

ALGORITMO DE EXTENSÃO EM FREQUÊNCIA BASEADO EM REDES
NEURAIIS E FILTRAGEM ÓTIMA

Ygor Ururahy de Carvalho

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

Aprovada por:

Prof. Mariane Rembold Petraglia, Ph.D.

Prof. Fernando Gil Vianna Resende Junior, Ph.D.

Prof. Jacques Szczupak, Ph.D.

RIO DE JANEIRO, RJ - BRASIL

JUNHO DE 2006

CARVALHO, YGOR URURAHY DE

Algoritmo de Extensão em Frequência
Baseado em Redes Neurais e Filtragem
Ótima [Rio de Janeiro] 2006

XIV, 77 p. 29,7 cm (COPPE/UFRJ,
M.Sc., Engenharia Elétrica, 2006)

Dissertação - Universidade Federal do
Rio de Janeiro, COPPE

1.Extensão em Frequência 2.Redes Neurais
3.Filtragem Ótima 4.Banco de Filtros
Modulados por Cosseno

I.COPPE/UFRJ II.Título (série)

Dedicatória

Feliz do homem que, ao longo de sua caminhada, pode contar com a companhia e o auxílio de pessoas tão especiais. Especialmente, de uma grande mulher. Obrigado, Cristia.

Dedico este trabalho a minha esposa Cristia, aos meus pais e avós, e ao meu irmão por toda ajuda, carinho e compreensão.

Agradecimentos

A Deus, por me dar todas as condições de realização deste trabalho.

À minha querida esposa, por estar sempre ao meu lado e dedicar seu tempo a me ajudar no que fosse possível, inclusive nas correções do texto. Pelo carinho, compreensão e paciência durante a realização do trabalho.

Aos familiares, pelo apoio e suporte.

À minha orientadora, Mariane Rembold Petraglia, por toda a ajuda e paciência durante todo o trabalho.

Aos amigos, pela compreensão e estímulos dados ao longo do trabalho.

À Eletronuclear S.A., por permitir que me ausentasse, no horário de trabalho, para ir à Universidade.

À FAPERJ, por ter me concedido uma bolsa de estudos no início do curso de mestrado.

A todas as pessoas que contribuíram direta ou indiretamente para a realização deste trabalho.

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

ALGORITMO DE EXTENSÃO EM FREQUÊNCIA BASEADO EM REDES
NEURAIIS E FILTRAGEM ÓTIMA.

Ygor Ururahy de Carvalho

Junho/2006

Orientador: Mariane Rembold Petraglia

Programa: Engenharia Elétrica

O sinal de voz da telefonia analógica atual é limitado em banda, sendo tipicamente utilizadas frequências entre 300 e 3400 Hz na sua transmissão. Desta forma, o sinal recebido tem sua qualidade comprometida. Uma solução para este problema seria a utilização de uma banda de transmissão maior, o que requereria a modificação do canal utilizado. Para se evitar alterações no sistema de transmissão existente, uma possível solução é a utilização de algoritmos de extensão em frequência na recepção do sinal de voz. Estes métodos têm como objetivo a adição de componentes de frequências mais altas aos sinais limitados em banda, baseados nas redundâncias contidas nas bandas de frequência dos sinais de voz. Este trabalho apresenta um novo método de extensão em frequência baseado na classificação de trechos dos sinais de voz por meio de redes neurais e na extensão de frequências altas destes sinais por meio de filtragem ótima. Testes subjetivos mostraram que o método proposto obteve bons resultados quando comparado a outros métodos de maior complexidade computacional.

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

FREQUENCY EXTENSION BASED ON NEURAL NETWORKS AND
OPTIMUM FILTERING.

Ygor Ururahy de Carvalho

June/2006

Advisor: Mariane Rembold Petraglia

Department: Electrical Engineering

The voice signal of the current analogic telephony system is limited in band. Typically frequencies between 300 and 3400 Hz are used in the transmission. In such a way, the received signal has its quality compromised. A solution for this problem would be the increase of the transmission band, which would require the modification of the used channel. To prevent alterations in the existing transmission system, a possible solution is the use of frequency extension algorithms in the reception. These methods add high frequency components to the bandlimited signals, based in the redundancies contained in the frequency bands of the speech signals. This work presents a new method of frequency extension based on the classification of voice frames by means of neural networks and on the extension of high frequencies of these signals by means of optimum filtering. Subjective tests had shown that the considered method got good results when compared to other methods of higher computational complexity.

Sumário

1	Introdução	1
2	Redes Neurais	3
2.1	Introdução	3
2.2	Fundamentos	3
2.3	Modelagem do Neurônio Artificial	4
2.4	Treinamento das Redes Neurais	7
2.4.1	O Algoritmo do Gradiente Descendente	9
2.4.2	O Algoritmo de Retropropagação do Erro - RPE	9
3	Método dos Mínimos Quadrados	15
3.1	Introdução	15
3.2	Estimação Linear	15
3.3	Filtragem Linear no Sentido dos Mínimos Quadrados	17
4	Bancos de Filtros Modulados por Cosseno	20
4.1	Introdução	20
4.2	Fundamentos	20
4.3	Expressões para os Filtros	22
5	Técnicas de Extensão em Frequência	24
5.1	Introdução	24
5.2	Conceitos Básicos sobre a Fala	24
5.3	Extensão em Frequência	26
6	Método Proposto de Extensão em Frequência	30
6.1	Introdução	30

6.2	Sinais de Voz Utilizados	30
6.3	Treinamento das Redes Neurais	35
6.4	Descrição do Método	35
6.5	Banco de Filtros	40
7	Resultados	43
7.1	Introdução	43
7.2	Filtros de Extensão em Frequência	43
7.3	Resultados do Algoritmo Proposto de Extensão em Frequência	48
7.4	Discussão dos Resultados	63
8	Conclusões e Trabalhos Futuros	74
8.1	Conclusões	74
8.2	Trabalhos Futuros	74
	Referências Bibliográficas	76

Lista de Figuras

1.1	Esquema de uma conexão telefônica com módulo de extensão em frequência (EF) implementado na recepção da conexão.	2
2.1	Representação de neurônios biológicos típicos.	4
2.2	Modelo matemático do neurônio artificial.	5
2.3	Função tangente hiperbólica.	6
2.4	Rede neural com duas camadas: camada intermediária e camada de saída.	7
3.1	Esquema de filtragem para a modelagem de processo.	17
4.1	Banco de filtros maximamente decimado com M canais.	21
4.2	Módulos das respostas em frequência típicas de um banco de filtros maximamente decimado com M canais.	21
5.1	Representação do trato vocal humano e seus principais constituintes.	25
5.2	Espectrograma do sinal correspondente à palavra <i>pencil</i> proferida por locutor masculino.	26
5.3	Modelo fonte-filtro do processo de geração da fala humana.	27
5.4	Estrutura básica dos algoritmos de extensão em frequência baseados no modelo fonte-filtro.	28
6.1	(a) Espectrograma de x_{f1}^{BE} ; (b) Espectrograma de x_{f1}^{BL}	31
6.2	(a) Espectrograma de x_{f2}^{BE} ; (b) Espectrograma de x_{f2}^{BL}	32
6.3	(a) Espectrograma de x_{f3}^{BE} ; (b) Espectrograma de x_{f3}^{BL}	32
6.4	(a) Espectrograma de x_{m1}^{BE} ; (b) Espectrograma de x_{m1}^{BL}	33
6.5	(a) Espectrograma de x_{m2}^{BE} ; (b) Espectrograma de x_{m2}^{BL}	33
6.6	(a) Espectrograma de x_{m3}^{BE} ; (b) Espectrograma de x_{m3}^{BL}	34

6.7	(a) Forma de onda da palavra <i>sheet</i> do sinal de banda larga x_{m3}^{BL} ; (b) espectrograma de (a).	36
6.8	Esquema do algoritmo proposto.	36
6.9	Espectrograma do sinal x_{m3}^{BE} com as regiões a serem separadas pelo banco de filtros em destaque.	38
6.10	Esquema da geração dos canais 8 a 16 do sinal x_{m3}^{BE}	39
6.11	Módulo da resposta em frequência do filtro protótipo de 96 coeficientes apresentados na Tabela 6.2.	42
6.12	Módulo da resposta em frequência do banco de filtros modulados por cosseno a partir do filtro protótipo da Tabela 6.2.	42
7.1	Módulo das respostas em frequência dos filtros ótimos utilizados na extensão dos <i>frames</i> dos sinais de locutores femininos classificados como vozeados. Canais: (a) 8, (b) 9, (c) 10, (d) 11, (e) 12, (f) 13, (g) 14, (h) 15 e (i) 16.	44
7.2	Módulo das respostas em frequência dos filtros ótimos utilizados na extensão dos <i>frames</i> dos sinais de locutores femininos classificados como não-vozeados. Canais: (a) 8, (b) 9, (c) 10, (d) 11, (e) 12, (f) 13, (g) 14, (h) 15 e (i) 16.	45
7.3	Módulo das respostas em frequência dos filtros ótimos utilizados na extensão dos <i>frames</i> dos sinais de locutores masculinos classificados como vozeados. Canais: (a) 8, (b) 9, (c) 10, (d) 11, (e) 12, (f) 13, (g) 14, (h) 15 e (i) 16.	46
7.4	Módulo das respostas em frequência dos filtros ótimos utilizados na extensão dos <i>frames</i> dos sinais de locutores masculinos classificados como não-vozeados. Canais: (a) 8, (b) 9, (c) 10, (d) 11, (e) 12, (f) 13, (g) 14, (h) 15 e (i) 16.	47
7.5	Curvas de erro médio quadrático de treinamento e de teste da rede neural.	50
7.6	(a) Espectrograma do sinal x_{f1}^{BE} estendido, (b) espectrograma de x_{f1}^{BL} destacando-se a classificação realizada pela rede neural.	52
7.7	(a) Espectrograma do sinal x_{f2}^{BE} estendido, (b) espectrograma de x_{f2}^{BL} destacando-se a classificação realizada pela rede neural.	52

7.8	(a) Espectrograma do sinal $x_{f_3}^{BE}$ estendido, (b) espectrograma de $x_{f_3}^{BL}$ destacando-se a classificação realizada pela rede neural.	53
7.9	(a) Espectrograma do sinal $x_{m_1}^{BE}$ estendido, (b) espectrograma de $x_{m_1}^{BL}$ destacando-se a classificação realizada pela rede neural.	53
7.10	(a) Espectrograma do sinal $x_{m_2}^{BE}$ estendido, (b) espectrograma de $x_{m_2}^{BL}$ destacando-se a classificação realizada pela rede neural.	54
7.11	(a) Espectrograma do sinal $x_{m_3}^{BE}$ estendido, (b) espectrograma de $x_{m_3}^{BL}$ destacando-se a classificação realizada pela rede neural.	54
7.12	Espectrograma do sinal $x_{f_1}^{BE}$ estendido utilizando-se ganho nos trechos identificados pela rede neural como vozeados e filtragem ótima nos não-vozeados.	55
7.13	Espectrograma do sinal $x_{f_2}^{BE}$ estendido utilizando-se ganho nos trechos identificados pela rede neural como vozeados e filtragem ótima nos não-vozeados.	56
7.14	Espectrograma do sinal $x_{f_3}^{BE}$ estendido utilizando-se ganho nos trechos identificados pela rede neural como vozeados e filtragem ótima nos não-vozeados.	56
7.15	Espectrograma do sinal $x_{m_1}^{BE}$ estendido utilizando-se ganho nos trechos identificados pela rede neural como vozeados e filtragem ótima nos não-vozeados.	57
7.16	Espectrograma do sinal $x_{m_2}^{BE}$ estendido utilizando-se ganho nos trechos identificados pela rede neural como vozeados e filtragem ótima nos não-vozeados.	57
7.17	Espectrograma do sinal $x_{m_3}^{BE}$ estendido utilizando-se ganho nos trechos identificados pela rede neural como vozeados e filtragem ótima nos não-vozeados.	58
7.18	Espectrograma do sinal $x_{f_1}^{BE}$ estendido utilizando-se ganho tanto nos trechos identificados pela rede neural como vozeados quanto nos não-vozeados.	59
7.19	Espectrograma do sinal $x_{f_2}^{BE}$ estendido utilizando-se ganho tanto nos trechos identificados pela rede neural como vozeados quanto nos não-vozeados.	60

7.20	Espectrograma do sinal x_{f3}^{BE} estendido utilizando-se ganho tanto nos trechos identificados pela rede neural como vozeados quanto nos não-vozeados.	60
7.21	Espectrograma do sinal x_{m1}^{BE} estendido utilizando-se ganho tanto nos trechos identificados pela rede neural como vozeados quanto nos não-vozeados.	61
7.22	Espectrograma do sinal x_{m2}^{BE} estendido utilizando-se ganho tanto nos trechos identificados pela rede neural como vozeados quanto nos não-vozeados.	61
7.23	Espectrograma do sinal x_{m3}^{BE} estendido utilizando-se ganho tanto nos trechos identificados pela rede neural como vozeados quanto nos não-vozeados.	62
7.24	Espectrograma do sinal x_{f1}^{BE} estendido utilizando-se um algoritmo baseado em predição linear.	63
7.25	Espectrograma do sinal x_{f2}^{BE} estendido utilizando-se um algoritmo baseado em predição linear.	64
7.26	Espectrograma do sinal x_{f3}^{BE} estendido utilizando-se um algoritmo baseado em predição linear.	64
7.27	Espectrograma do sinal x_{m1}^{BE} estendido utilizando-se um algoritmo baseado em predição linear.	65
7.28	Espectrograma do sinal x_{m2}^{BE} estendido utilizando-se um algoritmo baseado em predição linear.	65
7.29	Espectrograma do sinal x_{m3}^{BE} estendido utilizando-se um algoritmo baseado em predição linear.	66
7.30	Resultado do teste subjetivo para o sinal x_{f1}^{BE} estendido.	69
7.31	Resultado do teste subjetivo para o sinal x_{f2}^{BE} estendido.	70
7.32	Resultado do teste subjetivo para o sinal x_{f3}^{BE} estendido.	70
7.33	Resultado do teste subjetivo para o sinal x_{m1}^{BE} estendido.	71
7.34	Resultado do teste subjetivo para o sinal x_{m2}^{BE} estendido.	71
7.35	Resultado do teste subjetivo para o sinal x_{m3}^{BE} estendido.	72

Lista de Tabelas

1.1	Características da resposta em frequência do filtro passa-banda ITU-T G.712.	1
6.1	Frases utilizadas nos testes do algoritmos proposto de extensão em frequência.	31
6.2	Primeiros 48 coeficientes do filtro protótipo de tamanho 96.	41
7.1	Filtros ótimos utilizados na extensão em frequência dos canais 8-16 dos <i>frames</i> dos sinais de locutores femininos classificados como vozeados.	44
7.2	Filtros ótimos utilizados na extensão em frequência dos canais 8-16 dos <i>frames</i> dos sinais de locutores femininos classificados como não-vozeados.	45
7.3	Filtros ótimos utilizados na extensão em frequência dos canais 8-16 dos <i>frames</i> dos sinais de locutores masculinos classificados como vozeados.	46
7.4	Filtros ótimos utilizados na extensão em frequência dos canais 8-16 dos <i>frames</i> dos sinais de locutores masculinos classificados como não-vozeados.	47
7.5	Tabela dos ganhos utilizados na extensão em frequência para locutores masculinos e femininos e para sons vozeados e não-vozeados.	48

Lista de Algoritmos

2.1	Regra Delta	14
2.2	Batelada	14

Capítulo 1

Introdução

Quando o sistema de telefonia fixa analógico foi desenvolvido, os requisitos de qualidade de voz da época eram diferentes dos existentes nos dias de hoje. Atualmente, a qualidade de voz requerida pelas pessoas é maior, visto que há outras fontes de áudio de alta qualidade sonora, tais como o rádio e o disco compacto (CD - *Compact Disk*). A degradação da qualidade sonora do sistema de telefonia analógico ocorre pela introdução de um filtro passa-faixas antes do sinal de voz ser transmitido pelo canal.

A Tabela 1.1 mostra as características do filtro passa-banda ($H_{tel}(z)$) definido pela ITU-T (*International Telecommunication Union - Telecommunication Standardization Sector*) [1], pelo qual o sinal de voz é processado antes de ser transmitido pelo canal, ou seja, a linha telefônica. Este filtro faz com que o sinal transmitido

Tabela 1.1: Características da resposta em frequência do filtro passa-banda ITU-T G.712.

Frequência (Hz)	Atenuação (dB)
100 - 300	0 - 10
300 - 3400	0
3400 - 4000	0 - 18

pela linha telefônica tenha seu espectro de frequências reduzido a faixa de aproximadamente 300 a 3400 Hz, diminuindo a qualidade do sinal de voz transmitido.

Uma forma de melhorar a qualidade dos sinais transmitidos pela linha telefônica seria a transmissão destes sinais com uma banda maior (até 7 ou 8 kHz, por

exemplo). Porém, isto implicaria na substituição do canal de comunicação. Uma solução mais prática seria o uso de algoritmos de extensão em frequência para aumentar a banda do sinal após ser transmitido. A idéia básica destes algoritmos é estimar, a partir das componentes de frequência entre 300 e 3400 Hz, e inserir no sinal limitado em banda, as componentes de frequência acima de 3400 Hz e abaixo de 300 Hz. A Figura 1.1 mostra o esquema de telefonia atual com o algoritmo de extensão em frequência implementado na recepção de uma conexão.

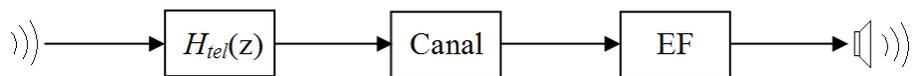


Figura 1.1: Esquema de uma conexão telefônica com módulo de extensão em frequência (EF) implementado na recepção da conexão.

Alguns métodos têm sido desenvolvidos objetivando a melhoria da qualidade de voz transmitida. Em [2] é feita uma comparação entre os métodos baseados no modelo fonte-filtro utilizando as técnicas de redes neurais e *codebook*. Em [3] é proposta uma técnica baseada em codificação por predição linear (LPC - *Linear Predictive Coding*) denominada *spectral shifting*, baseada em deslocamentos no espectro do sinal de banda estreita. Em [4] é proposto um método de extensão em frequência utilizando uma abordagem estatística baseado em um Modelo Escondido de Markov (HMM - *Hidden Markov Model*) na geração do envelope espectral do sinal de banda larga. Em [5] é apresentado um método baseado em um Modelo Escondido de Markov utilizando LSF (*Line Spectral Frequencies*) [2], a partir dos coeficientes LPC, para a realização da extensão em frequência. Em [1] é apresentado um método que combina equalização e estimação para a geração do sinal de banda larga. Em [6] é apresentada uma técnica baseada em predição linear e LSF na geração do envelope espectral do sinal de banda larga. Em [7] é introduzido o conceito de uma função custo no processo de estimação das altas frequências.

O objetivo deste trabalho é apresentar um novo método de extensão em frequência de baixa complexidade computacional baseado em filtragem ótima e redes neurais visando estender as componentes de frequências altas e melhorar a qualidade de sinais de voz limitados em banda.

Capítulo 2

Redes Neurais

2.1 Introdução

As redes neurais *feedforward* têm sido freqüentemente utilizadas para a resolução de problemas que envolvam tarefas de classificação não-linear e que realizem esta classificação a partir do treinamento de um conjunto de dados [8]. Este capítulo tem por objetivo a introdução dos principais conceitos a respeito das redes neurais *feedforward* bem como a descrição de alguns algoritmos visando o treinamento das redes. A Seção 2.2 apresenta os fundamentos das redes neurais, onde faz-se uma simples analogia ao sistema neuronal humano. Na Seção 2.3 é apresentada uma modelagem matemática para o neurônio artificial, que é o elemento básico na constituição de uma rede neural. Por fim, a Seção 2.4 descreve alguns algoritmos utilizados para o treinamento de redes neurais *feedforward* a partir de um conjunto de dados de treinamento.

2.2 Fundamentos

O desenvolvimento das Redes Neurais artificiais é baseado em conceitos biológicos, ou seja, a configuração da rede assim como os seus algoritmos geralmente baseiam-se na organização do cérebro humano. Porém, as comparações entre o cérebro humano e as redes neurais artificiais se limitam ao conhecimento do funcionamento do cérebro humano. Desta forma, a teoria das redes neurais artificiais baseia-se nos princípios do cérebro humano sem a pretensão de simulá-lo de forma

completa e precisa.

O sistema nervoso humano é constituído de células denominadas neurônios. Os neurônios possuem características comuns às outras células do corpo humano, entretanto possuem também a capacidade de receber, processar e transmitir sinais eletroquímicos por meio do sistema de comunicação do cérebro.

A Figura 2.1 mostra a representação de neurônios biológicos típicos, cujo funcionamento está descrito a seguir, de forma simplificada. Os dendritos se estendem do corpo celular até outros neurônios onde estes recebem os sinais em um ponto de conexão denominado sinapse. No lado de recepção das sinapses, estas entradas são conduzidas ao corpo celular onde são somadas. Algumas destas entradas tendem a estimular a célula e outras tendem a inibi-la. Quando esta excitação cumulativa no corpo celular ultrapassa um determinado limiar, a célula dispara, ou seja, envia um sinal por meio de seu axônio para outros neurônios.

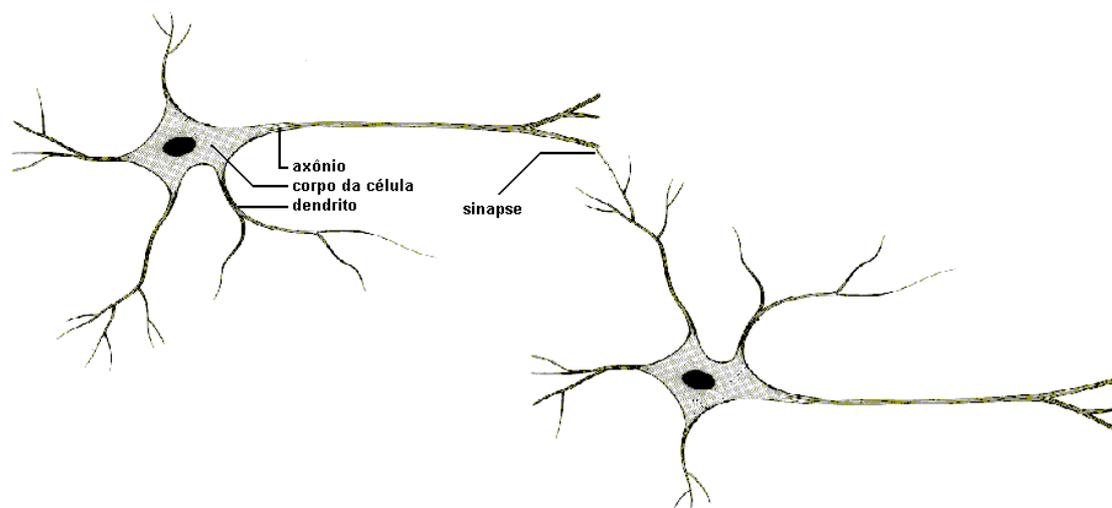


Figura 2.1: Representação de neurônios biológicos típicos.

2.3 Modelagem do Neurônio Artificial

O neurônio artificial foi desenvolvido para simular a característica de primeira ordem do neurônio biológico [9]. A Figura 2.2 representa o neurônio artificial o qual modela os princípios básicos do neurônio biológico descrito de forma sucinta na Seção 2.2. As entradas de um neurônio são as entradas da rede neural ou as saídas

de outros neurônios multiplicadas por pesos que representam as intensidades de cada sinapse. O sinal de *bias*, representado por y_0 na Figura 2.2, tem o efeito de aplicar uma transformação afim à saída u_k no modelo de neurônio não-linear [10]. As entradas e o *bias*, após serem multiplicados por seus respectivos pesos, são somados, simulando de forma simplificada o corpo celular de um neurônio biológico. Posteriormente, esta soma é processada por uma função denominada função de ativação do neurônio, gerando o sinal de saída que representa o seu nível de ativação. As equações (2.1) e (2.2) representam um modelo matemático para o neurônio artificial representado na Figura 2.2:

$$u_k = \sum_{j=0}^J w_{kj} y_j \quad (2.1)$$

$$o_k = f(u_k), \quad (2.2)$$

onde, u_k representa a soma das entradas y_j do neurônio ponderadas pelos pesos das sinapses w_{kj} , e f representa a função de ativação do neurônio k .

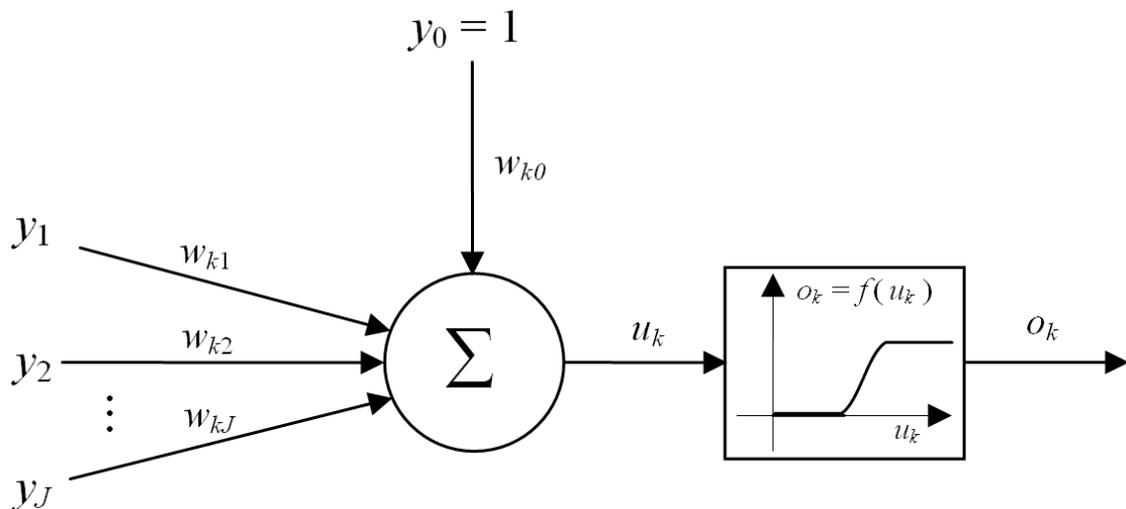


Figura 2.2: Modelo matemático do neurônio artificial.

De acordo com sua função de ativação, o neurônio artificial pode ser classificado como linear ou não-linear. O neurônio linear é caracterizado pela sua saída ser uma soma ponderada de suas entradas. O neurônio não-linear é utilizado para modelar de forma mais exata a característica não-linear do neurônio biológico, permitindo também a utilização de funções mais genéricas para a rede neural. Um exemplo desta função (geralmente utilizada) é a função tangente hiperbólica, cujo

gráfico é mostrado na Figura 2.3 e cuja expressão matemática é dada por:

$$\tanh(\alpha) = \frac{1 - e^{-2\alpha}}{1 + e^{-2\alpha}}. \quad (2.3)$$

Pode-se pensar na função de ativação como sendo um ganho não-linear para o neurônio artificial, onde para grandes valores de u_k , em módulo, o ganho é pequeno, ao passo que para pequenos valores de u_k , em módulo, o ganho é mais elevado. Desta forma, o neurônio consegue processar adequadamente uma ampla faixa de níveis de sinais de entrada.

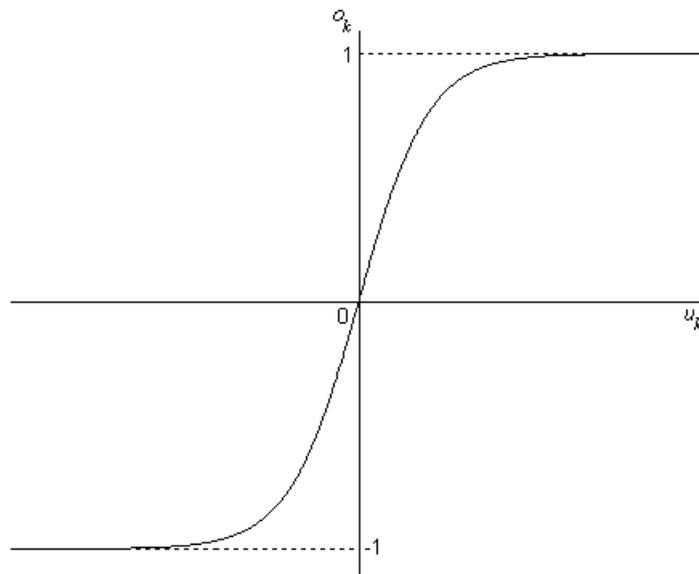


Figura 2.3: Função tangente hiperbólica.

Muitas características do neurônio biológico não são levadas em consideração por esta modelagem do neurônio artificial como, por exemplo, os atrasos que afetam a dinâmica do sistema, os efeitos de sincronismo e a função de modulação em frequência do neurônio biológico. Para alguns pesquisadores estas características seriam essenciais. Porém, a modelagem que é feita já leva em consideração algumas características do neurônio biológico, apresentando resultados significativos [9].

