

UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

CODIFICADORES PARAMÉTRICOS DE SINAIS DE VOZ
COM EXCITAÇÃO MULTI-PULSO

DISSERTAÇÃO SUBMETIDA À UNIVERSIDADE FEDERAL DE SANTA CATARINA
PARA OBTENÇÃO DE GRAU DE MESTRE EM ENGENHARIA ELÉTRICA

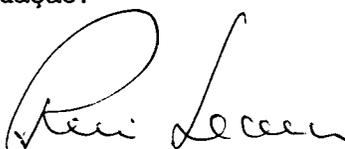
RICARDO FELIPE CUSTÓDIO

FLORIANÓPOLIS, NOVEMBRO - 1990.

CODIFICADORES PARAMÉTRICOS DE SINAIS DE VOZ
COM EXCITAÇÃO MULTI-PULSO

CANDIDATO: RICARDO FELIPE CUSTÓDIO

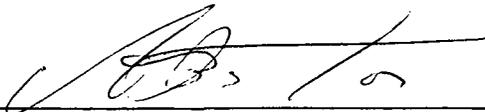
Esta dissertação foi julgada para obtenção do título - MESTRE
EM ENGENHARIA - Especialidade Engenharia Elétrica aprovada em sua forma final
pelo Curso de Pós-Graduação.



Prof. Rui Seara Dr. Ing.
ORIENTADOR



Prof. Sidnei Noceti Filho D. Sc.
CO-ORIENTADOR

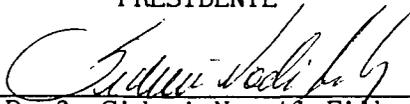


Prof. João Pedro Assumpção Bastos Dr. D'Etat
COORDENADOR DO CURSO DE PÓS GRADUAÇÃO EM
ENGENHARIA ELÉTRICA

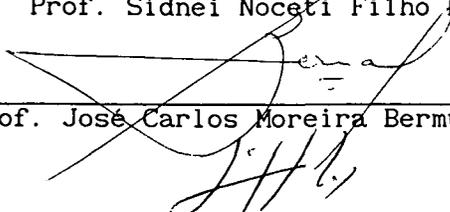
BANCA EXAMINADORA:



Prof. Rui Seara Dr. Ing.
PRESIDENTE



Prof. Sidnei Noceti Filho D. Sc.



Prof. José Carlos Moreira Bermudez Ph. D.

Prof. Luiz Pereira Calôba Dr. Ing.



Prof. Giles Lothar Istre Ph. D.

À minha esposa, Mari.

Aos meus pais, avós e irmãos.

AGRADECIMENTOS

Aos Professores Rui Seara e Sidnei Noceti Filho pela valiosa ajuda prestada na orientação deste trabalho.

À CAPES pelo apoio financeiro.

A todos que colaboraram para tornar possível a realização deste trabalho, e em particular à:

Mari Ângela Pauli Custódio

Irvando Luiz Speranzini

SUMÁRIO

LISTA DE FIGURAS.....	viii
LISTA DE TABELAS.....	ix
RESUMO.....	x
ABSTRACT.....	xi
<u>CAPÍTULO 1</u> - INTRODUÇÃO.....	01
<u>CAPÍTULO 2</u> - PREDIÇÃO LINEAR APLICADA A SINAIS DE VOZ - REVISÃO BIBLIOGRÁFICA	10
2.1 - Introdução.....	10
2.2 - Predição Linear de Sinais de Voz	10
2.3 - Determinação do Preditor Linear.....	12
2.3.1 - Método de Autocorrelação.....	15
2.3.2 - Método da Covariância.....	16
2.4 - Comparação entre os Métodos para Solução das Equações Geradas no Projeto do PCA-AR.....	17
2.5 - Modelos do Gerador de Excitação.....	19
2.6 - Projeto do Gerador de Excitação Multi-Pulso.....	22
2.7 - Projeto do PLA para Excitação Multi-Pulso.....	30
2.8 - Conclusão.....	34
<u>CAPÍTULO 3</u> - CODIFICADORES MULTI-PULSO COM PCA - ARMA.....	35
3.1 - Introdução.....	35

3.2 - Estudos e Considerações para a Determinação do Gerador de Excitação com PCA - ARMA.....	35
3.3 - Nova Técnica para Determinação das Localizações dos Impulsos do Gerador de Excitação Multi-Pulso.....	40
3.4 - Resultados.....	44
3.4.1 - Parâmetros para as Simulações.....	45
3.4.2 - Resultados Obtidos com um filtro PCA - AR.....	46
3.4.3 - Resultados Obtidos com um filtro PCA - ARMA.....	50
3.4.4 - Resultados Obtidos com um Filtro PCA - AR, Utilizando-se a Nova Proposta na Determinação da Localização dos Impulsos.....	56
3.5 - Conclusões.....	62
<u>CAPÍTULO 4</u> - CONCLUSÕES.....	65
REFERÊNCIAS BIBLIOGRÁFICAS.....	68
BIBLIOGRAFIAS CONSULTADAS.....	73
<u>APÊNDICE 1</u> - ALGORÍTMO DE LEVINSON-DURBIN PARA SOLUÇÃO DO SISTEMA DE EQUAÇÕES GERADAS PELO MÉTODO DA AUTOCORRELAÇÃO.....	78
<u>APÊNDICE 2</u> - PROGRAMA PARA DETERMINAÇÃO SUB-ÓTIMA DO GERADOR DE EXCITAÇÃO MULTI-PULSO.....	81
<u>APÊNDICE 3</u> - PROGRAMA PARA DETERMINAÇÃO SUB-ÓTIMA DO GERADOR DE EXCITAÇÃO MULTI-PULSO COM O NOVO ALGORÍTMO DE DETERMINAÇÃO DAS LOCALIZAÇÕES DOS IMPULSOS.....	85

