{|Y(e^{j\omega})|}$ e o valor médio do ruído $M(e^{j\omega})$ determinado durante os instantes de silêncio.

- **Retificação de meia onda:** depois de determinada a magnitude média, para cada freqüência onde a magnitude do espectro do sinal ruidoso for menor que a média da magnitude do espectro do ruído, sua saída é igualada a zero. Esta modificação no sinal é implementada pela retificação de meia onda. Desse modo, o estimador pode ser representado por:

$$\hat{X}(e^{j\omega}) = H_R(e^{j\omega})Y(e^{j\omega}) \tag{3.9}$$

sendo: $H_R(e^{j\omega}) = \dfrac{H(e^{j\omega}) + |H(e^{j\omega})|}{2}$.

O objetivo da retificação de meia onda é diminuir a magnitude do espectro em cada freqüência pela diminuição da média do ruído. A principal desvantagem da retificação é quando a soma do ruído e do sinal de voz em cada freqüência for menor do que a média do ruído.

- **Redução do erro residual:** depois de aplicada a retificação de meia onda percebe-se no sinal de voz (principalmente nos trechos de silêncio) a presença de um ruído residual. Para a redução desse ruído BOLL propõem a substituição da média do ruído determinada

durante os instantes de silêncio na eq. (3.8) por um valor mínimo determinado pela análise de janelas aplicadas ao sinal. Essa substituição deverá ocorrer quando a magnitude de $\hat{X}(e^{j\omega})$ for menor do que o máximo valor para o ruído residual determinado durante os instantes de silêncio.

- **Atenuação do sinal durante trechos de silêncio:** a fase final da redução de ruídos neste método proposto por BOLL consiste na redução de ruídos durante os trechos de silêncio do sinal de voz. BOLL definiu um detector para determinar a ausência de voz no sinal, ou seja, para determinar os instantes de silêncio no sinal e conseqüentemente eliminar o ruído neste trecho. A equação do detector é:

$$T = 20\log_{10}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{\hat{X}(e^{j\omega})}{M(e^{j\omega})}\right|d\omega\right] \tag{3.10}$$

sendo $\hat{X}(e^{j\omega})$ a magnitude da estimativa do sinal de voz, $M(e^{j\omega})$ é o valor médio da magnitude do ruído determinado durante os instantes de silêncio e $T$ um limiar (threshold) com valor $-12$dB determinado empiricamente para assegurar que somente os trechos contendo ruídos sejam atenuados. Se o valor de $T$ for menor que $-12$dB, a janela que está sendo analisada é classificada como não tendo atividades sonoras.

A figura 3.2 descreve o esquema do algoritmo proposto por BOLL.

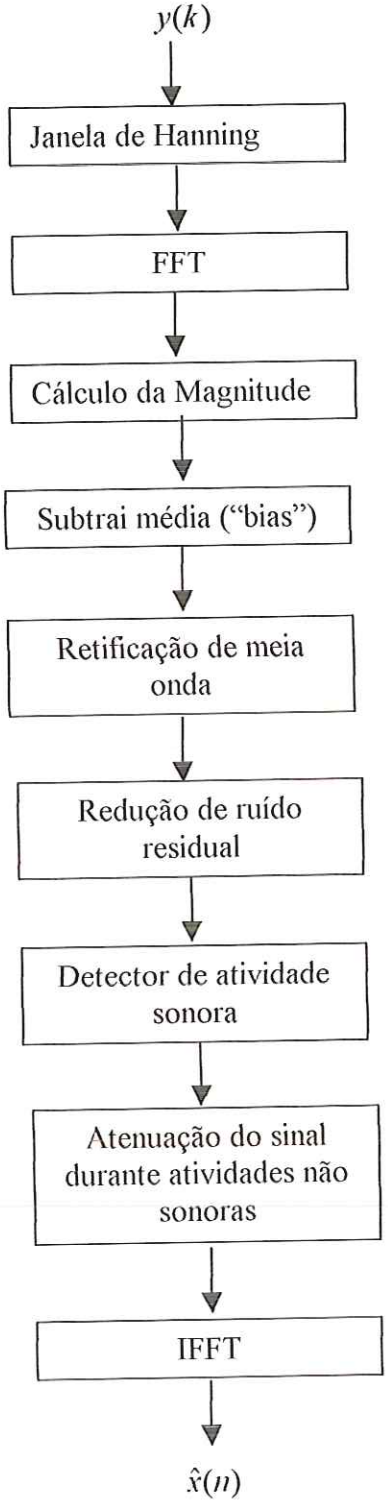



FIGURA 3.2 – Diagrama em blocos do sistema proposto por BOLL.

Este método desenvolvido foi aplicado a um sinal de voz amostrado a uma taxa de 6.67 kHz, colhido dentro de um helicóptero. O método proposto reduz de forma eficiente o ruído aditivo, mas algum ruído residual permanece. Tanto a inteligibilidade quanto à qualidade do sinal processado apresentam melhoras quando comparadas ao sinal de voz contaminado por ruído.

Trabalho semelhante ao proposto por BOLL foi publicado por VIEIRA-FILHO *et al.* (1986). Neste trabalho discute-se o método da subtração espectral de potências e filtragem de Wiener aplicados a redução de ruídos em sinais de voz captados em veículos. De acordo com os autores, três componentes básicas são responsáveis pela degradação de um sinal de voz no ambiente veicular: ruído do motor, o ruído do atrito dos pneus (ou ruído de rolamento) e o ruído provocado pelo vento. Se for levado em conta o fraco isolamento acústico dos veículos, considera-se também o ruído do ambiente por onde o veículo trafega. Outro fato importante considerado neste trabalho, característico de cada tipo de ruído e variando de acordo com cada tipo de veículo, é a estacionariedade do ruído.

A técnica analisada no trabalho em questão considera o uso de aparelhos telefônicos do tipo "*hands-free*" (HFT), através do qual o motorista utiliza um microfone, que geralmente é acoplado ao pára-sol do motorista (outros lugares também podem ser utilizados) e fone de ouvido. O sinal resultante, devido à distância do microfone e do fone de ouvido, será um sinal degradado, dificultando dessa forma o entendimento por parte do receptor, fazendo-se necessário o uso de técnicas de enriquecimento de sinais.

Para o desenvolvimento do método de redução de ruído baseado na subtração espectral de potências, no filtro de Wiener e no método EM (método de

EPHRAIM & MALAH (1984)), o sinal de voz corrompido por ruído é representado pela eq. (3.1). Calculando-se a subtração espectral de potências, tem-se que:

$$\left|X(e^{j\omega})\right|^2 = \left|Y(e^{j\omega})\right|^2 - \left|V(e^{j\omega})\right|^2 \tag{3.11}$$

A partir da eq. (3.11) pode-se obter o filtro redutor de ruído $H(e^{j\omega})$, definido por:

$$H(e^{j\omega}) = \begin{cases} \dfrac{Y(e^{j\omega})}{\left|Y(e^{J\omega})\right|^2}\sqrt{\left|Y(e^{j\omega})\right|^2 - \left|V(e^{j\omega})\right|^2} & \text{se } \left|Y(e^{j\omega})\right|^2 \geq \left|V(e^{j\omega})\right|^2 \\ 0 & \text{caso contrário} \end{cases} \tag{3.12}$$

Segundo o método EM, tanto o sinal de voz quanto o ruído são modelados estatisticamente como processos aleatórios estacionários tendo o seguinte filtro redutor de ruídos:

$$H(e^{j\omega}) = \Gamma(1{,}5)\frac{\sqrt{\eta(e^{j\omega})}}{SNR_post(e^{j\omega})}\exp\left[\frac{-\eta(e^{j\omega})}{2}\right]\left\{\left[1+\eta(e^{j\omega})\right]I_0\left[\frac{+\eta(e^{j\omega})}{2}\right]+\right.$$

$$\left.\eta(e^{j\omega})I_1\left[\frac{+\eta(e^{j\omega})}{2}\right]\right\} \tag{3.13}$$

sendo: $\eta(e^{j\omega}) = \frac{SNR_prior(e^{j\omega})}{1+SNR_prior(e^{j\omega})} SNR_post(e^{j\omega})$, $\Gamma(.)$ é a função gama, $I_0$ e $I_1$ representam as funções de Bessel modificadas do tipo um e de ordens zero e um, respectivamente.

Os valores para as expressões $SNR_prior(e^{j\omega})$ e $SNR_post(e^{j\omega})$ podem ser calculados da seguinte forma:

$$SNR_prior(e^{j\omega}) = \frac{\left|X(e^{j\omega})\right|^2}{\left|V(e^{j\omega})\right|^2} \tag{3.14}$$

$$SNR_post(e^{j\omega}) = \frac{\left|Y(e^{j\omega})\right|^2}{\left|V(e^{j\omega})\right|^2} \tag{3.15}$$

sendo os termos *SNR_prior* e *SNR_post* as relações sinal ruído a *priori* e *posteriori* respectivamente.

Outro ponto importante observado por VIEIRA-FILHO et al. (1986) é que a subtração espectral de potências pode ser representada por um filtro de Wiener generalizado. O filtro de Wiener clássico é dado por:

$$\left|H(e^{j\omega})\right| = \frac{E\left[\left|X(e^{j\omega})\right|^2\right]}{E\left[\left|X(e^{j\omega})\right|^2\right] + E\left[\left|V(e^{j\omega})\right|^2\right]} \tag{3.16}$$

Aplicando-se a subtração espectral de potências no filtro de Wiener clássico tem-se o filtro de Wiener em função da relação sinal ruído.

$$\left|H(e^{j\omega})\right| = \left\{\frac{SNR_prior(e^{j\omega})}{\alpha + SNR_post(e^{j\omega})}\right\}^{\beta} \tag{3.17}$$

os valores de $\alpha$ e $\beta$ definem o filtro. Geralmente os valores de $\alpha \geq 1$ e $\beta$ variam entre 0 e 1. Para um valor de $\alpha = 1$, a potência do ruído é exatamente igual a potência estimada; para um valor de $\alpha < 1$, tem-se uma subestimação da potência do ruído (menor redução do ruído) e para um valor de $\alpha > 1$ tem-se uma superestimação da potência do ruído, o que significa uma melhor redução do ruído mas aumentando o ruído musical (ruído que permanece depois que o sinal é filtrado). O valor de $\beta$ representa a atenuação do filtro e será tanto maior quanto mais próximo de 1 for o valor de $\beta$. Os valores de $\alpha$ e $\beta$ são determinados através de simulação.

Esse filtro foi denominado pelos autores de **Nova Subtração Espectral de Potências (NSEP).** Para o teste do método desenvolvido, os autores utilizaram uma frase em francês, de aproximadamente 5 segundos de duração, corrompida por ruído gravado no interior de um veículo. Os resultados obtidos pelos autores mostram que o método apresenta um bom desempenho na redução de ruídos sem provocar distorções no sinal resultante (filtrado).

NETO (1989), também emprega técnicas de subtração espectral de potências juntamente com o filtro de Wiener para a redução de ruídos em sinais de voz. Os métodos desenvolvidos são baseados no balanceamento espectral adaptativo

que consiste em fazer modificações na amplitude espectral a curtos intervalos de tempo no sinal de voz degradado.

Três métodos (técnicas) são desenvolvidos por NETO (1989), baseados no balanceamento espectral: balanceamento utilizando o filtro de Wiener-Kolmogoroff, balanceamento pela estimação das amplitudes espectrais e balanceamento pela estimação das potências espectrais.

Para o desenvolvimento das funções de transferências, NETO (1989) considera o sinal corrompido por ruído de acordo com a eq. (3.1). Para o método do balanceamento baseado no filtro de Wiener-Kolmogoroff, a seguinte função de transferência é desenvolvida:

$$H_1(e^{j\omega}) = \frac{Y(e^{j\omega}) - V(e^{j\omega})}{Y(e^{j\omega})} \tag{3.18}$$

sendo $Y(e^{j\omega})$ o espectro de amplitude para o sinal de voz disponível degradado por ruído e $V(e^{j\omega})$ é o espectro de amplitude para o ruído aditivo.

O balanceamento utilizando a estimação de amplitudes espectrais, como o próprio nome diz, consiste na estimação das amplitudes espectrais a curtos intervalos de tempo onde os sinais apresentam estacionaridade. Para este método é apresentada a seguinte função de transferência:

$$H_2(e^{j\omega}) = \frac{\left|Y(e^{j\omega})\right| - \hat{E}\left\|V(e^{j\omega})\right\|}{Y(e^{j\omega})} \tag{3.19}$$

sendo $\hat{E}\left[\left|V(e^{j\omega})\right|\right]$ a estimativa da amplitude espectral do ruído determinada por média temporal nos intervalos de pausas do sinal de voz, e finalmente, para o balanceamento pela estimação de potências espectrais, determinou-se a seguinte expressão para a função de transferência:

$$H_3(e^{j\omega}) = \frac{\left|Y(e^{j\omega})\right|^2 - \hat{E}\left[\left|V(e^{j\omega})\right|^2\right]^{\frac{1}{2}}}{\left|Y(e^{j\omega})\right|^2} \tag{3.20}$$

Utilizando as três funções de transferências apresentadas acima, NETO (1989) desenvolve os seguintes algoritmos de supressão de ruídos por balanceamento espectral adaptativo:

- **Balanceamento pelo filtro de Wiener-Kolmogoroff**

$$H_1(e^{j\omega}) = \begin{cases} 1 - D(e^{j\omega}) & \text{p/} \quad 0 \le D(e^{j\omega}) < 1 \\ 0 & \text{p/} \quad D(e^{j\omega}) \ge 1 \end{cases} \tag{3.21}$$

- **Balanceamento pela estimação da amplitude espectral**

$$H_2(e^{j\omega}) = \begin{cases} 1 - \sqrt{D(e^{j\omega})} & \text{p/} \quad 0 \le D(e^{j\omega}) < 1 \\ 0 & \text{p/} \quad D(e^{j\omega}) \ge 1 \end{cases} \tag{3.22}$$

- **Balanceamento pela estimação da potência espectral**

$$H_3(e^{j\omega}) = \begin{cases} \sqrt{1 - D(e^{j\omega})} & \text{p/} \quad 0 \le D(e^{j\omega}) < 1 \\ 0 & \text{p/} \quad D(e^{j\omega}) \ge 1 \end{cases} \tag{3.23}$$

sendo $D(e^{j\omega}) = \dfrac{\hat{E}\left[\left|V(e^{j\omega})\right|^2\right]}{\left|Y(e^{j\omega})\right|^2}$ o fator relativo de degradação.

Os três algoritmos apresentados acima foram aplicados a dois sinais de voz, sendo um deles degradados por ruído branco e o outro degradado por ruído de automóvel representados respectivamente por SP1 e SP2. Os resultados obtidos por NETO (1989) são mostrados nas tabelas 3.1 e 3.2.

TABELA 3.1 – Resultados dos algoritmos $H_1$, $H_2$ e $H_3$ para os sinais SP1 e SP2 degradados por ruído branco.

| Algoritmo | SNR (dB) | SNR-Seg (dB) | Sinal de Prova |
|---|---|---|---|
| $H_1$ | 4,15 | 7,85 | SP1 |
| $H_2$ | 5,29 | 9,41 | |
| $H_3$ | 3,30 | 6,82 | |
| $H_1$ | 4,93 | 15,99 | SP2 |
| $H_2$ | 5,77 | 16,72 | |
| $H_3$ | 4,21 | 15,35 | |

TABELA 3.2 – Resultados dos algoritmos $H_1$, $H_2$ e $H_3$ para os sinais SP1 e SP2 degradados por ruído de automóvel.

| Algoritmo | SNR (dB) | SNR-Seg (dB) | Sinal de Prova |
|---|---|---|---|
| $H_1$ | 5,18 | 8,54 | SP1 |
| $H_2$ | 5,91 | 9,56 | |
| $H_3$ | 4,03 | 7,44 | |
| $H_1$ | 6,94 | 17,26 | SP2 |
| $H_2$ | 9,15 | 18,40 | |
| $H_3$ | 5,43 | 16,32 | |

Pela análise dos resultados apresentados nas tabelas 3.1 e 3.2, nota-se claramente uma superioridade do algoritmo $H_2$ (balanceamento espectral por estimação da amplitude espectral). Dentre os fatores responsáveis por essa superioridade destacam-se o melhor desempenho deste algoritmo nos intervalos de pausas e melhor robustez com relação a possíveis erros de estimação espectral.

Semelhante ao trabalho desenvolvido por NETO (1989), VIEIRA-FILHO *et al.* (1995) desenvolveu uma análise teórica das técnicas clássicas para a filtragem de sinais de voz baseadas na subtração espectral a curto-termos (SECT), com implementação e avaliação dos resultados, destacando-se suas vantagens e limitações.

O seguinte filtro redutor de ruído foi desenvolvido:

$$H(e^{j\omega}) = 1 - \frac{1}{\sqrt{G(e^{j\omega})}} \tag{3.24}$$

sendo:

$$G(e^{j\omega}) = \frac{\left|Y(e^{j\omega})\right|^2}{\left|V(e^{j\omega})\right|^2} \tag{3.25}$$

A eq. (3.25) representa a medida instantânea da relação sinal ruído após a degradação do sinal. O filtro representado pela eq. (3.24) é baseado diretamente na amplitude dos sinais e o método resultante é denominado **SEA – Subtração Espectral em Amplitudes**. Se o termo $G(e^{j\omega})$ na eq. (3.24) for menor do que 1 o filtro será indeterminado. Como apresentado por BOLL, uma solução normalmente adotada é fazer uma retificação de meia onda depois da subtração. A eq. (3.24) pode ser representada matematicamente por:

$$\left|H(e^{j\omega})\right|_{SEA} = \begin{cases} 1 - \dfrac{1}{\sqrt{G(e^{j\omega})}}, & \text{se } G(e^{j\omega}) \geq 1 \\ 0, & \text{se } G(e^{j\omega}) < 1 \end{cases} \tag{3.26}$$

Além do filtro baseado na SEA também foi desenvolvido um novo filtro baseado na subtração espectral em potência denominado **SEP – Subtração Espectral em Potência** e pode ser representado por:

$$H(e^{j\omega}) = \sqrt{1 - \frac{1}{G(e^{j\omega})}} \tag{3.27}$$

De forma semelhante a eq. (3.24), o filtro descrito pela eq. (3.27) também será indeterminado caso $G(e^{j\omega})$ seja menor que 1. Fazendo-se uma retificação de meia onda na eq. (3.27) obtém-se:

$$\left|H(e^{j\omega})\right|_{SEP} = \begin{cases} \sqrt{1-\dfrac{1}{G(e^{j\omega})}}, & \text{se } G(e^{j\omega}) \geq 1 \\ 0, & \text{se } G(e^{j\omega}) < 1 \end{cases} \tag{3.28}$$

O filtro obtido pelo método SEP pode ser generalizado utilizando-se a filtragem de Wiener que representa um filtro ótimo para a estimação linear de voz minimizando o erro médio quadrático. O filtro de Wiener é representado por:

$$H(e^{j\omega}) = \frac{\left|V(e^{j\omega})\right|^2}{\left|Y(e^{j\omega})\right|^2} \tag{3.29}$$

Com base na eq. (3.2) e na eq. (3.25), o filtro de Wiener pode ser generalizado e representado por:

$$H(e^{j\omega}) = \left[1-\frac{1}{\beta G(e^{j\omega})}\right]^{\alpha} \tag{3.30}$$

Da eq. (3.30) pode-se observar que, para $\alpha = 0.5$ e $\beta = 1$, o filtro de Wiener é a técnica SEP representada na eq. (3.27). No caso de $\alpha = \beta = 1$, tem-se o filtro de Wiener padrão. Essa técnica baseada na eq. (3.30) com diferentes valores para $\alpha$ e $\beta$ é denominada **SEM – Subtração Espectral Modificada**.

Para a avaliação das técnicas apresentadas, VIEIRA-FILHO *et. al* (1995) restringiu-se os testes às técnicas de SEP e filtragem de Wiener clássica, regulando-se os parâmetros básicos através de testes auditivos. O sinal de voz aplicado nos testes foi um sinal em francês discretizado a uma taxa de amostragem de 16 kHz. A este sinal foi adicionado um ruído de veículo a passeio a uma velocidade de 80 km/h. O ruído apresenta características estacionárias. Dos resultados obtidos notou-se que o filtro de Wiener apresenta uma melhor SNR durante os trechos de silêncio e que ambas as técnicas se comportam da mesma maneira durante os intervalos de voz. Pela análise dos espectrogramas foi constatado que o processamento através da SEP apresenta uma melhor redução de ruído, mas ambos os métodos deixa um ruído residual nas baixas freqüências. Com relação aos testes subjetivos, os sinais processados pelos dois métodos são bem melhores em termos de inteligibilidade que o sinal ruidoso original. Foi verificado também que o sinal processado pelo filtro de Wiener é mais agradável auditivamente.

## 3.3 Filtros Ótimos

O termo filtro é usado freqüentemente para descrever um objeto físico ou programa de computador utilizado em um conjunto de dados ruidosos com o objetivo de extrair determinadas informações responsáveis pela degradação dos dados

HAYKIN (1986). Um exemplo de degradação de informações pode ser citado quando um sinal percorre um canal de comunicação. Este sinal, por sua vez, pode ser corrompido por ruídos do meio ambiente ou mesmo ruídos do próprio canal de comunicação, dificultando dessa maneira o entendimento por parte do receptor. Dessa forma, faz-se necessário o uso de filtros para a extração desses dados indesejáveis.