Uma rede neural é composta por um conjunto de neurônios dispostos em camadas. Cada camada é composta pelos valores dos pesos, ou sinapses, e pelos neurônios subsequentes que processam os sinais ponderados por estes pesos. A Figura 2.4 mostra uma rede neural de duas camadas: a camada intermediária, que processa os sinais de entrada, e a camada de saída, que gera as saídas da rede a

partir das saídas da camada intermediária. Esta rede é denominada de *feedforward* pois não apresenta realimentação (*feedback*), ou seja, a conexão de uma saída de um neurônio à entrada de outro pertence à mesma camada ou a uma camada anterior a ele.

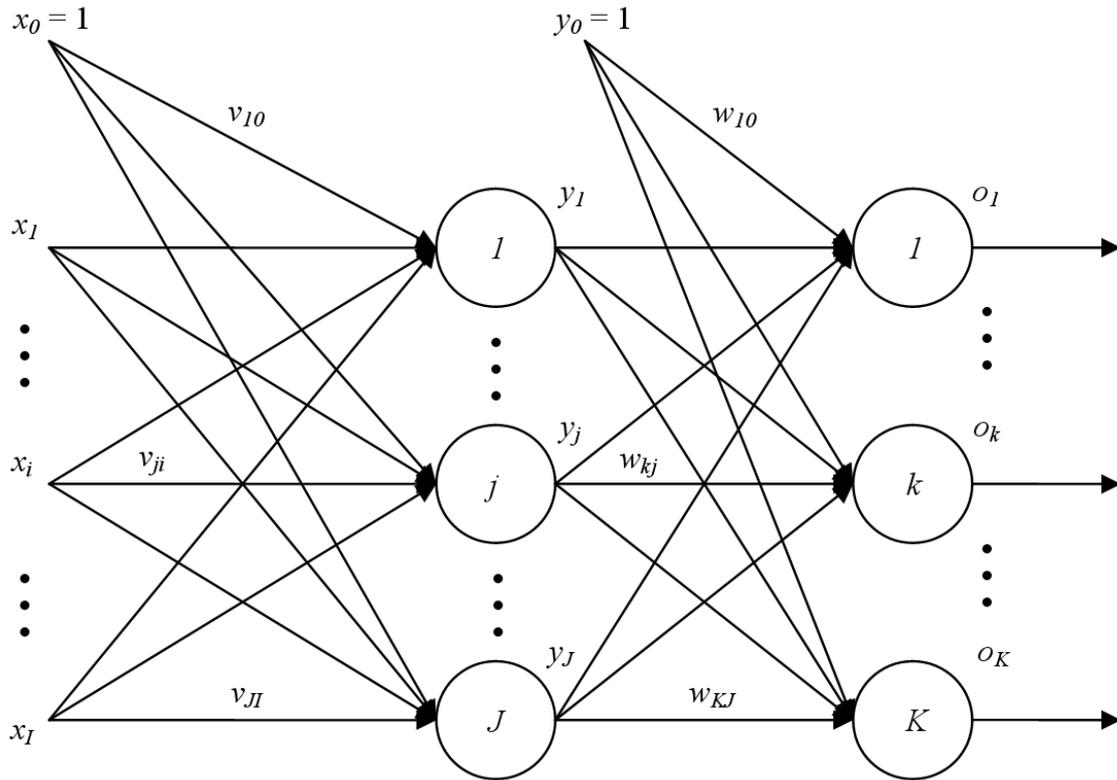


Figura 2.4: Rede neural com duas camadas: camada intermediária e camada de saída.

2.4 Treinamento das Redes Neurais

O treinamento de uma rede neural é realizado objetivando-se a obtenção de um conjunto de saídas a partir de um conjunto de entradas. Estes conjuntos iniciais de entradas e saídas são representados pelos seguintes vetores:

$$\mathbf{x}_p = [x_1 \dots x_i \dots x_I]_p^T, \quad (2.4)$$

$$\mathbf{d}_p = [d_1 \dots d_k \dots d_K]_p^T, \quad (2.5)$$

onde p representa o número do par entrada-saída.

Assim, a rede neural é treinada, ou seja, cada conjunto de entradas é aplicado à rede de forma que os seus pesos sejam ajustados para que o conjunto de saídas da rede (\mathbf{o}_p), representadas pelo vetor $\mathbf{o}_p = [o_1 \dots o_k \dots o_K]_p^T$, convirjam para o conjunto de valores desejados, \mathbf{d}_p , ou para valores próximos a ele.

O treinamento de uma rede neural pode ser classificado em supervisionado e não-supervisionado. O treinamento supervisionado requer o processamento de pares entrada-saída, chamados de pares de treinamento, onde as saídas representam o que se deseja obter quando as entradas correspondentes forem aplicadas à rede. Cada entrada é processada pela rede e gera um sinal de saída. Este último é então comparado à saída desejada, gerando um sinal de erro. De acordo com o sinal de erro gerado, o algoritmo de treinamento efetua correções sobre os valores das sinapses da rede de forma que o erro seja minimizado. Este procedimento é repetido sequencialmente até que o erro de todo o conjunto de treinamento atinja um valor adequado. Já no treinamento não-supervisionado, o treinamento é realizado sem os sinais de saída desejados.

As sinapses, ou seja, os pesos da rede, influenciam diretamente o desempenho do modelo treinado pela rede neural. Para que este desempenho possa ser avaliado, define-se uma medida da qualidade do treinamento realizado pela rede. Em termos matemáticos, isto é feito por uma função denominada de função objetivo (F). Esta função expressa uma medida do erro entre a saída desejada e a saída da rede.

Geralmente, os modelos a serem treinados, ou seja, aqueles representados pela rede neural, são aproximações não-lineares. Isto porque, a função objetivo, F , apresenta uma relação não-linear com as sinapses da rede, sendo estas diretamente ligadas ao processo de aprendizado da rede, ou seja, de treinamento. Isto significa que a superfície de erro pode não ser uma superfície convexa, com a existência garantida de um mínimo. Desta forma, a busca pelo conjunto de pesos que irá garantir a melhor representatividade do modelo treinado pela rede recai em um problema de otimização não-linear. Não há um processo de otimização não-linear geral para fins de aprendizado (treinamento) [8]. Contudo, um dos métodos mais conhecidos para a resolução deste problema é o método do gradiente descendente que realiza uma busca por um mínimo na superfície da função erro, ou função objetivo, F . Porém, este processo de treinamento não garante a obtenção do melhor grupo

de sinapses para a representação do modelo em questão.

2.4.1 O Algoritmo do Gradiente Descendente

As componentes do gradiente da função objetivo são dadas por:

$$\nabla_{kj} F = \frac{\partial F}{\partial w_{kj}}, \quad (2.6)$$

onde F é a função objetivo ou função custo. A direção do vetor gradiente, no ponto em que é calculado, é a direção de máximo crescimento da função sobre a qual o gradiente é calculado. Assim, para que se encontre o mínimo local de uma função, deve-se seguir o sentido contrário ao do gradiente. Desta forma, o gradiente da função objetivo em relação às sinapses da rede é utilizado na atualização das sinapses. A lei de aprendizado das sinapses utilizando-se o algoritmo do gradiente descendente é dada por:

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n) = w_{kj}(n) - \mu \frac{\partial F(n)}{\partial w_{kj}}, \quad (2.7)$$

onde n representa o número da iteração na qual ocorrerá a atualização das sinapses e μ é o passo (ou taxa) de treinamento e expressa o avanço em direção de um mínimo local ou do mínimo global da função objetivo. O módulo de μ não deve ser muito grande, pois pode acarretar problemas de instabilidade do algoritmo.

2.4.2 O Algoritmo de Retropropagação do Erro - RPE

O algoritmo de retropropagação do erro (*backpropagation*), ou simplesmente RPE, é um método de otimização de primeira ordem que utiliza o método do gradiente descendente para a correção do valor das sinapses da rede. É um dos mais utilizados para o treinamento de redes neurais. Sua idéia básica consiste na utilização dos sinais de erro provenientes da camada de saída da rede e sua posterior retropropagação pela rede para o cálculo da atualização das sinapses.

Como dito anteriormente, a função objetivo expressa uma medida do erro apresentado pela rede. Assim, considerando a rede neural apresentada na Figura 2.4 e de acordo com [11], a função objetivo pode ser escrita como:

$$F = \frac{1}{P} \sum_{p=1}^P \sum_{k=1}^K (e_{kp})^2, \quad (2.8)$$

$$e_{kp} = d_{kp} - o_{kp} = d_{kp} - f(u_{kp}), \quad (2.9)$$

onde K representa o número de neurônios da camada de saída da rede e P , o número de pares de treinamento da rede. Se a cada iteração, ou seja, a cada novo par entrada-saída, as sinapses forem atualizadas, então a Equação 2.8 apresentará apenas o somatório em k . Esta forma de atualização das sinapses é denominada de Regra Delta. No caso de a atualização ocorrer apenas após o processamento de todos os pares entrada-saída, então o método de correção das sinapses é denominado de Batelada.

A atualização das sinapses é realizada pelo método do gradiente descendente a partir da Equação 2.7. Aplicando-se a regra da cadeia à Equação 2.6, tem-se que o gradiente da função erro em relação às sinapses da rede é equivalente a:

$$\frac{\partial F}{\partial w_{kj}} = \frac{\partial F}{\partial u_k} \frac{\partial u_k}{\partial w_{kj}}, \quad (2.10)$$

$$\delta_{ok} = -\frac{1}{2} \frac{\partial F}{\partial u_k} = e_k f'(u_k), \quad (2.11)$$

onde δ_{ok} representa um sinal de erro, geralmente, diferente de e_k e o subscrito o indica que este sinal de erro pertence à camada de saída da rede [8]. A partir da Equação 2.1 verifica-se que a derivada parcial de u_k em relação a w_{kj} é igual a y_j . Assim, a Equação 2.10 pode ser reescrita utilizando-se novamente a regra da cadeia:

$$\frac{\partial F}{\partial w_{kj}} = \frac{\partial F}{\partial o_k} \frac{\partial o_k}{\partial u_k} y_j = -2e_k f'(u_k) y_j = -2\delta_{ok} y_j, \quad (2.12)$$

onde $f'(u_k)$ representa a derivada da função de ativação em relação a u_k do neurônio k da camada de saída da rede. Assim, a Equação 2.7 pode ser reescrita como:

$$\begin{aligned} w_{kj}(n+1) &= w_{kj}(n) + 2\mu e_k f'(u_k) y_j \\ &= w_{kj}(n) + 2\mu \delta_{ok} y_j. \end{aligned} \quad (2.13)$$

Esta é a equação de aprendizado das sinapses dos neurônios da camada de saída da rede neural representada pela Figura 2.4.

Devido à utilização da derivada da função de ativação no processo de treinamento da rede, a função de ativação dos neurônios deve ser diferenciável.

De forma análoga à camada de saída, pode-se deduzir a equação de aprendizado para os neurônios da camada intermediária da rede da Figura 2.4. Assim, o termo correspondente à atualização das sinapses da camada intermediária é dado por:

$$\Delta v_{ji} = -\mu \frac{\partial F}{\partial v_{ji}} = -\mu \frac{\partial F}{\partial t_j} \frac{\partial t_j}{\partial v_{ji}} = 2\mu \delta_{yj} x_i, \quad (2.14)$$

com

$$t_j = \sum_{i=0}^I v_{ji} x_i, \quad 1 \leq j \leq J, \quad (2.15)$$

$$y_j = f(t_j). \quad (2.16)$$

Verifica-se que a derivada de t_j em relação as sinapses da camada intermediária, v_{ji} , é igual a x_i e:

$$\delta_{yj} = -\frac{1}{2} \frac{\partial F}{\partial t_j}. \quad (2.17)$$

O subscrito y do sinal de erro significa que este pertence à camada intermediária da rede.

Durante a formulação deste método de atualização das sinapses, a maior dificuldade encontrada pelos pesquisadores que o desenvolviam foi o cálculo dos termos de erro da camada intermediária da rede. Esta dificuldade se dava devido ao termo t_j contribuir com os erros da camada de saída da rede.

Com a utilização da regra da cadeia, a Equação 2.17 pode ser reescrita como:

$$\delta_{yj} = -\frac{1}{2} \frac{\partial F}{\partial y_j} \frac{\partial y_j}{\partial t_j}, \quad (2.18)$$

onde a derivada parcial de y_j em relação a t_j é representada por $f'(t_j)$. A derivada parcial da função objetivo em relação a y_j pode ser escrita como:

$$\begin{aligned} \frac{\partial F}{\partial y_j} &= \frac{\partial}{\partial y_j} \left\{ \sum_{k=1}^K [d_k - f(u_k(\mathbf{y}))]^2 \right\} \\ &= -2 \sum_{k=1}^K [d_k - o_k] \frac{\partial}{\partial y_j} \{f(u_k(\mathbf{y}))\}, \end{aligned} \quad (2.19)$$

onde \mathbf{y} representa o vetor contendo as saídas dos neurônios da camada intermediária da rede e é dado por:

$$\mathbf{y} = [y_1 \dots y_j \dots y_J]^{\mathbf{T}}. \quad (2.20)$$

O vetor \mathbf{y} foi escrito na Equação 2.19 para destacar a dependência das saídas dos neurônios da camada de saída em relação às saídas da camada intermediária, representadas pelo vetor \mathbf{y} . Utilizando-se a regra da cadeia na derivada parcial da Equação 2.19 tem-se que:

$$\begin{aligned} \frac{\partial}{\partial y_j} f(u_k(\mathbf{y})) &= \frac{\partial}{\partial u_k} f(u_k(\mathbf{y})) \frac{\partial u_k}{\partial y_j} \\ &= f'(u_k) w_{kj}. \end{aligned} \quad (2.21)$$

Desta forma, a Equação 2.19 pode ser reescrita como:

$$\begin{aligned}\frac{\partial F}{\partial y_j} &= -2 \sum_{k=1}^K e_k f'(u_k) w_{kj} \\ &= -2 \sum_{k=1}^K \delta_{ok} w_{kj}.\end{aligned}\quad (2.22)$$

A partir das Equações 2.18 e 2.22 pode-se deduzir o termo representativo do erro da camada intermediária como sendo:

$$\delta_{yj} = f'(u_j) \sum_{k=1}^K \delta_{ok} w_{kj}.\quad (2.23)$$

A partir das Equações 2.14 e 2.23 pode-se deduzir a equação de aprendizado das sinapses da camada intermediária da rede da Figura 2.4 como sendo:

$$v_{ji}(n+1) = v_{ji}(n) + 2\mu x_i f'(u_j) \sum_{k=1}^K \delta_{ok} w_{kj}.\quad (2.24)$$

As derivadas das funções de ativação dos neurônios das camadas intermediária e de saída são necessárias para o cálculo dos erros, δ , destas camadas. No caso de neurônios lineares, a derivada da função de ativação é igual a 1. No caso de neurônios com função de ativação igual à tangente hiperbólica (Equação 2.3), sua derivada pode ser expressa, em termos da saída do neurônio, como:

$$f'(u) = 1 - o^2\quad (2.25)$$

$$f'(t) = 1 - y^2,\quad (2.26)$$

para a camada de saída e intermediária da rede, respectivamente.

O Algoritmo 2.1 resume o procedimento de cálculo do algoritmo RPE para o caso da atualização das sinapses ser efetuado utilizando-se a Regra Delta. Neste método de atualização, um conjunto de entradas é aplicado à rede e, na etapa de retropropagação do erro, as sinapses são atualizadas antes que o próximo conjunto de entradas seja aplicado à rede. De forma diferente do algoritmo Regra Delta, o algoritmo batelada utiliza a fórmula do erro dada pela Equação 2.8.

O Algoritmo 2.2 resume o procedimento de cálculo do algoritmo RPE para o caso da atualização das sinapses ser efetuado utilizando-se o método Batelada. Neste método, a atualização das sinapses só ocorre após todos os pares terem sido processados pela rede. Os testes de parada deste algoritmo foram omitidos por serem semelhantes aos do Algoritmo 2.1.

1. Escolher o passo de treinamento μ e definir o valor de erro desejado F_{des} ou estabelecer um número máximo de iterações, N_{iter} ;
2. Inicializar os valores das sinapses das camadas intermediária e de saída;
3. Treinar a rede com o próximo par entrada-saída $(\mathbf{x}_p, \mathbf{d}_p)$;
4. Calcular as saídas das camadas intermediária e de saída da rede:

$$y_{jp} = f_h(u_{jp}), \quad (2.27)$$

$$o_{kp} = f_o(u_{kp}); \quad (2.28)$$

5. Calcular a soma dos erros quadráticos de acordo com o par de treinamento utilizado e com as sinapses das camadas intermediária e de saída:

$$F_p = \sum_{k=1}^K (d_{pk} - o_{pk})^2. \quad (2.29)$$

Caso F_{des} tenha sido definido testar se $F_p < F_{des}$. Se verdadeiro terminar o algoritmo. Caso contrário ir para o item 6.

Etapa de retropropagação do erro:

6. Calcular os sinais de erro da camada de saída da rede:

$$\delta_{okp} = (d_{kp} - o_{kp}) f'_{ok}(u_{kp}), \quad k = 1, \dots, K; \quad (2.30)$$

7. Calcular os sinais de erro da camada intermediária da rede:

$$\delta_{yjp} = f'_{hj}(u_{jp}) \sum_{k=1}^K \delta_{okp} w_{kjp}, \quad j = 1, \dots, J; \quad (2.31)$$

8. Atualizar os valores das sinapses da camada de saída da rede:

$$w_{kj,p+1} = w_{kjp} + 2\mu \delta_{okp} y_{jp}; \quad (2.32)$$

9. Atualizar os valores das sinapses da camada intermediária da rede:

$$v_{ji,p+1} = v_{jip} + 2\mu \delta_{yjp} x_{ip}, \quad (2.33)$$

10. Se $p < P$ voltar para o item 3. Se $p = P$ fazer $n = n + 1$ e se $n \geq N$ terminar o algoritmo. Caso contrário voltar para o item 3.

Algoritmo 2.1: Regra Delta

Etapa de retropropagação do erro:

Os valores das sinapses w_{kj} e v_{ji} permanecem fixos durante o treinamento dos P pares, ou seja, para cada novo valor de n :

6. Calcular os valores das atualizações das sinapses da camada de saída da rede para $k = 1, \dots, K$ e para $j = 0, \dots, J$:

$$\Delta w_{kj} = \sum_{p=1}^P (d_{kp} - o_{kp}) f'_{ok}(u_{kp}) y_{jp}; \quad (2.34)$$

7. Calcular os valores das atualizações das sinapses da camada intermediária da rede para $j = 1, \dots, J$ e para $i = 0, \dots, I$:

$$\Delta v_{ji} = \sum_{p=1}^P x_{ip} f'_{hj}(u_{jp}) \sum_{k=1}^K \delta_{okp} w_{kjp}; \quad (2.35)$$

8. Atualizar os valores das sinapses da camada de saída da rede:

$$w_{kj} = w_{kj} + \frac{2}{P} \mu \Delta w_{kj}; \quad (2.36)$$

9. Atualizar os valores das sinapses da camada intermediária da rede:

$$v_{ji} = v_{ji} + \frac{2}{P} \mu \Delta v_{ji}. \quad (2.37)$$

Algoritmo 2.2: Batelada

Capítulo 3

Método dos Mínimos Quadrados

3.1 Introdução

Este capítulo tem por objetivo apresentar o método dos mínimos quadrados para a obtenção de filtros ótimos a partir de um conjunto de entradas e saídas. A Seção 3.2 apresenta uma breve descrição do método de estimação linear baseado no algoritmo dos mínimos quadrados. Por fim, a Seção 3.3 mostra o procedimento de obtenção do filtro ótimo pela minimização de uma função custo.

3.2 Estimação Linear

O Método dos Mínimos Quadrados é um método de estimação linear que não se baseia em estatísticas a respeito do sinal de entrada para a obtenção do filtro correspondente a este sinal de entrada. Supondo-se que se queira, a partir de um dado sinal de entrada, obter uma curva que se ajuste a estes dados de entrada de alguma maneira ótima, então, o método dos mínimos quadrados realiza esta tarefa minimizando a soma dos erros quadráticos entre o sinal de entrada e a curva que se deseje ajustar.

O método dos mínimos quadrados é uma alternativa determinística à teoria de filtragem de Wiener [12] que obtém os filtros ótimos de uma maneira probabilística, enquanto que o método tratado neste capítulo é determinístico.

Considerando um fenômeno físico que apresente um conjunto de entradas e saídas dadas por $x(i)$ e $d(i)$, respectivamente, de forma que, no instante i , a saída

$d(i)$ seja gerada por meio do conjunto de entradas $x(i), x(i-1), \dots, x(i-M+1)$, uma modelagem linear possível para determinado fenômeno pode ser expressa de forma matemática por:

$$d(i) = \sum_{k=0}^{M-1} h_m(k) x(i-k) + e_m(i), \quad (3.1)$$

onde $h_m(\cdot)$ representa os parâmetros do modelo que se deseja obter e $e_m(\cdot)$ representa o erro de medida cuja natureza estatística do fenômeno está relacionada [12].

O erro $e_m(\cdot)$ representa uma variável aleatória e é introduzido no modelo matemático para levar em conta a falta de acurácia na modelagem do sistema. Considerando-se o erro como sendo branco de média zero e variância σ^2 , então a Equação 3.1 pode ser escrita como:

$$E[d(i)] = \sum_{k=0}^{M-1} h_m(k) x(i-k), \quad (3.2)$$

onde o operador $E[\cdot]$ representa o valor esperado. Assim, uma média da resposta $d(\cdot)$ pode ser obtida conhecendo-se o modelo do processo.

Desta forma, o que se deseja é obter uma estimativa para os parâmetros $h_m(\cdot)$ utilizando-se os parâmetros $d(\cdot)$, saídas desejadas para o sistema, e $u(\cdot)$, entradas para o processo, conhecidos. Para isso, define-se a estrutura de filtragem linear da Figura 3.1 para representar o modelo de interesse. Utilizando-se o produto entre as entradas do filtro $x(i), x(i-1), \dots, x(i-M+1)$ e seus coeficientes $h(0), h(1), \dots, h(M-1)$ define-se o erro de estimação dado pela diferença entre a resposta desejada $d(i)$ do sistema e a saída do filtro $y(i)$:

$$e(i) = d(i) - y(i), \quad (3.3)$$

onde a saída do filtro $y(i)$ é definida por:

$$y(i) = \sum_{k=0}^{M-1} h(k) x(i-k). \quad (3.4)$$

A Figura 3.1 apresenta o esquema de filtragem graficamente, onde os termos z^{-1} representam atrasos no tempo.

Como dito anteriormente, no método dos mínimos quadrados minimiza-se uma função custo dada pela soma dos erros quadráticos, ou seja:

$$F(h(0), \dots, h(M-1)) = \sum_{i=M}^N |e(i)|^2, \quad (3.5)$$

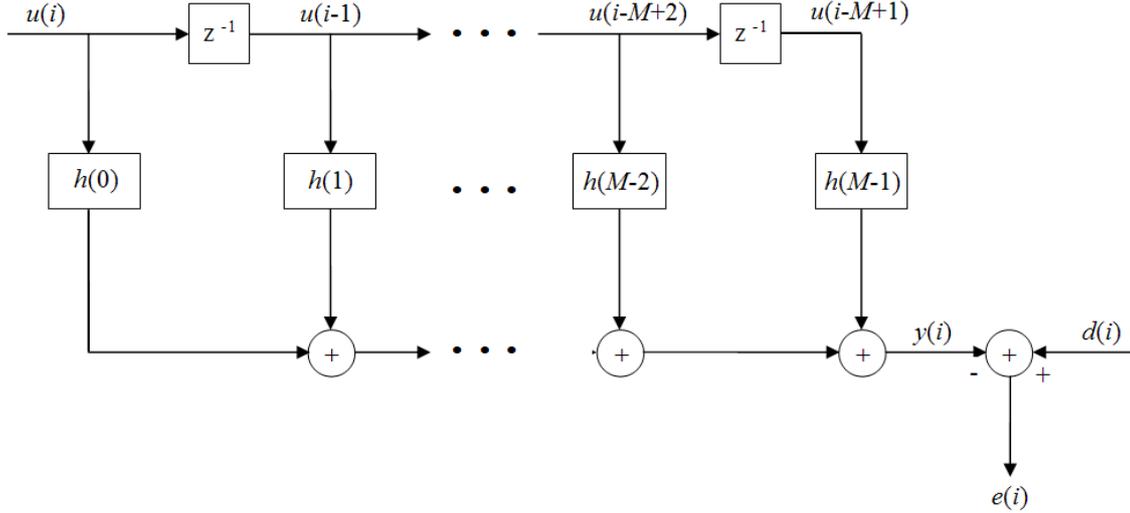


Figura 3.1: Esquema de filtragem para a modelagem de processo.

onde M é o número de coeficientes do filtro que se deseja obter e N é o número de amostras da entrada utilizadas para a geração do sinal erro, e representam os limites utilizados para realizar a minimização da função custo.

3.3 Filtragem Linear no Sentido dos Mínimos Quadrados

Objetivando a minimização da função custo da Equação 3.5, calculando-se o gradiente desta função e igualando-o a zero [12] é possível obter a equação:

$$\sum_{i=M}^N x(i-k) e_{min}(i) = 0, \quad k = 0, 1, \dots, M-1, \quad (3.6)$$

onde $e_{min}(i)$ representa o erro que se obtém pela modelagem com a minimização do função custo. A Equação 3.6 representa o princípio da ortogonalidade na versão temporal [12]. O termo da esquerda da Equação 3.6 é a correlação cruzada entre a entrada $x(i-k)$ e o erro de estimação mínimo $e_{min}(i)$. Estas duas variáveis são ortogonais quando os coeficientes do filtro estão operando na condição de mínimos quadrados.

A partir da combinação das Equações 3.3 e 3.4 com os coeficientes do filtro

($\hat{h}(\cdot)$) obtidos pelo método dos mínimos quadrados, tem-se que:

$$e_{min}(i) = d(i) - \sum_{l=0}^{M-1} \hat{h}(l) x(i-l). \quad (3.7)$$

Combinando as Equações 3.6 e 3.7 obtém-se um sistema de M equações dadas por:

$$\sum_{l=0}^{M-1} \hat{h}(l) \sum_{i=M}^N x(i-k) x(i-l) = \sum_{i=M}^N x(i-k) d(i), \quad k = 0, \dots, M-1. \quad (3.8)$$

O somatório em i do termo da esquerda da Equação 3.8 é a função de autocorrelação determinística das entradas do filtro e é representada por:

$$\phi(l, k) = \sum_{i=M}^N x(i-k) x(i-l), \quad 0 \leq l, k \leq M-1. \quad (3.9)$$

O somatório, também em i , do termo da direita da Equação 3.8 é a correlação cruzada determinística entre as entradas do filtro e o sinal de saída desejado:

$$r(-k) = \sum_{i=M}^N x(i-k) d(i), \quad 0 \leq k \leq M-1. \quad (3.10)$$

Assim, pode-se reescrever a Equação 3.8 como:

$$\sum_{l=0}^{M-1} \hat{h}(l) \phi(l, k) = r(-k), \quad k = 0, \dots, M-1. \quad (3.11)$$

De forma a facilitar o cálculo do filtro ótimo no sentido dos mínimos quadrados, pode-se reescrever a Equação 3.11 em forma matricial. Para isso, são definidas três matrizes, enumeradas a seguir:

1. A matriz de correlação determinística ($M \times M$) das entradas do filtro $u(i)$, $u(i-1)$, \dots , $u(i-M+1)$:

$$\Phi = \begin{bmatrix} \phi(0,0) & \phi(1,0) & \dots & \phi(M-1,0) \\ \phi(0,1) & \phi(1,1) & \dots & \phi(M-1,1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(0,M-1) & \phi(1,M-1) & \dots & \phi(M-1,M-1) \end{bmatrix}. \quad (3.12)$$

2. O vetor de correlação cruzada determinística ($M \times 1$) entre as entradas do filtro $u(i)$, $u(i-1)$, \dots , $u(i-M+1)$ e a saída desejada $d(i)$:

$$\mathbf{r} = \begin{bmatrix} r(0) & r(-1) & \dots & r(-M+1) \end{bmatrix}^T. \quad (3.13)$$

3. O vetor ($M \times 1$) contendo os coeficientes do filtro linear ótimo no sentido dos mínimos quadrados:

$$\hat{\mathbf{w}} = \left[\hat{h}(0) \quad \hat{h}(1) \quad \dots \quad \hat{h}(M-1) \right]^T. \quad (3.14)$$

A partir das definições das Equações 3.12-3.14 obtém-se então a equação matricial:

$$\mathbf{\Phi} \hat{\mathbf{w}} = \mathbf{r}. \quad (3.15)$$

Assim, é possível encontrar o filtro ótimo no sentido dos mínimos quadrados dado pela equação matricial abaixo:

$$\hat{\mathbf{w}} = \mathbf{\Phi}^{-1} \mathbf{r}. \quad (3.16)$$

Capítulo 4

Bancos de Filtros Modulados por Cosseno

4.1 Introdução

Este capítulo tem por finalidade apresentar a obtenção de bancos de filtros modulados por cosseno a partir de um filtro protótipo. A Seção 4.2 mostra os requisitos de reconstrução perfeita que a função de transferência de um banco de filtros maximamente decimado deve apresentar. Na Seção 4.3 é mostrada a obtenção dos filtros modulados por cosseno.