LISTA DE FIGURAS

1.1 - Representações do Sinal de Voz.....	04
1.2 - Comparação das Diferentes Formas de Representações da Voz Humana em Termos de Taxa de Bits.....	05
2.1 - Diagrama de Blocos da Representação do Erro de Predição.....	11
2.2 - Codificadores Paramétricos de Sinais de Voz.....	11
2.3 - Modelo de Análise para Preditores na Forma Transversal.....	14
2.4 - Modelo Simplificado de Excitação dos Codificadores Paramétricos.....	20
2.5 - Diagrama de Blocos da Etapa de Análise do Codificador Multi-Pulso.....	23
2.6 - Determinação do Sinal Erro Ponderado.....	25
2.7 - Codificador Multi-Pulso com PLA.....	31
3.1 - Modelo de Análise para Preditores Transversais com Filtro de Predição do Tipo ARMA.....	36
3.2 - Modelo de Análise para a Determinação da Excitação Multi-Pulso com um Filtro Preditor do Tipo ARMA.....	37
3.3 - Determinação do Erro Ponderado.....	38
3.4 - Curva Típica da Função de Correlação Cruzada Entre o Sinal de Voz Original e a Resposta ao Impulso do Filtro Preditor para um Quadro de Voz.....	42

LISTA DE TABELAS

1.1 - Codificadores Padrões de Voz em Telefonia.....	06
1.2 - Aplicação do Processamento Digital da Voz em Diversas Áreas.....	08
2.1 - Comparação dos Métodos de Solução das Equações Geradas no Projeto dos Preditores PLA e PCA.....	18
3.1 - Estatística da Ordem do Preditor para $OD \leq 10$	47
3.2 - Estatística da Ordem do Preditor para $OD \leq 10$ com $RSR \geq RSR$ para PCA-AR com $OD=10$	48
3.3 - Estatística da Ordem do Preditor para $OD \leq 13$	49
3.4 - Estatística da Ordem do Preditor para $ON + OD \leq 10$	51
3.5 - Estatística da Ordem do Preditor para $ON + OD \leq 12$	53
3.6 - Estatística da Ordem do Preditor para $ON + OD \leq 10$ com $RSR \geq RSR$ do PCA - AR com $OD=10$	55
3.7 - Estatística da Ordem do Preditor para $OD \leq 10$, com $N_f=3$ e $N_p=60$	57
3.8 - Estatística da Ordem do Preditor para $OD \leq 10$ com $RSR \geq RSR$ do PCA-AR com $OD=10$, $N_f=3$ e $N_p=60$	58
3.9 - Estatística da Ordem do Preditor para $OD \leq 13$, $N_f=3$ e $N_p=60$	59
3.10 - Estatística da Ordem do Preditor para $OD \leq 10$, $N_f=2$ e $N_p=80$	60
3.11 - Estatística da Ordem do Preditor para $OD \leq 10$, $N_f=2$ e $N_p=80$ com $RSR \geq RSR$ para $OD=10$	61
3.12 - Estatística da Ordem do Preditor para $OD \leq 13$, $N_f=2$ e $N_p=80$	62
3.13 - Resultados do Método para a Determinação das Localizações Proposto por SINGHAL & ATAL[1].....	63
3.14 - Resultados do Novo Método para a Determinação das Localizações com $N_f=3$ e $N_p=60$	63
3.15 - Resultados do Novo Método para a Determinação das Localizações com $N_f=2$ e $N_p=80$	64

R E S U M O

Neste trabalho é apresentado um estudo e algumas considerações a respeito da implementação de codificadores de sinais de voz com predição linear e excitação multi-pulso ("MPELPC").

Para o projeto do preditor é proposta a utilização do "melhor" filtro preditor para a excitação multi-pulso, ao invés da utilização do clássico filtro só pólos. É mostrado, através de tabelas contendo resultados de diversas simulações que a utilização do "melhor" filtro preditor melhora a relação sinal-ruído da voz gerada.

Para a determinação da excitação é proposta uma nova técnica com menor complexidade computacional do que a proposta por SINGHAL & ATAL[1]. São mostrados os resultados obtidos nas diversas simulações comparando os dois métodos.

A B S T R A C T

This work presents a study and some considerations about the implementation of speech signal coders using linear prediction and multi-pulse (MPELPC) excitation.

The employment of a "better" predictor filter for the multi-pulse excitation is proposed. It is shown through various simulation results that using the "better" predictor filter improves the signal-to-noise ratio of the speech signal generated.

A new technique is proposed for the determination of the excitation. This technique presents a reduced computational complexity when compared with that proposed by SINGHAL AND ATAL[1]. Many simulation results are shown along with a comparison of both methods.

