



**DEPARTAMENTO DE
ENGENHARIA ELECTROTÉCNICA E DE COMPUTADORES**

**MODELIZAÇÃO PARAMÉTRICA
DE SINAIS PARA APLICAÇÃO EM
SISTEMAS DE CONVERSÃO
TEXTO-FALA**

004(043) / TEX / MOD

**FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO**

Rua dos Bragas, 4099 Porto Codex – PORTUGAL

UNIVERSIDADE DO PORTO

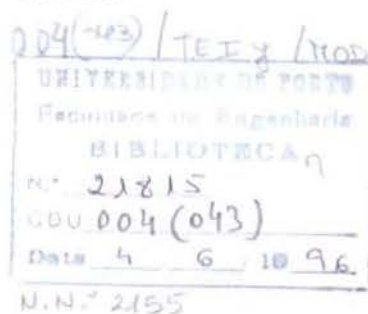
FACULDADE DE ENGENHARIA

**DEPARTAMENTO DE ENGENHARIA ELECTROTÉCNICA
E DE COMPUTADORES**

Dissertação de Mestrado

**MODELIZAÇÃO PARAMÉTRICA
DE SINAIS PARA APLICAÇÃO EM
SISTEMAS DE CONVERSÃO**

TEXTO-FALA



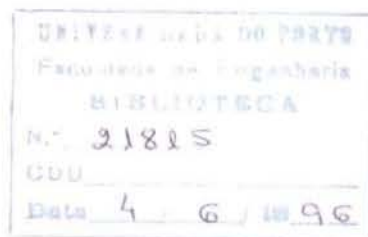
**ENGENHARIA ELECTROTÉCNICA E DE COMPUTADORES
ÁREA CIENTÍFICA DE INFORMÁTICA INDUSTRIAL**

João Paulo Ramos Teixeira

Porto
Outubro 1995

**MODELIZAÇÃO PARAMÉTRICA
DE SINAIS PARA APLICAÇÃO EM
SISTEMAS DE CONVERSÃO
TEXTO-FALA**

Tese realizada sob orientação do Professor Doutor
Diamantino Rui da Silva Freitas



Professor Auxiliar do Departamento de Engenharia
Electrotécnica e de Computadores da
Faculdade de Engenharia da Universidade do Porto

**Aos meus pais
e esposa**

RESUMO

Neste trabalho foi desenvolvido um sistema de extração automática de parâmetros de sinais de fala recorrendo a ferramentas de análise cepstral, de predição linear quer pela matriz autocorrelação quer pela matriz covariância, e ao método de análise síncrona com o período fundamental.

Realiza-se uma segmentação e classificação dos sinais em vocalizados, não vocalizados ou silêncio. Aos segmentos com conteúdo de fala atribuem-se modelos baseados em formantes.

Os parâmetros definidos pelo modelo para a fala vocalizada são 4 formantes e respectivas larguras de banda, frequência fundamental e amplitude. Para os sons não vocalizados considerou-se um modelo com um pólo, um zero e excitação com sinal de ruído aleatório.

O método de análise cepstral segmenta o sinal com comprimentos fixos e analisa individualmente cada segmento. A análise de cada segmento consiste na separação das características do trato vocal e da fonte excitadora, recorrendo a uma função de "lifteragem" nas quefrências sendo determinada a frequência fundamental da fonte excitadora ("pitch") e alisado o espectro relativo ao trato vocal. A partir deste espectro alisado ou envelope espectral é aplicado um algoritmo de determinação dos picos para extrair as frequências formantes das ressonâncias do trato vocal obedecendo a restrições respeitantes às regiões de frequências de cada formante e às amplitudes relativas dos respectivos picos. São também determinadas as correspondentes larguras de banda a 3 dB a partir do envelope espectral.