4.2 Fundamentos

A Figura 4.1 mostra um esquema de um banco de filtros maximamente decimado de M canais, usualmente denominado de QMF (*Quadrature Mirror Filter*) de M canais [13]. Primeiramente, o sinal $x(n)$ é filtrado pelos M filtros de análise, $H_k(z)$, dando origem aos sinais das M subbandas $x_k(n)$. A Figura 4.2 apresenta os módulos das respostas em frequência típicas dos filtros de análise. Posteriormente, os sinais $x_k(n)$ são subamostrados por M gerando os sinais $v_k(n)$. Estes sinais são então superamostrados por M , $u_k(n)$, e, por fim, são recombinados para obter o sinal de saída $\hat{x}(n)$. O sinal de saída do banco de filtros pode ser escrito, em

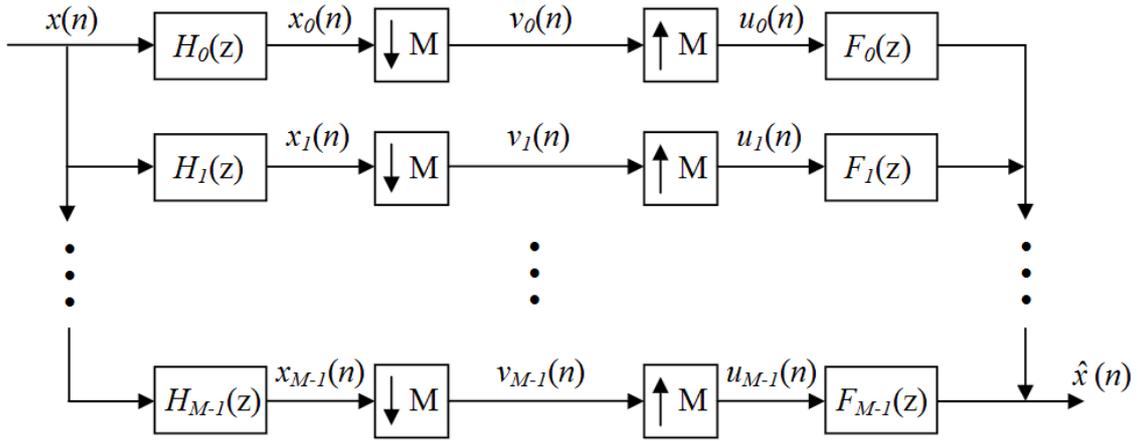


Figura 4.1: Banco de filtros maximamente decimado com M canais.

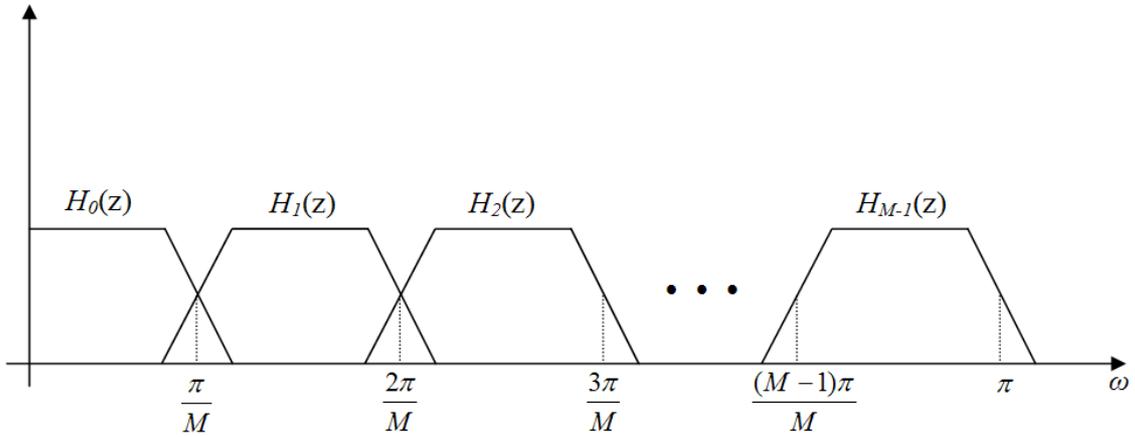


Figura 4.2: Módulos das respostas em frequência típicas de um banco de filtros maximamente decimado com M canais.

termos da Transformada Z, como [13]:

$$\hat{X}(z) = \sum_{l=0}^{M-1} A_l(z) X(zW^l), \quad (4.1)$$

onde

$$A_l(z) = \frac{1}{M} \sum_{k=0}^{M-1} H_k(zW^l) F_k(z), \quad 0 \leq l \leq M-1, \quad (4.2)$$

e $W = e^{-j2\pi/M}$, com $j = \sqrt{-1}$. Desta forma, o espectro do sinal ($\hat{X}(e^{j\omega})$) reconstruído é constituído por uma combinação de $X(e^{j\omega})$ com suas $M-1$ versões deslocadas na frequência.

De maneira geral, o sinal $\hat{x}(n)$ difere de $x(n)$ devido a alguns fatores, tais

como: *aliasing*, *imaging*, distorções de fase e de amplitude. A presença dos espectros deslocados de $X(z)$ é resultante dos fenômenos de *aliasing* e *imaging* oriundos dos processos de redução e expansão da taxa de amostragem, respectivamente. Para que não haja distorção devido a estes fenômenos, a seguinte condição deve ser satisfeita:

$$A_l(z) = 0 \quad 1 \leq l \leq M - 1. \quad (4.3)$$

Sendo esta condição satisfeita, então o sinal reconstruído resulta em:

$$\hat{X}(z) = T(z)X(z), \quad (4.4)$$

onde

$$T(z) = A_0(z) = \frac{1}{M} \sum_{k=0}^{M-1} H_k(z) F_k(z), \quad (4.5)$$

sendo $T(z)$ a função de distorção. Denomina-se distorção de amplitude quando $|T(e^{j\omega})|$ não é constante. A distorção de fase ocorre quando a fase de $T(e^{j\omega})$ não é linear. Um banco de filtros possui a propriedade de reconstrução perfeita quando os filtros $H_k(z)$ e $F_k(z)$ são projetados de forma que as distorções de *aliasing* e *imaging* são completamente canceladas e quando $T(z) = cz^{-n_0}$, ou seja, a função de distorção corresponde a apenas um atraso. Assim, o sinal reconstruído é $\hat{x}(n) = cx(n - n_0)$, onde c é uma constante diferente de zero e n_0 é uma constante inteira. Existem sistemas em que os filtros de análise e síntese são projetados para eliminar as distorções de *aliasing* e *imaging* mas que apenas aproximam $T(z)$ de um atraso. Estes bancos de filtros são denominados de sistemas de reconstrução quase-perfeita ou pseudo QMF [13].

4.3 Expressões para os Filtros

Em um banco de filtros modulados por cosseno os filtros de análise são derivados de um filtro protótipo $P_0(z)$ a partir de modulação por cosseno. O filtro protótipo pode ser projetado para que o banco de filtros seja de reconstrução perfeita ou de reconstrução quase-perfeita (pseudo QMF).

A idéia do banco de filtros modulados por cosseno é de, inicialmente, se obter $2M$ filtros complexos obtidos por modulação exponencial e, posteriormente, combinar os pares de filtros apropriados para a obtenção de M obtidos por modulação por

cossenos. Esta combinação é feita de tal forma que os filtros resultantes, de análise e de síntese, são calculados para eliminar as distorções de *aliasing* e *imaging*. Com estas distorções praticamente eliminadas, a função de distorção pode ser expressa pela Equação 4.5. Considerando-se que o banco de filtros é constituído por filtros FIR (*Finite Impulse Response*), para que a distorção de fase seja eliminada, $T(e^{j\omega})$ deve apresentar fase linear. Esta característica pode ser conseguida com a condição:

$$f_k(n) = h_k(N - n), \quad (4.6)$$

onde N representa a ordem dos filtros. Considerando-se $h_k(n)$ com coeficientes reais, então, a função de distorção da Equação 4.5 pode ser reescrita como:

$$T(e^{j\omega}) = \frac{e^{-j\omega N}}{M} \sum_{k=0}^{M-1} |H_k(e^{j\omega})|^2, \quad (4.7)$$

o que mostra que $T(e^{j\omega})$ possui fase linear para esta escolha dos filtros.

Com os requisitos de eliminação dos termos de *aliasing* e de distorção de fase, os filtros de análise e de síntese podem ser obtidos a partir do filtro protótipo $p_0(n)$ por modulação por cosseno da seguinte forma [13]:

$$h_k(n) = 2p_0(n) \cos\left(\frac{\pi}{M}\left(k + \frac{1}{2}\right)\left(n - \frac{N}{2}\right) + (-1)^k \frac{\pi}{4}\right), \quad (4.8)$$

$$f_k(n) = 2p_0(n) \cos\left(\frac{\pi}{M}\left(k + \frac{1}{2}\right)\left(n - \frac{N}{2}\right) - (-1)^k \frac{\pi}{4}\right). \quad (4.9)$$

As distorções de amplitude podem ser reduzidas dando origem ao banco pseudo QMF, ou totalmente eliminadas, resultando no banco QMF com reconstrução perfeita [13], dependendo do projeto do filtro protótipo $p_0(n)$.

Capítulo 5

Técnicas de Extensão em Frequência

5.1 Introdução

Este capítulo apresenta alguns conceitos básicos sobre a voz, utilizados no desenvolvimento do presente trabalho, bem como as principais técnicas de extensão em frequência geralmente utilizadas. A Seção 5.2 apresenta o mecanismo da geração de sons pelo sistema fonador humano. A Seção 5.3 mostra uma breve descrição dos principais métodos utilizados no processo de extensão em frequência.

5.2 Conceitos Básicos sobre a Fala

A Figura 5.1 destaca as partes principais do sistema fonador humano. O trato vocal começa nas cordas vocais e termina nos lábios, compreendendo a conexão do esôfago com a boca e a cavidade oral. O sistema subglotal, composto pelos pulmões, brônquios e traquéia, funciona como fonte de energia para a produção da fala. A fala é a onda acústica oriunda deste sistema quando o ar é expelido pelos pulmões e o fluxo de ar resultante sofre perturbações por alguma constrição em algum lugar do trato vocal.

Os sons que compõem a fala são classificados em três classes, segundo o seu modo de excitação. Os sons vozeados (som /i:/, por exemplo) são produzidos pela passagem de ar pelas cordas vocais, quando estas estão tensionadas, fazendo com que elas vibrem, produzindo pulsos de ar quase periódicos que excitam o trato vocal. Os sons não-vozeados (som /f/, que representa o som "sh", por exemplo)

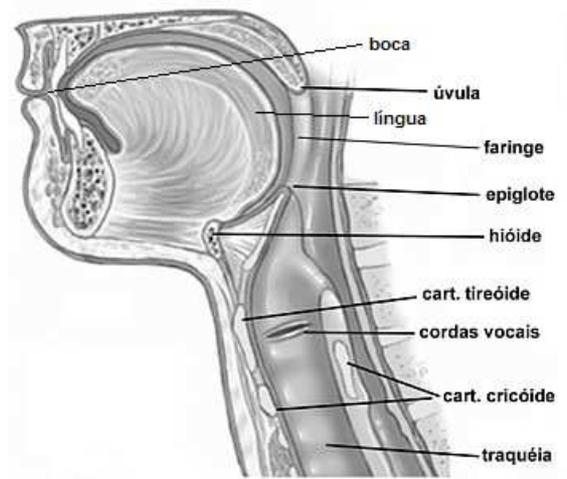


Figura 5.1: Representação do trato vocal humano e seus principais constituintes.

são gerados pela formação de uma constrição em algum ponto do trato vocal e pela passagem forçada de ar por essa constrição de forma a produzir uma turbulência. Este modo de excitação cria um ruído de banda larga que excita o trato vocal. Os sons plosivos (som /t/, por exemplo) resultam de um completo fechamento, perto da boca, seguido de um aumento de pressão atrás deste fechamento e, em seguida, pela liberação abrupta do ar. Os sons plosivos são precedidos de um pequeno período de quase ausência sonora devido ao tempo de fechamento completo do trato vocal.

Devido à propagação do som pelas cavidades do sistema fonador humano, o espectro de frequências do som é modelado de acordo com a seletividade deste sistema. As diferenças entre os sons são devidas às alterações do formato do trato vocal. Desta forma, as propriedades espectrais de um sinal de voz variam com o tempo de acordo com a mudança do trato vocal. As características espectrais variantes no tempo do sinal de voz podem ser visualizadas por um gráfico bi-dimensional denominado espectrograma. O eixo vertical do espectrograma é o eixo das frequências e o eixo horizontal representa o tempo. Um exemplo de espectrograma é mostrado na Figura 5.2. As cores no espectrograma são proporcionais à energia do sinal de voz. Quanto mais próximo do vermelho maior a energia, e quanto mais próximo do azul escuro menor é a energia em uma determinada região do espectrograma.

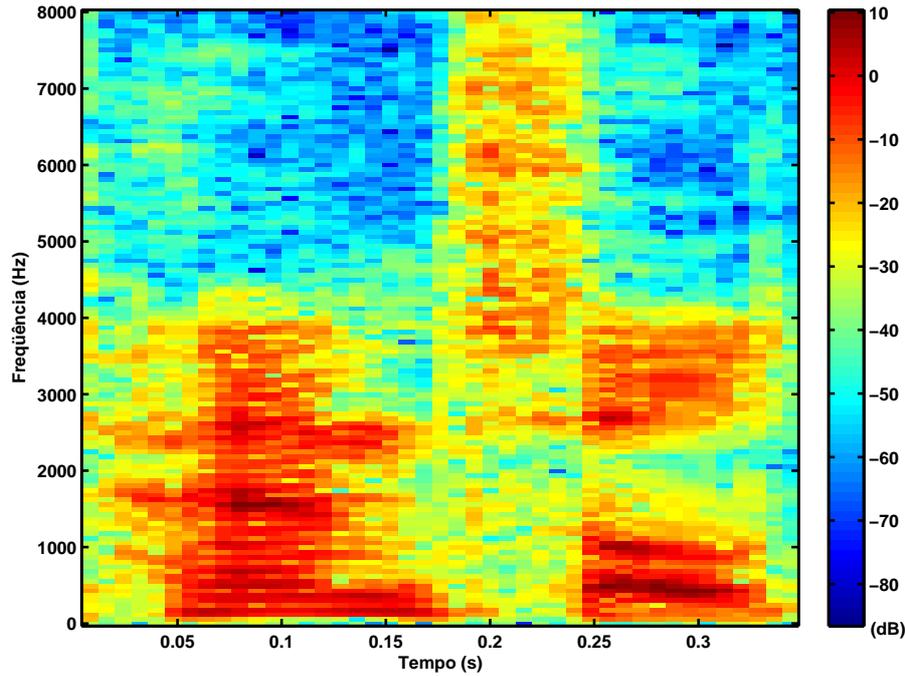


Figura 5.2: Espectrograma do sinal correspondente à palavra *pencil* proferida por locutor masculino.

5.3 Extensão em Frequência

As primeiras tentativas de utilização da extensão em frequência não se baseavam em nenhum modelo particular do processo de geração de voz e também não consideravam as propriedades da voz. Tais abordagens baseavam-se em duas metodologias principais:

- a primeira faz uso do fenômeno denominado *imaging* quando do aumento da taxa de amostragem do sinal de voz;
- a segunda faz uso da aplicação de características não-lineares a sinais periódicos.

A maioria dos algoritmos de extensão em frequência utiliza o modelo fonte-filtro para a geração das componentes de frequência perdidas na transmissão do sinal de voz. Este modelo é baseado no trato vocal humano e tenta modelar dois cenários distintos que ocorrem a partir do fluxo de ar oriundo dos pulmões. Este fluxo pressiona as cordas vocais podendo ocorrer duas situações distintas:

- cordas vocais relaxadas causam um fluxo de ar turbulento, semelhante a um ruído;

- cordas vocais tencionadas e fechadas, fazendo com que a pressão do ar oriundo dos pulmões aumente até a abertura das cordas vocais. Posteriormente, a pressão decai rapidamente e as cordas vocais voltam a se fechar. O resultado deste fenômeno é um sinal periódico.

O sinal de banda larga observado antes das cordas vocais é denominado de sinal de excitação. Este sinal sofre alterações em seu espectro de frequências ao passar pelo trato vocal humano.

O modelo fonte-filtro tenta modelar os dois fenômenos descritos acima por meio da geração de dois sinal distintos:

- um gerador de ruído para reconstruir sons não-vozeados;
- um gerador de pulsos para reconstruir sons vozeados.

A Figura 5.3 apresenta um diagrama de blocos do modelo fonte-filtro do processo de geração da fala humana, representada nesta figura pelo sinal $x_{voz}(n)$. Nesta figura, a frequência do *pitch*, f_{pitch} , é dada pelo inverso da duração do pulso e a função $g(n)$ permite acessar os sinais provenientes dos geradores e pode ser discreta, $\{0,1\}$, ou contínua, $0 \leq g(n) \leq 1$. A influência das cavidades do trato vocal é modelada a partir do filtro $G/A(z)$, onde G é constante e os coeficientes de $A(z)$ são denominados coeficientes LPC (*Linear Predictive Coding*). Como o sinal de excitação é de banda larga, a função de transferência do filtro $G/A(z)$ representa o envelope espectral do sinal de voz [2].

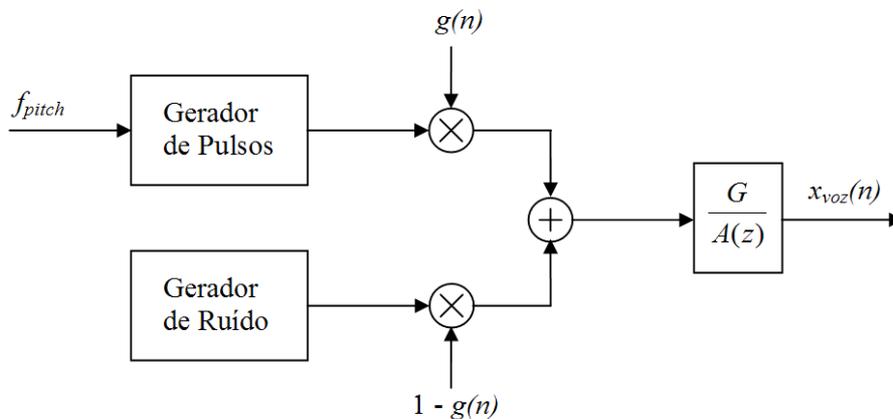


Figura 5.3: Modelo fonte-filtro do processo de geração da fala humana.

A Figura 5.4 apresenta um diagrama de blocos da estrutura básica utilizada pelos algoritmos de extensão em frequência baseados no modelo fonte-filtro descrito acima. Estes algoritmos devem gerar estimativas do sinal de excitação de banda larga ($e_{BL}(n)$) e do envelope espectral (representado pelos coeficientes $a_{BL}(n)$) do sinal de voz de banda larga. Posteriormente, o sinal de banda larga estimado é adicionado de forma complementar ao sinal original de banda estreita resultando no sinal estendido ($x_{ext}(n)$). O sinal de excitação de banda estreita ($e_{BE}(n)$) é obtido pela aplicação do sinal de voz de banda estreita ao filtro $A(z)$. Assim, o sinal de excitação de banda larga é obtido a partir do sinal de excitação de banda estreita. As principais técnicas utilizadas na geração do sinal de excitação de banda larga são:

- modular o sinal de excitação de banda estreita por uma função cosseno de frequência variável segundo o *pitch* do sinal processado [2, 14];
- realizar operações não-lineares no sinal de excitação de banda estreita como, por exemplo, elevá-lo ao quadrado ou ao cubo, para acrescentar harmônicos ao sinal [2].

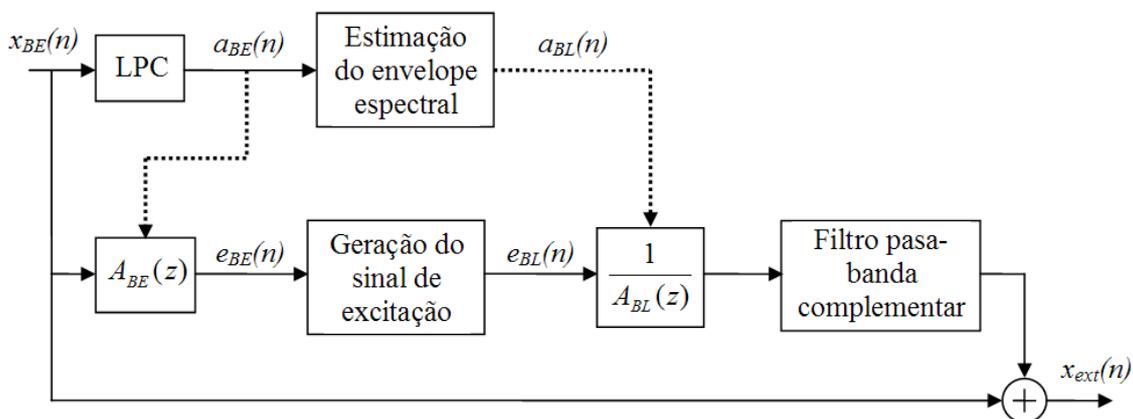


Figura 5.4: Estrutura básica dos algoritmos de extensão em frequência baseados no modelo fonte-filtro.

Após a geração do sinal de excitação de banda larga, deve-se estimar o envelope espectral do sinal de voz de banda larga. As principais técnicas utilizadas nesta estimativa são:

- mapeamento linear, que consiste em estimar os coeficientes que contém informações a respeito do envelope espectral do sinal de banda larga, $a_{BE}(n)$, a partir dos coeficientes que contém informações do envelope espectral do sinal de banda estreita, $a_{BL}(n)$, por meio de uma relação linear [2, 15];
- redes neurais, cujas entradas são os coeficientes $a_{BE}(n)$ e as saídas são os coeficientes $a_{BL}(n)$. A rede é previamente treinada por pares $a_{BE}(n)$ - $a_{BL}(n)$, e pode representar relações lineares e não-lineares entre estes coeficientes [2];
- *codebook*, onde os coeficientes $a_{BE}(n)$ são mapeados em um dos conjuntos de coeficientes de banda estreita existentes na base de dados, de acordo com uma medida de distância predefinida. Esta base de dados é formada por pares $a_{BE}(n)$ - $a_{BL}(n)$. Cada entrada no *codebook* tem sua saída correspondente [2, 16].

Capítulo 6

Método Proposto de Extensão em Frequência

6.1 Introdução

Este capítulo apresenta um novo método de extensão em frequência baseado na classificação de trechos de voz por meio de redes neurais e pela utilização de filtragem ótima. A Seção 6.2 apresenta os sinais de voz utilizados nos testes do método proposto. A Seção 6.3 mostra o procedimento de obtenção do conjunto de treinamento da rede neural utilizada na classificação dos trechos dos sinais de voz. A Seção 6.4 descreve o algoritmo de extensão em frequência desenvolvido neste trabalho. Por fim, a Seção 6.5 mostra o filtro protótipo utilizado na obtenção do banco de filtros modulados por cosseno usado para a decomposição dos sinais de voz.

6.2 Sinais de Voz Utilizados

Para a realização dos testes com o algoritmo de extensão em frequência proposto, foram utilizados neste trabalho três sinais de vozes de locutores masculinos e três sinais de vozes de locutores femininos, todas em inglês. Os sinais originais são sinais com taxa de amostragem de 16 kHz com banda de até 8 kHz. Estes sinais serão denominados de sinais de banda larga. Os sinais de banda estreita, sinais adequados para a transmissão em um sistema de telefonia fixa, foram filtrados para

ocuparem uma banda de aproximadamente 3,4 kHz. A Tabela 6.1 apresenta as frases utilizadas juntamente com a identificação de cada sinal de voz de banda larga e de banda estreita utilizados. As Figuras 6.1-6.6 mostram os espectrogramas dos sinais de banda estreita (a) e dos sinais de banda larga (b) das frases apresentadas na Tabela 6.1.

Tabela 6.1: Frases utilizadas nos testes do algoritmos proposto de extensão em frequência.

Identificação dos Sinais		Frase
Banda Larga	Banda Estreita	
x_{f1}^{BL}	x_{f1}^{BE}	<i>Our janitor sweeps the floor every night.</i>
x_{f2}^{BL}	x_{f2}^{BE}	<i>There isn't enough paint to finish the room.</i>
x_{f3}^{BL}	x_{f3}^{BE}	<i>The fruit peel was cut in six slices.</i>
x_{m1}^{BL}	x_{m1}^{BE}	<i>Use a pencil to write the first draft.</i>
x_{m2}^{BL}	x_{m2}^{BE}	<i>If your tooth hurts that much you should see a dentist.</i>
x_{m3}^{BL}	x_{m3}^{BE}	<i>Tuck the sheet under the edge of the mat.</i>

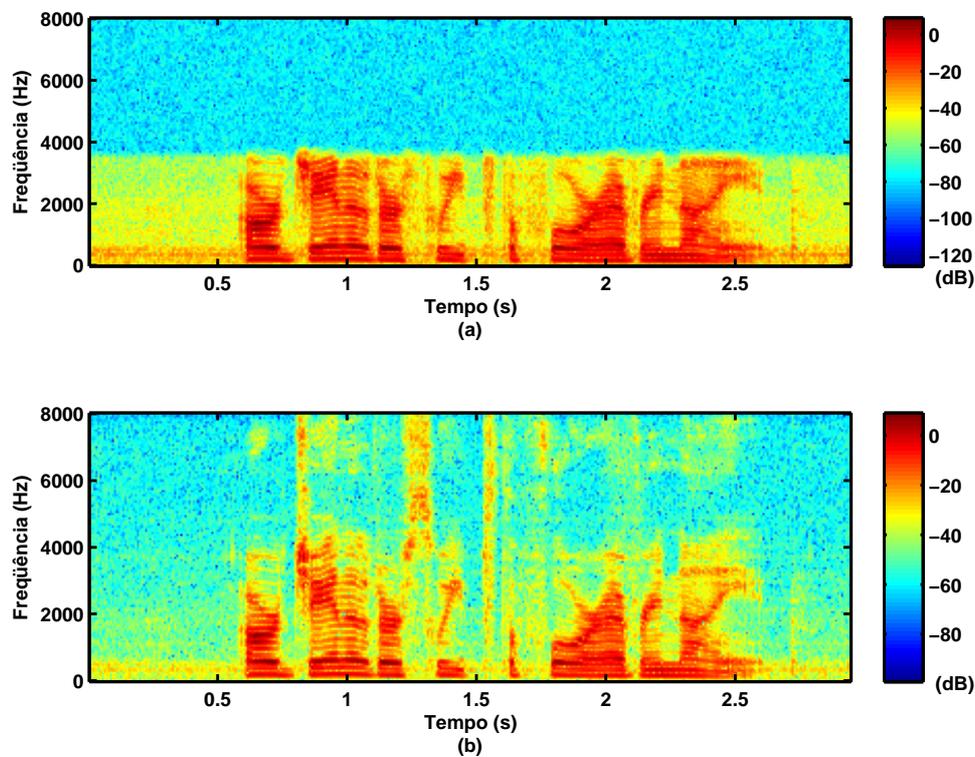


Figura 6.1: (a) Espectrograma de x_{f1}^{BE} ; (b) Espectrograma de x_{f1}^{BL} .

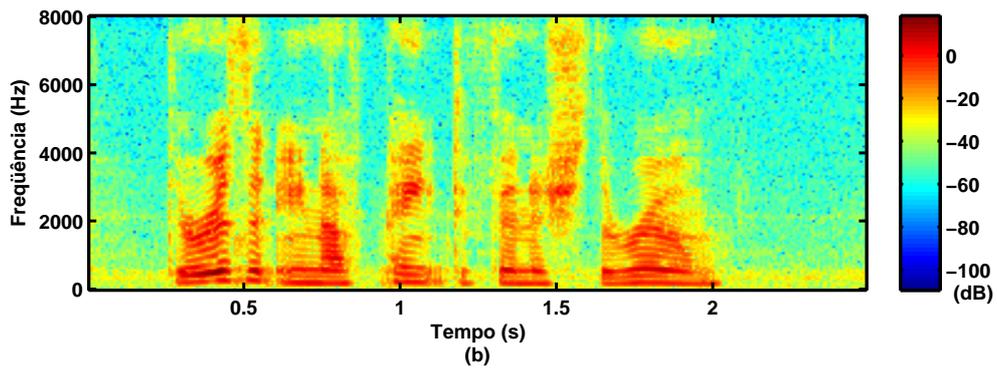
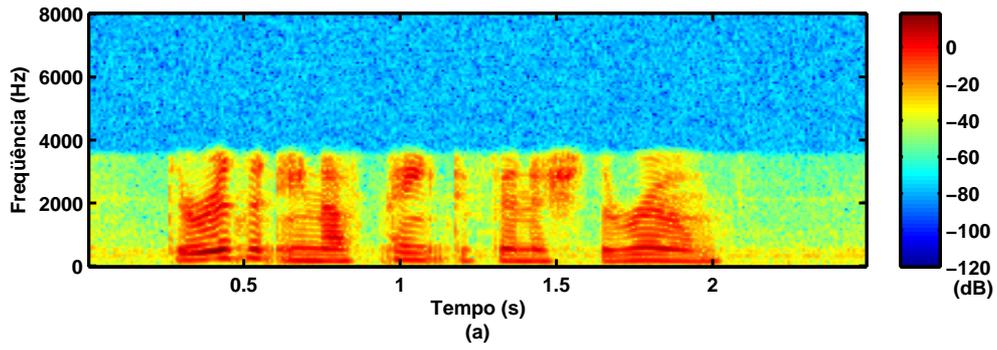


Figura 6.2: (a) Espectrograma de $x_{f_2}^{BE}$; (b) Espectrograma de $x_{f_2}^{BL}$.