Um filtro pode ser responsável pelas seguintes operações:

- **filtragem:** consiste na extração de uma informação indesejável usando valores correntes e passados do sinal.

- **predição:** neste caso, o filtro produz uma estimativa de valores futuros do sinal, combinando os valores presentes e passados.

A teoria de filtros ótimos é desenvolvida com base em um sinal de erro. Este sinal de erro é definido como a diferença entre o sinal desejado e o sinal estimado (saída do filtro). Vários métodos foram desenvolvidos visando uma otimização do sinal de erro. Dentre esses métodos pode-se destacar o método dos mínimos quadrados.

Quando o sinal de entrada (do filtro) for um sinal estacionário, pode-se empregar a teoria do filtro de Wiener, mas quando o sinal de entrada não for estacionário, o filtro de Wiener não pode ser aplicado e deverá ser aplicado um outro filtro que assuma uma forma variante no tempo. Este filtro é conhecido como filtro

de Kalman. Ambos os filtros são baseados na teoria dos mínimos quadrados, que consiste em minimizar a soma dos erros quadráticos médios.

### 3.3.1. Filtro de Wiener

Para o desenvolvimento das equações que compõem o filtro de Wiener, para a estimação de um sinal $x(k)$, considera-se a seguinte equação de observação HAYES (1996):

$$y(k) = x(k) + v(k) \tag{3.31}$$

sendo $y(k)$ a única informação disponível do sinal, $x(k)$ o sinal de voz livre de ruído e $v(k)$ o sinal de ruído aditivo. Considera-se que os sinais $y(k)$ e $v(k)$ são processos aleatórios estacionários em sentido amplo com funções de autocorrelação conhecidas $r_y(k)$ e $r_x(k)$, e correlação cruzada $r_{xy}(k)$. A resposta em freqüência do filtro de Wiener é:

$$W(z) = \sum_{n=0}^{P-1} a(n) z^{-n} \tag{3.32}$$

sendo $a(n)$ os coeficientes do filtro, $z^{-1}$ representa o operador de deslocamento e $P$ é a ordem do filtro.

A estimativa do sinal $x(k)$ é dada pela seguinte relação:

$$\hat{x}(k) = \sum_{n=0}^{P-1} a(n) y(k-n) \tag{3.33}$$

De acordo com a relação acima, tendo-se o sinal $y(k)$ corrompido por ruído aditivo, é necessário calcular os coeficientes $a(n)$ que são responsáveis pela minimização do erro quadrático médio:

$$\xi = E\left[|e(k)|^2\right] = E\left[|x(k) - \hat{x}(k)|^2\right] \tag{3.34}$$

Para a minimização do erro quadrático médio, deve-se derivar o erro em relação aos coeficientes $a(n)$, igualar o resultado a zero e resolver a equação para $n = 0,1,2,...P-1$.

$$\frac{\partial \xi}{\partial a^*(n)} = \frac{\partial}{\partial a^*(n)} E\{e(k)e^*(k)\} = E\left\{e(k)\frac{\partial e^*(k)}{\partial a^*(n)}\right\} = 0 \tag{3.35}$$

com

$$e(k) = x(k) - \sum_{n=0}^{P-1} a(n) y(k-n) \tag{3.36}$$

Desenvolvendo a equação acima, obtém-se:

$$E\{e(k)x^*(k-n)\}=0\,; \quad \text{para } n = 0, 1, 2, 3,..., P\text{-}1 \qquad (3.37)$$

A eq. (3.37) é conhecida como *princípio da ortogonalidade* ou *teorema da projeção*. Substituindo a eq. (3.36) na eq. (3.37), tem-se:

$$E\{x(k)x^*(k-n)\}-\sum_{l=0}^{P-1} a(l)E\{y(k-l)y^*(k-n)\}=0 \qquad (3.38)$$

Como $y(k)$ e $x(k)$ são processos estacionários em sentido amplo, então: $E\{x(k)y^*(k-n)\}=r_{xy}(n)$ e $E\{y(k-l)y^*(k-n)\}=r_y(n-l)$. A eq. (3.38) pode ser escrita como:

$$\sum_{l=0}^{P-1} a(l)r_y(n-l)=r_{xy}(n)\,; \quad \text{para } n = 0, 1, 2, 3,..., P\text{-}1 \qquad (3.39)$$

A eq. (3.39) é um conjunto de $p$ equações lineares em $p$ variáveis $a(n)$, para $n$ = 0, 1, 2, 3,..., $P$-1. Para a representação matricial considera-se que a seqüência de autocorrelação é simétrica conjugada, $r_y(n)=r_y(-n)$:

$$\begin{bmatrix} r_y(0) & r_y{}^*(1) & \cdots & r_y{}^*(p-1) \\ r_y(1) & r_y(0) & \cdots & r_y{}^*(p-2) \\ r_y(2) & r_y(1) & \cdots & r_y{}^*(p-3) \\ \vdots & \vdots & \cdots & \vdots \\ r_y(p-1) & r_y{}^*(p-2) & \cdots & r_y(0) \end{bmatrix} \begin{bmatrix} a(0) \\ a(1) \\ a(2) \\ \vdots \\ a(p-1) \end{bmatrix} = \begin{bmatrix} r_{xy}(0) \\ r_{xy}(1) \\ r_{xy}(2) \\ \vdots \\ r_{xy}(p-1) \end{bmatrix} \quad (3.40)$$

Esta equação é chamada de Wiener-Hopf e também pode ser escrita na seguinte forma:

$$\boldsymbol{R}_x \boldsymbol{a} = \boldsymbol{r}_{xy} \quad (3.41)$$

A expressão, para o cálculo do erro quadrático médio é dada por:

$$\xi_{\min} = r_x(0) - \sum_{l=0}^{P-1} a(l) r_{xy}{}^*(l) \quad (3.42)$$

O conjunto de equações que compõem o filtro de Wiener FIR é resumido na tabela 3.3.

TABELA 3.3 – Equações para o filtro de Wiener.

| | |
|---|---|
| **Equação de Wiener-Hopf:** | $\sum_{l=0}^{P-1} a(l) r_y(n-l) = r_{xy}(n)$ $k$=0,1, 2, 3,...$P$-1 |
| **Correlação** | $r_y(n) = E\{y(k)y^*(k-n)\}$ <br> $r_{xy}(n) = E\{x(k)y^*(k-n)\}$ |
| **Erro médio quadrático** | $\xi_{\min} = r_x(0) - \sum_{l=0}^{P-1} a(l) r_{xy}^*(l)$ |

O filtro de Wiener quando aplicado à estimação de sinais de voz corrompidos por ruído aditivo produz resultados satisfatórios, mas apresenta uma limitação quanto aos dados de entrada que devem ser estacionários. Quando o sinal de entrada não for estacionário será empregado o filtro de Kalman que será apresentado no capítulo 4.

## 3.4. Considerações Finais

Neste capítulo foram apresentadas duas técnicas empregadas em processamento digital de sinais para a redução de ruídos em sinais de voz: subtração espectral de potências e filtragem de Wiener.

Tanto a subtração espectral de potências quanto a filtragem de Wiener são técnicas relativamente simples e fáceis de ser implementadas em computador. Apesar dessa facilidade ambas apresentam limitações. A subtração espectral quando

aplicada a filtragem de sinais apresenta um ganho baixo em relação aos filtros ótimos e as vezes provoca fortes distorções no sinal filtrado.

Quanto à filtragem de Wiener, os resultados são satisfatórios e o ganho alto. A sua limitação é com relação aos dados de entrada que devem ser estacionários, caso contrário o filtro não produzirá bons resultados.

# Capítulo 4

# Filtro de Kalman

## 4.1. Introdução

As técnicas de filtragem têm estado presente na vida do homem desde os tempos remotos assegurando a sua sobrevivência. Durante os anos de 1930 e 1940, este conceito se estendeu ao processo de separação de sinais a partir de medidas ruidosas (sinais corrompidos por ruído), surgindo dessa forma o conceito de filtragem de sinais. Dentre os filtros digitais destacam-se os filtros de freqüência seletivas (filtros de Butterworth, Chebyshev, Elíptico) e os filtros ótimos (filtros de Wiener e de Kalman).

O filtro de Kalman foi publicado pela primeira vez em 1960 por Rudolf Emil Kalman KALMAN (1960). O filtro de Wiener ainda largamente utilizado em vários trabalhos e pesquisas apresenta limitações em relação aos dados de entrada, que devem ser estacionários. Como na natureza nem todos os fenômenos são estacionários, o filtro de Wiener não consegue solucionar tais problemas.

A formulação do filtro de Kalman utiliza espaço de estados, através dos quais o comportamento de um sistema dinâmico é descrito por um conjunto de variáveis de estados. As variáveis de estados contêm todas as informações necessárias sobre o comportamento do sistema de forma que dados os valores presentes e passados da entrada, pode-se determinar os estados futuros do sistema KAILATH (1980).

O advento das comunicações digitais, principalmente o surgimento e evolução dos computadores digitais, provocou uma grande utilização do algoritmo de Kalman, notadamente na área científica. Dentre os fatores que favoreceram a popularidade do filtro de Kalman, pode-se destacar:

- utilização de estruturas recursivas;
- manipulação de álgebra matricial simples, que possibilita sua implementação em computador digital;
- ampla aplicabilidade em várias áreas da engenharia (onde a estimação de estado é encontrada com freqüência);
- adaptabilidade a aplicações não-lineares.

De acordo com NASCIMENTO JR. (1988), algumas dificuldades podem surgir quando técnicas de filtragem de Kalman são utilizadas. Dentre essas dificuldades, destacam-se:

- **sintonização:** consiste no ajuste adequado dos parâmetros. Os parâmetros dependem da aplicação utilizada, bem como um conhecimento do sistema.

- **condicionamento numérico:** de acordo com alguns autores, o algoritmo do filtro de Kalman é numericamente instável (mal condicionado), pois pode-se ter resultados incoerentes como resposta do filtro, devido a problemas numéricos.

- **limitação de recursos computacionais:** aplicações em tempo real geralmente envolvem limitação da quantidade de memória, comprimento da palavra utilizada nos cálculos e tempo de processamento, podendo provocar dessa forma divergência do filtro.

Inicialmente a grande área de aplicação do filtro de Kalman foi no controle de sistemas dinâmicos complexos tais como: aeronaves, navios e processos de produção contínuo. O filtro de Kalman é também empregado na predição de sistemas dinâmicos, como:

- navegação de mísseis e aviões;
- estimação de sinais (de voz, de eletrocardiograma);
- processamento de imagens (radar, tomografia e cartografia);
- determinação das órbitas dos planetas;
- trajetória de corpos celestes;
- identificação de sistemas.

## 4.2. Formulação do Filtro de Kalman

Como mencionado na seção anterior, o filtro de Kalman visa estimar os estados de um sistema dinâmico. Por estimação, entende-se o processo pelo qual dadas as medidas do sistema, seu modelo e as estatísticas dos ruídos que atuam tanto nas medidas quanto no modelo do sistema, determina-se, no decorrer do tempo, uma estimativa dos valores dos estados do sistema.

Considere um sistema cuja dinâmica é representada pela seguinte equação estocástica:

$$x(k) = A(k-1)x(k-1) + w(k) \tag{4.1}$$

sendo:

$x$ = vetor de estados do sistema;

$A$ = matriz de transição de estados variante no tempo. É uma matriz quadrada de dimensão $nxn$ que representa a transição entre os estados $x(k-1)$ e $x(k)$.

$w$ = vetor de dimensão $nx1$ que representa um processo de ruído, conhecido como ruído de estado, caracterizado estatisticamente por:

$$E[w(k)] = 0\,; \qquad \text{(média zero)}$$

$$E[w(k)w^T(i)] = \begin{cases} Q(k) & k = i \\ 0 & k \neq i \end{cases} \tag{4.2}$$

Este vetor pode ser referenciado por "ruído de estado", pois exerce influência direta na dinâmica dos estados do sistema. O fato de o ruído de estado ser descorrelacionado no tempo $(k \neq i)$ indica que ele é branco (independentemente em relação a si mesmo no tempo) NASCIMENTO JR. (1988).

A única informação disponível do sinal é a medida discreta $y(k)$ corrompida por ruído branco gaussiano dada pela seguinte equação:

$$y(k) = C(k)x(k) + v(k) \tag{4.3}$$

sendo:

$y$ = vetor de medidas discretas do sistema com dimensão $mx1$;

$C$ = matriz de medidas com dimensão $mxn$;

$v$ = vetor coluna de dimensão $mx1$ que representa ruído branco gaussiano aditivo cujas estatísticas são dadas por:

$$E[v(k)] = 0 \text{; (média zero)}$$

$$E[v(k)v^T(i)] = \begin{cases} R(k) & k = i \\ 0 & k \neq i \end{cases} \tag{4.4}$$

$$E[v(k)w^T(i)] = 0 \text{;} \quad \text{para todo } k \text{ e } i$$

as matrizes $Q(k)$ e $R(k)$ representam as matrizes de covariância para os ruídos $w(k)$ e $v(k)$ respectivamente.

Para o início do desenvolvimento das equações, considera-se a existência de uma estimativa inicial do estado $x$ que é dada por:

$$\hat{x}^{-}(0) = E[x(0)] \tag{4.5}$$

sendo que o sinal de menos ("-") indica uma estimativa inicial antes da assimilação de uma nova medida. O erro de estimação associado a estimativa inicial é dado pela seguinte expressão:

$$e^{-}(0) = x(0) - \hat{x}^{-}(0) \tag{4.6}$$

e a matriz de covariância do erro é:

$$P^{-}(0) = E[e^{-}(0)e^{-T}(0)] = E\left[\left(x(0) - \hat{x}^{-}(0)\right)\left(x(0) - \hat{x}^{-}(0)\right)^{T}\right] \tag{4.7}$$

Depois de definida as estimativas iniciais, o processo de estimação é repetido para cada nova observação $y(k)$ disponível, com o objetivo de determinar a estimativa $\hat{x}(k)$ que minimiza o erro quadrático médio. A determinação de $\hat{x}(k)$ envolve dois passos: primeiro determina-se a estimativa $\hat{x}^{-}(k)$ sem a utilização de uma nova medida $y(k)$ e no segundo passo, quando $y(k)$ torna-se disponível calcula-se $\hat{x}(k)$.

Inicialmente, nenhuma medida está disponível para o cálculo de $\hat{x}^{-}(k)$, que é dada pela seguinte equação:

$$\hat{x}^{-}(k) = A(k-1)\hat{x}(k-1) \tag{4.8}$$

Para a determinação da matriz de covariância para o erro associado com a estimativa $\hat{x}^{-}(k)$, é necessário calcular primeiro a expressão para o erro associado com o valor da estimativa predita:

$$\begin{aligned} e^{-}(k) &= x(k) - \hat{x}^{-}(k) \\ &= \big(A(k-1)x(k-1) + w(k)\big) - A(k-1)\hat{x}(k-1) \\ &= A(k-1)e^{-}(k-1) + w(k) \end{aligned} \tag{4.9}$$

A estimativa do erro $e^{-}(k-1)$ é descorrelacionada com o ruído de estado $w(k)$ porque este é um processo aleatório branco. A expressão para a matriz de covariância associada com a estimativa $\hat{x}^{-}(k)$ é dada por:

$$\begin{aligned} P^{-}(k) &= E\left[e^{-}(k-1)e^{-}(k-1)^{T}\right] \\ &= E\left[\big(A(k-1)e(k-1) + w(k)\big)\big(A(k-1)e(k-1) + w(k)\big)^{T}\right] \\ &= A(k-1)P(k-1)A^{T}(k-1) + Q(k) \end{aligned} \tag{4.10}$$

Depois de determinada as estimativas iniciais, o passo seguinte consiste em usar a medida $y(k)$ para a determinação de $\hat{x}(k)$ da seguinte forma:

$$\hat{x}(k) = \hat{x}^{-}(k) + K(k)\left[y(k) - C(k)\hat{x}^{-}(k)\right] \tag{4.11}$$

sendo:

$\hat{x}(k)$, a estimativa atualizada;

$K(k)$, o fator de combinação (ainda não determinado).

A diferença $\left[y(k) - C(k)\hat{x}^{-}(k)\right]$ na equação (4.11) é chamada de medida de inovação e reflete a discrepância entre o que se mediu $y(k)$ menos o que se esperava medir $\left[C(k)\hat{x}^{-}(k)\right]$.

Para que a estimativa $\hat{x}(k)$ da equação (4.11) seja realizada é necessário encontrar o valor do fator de combinação $K(k)$ que produz uma estimativa atualizada ótima e é denominado ganho de Kalman. Para o cálculo do ganho de Kalman é necessário calcular a matriz de covariância do erro associada com a estimativa atualizada:

$$P(k) = E\left[e(k)e(k)^{T}\right] = E\left[\left(x(k) - \hat{x}(k)\right)\left(x(k) - \hat{x}(k)\right)^{T}\right] \tag{4.12}$$

Substituindo a eq. (4.3) na eq. (4.11) tem-se:

$$P(k) = E\left\{\left[\left(x(k) - \hat{x}^{-}(k)\right) - K(k)\left(C(k)x(k) + v(k) - C(k)\hat{x}^{-}(k)\right)\right]\right.$$
$$\left.\left[\left(x(k) - \hat{x}^{-}(k)\right) - K(k)\left(C(k)x(k) + v(k) - C(k)\hat{x}^{-}(k)\right)\right]^{T}\right\} \tag{4.13}$$

Calculando os valores indicados e notando que $\left[x(k) - \hat{x}^{-}(k)\right]$ é o erro da estimativa anterior que é descorrelacionado com a medida do erro $v(k)$, tem-se:

$$P(k) = [I - K(k)C(k)]P^{-}(k)[I - K(k)C(k)]^{T} + K(k)R(k)K^{T}(k) \qquad (4.14)$$

O ganho de Kalman é responsável pela minimização dos termos ao longo da diagonal principal de $P(k)$, pois estes representam as variâncias dos erros que estão associados a estimação dos elementos do vetor de estado $x$.

O processo de otimização pode ser realizado de diversas maneiras. Neste trabalho a derivação da expressão para o ganho de Kalman será desenvolvida utilizando métodos de cálculo diferencial, mais especificamente fórmulas de diferenciação de matrizes. A demonstração completa das expressões não será desenvolvida e pode ser encontrada em MAYBECK (1982) e BROWN & HWANG (1992).

Para as fórmulas de diferenciação de matrizes, considere três matrizes quadradas $A$, $B$ e $C$, sendo que a matriz $C$ é simétrica. O traço ("*trace*") de uma matriz é definido matematicamente como a soma dos elementos da diagonal principal. As fórmulas de diferenciação são dadas por:

$$\frac{d[traço(AB)]}{dA} = B^{T} \qquad (4.15)$$

$$\frac{d[traço(ACA^{T})]}{dA} = 2AC \qquad (4.16)$$

e a derivada de um escalar em relação a matriz é definida como:

$$\frac{ds}{dA} = \begin{bmatrix} \frac{ds}{da_{11}} & \frac{ds}{da_{12}} & \cdots \\ \frac{ds}{da_{21}} & \frac{ds}{da_{21}} & \cdots \\ \vdots & \vdots & \\ & & \end{bmatrix} \tag{4.17}$$

Expandindo a eq. (4.14), obtém-se a forma:

$$P(k) = P^{-}(k) - K(k)C(k)P^{-}(k) - P^{-}(k)C^{T}(k)K^{T}(k) +$$
$$+ K(k)\left[C(k)P^{-}(k)C^{T}(k) + R(k)\right]K^{T}(k) \tag{4.18}$$

A determinação do valor de $K(k)$ que fornece um valor mínimo para os erros de estimação, é equivalente a minimizar a soma dos elementos da diagonal principal da matriz $P$. Aplicando as fórmulas de diferenciação de matrizes para calcular a derivada parcial do traço de $P(k)$ em relação a $K(k)$, tem-se:

$$\frac{d(traçoP(k))}{dK(k)} = -2\left[C(k)P^{-}(k)\right]^{T} +$$
$$+ 2K(k)\left[C(k)P^{-}(k)C^{T}(k) + R(k)\right] \tag{4.19}$$

Igualando-se a eq. (4.19) a zero e resolvendo a equação, determina-se a expressão para o ganho de Kalman $K(k)$, que é responsável pela minimização do erro quadrático médio de estimação. A expressão resultante é:

$$K(k) = P^{-}(k)C^{T}(k)\left[C(k)P^{-}(k)C^{T}(k) + R(k)\right]^{-1} \tag{4.20}$$

Uma vez determinada a expressão para o ganho de Kalman e calculada a estimativa do estado $x$ no instante de tempo $k$, que é dada pela eq. (4.11), é necessário calcular a matriz de covariância do erro de estimação. A partir da eq. (4.14) tem-se:

$$P(k) = [I - K(k)C(k)]P^{-}(k)[I - K(k)C(k)]^{T} + K(k)R(k)K^{T}(k) \quad (4.21)$$

$$= P^{-}(k) - K(k)C(k)P^{-}(k) - P^{-}(k)C^{T}(k)K^{T}(k) +$$

$$+ K(k)[C(k)P^{-}(k)K^{T}(k) + R(k)]K^{T}(k) \quad (4.22)$$

Substituindo a eq. (4.20) em (4.22) e rearranjando os termos, obtém-se as seguintes expressões para $P(k)$:

$$P(k) = P^{-}(k) - P^{-}(k)C^{T}(k)[C(k)P^{-}(k)C^{T}(k) + R(k)]^{-1}C(k)P^{-}(k) \quad (4.23)$$

ou

$$P(k) = P^{-}(k) - K(k)[C(k)P^{-}(k)C^{T}(k) + R(k)]K^{T}(k) \quad (4.24)$$

ou

$$P(k) = [I - K(k)C(k)]P^{-}(k) \quad (4.25)$$

Das três expressões apresentadas para $P(k)$ a eq. (4.25) é a mais simples e a mais utilizada. Apresentadas estas equações, pode-se assimilar a medida no instante de tempo $k$ pelo uso da eq. (4.11) com ganho $K(k)$ que é dado pela eq. (4.20), sendo necessário o valor de $\hat{x}^-(k-1)$ e $P^-(k)$ para assimilar uma nova medida.