CAPÍTULO 1

INTRODUÇÃO

A primeira máquina destinada a realizar a codificação da voz com objetivo de reduzir a taxa para fins de transmissão foi concebida em 1936 pelo americano Homer Dudley[25]. Foi o primeiro também a utilizar o termo "vocoder", aglutinação do termo "voice coder", do inglês codificador de voz. Porém, os primeiros "vocoders" comerciais só passaram a ser utilizados na década de 60. Os primeiros codificadores tinham como desvantagem suas grandes dimensões, que dificultavam sua utilização em muitas aplicações.

Com o acelerado desenvolvimento da microeletrônica, foi se tornando viável a construção de "vocoders" com dimensões cada vez menores e mais eficientes.

A maior dificuldade na construção desses equipamentos era a necessidade de um grande número de filtros, com o objetivo de simular todas as nuances do carácter tonal da voz humana. A quantidade desses filtros, por razões de volume do "hardware", era limitada, prejudicando a qualidade da voz gerada. Outra significativa dificuldade era a necessidade de uma enorme quantidade de memória para o armazenamento da voz.

Com o advento da eletrônica digital associado ao grande desenvolvimento das técnicas de processamento digital de sinais, circuitos digitais passaram a ser incorporados aos "vocoders", substituindo, com vantagem, diversas partes analógicas destes. Com isso, os "vocoders" passaram a ter:

- O computador como ferramenta de auxílio à análise, ao projeto e à implementação;
- Diminuição do número de filtros analógicos necessários, em razão da substituição destes por filtros digitais variantes no tempo;
- Diminuição do tamanho físico;
- Custo total do hardware menor;
- Melhor qualidade da voz gerada.

A principal característica do "vocoder" é a sua capacidade de analisar e sintetizar, por meios eletrônicos, o som da voz humana. A análise é realizada através da caracterização da voz em termos do sinal portador da informação. Os sons são formados na realidade pela composição de um conjunto de frequências diferentes. Os sons ditos vozeados, são os que possuem uma característica em frequência mais bem definida, e os não vozeados, são os que contém uma quantidade enorme de frequências, caracterizando-se por serem mais ruidosos.

A ciência que estuda os sons da voz, como são produzidos pelo homem, as vibrações propagadas e captadas pelo ouvinte e como os ouvintes percebem esses sons é denominada fonética.

No caso dos sons vozeados, o ar vindo dos pulmões, é forçado a passar pelas cordas vocais situadas na traquéia, fazendo-as vibrar e conseqüentemente, excitando as cavidades bucal e nasal com uma coluna de ar pulsante. São exemplos desses sons os fonemas relativos às vogais.

No caso dos sons não vozeados, o ar vindo dos pulmões, ao contrário, não é forçado através das cordas vocais, mas simplesmente expelido pela boca. São exemplos desses sons os fonemas relativos ao "f" e o "s".

Tanto nos sons vozeados, quanto nos não vozeados, o formato das cavidades bucal e nasal é que determina o timbre dos sons emitidos. Por outro lado, as variações das ressonâncias dessas cavidades, pelo movimento da língua e dos lábios, controla o conteúdo harmônico da voz. Os lábios, por exemplo, desempenham um papel importante nos sons que são individualizados pelas suas características de amplitude dinâmica, tal como no transiente percursivo e explosivo do "p" em "papel".

Dessa forma, o aparelho fonador humano pode ser encarado como um complexo instrumento gerador de sons, composto por um oscilador controlado em frequência e amplitude, um gerador de ruído e uma série de filtros de tonalidade.

A codificação visa aproveitar as redundâncias do sinal de voz, para compactar sua representação. Este é um dos principais motivos da importância das pesquisas em processamento de sinais para o estudo de técnicas e algoritmos de representação do sinal de voz.

Pode-se classificar estas técnicas em três principais grupos como mostra a Figura 1.1.

As técnicas de codificação utilizadas para os "vocoders" enquadram-se na representação paramétrica[2].

O princípio da representação pela forma de onda é a amostragem em intervalos fixos da amplitude do sinal de voz, quantização e a compactação dos dados. São exemplos desta representação a modulação por codificação de pulsos (PCM), a modulação por codificação de pulsos diferencial (DPCM), a modulação delta adaptável (ADM), a modulação por codificação de pulsos diferencial adaptável (ADPCM), entre outras técnicas.

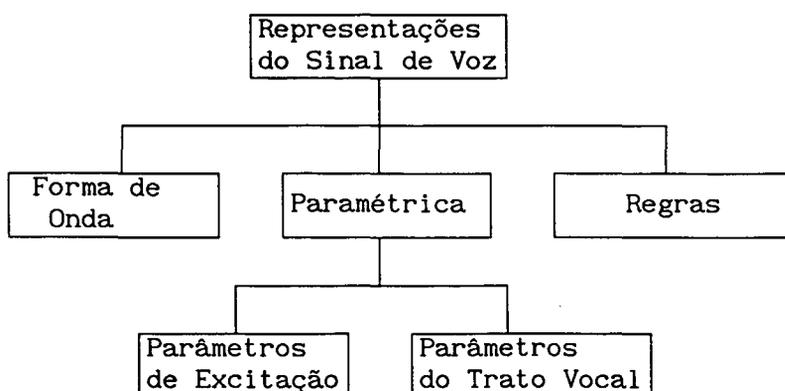


FIGURA 1.1 - Representações do Sinal de Voz.