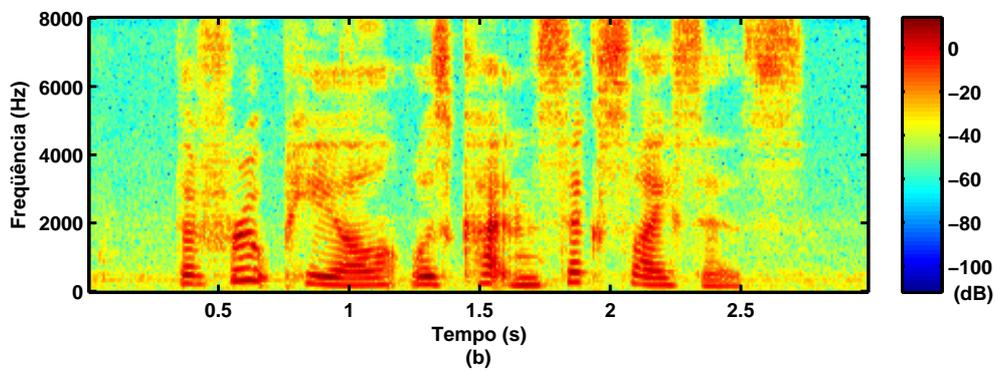
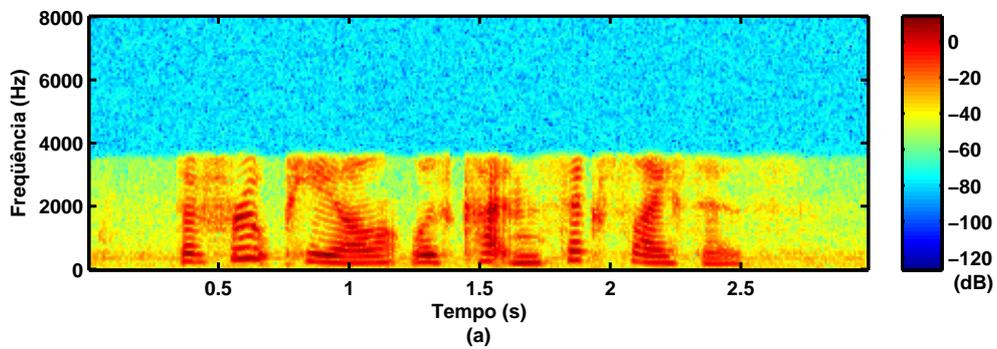


Figura 6.3: (a) Espectrograma de $x_{f_3}^{BE}$; (b) Espectrograma de $x_{f_3}^{BL}$.

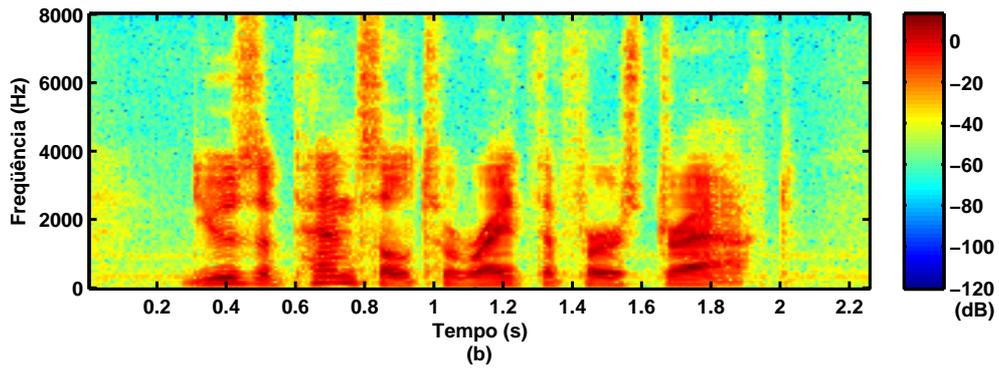
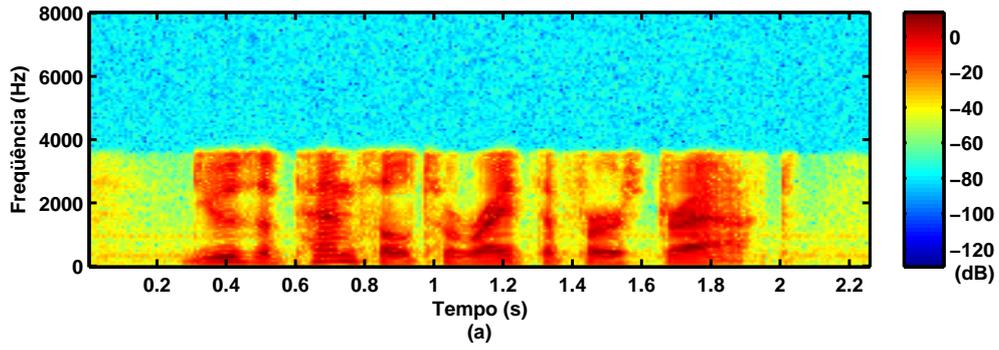


Figura 6.4: (a) Espectrograma de x_{m1}^{BE} ; (b) Espectrograma de x_{m1}^{BL} .

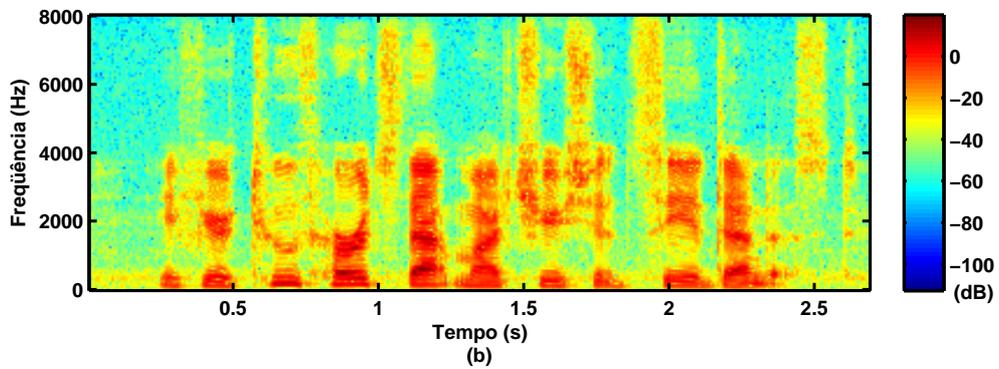
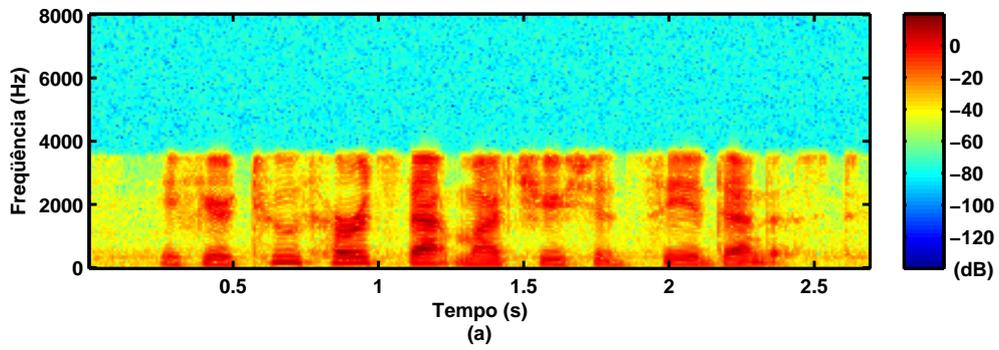


Figura 6.5: (a) Espectrograma de x_{m2}^{BE} ; (b) Espectrograma de x_{m2}^{BL} .

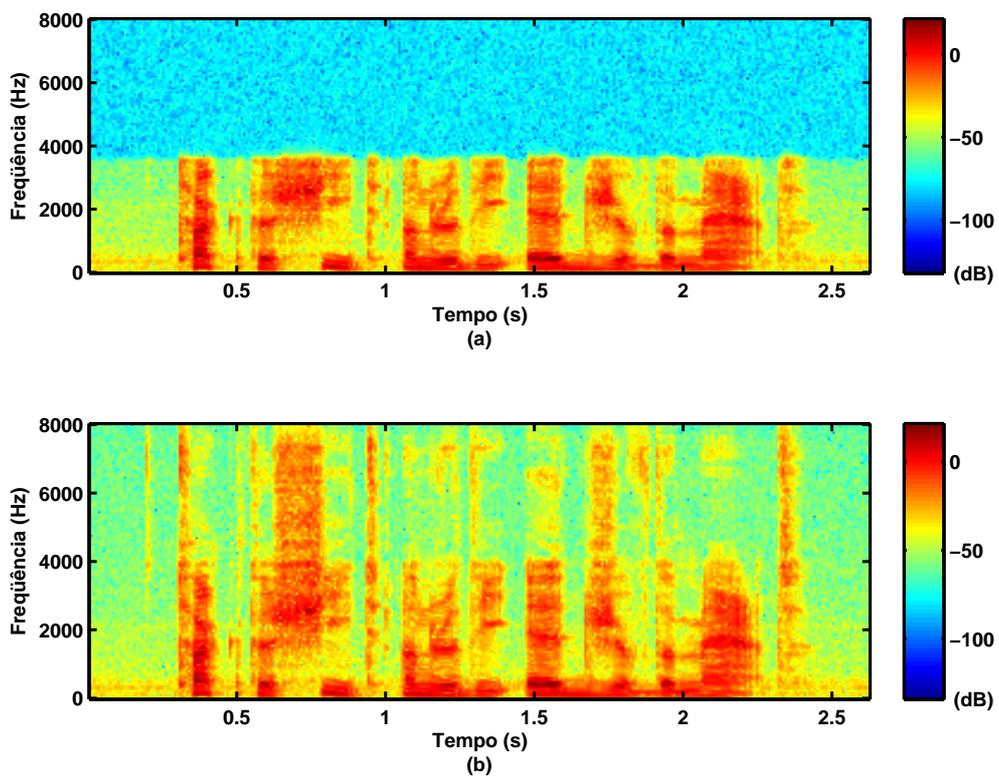


Figura 6.6: (a) Espectrograma de x_{m3}^{BE} ; (b) Espectrograma de x_{m3}^{BL} .

6.3 Treinamento das Redes Neurais

Os conjuntos de dados utilizados nos treinamentos e testes das redes neurais empregadas neste trabalho foram obtidos a partir de trechos selecionados manualmente dos sinais das vozes (coluna Banda Estreita) listados na Tabela 6.1. A idéia de se utilizar a rede neural no algoritmo de extensão de frequência é a de fazer uma classificação dos *frames* consecutivos do sinal de voz de acordo com características comuns entre os trechos selecionados durante a etapa de treinamento da rede. Assim, os trechos dos sinais de voz são divididos em duas grandes classes de sons:

- os vozeados, que apresentam grande parte de sua energia localizada nas baixas frequências, ou seja, possuem características passa-baixa [17, 4];
- os não-vozeados, que apresentam grande parte de sua energia localizada em altas frequências, ou seja, possuem características passa-alta [17, 4].

Foram extraídos parâmetros destes trechos, utilizando-se as técnicas de processamento de sinais descritas na Seção 6.4, para a formação do conjunto de treinamento da rede neural. Como exemplo destes trechos, a Figura 6.7 mostra a forma de onda da palavra *sheet* do sinal de banda larga x_{m3}^{BL} juntamente com seu espectrograma. Nesta figura, são destacadas as delimitações dos fonemas utilizados na obtenção dos trechos retirados do sinal de banda estreita x_{m3}^{BE} para o treinamento da rede neural.

6.4 Descrição do Método

Neste trabalho, foram utilizados os conceitos de redes neurais *feedforward*, de filtragem ótima no sentido dos mínimos quadrados e de bancos de filtros para a realização de um algoritmo de extensão em frequência. A Figura 6.8 mostra o esquema do algoritmo proposto. Primeiramente, um trecho (x_f) do sinal de voz de banda estreita é decomposto por um banco de filtros de 16 canais de tal forma que cada canal tenha uma banda de 500 Hz. Estes trechos são denominados *frames* e correspondem a regiões consecutivas do sinal de voz com duração de 20 ms. Posteriormente, são calculadas as variâncias das saídas dos sete primeiros canais do banco

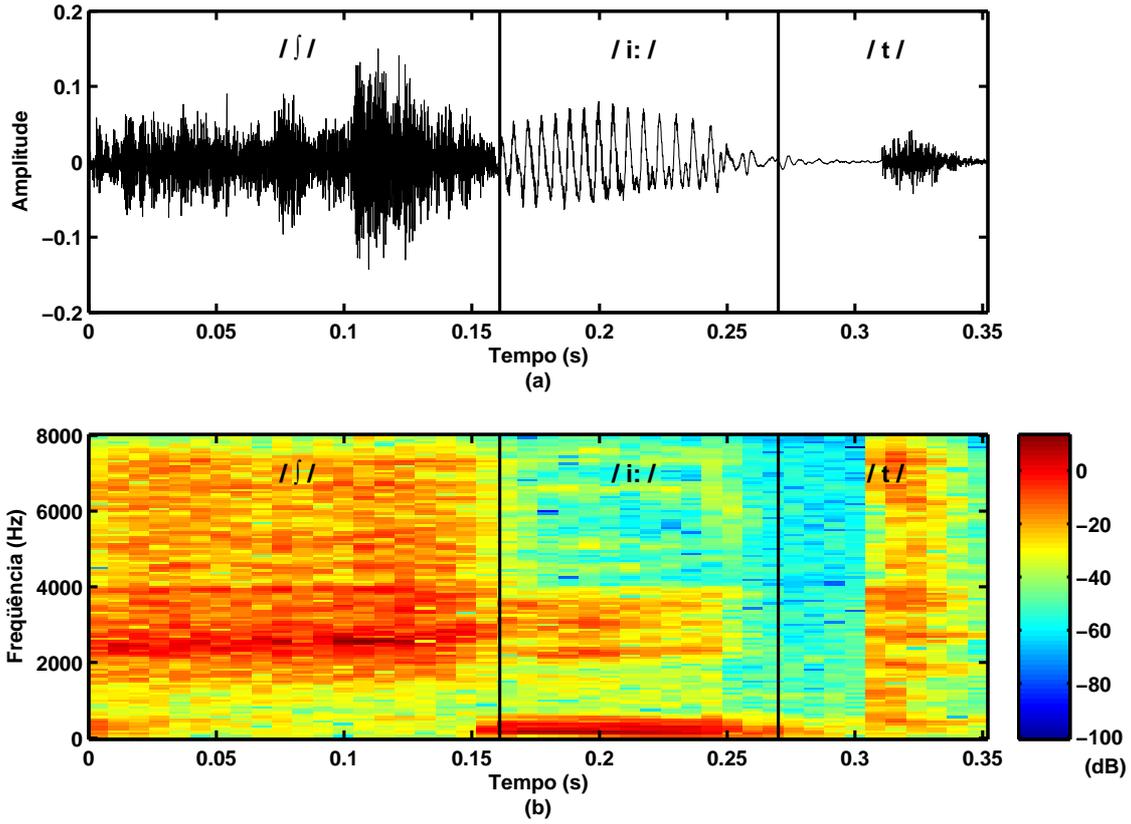


Figura 6.7: (a) Forma de onda da palavra *sheet* do sinal de banda larga x_{m3}^{BL} ; (b) espectrograma de (a).

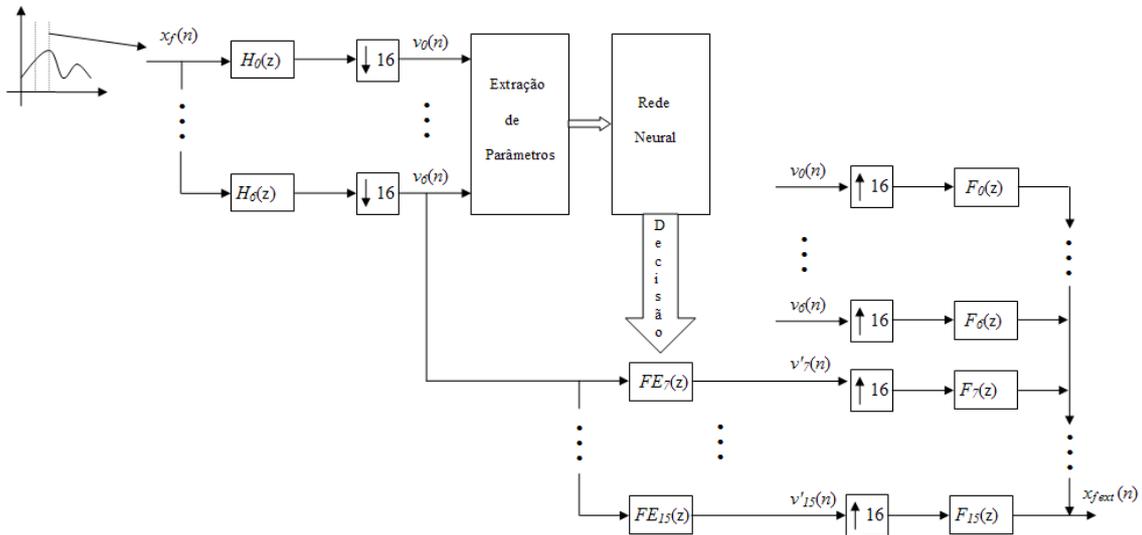


Figura 6.8: Esquema do algoritmo proposto.

de filtros de análise de acordo com a Equação 6.1:

$$\sigma^2 = \frac{1}{L-1} \sum_{n=1}^L (x(n) - \bar{x})^2, \quad (6.1)$$

onde L representa o tamanho da sequência $x(n)$ e \bar{x} é a média da sequência $x(n)$, que é dada por:

$$\bar{x} = \frac{1}{L} \sum_{n=1}^L x(n). \quad (6.2)$$

Como os sinais de banda estreita utilizados apresentam informações relevantes apenas até 3,5 kHz, aproximadamente, então utilizam-se apenas as sete primeiras bandas do sinal decomposto na próxima etapa do algoritmo. As variâncias definidas na Equação 6.1 são utilizadas como entradas, após terem sido normalizadas pela soma das variâncias dos sete canais, de uma rede neural previamente treinada com o intuito de que a rede classifique os *frames* adequadamente. Também foram utilizados como parâmetros de entrada da rede neural as seguintes características extraídas dos *frames* dos sinais de banda estreita [4, 18]:

- o valor da autocorrelação para o atraso igual a 1;
- o valor da autocorrelação para o atraso igual ao período do *pitch*;
- a taxa de cruzamento de zero, definida como a taxa de troca de sinal das amostras do *frame*, ou seja, o número de trocas de sinal dividido pelo número de amostras consideradas;
- o parâmetro *kurtosis* local, definido como:

$$x_{LK} = \log \frac{1}{N} \sum_{\kappa=0}^{N-1} x_f^4(k) - 2 \log \frac{1}{N} E(m), \quad (6.3)$$

$$E(m) = \sum_{k=0}^{N-1} x_f^2(k), \quad (6.4)$$

onde m é o índice dos *frames* e N é o tamanho de cada *frame*;

- o centróide espectral, definido pela expressão:

$$x_{SC} = \frac{\sum_{i=0}^{N/2} i \cdot |X_f(e^{j\omega_i})|}{(\frac{N}{2} + 1) \sum_{i=0}^{N/2} |X_f(e^{j\omega_i})|}, \quad (6.5)$$

sendo $X_f(e^{j\omega_i})$ o i -ésimo coeficiente da transformada discreta de Fourier de x_f .

O *pitch* é estimado como sendo o valor máximo da função de autocorrelação do sinal de voz, a partir do atraso igual a 20 desta função [18]. Estes parâmetros foram

selecionados a partir de testes realizados visando um melhor desempenho da rede neural.

Após a classificação pela rede neural, utiliza-se um conjunto de filtros adequados, de acordo com a classificação resultante da rede neural, para realizar a extensão em frequência do *frame*. Cada conjunto de filtros é utilizado para a geração das altas frequências do sinal estendido. A Figura 6.9 apresenta o espectrograma do sinal de voz de banda estreita x_{m3}^{BE} com regiões de frequências a serem processadas por cada canal (de 1 a 16) do banco de filtros. Estes intervalos correspondem à informação

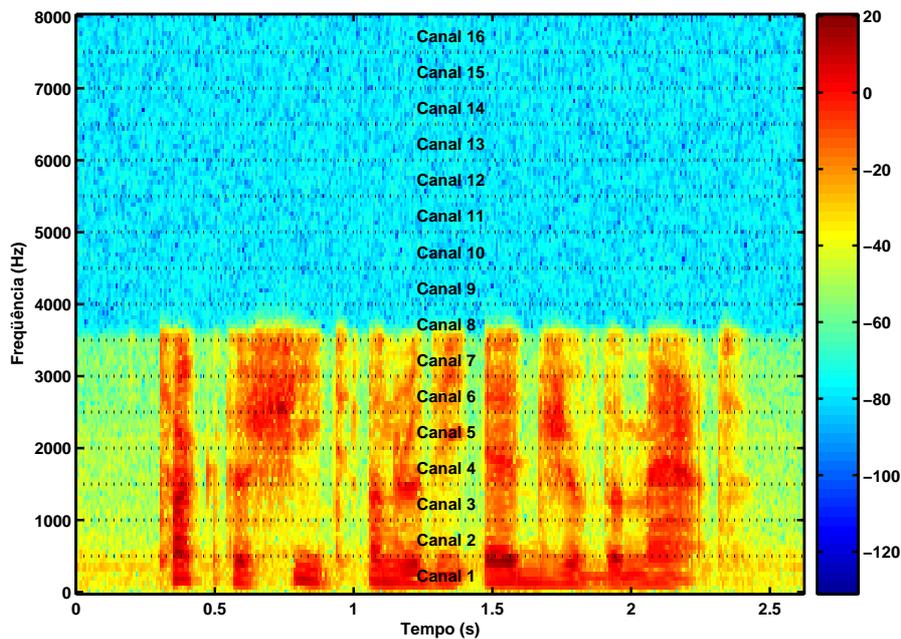


Figura 6.9: Espectrograma do sinal x_{m3}^{BE} com as regiões a serem separadas pelo banco de filtros em destaque.

que aparecerá em cada canal do banco de filtros, considerando os filtros de análise ideais. Desta forma, pode-se observar que a partir do canal 8 quase não há informação relevante no sinal de voz de banda estreita. Então, o sinal oriundo do sétimo canal do banco de análise, sinal que compreende frequências de 3,0 a 3,5 kHz, é aquele que contém mais informações de alta frequência. A partir deste sinal, após ser processado por cada filtro de extensão de frequência correspondente a cada canal, é que são gerados os sinais dos canais 8 a 16 que substituirão os sinais originais destes canais provenientes do banco de análise. Este processo de extensão de frequência é esquematizado na Figura 6.10. Após esta substituição os sinais originais do sinal de

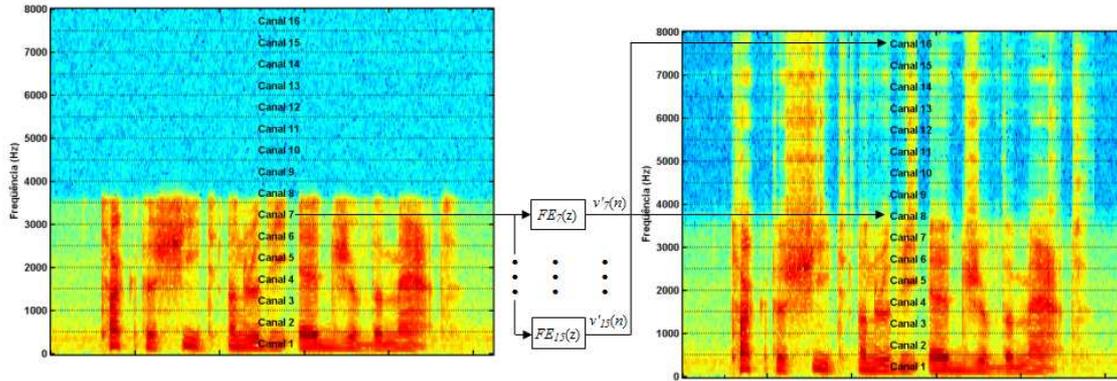


Figura 6.10: Esquema da geração dos canais 8 a 16 do sinal x_{m3}^{BE} .

voz de banda estreita dos canais de 1 a 7 e os outros sinais sintetizados dos canais de 8 a 16 serão processados pelo banco de síntese dando origem a uma versão do sinal original com as componentes de alta freqüências sintetizadas.

Os filtros utilizados na extensão em freqüência foram obtidos pelo método dos mínimos quadrados, exposto no Capítulo 3. Cada conjunto de filtros corresponde aos filtros $FE_7(z), \dots, FE_{15}(z)$. Foram utilizados quatro conjuntos de acordo com a classificação de cada *frame*, dada pela rede neural, e com o gênero do locutor. Assim, cada conjunto de filtros será aplicado a um *frame*, de acordo com a classificação abaixo:

- sons vozeados pertencentes a locutores masculinos;
- sons não-vozeados pertencentes a locutores masculinos;
- sons vozeados pertencentes a locutores femininos;
- sons não-vozeados pertencentes a locutores femininos.

Para a obtenção de cada conjunto de filtros, foram selecionados seis trechos, seguindo a classificação acima, dos sinais de banda larga definidos na Tabela 6.1, de acordo com a Equação 3.16. Cada trecho foi decomposto pelo banco de filtros de 16 canais mostrado na Figura 6.12. O canal de número 7 foi utilizado como sinal de entrada, x (Figura 3.1), e cada sinal proveniente dos canais subseqüentes (8 à 16) foi utilizado como sinal desejado, d (Figura 3.1), visando a geração dos filtros de extensão em freqüência. Ao final, foi obtida uma média dos filtros gerados a partir dos seis trechos.

A etapa de extensão em frequência também foi avaliada considerando-se apenas ganhos no lugar de filtros ótimos. De forma análoga à utilização de filtragem ótima, os quatro conjuntos de seis trechos dos sinais de banda larga, selecionados para a obtenção dos filtros, também foram utilizados para a obtenção dos ganhos. Considerando um dos quatro conjuntos, cada trecho foi decomposto pelo mesmo banco de filtros utilizado na obtenção dos filtros ótimos. O ganho de cada canal (8 à 16) foi obtido pela média dos ganhos calculados nos seis trechos de cada conjunto. Cada ganho foi calculado pela divisão do desvio padrão de cada canal pelo desvio padrão do sétimo canal.

6.5 Banco de Filtros

Para a decomposição dos sinais envolvidos no algoritmo de extensão de frequência foi utilizado um banco de filtros modulado por cosseno. Este banco foi obtido a partir de um filtro protótipo ($h_p(n)$) com 96 coeficientes cujos 48 primeiros coeficientes são mostrados na Tabela 6.2 [19], de cima para baixo, da esquerda para a direita. Esta tabela mostra apenas os primeiros 46 coeficientes do filtro protótipo, já que o mesmo é um filtro de fase linear do tipo 2 [20], ou seja, atende à seguinte equação:

$$h_p(n) = h_p(95 - n), \quad 0 \leq n \leq 95. \quad (6.6)$$

A Figura 6.11 apresenta o módulo da resposta em frequência do filtro protótipo da Tabela 6.2. A Figura 6.12 apresenta os módulos das respostas em frequência do banco de filtros de análise obtidos a partir do filtro protótipo por meio da Equação 4.8.

Tabela 6.2: Primeiros 48 coeficientes do filtro protótipo de tamanho 96.