As eq. (4.8), (4.10), (4.11), (4.20) e (4.25) compõem o conjunto de equações recursivas do filtro de Kalman.

## 4.3. Modelo Recursivo do Filtro de Kalman

O filtro de Kalman realiza uma recursão utilizando o conjunto de equações apresentadas na seção anterior. O estado de um processo é estimado usando uma forma de realimentação, ou seja, o filtro calcula a estimativa em um determinado instante de tempo e então uma nova medida é adicionada ao sistema para corrigir a estimativa inicial.

As equações que compõem o filtro de Kalman podem ser divididas em dois grupos: ***equações de propagação (projeção)*** e ***equações de atualização (correção)*** NASCIMENTO JR. (1988). As equações de propagação são responsáveis por projetar a estimativa do estado corrente e da matriz de covariância, enquanto que as equações de atualização são responsáveis pela realimentação, ou seja, incorporam uma nova medida para obter uma estimativa atualizada (melhorada). As equações de atualização podem ser consideradas equações corretoras.

EESC - USP
Biblioteca
CAMPUS DE S. CARLOS

Resumindo, pode-se dizer que o filtro de Kalman é apenas um algoritmo computacional para o mapeamento de medidas discretas (entradas) em estimativas ótimas (saídas). A figura 4.1 mostra o ciclo de execução das equações.

A cada instante de tempo $k$, o par projeção (predição) e atualização (correção) é repetido utilizando as estimativas iniciais $\hat{x}^-(k-1)$ e $P^-(k-1)$ para determinar o valor da estimativa corrente $\hat{x}(k)$.

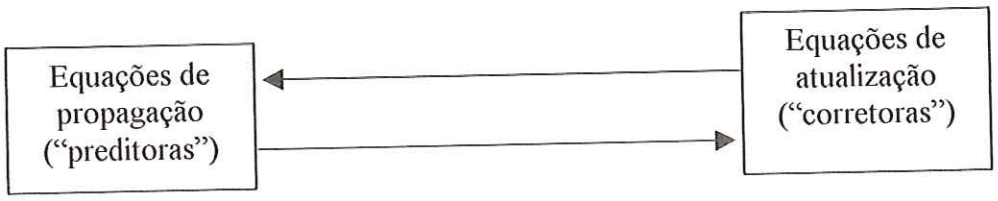



FIGURA 4.1 - Ciclo de execução das equações para o filtro de Kalman.

De forma esquemática, o algoritmo recursivo do filtro de Kalman, processa as medidas da seguinte forma:

1. Cálculo das estimativas iniciais $\hat{x}^-(0)$ e $P^-(0)$;
2. Início

   Para $k = 1, 2, 3, \ldots$, faça

$$\hat{x}^-(k) = A(k-1)\hat{x}(k-1)$$

$$P^-(k) = A(k-1)P(k-1)A^T(k-1) + Q(k)$$

$$K(k) = P^-(k)C^T(k)\left[C(k)P^-(k)C^T(k) + R(k)\right]^{-1}$$

$$\hat{x}(k) = \hat{x}^-(k) + K(k)\left[y(k) - C(k)\hat{x}^-(k)\right]$$

$$P(k) = [I - K(k)C(k)]P^{-}(k)$$

Fim

As equações específicas para a propagação das estimativas e de atualização para o filtro de Kalman discreto estão resumidas nas tabelas 4.1 e 4.2 respectivamente.

TABELA 4.1 – Equações de propagação das estimativas para o filtro de Kalman discreto.

| | |
|---|---|
| **Propagação da estimativa do estado** | $\hat{x}^{-}(k) = A(k-1)\hat{x}(k-1)$ |
| **Propagação do Erro de Covariância** | $P^{-}(k) = A(k-1)P(k-1)A^{T}(k-1) + Q(k)$ |

TABELA 4.2 – Equações de atualização das estimativas para o filtro de Kalman discreto.

| | |
|---|---|
| **Ganho** | $K(k) = P^{-}(k)C^{T}(k)[C(k)P^{-}(k)C^{T}(k) + R(k)]^{-1}$ |
| **Estimativa do estado** | $\hat{x}(k) = \hat{x}^{-}(k) + K(k)[y(k) - C(k)\hat{x}^{-}(k)]$ |
| **Erro de covariância** | $P(k) = [I - K(k)C(k)]P^{-}(k)$ |

A operação completa do filtro de Kalman, combinado a figura 4.1 e as tabelas 4.1 e 4.2 é mostrada na figura 4.2.

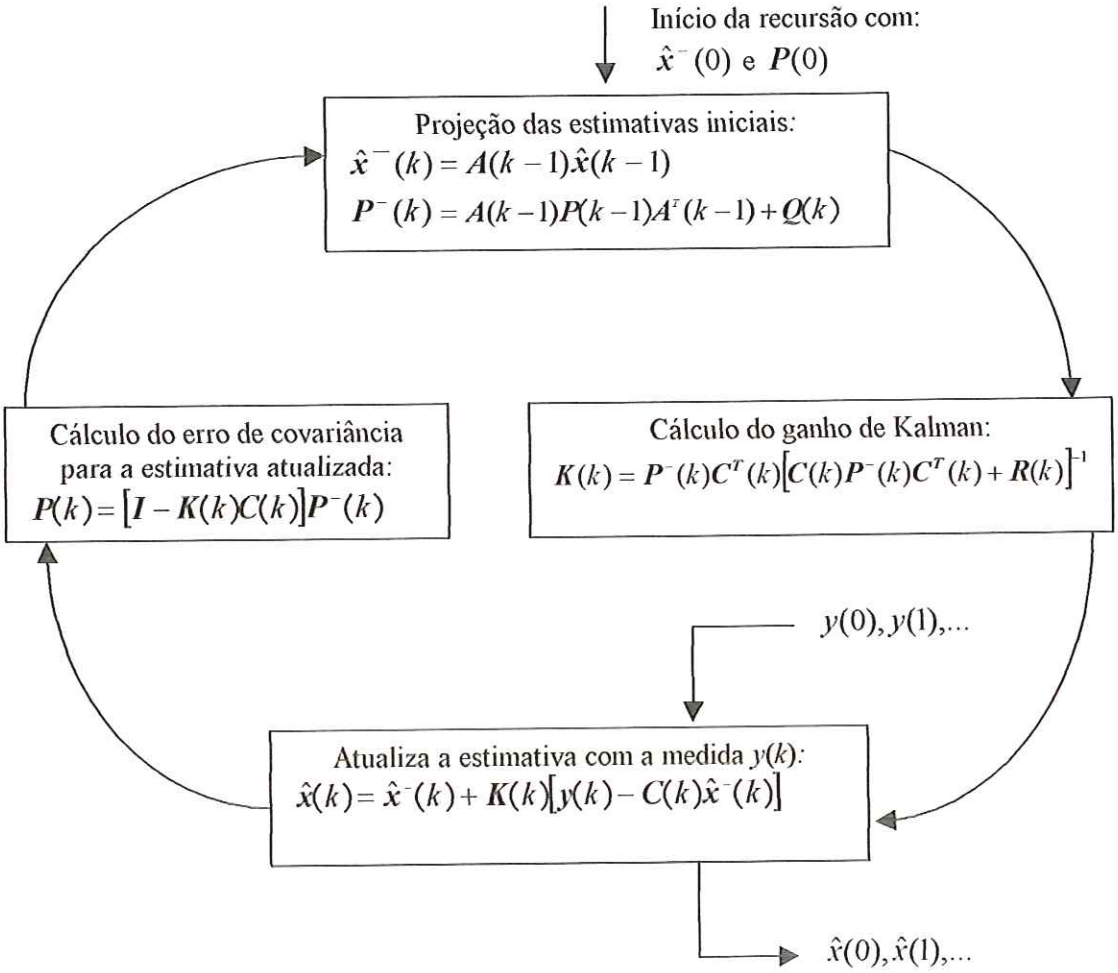



FIGURA 4.2 – Ciclo de operação do filtro de Kalman discreto.

## 4.4. Forma Alternativa do Filtro Discreto de Kalman

As equações recursivas que compõem o filtro de Kalman podem ser manipuladas de diversas maneiras MEDITCH (1969), GELB (1974), SORENSON

(1966). Nesta seção serão apresentadas algumas formas alternativas para algumas das expressões apresentadas na seção anterior.

Inicialmente será apresentada a forma reduzida para a equação da matriz de covariância. Para a demonstração das equações será omitido o índice de tempo $k$ para facilitar a manipulação.

Considere a eq. (4.25), que representa a matriz de covariância associada ao erro de estimação, como apresentado na seção anterior:

$$P = (I - KC)P^{-} \tag{4.26}$$

lembrando que o ganho de Kalman é dado pela eq. (4.20), tem-se:

$$K = P^{-}C^{T}\left[CP^{-}C^{T} + R\right]^{-1} \tag{4.27}$$

substituindo a eq. (4.27) em (4.26), obtém-se:

$$P = P^{-} - P^{-}C^{T}\left[CP^{-}C^{T} + R\right]^{-1}CP^{-} \tag{4.28}$$

Analisando a eq. (4.28) e considerando a existência das matrizes inversas de: $P$, $P^{-}$ e $R$, pode-se escrever a seguinte expressão para $P^{-1}$:

$$P^{-1} = \left[P^{-}\right]^{-1} + C^{T}R^{-1}C \tag{4.29}$$

A eq. (4.29) representa uma forma alternativa para a matriz de covariância associada ao erro de estimação. A sua justificativa pode ser dada da seguinte forma: multiplicando-se o lado direito das eq. (4.28) e (4.29), isto se reduz a matriz identidade.

$$
\begin{aligned}
&\left[P^- - P^- C^T (CP^- C^T + R)^{-1} CP^-\right]\left[(P^-)^{-1} + C^T R^{-1} C\right] \\
&= I - P^- C^T \left[(CP^- C^T + R)^{-1} - R^{-1} + (CP^- C^T + R)^{-1} CP^- C^T R^{-1}\right] C \\
&= I - P^- C^T \left[(CP^- C^T + R)^{-1} (I + CP^- C^T R^{-1}) - R^{-1}\right] C \\
&= I - P^- C^T \left[R^{-1} - R^{-1}\right] C \\
&= I
\end{aligned}
\tag{4.30}
$$

Também pode ser derivada uma expressão alternativa para a expressão do ganho de Kalman. Tomando-se a eq. (4.27) como ponto de partida tem-se:

$$K = P^- C^T \left[CP^- C^T + R\right]^{-1} \tag{4.31}$$

Inserindo os termos $PP^{-1}$ e $R^{-1}R$ não haverá mudança no ganho e a equação anterior será escrita da seguinte forma:

$$
\begin{aligned}
K &= PP^{-1} P^- C^T R^{-1} R \left[CP^- C^T + R\right]^{-1} \\
&= PP^{-1} P^- C^T R^{-1} \left[CP^- C^T R^{-1} + R + I\right]^{-1}
\end{aligned}
\tag{4.32}
$$

Utilizando a eq. (4.29), tem-se:

$$
\begin{aligned}
K &= P\left[\left(P^{-}\right)^{-1} + C^{T}R^{-1}C\right]P^{-}C^{T}R^{-1}\left[CP^{-}C^{T}R^{-1} + I\right]^{-1} \\
&= P\left[I + C^{T}R^{-1}CP^{-}\right]C^{T}R^{-1}\left[CP^{-}C^{T}R^{-1} + I\right]^{-1} \\
&= PC^{T}R^{-1}\left[I + CP^{-}C^{T}R^{-1}\right]\left[CP^{-}C^{T}R^{-1} + I\right]^{-1} \\
&= PC^{T}R^{-1} \qquad (4.33)
\end{aligned}
$$

Dessa forma, obtém-se duas expressões alternativas: uma para o ganho de Kalman e outra para a matriz de covariância associada ao erro de estimação. Tomando-se novamente as eqs. (4.29) e (4.30) e reescrevendo com os índices, obtém-se:

$$P^{-1}(k) = \left[P^{-}(k)\right]^{-1} + C^{T}(k)R^{-1}(k)C(k) \qquad (4.34)$$

$$K(k) = P(k)C^{T}(k)R^{-1}(k) \qquad (4.35)$$

O processamento do filtro de Kalman usando as formas alternativas é semelhante ao filtro de Kalman clássico e pode ser dado da seguinte maneira:

1. Cálculo das estimativas iniciais: $\hat{x}^{-}(0)$ e $P^{-}(0)$;
2. Início

   Para $k = 1, 2, 3, \ldots.$ faça

   $$\hat{x}^{-}(k) = A(k-1)\hat{x}(k-1)$$

   $$P^{-}(k) = A(k-1)P(k-1)A^{T}(k-1) + Q(k)$$

$$P^{-1}(k) = \left[P^{-}(k)\right]^{-1} + C^{T}(k)R^{-1}(k)C(k) \text{ e inverte para obter } P(k)$$

$$K(k) = P(k)C^{T}(k)R^{-1}(k)$$

$$\hat{x}(k) = \hat{x}^{-}(k) + K(k)\left[y(k) - C(k)\hat{x}^{-}(k)\right]$$

Fim

A figura 4.3 mostra de forma resumida as expressões que compõem a recursão para o filtro de Kalman, com as duas expressões alternativas derivadas.

Comparando a figura 4.3 com a figura 4.2 percebe-se que utilizando as expressões alternativas que a matriz de covariância associada ao erro de estimação pode ser calculada antes de se calcular o ganho, o que não ocorre quando se utiliza as expressões da figura 4.2. Outra alteração ocorre na expressão para o ganho que agora utiliza $P$ e dessa forma o cálculo $P$ deve ser realizado antes do cálculo do ganho, o que não ocorre com as equações da figura 4.2.

Pela análise da figura 4.3 é fácil de se notar que a recursão utilizando as expressões alternativas requer a inversão de matrizes quadradas para cada ciclo de operação do algoritmo do filtro de Kalman. Se a ordem das matrizes for grande, isso representa um problema computacional.

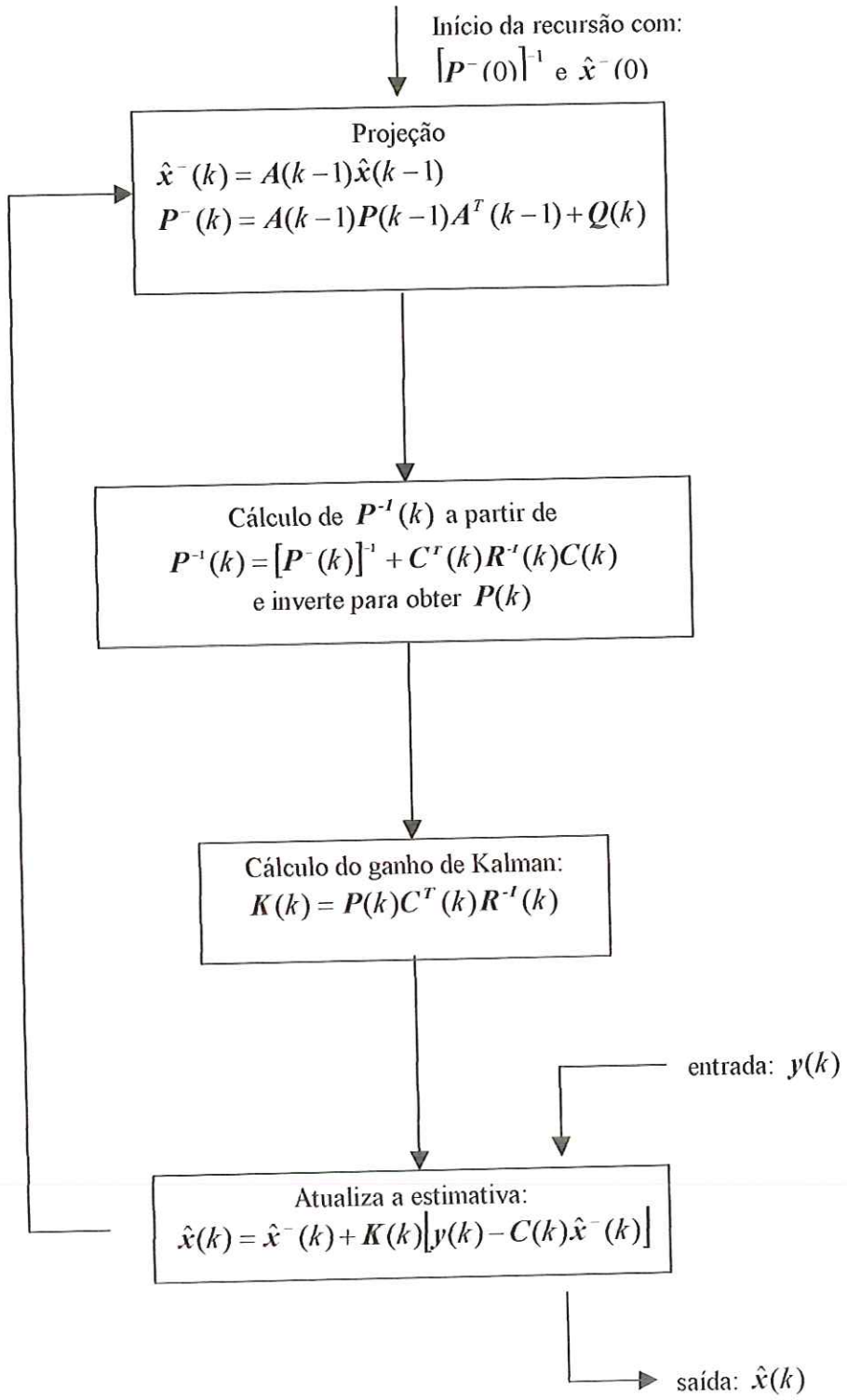



FIGURA 4.3 – Ciclo de operação do filtro de Kalman discreto usando formas alternativas.

## 4.5. Divergência do Filtro de Kalman

Desde a publicação do filtro de Kalman em 1960, este se tornou uma técnica de filtragem poderosa e largamente utilizada em várias áreas e ao mesmo tempo assunto de pesquisas científicas. Com a intensificação dos estudos sobre o filtro de Kalman e com várias aplicações, notou-se que em certas condições o filtro poderia divergir FITZGERALD (1971).

Por divergência considera-se que os erros de estimação passam a possuir um valor médio não nulo, divergindo dessa forma do valor esperado:

$$E[x - \hat{x}] = E[\text{erro de estimação}] = 0$$

e conseqüentemente a estimativa do estado diverge do estado verdadeiro do sistema, levando dessa forma a resultados bem diferentes do esperado.

Segundo GELB (1974) e FITZGERALD (1971), dois tipos de divergências podem ser encontradas:

- **Divergência aparente:** o valor do erro de estimação deixa de ter média nula e seu valor permanece constante (estacionário) no decorrer do tempo.

- **Divergência verdadeira:** neste caso, além do erro de estimação deixar de ter média nula, o seu valor cresce no decorrer do tempo.

De acordo com BROWN & HWANG (1992), duas são as principais causas da divergência do filtro:

- **Problemas numéricos:** este tipo de problema é inevitável quando operações aritméticas são realizadas usando computador digital. Os resultados estão sujeitos aos erros de arredondamento ("*roundoff errors*") devido ao comprimento de palavra utilizada pelo computador para fazer a representação. Dessa forma, os erros de arredondamento são cumulativos e podem distorcer os resultados finais do filtro de Kalman.

- **Problemas de modelagem:** também são responsáveis por divergências do filtro, uma vez que este é muito sensível aos erros de modelagem.

## 4.6. Filtro de Kalman Estendido

Quando o filtro de Kalman é aplicado a sistemas lineares, este se comporta de maneira ótima minimizando o valor do erro de estimação. O conjunto de equações apresentado nas seções anteriores considera tanto o modelo do sinal quanto a medida disponível como sendo lineares, produzindo dessa forma uma estimativa linear ótima. Em algumas situações a modelagem linear da dinâmica ou das medidas do sistema pode não ser adequada para representar o sistema NASCIMENTO JR. (1988).

Levando-se em conta essas observações, o conjunto de equações que compõem o filtro de Kalman é ajustado com o objetivo de solucionar o problema da estimação de estados em sistemas não lineares e portanto algumas alterações são

realizadas no algoritmo do filtro de Kalman MAYBECK (1982). De acordo com NASCIMENTO JR. (1988), a estrutura linear do filtro de Kalman é utilizada para a propagação e atualização da matriz de covariância dos erros de estimação. Por outro lado, a propagação da estimativa dos estados é realizada usando a estrutura não linear do sistema.