A representação paramétrica é caracterizada pela obtenção de parâmetros da forma de onda do sinal de voz com o objetivo de síntese. A codificação por predição linear é um típico exemplo.

Na representação por regras, grupos de fonemas representados por pequenas quantidades de dados são justapostos para reproduzir qualquer som de voz desejado. Esta representação requer que se considere a entonação, a acentuação e a duração dos sons para se obter uma voz "natural". Estes requisitos, no entanto, ainda merecem consideráveis estudos, de um lado devido à grande quantidade de idiomas diferentes existentes e do outro, devido à grande variação dos sons existentes em um mesmo idioma.

A escolha de uma destas representações para a codificação da voz dependerá da qualidade e da complexidade requerida. Uma maior complexidade do codificador implica num maior esforço computacional para implementá-lo.

A Figura 1.2 mostra uma comparação das representações supracitadas em termos de taxa de bits do sinal de voz codificado.

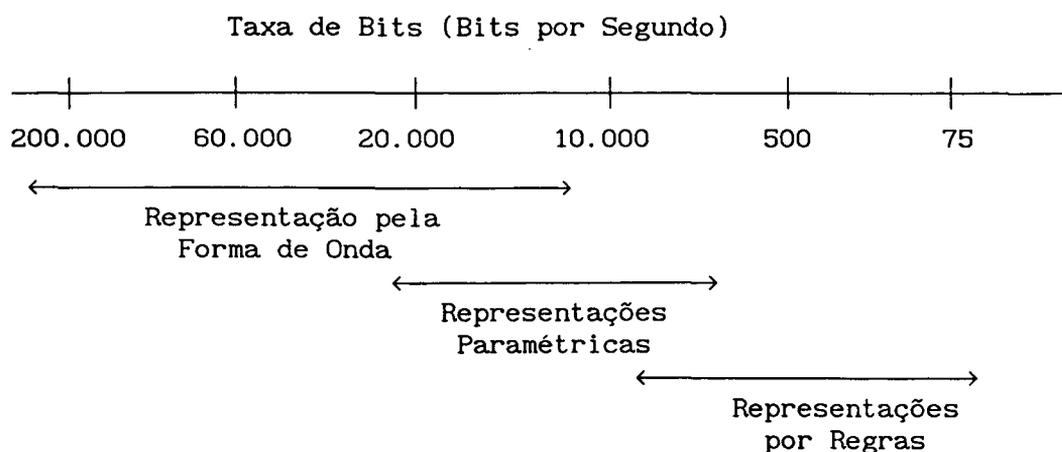


FIGURA 1.2 - Comparação das Diferentes Formas de Representações da Voz Humana em Termos de Taxa de Bits

A Tabela 1.1 apresenta as técnicas de codificação de voz em termos de padrões, taxa de bits, aplicações típicas e qualidade da voz decodificada.

Nesta Tabela, a qualidade da voz é medida de forma subjetiva, utilizando-se para tal a escala MOS ("Mean Opinion Score") que varia de 1 a 5 onde 5 significa excelente, 4 boa, 3 razoável, 2 ruim e 1 inaceitável[2].

Resumindo, o uso destas técnicas especiais na codificação da voz tem os seguintes propósitos:

- Reduzir a quantidade de memória de massa para armazenar a voz;
- Promover uma redução na largura de banda do sinal codificado em relação a do sinal de voz original;
- Melhorar a relação sinal ruído do sinal de voz quando este apresentar alguma forma de degradação;
- Compartilhar um mesmo canal de comunicação para transmissão de outros sinais durante os períodos de silêncio na transmissão de voz em tempo real.

PADRÕES	CCITT	CCITT	CCITT	GSM	CTIA
DATA	1972	1984	1991 *	1988	1989
TAXA DE BITS(kb/s)	64	32	16	8	
APLICAÇÕES	COMUNICAÇÕES			RÁDIO CELULAR	
QUALIDADE	4.0 a 4.5			3.5 a 4.0	

CCITT: International Consultative Committee for Telephone and Telegraph

GSM: Group Speciale Mobile

CTIA: Cellular Technology Industry Association

* em fase de padronização

TABELA 1.1 - Codificadores Padrões de Voz em Telefonia[2].

Em termos de aplicações, estas técnicas de codificação da voz podem ser utilizadas em várias situações de forma a aproveitar melhor os meios de armazenamento, transmissão e reprodução do sinal de voz.

Os equipamentos ditos falantes convencionais, como por exemplo os toca-discos e os toca-fitas, que reproduzem a voz através de sistemas de armazenamento analógicos (discos e fitas convencionais), podem ser totalmente ou em parte, substituídos por compactos sistemas digitais. Como vantagens, estes apresentariam: um rápido acesso a partes da voz gravada e uma maior vida útil devido à ausência de partes mecânicas, comuns nos sistemas analógicos. Em relação ao tipo de armazenamento, a degradação com o envelhecimento e o uso seria mínima ou praticamente inexistente.

Também é possível utilizar estas técnicas na construção de sistemas de identificação e verificação de locutores. As técnicas de identificação e verificação de locutores envolvem a autenticação ou identificação de um locutor num conjunto de possíveis locutores.

Em muitas situações os sinais de voz são de tal forma degradados que chegam a comprometer sua inteligibilidade. Nestes casos, as técnicas de processamento de sinais podem ser aplicadas para melhorar a qualidade da voz.