Os métodos de predição linear analisam também individualmente cada segmento de comprimento fixo do sinal de fala, obedecendo a um modelo só com pólos, determinando os coeficientes de predição linear por multiplicação matricial. A partir destes coeficientes são determinados os pólos. Cada par de pólos complexos conjugados é considerado um possível formante, sendo posteriormente seleccionados justamente 4 formantes por um processo de eliminação das frequências formantes que não têm a correspondência de um pico na função de transferência do sistema.

O método de análise síncrona com o período fundamental determina o sincronismo com o impulso glotal segmentando o sinal em troços de duração de um período, sendo estes posteriormente analisados pelo método de predição linear (matriz covariância).

Posteriormente a sequência de parâmetros é sujeita a um alisamento não linear para corrigir eventuais pontos fora de uma linha definida pelos valores dos parâmetros anteriores e posteriores ("outliers").

Todos estes métodos determinam com razoável fidelidade as frequências formantes dos sinais de fala, contudo, as larguras de banda são mais correctamente determinadas pelo método de predição linear pela matriz covariância.

É ainda apresentado o desenvolvimento de um conversor texto-fala para o português baseado num sintetizador de formantes com o mesmo modelo usado na análise para os sinais vocalizados. Os principais resultados obtidos foram a realização acústica de uma lista de 37 fonemas fundamentais, regras de conversão grafema-som na forma tabular, um grupo de regras de concatenação para as estruturas acústica e temporal inerentes aos sons, regras prosódicas elementares e, pronuncia de acrónimos e numerais.

Foram ainda desenvolvidas várias ferramentas complementares à análise dos sinais de fala como sejam um espectrógrafo e um outro sintetizador de formantes, exclusivamente computacional e para testes, baseado no modelo com os mesmos parâmetros.

Os métodos desenvolvidos foram testados com sinais de fala adequadamente seleccionada e recolhida em sala insonorizada e, registados magneticamente com aparelhagem adequada.

Os resultados atingidos satisfazem os objectivos inicialmente propostos para este trabalho.

Palavras-chave: processamento digital de sinais, análise de fala, síntese de fala, conversores texto-fala, modelização de sinais de fala, análise cepstral, predição linear.

ABSTRACT

It has developed in this work a system for automatic feature extraction of speech signal parameters using tools like cepstral analysis, linear prediction by autocorrelation and covariance matrix and the pitch synchronous analysis method.

A segmentation and classification of this signal in voiced, unvoiced or silence as done. The segments with speech content are considered as a formant based model.

The parameters defined by the model for voiced speech consist in a 4 formants and corresponding bandwidths, fundamental frequency or pitch and magnitude. For the unvoiced sounds was considered a model with one pole, one zero and white noise excitation source signal.

In the cepstral analysis method the signal are segmented in units of fixed duration to be individually analysed. The analysis of each segment consist in the separation of the vocal tract characteristics from the excitation source characteristics applying a lifter function to the quefrequency domain signal resulting in the smoothed spectrum of the vocal tract, meanwhile the pitch are resolved. To the smoothed spectrum or spectral envelop are applied an algorithm to extract the formant frequencies corresponding to the vocal tract resonance's obeying at a some constraints related to the frequency regions of each formant and magnitude of the picks. Also using the spectral envelop are resolved the corresponding bandwidths at 3 dB.

The LPC methods also individually analyze each segment of speech signal with fixed duration obeying a all poles model, determining the LPC coefficients by matricial multiplication. The poles are resolved using this coefficients. Each pair of conjugated complex poles is considered as a possible formant, being latter selected just 4 formants, by a process of eliminating formant frequencies which do not have the corresponding pick in the system transfer function.

The pitch synchronous analysis method adjust the synchronism with glotal pulse doing the segmentation of signal in frames of one period duration, to future analyze by the LPC method (covariance).

After the extraction, the parameters sequence are passed through a non linear smoothing in order to correct eventual outlier points.

All of this methods resolve with reasonable fidelity the formant frequencies of speech signals, although, the bandwidths are most correctly resolved by the LPC covariance method.