A estrutura inicial para o filtro de Kalman estendido é semelhante a estrutura para o filtro discreto linear com algumas modificações. Considere um sistema cuja dinâmica é representada pela seguinte equação não-linear:

$$x(k) = f(\mathbf{x}(k-1)) + w(k) \tag{4.36}$$

sendo $f(.)$ uma função não-linear que relaciona o estado $x$ no instante de tempo $k$ e $k-1$, $x(k)$ é o vetor de estados e $w(k)$ é o ruído de estado. Considere também que a única medida disponível do sistema é corrompida por ruído aditivo e é dada pela equação:

$$y(k) = h(x(k)) + v(k) \tag{4.37}$$

sendo $y(k)$ a única informação disponível, $h(.)$ uma função não-linear que representa as conexões entre o vetor de estados $x$ e a medida de ruído $v$. Tanto $w(k)$ quanto $v(k)$ são ruídos brancos gaussianos, definidos em tempo discreto, tendo média zero e variância dada por:

$$E\left[w(k)w^T(i)\right] = Q(k) \tag{4.38}$$

$$E\left[v(k)v^T(i)\right] = \boldsymbol{R}(k) \tag{4.39}$$

Dado o conjunto de equações que descrevem a dinâmica, as medidas do sistema e as estatísticas do ruído, o passo seguinte é calcular a estimativa de variância mínima do estado. Lembrando que, a estimativa de variância mínima é a média condicional do vetor de estados, indiferente de sua função densidade de probabilidade GELB (1974). Expandindo $f$ em uma série de Taylor sobre a estimativa corrente (média condicional) do vetor de estados, tem-se:

$$f(x(k)) = f(\hat{x}(k)) + \frac{\partial f}{\partial x}\Big|_{x=\hat{x}}(x-\hat{x}) + \frac{\partial^2 f}{\partial x^2}\Big|_{x=\hat{x}}\frac{(x-\hat{x})^2}{2!} + \ldots \tag{4.40}$$

É assumido que as derivadas parciais existem. Calculando a média em ambos os lados da equação acima, obtém-se:

$$\hat{f}(x(k)) = f(\hat{x}(k)) + 0 + 0 + \ldots \tag{4.41}$$

De acordo com GELB (1974), a equação para a propagação das estimativas do estado e do erro de covariância pode ser determinada considerando-se aproximações de primeira ordem para $\hat{f}(x(k))$. A equação para a propagação é dada pela seguinte relação:

$$\hat{x}^-(k) = f(\hat{x}(k)) \tag{4.42}$$

e a equação para a propagação do erro de covariância é dada por:

$$\boldsymbol{P}^{-}(k) = F(\hat{x}(k-1))\boldsymbol{P}(k-1) + \boldsymbol{P}(k-1)F^{T}(\hat{x}(k-1)) + \boldsymbol{Q}(k) \quad (4.43)$$

sendo $F(\hat{x}(k))$ a matriz cujos elementos $ij^{th}$ é dado por GELB (1974):

$$f_{ij}(\hat{x}(k)) = \frac{\partial f_i(x(k))}{\partial x_j(k)}\Big|_{x(k)=\hat{x}(k)} \quad (4.44)$$

As eqs. (4.42) e (4.43) são as equações para a propagação da estimativa corrente e da matriz de covariância associada ao erro de estimação. Quando uma nova medida $y(k)$ se torna disponível, a estimativa corrente é atualizada de forma semelhante ao filtro discreto de Kalman. A expressão para o cálculo da estimativa corrente é:

$$\hat{x}(k) = \hat{x}^{-}(k) + \boldsymbol{K}(k)\left[y(k) - \hat{h}(x(k))\right] \quad (4.45)$$

Analisando a eq. (4.45), pode-se observar claramente que para o cálculo da estimativa do estado $\hat{x}(k)$ é necessário o cálculo da matriz do ganho de Kalman. O ganho de Kalman apresenta a mesma função que no algoritmo para as equações lineares, ou seja, é responsável pela minimização do erro de estimação. O ganho de Kalman e a covariância do erro de estimação são representados pelas seguintes expressões:

$$K(k) = P^{-}(k)C^{T}(k)\hat{x}^{-}(k)\left[C(k)\hat{x}^{-}(k)P^{-}(k)C^{T}(k)\hat{x}^{-}(k) + R(k)\right]^{-1} \quad (4.46)$$

$$P(k) = \left[I - K(k)C(k)\hat{x}^{-}(k)\right]P^{-}(k) \quad (4.47)$$

Dessa forma, o procedimento usado pelo filtro de Kalman estendido é o seguinte:

1. Inicialmente calcula-se as estimativas iniciais do estado $\hat{x}^{-}(0)$ e da matriz de covariância do erro $P(0)$;
2. Início

   Para $k = 1, 2, 3, \ldots$ faça

$$\hat{x}^{-}(k) = f(\hat{x}(k-1))$$

$$F(\hat{x}(k)) = \left.\frac{\partial f(x(k))}{\partial x(k)}\right|_{x(k)=\hat{x}(k)}$$

$$P^{-}(k) = F(\hat{x}(k-1))P(k-1) + P(k-1)F^{T}(\hat{x}(k-1)) + Q(k)$$

$$H(\hat{x}(k)) = \left.\frac{\partial h(x(k))}{\partial x(k)}\right|_{x(k)=\hat{x}^{-}(k)}$$

$$K(k) = P^{-}(k)H^{T}\left(\hat{x}^{-}(k)\right)\left[H(\hat{x}^{-}(k))P^{-}(k)H^{T}(\hat{x}^{-}(k)) + R(k)\right]^{-1}$$

$$\hat{x}(k) = \hat{x}^{-}(k) + K(k)\left[y(k) - h(\hat{x}(k))\right]$$

$$P(k) = \left[I - K(k)H(\hat{x}^{-}(k))\right]P^{-}(k)$$

   Fim

O conjunto das equações que compõem o filtro de Kalman estendido é sumarizado nas tabelas 4.3 e 4.4.

TABELA 4.3 – Equações de propagação das estimativas para o filtro de Kalman estendido.

| | |
|---|---|
| **Propagação da estimativa do estado** | $\hat{x}^{-}(k) = f(\hat{x}(k-1))$ |
| **Propagação do Erro de Covariância** | $P^{-}(k) = F(\hat{x}(k-1))P(k-1) + P(k-1)F^{T}(\hat{x}(k-1)) + Q(k)$ |

TABELA 4.4 – Equações de atualização das estimativas para o filtro de Kalman estendido.

| | |
|---|---|
| **Ganho** | $K(k) = P^{-}(k)H^{T}(\hat{x}^{-}(k))\left[H(\hat{x}^{-}(k))P^{-}(k)H^{T}(\hat{x}^{-}(k)) + R(k)\right]^{-1}$ |
| **Estimativa do estado** | $\hat{x}(k) = \hat{x}^{-}(k) + K(k)[y(k) - h(\hat{x}(k))]$ |
| **Erro de covariância** | $P(k) = [I - K(k)H(\hat{x}^{-}(k))]P^{-}(k)$ |

O filtro de Kalman estendido tem sido aplicado em várias áreas e em diversas aplicações, sendo computacionalmente superior a filtros não lineares "ótimos" que se baseiam na teoria da filtragem não linear NASCIMENTO JR. (1988) e MAYBECK (1982). A operação completa do filtro de Kalman estendido é mostrada na figura 4.4.

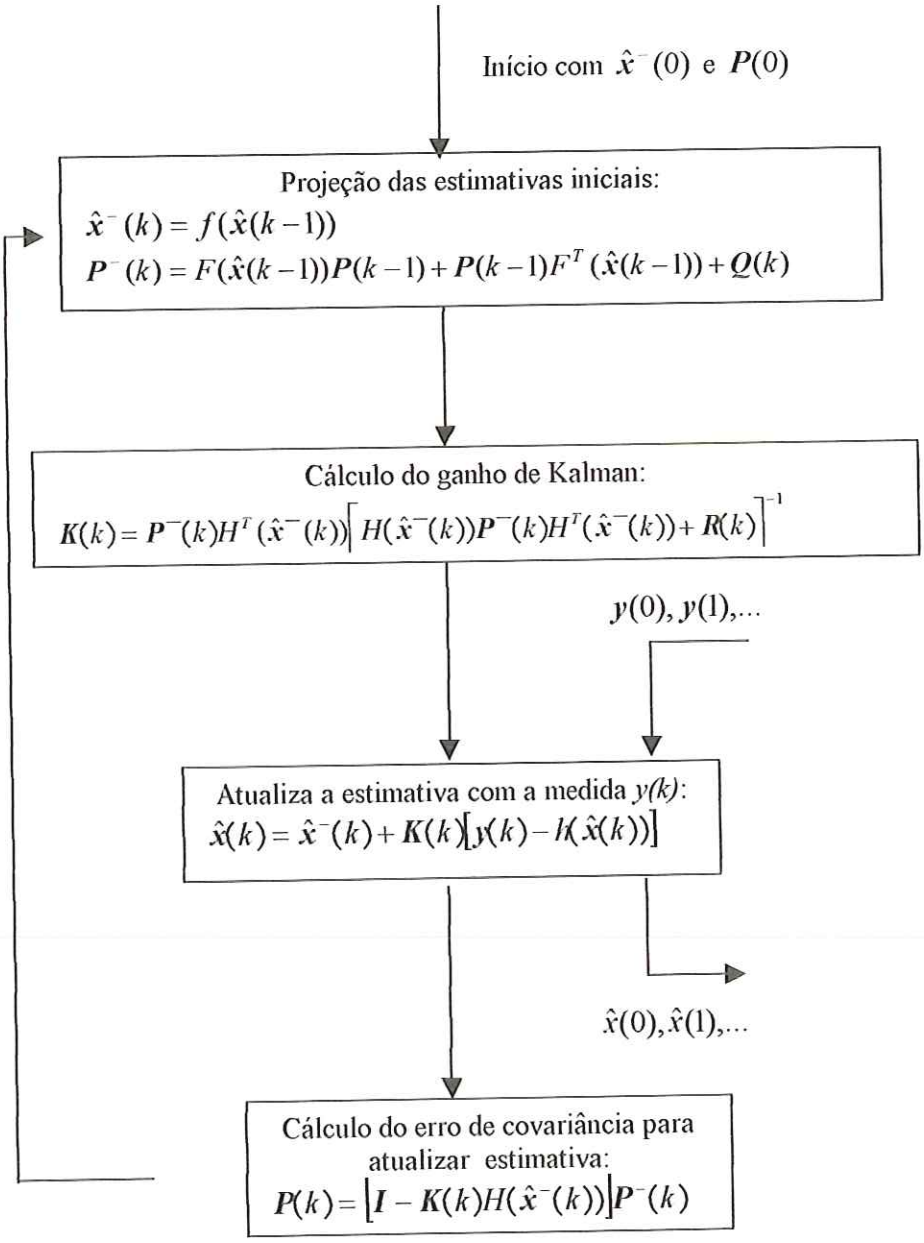



FIGURA 4.4 – Ciclo de operação do filtro de Kalman estendido.

## 4.7. Considerações Finais

Neste capítulo foi apresentada a teoria do filtro de Kalman aplicada a sistemas lineares e não lineares (filtro de Kalman estendido).

Além do desenvolvimento das equações que compõem ambos os filtros, o capítulo também apresenta algumas dificuldades que podem surgir quando as técnicas do filtro de Kalman são empregados, dentre elas: sintonização, condicionamento numérico e limitação de recursos computacionais.

Outro problema que ocorre com freqüência é a divergência, ou seja, o erro de estimação passa a ter um valor médio diferente de zero. A divergência pode ser dividida basicamente em dois grupos: divergência aparente e divergência verdadeira, e pode ser provocada por problemas numéricos e de modelagem.

Quando o filtro de Kalman é aplicado a sistemas lineares este produz estimativas ótimas minimizando o valor do erro de estimação. Para a aplicação em sistemas não lineares, o mesmo não ocorre, dessa forma, define-se o filtro de Kalman estendido e apresentou-se as modificações necessárias no conjunto de equações para que este possa ser utilizado em sistemas não lineares.

# Capítulo 5

# Redes Neurais Artificiais

## 5.1. Introdução

Neste capítulo é apresentada uma nova técnica que vem ganhando muita atenção entre os pesquisadores da área de processamento digital de sinais: as **Redes Neurais Artificiais**.

Inicialmente, o capítulo traz um breve histórico sobre o surgimento das redes neurais artificiais, o modelo do neurônio artificial e a descrição das principais arquiteturas empregadas.

O capítulo se encerra com o estudo de duas classes importantes de redes neurais: as redes com múltiplas camadas treinadas pelo algoritmo de retropropagação (*backpropagation*) e as redes de função de base radial (*radial basis function – RBF*).

## 5.2. Histórico

A criação de um modelo artificial para a representação do cérebro e a conseqüente reprodução das atividades neurais têm sido assunto de pesquisas há muitas décadas. O primeiro trabalho reportando um modelo artificial de um neurônio biológico data de 1943, publicado por um psiquiatra e neuroanatomista Warren McCulloch e por um matemático recém formado Walter Pitts MCCULLOCH & PITTS (1943).

Neste trabalho, os autores reportam a utilização de redes neurais artificiais e métodos de aprendizagem para que os neurônios possam executar atividades semelhantes às do cérebro humano.

Alguns anos depois do trabalho de McCulloch e Pitts, Donald Hebb HEBB (1949) publica seu trabalho apresentando uma nova regra de aprendizagem para as redes neurais. De acordo com esta regra, conhecida como regra de Hebb, a aprendizagem de redes neurais é conseguida através da variação dos pesos de entrada dos neurônios artificiais.

Em 1960, uma nova regra de aprendizagem surge na comunidade de RNA (redes neurais artificiais). Esta nova regra, conhecida como regra de Widrow-Hoff WIDROW & HOFF (1960) ou regra delta, é baseada no método do gradiente para a minimização do erro na saída de um neurônio.

Uma das primeiras redes neurais artificiais com capacidade de aprendizagem surge em 1958 publicado por Frank Rosenblatt ROSENBLATT (1958). O modelo de Rosenblatt era baseado no modelo proposto por McCulloch e Pitts e foi chamado de perceptron. O perceptron poderia ser treinado para classificar certos tipos de padrões linearmente separáveis, ou seja, problemas cuja solução pode ser obtida dividindo-se o espaço de entrada em duas regiões através de uma reta. O perceptron de Rosenblatt era formado por três camadas: a camada de entrada, que recebe as informações do meio externo; a segunda camada ou camada oculta, que recebe os impulsos da primeira camada através dos pesos e a camada de saída que gera a saída da rede neural.

Alguns anos mais tarde, o trabalho publicado por Minsky e Papert, MINSKY & PAPERT (1969) causa grande repercussão na comunidade de RNAs. Neste trabalho, Minsky e Papert mostram algumas tarefas que o perceptron não era capaz de executar, mas que representam uma grande classe que não pode ser desprezada. Este fato foi responsável por uma baixa produção científica na década de 70.

A década de 80 é marcada pelo ressurgimento das redes neurais na comunidade internacional devido a vários fatores. O primeiro deles foi a publicação do trabalho de John Hopfield em 1982 HOPFIELD (1982) que mostrou a relação entre as redes recorrentes associativas e sistemas físicos, abrindo dessa forma espaço para teorias da Física estudar tais modelos. Outro fator responsável pela retomada de interesse nas RNAs foi o avanço da tecnologia, principalmente da microeletrônica, permitindo dessa forma a realização física de alguns modelos de RNAs.

## 5.3 Neurônios Biológicos e Artificiais

A grande motivação para o estudo e desenvolvimento das RNAs está inspirada no cérebro humano. O cérebro por sua vez é formado por neurônios, tendo aproximadamente $10^{11}$ neurônios.

O objetivo de uma rede neural é tentar reproduzir as funções das redes biológicas, tais como: percepção, reconhecer e relacionar padrões, usar e armazenar conhecimento por experiência e interpretar observações BRAGA *et. al* (1999).

O neurônio pode ser dividido basicamente em três partes: o corpo celular ou soma, o axônio e os dendritos. A figura 5.1 mostra o modelo anatômico simplificado de um neurônio biológico.

Os dendritos são responsáveis por receber as informações (impulsos nervosos) a partir dos axônios de outros neurônios. Estas informações recebidas são processadas pelo corpo celular, que por sua vez gera novos impulsos que são transmitidos a outros neurônios através do axônio. Os axônios podem ser considerados as "linhas de transmissão" do neurônio. Estes axônios estão ligados aos dendritos de outros neurônios e o ponto de contato é chamado de sinapse. É através das sinapses que os neurônios se unem formando redes neurais.

Uma rede neural assim como o cérebro humano, pode ser considerada um processador paralelo e distribuído, ou seja, a informação é processada paralelamente por vários neurônios e depois distribuída para outros neurônios para gerar a saída da rede.

A aquisição de "conhecimento" pela rede neural é realizada através de um processo de aprendizagem e este "conhecimento" é armazenado pelos pesos

sinápticos. Este processo se dá através de um algoritmo de aprendizado através do qual os pesos sinápticos são modificados objetivando saídas desejadas.

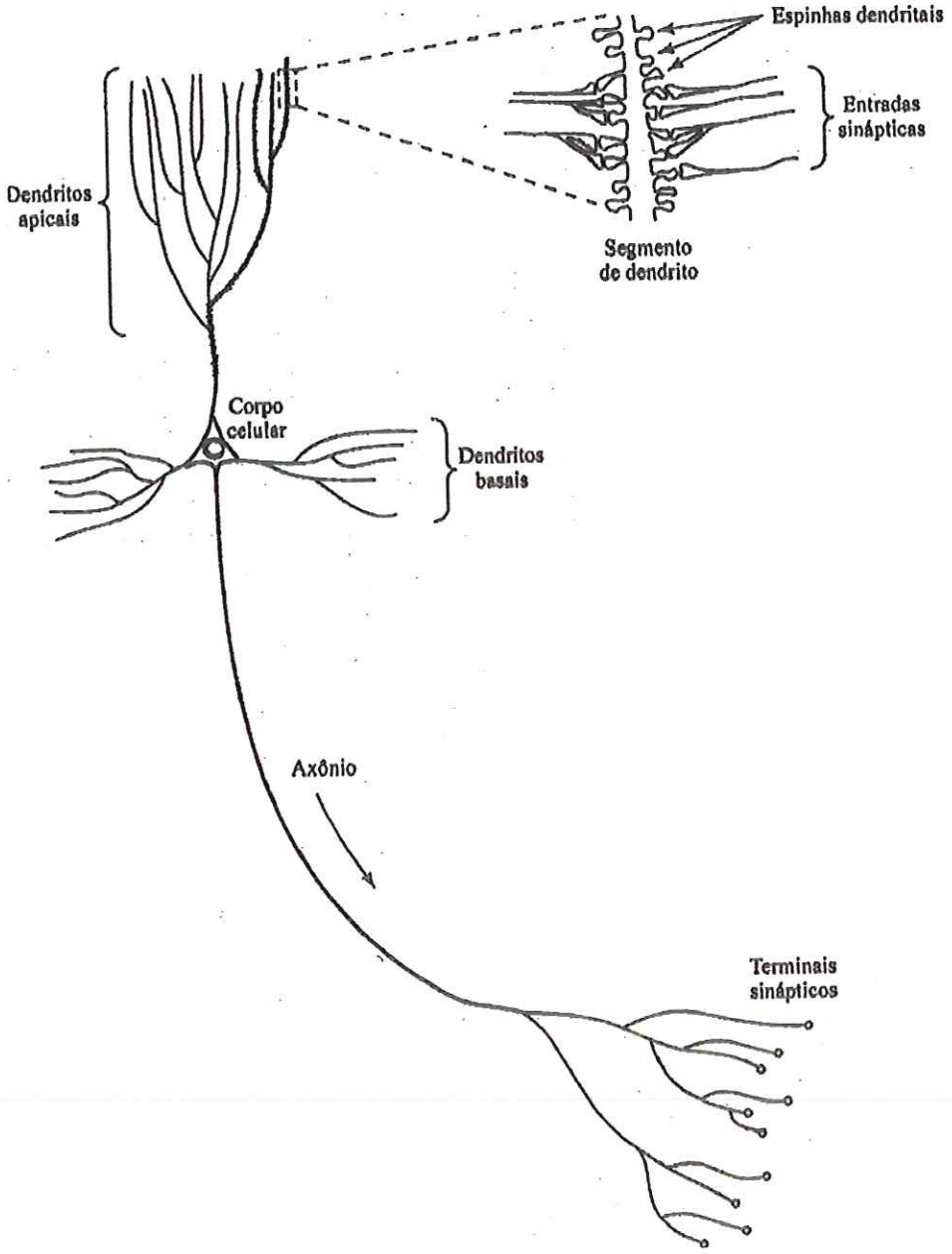



FIGURA 5.1 – Modelo anatômico simplificado de um neurônio biológico.

Como exposto na seção anterior, o neurônio constitui a unidade básico fundamental para a operação de uma rede neural. A figura 5.2 mostra o modelo de

um neurônio artificial.