A Tabela 1.2 mostra um resumo de algumas possíveis aplicações do processamento de voz.

Dentre as principais técnicas de codificação de voz, a representação paramétrica destaca-se, principalmente, pelo excelente compromisso apresentado entre a complexidade de implementação, taxa de bits de codificação e qualidade do sinal de voz sintetizado.

Uma das técnicas que tem encontrado uma considerável aceitação na atualidade é a codificação paramétrica com excitação multi-pulso.

Os codificadores paramétricos com excitação multi-pulso, inicialmente propostos por ATAL & REMDE [6,7], sintetizam a voz com alta qualidade, razão pela qual, nos últimos anos, um grande esforço de pesquisa tem sido dedicado a estes codificadores[1,6,7]. Em relação à complexidade computacional com respeito à determinação da excitação para estes codificadores, diversos trabalhos de pesquisa têm sido publicados, destacando-se os codificadores com excitação multi-pulso regularmente espaçados[14], os codificadores com excitação residual[13] e os codificadores com excitação multi-pulso com otimização das amplitudes dos pulsos[1]. Trabalhos recentes têm utilizado a excitação multi-pulso nos codificadores CELP ("Code Excited Linear Prediction")[8-11].

Área de aplicação	Forma de aplicação
Militar	<ul style="list-style-type: none"> - Operação de equipamento militar - Sistemas de avisos - Ajuda em operações e serviços - Simulação e treinamento. - Segurança em comunicação
Computador	Interface entre o homem e a máquina
Consumidor	<ul style="list-style-type: none"> - Dispositivos falantes - Dispositivos de ensino - Brinquedos - Dispositivos para cegos
Telecomunicações	<ul style="list-style-type: none"> - Síntese de voz para mensagens telefônicas - Menor largura de banda para o sinal de voz codificado visando a reduzir os custos de comunicação - Correio eletrônico falante
Indústria	<ul style="list-style-type: none"> - Instrumentos falantes - Alarmes - Controle de processos industriais - Avisos em locais públicos

TABELA 1.2 - Aplicação do Processamento Digital da Voz em Diversas Áreas.

Assim, para esta dissertação, propõe-se estudar os codificadores paramétricos de sinais de voz com excitação multi-pulso ("MPELPC").

Como resultado deste estudo, é proposto um melhoramento na estrutura do PCA (Preditor de Curto Atraso) que levou a um aumento na relação sinal ruído do sinal de voz sintetizado, como também uma técnica alternativa para a determinação da excitação multi-pulso, a qual exhibe uma menor complexidade computacional em relação à proposta na referência [1].

No Capítulo 2 é apresentada uma revisão bibliográfica dos codificadores paramétricos de sinais de voz onde são discutidos os diversos modelos de excitação utilizados nestes codificadores. São revistas as principais técnicas de projeto do PCA-AR (Preditor de Curto Atraso-"AutoRegressive") e é apresentado um algoritmo para determinação sub-ótima da excitação multi-pulso.

No Capítulo 3 é refeito o equacionamento do algoritmo descrito na referência [1], considerando-se também zeros na função de transferência do filtro PCA, ao invés de unicamente pólos. Diversas tabelas com os resultados das simulações que comparam o tradicional PCA-AR com o PCA-ARMA (Preditor de Curto Atraso-"AutoRegressive Moving Average") são apresentadas. É proposta também uma nova técnica para a determinação da excitação multi-pulso e são apresentados os resultados obtidos, em termos da relação sinal-ruído e da complexidade computacional, com a aplicação desta.

No Apêndice 1, apresenta-se o algoritmo de LEVINSON-DURBIN[3,16-18] utilizado no projeto do PCA-AR.

No Apêndice 2, apresenta-se um programa com a implementação do algoritmo de SINGHAL & ATAL[1], utilizado na determinação da excitação multi-pulso.

No Apêndice 3, apresenta-se um programa com a implementação da nova técnica proposta para a determinação das localizações dos impulsos da excitação multi-pulso.

C A P Í T U L O 2

PREDIÇÃO LINEAR APLICADA A SINAIS DE VOZ - REVISÃO BIBLIOGRÁFICA

2.1. INTRODUÇÃO

Este capítulo apresenta uma revisão bibliográfica das principais técnicas paramétricas utilizando-se predição linear para síntese de sinais de voz. São discutidos os principais modelos do gerador de excitação e do preditor linear.

São apresentadas as técnicas de determinação da excitação multi-pulso e as técnicas de projeto do filtro preditor AR ("AutoRegressive").

Finalmente é apresentada uma recente configuração do codificador, que consiste em inserir um preditor de longo atraso entre a excitação propriamente dita e o preditor AR com a finalidade de aproveitar a forte correlação que existe entre amostras dos sinais de voz, espaçadas por um intervalo da ordem de grandeza do período de "PITCH".

2.2. PREDIÇÃO LINEAR DE SINAIS DE VOZ

A Figura 2.1 apresenta um diagrama de uma estrutura de predição linear que é intensamente utilizada para estimação de parâmetros em sinais de voz. Nesta Figura, $s(n)$ e $\tilde{s}(n)$ representam os sinais de voz original e sua predição, respectivamente, e $e(n)$ o erro instantâneo de predição.