Also is presented the text-to-speech converter system for Portuguese development in a formant synthesizer using the same model that as used in voiced speech analysis. The principal results was an acoustic realization of a set of 37 fundamental phonemes, a grapheme-sound conversion rules in a tabular format, a set of concatenation rules for the inherent time and acoustic structures of sounds, elementary prosodic rules and, numerals and acronyms pronunciation.

Also have been developed complementary tools to speech analysis like a spectrograph and another formant synthesizer, exclusively computational for testing, based on same model.

The developed methods was tested with a selected speech signals recorded in an appropriated room and magnetically registered with adequate devices.

The achieved results corresponds to the initially proposed objectives of this work.

Keywords: digital signal processing, speech analysis, speech synthesis, text-to-speech systems, speech signals modeling, cepstral analysis, linear prediction.

SOMMAIRE

Dans le présent travail on a développé un système d'extraction automatique de paramètres des signaux de parole, par recours à des outils d'analyse par cepstre, par prédiction linéaire soit par la matrice autocorrélation soit par la matrice covariance et à la méthode d'analyse synchrone avec la période fondamentale.

Les signaux sont segmentés et classifiés selon les catégories de vocalisé, non vocalisé et silence. Aux segments avec un contenu de parole on confère des modèles basés en fréquences formantes.

Les paramètres définis par le modèle pour les segments vocalisés sont 4 formantes et respectives largeurs de bande, la fréquence fondamentale et l'amplitude. Pour les sons non-vocalisés on a pris un modèle avec un pôle, un zéro et une excitation avec un signal de bruit aléatoire.

La méthode d'analyse cepstrale segmente le signal avec des longueurs fixes et fait en suite l'analyse individuelle. L'analyse de chaque segment consiste dans la séparation des caractéristiques du trait vocal et de la source excitatrice, par recours à une fonction de "liffrage" dans les fréquences, calculant la fréquence fondamentale de la source excitatrice ("pitch") et lissant le spectre relatif au trait vocal. À partir de ce spectre lissé ou enveloppe spectrale on applique un processus pour la détermination des maxima de façon à extraire les fréquences formantes des résonances du trait vocal en obéissant à des constrictions relatives aux régions fréquentielles de chaque formante et aux amplitudes relatives des maxima respectifs. Les largeurs de bande à -3 dB correspondantes sont aussi déterminées d'après l'enveloppe spectrale.

Les méthodes de prédiction linéaire analysent aussi individuellement chacun des segments de longueur fixe du signal de parole, en obéissant à un modèle seulement avec des pôles, et en déterminant les coefficients de prédiction linéaire par produit matriciel. Après ces coefficients on calcule les pôles. Chaque paire de pôles complexes conjugués est considéré comme une possible formante, candidat à la sélection de justement 4 formantes par un processus d'élimination des fréquences formantes qui n'ont pas de correspondance à un pic dans la fonction de transfert du système.

La méthode d'analyse synchrone avec la période fondamentale détermine le synchronisme avec l'impuls glottal, en segmentant le signal en traits avec la durée d'une période, pour postérieure analyse par la méthode de prédiction linéaire (matrice covariance).

Postérieurement la séquence des paramètres est soumise à un lissage non-linéaire pour corriger les points éventuellement au dehors d'une ligne définie par les valeurs antérieures et postérieures des paramètres ("outliers").

Toutes ces méthodes sont capables de déterminer avec une raisonnable fidélité les fréquences formantes des signaux de parole, cependant, les largeurs de bande sont plus correctement déterminées par la méthode de prédiction linéaire par la matrice de covariance.

On présente en suite le développement d'un convertisseur texte-parole pour le portugais basé dans un synthétiseur de formantes pourvu du même modèle employé dans l'analyse des signaux vocalisés. Les principaux résultats obtenus sont la réalisation acoustique d'une liste de 37 phonèmes fondamentaux, règles de conversion graphème-son en forme tabulaire, un groupe de règles de concaténation pour les structures acoustique et temporelle des sons, des règles prosodiques élémentaires et prononciation d'acronymes et numéraux.