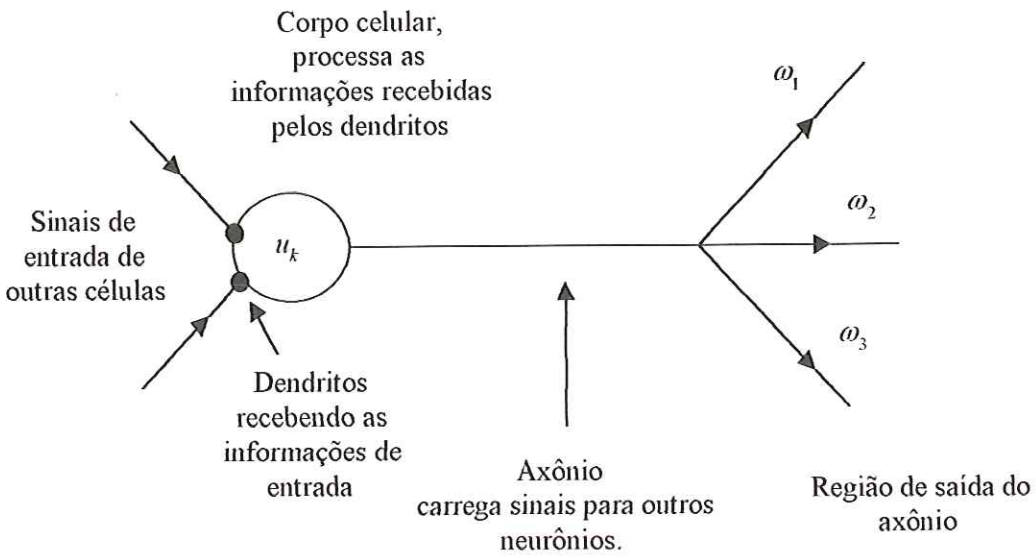



FIGURA 5.2 – Modelo de um neurônio artificial.

Analisando a figura 5.2, pode-se destacar três elementos básicos, como aqueles destacados para o neurônio biológico:

- **região de entrada:** região que recebe informações de outros neurônios;

- **corpo celular:** região onde ocorre o processamento das informações que chegam até o neurônio; as operações descritas nesta região constituem uma combinação linear;

- **função de ativação ($u_k$):** esta função é responsável por limitar a amplitude da saída do neurônio.

A figura 5.3 descreve o modelo não-linear de um neurônio.

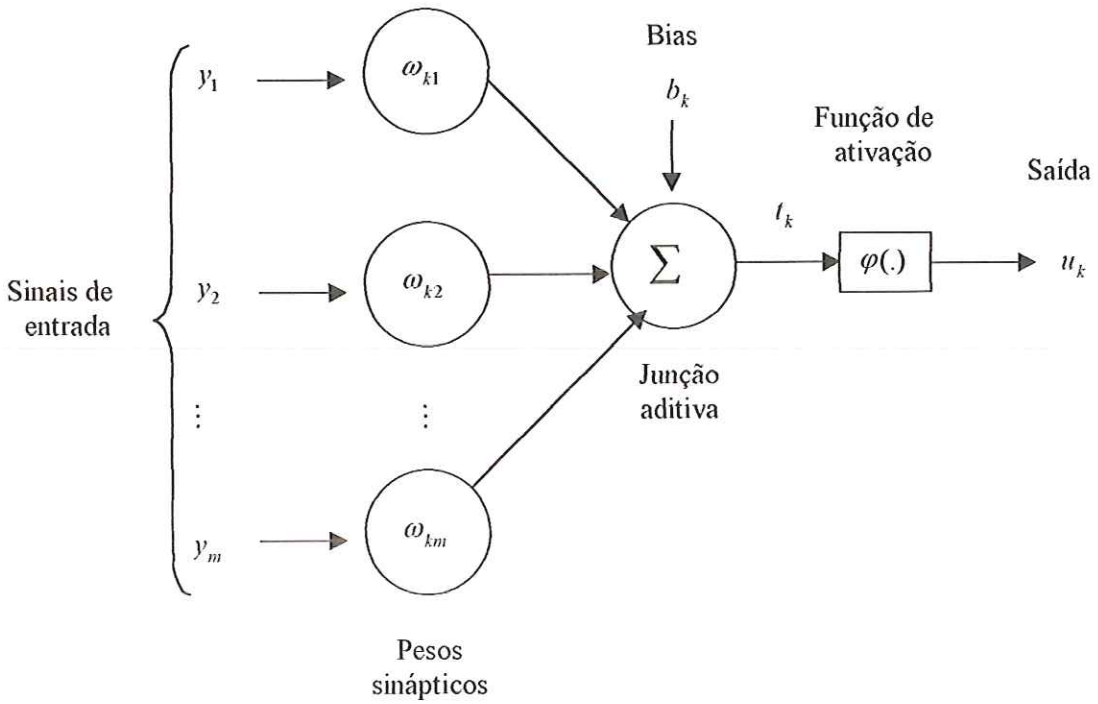



FIGURA 5.3 – Modelo não linear de um neurônio artificial.

sendo $y_1, y_2, \ldots y_j$ os sinais de entrada para a rede neural; $\omega_{k1}, \omega_{k2}, \ldots, \omega_{kj}$ são os pesos sinápticos que conectam o neurônio $j$ ao neurônio $k$; $u_k$ é a combinação linear para o neurônio $k$.

## 5.4. Principais Arquiteturas de Redes Neurais

A arquitetura de uma rede neural refere-se ao modo pelo qual os neurônios da rede estão estruturados. Na literatura clássica sobre redes neurais distinguem-se basicamente três classes de arquiteturas de redes: redes com camada única, redes com múltiplas camadas e redes recorrentes.

### 5.4.1. Redes com Camada Única

A figura 5.4 mostra o esquema de uma rede com camada única. Esta rede é dita ter camada única porque a camada de entrada não é contada por não haver processamento nela.

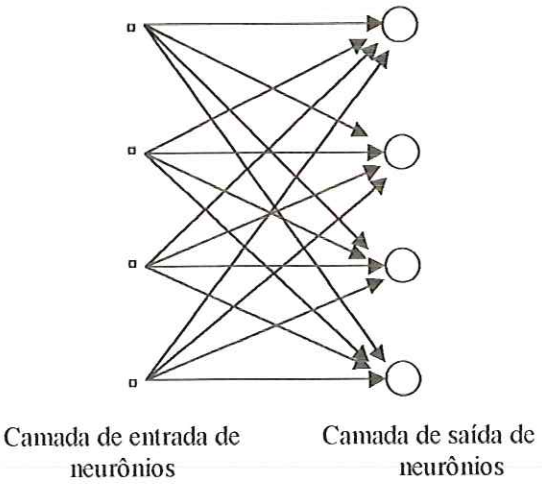



FIGURA 5.4 – Rede neural com apenas uma camada de neurônios.

### 5.4.2. Redes de Múltiplas Camadas

Nesta estrutura, a rede neural apresenta uma ou mais camadas ocultas, que são responsáveis pelo processamento da rede neural. Além do processamento, as

camadas ocultas (ou unidades ocultas) também são responsáveis pela extração de estatísticas de ordem elevada que são utilizadas para mapear entrada-saída. A figura 5.5 ilustra uma rede com múltiplas camadas.

Como pode ser observado na figura 5.5, a camada de entrada apenas recebe o sinal a ser processado do meio externo e passa esse sinal para a primeira camada oculta. Os sinais de saída da primeira camada oculta são utilizados como entrada para a próxima camada oculta e assim por diante até a camada de saída da rede neural. O sinal de saída da última camada da rede neural representa a resposta da rede para o sinal aplicado na entrada.

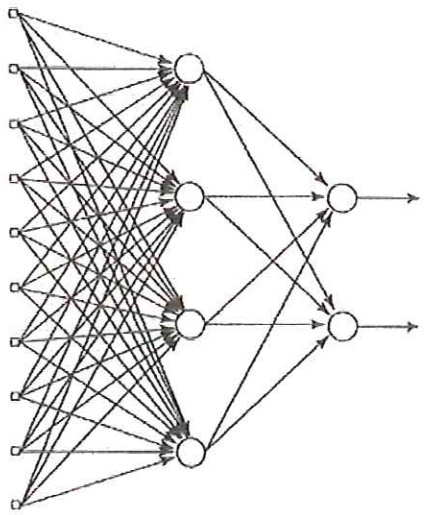


FIGURA 5.5 – Rede neural com múltiplas camadas.

As redes neurais com múltiplas camadas podem ser classificadas em: redes neurais totalmente conectadas e redes neurais parcialmente conectadas. Nas redes neurais totalmente conectadas, cada um dos neurônios de uma camada está conectado com todos os outros neurônios da camada seguinte. Para as redes

parcialmente conectadas, nem todos os neurônios de uma determinada camada estão conectados aos neurônios da camada seguinte.

### 5.4.3. Redes Recorrentes

A principal característica de uma rede neural recorrente é a presença de pelo menos um laço de realimentação. As redes neurais recorrentes podem apresentar apenas uma ou várias camadas de neurônios. Além do número de camadas pode ocorrer também a presença de laços de auto-realimentação; auto-realimentação se refere a situação onde a saída de um neurônio é realimentada para a sua própria entrada.

A figura 5.6 mostra o esquema de uma rede neural recorrente sem laços de auto-realimentação e sem neurônios ocultos. A figura 5.7 mostra uma rede neural recorrente com neurônios ocultos e com laços de realimentação.

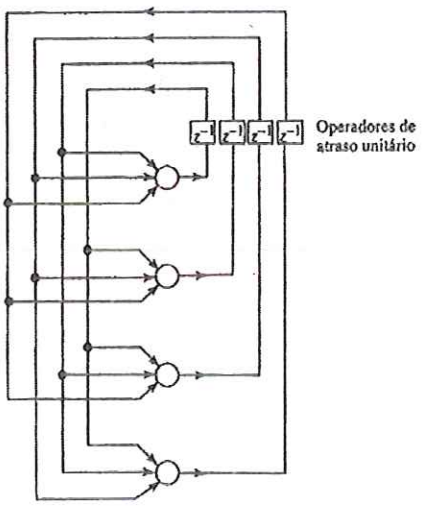



FIGURA 5.6 – Rede neural recorrente sem neurônios ocultos e sem laços de auto-realimentação.

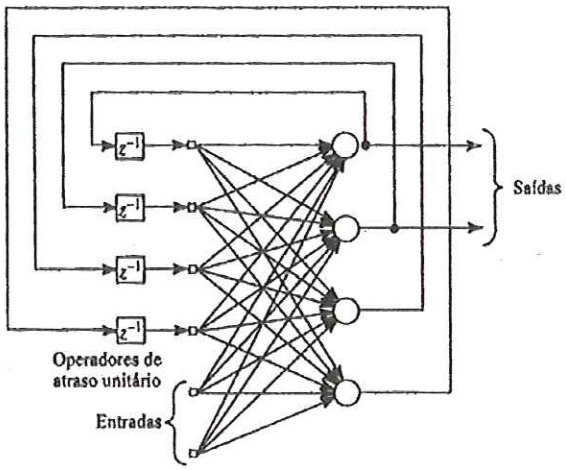



FIGURA 5.7 – Rede neural recorrente com neurônios ocultos e laços de auto-realimentação.

## 5.5. Aprendizagem em Redes Neurais Artificiais

As principais características das redes neurais responsáveis por seu próprio desenvolvimento são sua capacidade de aprendizagem e generalização. Segundo MENDEL & MCCLAREN (1970), aprendizagem é um processo pelo qual os parâmetros livres (pesos sinápticos) de uma rede são adaptados através de um processo de estimulação pelo ambiente no qual a rede está inserida.

Vários algoritmos de aprendizagem foram desenvolvidos desde o surgimento das redes neurais e podem ser classificados em dois grupo: aprendizagem supervisionada e aprendizagem não-supervisionada.

### 5.5.1. Aprendizagem Supervisionada

O método de aprendizagem supervisionada é o método mais comum

utilizado para o treinamento de redes neurais. Neste método, tanto a entrada quanto a saída desejada para a rede são fornecidas por um professor, e o algoritmo de aprendizagem procura criar uma relação entre a entrada e a saída fornecidas através da adaptação dos pesos da rede.

Para cada padrão de entrada fornecido à rede, esta calcula a saída e compara com a saída desejada gerando dessa forma um sinal de erro. Este sinal de erro por sua vez é utilizado para ajustar os pesos da rede, com o objetivo de aproximar o sinal de saída calculado pela rede com o sinal desejado.

A grande desvantagem da aprendizagem supervisionada é que na ausência do professor a rede neural não conseguirá aprender novas situações. Como exemplos de aprendizagem supervisionada pode-se citar a regra delta WIDROW & HOFF (1960) e o algoritmo de retropropagação RUMELHART *et. al.* (1986), que é uma generalização da regra delta para redes com múltiplas camadas.

### 5.5.2. Aprendizagem Não-Supervisionada

Este método de aprendizagem se comporta de forma diferente da aprendizagem supervisionada. Como o próprio nome sugere, não existe neste método um professor para supervisionar o aprendizado da rede. Os padrões de entrada e saída que eram fornecidos para o treinamento da rede agora não estão mais disponíveis.

Para o treinamento, a rede recebe apenas os padrões de entrada e estabelece regularidades estatísticas da entrada dos dados, desenvolvendo dessa forma uma habilidade de formar representações internas para codificar características

da entrada e criar novas classes BRAGA *et. al.* (1999). Este tipo de aprendizagem é utilizado quando os dados de entrada da rede neural apresentam redundâncias. Um exemplo de aprendizagem não-supervisionada é a regra de aprendizagem competitiva.

### 5.5.3. Aprendizagem por Reforço

A aprendizagem por reforço pode ser considerada um caso particular de aprendizagem supervisionada. Em ambos os casos existe uma interação com o meio externo. Cada um dos métodos de aprendizagem utiliza uma medida de desempenho, que é a principal diferença entre eles. Na aprendizagem supervisionada, o desempenho do sistema é baseado em um sinal de erro que é gerado comparando-se a saída corrente (calculada) da rede com a saída desejada. Para a aprendizagem por reforço, não existe um sinal de erro utilizado para ajustar os parâmetros da rede e sim um sinal de reforço gerado por um *crítico*. O crítico recebe um sinal de reforço (sinal de reforço primário) do ambiente e converte em um sinal de reforço heurístico, indicando se a saída está correta ou não. A figura 5.8 ilustra a aprendizagem por reforço.

O sinal de reforço (sim/não) enviado para a rede neural em resposta a uma ação produzida pela rede apenas avalia o desempenho da rede. Em momento algum, o sinal de reforço instrui a rede em como melhorar o desempenho.

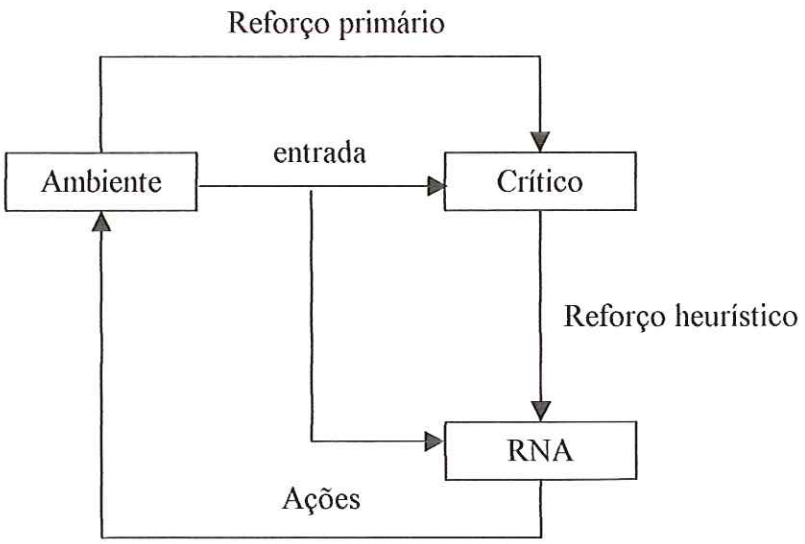



FIGURA 5.8 – Modelo de aprendizagem por reforço HAYKIN (1994).

## 5.6. Redes com Múltiplas Camadas

Na seção 5.4 foram apresentadas as principais arquiteturas de redes, dentre elas: redes com camada única, redes com múltiplas camadas e redes recorrentes. As redes com camada única (camada computacional consistindo de um único neurônio) somente realizam classificação de padrões quando estes são linearmente separáveis, daí sua limitação.

Nesta seção serão apresentadas as redes com múltiplas camadas (*MLP – Multilayer Perceptrons*) treinadas pelo algoritmo de retropropagação, que são largamente empregadas em processamento digital de sinais de voz.

As redes de múltiplas camadas surgiram para resolver problemas que os perceptrons de camada única não conseguiam resolver (problemas que não são linearmente separáveis). A figura 5.9 apresenta uma rede de múltiplas camadas com duas camadas ocultas e uma camada de saída. Como também já foi mencionado na

seção 5.4, as camadas ocultas são responsáveis pela extração de estatísticas de ordem elevada. Segundo CYBENKO (1989), uma rede com duas camadas intermediárias pode implementar qualquer função, linear ou não.

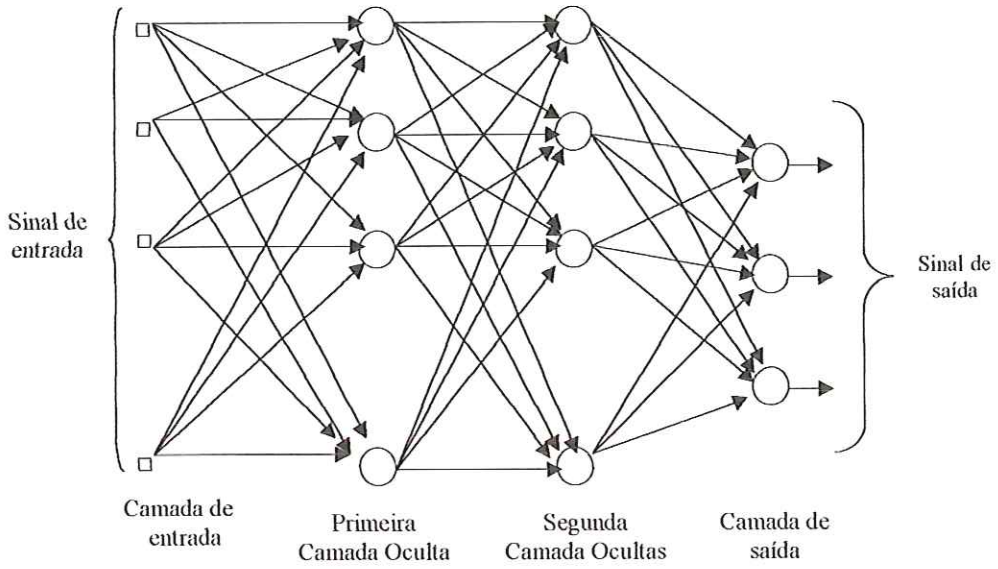



FIGURA 5.9 – Rede neural com duas camadas ocultas e uma camada de saída.

O treinamento dos perceptrons com múltiplas camadas é realizado pelo algoritmo de retropropagação, também conhecido como propagação do erro (*error backpropagation*).

### 5.6.1. Algoritmo de Retropropagação

O treinamento de uma rede neural pelo algoritmo de retropropagação é realizado em dois passos: o passo para a frente (propagação) e o passo para trás (retropropagação).

No passo para frente, o vetor de entrada aplicado à camada de entrada se propaga pelas camadas ocultas até a camada de saída, produzindo a saída corrente (calculada). Considere o vetor $y$ de comprimento $m$ como sendo o sinal de entrada da rede, portanto a saída produzida no neurônio $j$ é calculada pela equação:

$$u_j(k) = \varphi(t_j(k)) \tag{5.2}$$

sendo $\varphi(.)$ a função de ativação do tipo sigmoidal e $t_j(k)$ o campo local do neurônio $j$, definido por:

$$t_j(k) = \sum_{i=0}^{m-1} \omega_{ji}(k) u_i(k) \tag{5.3}$$

sendo $\omega_{ji}$ os pesos sinápticos que conectam o neurônio $i$ ao neurônio $j$ e $u_i$ é o sinal de entrada do neurônio $j$, ou seja, é o sinal produzido na saída do neurônio $i$. Caso o neurônio $j$ esteja na primeira camada oculta da rede, sua entrada será o vetor $y$:

$$u_j(k) = y_i(k) \tag{5.4}$$

sendo $y_i(k)$ o *i-ésimo* elemento do vetor de entrada. Da mesma forma, caso o neurônio $j$ esteja na camada de saída da rede, $u_j$ será a saída da rede neural, e será representada da seguinte forma:

$$u_j(k) = o_j(k) \tag{5.5}$$

O sinal de saída $o_j(k)$ é comparado com a saída desejada $d_j(k)$ gerando o sinal de erro $e_j(k)$ para o neurônio $j$, definido por:

$$e_j(k) = o_j(k) - d_j(k) \tag{5.6}$$

Durante a fase de propagação, os pesos sinápticos da rede neural permanecem inalterados. Com o cálculo do sinal de erro na saída da rede, termina a fase de propagação e inicia-se a fase de retropropagação, através da qual o sinal de erro é retropropagado entre as camadas ocultas e os pesos sinápticos são alterados. A cada camada de retropropagação é necessário calcular o gradiente ($\delta$) para cada neurônio oculto.