$1.0261438E - 006$	$1.4885292E - 005$	$4.3102876E - 005$
$8.2492747E - 005$	$8.9167054E - 005$	$1.1967747E - 005$
$-2.0883507E - 004$	$-5.7484056E - 004$	$-1.0486970E - 003$
$-1.5400399E - 003$	$-1.9801429E - 003$	$-2.3227495E - 003$
$-2.5761991E - 003$	$-2.7568799E - 003$	$-2.8822179E - 003$
$-2.9559677E - 003$	$-2.9583428E - 003$	$-2.8633102E - 003$
$-2.6109817E - 003$	$-2.1765068E - 003$	$-1.5249488E - 003$
$-7.2596283E - 004$	$1.8599629E - 004$	$1.1138847E - 003$
$2.1088880E - 003$	$3.1741266E - 003$	$4.4456413E - 003$
$5.8977043E - 003$	$7.5946070E - 003$	$9.4135817E - 003$
$1.1390701E - 002$	$1.3406733E - 002$	$1.5532140E - 002$
$1.7659200E - 002$	$1.9858866E - 002$	$2.2009998E - 002$
$2.4180229E - 002$	$2.6249257E - 002$	$2.8279622E - 002$
$3.0158205E - 002$	$3.1940047E - 002$	$3.3525497E - 002$
$3.4959901E - 002$	$3.6162817E - 002$	$3.7168051E - 002$
$3.7919123E - 002$	$3.8436901E - 002$	$3.8692263E - 002$

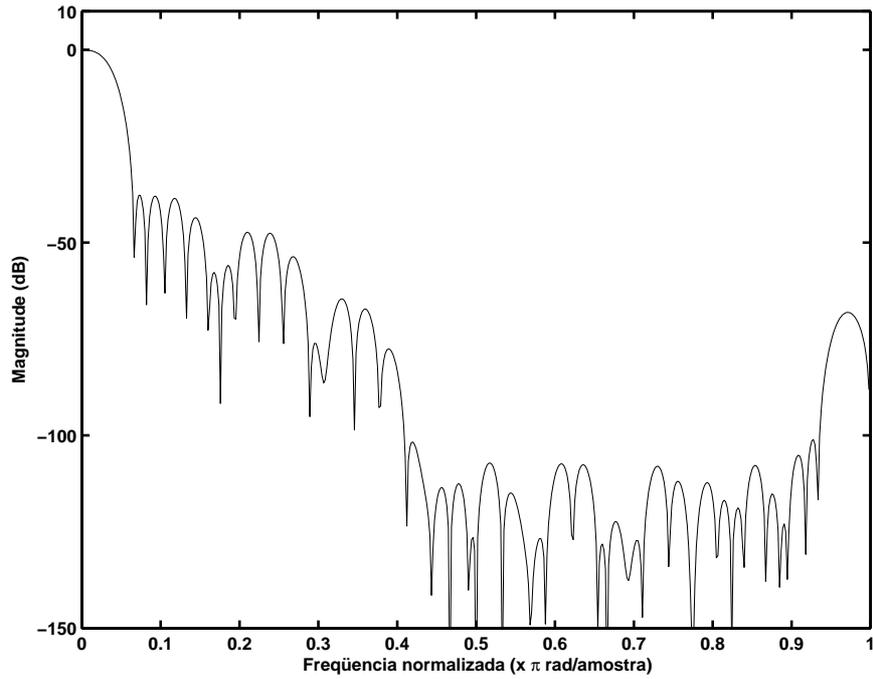


Figura 6.11: Módulo da resposta em frequência do filtro protótipo de 96 coeficientes apresentados na Tabela 6.2.

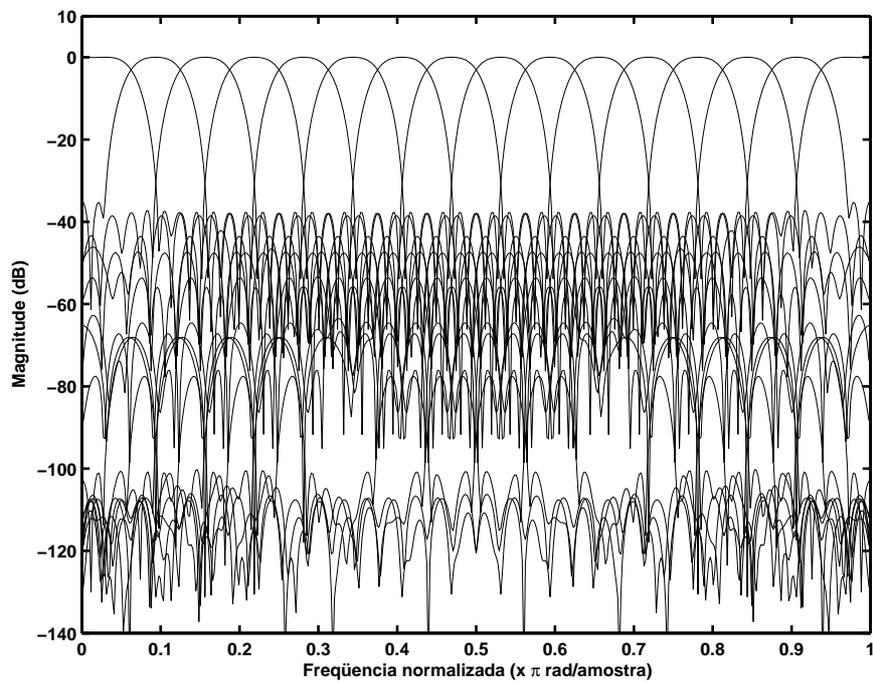


Figura 6.12: Módulo da resposta em frequência do banco de filtros modulados por cosseno a partir do filtro protótipo da Tabela 6.2.

Capítulo 7

Resultados

7.1 Introdução

Neste capítulo são apresentados os resultados obtidos com o método proposto de extensão em frequência, assim como a análise destes resultados. A Seção 7.2 apresenta os filtros de extensão em frequência obtidos. A Seção 7.3 mostra os resultados do algoritmo proposto de extensão em frequência. Por fim, a Seção 7.4 apresenta uma discussão dos resultados obtidos pelo método proposto e por um método baseado em predição linear.

7.2 Filtros de Extensão em Frequência

Os resultados do algoritmo de extensão em frequência proposto neste trabalho foram obtidos utilizando-se filtros ótimos com 8 coeficientes. A Tabela 7.1 apresenta os coeficientes dos filtros ótimos utilizados (h_{fv} - para cada canal) na extensão em frequência dos *frames* dos sinais dos locutores femininos classificados como vozeados pela rede neural. Os módulos das respostas em frequência destes filtros ótimos são mostrados na Figura 7.1.

Tabela 7.1: Filtros ótimos utilizados na extensão em frequência dos canais 8-16 dos *frames* dos sinais de locutores femininos classificados como vozeados.

Canais	8	9	10	11	12	13	14	15	16
$h_{fv}(0)$	-2,69E-2	-6,53E-4	6,36E-3	2,67E-2	-1,34E-2	-1,63E-3	3,43E-3	-1,28E-2	1,83E-3
$h_{fv}(1)$	-7,24E-2	-4,50E-2	1,93E-2	3,01E-2	-7,45E-3	-4,10E-2	1,53E-2	1,02E-2	-2,23E-2
$h_{fv}(2)$	-7,69E-2	5,80E-2	-3,06E-2	9,16E-3	1,65E-2	1,43E-3	-1,58E-2	2,41E-2	2,85E-2
$h_{fv}(3)$	-1,79E-2	8,03E-2	-2,48E-2	-3,46E-2	8,16E-3	7,99E-3	-9,73E-3	-6,23E-3	-9,20E-3
$h_{fv}(4)$	1,00E-1	-1,72E-2	4,40E-2	1,63E-2	-6,49E-3	5,50E-3	1,04E-2	1,38E-2	4,87E-3
$h_{fv}(5)$	1,47E-3	-3,04E-2	1,59E-2	2,88E-2	-9,07E-3	3,09E-2	1,99E-2	1,05E-2	-7,70E-3
$h_{fv}(6)$	-9,61E-3	2,66E-2	-4,74E-3	2,66E-2	-1,69E-2	-8,50E-3	-2,00E-2	-2,04E-2	-1,29E-2
$h_{fv}(7)$	-9,60E-2	4,92E-3	-2,84E-2	1,37E-2	5,18E-3	-2,29E-2	-7,02E-3	2,02E-2	2,42E-2

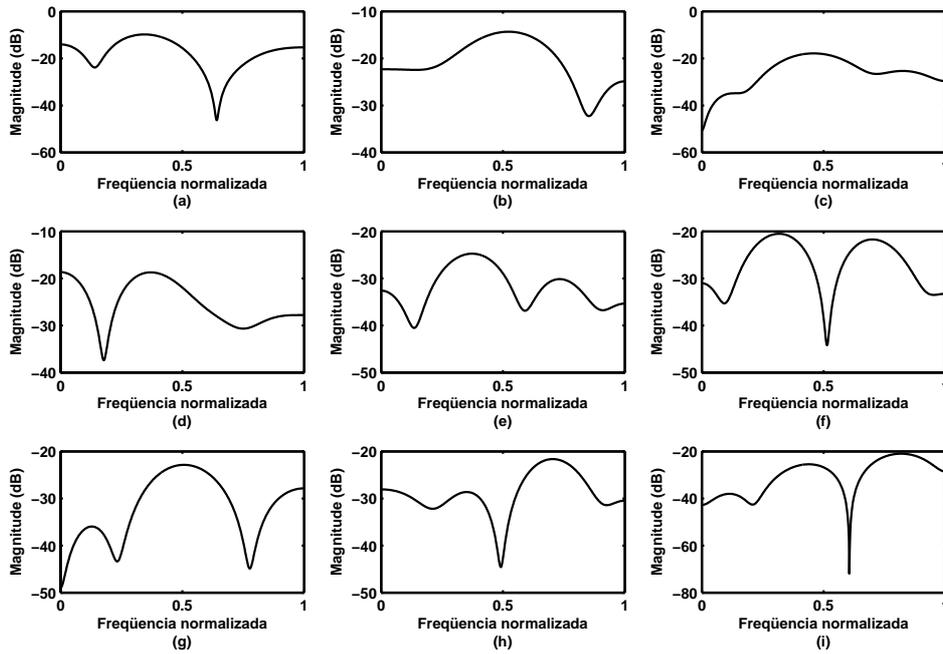


Figura 7.1: Módulo das respostas em frequência dos filtros ótimos utilizados na extensão dos *frames* dos sinais de locutores femininos classificados como vozeados. Canais: (a) 8, (b) 9, (c) 10, (d) 11, (e) 12, (f) 13, (g) 14, (h) 15 e (i) 16.

A Tabela 7.2 apresenta os coeficientes dos filtros ótimos utilizados (h_{ff} - para cada canal) na extensão em frequência dos *frames* dos sinais dos locutores femininos classificados como não-vozeados pela rede neural. Os módulos das respostas em frequência destes filtros ótimos são mostrados na Figura 7.2.

Tabela 7.2: Filtros ótimos utilizados na extensão em frequência dos canais 8-16 dos *frames* dos sinais de locutores femininos classificados como não-vozeados.

Canais	8	9	10	11	12	13	14	15	16
$h_{ff}(0)$	-6,59E-2	1,65E-2	9,71E-3	-2,08E-2	7,18E-2	-5,02E-3	9,53E-3	-9,33E-2	-6,18E-2
$h_{ff}(1)$	2,27E-1	7,04E-2	3,48E-2	-8,85E-3	-5,15E-2	-2,14E-2	-5,38E-2	9,37E-2	-5,00E-2
$h_{ff}(2)$	-2,60E-2	-3,27E-2	-1,07E-1	7,96E-2	2,37E-2	1,98E-2	4,60E-2	-8,75E-2	3,06E-2
$h_{ff}(3)$	-1,21E-1	-1,57E-1	-4,80E-3	-8,68E-3	-6,47E-2	3,73E-3	-8,97E-2	-8,70E-2	-5,53E-2
$h_{ff}(4)$	-2,61E-2	-3,64E-2	-3,09E-2	5,74E-3	4,70E-2	3,63E-2	6,48E-2	-7,23E-2	-9,54E-3
$h_{ff}(5)$	-1,15E-2	-4,37E-2	6,05E-2	-1,52E-3	1,96E-2	3,55E-2	-5,93E-2	-3,48E-2	1,05E-1
$h_{ff}(6)$	-8,24E-2	1,24E-1	3,49E-2	7,28E-2	-3,16E-2	-2,03E-2	6,99E-2	-1,36E-1	-1,02E-1
$h_{ff}(7)$	1,40E-1	-9,98E-2	7,21E-2	1,48E-2	-2,52E-2	-3,62E-3	1,20E-1	6,00E-3	-3,06E-2

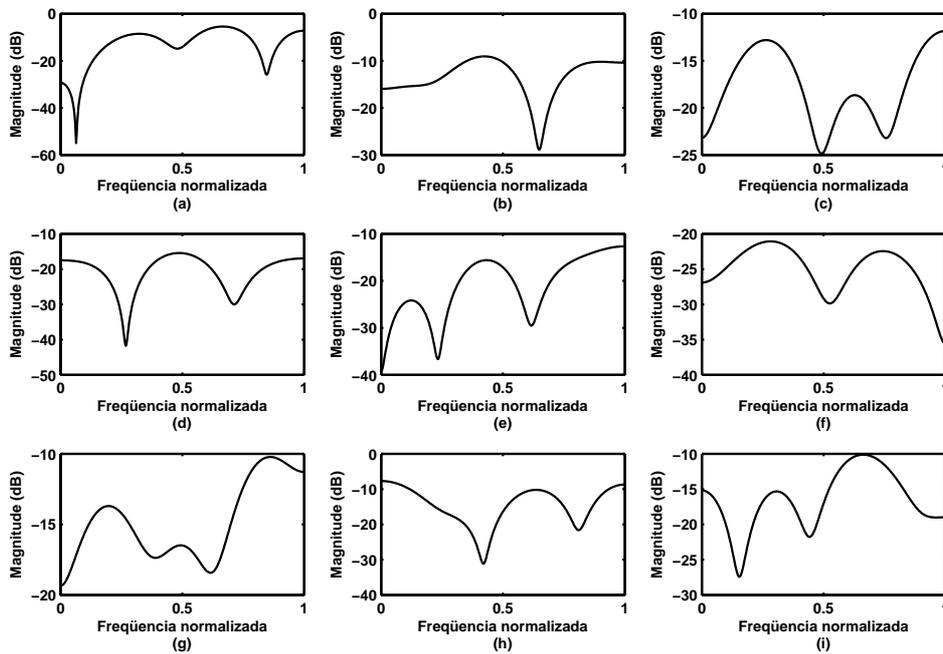


Figura 7.2: Módulo das respostas em frequência dos filtros ótimos utilizados na extensão dos *frames* dos sinais de locutores femininos classificados como não-vozeados. Canais: (a) 8, (b) 9, (c) 10, (d) 11, (e) 12, (f) 13, (g) 14, (h) 15 e (i) 16.

A Tabela 7.3 apresenta os coeficientes dos filtros ótimos utilizados (h_{mv} - para cada canal) na extensão em frequência dos *frames* dos sinais dos locutores masculinos classificados como vozeados pela rede neural. Os módulos das respostas em frequência destes filtros ótimos são mostrados na Figura 7.3.

Tabela 7.3: Filtros ótimos utilizados na extensão em frequência dos canais 8-16 dos *frames* dos sinais de locutores masculinos classificados como vozeados.

Canais	8	9	10	11	12	13	14	15	16
$h_{mv}(0)$	1,32E-2	8,97E-3	-3,11E-3	3,14E-3	2,18E-3	-1,37E-2	-7,03E-4	-5,16E-4	8,97E-4
$h_{mv}(1)$	-5,29E-2	1,13E-2	1,86E-3	-2,67E-3	-6,37E-3	-7,93E-3	-7,12E-4	6,85E-4	-1,33E-3
$h_{mv}(2)$	4,41E-2	1,23E-2	1,78E-3	-4,79E-3	-3,50E-3	4,72E-3	-1,12E-2	-9,96E-4	-4,20E-3
$h_{mv}(3)$	2,31E-2	-8,65E-005	3,39E-3	1,56E-3	-4,50E-3	1,21E-3	-2,77E-2	-1,27E-2	-8,68E-4
$h_{mv}(4)$	-1,19E-2	6,05E-3	-1,76E-3	6,77E-005	-5,87E-4	-2,54E-3	-1,54E-2	-1,09E-2	7,54E-4
$h_{mv}(5)$	-4,29E-2	1,14E-2	-4,01E-3	-4,06E-3	4,61E-3	1,25E-2	-6,75E-3	-1,14E-2	1,13E-2
$h_{mv}(6)$	1,76E-2	5,14E-3	1,38E-3	-3,13E-3	6,31E-3	4,24E-3	6,10E-3	-9,77E-3	-2,80E-3
$h_{mv}(7)$	2,97E-2	5,32E-3	-2,01E-3	1,38E-3	5,97E-3	-1,20E-2	-2,04E-4	-7,68E-3	2,33E-3

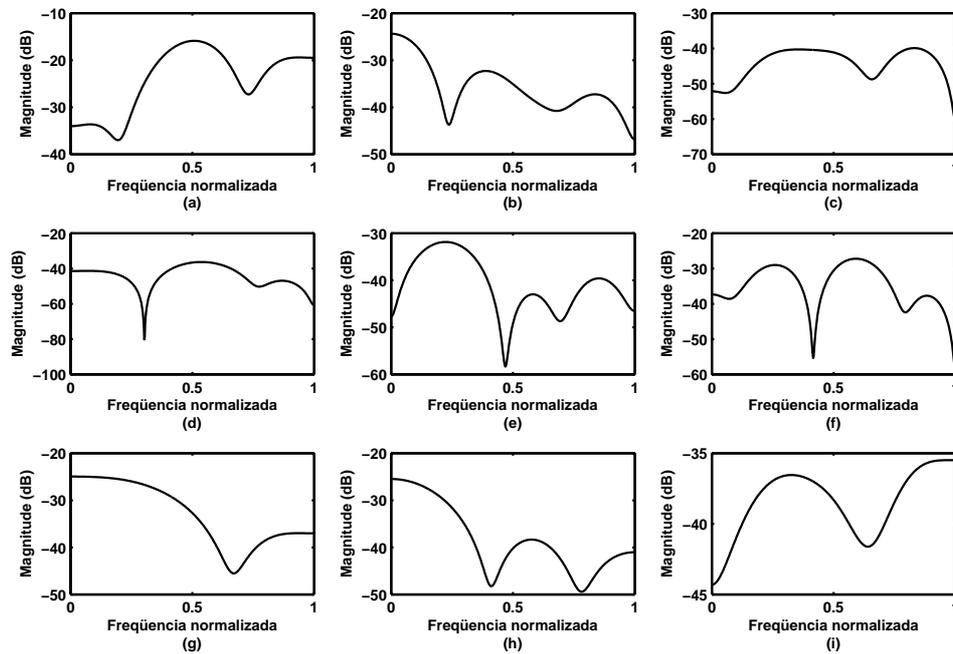


Figura 7.3: Módulo das respostas em frequência dos filtros ótimos utilizados na extensão dos *frames* dos sinais de locutores masculinos classificados como vozeados. Canais: (a) 8, (b) 9, (c) 10, (d) 11, (e) 12, (f) 13, (g) 14, (h) 15 e (i) 16.

A Tabela 7.4 apresenta os coeficientes dos filtros ótimos utilizados (h_{mf} - para cada canal) na extensão em frequência dos *frames* dos sinais dos locutores masculinos classificados como não-vozeados pela rede neural. Os módulos das respostas em frequência destes filtros ótimos são mostrados na Figura 7.4.

Tabela 7.4: Filtros ótimos utilizados na extensão em frequência dos canais 8-16 dos *frames* dos sinais de locutores masculinos classificados como não-vozeados.

Canais	8	9	10	11	12	13	14	15	16
$h_{mf}(0)$	-2,10E-1	-5,26E-2	-1,12E-2	8,42E-2	-8,61E-2	7,67E-2	8,54E-2	2,59E-2	-6,57E-2
$h_{mf}(1)$	-6,36E-2	3,70E-2	-1,62E-2	1,94E-1	-9,97E-2	1,10E-1	9,46E-2	-2,42E-2	-3,79E-3
$h_{mf}(2)$	1,93E-2	3,59E-2	1,08E-1	8,78E-2	1,79E-2	8,80E-2	8,52E-2	-4,88E-2	2,45E-2
$h_{mf}(3)$	3,18E-1	1,10E-1	-7,67E-2	9,19E-2	5,56E-2	5,34E-2	-8,89E-3	-2,08E-2	1,26E-2
$h_{mf}(4)$	1,02E-1	7,33E-3	-5,46E-2	5,76E-2	2,08E-1	-1,44E-1	3,89E-2	6,22E-3	-1,50E-2
$h_{mf}(5)$	-9,40E-2	8,37E-2	-1,27E-1	3,35E-2	5,55E-2	-7,38E-2	1,77E-2	1,48E-2	-4,37E-2
$h_{mf}(6)$	-8,56E-2	2,44E-3	9,28E-2	-5,09E-2	2,10E-2	6,58E-2	1,26E-1	3,30E-2	4,45E-2
$h_{mf}(7)$	-1,83E-1	-2,27E-1	-7,17E-2	-1,49E-1	-3,45E-3	1,11E-1	1,04E-2	-9,76E-3	-4,22E-2

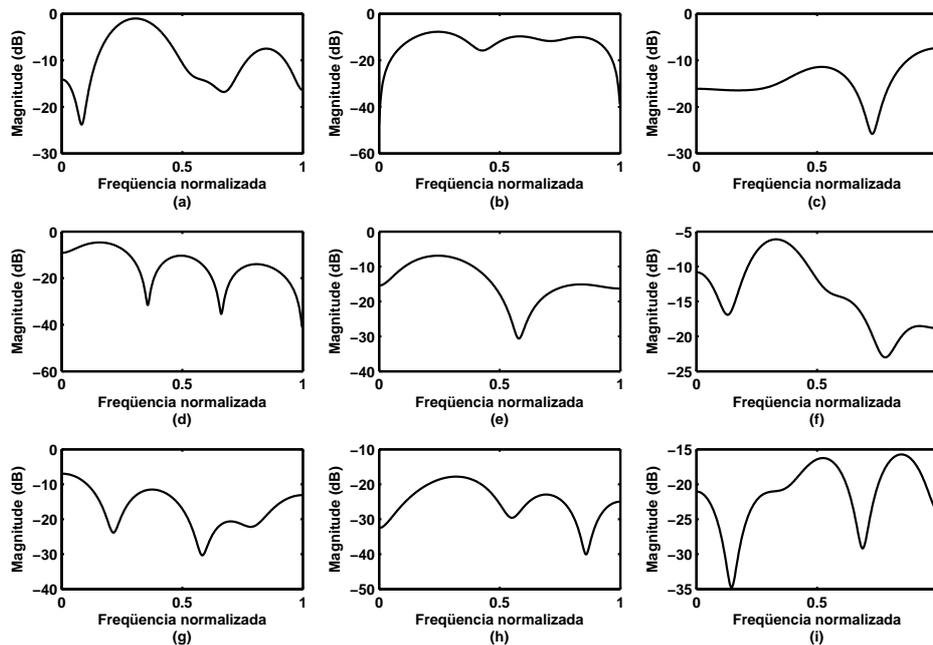


Figura 7.4: Módulo das respostas em frequência dos filtros ótimos utilizados na extensão dos *frames* dos sinais de locutores masculinos classificados como não-vozeados. Canais: (a) 8, (b) 9, (c) 10, (d) 11, (e) 12, (f) 13, (g) 14, (h) 15 e (i) 16.

A Tabela 7.5 apresenta os ganhos utilizados no lugar dos filtros ótimos, como opção para a realização da extensão em frequência propriamente dita. Considerando esta tabela da esquerda para a direita, a primeira coluna foi utilizada para os trechos das vozes femininas classificados como sons vozeados pela rede neural. A segunda coluna foi utilizada para os trechos das vozes femininas classificados como sons não-vozeados. A terceira coluna foi utilizada para os trechos das vozes masculinas classificados como sons vozeados e a quarta coluna foi utilizada para os trechos das vozes masculinas classificados como sons não-vozeados.

Tabela 7.5: Tabela dos ganhos utilizados na extensão em frequência para locutores masculinos e femininos e para sons vozeados e não-vozeados.

	Feminina Vozeados	Feminina Não-Vozeados	Masculino Vozeados	Masculino Não-Vozeados
Canal 8	8,8364568E-001	1,2543078E+000	9,6204792E-001	1,4021594E+000
Canal 9	4,5158929E-001	1,0805934E+000	1,4493045E-001	8,8070024E-001
Canal 10	2,2734197E-001	7,8482846E-001	5,3217999E-002	7,0477591E-001
Canal 11	1,9250554E-001	1,0459498E+000	5,2116984E-002	6,3273288E-001
Canal 12	1,3336593E-001	9,3912670E-001	5,5001248E-002	7,3664723E-001
Canal 13	2,2069788E-001	7,5014362E-001	8,9654445E-002	9,6239078E-001
Canal 14	1,7501288E-001	1,1815214E+000	9,4833402E-002	8,7296219E-001
Canal 15	1,9280490E-001	1,4953606E+000	6,7899327E-002	8,5315208E-001
Canal 16	1,6970854E-001	9,8430083E-001	4,1233186E-002	4,3412773E-001

7.3 Resultados do Algoritmo Proposto de Extensão em Frequência

Para a obtenção dos resultados apresentados a seguir, de acordo com o procedimento apresentado no Capítulo 6, foi treinada uma rede neural com uma base de dados de 40 pares entrada-saída, sendo que 20 pares foram obtidos de sons vozeados e 20 pares de sons não-vozeados. Cada par é composto por 12 entradas e 2 saídas. Sete entradas são provenientes das variâncias obtidas dos sete primeiros canais

da decomposição do som por um banco de filtros modulado por um cosseno de 16 canais com filtros de 96 coeficientes. As outras cinco entradas são as características extraídas dos trechos de voz, definidas na Seção 6.4, e estão apresentadas abaixo:

- o valor da autocorrelação para o atraso igual a 1;
- o valor da autocorrelação para o atraso igual ao período do *pitch*;
- a taxa de cruzamento de zero;
- o parâmetro *kurtosis* local;
- o centróide espectral.

Os trechos dos sinais de voz utilizados para a extração dos parâmetros de entrada, que foram utilizados no treinamento da rede neural, foram selecionados manualmente dos sinais de banda estreita apresentados na Tabela 6.1. Desta forma, os trechos selecionados manualmente possuem tamanhos diferentes. As entradas da rede foram calculadas a partir de cada um destes trechos considerando-se toda a duração de cada trecho.

Visando um melhor desempenho no treinamento da rede neural, foram testados alguns tipos de normalização nas entradas da rede. Uma delas foi a descrita em [11], que propõe que a normalização seja realizada considerando cada entrada. Isto é feito, primeiramente, pela subtração da entrada pela média e, posteriormente, pela divisão por duas vezes o desvio padrão. Porém, o tipo de normalização utilizada neste trabalho foi a divisão das sete primeiras entradas pela soma delas. Isto porque esta normalização foi a que obteve melhores resultados no aprendizado da rede neural. As duas saídas da rede foram definidas pelo conjunto $\{1,-1\}$ para a classe dos sons vozeados e pelo conjunto $\{-1,1\}$ para a classe dos sons não-vozeados.

A rede neural utilizada é do tipo *feedforward* com a atualização das sinapses realizada pelo algoritmo de batelada (descrito pelo Algoritmo 2.2, apresentado no Capítulo 2), com 1 neurônio não-linear na camada intermediária e dois neurônios não-lineares na camada de saída. A função de ativação dos neurônios da camada intermediária e de saída é do tipo tangente hiperbólica, representada pela Equação 2.3 e mostrada na Figura 2.3. Foram utilizados neurônios não-lineares na camada

de saída para melhorar o desempenho da classificação realizada pela rede. A inicialização foi de 0,15 para todas as sinapses e o passo de treinamento utilizado foi $\mu = 0,1$. Estes parâmetros de configuração da rede neural foram os que resultaram em um melhor desempenho após diversos testes realizados. A Figura 7.5 mostra as curvas de erro médio quadrático de treinamento e de teste obtidas considerando os parâmetros descritos acima, onde são destacadas as últimas iterações para uma melhor visualização dos erros de treinamento e de teste da rede. Para evitar a ocorrência do fenômeno de *overtraining*, buscou-se utilizar o conjunto de sinapses da iteração que apresentou o menor erro de teste da rede [11]. Pela análise do gráfico da Figura 7.5, verifica-se que o menor erro de teste ocorre na última iteração do algoritmo de treinamento da rede, ou seja, o erro de teste continua diminuindo. Assim, as sinapses desta iteração foram as utilizadas na obtenção dos resultados do algoritmo de extensão de frequência apresentados a seguir. Também foram realizados testes com um maior número de iterações; porém, não houve melhora significativa no desempenho da rede neural. O erro médio quadrático (EMQ), citado na Figura 7.5, é

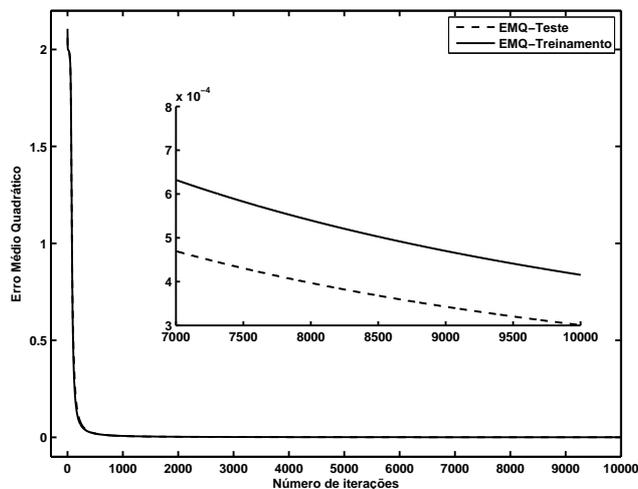


Figura 7.5: Curvas de erro médio quadrático de treinamento e de teste da rede neural.

a média realizada entre os resultados da Equação 2.29 para cada par, dentro de uma mesma iteração. A diferença entre a etapa de treinamento e a etapa de teste é que na primeira há a atualização das sinapses enquanto que na segunda são utilizadas as sinapses calculadas na etapa de treinamento para cada interação, ou seja, não há atualização das sinapses.