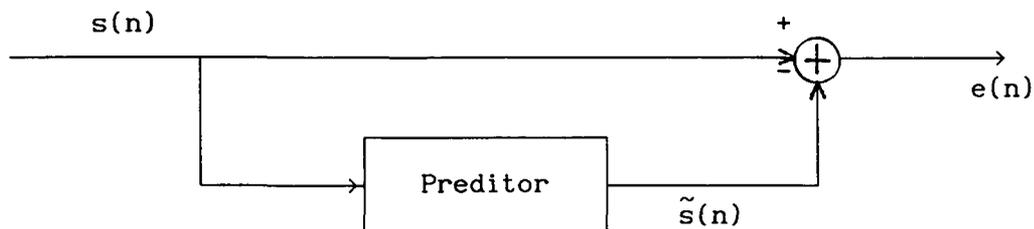


FIGURA 2.1 - Diagrama de Blocos da Representação do Erro de Predição.

É sabido que para pequenos segmentos, doravante denominados quadros de voz, os sinais de voz podem ser considerados praticamente estacionários[3]. Esta consideração permite que se determine um conjunto de parâmetros para cada um destes segmentos cuja quantidade de bits necessária para representá-los é menor que, quando comparada por exemplo, às das técnicas de codificação pela forma de onda. Este fato traduz-se em uma significativa redução da taxa de bits e economia de memória para transmitir e/ou armazenar um sinal de voz. Assim, para sintetizar um quadro de voz, basta conhecer os correspondentes parâmetros de codificação para este quadro.

A estrutura mais utilizada em codificadores paramétricos de sinais de voz é a formada basicamente por um filtro de predição e um gerador de excitação, conforme é mostrado na Figura 2.2, onde $u(n)$ e $\hat{s}(n)$ são os sinais de excitação do filtro preditor, e o sinal de voz sintetizado, respectivamente.

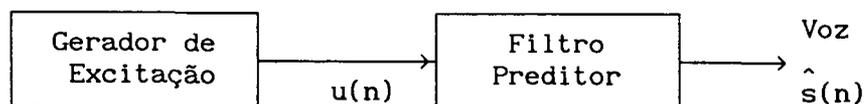


FIGURA 2.2 - Codificadores Paramétricos de Sinais de Voz

Por razões de ordem prática, a determinação do filtro preditor é realizada, na maioria dos casos, independentemente da excitação. Após, determina-se o gerador de excitação, considerando o preditor fixo. Esta técnica de determinação do preditor e da excitação leva a preditores "sub-ótimos", quando a excitação $u(n)$ for diferente daquela que foi utilizada na determinação do preditor.

2.3 - DETERMINAÇÃO DO PREDITOR LINEAR

No modelo de produção de voz supracitado, a composição dos efeitos do espectro de radiação, das cavidades bucal e nasal e da excitação glotal são representados pelo filtro ARMA ("AutoRegressive Moving Average") invariante no tempo para cada quadro de voz, cuja equação de diferenças tem a seguinte forma:

$$s(n) = \sum_{k=1}^{OD} a_k \cdot s(n-k) + \sum_{k=0}^{ON} b_k \cdot u(n-k) \quad (2.1)$$

onde $s(n)$ é o sinal de voz e $u(n)$ a excitação do filtro.

A transformação em z da Equação (2.1) leva à seguinte função de transferência geral de um filtro tipo ARMA.

$$H(z) = \frac{S(z)}{U(z)} = \frac{\sum_{k=0}^{ON} b_k \cdot z^{-k}}{1 - \sum_{k=1}^{OD} a_k \cdot z^{-k}} \quad (2.2)$$

O polinômio do numerador da Equação (2.2) representa um filtro do tipo MA ("Moving Average"), enquanto que o polinômio do denominador representa um filtro do tipo AR, sendo que estes dois, em cascata, constituem um sistema tipo ARMA.

Devido à não linearidade do sistema de equações obtido na fase de análise, para a estimação de sinais, em uma estrutura tipo ARMA[24], tem-se normalmente utilizado para modelar o trato vocal, filtros tipo AR com 10 a 12 pólos em sua função de transferência[3].

Com isso a função de transferência do filtro preditor terá a seguinte forma:

$$H(z) = \frac{b_0}{1 - \sum_{k=1}^{OD} a_k \cdot z^{-k}} \quad (2.3)$$

Em um sinal de voz, amostras próximas entre si e amostras separadas por um período de "PITCH" são altamente correlacionadas. Para aproveitar esta forte correlação, utiliza-se dois filtros preditores cascadeados. O primeiro é denominado filtro preditor de curto atraso (PCA) e o segundo, filtro preditor de longo atraso (PLA).

Um preditor linear com coeficientes de predição a_k , é definido como um sistema cuja função de transferência é

$$P(z) = \sum_{k=1}^{OD} a_k \cdot z^{-M_k} \quad (2.4)$$

onde M_k caracteriza o atraso de cada derivação do preditor $P(z)$.