O cálculo do gradiente $\delta$ envolve a derivada da função de ativação ALEKSANDER & MORTON (1990), HAYKIN (1999). Para a função de ativação sigmoidal logística, o valor do $\delta$ é dado por:

$$\begin{aligned} \delta_j(k) &= e_j(k)\varphi'_j(u_j(k)) \\ &= e_j(k)o_j(k)[1 - o_j(k)] \end{aligned} \tag{5.7}$$

sendo a eq. (5.7) utilizada quando o neurônio $j$ é um neurônio localizado na camada de saída. Para os neurônios localizados nas camadas ocultas, o cálculo do gradiente é dado por:

$$\delta_j(k) = \varphi'_j\left(u_j(k)\right)\sum_i \delta_i(k)\omega_{ij}(k)$$

$$= u_j(k)\left[1 - u_j(k)\right]\sum_i \delta_i(k)\omega_{ij}(k) \qquad (5.8)$$

sendo $u_j(k)$ a saída para o neurônio $j$ localizado em uma camada oculta, $\omega_{ij}(k)$ é o peso sináptico que conecta o neurônio $j$ ao neurônio $i$ e $\delta_i(k)$ é o gradiente local para o neurônio $i$.

Depois de calculados os gradientes locais para os neurônios da camada oculta, a atualização dos pesos sinápticos pode ser realizada pela seguinte equação:

$$\omega_{ji}(k+1) = \omega_{ji}(k) + \alpha\left[\omega_{ji}(k-1)\right] + \eta\delta_j(k)u_i(k) \qquad (5.9)$$

sendo $\eta$ a taxa de aprendizagem e $\alpha$ é a constante de momento.

A taxa de aprendizagem $\eta$ é um parâmetro utilizado para controlar as variações dos pesos sinápticos, ou seja, quanto menor a taxa de aprendizagem, menor serão as variações dos pesos sinápticos da rede, caso contrário, grandes modificações ocorrerão nos pesos sinápticos podendo tornar a rede instável. Para evitar que a rede se torne instável com o aumento do parâmetro da taxa de aprendizagem, pode-se incluir na rede o termo de momento. O termo de momento aumenta a velocidade de aprendizagem dos padrões, reduzindo o problema da instabilidade.

### 5.6.2. Resumo do Algoritmo de Retropropagação

1. **Inicialização:** inicializar os pesos e os limiares da rede;

2. **Apresentação dos exemplos de treinamento:** escolha um exemplo de treinamento aleatório e apresente a rede;

3. **Propagação:** para o par de entrada $(y(k), d(k))$, sendo $y(k)$ o vetor de entrada e $d(k)$ a resposta desejada para o padrão $y(k)$, calcule os campos locais utilizando a eq. (5.3). Caso o neurônio $j$ esteja na primeira camada, substituir a eq. (5.3) pela eq. (5.4). Em seguida, calcule a saída para cada neurônio utilizando a eq. (5.2). Caso o neurônio $j$ seja um neurônio de saída, sua saída será calculada pela eq. (5.5). Depois de calculada todas as saídas para os neurônios, calcular os erros para cada neurônio da camada de saída com auxílio da eq. (5.6) ;

4. **Retropropagação:** calcule os gradientes locais $(\delta)$ para cada neurônio da rede. Se o neurônio $j$ estiver na camada de saída, utilize a eq. (5.7), caso contrário utilize a eq. (5.8). Depois de calculado o gradiente local, ajuste os pesos sinápticos com a eq. (5.9);

5. **Iteração:** apresente novos padrões para a rede e execute os passos de 2 a 4 até que um critério de parada seja satisfeito.

## 5.7. Redes de Função de Base Radial

Uma grande quantidade de estudos e aplicações realizados com as redes neurais MLP com função de ativação do tipo sigmoidal têm assegurado a efetiva validade e contribuição do paradigma. Entretanto, estudos têm mostrado que as MLP não estão livres de problemas. Em particular, elas requerem um extenso tempo de treinamento, são suscetíveis a problemas de mínimos locais e suas performances dependem da arquitetura utilizada. As redes neurais RBF surgem então como uma alternativa para competir com o espaço de pesquisas aberto pelas redes MLP.

Inicialmente as redes RBF foram utilizadas em 1988 por BROOMHEAD & LOWE (1988).

A arquitetura empregada nas redes RBF é semelhante à empregada nas redes MLP, ou seja, apresentam camadas de entrada, camadas escondidas (intermediárias) e camadas de saída. Geralmente, o número de camadas escondidas é igual a um, embora redes RBF com mais de uma camada escondida já tenham sido empregadas na literatura HE & LAPEDES (1991). A figura 5.10 mostra a arquitetura típica de uma rede RBF.

O número de neurônios em cada camada pode variar de acordo com a aplicação. Embora a figura 5.10 esteja apresentando uma rede com apenas um neurônio de saída, esta pode apresentar qualquer quantidade de neurônios.

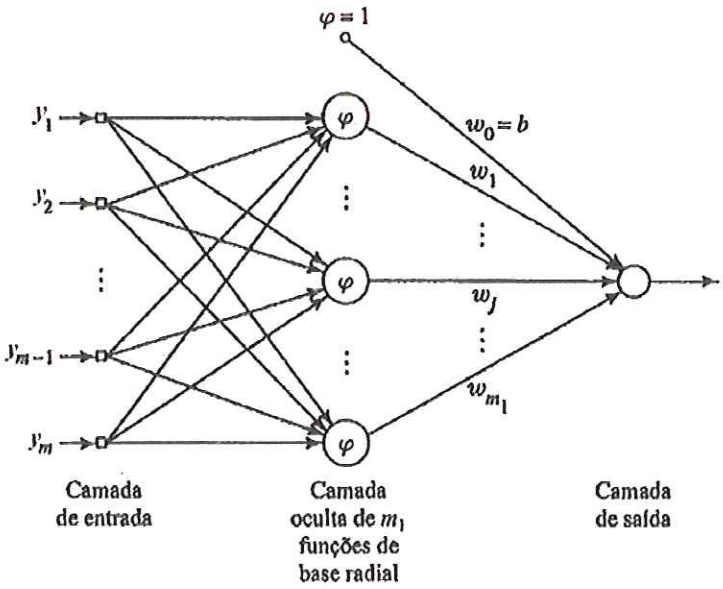



FIGURA 5.10 – Rede de função de base radial.

Com relação à função de ativação, as redes RBF utilizam funções radiais. As funções radiais são uma classe de funções onde seus valores podem aumentar (ou diminuir) em relação a distância de um ponto central MARAR (1997). Esta classe de funções é caracterizada por dois parâmetros: o centro e o raio. Uma função radial típica é a função gaussiana, que é dada pela seguinte equação:

$$\varphi(y) = e^{-\frac{(y-c)^2}{2r^2}} \tag{5.10}$$

sendo $y$ o vetor de entrada, $c$ é o centro da função e $r$ é o raio. Outras funções radiais também são empregadas, dentre elas a função multiquadrática e a multiquadrática inversa.

A função multiquadrática é dada pela equação:

$$\varphi(y) = \frac{\sqrt{r^2 + (y-c)^2}}{r} \tag{5.11}$$

e a multiquadrática inversa é dada pela equação:

$$\varphi(y) = \frac{r}{\sqrt{r^2 + (y-c)^2}} \tag{5.12}$$

A saída para um neurônio $j$ é dada pela expressão:

$$u_j = \sum \omega_{ji} \varphi(\|y - \mu_i\|) \tag{5.13}$$

sendo $\omega_{ji}$ os pesos sinápticos que conectam o neurônio $i$ ao neurônio $j$, $\varphi(.)$ é um conjunto de funções radiais e $\|.\|$ denota a norma euclidiana.

### 5.7.1. Treinamento das Redes RBF

O processo de treinamento de uma rede RBF pode ser realizado por vários métodos. Basicamente, os métodos de aprendizagem são divididos em duas fases. Na primeira fase são determinados os raios e os centros das funções de base radial e na segunda fase do processo de aprendizagem são ajustados os pesos.

- **Primeira fase do treinamento:**

Nesta primeira fase são determinados os números de funções de base radial, os centros e os raios. Para a determinação dos centros, várias abordagens podem ser empregadas. Uma dessas abordagens consiste em selecionar aleatoriamente os centros a partir do conjunto de treinamento. Uma segunda abordagem considera a determinação dos centros e dos raios como um processo de aprendizagem supervisionada baseado na correção de erro. A terceira abordagem emprega o algoritmo de agrupamento de *k*-médias ("*k-means*") DELLER (1993), HAYKIN (1999).

Para a aplicação do algoritmo *k*-médias, inicialmente o espaço de padrões de entrada é dividido em *k* regiões. Escolhe-se *k* vetores para serem os centros de *k* grupos. Os outros vetores do conjunto de treinamento são atribuídos ao grupo que tiver o centro mais próximo. Depois que os vetores forem atribuídos aos grupos que tiverem o centro mais próximo, os centros dos grupos são recalculados. Um novo conjunto de treinamento é apresentado ao algoritmo, calcula-se novamente a distância em relação aos grupos e em seguida os centros são recalculados. O algoritmo *k*-médias será encerrado quando os centros não sofrerem mais alterações. A tabela 5.1 descreve a recursão do algoritmo *k*-médias.

Depois de determinados os centros das funções de base radial, outro parâmetro a ser determinado na primeira fase de treinamento são os raios (larguras) das funções. De acordo com MOODY & DARKEN (1989), cada raio pode ser determinado como a média entre todas as distâncias de cada centro da unidade $n_i$ e o centro da unidade $n_j$ mais próxima. Em termos matemáticos a equação é representada por:

$$\sigma_j = \frac{1}{m}\sum_{i=1}^{m}\left\|\mu_j - \mu_i\right\| \tag{5.14}$$

sendo que $\mu_j$ é o centro mais próximo de $\mu_i$.

Tabela 5.1 – Algoritmo *k*-médias.

**Inicialização:** escolha *k* partições do vetor de treinamento *y*, denotados por $\Lambda_k$, *k* = 1, 2, 3, ..., *K* e calcule o vetor média de cada partição $\bar{y}_k$, *k* = 1, 2, 3, ..., *K*.

**Recursão**

1. Para cada vetor do conjunto de treinamento, atribua-o a partição que tiver a menor distância em relação ao centro

$$k = \text{argmin}\ \left\|y, \bar{y}_k\right\|$$

2. Recalcule os centros das partições.

3. Retorne ao passo 1 enquanto os centros sofrem alterações.

- **2ª fase do treinamento:**

Nesta fase, os pesos sinápticos que conectam cada neurônio da camada escondida com os neurônios da camada de saída são ajustados. Os ajustes dos pesos podem ser determinados pela regra delta WIDROW & HOFF (1960).

$$\omega(k+1) = \omega(k) + \eta e(k) x(k) \tag{5.15}$$

sendo $\eta$ o parâmetro da taxa de aprendizagem.

## 5.8. Comparação entre Redes RBF e Redes MLP

As redes RBF e as redes MLP são exemplos de redes alimentadas adiante (*feedforward*) e não lineares. Ambas são consideradas aproximadores universais de funções. Na prática, sempre existe uma rede RBF capaz de imitar precisamente uma MLP específica, ou vice-versa. Apesar dessa equivalência, algumas diferenças podem ser destacasas:

- as duas redes apresentam número de camadas intermediárias diferentes. Enquanto as redes MLP podem apresentar qualquer número de camadas intermediárias, as redes RBF apresentam apenas uma camada;
- o argumento para a função de ativação nas redes RBF é a norma (distância) entre os padrões de entrada e os centros das unidades. Para as redes MLP, o argumento para a função de ativação é o produto escalar entre o vetor de entrada e os pesos;
- as redes MLP apresentam unidades (neurônios) não-lineares. As redes RBF apresentam neurônios ocultos (intermediários) não-lineares e neurônios de saída lineares.
- O treinamento para as redes RBF é mais rápido do que para as redes MLP.

## 5.9. Considerações Finais

Este capítulo foi dedicado ao estudo de uma nova tecnologia que vem ganhando espaço na área de processamento digital de sinais: **as redes neurais artificiais**.

Inicialmente foi apresentado um breve histórico sobre o surgimento das redes neurais, o modelo de neurônio artificial comumente empregado e as principais arquiteturas existentes.

O capítulo também apresenta os principais métodos de aprendizagem encontrados na literatura clássica, destacando as principais diferenças entre eles, vantagens e desvantagens de cada um.

A seção 6 descreve as redes neurais com múltiplas camadas que com o desenvolvimento do seu algoritmo de treinamento (retropropagação) foi responsável pelo ressurgimento das pesquisas na área de redes que até então estavam adormecidas, principalmente pela publicação do trabalho de Minsky e Papert. A aprendizagem por retropropagação tornou-se padrão para o treinamento de redes com múltiplas camadas.

O capítulo se encerra com a teoria das redes RBF, que são as redes utilizadas no desenvolvimento deste trabalho de pesquisa. As redes RBF apresentam-se de forma semelhante as redes MLP treinadas pelo algoritmo de retropropagação, mas algumas diferenças são destacadas, sendo uma das principais o tempo de treinamento que é menor para as redes RBF.

# Capítulo 6

# Filtro Duplo de Kalman Estendido

## 6.1. Introdução

Baseado nos capítulos 4 e 5, este capítulo apresenta a descrição do algoritmo do **Filtro Duplo de Kalman Estendido** desenvolvido para a aplicação em filtragem de sinais de voz contaminados por ruído. Além da descrição, será apresentado também o resultado das simulações realizadas para avaliar o algoritmo utilizando dois tipos de ruído: o ruído branco gaussiano e o ruído branco não estacionário.

## 6.2. Descrição do Algoritmo

Como apresentado no capítulo 5, as redes neurais artificiais têm sido amplamente utilizadas em processamento digital de sinais, sendo empregadas principalmente em sistemas de reconhecimento e estimação (filtragem) de sinais de voz a partir de medidas corrompidas por ruídos, classificação de sinais bioelétricos, particularmente EKG (eletrocardiograma), EEG (eletroencefalograma), EMG (eletromiograma) e ENG (eletroneurograma) com a finalidade de se extrair informações patológicas dos órgãos.

Baseado na teoria do filtro de Kalman e nos métodos de redes neurais artificiais, WAN & NELSON (1997), (1998a) e (1998b) desenvolveram o algoritmo para o **Filtro Duplo de Kalman Estendido - FDKE** (*Dual Extended Kalman Filtering*). Neste algoritmo são empregados dois filtros de Kalman e uma rede neural artificial para a estimação de sinais de voz a partir de medidas corrompidas por ruído.

O modelo considerado admite que o sinal de voz é gerado por um processo auto-regressivo não linear, alimentado por um ruído de estado $w(k)$ de acordo com a eq. (6.1).

$$x(k) = f(x(k-1), x(k-2), \ldots, x(k-M), \omega) + w(k) \qquad (6.1)$$

O sinal disponível é contaminado por ruído aditivo e é dado pela eq. (6.2)

$$y(k) = x(k) + v(k) \qquad (6.2)$$

sendo $x(k)$ o vetor que representa o sinal de voz alimentado pelo ruído de estado $w(k)$, $M$ a ordem do processo, $f(.)$ uma função não linear, neste caso uma rede neural com valores anteriores a $x(k)$ parametrizados por $\omega$, que é o vetor contendo os pesos da rede neural. Os pesos são determinados durante o processo de treinamento da rede. O sinal $y(k)$ representa o sinal de voz disponível contaminado por ruído aditivo $v(k)$.

Como apresentado no capítulo 4, para a aplicação das técnicas de filtragem de Kalman é necessário que as eq. (6.1) e (6.2) estejam na forma de espaços de estados. Considere os seguintes modelos de espaços de estados para as eq. (6.1) e (6.2):

$$x(k) = F[x(k-1)] + w(k) \tag{6.3}$$

$$y(k) = C(k)x(k) + v(k) \tag{6.4}$$

sendo: $x(k) = \begin{bmatrix} x(k) \\ x(k-1) \\ \vdots \\ x(k-M+1) \end{bmatrix}$, $F[x(k)] = \begin{bmatrix} f(x(k), x(k-1), \ldots, x(k-M+1), \omega) \\ x(k) \\ \vdots \\ x(k-M+2) \end{bmatrix}$

$$C = [1,0,0,\ldots,0]$$

Para um modelo linear, $F[x(k)]$ pode ser escrito como $Ax(k)$, sendo $A$ a matriz de transição de estados e $w(k)$ e $v(k)$ são ruídos com variâncias conhecidas $\sigma_w^2$ e $\sigma_v^2$.

Para o desenvolvimento do algoritmo FDKE, WAN & NELSON (1998b) se basearam nos trabalhos desenvolvidos por CONNOR *et al.* (1994) e PUSKORIOUS & FELDKAMP (1994). CONNOR *et al.* (1994) utilizam um filtro de Kalman estendido juntamente com uma rede neural para a estimação dos estados de um sistema dinâmico e PUSKORIOUS & FELDKAMP (1994) empregaram um filtro de Kalman para a estimação dos pesos de uma rede neural. O esquema do algoritmo é mostrado na figura 6.1.

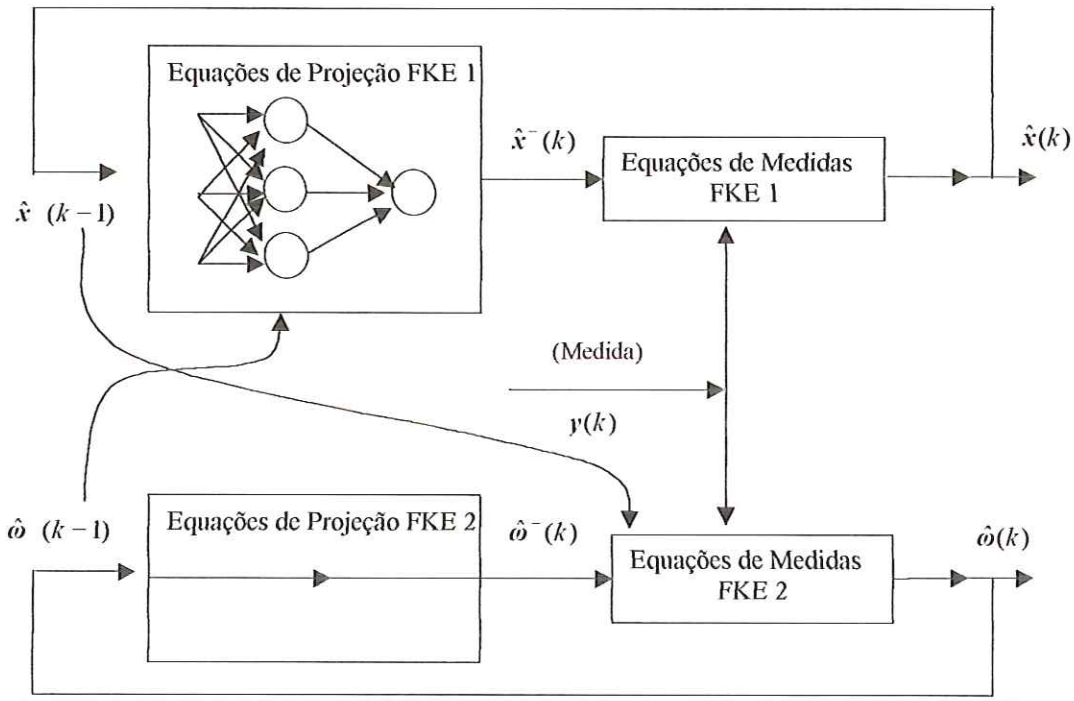



FIGURA 6.1 – Diagrama em blocos do algoritmo proposto por WAN & NELSON (1998b).

O algoritmo FDKE consiste em fazer uma dupla estimativa, ou seja, estimar os estados do sistema e estimar os pesos da rede. Pela análise da figura 6.1

observa-se que a cada instante de tempo $k$, a estimativa do estado $x$ é utilizada pelo filtro dos pesos e a estimativa dos pesos é utilizada pelo filtro de estados. Para um conjunto de dados finitos, o algoritmo trabalha em paralelo até que os pesos da rede convirjam.

Da teoria do filtro de Kalman clássico, para um modelo linear com parâmetros conhecidos, o filtro de Kalman pode ser aplicado diretamente para estimar os estados de um sistema caso o modelo empregado seja linear com parâmetros conhecidos. A cada instante de tempo $k$, o filtro calcula a estimativa $\hat{x}^-(k)$ bem como a covariância do erro, $P^-(k)$, associada com a estimativa do estado. Quando uma nova medida $y(k)$ torna-se disponível, o filtro de Kalman atualiza a estimativa $\hat{x}^-(k)$ obtendo $x(k)$ e a sua covariância associada ao erro de estimação $P(k)$.

## 6.3. Alterações no Algoritmo

Com base nos trabalhos desenvolvidos por BROOMHEAD & LOWE (1998) e KADIRKAMANATHAN *et al.* (1991), onde redes RBF foram empregadas com sucesso para a predição de séries temporais caóticas, e os trabalhos desenvolvidos por CHEN *et al.* (1992a) (1992b) que empregaram redes RBF para a estimação de sinais que foram corrompidos na transmissão, este trabalho consiste em substituir a rede neural treinada pelo algoritmo de retropropagação no trabalho desenvolvido por WAN & NELSON (1998b) por redes do tipo RBF com função de ativação gaussiana.

Como exposto no capítulo 5, as redes do tipo RBF apresentam várias vantagens em relação às redes MLP, sendo que uma das principais é sua capacidade de aprendizagem mais rápida, além do custo computacional ser reduzido.

Neste algoritmo FDKE também é considerada a inexistência de sinais de voz livres de ruídos para o treinamento da rede neural. Esta por sua vez utiliza os parâmetros determinados pelo filtro de Kalman durante o processo de estimação

## 6.4. Estimação de Estados

Para a estimação dos estados do sistema são utilizados um filtro de Kalman e uma rede neural. A rede é responsável pela projeção, ou seja, predição das estimativas iniciais que serão utilizadas pelo filtro de Kalman para atualização.