On a aussi développé plusieurs outils complémentaires à l'analyse des signaux de parole comme un spectrographe et un autre synthétiseur de formantes, exclusivement computationnel et pour testes, basé sur le modèle aux mêmes paramètres.

Les méthodes développées ont passé des testes avec des signaux de parole choisis et acquis dans une chambre insonorisée par enregistrement magnétique. Les résultats obtenus accomplissent les objectifs proposés au début pour ce travail.

Mots-clefs: Processement digital des signaux, analyse de la parole, synthèse de la parole, convertisseurs texte-parole, modélisation des signaux de parole, analyse cepstrale, prédiction linéaire.

AGRADECIMENTOS

Dirijo o meu primeiro agradecimento ao orientador deste trabalho, Prof. Diamantino Rui da Silva Freitas, pela forma cativante e amigável com que sempre se dispôs a discutir assuntos relacionados com esta dissertação, incentivando ao rigor científico e propondo novos desafios e ideias que resultaram em contribuições valiosas para este trabalho. Agradeço-lhe ainda o "empréstimo" da sua voz no processo de recolha dos sinais de fala.

Aos Prof. Espain Oliveira, Rui Araújo e à Anabela, ao Rui Pessegueiro e ao Manuel Cardoso agradeço as constantes trocas de impressões sobre matérias relacionadas com o tema da dissertação.

Aos Prof. Géza Németh e Doutor Gábor Olaszy agradeço o apoio e boas condições criadas no decurso do trabalho em Budapest, tendo resultado num forte estímulo para a prossecução deste.

Aos colegas e amigos Jorge Reis, Avelino Marques e Henrique Cabral entre outros, o meu muito obrigado pelo ambiente de boa disposição e entre ajuda que souberam criar.

Agradeço aos meus pais todo o apoio que sempre me deram até ao atingir desta etapa.

Por último mas não menos importante quero deixar uma palavra de carinho muito especial para a minha esposa pelo apoio, confiança e amor incondicionais que me dedicou fazendo com que este trabalho fosse possível.

O meu obrigado à RDP Porto pelas instalações e condições técnicas cedidas para a recolha de sinais de fala.

O desenvolvimento desta dissertação foi subsidiada com uma bolsa de Mestrado pelo programa PRAXIS XXI, facto pelo qual manifesto a minha gratidão.

Índice

1. INTRODUÇÃO	1
1.1 Objectivos e Enquadramento Deste Trabalho	2
1.2 Aspectos de Organização da Dissertação	3
2. CARACTERIZAÇÃO FONÉTICA DA LÍNGUA PORTUGUESA	5
2.1 Introdução	6
2.2 Produção da Fala	6
2.2.1 Fonte Sonora	8
2.2.2 Frequência Fundamental	8
2.2.3 Formantes	9
2.3 Fonemas Do Português	10
2.3.1 Caracterização de Alguns Traços Fonéticos	10
2.3.1.1 Traços de Sonoridade	10
2.3.1.2 Traços de Tonalidade	13
2.3.2 Parâmetros Classificatórios das Consoantes	13
2.3.2.1 Modo de Articulação	14
2.3.2.1.1 Consoantes Oclusivas	14
2.3.2.1.2 Consoantes Fricativas	14
2.3.2.1.3 Consoantes Laterais	15
2.3.2.1.4 Consoantes Vibrantes	15
2.3.2.1.5 Consoantes Africadas	15
2.3.2.2 Ponto de Articulação	16
2.3.3 Lista de Fonemas do Português	16
3. RECOLHA DO SINAL DE FALA	18
3.1 Introdução	19
3.2 Elementos Importantes na Aquisição do Sinal de Fala	20
3.2.1 Câmara Insonorizada	20
3.2.2 Microfone	20
3.2.3 Pré-amplificador	20
3.2.4 Filtro Anti-Aliasing	21
3.2.5 Frequência de Amostragem	21
3.2.6 Número de Bits	21
3.2.7 Falante	22
3.2.8 Texto	22
3.3 Condições em que Decorreu a Recolha/Aquisição dos Sinais de Fala Neste Trabalho	23
3.3.1 Câmara Insonorizada	23
3.3.2 Armazenamento do Sinal em Fita Magnética	24