Os pares de teste foram utilizados para a validação da rede. Para o conjunto ótimo de sinapses obtido, a rede classificou corretamente todos os pares de teste.

Os filtros de extensão em frequência utilizados foram os filtros com 8 coeficientes apresentados anteriormente na Seção 7.2. Para a decomposição dos trechos de voz utilizados na geração destes filtros de extensão, também foi utilizado o banco de filtros modulado por cosseno, proveniente do filtro protótipo mostrado na Tabela 6.2.

A partir da obtenção da rede neural e dos filtros ótimos, aplicou-se o algoritmo de extensão de frequência, definido na Seção 6.4, nos sinais de voz de banda estreita relacionados na Tabela 6.1. As Figuras 7.6(a)-7.11(a) mostram os espectrogramas dos sinais estendidos a partir dos sinais de banda estreita x_{f1}^{BE} , x_{f2}^{BE} , x_{f3}^{BE} , x_{m1}^{BE} , x_{m2}^{BE} e x_{m3}^{BE} , respectivamente. É possível observar a extensão realizada nas frequências mais altas, entre 3,7 kHz e 8 kHz, quando estes espectrogramas são comparados com aqueles apresentados nas Figuras 6.1(b)-6.6(b), respectivamente. Estas figuras também apresentam os espectrogramas dos sinais de banda larga das mesmas frases juntamente com um gráfico com dois valores distintos. As regiões do gráfico com os valores maiores indicam trechos classificados pela rede neural como trechos não-vozeados. Já as regiões do gráfico com os valores menores indicam trechos classificados pela rede neural como trechos vozeados. Estes gráficos, em (b), permitem avaliar a classificação realizada pela rede neural nos sinais de banda estreita de acordo com as regiões correspondentes no espectrograma dos sinais de banda larga.

Pela análise das Figuras 7.6-7.11, pode-se verificar que, de maneira geral, a rede neural classificou os sons contendo frequências mais altas como não-vozeados. Analisando-se a Figura 7.7, verifica-se que, por volta de 0,5 s (som não-vozeado /s/ da palavra *isn't*) houve um erro na classificação da rede neural, ou seja, a rede classificou como vozeado um trecho não-vozeado. Pela análise das figuras, também pode-se observar que entre as palavras de uma frase, ou seja, nas regiões de silêncio, não há um padrão na classificação da rede neural. Outro fato a ser observado é que a rede classificou, de maneira geral, sons plosivos como sons não-vozeados. Isto porque os sons plosivos apresentaram componentes de altas frequências similares a dos sons não-vozeados.

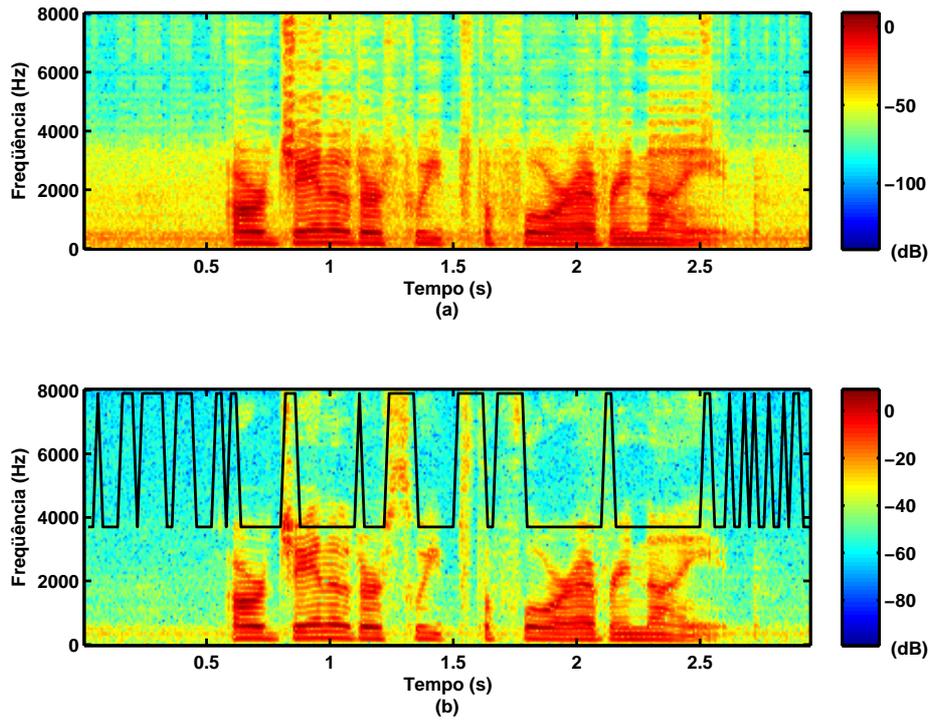


Figura 7.6: (a) Espectrograma do sinal x_{f1}^{BE} estendido, (b) espectrograma de x_{f1}^{BL} destacando-se a classificação realizada pela rede neural.

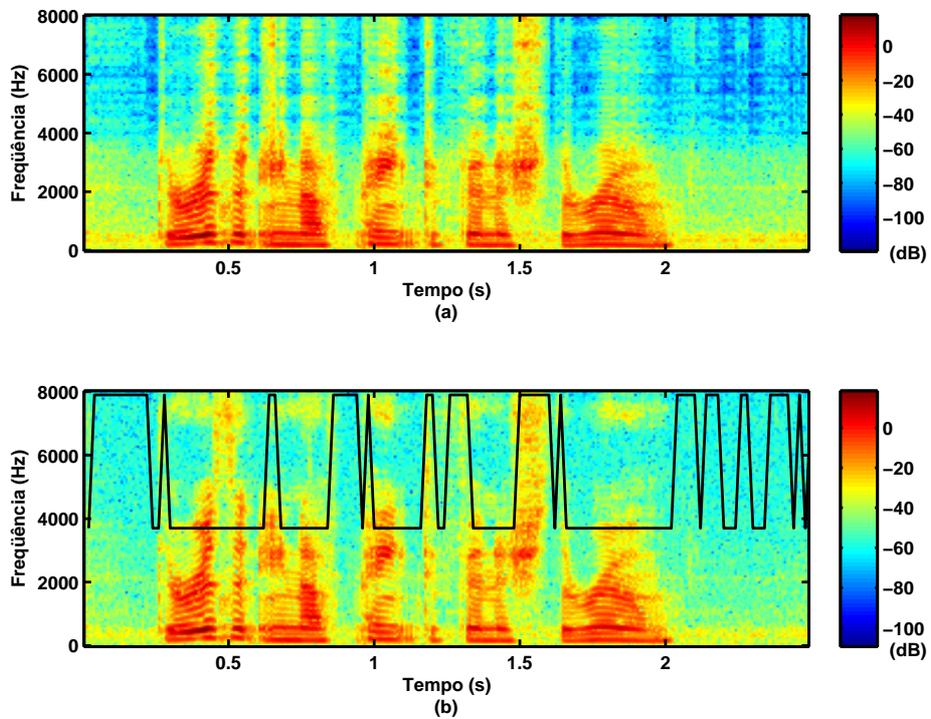


Figura 7.7: (a) Espectrograma do sinal x_{f2}^{BE} estendido, (b) espectrograma de x_{f2}^{BL} destacando-se a classificação realizada pela rede neural.

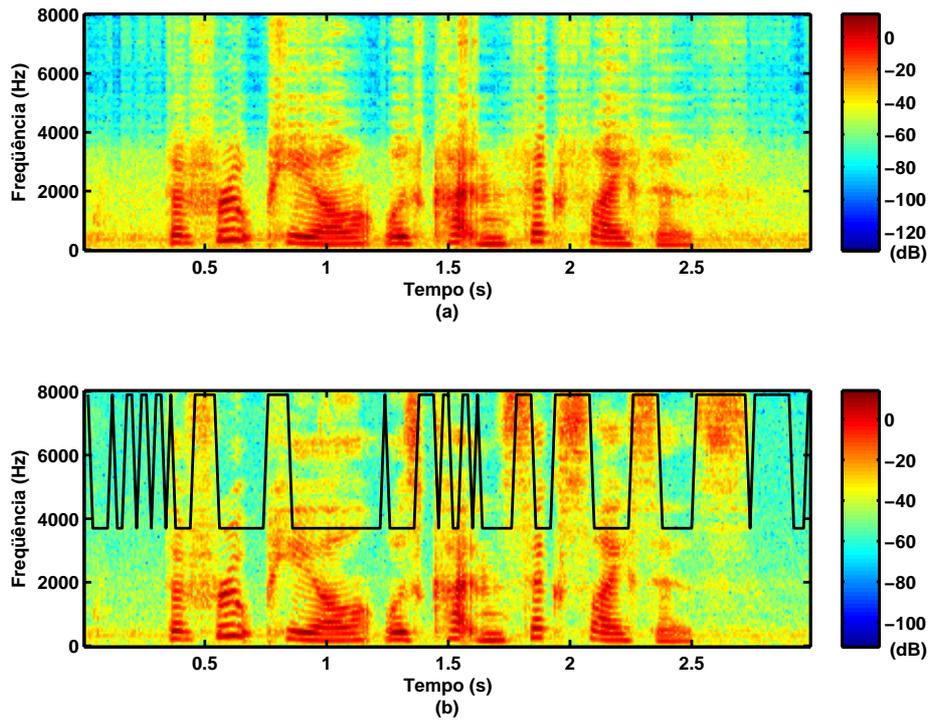


Figura 7.8: (a) Espectrograma do sinal x_{f3}^{BE} estendido, (b) espectrograma de x_{f3}^{BL} destacando-se a classificação realizada pela rede neural.

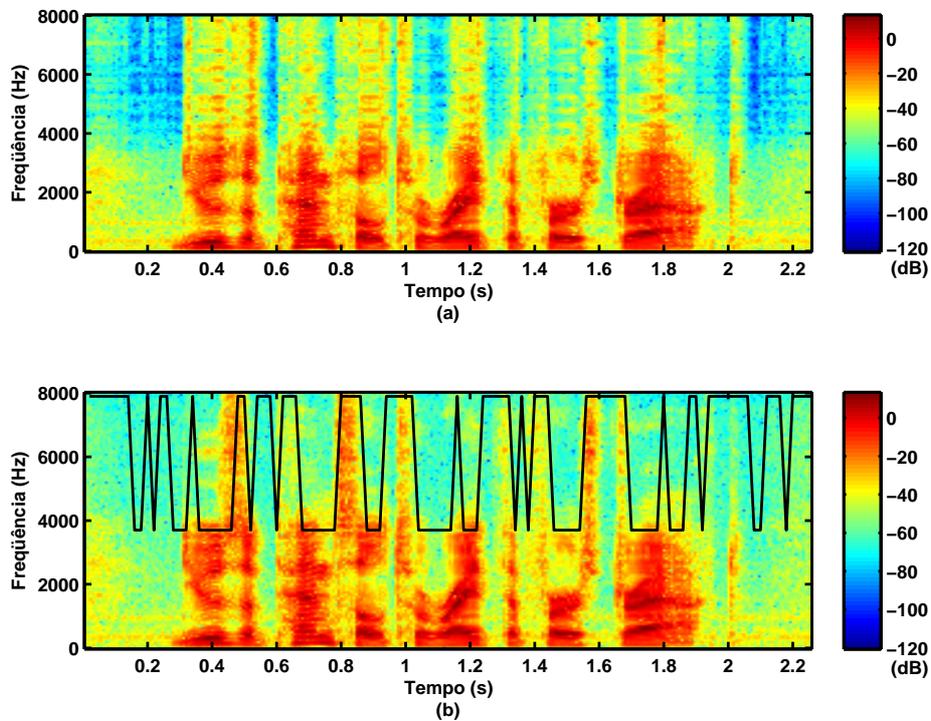


Figura 7.9: (a) Espectrograma do sinal x_{m1}^{BE} estendido, (b) espectrograma de x_{m1}^{BL} destacando-se a classificação realizada pela rede neural.

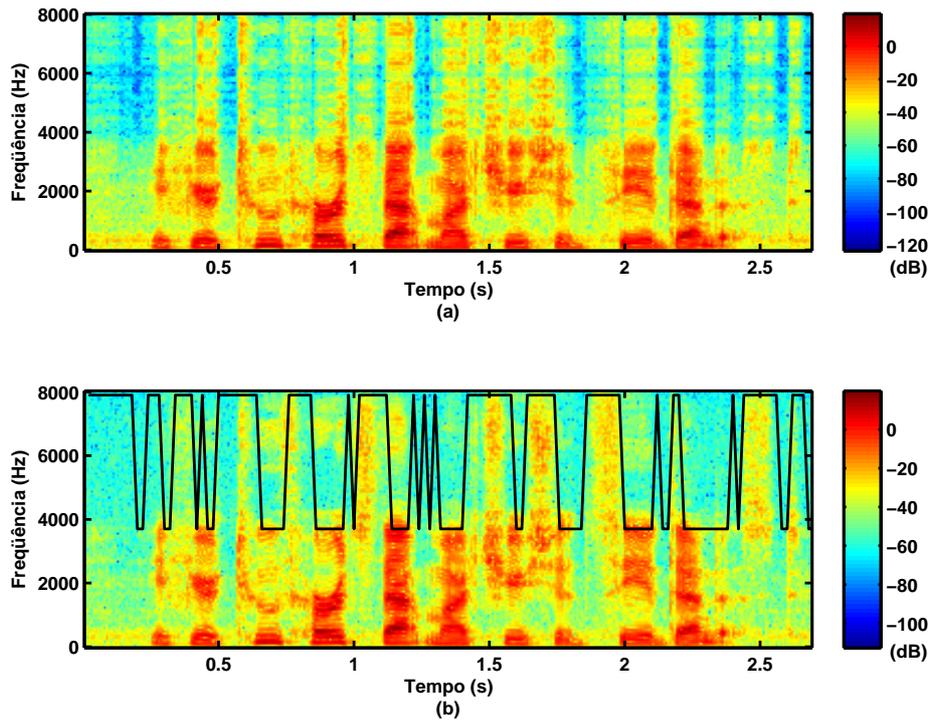


Figura 7.10: (a) Espectrograma do sinal x_{m2}^{BE} estendido, (b) espectrograma de x_{m2}^{BL} destacando-se a classificação realizada pela rede neural.

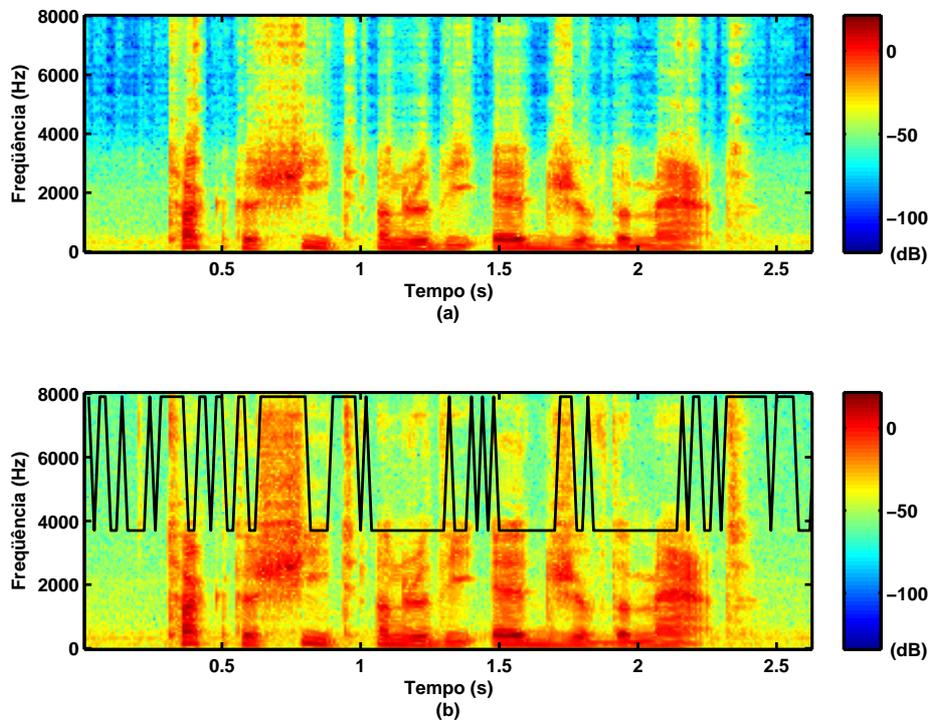


Figura 7.11: (a) Espectrograma do sinal x_{m3}^{BE} estendido, (b) espectrograma de x_{m3}^{BL} destacando-se a classificação realizada pela rede neural.

A seguir, são mostrados os resultados do algoritmo proposto, utilizando filtragem ótima em trechos classificados como não-vozeados e ganhos em trechos classificados como vozeados pela rede neural. As Figuras 7.12-7.17 apresentam os sinais x_{f1}^{BE} , x_{f2}^{BE} , x_{f3}^{BE} , x_{m1}^{BE} , x_{m2}^{BE} e x_{m3}^{BE} estendidos, respectivamente, utilizando este método.

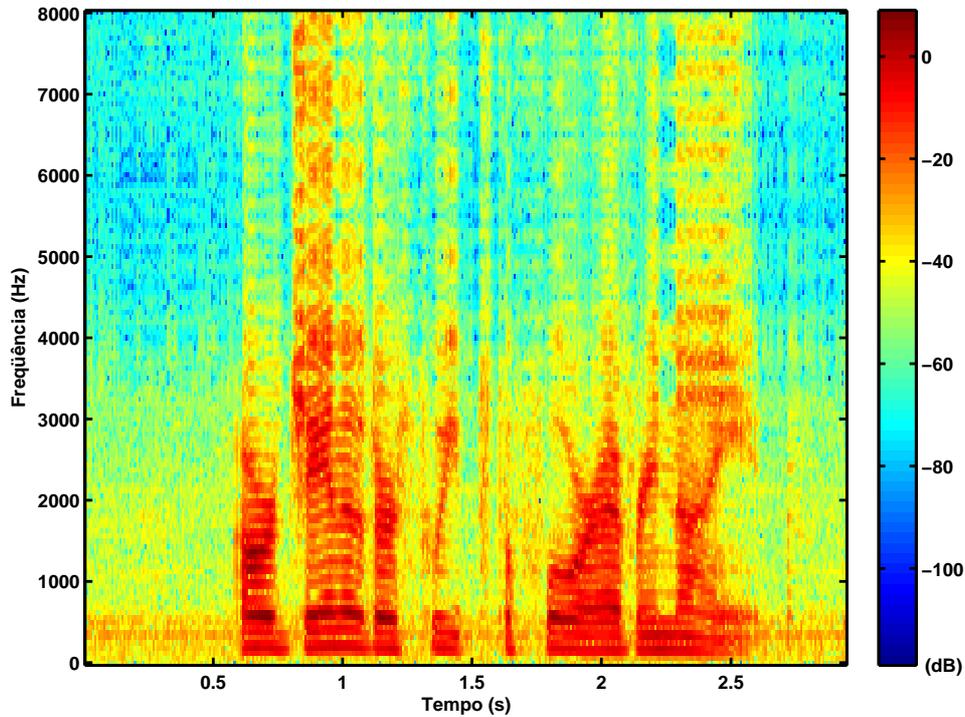


Figura 7.12: Espectrograma do sinal x_{f1}^{BE} estendido utilizando-se ganho nos trechos identificados pela rede neural como vozeados e filtragem ótima nos não-vozeados.

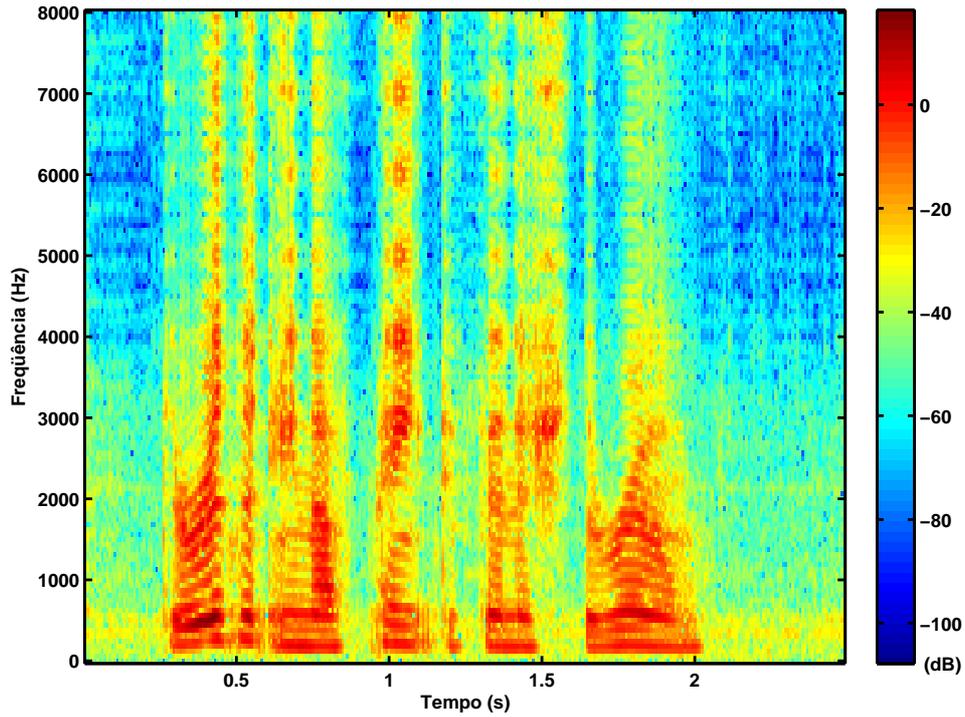


Figura 7.13: Espectrograma do sinal x_{f2}^{BE} estendido utilizando-se ganho nos trechos identificados pela rede neural como vozeados e filtragem ótima nos não-vozeados.

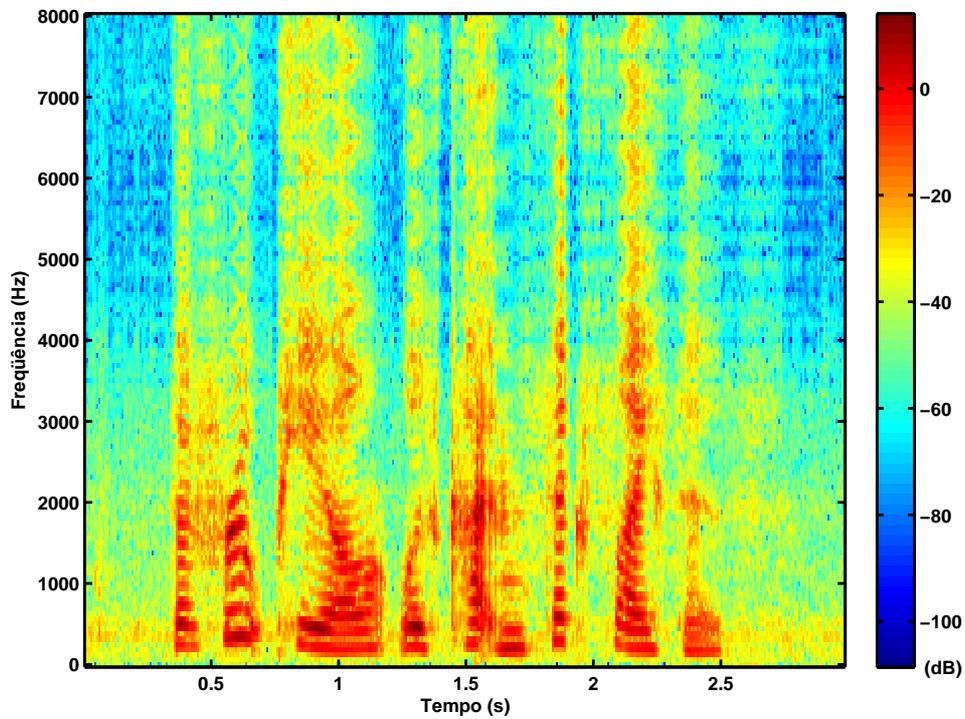


Figura 7.14: Espectrograma do sinal x_{f3}^{BE} estendido utilizando-se ganho nos trechos identificados pela rede neural como vozeados e filtragem ótima nos não-vozeados.

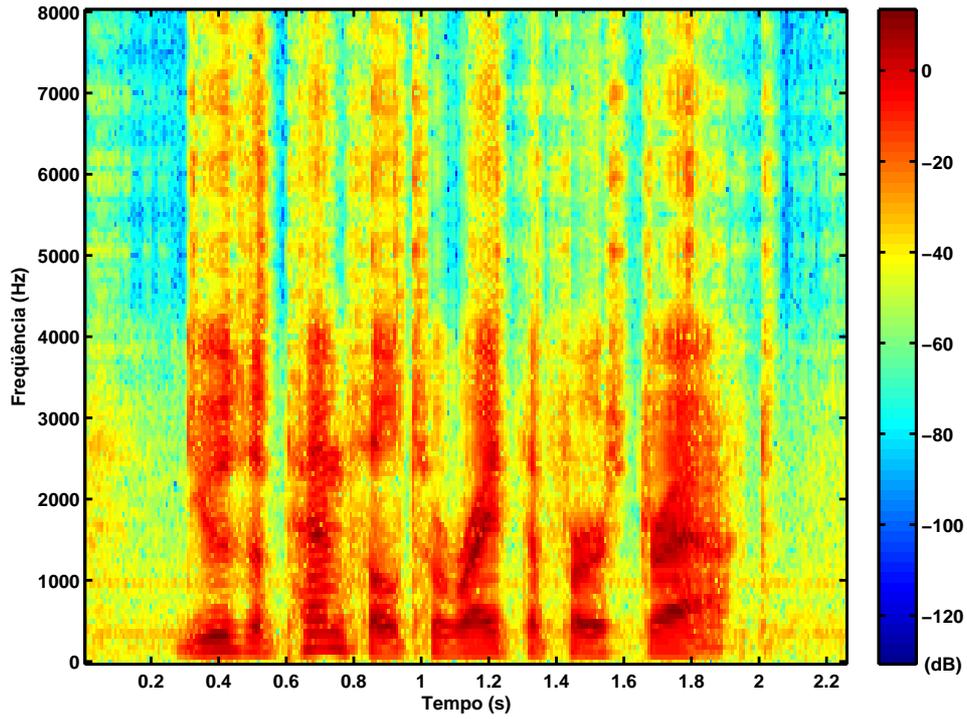


Figura 7.15: Espectrograma do sinal x_{m1}^{BE} estendido utilizando-se ganho nos trechos identificados pela rede neural como vozeados e filtragem ótima nos não-vozeados.

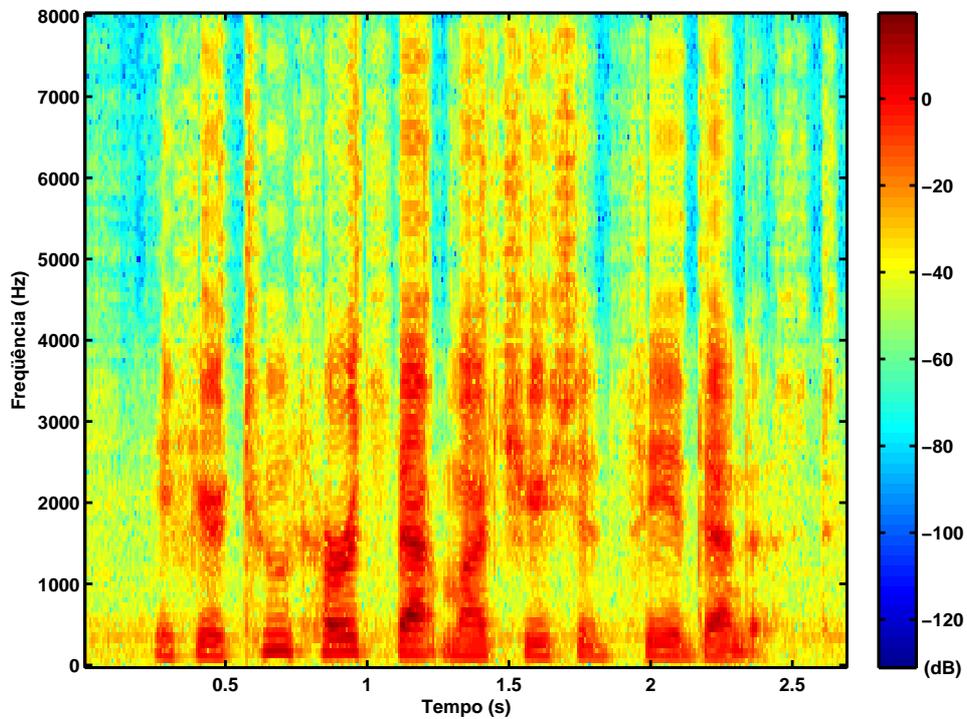


Figura 7.16: Espectrograma do sinal x_{m2}^{BE} estendido utilizando-se ganho nos trechos identificados pela rede neural como vozeados e filtragem ótima nos não-vozeados.