Como o modelo do sinal de voz apresentado na eq. (6.1) não é linear, o filtro de Kalman não pode ser aplicado diretamente. Dessa forma, é necessário linearizar o modelo a cada instante de tempo $k$. As equações que compõem o filtro de Kalman estendido para a estimação dos estados são apresentadas abaixo:

$$\hat{x}^{-}(k) = F\left[\hat{x}(k-1), \omega\right] \tag{6.5}$$

$$P^{-}(k) = B(k)P(k-1)B^{T}(k) + \sigma_{w}^{2} \tag{6.6}$$

$$B(k) = \frac{\partial F\left[\hat{x}, \omega\right]}{\partial \hat{x}(k-1)} \tag{6.7}$$

$$K(k) = P^{-}(k)C^{T}(k)\left[C(k)P^{-}(k)C^{T}(k) + \sigma_{v}^{2}\right]^{-1} \tag{6.8}$$

$$P(k) = \left[I - K(k)C(k)\right]P^{-}(k) \tag{6.9}$$

$$\hat{x}(k) = \hat{x}^{-}(k) + K(k)\left[y(k) - C\hat{x}^{-}(k)\right] \tag{6.10}$$

No conjunto de equações exposto acima, as equações (6.5) e (6.6) correspondem às equações de projeção do filtro enquanto que as equações (6.8), (6.9) e (6.10) são as equações de atualização das estimativas. A eq. (6.7) é responsável pela linearização da rede neural a cada instante de tempo $k$. Para o cálculo da matriz de linearização da rede neural é assumido que as derivadas parciais existem.

## 6.5. Estimação dos Pesos

Como o algoritmo estudado utiliza redes neurais, é necessária a determinação dos pesos da rede. Como exposto anteriormente, nenhum sinal de voz livre de ruído está disponível para o treinamento da rede neural. Dessa forma, um segundo filtro de Kalman é utilizado para a determinação dos pesos.

Para a aplicação do filtro de Kalman para a determinação dos pesos é necessário que o modelo das equações esteja num espaço de estados. O modelo de espaço de estados dos pesos é dado por:

$$\omega(k) = \omega(k-1) \tag{6.11}$$

$$y(k) = f(x(k-1), \omega(k)) + w(k) + v(k) \tag{6.12}$$

As equações utilizadas pelo filtro de Kalman para a estimação dos pesos da rede neural são dadas por:

$$\hat{\omega}^{-}(k) = \hat{\omega}(k-1) \tag{6.13}$$

$$P^{-}(k) = P(k-1) \tag{6.14}$$

$$K(k) = P^{-}(k)H^{T}(k)\left[H(k)P^{-}(k)H^{T}(k) + \sigma_{w}^{2} + \sigma_{v}^{2}\right]^{-1} \tag{6.15}$$

$$P(k) = \left[I - K(k)H(k)\right]P^{-}(k) \tag{6.16}$$

$$H(k) = \frac{\partial \hat{x}^{-}(k)}{\partial \hat{\omega}} \tag{6.17}$$

$$\hat{\omega}(k) = \hat{\omega}^{-}(k) + K(k)\left[y(k) - \hat{x}^{-}(k)\right] \tag{6.18}$$

De forma semelhante a eq. (6.7), a eq. (6.17) é responsável pela linearização da rede a cada ciclo de operação do filtro de Kalman.

## 6.6. Implementação e Avaliação do Algoritmo

O algoritmo descrito acima foi implementado usando o ambiente de programação Matlab. Os sinais a serem processados foram colhidos na placa *sound blaster* e corrompidos por ruído. Dois tipos de ruídos são considerados no processamento: ruído branco gaussiano e ruído branco não estacionário.

Para o processamento do algoritmo várias palavras foram utilizadas dentre elas: "casa", "vela", "elétrica" e "cedo". Todas estas palavras foram colhidas na placa *sound blaster* e amostradas a uma taxa de 11025 Hz. Em ambos os resultados dos processamentos apresentados neste trabalho foi utilizada a palavra "elétrica". O algoritmo foi aplicado a janelas retangulares sucessivas de 250 amostras (aproximadamente 23 ms), sem superposição.

Para a rede neural, esta foi processada com três camadas. Variou-se diversas vezes o número de neurônios tanto na camada de entrada quanto na camada

oculta e o melhor ajuste encontrado foi 15 neurônios na camada de entrada e 5 neurônios na camada oculta. O número de neurônios na camada de saída foi fixado em 1.

Os neurônios da camada oculta utilizam funções de base radial gaussiana. A expressão para esta função foi apresentada no capítulo 5 e é dada por:

$$\varphi(y) = e^{-\frac{(y-c)^2}{2r^2}} \tag{6.19}$$

Testes com as outras funções de base radial (multiquadrática e multiquadrática inversa) também foram realizados, porém notou-se um aumento no tempo de processamento. De acordo com a eq. (6.19), antes do processamento da rede neural é necessária a determinação dos centros e dos raios da função de base radial. Para o cálculo dos centros e dos raios foi utilizado o algoritmo *k*-médias descrito na tabela 5.1.

A simulação foi realizada num computador com processador pentium III 333 MHz, com 64 MB de memória RAM. O tempo de processamento variou entre 15 e 20 minutos, dependendo da configuração da rede neural e do sinal processado.

### 6.6.1. Processamento Utilizando Ruído Branco Gaussiano

Neste primeiro teste, foi considerado um ruído aditivo branco gaussiano (média constante zero e desvio padrão 0.05). Este ruído foi gerado no Matlab e adicionado ao sinal de voz. A figura 6.2 mostra a forma de onda do sinal original antes de ser corrompido pelo ruído branco gaussiano que por sua vez está

representado na figura 6.3. Depois de corrompido, a forma de onda do sinal é mostrada na figura 6.4. Este sinal contaminado por ruído é aplicado ao FDKE e o resultado pode ser observado na figura 6.5.

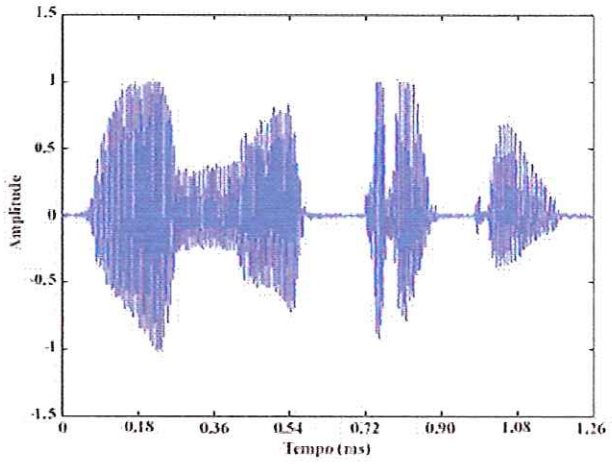



FIGURA 6.2 – Forma de onda da palavra "elétrica".

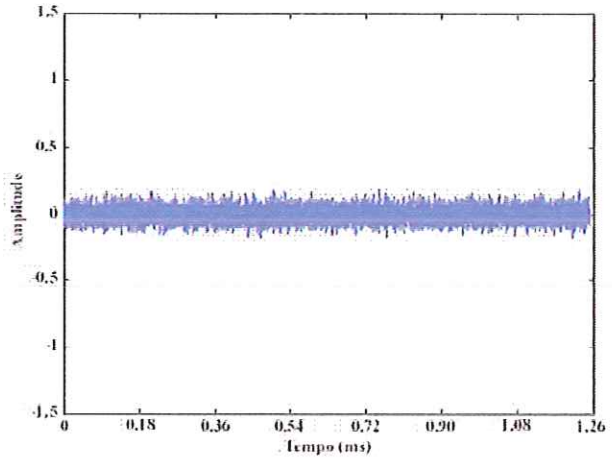



FIGURA 6.3 – Ruído branco gerado para contaminar o sinal da figura 6.2.

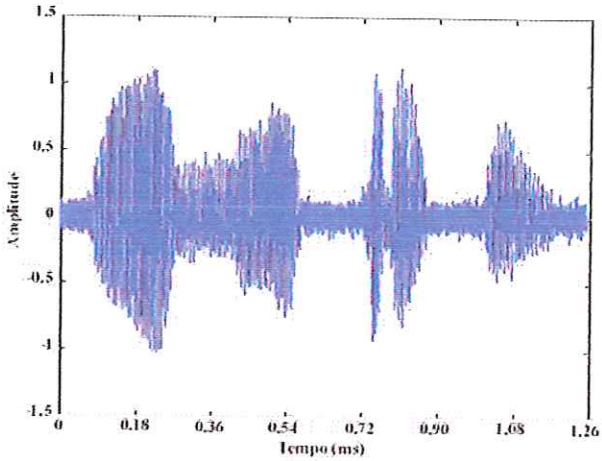



FIGURA 6.4 – Sinal de voz contaminado por ruído branco gaussiano.

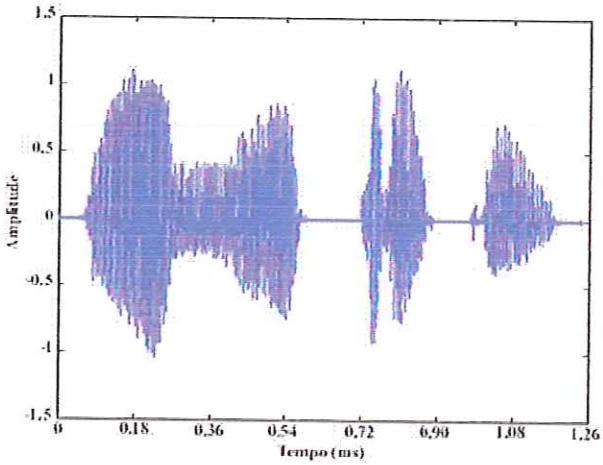



FIGURA 6.5 – Sinal de voz processado pelo FDKE.

Comparando a figura 6.5 com a figura 6.4 nota-se claramente uma grande redução do ruído pelo algoritmo FDKE, principalmente nos trechos de silêncio do sinal. Além da comparação entre as formas de onda, foram realizados também testes

subjetivos para avaliar a qualidade do sinal processado. Estes testes comprovam a redução de ruídos em relação ao sinal de voz contaminado e o sinal original.

Outro fator que pode ser levado em conta para avaliar o sinal processado é a relação sinal ruído (SNR). Para este teste utilizando ruído branco gaussiano houve um aumento de 25 dB na SNR. Apesar desse aumento na SNR, um ruído residual permanece e sua eliminação é quase impossível.

Além da comparação entre as formas de onda no tempo e da relação sinal ruído, outra ferramenta que também pode ser utilizada para analisar o sinal processado é o espectrograma. Os espectrogramas são gráficos bidimensionais na qual a dimensão vertical corresponde à freqüência e a dimensão horizontal ao tempo. As figuras 6.6 e 6.7 mostram os espectrogramas do sinal de voz original e do sinal de voz processado respectivamente. Conforme observado para as formas de onda das figuras 6.5 e 6.2, a redução do ruído pode ser observada nos trechos de silêncio.

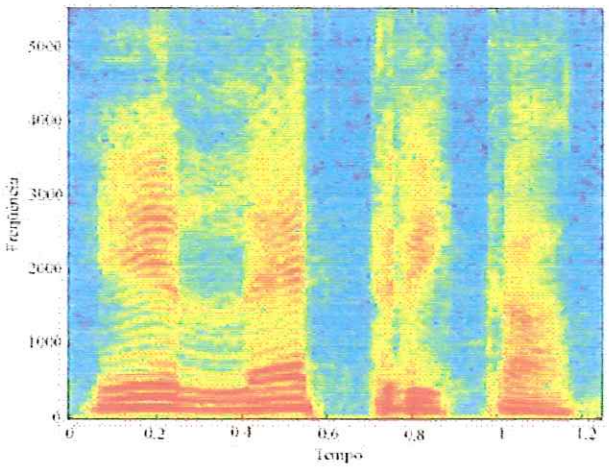



FIGURA 6.6 – Espectrograma do sinal original.

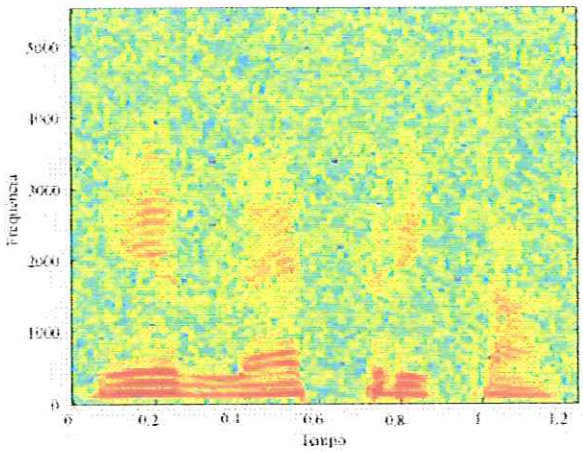



FIGURA 6.7 – Espectrograma do sinal corrompido por ruído branco gaussiano.

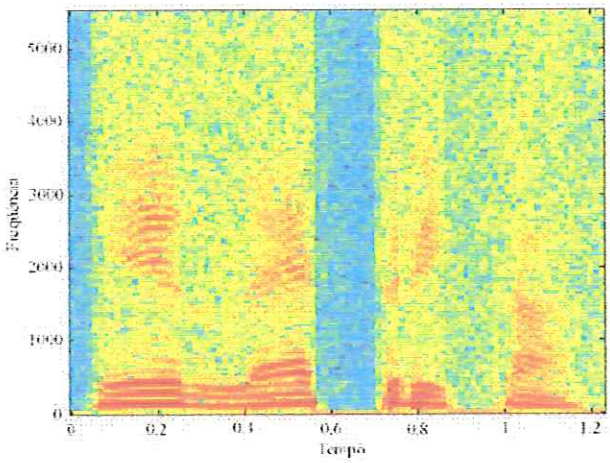



FIGURA 6.8 – Espectrograma do sinal filtrado.

Conforme análise dos três espectrogramas percebe-se na figura 6.8 (sinal filtrado) que nas altas freqüências (entre 2000 e 3000 Hz) a intensidade do sinal foi ressaltada pelo algoritmo de filtragem se comparado com o espectrograma da figura 6.7 (sinal corrompido por ruído). Pode-se perceber também que apesar de uma redução da intensidade do ruído, um ruído residual permanece.

### 6.6.2. Processamento Utilizando Ruído Branco Não Estacionário

Conforme apresentado na seção anterior, o sinal de voz resultante do processamento pelo FDKE, quando corrompido por ruído branco gaussiano apresenta uma grande redução de ruído e um aumento de 25 dB na SNR. Como os sinais na natureza nem sempre apresentam características estacionárias, este segundo teste considera o sinal de voz corrompido por ruído branco não estacionário. De forma semelhante ao processamento utilizando ruído branco gaussiano, o sinal original representado na figura 6.2 é corrompido por ruído branco não estacionário. A forma de onda no tempo para o ruído branco não estacionário está representada na figura 6.9 e o sinal corrompido na figura 6.10.

As características da rede neural são as mesmas que foram utilizadas para o ruído branco. O sinal processado pode ser observado na figura 6.11.

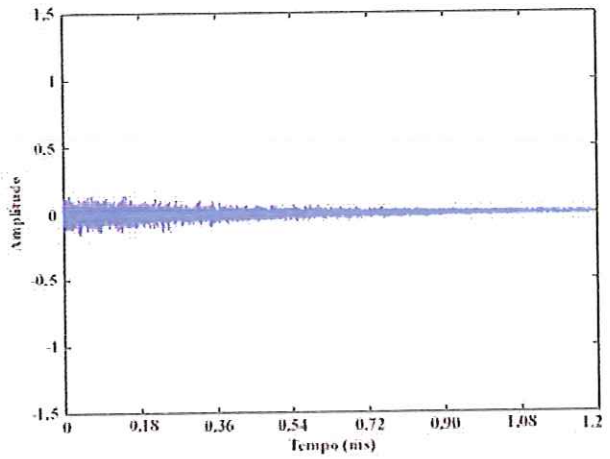



FIGURA 6.9 – Ruído não estacionário.

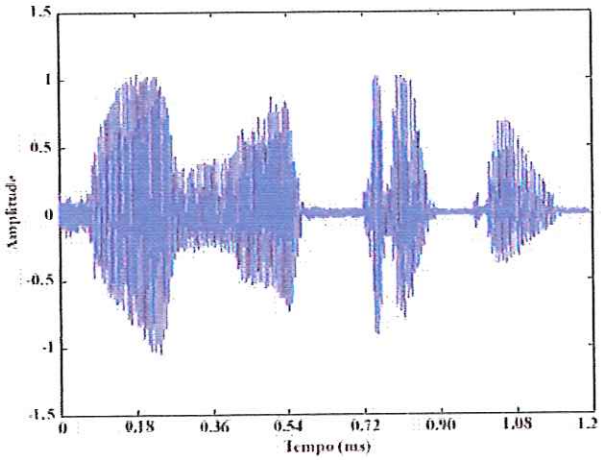



FIGURA 6.10 – Sinal de voz contaminado com ruído branco não estacionário.

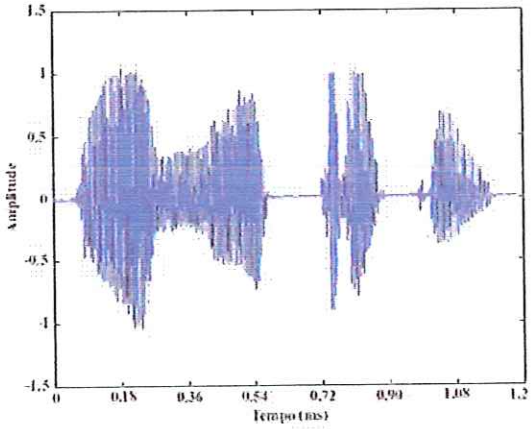



FIGURA 6.11 – Resultado da filtragem do sinal contaminado com ruído branco não estacionário.

Como para o ruído branco estacionário, o sinal resultante do processamento do sinal corrompido por ruído branco não estacionário apresentou

uma redução considerável do ruído aditivo, fácil de se notar nos trechos de silêncio. A inteligibilidade do sinal é bem melhor em relação ao sinal corrompido. A figura 6.12 mostra o espectrograma para o sinal corrompido por ruído branco não estacionário e a figura 6.13 mostra o espectrograma para o sinal resultante do processamento.

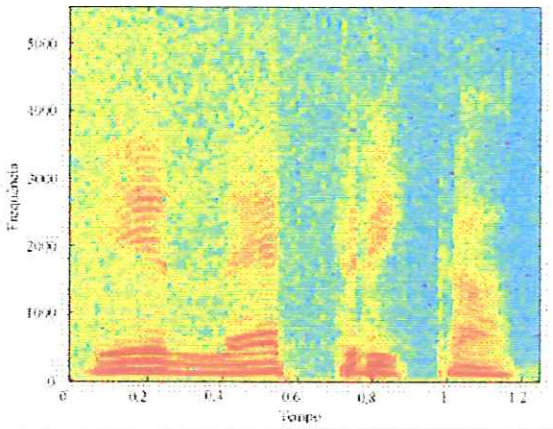



FIGURA 6.12 – Espectrograma do sinal de voz corrompido com ruído branco não estacionário.

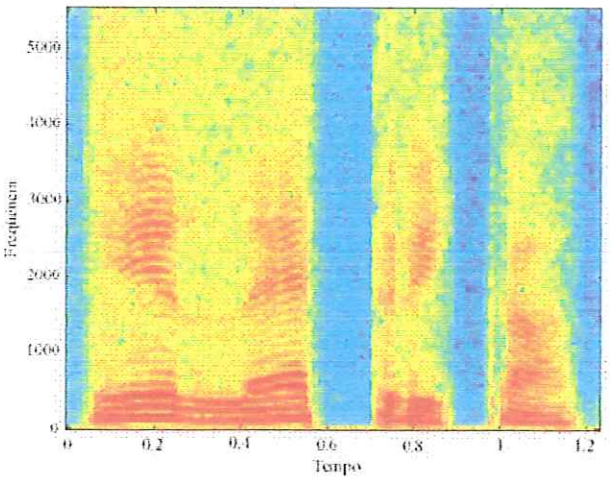



FIGURA 6.13 – Espectrograma do sinal resultante do processamento do sinal corrompido com ruído branco não estacionário.

Comparando-se as figuras 6.12 e 6.13 observa-se uma melhora do sinal filtrado em relação ao sinal corrompido por ruído, que pode ser constatado principalmente nos trechos de silêncio do sinal e nas altas freqüências (entre 2000 e 4000 Hz), onde as componentes do sinal foram ressaltadas. Para este processamento houve um aumento de 15 dB na SNR.

### 6.6.3. Pré-Filtragem Utilizando o Filtro de Média

Nos testes anteriores, para a determinação dos parâmetros do filtro de Kalman e das redes neurais, o algoritmo de filtragem é aplicado diretamente ao sinal de voz corrompido por ruído. Desse modo, tanto os parâmetros do filtro de Kalman quanto da rede neural não são determinados eficientemente devido a intensidade do ruído e conseqüentemente apesar do sinal resultante apresentar uma grande redução do ruído aditivo notou-se nos testes auditivos a presença de um ruído residual.