3.3.3 Reprodução e Conversão Analógico/Digital com Recurso a Uma Placa de Som	24
3.3.3.1 Facilidades Oferecidas Pela Placa de Som	24
3.3.3.2 Frequência de Amostragem	24
3.3.3.3 Resolução	25
3.3.3.4 Filtro Anti-Aliasing	25
3.3.3.5 Armazenamento do Sinal	26
3.3.4 Texto Escolhido	26
3.3.5 Falantes	27
4. MODELIZAÇÃO DOS SINAIS DE FALA / CODIFICAÇÃO	
PARAMÉTRICA	28
4.1 Introdução	29
4.2 Modelos de Produção da Fala	30
4.2.1 Modelos de Tubos Sem Perdas	30
4.2.2 Modelo de Engenharia	31
4.2.2.1 Trato Vocal	32
4.2.2.2 Efeito de Radiação	34
4.2.2.3 Excitação	35
4.2.2.4 O Modelo Completo	37
4.2.3 Modelamento Sinusoidal da Fala	38
4.3 Parâmetros Para os Modelos dos Sinais de Fala	39
4.3.1 Parametrização por Formantes	39
4.4 Técnicas de Análise Para Obtenção dos Parâmetros dos Sinais de Fala Baseadas no Modelo de Formantes	40
4.4.1 Análise por Síntese	40
4.4.2 Análise Síncrona com o Período Fundamental	42
4.4.3 Análise Cepstral	43
4.4.4 Análise por LPC's	45
4.4.4.1 Princípios Básicos da Análise por Predição Linear	46
4.4.4.2 Método da Autocorrelação	50
4.4.4.3 Método da Covariância	52
4.5 Codificação Digital Com Qualidade de Telefonia da Forma de Onda da Voz	53
5. CONVERSORES TEXTO-FALA	55
5.1 Introdução	56
5.2 Sistemas de Conversão Texto-Fala	56
5.2.1 Processamento Linguístico	57
5.2.2 Processamento Acústico	59
5.3 O Conversor Texto-Fala MULTIVOX	60
5.3.1 Blocos Constituintes	61

5.3.2 A Conversão Grafema Fonema Para o Português	63
5.3.3 Regras de Concatenação de Fonemas	67
5.3.4 Regras de Prosódia	68
6. FERRAMENTAS USADAS/CRIADAS COMO SUPORTE À DETERMINAÇÃO AUTOMÁTICA DE PARÂMETROS	70
6.1 Introdução	71
6.2 Amplitude Média Deslizante	71
6.3 Energia Média Deslizante	74
6.4 Taxa de Passagem Por Zero	76
6.5 Classificação Quanto ao Modo de Excitação / Segmentação	78
6.6 Transformada de Fourier de Curta Duração "Short Time"	85
6.7 Chirp Z Transform	86
6.8 Cepstro	88
6.9 Espectrógrafo	91
6.10 Codificação por Predição Linear (LPC)	92
6.10.1 Matriz Autocorrelação	93
6.10.2 Matriz Covariância	93
6.11 Sintetizador	95
7. DETERMINAÇÃO AUTOMÁTICA DOS PARÂMETROS DOS SINAIS DE FALA	98
7.1 Introdução	99
7.2 Determinação da Frequência Fundamental	100
7.2.1 Introdução	100
7.2.2 Método do Cepstro	100
7.2.3 Determinação da Frequência Fundamental no Erro Residual de Predição Linear	101
7.2.4 Processamento no Domínio Temporal	102
7.3 Determinação Automática dos Parâmetros Para a Fala Vocalizada	108
7.3.1 Estrutura de Análise	108
7.3.1.1 Método de Análise Cepstral	110
7.3.1.2 Método da Predição Linear - Matriz Autocorrelação e Matriz Covariância	117
7.3.2 Análise Síncrona Com o Período Fundamental	123
7.3.3 Alisamento Não Linear Aplicado à Sequência de Parâmetros Estimados	127
7.4 Análise de Fala Não Vocalizada	129