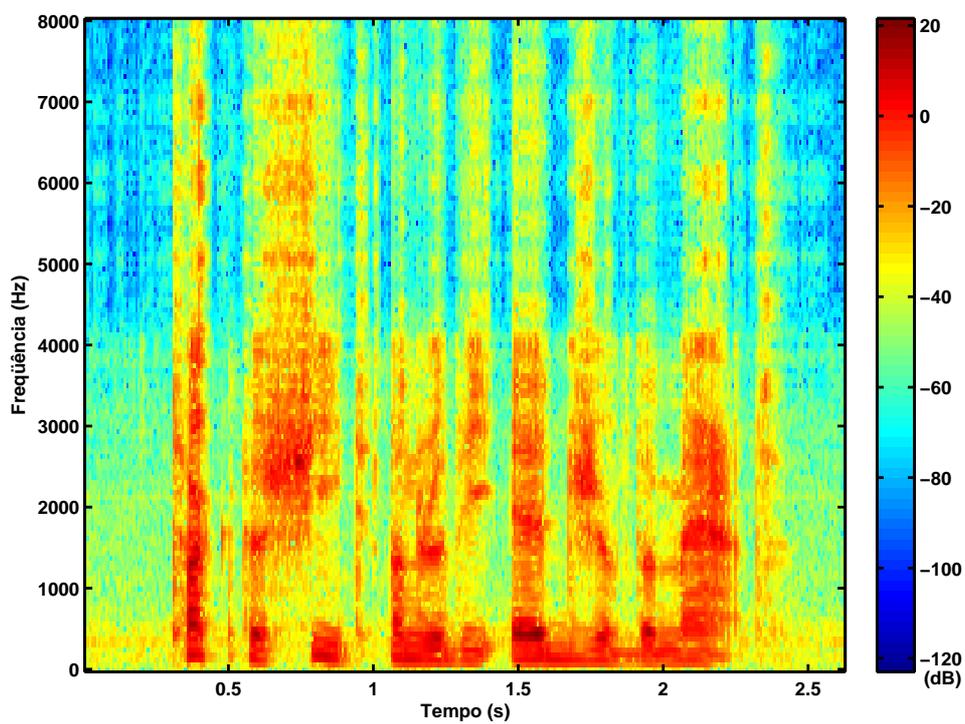


Figura 7.17: Espectrograma do sinal x_{m3}^{BE} estendido utilizando-se ganho nos trechos identificados pela rede neural como vozeados e filtragem ótima nos não-vozeados.

A seguir, são mostrados os resultados do algoritmo proposto, utilizando apenas ganhos, tanto nos trechos classificados como não-vozeados, quanto nos trechos classificados como vozeados. As Figuras 7.18-7.23 apresentam os sinais x_{f1}^{BE} , x_{f2}^{BE} , x_{f3}^{BE} , x_{m1}^{BE} , x_{m2}^{BE} e x_{m3}^{BE} estendidos, respectivamente, utilizando este método.

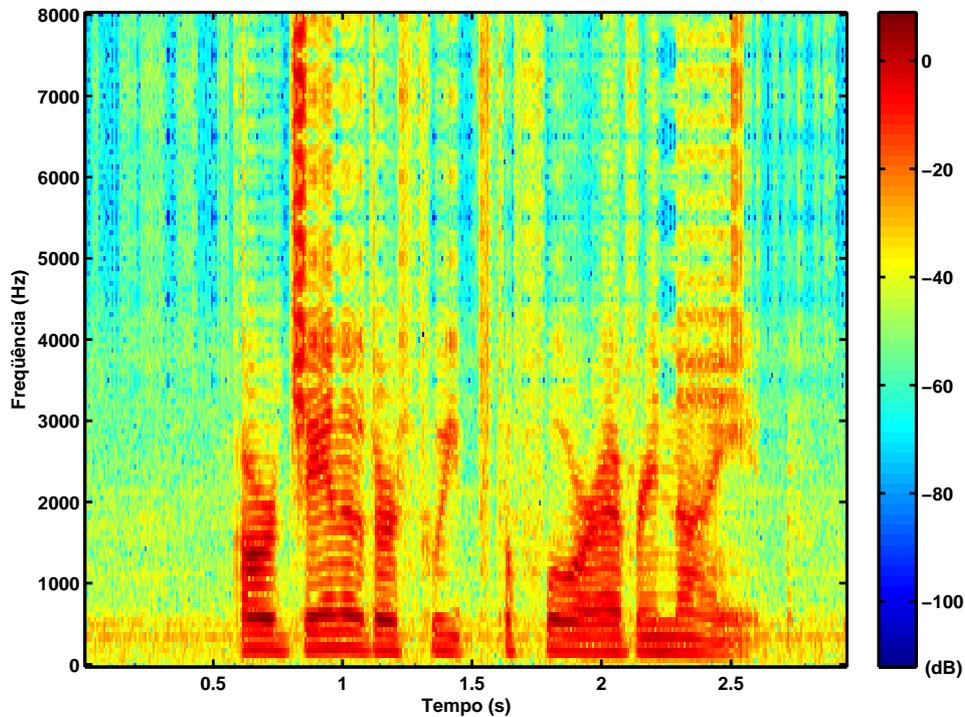


Figura 7.18: Espectrograma do sinal x_{f1}^{BE} estendido utilizando-se ganho tanto nos trechos identificados pela rede neural como vozeados quanto nos não-vozeados.

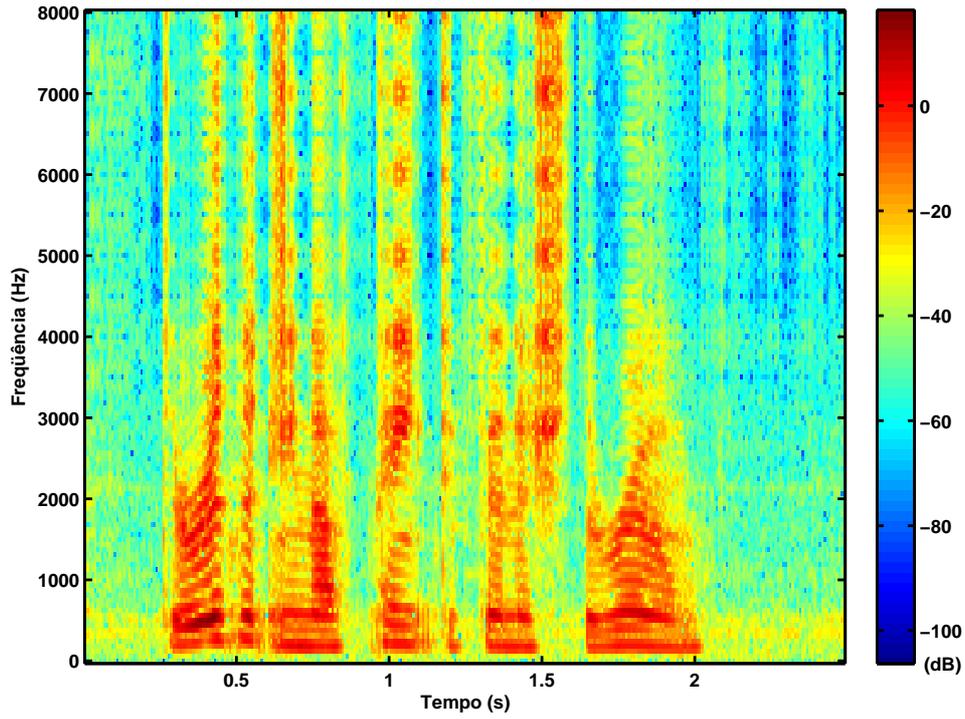


Figura 7.19: Espectrograma do sinal x_{f2}^{BE} estendido utilizando-se ganho tanto nos trechos identificados pela rede neural como vozeados quanto nos não-vozeados.

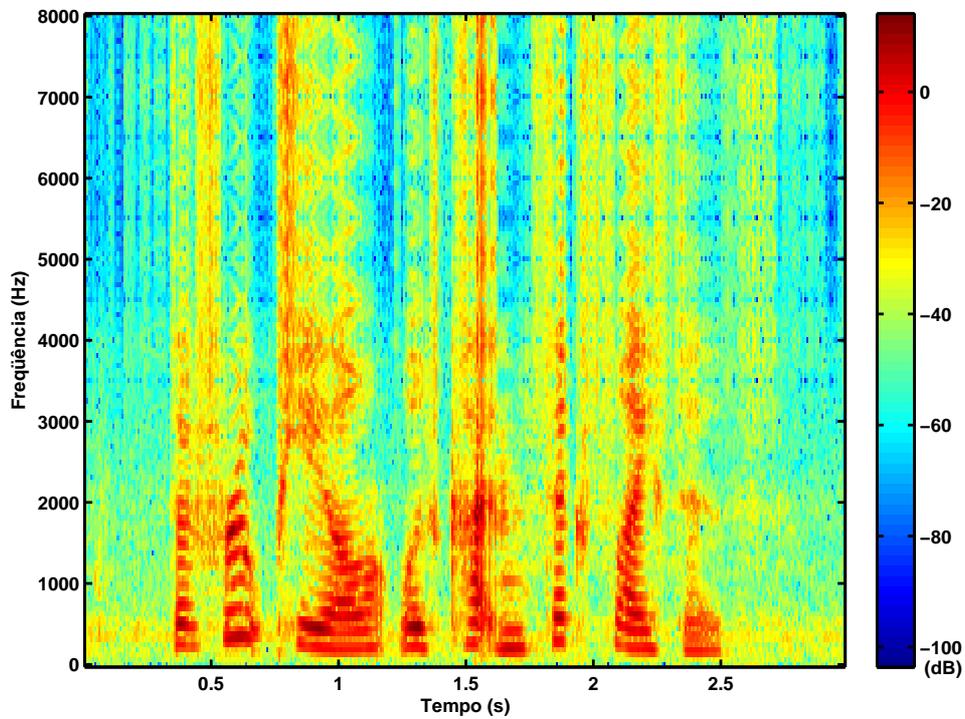


Figura 7.20: Espectrograma do sinal x_{f3}^{BE} estendido utilizando-se ganho tanto nos trechos identificados pela rede neural como vozeados quanto nos não-vozeados.

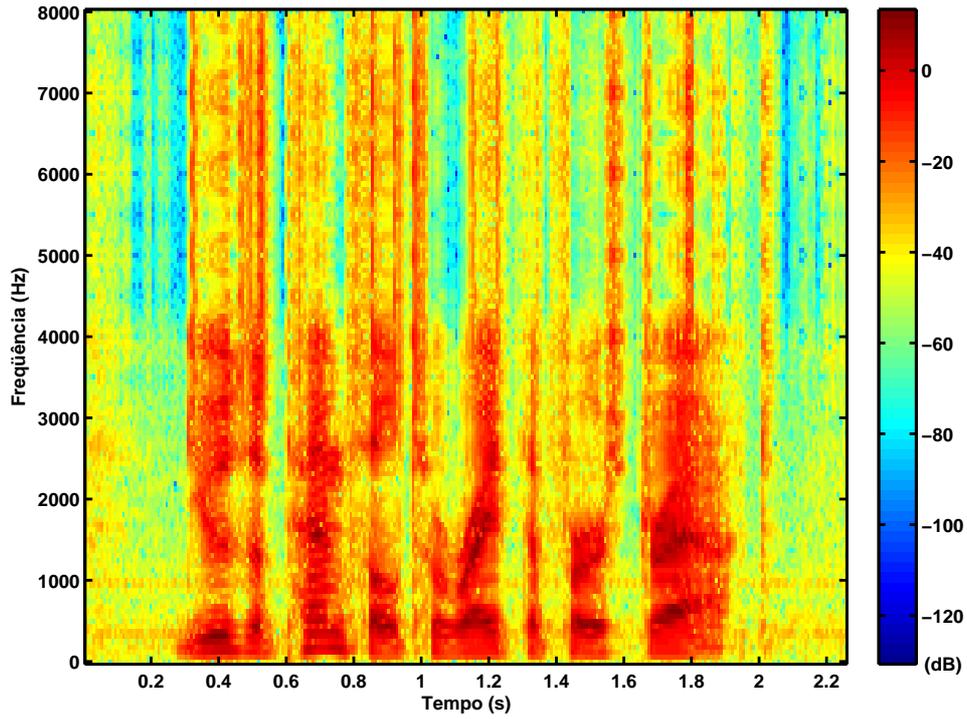


Figura 7.21: Espectrograma do sinal x_{m1}^{BE} estendido utilizando-se ganho tanto nos trechos identificados pela rede neural como vozeados quanto nos não-vozeados.

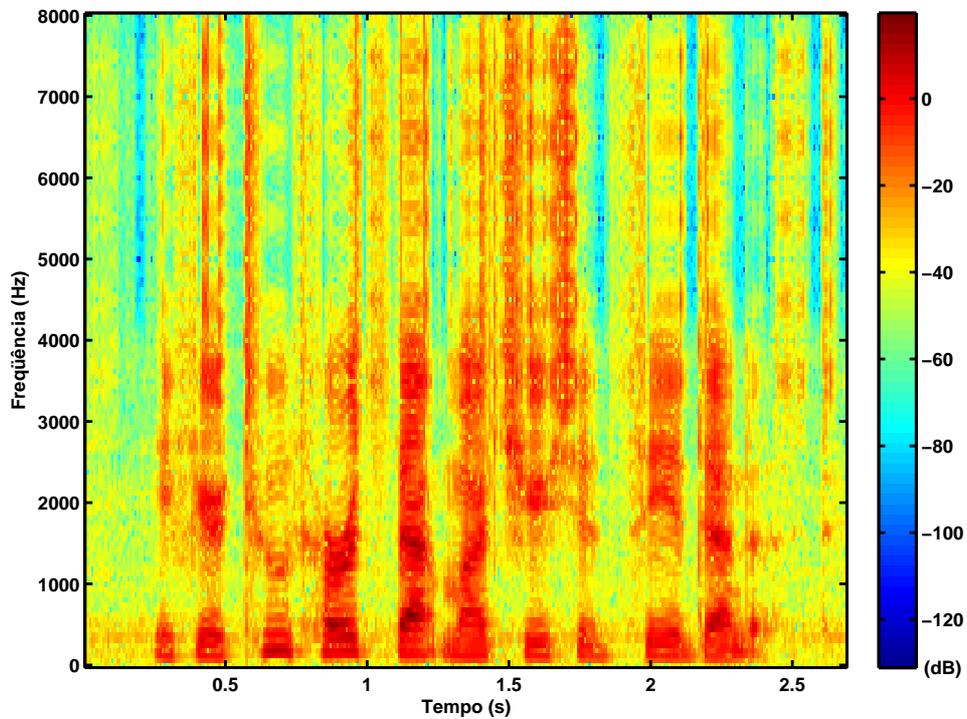


Figura 7.22: Espectrograma do sinal x_{m2}^{BE} estendido utilizando-se ganho tanto nos trechos identificados pela rede neural como vozeados quanto nos não-vozeados.

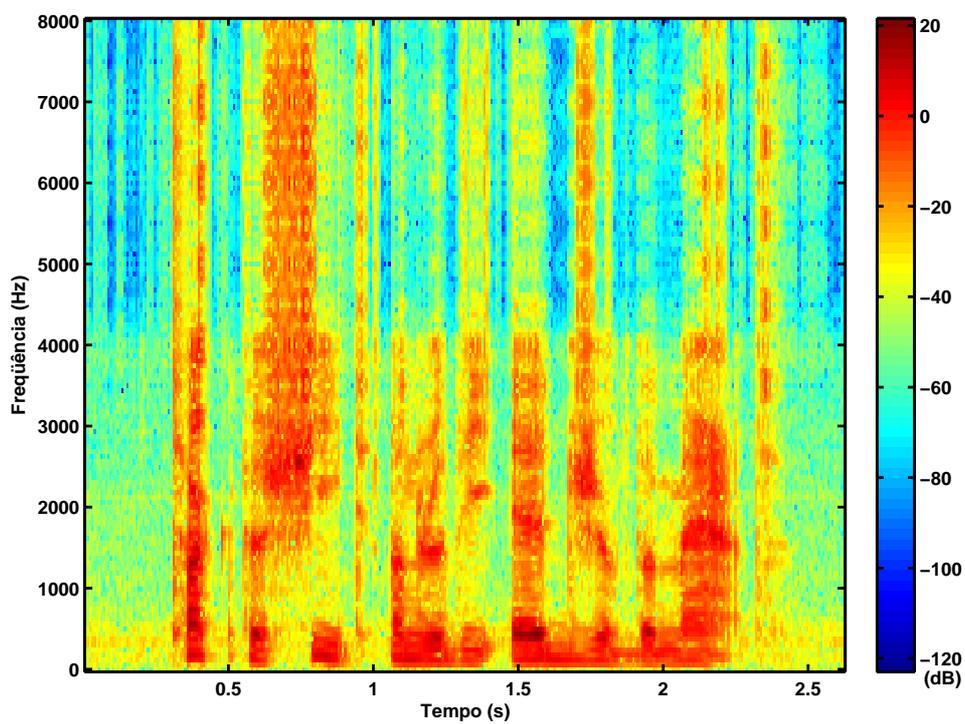


Figura 7.23: Espectrograma do sinal x_{m3}^{BE} estendido utilizando-se ganho tanto nos trechos identificados pela rede neural como vozeados quanto nos não-vozeados.

7.4 Discussão dos Resultados

Também foram utilizados, para efeito de comparação, sinais estendidos por um método baseado no mapeamento dos coeficientes LPC do sinal de banda estreita para a geração dos coeficientes do sinal de banda larga. As Figuras 7.24-7.29 mostram os espectrogramas dos sinais x_{f1}^{BE} , x_{f2}^{BE} , x_{f3}^{BE} , x_{m1}^{BE} , x_{m2}^{BE} e x_{m3}^{BE} estendidos, respectivamente, utilizando este método. Observa-se que, neste método, a extensão em frequência foi realizada até 7,5 kHz, aproximadamente.

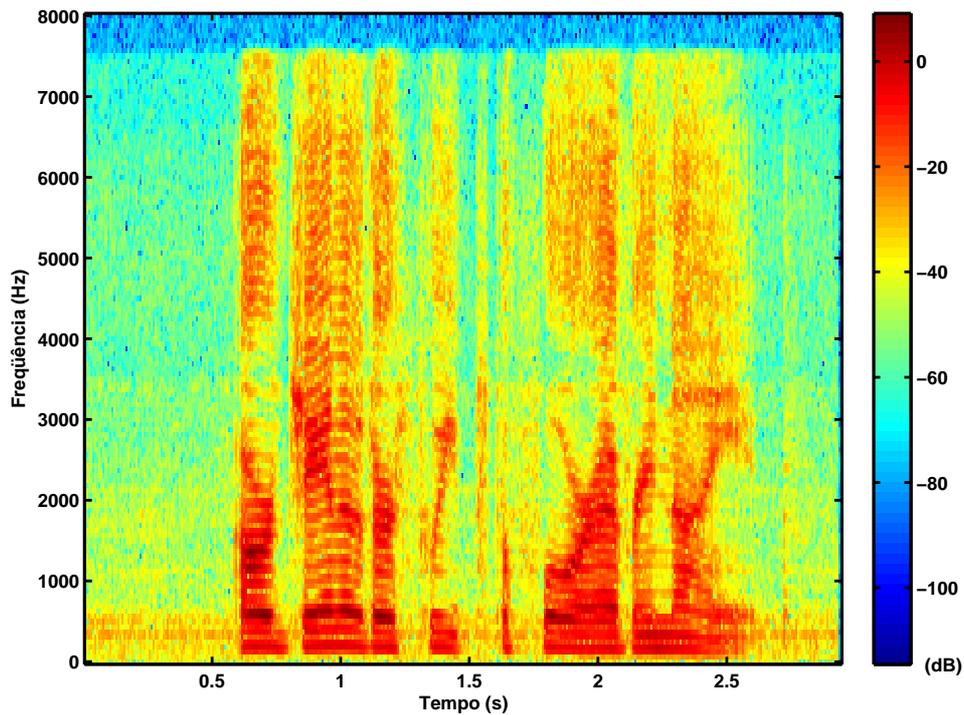


Figura 7.24: Espectrograma do sinal x_{f1}^{BE} estendido utilizando-se um algoritmo baseado em predição linear.

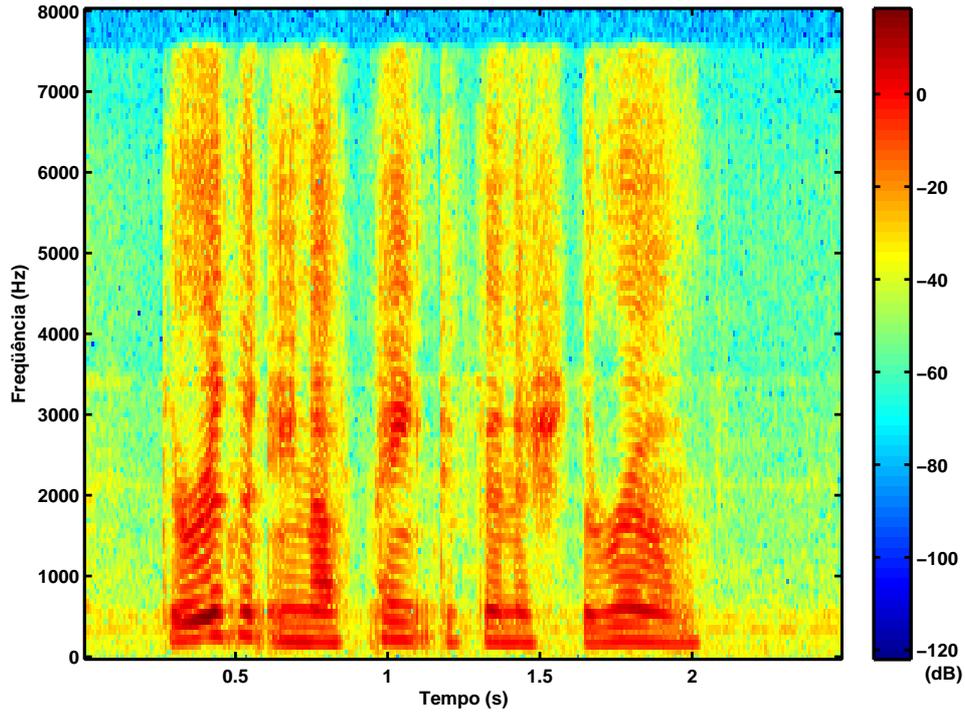


Figura 7.25: Espectrograma do sinal x_{f2}^{BE} estendido utilizando-se um algoritmo baseado em predição linear.

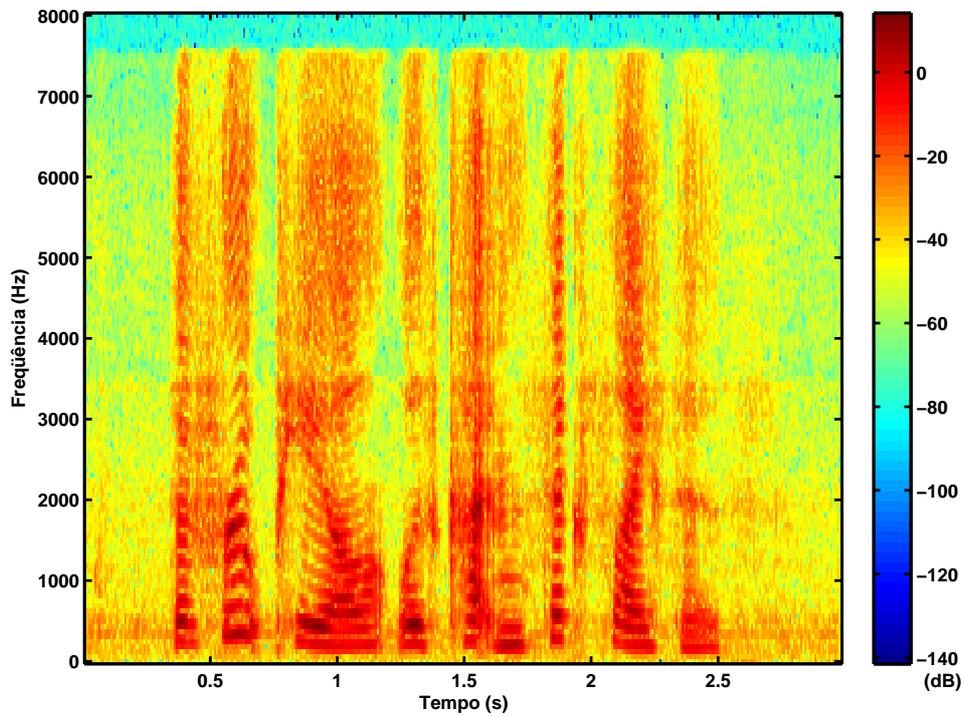


Figura 7.26: Espectrograma do sinal x_{f3}^{BE} estendido utilizando-se um algoritmo baseado em predição linear.

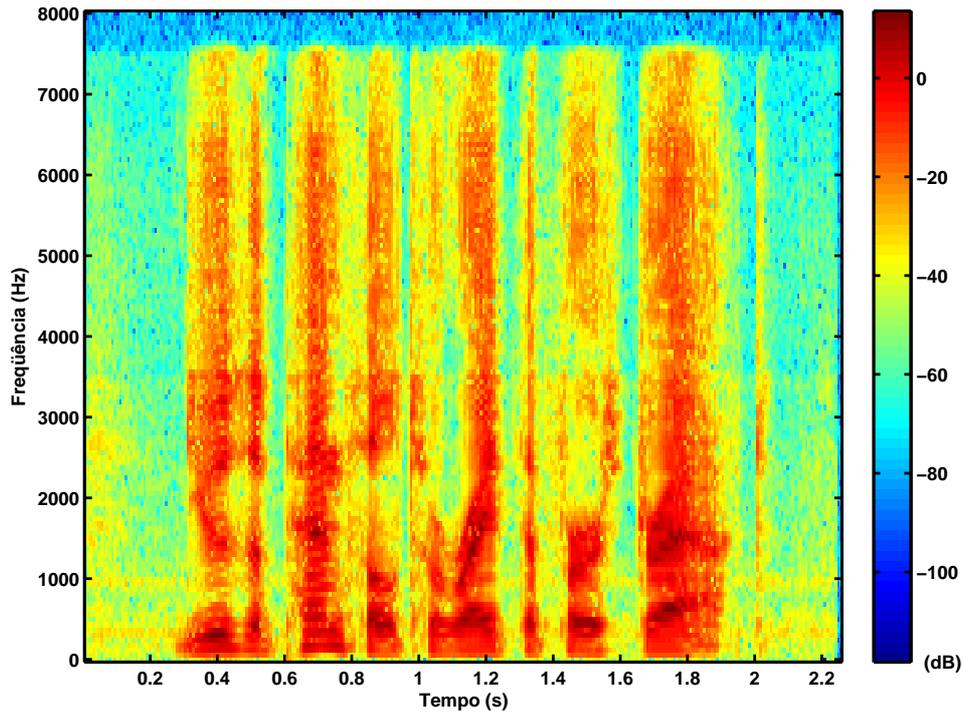


Figura 7.27: Espectrograma do sinal x_{m1}^{BE} estendido utilizando-se um algoritmo baseado em predição linear.

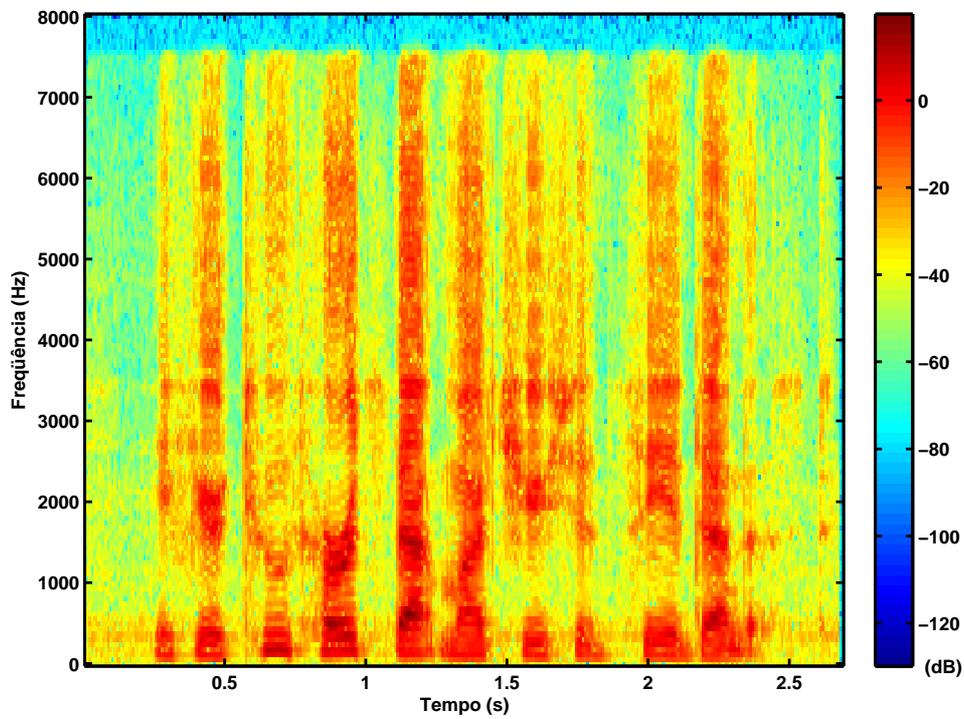


Figura 7.28: Espectrograma do sinal x_{m2}^{BE} estendido utilizando-se um algoritmo baseado em predição linear.