A determinação dos coeficientes do preditor para uma implementação transversal pode ser realizada pelo modelo ilustrado na Figura 2.3. Nesta Figura o sinal de entrada $s(n)$ é multiplicado por uma janela $w(n)$ gerando um sinal $s_w(n)$. O erro de predição, neste caso, é dado por

$$e(n) = s_w(n) - \sum_{k=1}^{OD} a_k \cdot s_w(n - M_k) \quad (2.5)$$

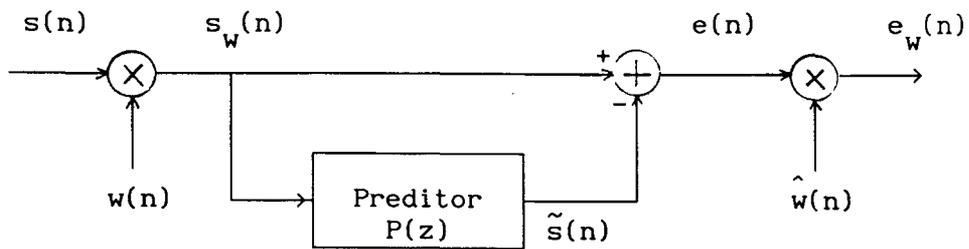


FIGURA 2.3 - Modelo de Análise para Preditores na Forma Transversal

A determinação dos coeficientes ótimos de predição a_k é realizada através da minimização do erro médio quadrático na saída do sistema da Figura 2.3.

Assim,

$$\frac{\partial \varepsilon}{\partial a_k} = 0 \quad (2.6)$$

onde

$$\varepsilon = \sum_{n=-\infty}^{\infty} e_w^2(n) \quad (2.7)$$

Esta otimização leva a um conjunto de equações lineares que podem ser escritas de forma matricial ($\underline{\phi}\underline{a}=\underline{g}$), ou seja

$$\begin{bmatrix} \phi(M_1, M_1) & \phi(M_1, M_2) & \dots & \phi(M_1, M_{OD}) \\ \phi(M_2, M_1) & \phi(M_2, M_2) & \dots & \phi(M_2, M_{OD}) \\ \phi(M_3, M_1) & \phi(M_3, M_2) & \dots & \phi(M_3, M_{OD}) \\ \dots & \dots & \dots & \dots \\ \phi(M_{OD}, M_1) & \phi(M_{OD}, M_2) & \dots & \phi(M_{OD}, M_{OD}) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \dots \\ a_P \end{bmatrix} = \begin{bmatrix} \phi(0, M_1) \\ \phi(0, M_2) \\ \phi(0, M_3) \\ \dots \\ \phi(0, M_{OD}) \end{bmatrix} \quad (2.8)$$

onde

$$\phi(i, j) = \sum_{n=-\infty}^{\infty} \hat{w}^2(n) \cdot s_w(n-i) \cdot s_w(n-j) \quad (2.9)$$

Para o caso dos PCA (Preditores de Curto Atraso), $M_k=k$ para $k=1$ até OD. Para o PLA (Preditor de Longo Atraso), $M_k=M+k$, para $k=0$ até $OD=OL-1$, onde OL representa o número de derivações do PLA.

A seguir descreve-se os métodos da autocorrelação e da covariância, mais utilizados na determinação dos coeficientes destes preditores na estimação de sinais de voz.

2.3.1 - MÉTODO DA AUTOCORRELAÇÃO

Caso a janela $\hat{w}(n)$ seja igual a 1 para todo n , ou seja, somente $s(n)$ seja janelado, resulta no procedimento conhecido como o método da autocorrelação. A janela $w(n)$ define o tamanho do quadro de voz e reduz os efeitos das discontinuidades nas extremidades do intervalo. São tipicamente utilizados para tal, janelas clássicas como as de Hanning, Hamming, Blackman, etc.

Uma consideração importante é a garantia de fase mínima para a função de predição $A(z) = 1 - P(z)$. Se $A(z)$ é fase mínima, o filtro inverso $1/A(z)$, usado na síntese, será estável. Pode-se mostrar que o preditor $P(z)$, obtido pelo método da autocorrelação leva sempre a ter $A(z)$ função fase mínima[18].

Para o caso do PLA, a propriedade de fase mínima não é garantida. Uma exceção ocorre quando os coeficientes são uniformemente espaçados, ou seja, $M_k = k \cdot M_1$.

A matriz ϕ é sempre simétrica e definida positiva. Para o PCA, ela é também uma matriz TOEPLITZ. Para o PLA, ela só é TOEPLITZ, caso os atrasos entre coeficientes sejam iguais. Caso ϕ não seja TOEPLITZ, pode-se utilizar para a solução do sistema de equações lineares, os métodos de LEVINSON[15], ROBINSON[16] ou a decomposição de CHOLESKY[3].

Para o caso da matriz ser TOEPLITZ, DURBIN[17] propôs um eficiente algoritmo para solução do sistema de equações que reduz substancialmente a complexidade computacional requerida em comparação com outros métodos. No Apêndice 1, apresenta-se uma sucinta descrição deste algoritmo.

2.3.2 - MÉTODO DA COVARIÂNCIA

Caso a janela $w(n)$ seja igual a 1 para todo n e $\hat{w}(n)$ seja igual a 1 para $0 \leq n < N$, ou seja, somente $e(n)$ seja janelado, resulta no procedimento conhecido como método da covariância.

A propriedade de fase mínima da função de predição $A(z)$ não é garantida quando utilizado o método da covariância[18]. ATAL & SCHROEDER[22] propuseram uma alteração neste método para garantir a estabilidade do filtro de predição.