8. TESTE E AVALIAÇÃO DOS RESULTADOS	131
8.1 Introdução	132
8.2 Conversor Texto-Fala para o Português	132
8.3 Determinação da Frequência Fundamental com Processamento no Domínio Temporal	133
8.4 Extração Automática de Parâmetros dos Sinais de Fala	133
8.4.1 Método de Análise Cepstral	135
8.4.2 Método de Predição Linear - Matriz Autocorrelação	138
8.4.3 Método de Predição Linear - Matriz Covariância	141
8.4.4 Método de Análise Síncrona com o Período Fundamental	144
8.4.5 Comparação de Resultados dos Diferentes Métodos de Análise	146
9. CONCLUSÕES E DESENVOLVIMENTOS FUTUROS	148
9.1 Conclusões	149
9.2 Desenvolvimentos Futuros	151
BIBLIOGRAFIA	153
ANEXO A	159
ANEXO B	174

Lista de Figuras

Capítulo 2

Figura 2.1 - Aparelho fonador humano	6
Figura 2.2 - Glote	7
Figura 2.3 - Modelo de Engenharia.	8
Figura 2.4 - Espectro alisado de um segmento de som da vogal [i]	9

Capítulo 3

Figura 3.1 - Sistema de Recolha	20
Figura 3.2 - Recolha do Sinal	23
Figura 3.3 - Amostragem do Sinal	23
Figura 3.4 -Função de Transferência medida do filtro anti-aliasing da placa de som SOUND BLASTER	25
Figura 3.5 - Função de Transferência do filtro passa-baixo usado na decimação de 2:1	26

Capítulo 4

Figura 4.1 - Sinal de fala correspondente à locução de "leitura de um parágrafo"	29
Figura 4.2 - Concatenação de 5 tubos acústicos sem perdas	30
Figura 4.3 - Modelo de engenharia	31
Figura 4.4 - Diagrama de blocos representando: a) modelo de tubos sem perdas; b) modelo de engenharia	31
Figura 4.5 - Implementação directa da função de transferência do modelo só com pólos	33
Figura 4.6 - Implementação em cascata de factores de 2ª ordem da função de transferência do modelo só com pólos ($G_k = 1 - 2 z_k \cos\theta_k + z_k ^2$)	34
Figura 4.7 Modelo de engenharia incluindo os efeitos de radiação	34
Figura 4.8 - Impulso glotal humano	35
Figura 4.9 - Gerador do sinal de excitação para a fala vocalizada	35
Figura 4.10 - Forma de onda do impulso glotal sintético $G(z) = \frac{-ae \ln(a)z^{-1}}{(1 - az^{-1})^2}$ com $a=0,90$	36
Figura 4.11 - Espectro do impulso glotal sintético $G(z) = \frac{-ae \ln(a)z^{-1}}{(1 - az^{-1})^2}$ com $a = 0,90$	37

Figura 4.12 - Modelo genérico para a produção da fala	37
Figura 4.13 - Sistema de análise por síntese	41
Figura 4.14 - Modelo simples de produção da fala no domínio temporal	43
Figura 4.15 - Diagrama de fluxo da estimação do cepstro	44
Figura 4.16 - Cepstro de um segmento de fala da vogal [a]	44
Figura 4.17 - Espectro e envelope espectral (pelo método do cepstro) de um segmento de fala da vogal [a]	45
Figura 4.18 - Diagrama de blocos do modelo simplificado de produção da fala	46

Capítulo 5

Figura 5.1 - Diagrama de blocos genérico de um sistema de conversão texto-fala	57
Figura 5.2 - Diferentes tarefas do processamento linguístico	58
Figura 5.3 - Diagrama de blocos do processamento acústico	59
Figura 5.4 - Módulos constituintes do conversor MULTIVOX	62