Para a redução desse ruído residual, uma pré-filtragem é empregada no sinal disponível $y(k)$. Duas técnicas de pré-filtragem foram utilizadas: filtro de média e filtro de mediana, sendo que os melhores resultados foram obtidos com o filtro de média. Para o filtro de média cada amostra do sinal corrompido por ruído é a média aritmética entre amostras anteriores e posteriores em relação a uma determinada amostra. A ordem do filtro de média (número de pontos utilizados para calcular a média) foi variada em 3, 5 e 7 e os melhores resultados foram obtidos com 5 pontos. As figuras 6.14 a 6.16 mostram as formas de onda para o sinal corrompido por ruído antes da pré-filtragem, sinal resultante da pré-filtragem e o sinal resultante do processamento respectivamente.

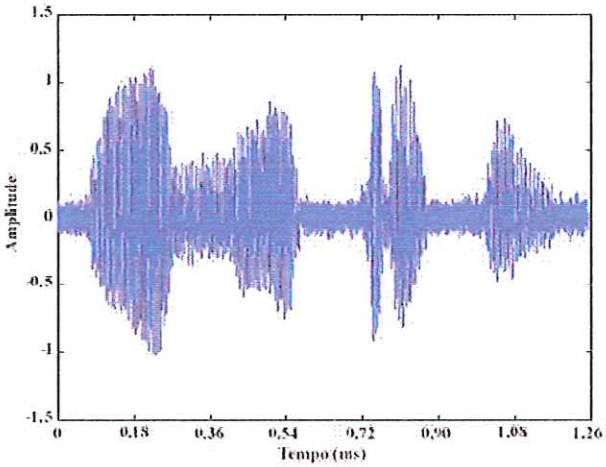



FIGURA 6.14 – Sinal de voz contaminado por ruído branco gaussiano antes da pré-filtragem.

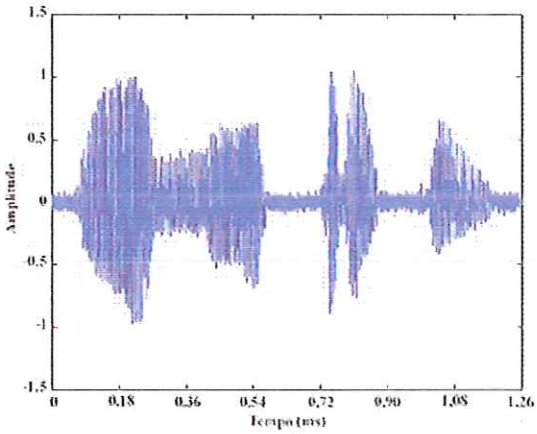



FIGURA 6.15 – Sinal resultante da pré-filtragem.

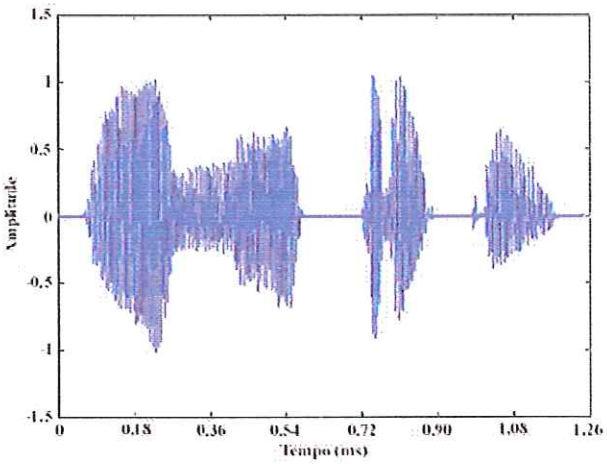



FIGURA 6.16 – Resultado da filtragem do sinal pré-filtrado.

De todos os testes realizados o sinal apresentado na figura 6.16 representa o melhor processamento realizado pelo FDKE com um aumento de aproximadamente 28 dB na SNR. Quando comparado este teste (utilizando pré-filtragem) com o primeiro teste realizado em que o FDKE foi aplicado diretamente ao sinal corrompido por ruído, o processamento com pré-filtragem apresentou um aumento de 3 dB na SNR em relação ao primeiro teste. Com relação a qualidade do som, é o que mais se aproxima do sinal original.

As figuras 6.17 e 6.18 representam os espectrogramas para o sinal corrompido por ruído branco gaussiano antes da pré-filtragem e para o sinal resultante do processamento pelo FDKE, respectivamente. Comparando-se as figuras 6.18 e 6.6, pode-se observar uma redução considerável do ruído. Nas altas freqüências (entre 2000 e 3000 Hz) nota-se que as componentes do sinal foram recuperadas, mas ainda permanece um ruído residual. A eficiência do algoritmo deve-se principalmente a pré-filtragem inicial que reduziu a potência do ruído.

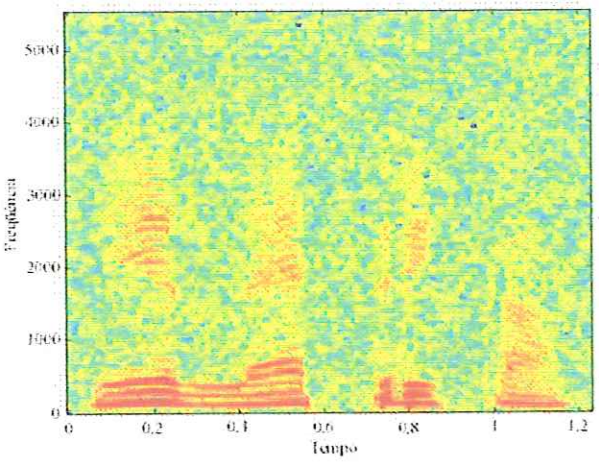



FIGURA 6.17 - Espectrograma do sinal de voz corrompido com ruído branco gaussiano.

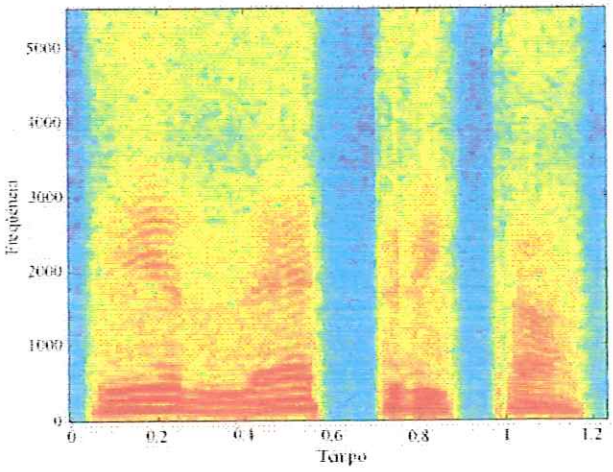



FIGURA 6.18 – Espectrograma do sinal resultante do processamento pelo FDKE, sendo que a pré-filtragem foi realizada pelo filtro de média.

## 6.7. Considerações Finais

Neste capítulo foi apresentado o algoritmo para o Filtro Duplo de Kalman Estendido (FDKE) que foi estudado e analisado. O algoritmo original

consiste em uma rede neural, treinada pelo algoritmo de retropropagação e dois filtros de Kalman estendidos. A rede neural e um filtro de Kalman são responsáveis pela estimação dos estados enquanto que o segundo filtro é responsável pela estimação dos pesos da rede neural.

Neste trabalho, a rede MLP treinada pelo algoritmo de retropropagação foi substituída por redes RBF com função de ativação do tipo gaussiana.

Para a avaliação do algoritmo foram realizados vários testes com diferentes sinais de voz corrompidos por dois tipos de ruído: ruído branco gaussiano e ruído branco não estacionário. Para o teste em que o sinal de voz foi corrompido por ruído branco gaussiano houve um aumento de 25 dB na SNR. Para o sinal corrompido por ruído branco não estacionário o aumento foi de 15 dB. Todos os testes apresentados neste trabalho foram para a palavra "elétrica". Para a comprovação desses resultados outros sinais foram processados.

Apesar da grande redução de ruído, um ruído residual (ruído musical) permanece. Para uma redução da potência deste ruído residual foi aplicada ao sinal ruidoso uma pré-filtragem com o filtro de média e logo em seguida o filtro duplo de Kalman estendido. O sinal processado é o que mais se aproxima do sinal original (sem ruído aditivo) com um aumento de 28 dB na SNR. Com o emprego da pré-filtragem houve um aumento de 3 dB em relação ao teste sem pré-filtragem. Além do filtro de média na pré-filtragem foi utilizado também o filtro de mediana mas o melhor resultado foi para o filtro de média.

# Capítulo 7

# Conclusões

As técnicas de filtragem de sinais de voz, mais especificamente a filtragem de Kalman e Redes Neurais Artificiais foram o enfoque deste trabalho. Várias técnicas comumente empregadas na literatura clássica foram ressaltadas, mas muitas vezes apresentam um ganho baixo e podem até provocar distorções no sinal filtrado.

Neste trabalho foi estudado e implementado o algoritmo para o Filtro Duplo de Kalman Estendido (FDKE). Este algoritmo utiliza dois filtros de Kalman e uma rede neural. O filtro de Kalman juntamente com a rede neural realiza a estimação dos estados enquanto que o segundo filtro é responsável pela estimação

dos pesos da rede, tendo em vista que foi considerada a inexistência de sinais livre de ruídos para o treinamento da rede neural.

O algoritmo original foi implementado com redes MLP treinadas pelo algoritmo de retropropagação, enquanto que este trabalho propôs a substituição da rede MLP por redes RBF. As redes RBF são redes equivalentes, mas apresentam várias diferenças e vantagens, sendo a principal o seu tempo de treinamento que é menor.

A avaliação do algoritmo foi comparativa e os sinais de voz foram colhidos na placa *sound blaster* amostrados a uma taxa de 11.025 Hz e corrompidos com ruído aditivo gerado através da ferramenta Matlab. Para os testes foram gerados dois tipos de ruído: o ruído branco gaussiano e o ruído branco não estacionário. Os resultados apresentados neste trabalho foram para a palavra "elétrica". Para a confirmação dos resultados outros sinais foram testados dentre eles as palavras "casa", "cedo" e "vela".

A avaliação dos sinais processados foi realizada através da observação da forma de onda no tempo, análise dos espectrogramas, determinação das relações sinal ruído (SNR) e testes subjetivos. Tanto os testes com ruído branco gaussiano quanto com ruído branco não estacionário produziram resultados satisfatórios e mostraram a eficiência do algoritmo implementado. Pela análise das formas de onda no tempo percebe-se a redução do ruído principalmente nos instantes de silêncio, mas um ruído residual permanece.

Além das comparações entre as formas de onda dos sinais no tempo, a relação sinal ruído representa uma ferramenta importante para a análise de sinais. O sinal de voz corrompido com ruído branco gaussiano apresentou um aumento de

aproximadamente 25 dB na SNR. Para o sinal corrompido com ruído branco não estacionário o aumento na SNR foi de aproximadamente 15 dB.

Quando o algoritmo de filtragem é aplicado diretamente ao sinal de voz corrompido por ruído, tanto os parâmetros do filtro de Kalman quanto da rede neural não são determinados eficientemente principalmente devido a intensidade do ruído e conseqüentemente nota-se no sinal filtrado a presença do um ruído musical.

Com o objetivo de melhorar as estimativas iniciais do sinal para a determinação dos parâmetros da rede neural e do filtro de Kalman e reduzir a intensidade do ruído musical, foi utilizado no sinal corrompido por ruído uma pré-filtragem com o filtro de média. Dentre os testes realizados, o sinal filtrado pelo algoritmo com pré-filtragem apresentou o melhor resultado com um aumento de aproximadamente 28 dB na SNR, ou seja, uma melhora de 3 dB em relação ao processamento sem pré-filtragem.

Além da análise das formas de onda no tempo, análise dos espectrogramas e determinação da relação sinal ruído também foram realizados testes subjetivos para a análise da qualidade auditiva do sinal filtrado. Todos os sinais filtrados são auditivamente melhores que o sinal corrompido por ruído. O sinal processado pelo FDKE e com pré-filtragem apresenta ruído musical menor que os outros dois.

Com relação a rede neural, esta foi implementada com três camadas: camada de entrada, escondida e de saída. O número de neurônios nas camadas de entrada e escondida foi variado diversas vezes e o melhor ajuste encontrado foi 15 neurônios na camada de entrada e 5 neurônios na cama oculta. O número de neurônios na camada de saída foi fixado em 1.

Dessa forma, o projeto foi concluído após um estudo do algoritmo de filtragem de Kalman e de redes neurais artificiais. Estes algoritmos foram implementados e os testes realizados provaram sua eficiência.

Durante o desenvolvimento deste projeto algumas idéias surgiram, mas não puderam ser realizadas. Algumas delas são possíveis de ser implementadas a curto prazo, porém outras exigiriam um estudo mais detalhado. Abaixo algumas idéias.

1. Estimativa das variâncias dos ruídos de estado e aditivo. Neste trabalho foi considerado que as estimativas eram conhecidas;
2. Testes com outras estruturas de redes como, por exemplo, redes recorrentes.
3. Alteração no modelo dos pesos da rede neural que ao invés de utilizar a medida no instante de tempo anterior será utilizada uma média de $n$ amostras anteriores.
4. Determinação dos centros da rede RBF substituindo o algoritmo $k$-médias pela recursão ortogonal mínima quadrática, que dentre várias vantagens apresenta um custo computacional reduzido.

# BIBLIOGRAFIA

ALEKSANDER, I.; MORTON, H. (1990). *An Introduction to Neural Computing*, London: Chapman and Hall.

ANDERSON, B. D. O.; MOORE, J. B. (1979). *Optimal Filtering*. Englewood Cliffs, N. J.: Prentice-Hall.

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

BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. (1999). *Redes Neurais Artificiais: Teoria e Aplicação*. Editora LTC.

BROOMHEAD, D. S; LOWE, D. (1998). *Multivariable functional interpolation and adaptive networks.*, Complex Systems 2, pg. 321-355.

BROWN, R. G.; HWANG, P. Y. C. (1992). *Introduction to Random Signals and Applied Kalman Filtering*, Second Edition, John Wiley & Sons, Inc.

CHEN, S.; MULGREW, B.; MCLAUGHLIN, S.; GRANT, P. M. (1992a). Adaptive Bayesian feedback equalizer based on radial basis function network, IEEE International Conference on Communications 3, pg. 1267-1271, Chicago, IL.

CHEN, S.; MULGREW, B.; MCLAUGHLIN, S.; GRANT, P. M. (1992b). Adaptive Bayesian equalizer with feedback for mobile radio channels. Workshop on Adaptive Algorithms in Communications, Bordeaux, France.

CONNOR, J.; MARTIN, R. AND ATLAS, L. (1994). Recurrent Neural Networks and robust time series prediction, IEEE Tr. on Neural Networks.

CYBENKO, G. (1989). Aproximation by superpositions of a sigmoidal function, *Mathematics of Control, Signals, and Systems*, vol. 2, pp. 303-314.

DELLER, J. R.; PROAKIS, J. G.; HANSEN, J. H. L. (1993). *Discrete-Time Processing of Speech Signals.* Englewood Cliffs, N.J.: Prentice Hall.

EPHRAIM, Y.; MALAH, D. (1984). Speech Enhancement Using Minimum Mean Square Error Short-Time Espectral Amplitude Estimator. IEEE Trans. on Acoust., Speech Signal Processing, vol. 32, nº 6, december.

FANT, C. G. M. (1960). *Acoustic Theory of Speech Production.* The Hague, The Netherlands: Mounton.

FITZGERALD, R. J. (1971). Divergence of the Kalman Filter, IEEE Trans. Automatic Control., p. 736-747, vol AC-16, nº 6, december.

EESC - USP
Biblioteca
CAMPUS DE S. CARLOS

FLANAGAN, J. L. (1972). *Speech Analysis, Synthesis, and Perception,* 2nd ed. New York: Springer-Verlag.

GELB, A. (1974). *Applied Ottimal Estimation*. The MIT Press, Cabridge.

HAYES, M. H. (1996). *Statistical Digital Signal Pocessing and Modeling*. John Wiley & Sons, Inc.

HAYKIN, S. (1996). *Adaptive Filter Theory,* 2nd ed. Englewood Cliffs, N.J.: Prentice Hall.

HAYKIN, S. (1999). *Neural Networks – A Comprehensive Foundation*. Prentice-Hall.

HE, X.; LAPEDES, A. (1991). Nonlinear Modeling and prediction by sucessive aproximations using radial basis functions. Technical report, Los Alamos National Laboratory.

HEBB, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*, New York: Wiley.

HOPFIELD, J. J. (1982). Neural networks and physical systems with emergent colletive computational abilities, *Proceedings of the National Academy of Sciences*, USA, vol. 81, pp. 3088-3092.

JOAQUIM, M. B. (1988). *Representação Senoidal da Voz*. 147p. Tese (Doutorado), EP-Escola Politécnica, São Paulo.

KADIRKAMANATHAN, V.; NIRAJAN, M.; FALLSIDE, F. (1991). The reorganization of somatosensory cortex following peripheral nerve demage in adult and developing mammals., Anual Review of Neurosciences 6, pg. 325-356.

KAILATH, T. (1980). *Linear Systems*, Englewood Cliffs, NJ: Prentice-Hall.

KALMAN, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems, Transaction of the ASME – Journal of Basic Enginerring, pp. 35-45, March.

MARAR, J. F., (1997). Polinômios Potência de Sigmóide (PPS): Uma Nova Técnica para Aproximação de Funções, Construções de Wavenets e suas Aplicações em Processamento de Imagens e Sinais, 180p. Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN. Informática, Recife.

MAYBECK, P. S. (1982). *Stochastic Models, Estimation, and Control*, Volume 1, Academic Press, Inc.

MCCULLOCH, W. S.; PITTS, W., (1943). A logical calculus od the ideas immanent in nervous activity, Bulletin of Mathematical Biophysics, vol. 5, pp. 115-133.

MEDITCH, J. S. (1969). *Stochastic Optimal Linear Estimation and Control,* MacGraw-Hill Book Company, New York.

MENDEL, J. M. (1971). Computational Requeriments for a Discrete Kalman Filter. IEEE Trans. Automat. Contr., vol. AC-16, p.748-758, december.

MENDEL, J. M., MCLAREN, R. W. (1970). Reinforcement-learning control and pattern recognition systems, in *Adaptive, Learning, and Pattern Recognition Systems: Theory and Applications*, vol. 66, J. M. Mendel and K. S. Fu, eds., pp.287-318, New York: Academic Press.

MINSKY, M. L.; PAPERT, S. A. (1990). *Perceptrons*. Cambridge: MIT Press. Expanded edition.

MOODY, J.; DARKEN, C., J. (1989). Fast learning in networks of locally-tuned processing units, *Neural Computation*, vol. 1, pp. 281-294.

NASCIMENTO JR., C. L. (1988). *Estudo Comparativo de Métodos de Combate à Divergência de Filtros de Kalman*, 248p. Dissertação (Mestrado), ITA.

NETO, A. B. G. (1989). Melhoramento de Voz Degradada por Método Baseado em Subtração Espectral Adaptativa, in Anais do 7º Simpósio Brasileiro de Telecomunicações, Florianópolis-SC, p. 54-59, setembro.

PARK, J.; SANDBERG, I. W. (1991). Universal Aproximation using radial basis function networks. *Neural Computation* 3 (2), 246-257.

PARSONS, T. W. (1987). *Voice and Speech Processing*. McCraw-Hill, New York.

PUSKORIOUS, G.; FELDKAMP, L. (1994). Neural control of nonlinear dynamic systems with Kalman filter trained recurrent networks, IEEE Trn. on NN, vol. 5, n.° 2.

RABINER, L. R.; SCHAFER, R. W. (1978). *Digital Processing of Speech Signals.* Englewood Cliffs, N.J.: Prentice Hall.

ROSENBLATT, F. (1958). The Perceptron: A probabilistc model for information storage and organization in the brain, *Psychological Review*, vol. 65, pp.386-408.

RUMELHART, D. E.; MCCLELLAND, J. L. (1986). *Paralell Distributed Processing: Explorations in the microstructure of cognition*, Vol. 1, The MIT Press.

SORENSON, H. W. (1966). Least-squares estimation: from Gauss to Kalman. IEEE Spectrum, vol. 7, p. 63-68.

VIEIRA-FILHO, J. V. (1996). *Redução de Ruídos em Sinais de Voz em Sistemas Rádio-Móveis Veiculares*, Tese de Doutorado. UNICAMP, agosto.

VIEIRA-FILHO, J. V; SCARLAT, P.; CHIQUITO, J. G. (1995). Redução de Ruídos em Sinais de Voz: Análise e avaliação das técnicas clássicas baseadas na subtração espectral a curto termos (SECT), in 13° SBT, Águas de Lindóia, p. 685-689, setembro.

WAN, E.; NELSON, A. (1997). Neural Speech Enhancement Using Dual Extended Kalman Filtering, International Conference on Neural Networks.

WAN, E.; NELSON, A. (1998a). Networks for Speech Enhancement, in *Handbook of Neural Networks for Speech Processing*, Edited by Shigeru Katagiri, Artech House, Boston.

WAN, E.; NELSON, A. (1998b). Removal of noise from speech using the dual extended Kalman filtering algorithn. ICASSP.

WELCH, G.; BISHOP, G. (1996). An Introduction to the Kalman Filter.

WIDROW, B.; HOFF-JR, M. E. (1960). "Adaptative switching circuits", IRE WESCON Conventional Record, pp. 96-104.