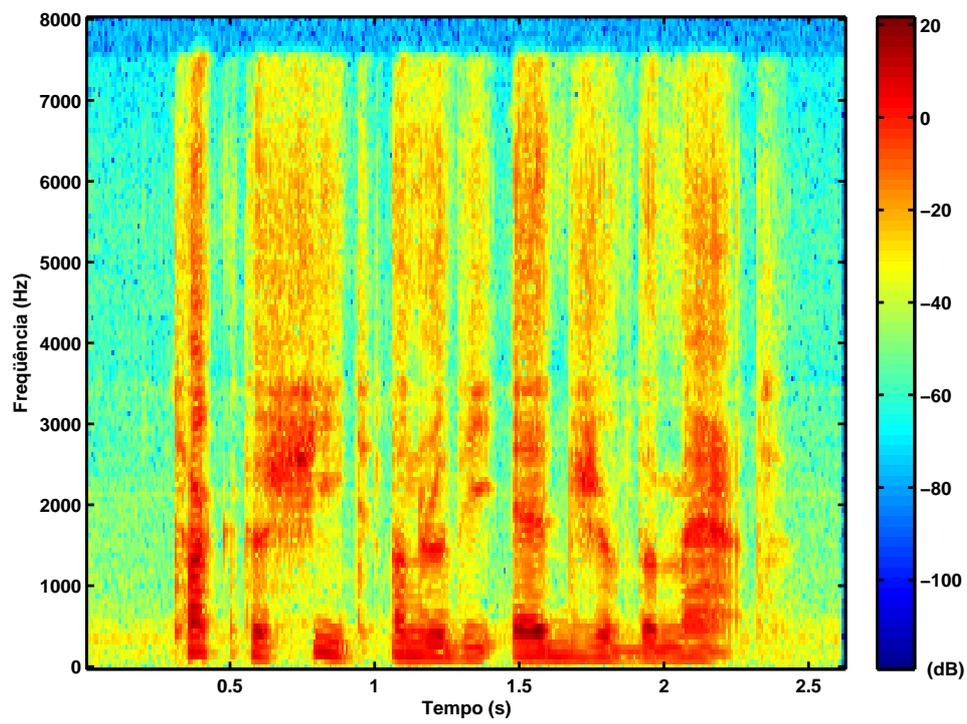


Figura 7.29: Espectrograma do sinal x_{m3}^{BE} estendido utilizando-se um algoritmo baseado em predição linear.

Para efeito de avaliação da qualidade auditiva e da extensão dos sinais estendidos apresentados na Seção 7.3, foram realizados testes subjetivos com os mesmos e com os sinais estendidos por meio do método de predição linear (Figuras 7.24-7.29).

Os testes subjetivos foram realizados com um total de 30 ouvintes, os quais conferiram notas subjetivas para os seis sinais de voz estendidos citados anteriormente. Cada ouvinte foi orientado a conferir duas notas subjetivas, de 1 a 5, para cada sinal de voz: uma nota para a extensão, ou seja, o quanto cada sinal foi estendido em relação aos sinais de banda estreita, e a outra para avaliar a qualidade dos sinais estendidos em relação aos sinais de banda estreita. As Figuras 7.30-7.35 apresentam o resultado da média das notas conferidas pelos 30 ouvintes para a extensão e a qualidade dos sinais estendidos. As notas são comparativas em relação aos sinais de banda estreita e têm o seguinte significado:

- nota 1 - o sinal estendido é muito pior do que o sinal de banda estreita,
- nota 2 - o sinal estendido é pior do que o sinal de banda estreita,
- nota 3 - o sinal estendido é semelhante ao sinal de banda estreita,
- nota 4 - o sinal estendido é melhor do que o sinal de banda estreita,
- nota 5 - o sinal estendido é muito melhor do que o sinal de banda estreita.

Os algoritmos utilizados na extensão dos sinais de banda estreita considerados nos testes subjetivos foram:

- Algoritmo 1 - algoritmo proposto com a utilização de filtragem ótima para os *frames* classificados tanto como vozeados quanto como não-vozeados;
- Algoritmo 2 - algoritmo proposto com a utilização de ganhos para a extensão dos *frames* classificados como vozeados e filtragem ótima para os *frames* classificados como não-vozeados;
- Algoritmo 3 - algoritmo proposto com a utilização de ganhos para os *frames* classificados tanto como vozeados quanto como não-vozeados;
- Algoritmo 4 - método baseado no mapeamento dos coeficientes LPC do sinal de banda estreita para a geração dos coeficientes do sinal de banda larga.

Pela análise da Figura 7.30 pode-se observar que o Algoritmo 4 obteve o melhor resultado quanto à extensão. Em seguida, o Algoritmo 2 obteve o segundo melhor resultado de extensão. Os Algoritmos 1 e 3 realizaram pouca extensão em relação ao sinal de banda estreita. Considerando a qualidade dos sinais estendidos, verifica-se que os Algoritmos 1 e 4 obtiveram os melhores resultados e estes foram semelhantes. O Algoritmo 3 apresentou qualidade inferior à do sinal de banda estreita.

Pela análise da Figura 7.31 pode-se observar que os Algoritmos 2, 3 e 4 obtiveram resultados semelhantes quanto à extensão realizada, enquanto que o Algoritmo 1 apresentou um resultado um pouco inferior em relação ao sinal de banda estreita. Considerando a qualidade dos sinais estendidos, verifica-se que o Algoritmo 1 obteve o melhor resultado, seguido pelos Algoritmos 4, 2 e 3. Novamente, o Algoritmo 3 apresentou qualidade inferior à do sinal de banda estreita.

Pela análise da Figura 7.32 pode-se observar que o Algoritmo 4 obteve o melhor resultado quanto à extensão. Porém, os Algoritmos 2 e 3 também obtiveram bons resultados. O Algoritmo 1 obteve pouca extensão em relação ao sinal de banda estreita. Em relação à qualidade, verifica-se que o Algoritmo 1 obteve o melhor resultado e foi o único considerado com melhor qualidade do que o sinal de banda estreita. Os Algoritmos 2 e 4 apresentaram resultados semelhantes para a qualidade, porém, considerados com qualidade um pouco inferior à do sinal de banda estreita. O Algoritmo 3 apresentou, novamente, qualidade inferior à do sinal de banda estreita.

Pela análise da Figura 7.33 pode-se observar que os Algoritmos 3 e 4 obtiveram melhores resultados quanto à extensão. Porém, os Algoritmos 1 e 2 também obtiveram bons resultados considerando a extensão. Em relação à qualidade, verifica-se que os Algoritmos 1 e 4 obtiveram melhores resultados. Os Algoritmos 2 e 3 apresentaram resultados semelhantes para a qualidade, porém, foram considerados com qualidade um pouco inferior à do sinal de banda estreita.

Pela análise da Figura 7.34 pode-se observar que o Algoritmo 4 obteve o melhor resultado quanto a extensão. Porém, os Algoritmos 1 e 2 também obtiveram bons resultados considerando a extensão. O Algoritmo 3 obteve pouca extensão em relação ao sinal de banda estreita. Em relação à qualidade, verifica-se que os

Algoritmos 1 e 4 obtiveram os melhores resultados e com melhor qualidade do que o sinal de banda estreita. O Algoritmo 2 apresentou um resultado considerado com qualidade um pouco inferior à do sinal de banda estreita. O Algoritmo 3 apresentou, novamente, qualidade inferior à do sinal de banda estreita.

Pela análise da Figura 7.35 pode-se observar que os Algoritmos 2 e 4 obtiveram os melhores resultados quanto à extensão. Porém, os Algoritmos 1 e 3 também obtiveram bons resultados considerando a extensão, sendo que o Algoritmo 3 obteve um resultado um pouco melhor do que o Algoritmo 1. Em relação à qualidade, verifica-se que os Algoritmos 1, 2 e 4 obtiveram os melhores resultados. O Algoritmo 3 apresentou, novamente, qualidade inferior à do sinal de banda estreita.

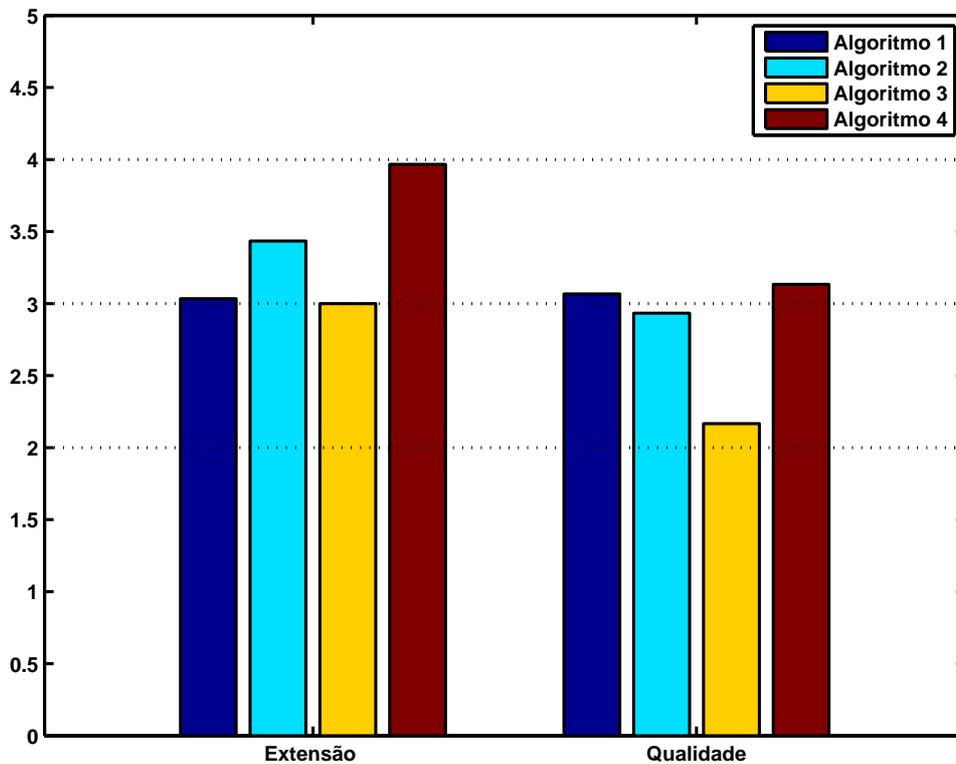


Figura 7.30: Resultado do teste subjetivo para o sinal x_{f1}^{BE} estendido.

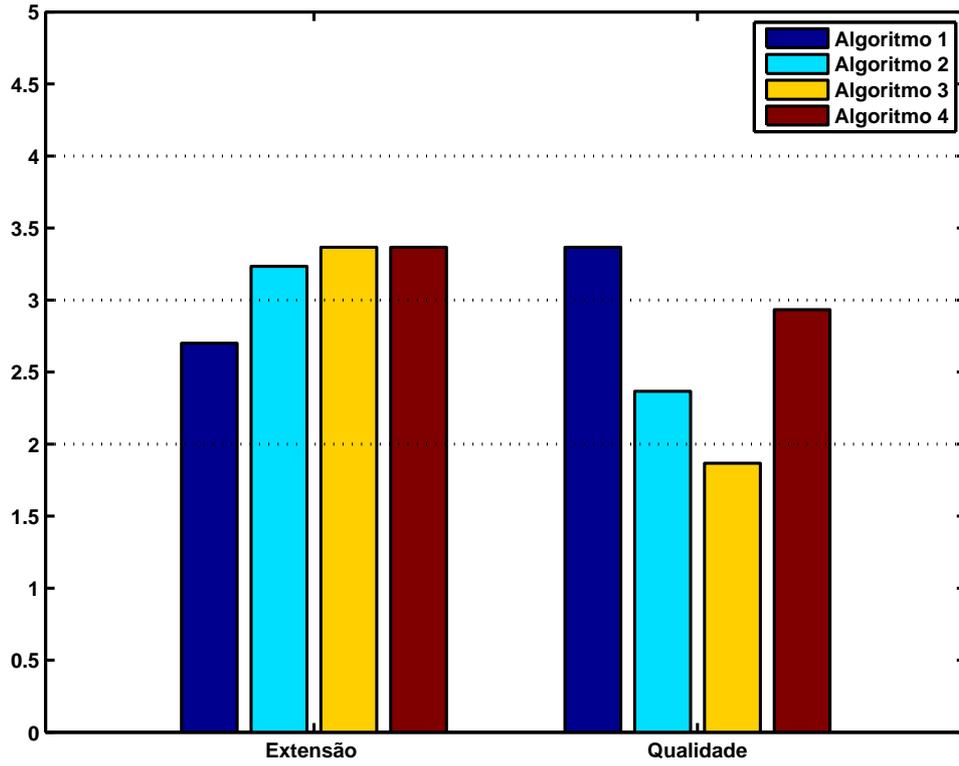


Figura 7.31: Resultado do teste subjetivo para o sinal $x_{f_2}^{BE}$ estendido.

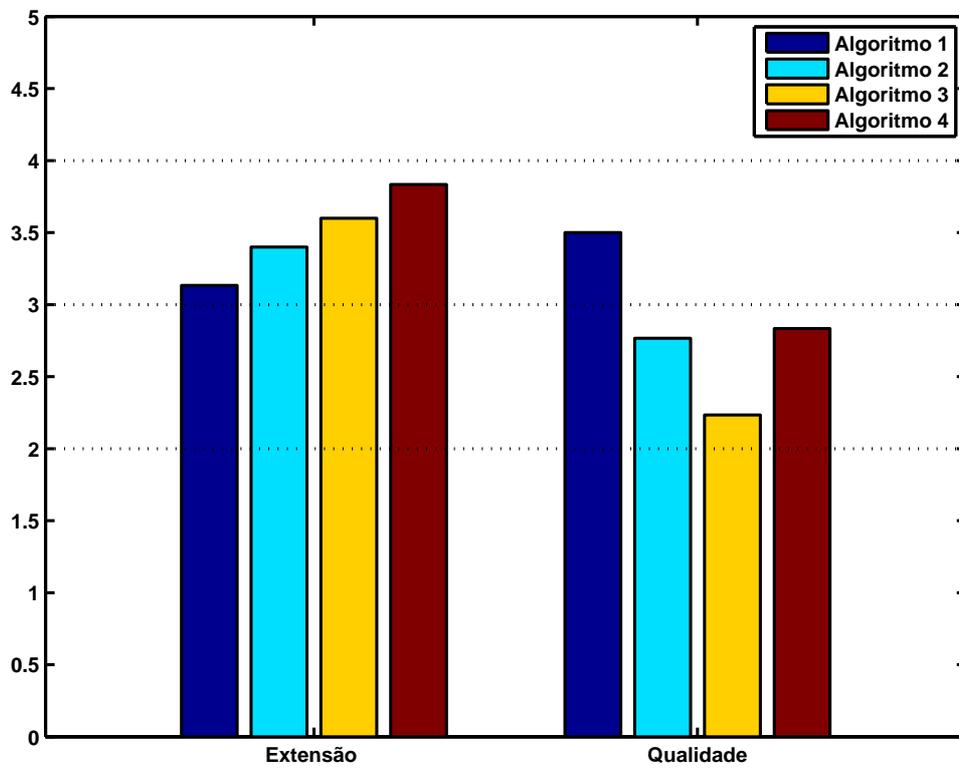


Figura 7.32: Resultado do teste subjetivo para o sinal $x_{f_3}^{BE}$ estendido.

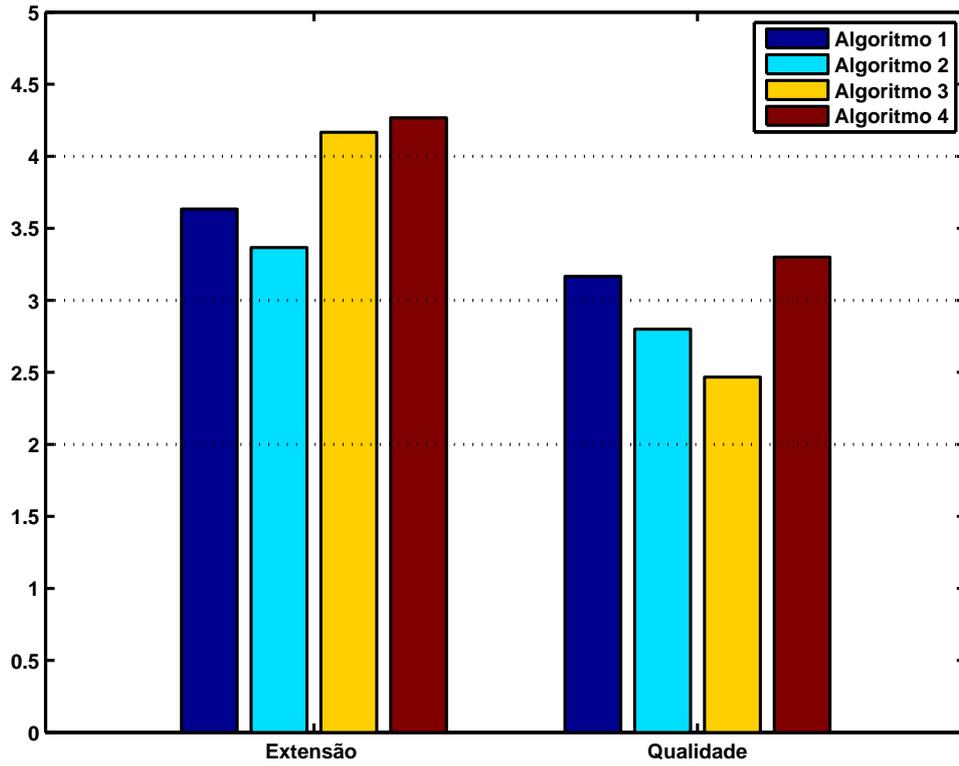


Figura 7.33: Resultado do teste subjetivo para o sinal x_{m1}^{BE} estendido.

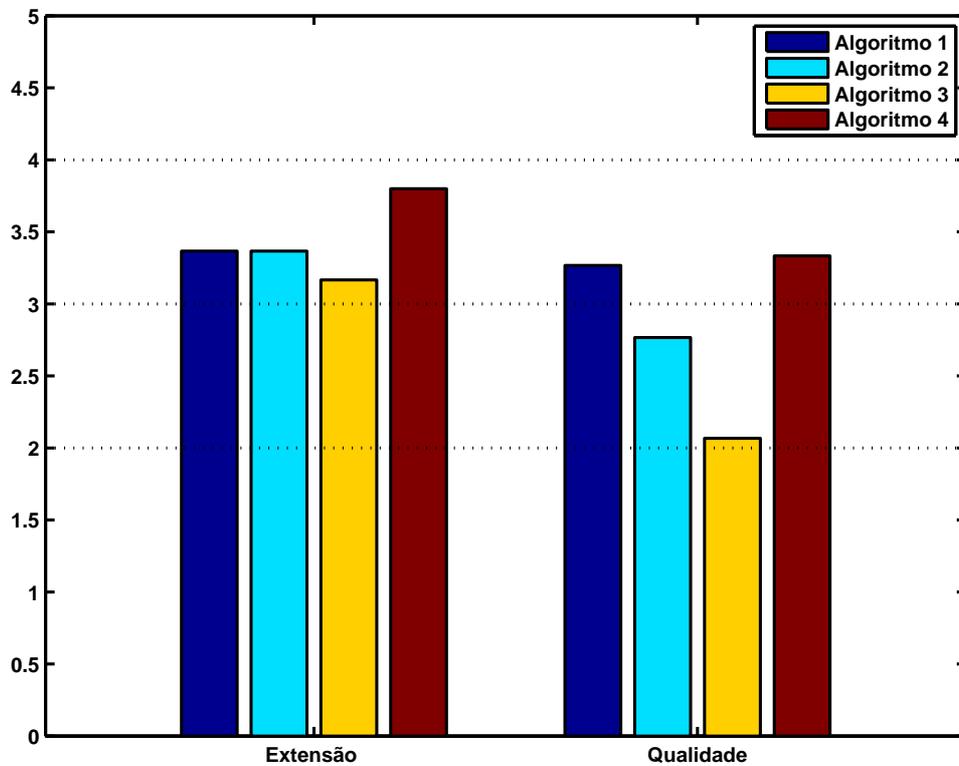


Figura 7.34: Resultado do teste subjetivo para o sinal x_{m2}^{BE} estendido.

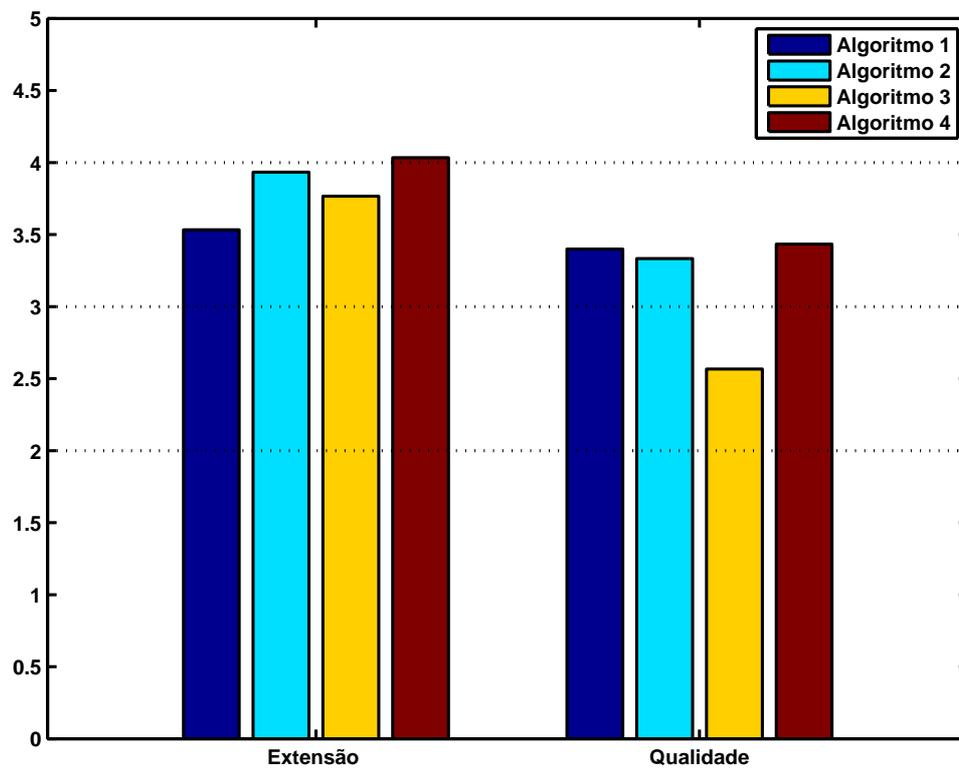


Figura 7.35: Resultado do teste subjetivo para o sinal x_{m3}^{BE} estendido.

Avaliando apenas o método proposto neste trabalho e suas variações (Algoritmos 1, 2 e 3), pode-se concluir que, considerando a extensão e a qualidade dos sinais resultantes da aplicação destes algoritmos e observando os resultados dos testes subjetivos (Figuras 7.30-7.35), o Algoritmo 1 (algoritmo que utiliza filtragem ótima para os *frames* classificados pela rede neural como vozeados e não-vozeados) foi o que obteve melhores resultados. Considerando apenas o critério da extensão, os Algoritmos 2 e 3 obtiveram os melhores resultados. Considerando apenas o critério da qualidade, o Algoritmo 1 obteve o melhor resultado.

Em relação às notas conferidas pelos ouvintes, os melhores resultados citados acima estiveram quase sempre entre as notas 3 e 4, ou seja, foram julgados como sendo um pouco melhores que os respectivos sinais de banda estreita.

Considerando o Algoritmo 4 em relação aos outros, o primeiro apresentou resultados melhores, considerando-se o critério da extensão. Considerando o critério da qualidade, o Algoritmo 4 apresentou resultados ora equivalentes, ora um pouco piores do que os resultados do Algoritmo 1.

Assim, considerando o objetivo de desenvolvimento de um método de menor complexidade computacional em relação aos métodos existentes, dentre os quais alguns foram citados no Capítulo 1, os resultados foram satisfatórios. Comparado com um método de complexidade computacional superior (Algoritmo 4), o Algoritmo 1 apresentou resultados melhores ou equivalentes no critério qualidade e piores resultados segundo o critério de extensão. Porém, quase sempre melhores que os sinais de banda estreita.

Capítulo 8

Conclusões e Trabalhos Futuros

8.1 Conclusões

Pela análise dos resultados dos testes subjetivos apresentados na Seção 7.3, conclui-se que o método desenvolvido neste trabalho apresentou resultados satisfatórios, tanto no critério de extensão quanto no de qualidade dos sinais estendidos.

De maneira geral, o Algoritmo 1 (algoritmo que utiliza filtragem ótima na extensão dos trechos classificados tanto como vozeados quanto como não-vozeados) obteve os melhores resultados em relação ao Algoritmo 2 (algoritmo que utiliza ganhos na extensão dos trechos classificados como vozeados e filtragem ótima nos não-vozeados) e ao Algoritmo 3 (algoritmo que utiliza ganhos na extensão dos trechos classificados tanto como vozeados quanto como não-vozeados). Os Algoritmos 2 e 3 apresentaram melhores resultados de extensão, em relação ao Algoritmo 1. Porém, este último apresentou melhores resultados no critério de qualidade.

Quando comparado ao Algoritmo 4, que é um método de maior complexidade computacional do que aquele desenvolvido neste trabalho, o Algoritmo 1 apresentou resultados semelhantes ou melhores no critério de qualidade. Contudo, o Algoritmo 1 obteve resultados piores no critério de extensão.

8.2 Trabalhos Futuros

O método desenvolvido neste trabalho foi realizado utilizando duas classes (vozeados e não-vozeados) na classificação dos *frames* pela rede neural. Pode-se

utilizar uma terceira classe para contemplar os sons plosivos [17].

Outras combinações de parâmetros podem ser utilizadas como conjunto de treinamento da rede neural para tentar melhorar o desempenho da rede com três classes, tais como: a energia normalizada do *frame*, *gradient index*, *spectral flatness*, dentre outros [18].

Referências Bibliográficas

- [1] QIAN, Y., KABAL, P., “Combining Equalization and Estimation for Bandwidth Extension of Narrowband Speech”, *Proceeding IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, v. 1, pp. I-713–I-716, May 2004.
- [2] ISER, B., SCHMIDT, G., “Bandwidth Extension of Telephony Speech”, *EURASIP Newsletter - Short Tutorials*, v. 1, pp. 1–24, 2005.
- [3] JASON A. FUEMMELE, R. C. H., GARDNER, W. R., “Techniques for the Regeneration of Wideband Speech from Narrowband Speech”, *EURASIP Journal on Applied Signal Processing*, pp. 266–274, 2001.
- [4] PETER JAX, P. V., “On Artificial Bandwidth Extension of Speech Signals”, *Signal Processing*, v. 83, n. 8, pp. 1707–1719, August 2003.
- [5] GUO CHEN, V. P., “HMM-Based Frequency Bandwidth Extension for Speech Enhancement Using Line Spectral Frequencies”, *Proceeding IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. I-709–I-712, 2004.
- [6] S. CHENNOUKH, A. GERRITS, G. M., SLUIJTER, R., “Speech Enhancement via Frequency Bandwidth Extension Using Line Spectral Frequencies”, *Proceeding IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, , May 2001.
- [7] NILSSON, M., KLEIJN, W. B., “Avoiding Over-Estimation in Bandwidth Extension of Telephony Speech”, *Proceeding IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 869–872, May 2001.

- [8] KECCMAN, V., *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*. The MIT Press, 2001.
- [9] WASSERMAN, P. D., *Neural Computing: Theory and Practice*. New York, USA, Van Nostrand Reinhold, 1989.
- [10] HAYKIN, S., *Neural Networks: A Comprehensive Foundation*. MacMillan Publishing Company, 1994.
- [11] CALÔBA, L., “Notas de Aula do Curso: Redes Neurais Feedforward”, 2003, Departamento de Eletrônica, Universidade Federal do Rio de Janeiro.
- [12] HAYKIN, S., *Adaptive Filter Theory*. Prentice Hall.
- [13] VAIDYANATHAN, P. P., *Multirate Systems and Filter Banks*. Prentice Hall, 1993.
- [14] KORNAGEL, U., “Spectral Widening of the Excitation Signal for Telephone-Band Speech Enhancement”, *Proc. IWAENC '01*, pp. 215–218, September 2001.
- [15] EPPS, J., HOLMES, W. H., “A New Technique for Wideband Enhancement of Coded Narrowband Speech”, *Proc. IEEE Speech Coding Workshop*, pp. 174–176, 1999.
- [16] CARL, H., HEUTE, U., “Bandwidth Enhancement of Narrow-Band Speech Signals”, *European Signal Processing Conf.*, v. 2, pp. 1178–1181, 1994.
- [17] RABINER, L. R., SCHAFER, R. W., *Digital Processing of Speech Signals*. Prentice-Hall, 1978.
- [18] JAX, P., VARY, P., “Feature Selection for Improved Bandwidth Extension of Speech Signals”, *Pro. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, v. 1, pp. 667–700, 2004.
- [19] NGUYEN, T. Q., “Digital Filter Bank Design Quadratic-Constrained Formulation”, *IEEE Transactions on Signal Processing*, v. 43, n. 9, pp. 2103–2108, September 1995.
- [20] ALAN V. OPPENHEIM, R. W. S. W. J. R. B., *Discrete-Time Signal Processing*. Upper Saddle River, New Jersey, Prentice Hall, 1998.