Capítulo 6

Figura 6.1 - Aplicação da amplitude média a um sinal de fala vocalizado	72
Figura 6.2 - Aplicação da amplitude média deslizante	73
Figura 6.3 - Aplicação da energia média deslizante a um sinal de fala vocalizada	75
Figura 6.4 - Aplicação da energia média	76
Figura 6.5 - Aplicação da taxa de passagem por zero	77
Figura 6.6 - Algoritmo para classificação de semi-segmentos de sinais de fala	80
Figura 6.7 - Aplicação da TPZ e energia média deslizante	81
Figura 6.8 - "Domínio de decisão" para classificar sons baseado na energia e taxa de passagem por zero	82
Figura 6.9 - Classificação do sinal "isto"	83
Figura 6.10 - Classificação do sinal "pato"	84
Figura 6.11 - Peso de cada uma das amostras com a aplicação de 50% de sobreposição às janelas a) Hanning, b) Hamming	86
Figura 6.12 - Exemplo do "zoom" realizado pelo CZT	87
Figura 6.13 - Função da janela $l(n)$ usada no alisamento espectral pelo método do cepstro	90
Figura 6.14 - Espectrograma de banda larga do sinal de fala "ama" obtido com alisamento espectral e a função <i>espectro()</i>	92
Figura 6.15 - Artificio de cálculo para a determinação da matriz covariância	94
Figura 6.16 - Sequência de segmentos de impulsos glotais gerados pela função <i>fgerimp()</i>	96

Capítulo 7

Figura 7.1 - Variação da frequência fundamental	101
Figura 7.2 - Modelo de predição linear só com pólos	101
Figura 7.3 - Diagrama de blocos do processamento paralelo no domínio temporal do detector de frequência fundamental	103
Figura 7.4 - As seis sequências m_1 a m_6 de impulsos obtidas a partir do sinal $[i]$ alisado com uma janela de média de 80 amostras e espaçamento unitário	104, 105
Figura 7.5 - Operação básica de cada estimador individual do período fundamental no domínio temporal	106
Figura 7.6 - Frequência fundamental estimada pelos seis estimadores para o sinal $[i]$	107
Figura 7.7 - Fluxograma da estrutura de análise para o modelo de sinais vocalizados	109
Figura 7.8 - Diagrama de blocos para determinação da frequência fundamental e envelope espectral	110
Figura 7.9 - Função de equalização das amplitudes dos formantes	111
Figura 7.10 - Lugares das frequências dos 4 formantes	112
Figura 7.11 - a) Modelo digital para a fala vocalizada b) Diagrama detalhado do modelo de trato vocal	113
Figura 7.12 - Curva limite da relação das amplitudes logarítmicas entre F2 e F1	113
Figura 7.13 - Algoritmo usado para extrair os formantes e larguras de banda do envelope espectral	114, 115
Figura 7.14 - Espectro alisado pelo método do cepstro de um segmento da vogal [a]	117
Figura 7.15 - Fluxograma do processamento de extracção automática dos formantes e larguras de banda pelo método de predição linear - matriz autocorrelação e matriz covariância	118, 119
Figura 7.16 - Função de transferência do trato vocal obtida pelo método da matriz autocorrelação para um segmento da vogal [a]	122
Figura 7.17 - Função de transferência do trato vocal obtida pelo método da matriz covariância para um segmento da vogal [a]	123
Figura 7.18 - Detecção de sincronismo realizada com a função <i>fdetsinc()</i> aplicada ao sinal [a]	124
Figura 7.19 - Fluxograma do algoritmo desenvolvido para extracção automática dos parâmetros do modelo de fala vocalizada por um processo de análise síncrona com o período fundamental	126
Figura 7.20 - Função de transferência do trato vocal para um segmento de fala da vogal [a] obtida pelo algoritmo descrito para análise síncrona	127

Figura 7.21 - Situações de parâmetros fora de uma linha consideradas como incorrectas 128

Figura 7.22 - Espectro e espectro alisado pelo método do cepstro de um segmento de fala do som do fonema [j] retirado da palavra "isto" 130

Capítulo 8

Figura 8.1 - Sinais usados na análise 133, 134

Figura 8.2 - Parâmetros extraídos automaticamente pelo método de análise cepstral para o sinal [i] 135

Figura 8.3 - Parâmetros extraídos automaticamente pelo método de análise cepstral para o sinal "ama" 136

Figura 8.4 - Parâmetros extraídos automaticamente pelo método de análise cepstral para o sinal sintetizado 137

Figura 8.5 - Parâmetros extraídos automaticamente pelo método de LPC - matriz autocorrelação para o sinal [i] 138, 139

Figura 8.6 - Parâmetros extraídos automaticamente pelo método de LPC - matriz autocorrelação para o sinal "ama" 139, 140

Figura 8.7 - Parâmetros extraídos automaticamente pelo método de LPC - matriz autocorrelação para o sinal sintetizado 140

Figura 8.8 - Parâmetros extraídos automaticamente pelo método de LPC - matriz covariância para o sinal [i] 141, 142

Figura 8.9 - Parâmetros extraídos automaticamente pelo método de LPC - matriz covariância para o sinal "ama" 142

Figura 8.10 - Parâmetros extraídos automaticamente pelo método de LPC - matriz covariância para o sinal sintetizado 143

Figura 8.11 - Parâmetros extraídos automaticamente pelo método de análise síncrona (LPC - matriz covariância) para o sinal [i] 144

Figura 8.12 - Parâmetros extraídos automaticamente pelo método de análise síncrona (LPC - matriz covariância) para o sinal "ama" 144, 145

Figura 8.13 - Parâmetros extraídos automaticamente pelo método de análise síncrona (LPC - matriz covariância) para o sinal sintetizado 145

Lista de Tabelas

Capítulo 2

Tabela 2.1 - Traços de sonoridade	10
Tabela 2.2 - Traços de tonalidade	13
Tabela 2.3 - Oclusivas orais para o português	14
Tabela 2.4 - Oclusivas nasais para o português	14
Tabela 2.5 - Fricativas para o Português	15
Tabela 2.6 - Laterais para o português	15
Tabela 2.7 - Vibrantes para o português	15
Tabela 2.8 - Ponto de articulação	16

Capítulo 5

Tabela 5.1 - Símbolos fonéticos e respectivos códigos usados no conversor texto-fala MULTIVOX (versão portuguesa)	64
Tabela 5.2 - Função prosódica relativa a cada código de marca no nível 1 de representação do conversor texto-fala MULTIVOX	66

Capítulo 7

Tabela 7.1 - Parâmetros estimados para um segmento da vogal [a] pelo método de predição linear, matrizes autocorrelação e covariância	123
Tabela 7.2 - ΔT usado para realizar o alisamento não linear	129

Lista de Abreviaturas e Símbolos

ABU - Acoustic Building Unit

ADM - Adaptative Delta Modulation

ADPCM - Adaptative Differential Pulse Code Modulation

APCM - Adaptive Pulse Code Modulation

CZT - Chirp Z Transform

DM - Delta Modulation

DPCM - Differential Pulse Code Modulation

e - número de elementos abaixo de LSES

F0 - Frequência Fundamental

FFT - First Fourier Transform

F_i - formante de ordem i

F_{imax} - limite superior da região de frequências para o formante F_i

F_{imin} - limite inferior da região de frequências para o formante F_i

F_n - Frequência de Nyquist

F_s - Frequência de amostragem

$G(z)$ - Função de transferência do impulso glotal

IDFT - Inverse Discrete Fourier Transform

LIZ - Limite Inferior da taxa de passagem por Zero

LSES - Limite Superior da Energia do Sinal

PCM - Pulse Code Modulation

$R(z)$ - Função de transferência para o efeito de radiação nos lábios

RDP - Rádio Difusão Portuguesa

TPZ ou tpz - Taxa de Passagem por Zero

$V(z)$ - Função de transferência do trato vocal

z - Número de elementos acima de LIZ