Recentemente, NG & MILENKOVIC[21] propuseram um modelo simplificado considerando a não estacionaridade do sinal de voz no interior dos quadros de voz, mais propriamente nas vizinhanças dos sons consonantais explosivos. Estas não estacionaridades levam normalmente a um mal condicionamento da matriz ϕ implicando na determinação de PCA não estáveis. As vizinhanças dos sons consonantais são caracterizadas pelas rápidas mudanças de amplitude e, conseqüentemente, dos espectros dos sinais de voz. Esta condição foi modelada através de um sinal de baixo nível, representando um som consonantal, adicionado de um impulso de grande amplitude, representando o primeiro período de "PITCH" imediatamente seguinte ao som consonantal.

2.4. COMPARAÇÃO ENTRE OS MÉTODOS PARA SOLUÇÃO DAS EQUAÇÕES GERADAS NO PROJETO DO PCA - AR

A Tabela 2.1 mostra o número de operações aritméticas e a quantidade de memória necessária para cada um dos métodos estudados acima. Para a construção desta Tabela, considera-se que o número N de amostras do sinal utilizado e a ordem OD do preditor são os mesmos para ambos os métodos.

Em termos de armazenamento, os métodos da autocorrelação e covariância requerem a mesma quantidade de memória. Em termos de quantidade de operações aritméticas, o método da autocorrelação requer uma menor quantidade. Porém, como o cálculo da matriz correlação requer uma quantidade de operações aritméticas muito maior do que a solução do sistema de equações lineares ($N \gg OD$), os métodos são praticamente equivalentes em termos de complexidade computacional.

Finalmente, cabe colocar que o método da autocorrelação poderá gerar um filtro preditor instável quando da quantização dos seus coeficientes com um número finito de bits. Uma forma alternativa de determinação dos coeficientes com um número finito de bits de quantização é o projeto em uma estrutura LATTICE. Neste caso, o filtro preditor com coeficientes quantizados será estável[3].

	Covariância	Autocorrelação
Armazenamento		
* Dados	$O(N)$	$O(N)$
* Matriz	$O(OD^2)$	$O(OD)$
* Janela	-	$O(N)$
Operações Aritméticas		
* Janelamento	-	$O(N)$
* Correlação	$O(N \cdot OD)$	$O(N \cdot OD)$
* Solução do sistema de equações	$O(OD^3)$	$O(OD^2)$

OD - Ordem do preditor

$O(.)$ - Ordem de Complexidade

TABELA 2.1 - Comparação dos Métodos de Solução das Equações Geradas no Projeto dos Preditores PLA e PCA

2.5. MODELOS DO GERADOR DE EXCITAÇÃO

Um dos primeiros modelos de excitação foi proposto com base no mecanismo de produção real da voz humana[3]. Neste, a voz é produzida pela excitação de um tubo acústico, denominado trato vocal.

Como anteriormente descrito, devido à natureza estacionária para pequenos segmentos de voz, o trato vocal pode ser modelado como um sistema linear invariante no tempo.

A geração da voz é obtida basicamente através da produção de três sons básicos:

- Sons vozeados são produzidos pela excitação do trato vocal por impulsos pseudo-periódicos do fluxo de ar proveniente dos pulmões causando uma vibração nas cordas vocais.

- Sons fricativos são produzidos pela passagem forçada do ar numa constricção em uma região do trato vocal, produzindo neste estreitamento uma turbulência aleatória com características semelhantes às de um ruído branco.

- Sons explosivos são produzidos pelo fechamento completo da saída do trato vocal, seguido pelo aumento da pressão no seu interior e sua abertura abrupta.

O trato vocal é caracterizado por suas frequências naturais (denominadas por formantes) correspondendo às frequências de ressonância da transmissão sonora do trato vocal.

Se for assumido que as fontes de excitação e a forma do trato vocal são relativamente independentes, um modelo altamente compacto e elegante para a excitação pode ser usado, o qual é ilustrado na Figura 2.4.

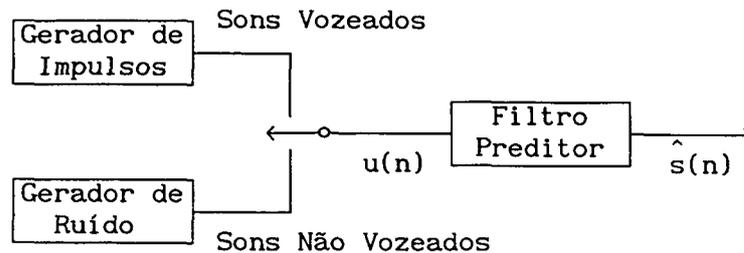


FIGURA 2.4 - Modelo Simplificado de Excitação dos Codificadores Paramétricos

Para os sons vozeados, é proposta a utilização de um gerador de impulsos periódicos com período igual ao "PITCH" do sinal de voz. Para os sons não vozeados, a excitação consiste de um gerador de ruídos.

Testes efetuados com este modelo de excitação[1,2,5], mostraram que a voz gerada é inteligível, mas não natural, mesmo quando usadas taxas de bits elevadas. Uma forma alternativa encontrada para suplantar as deficiências deste modelo foi proposta por ATAL & REMDE[6,7], a qual consiste na utilização de uma seqüência de impulsos de amplitudes variáveis tanto para os sons vozeados quanto para os não vozeados.

Neste modelo, em vez de um único impulso por período de "PITCH" para sons vozeados e ruído para os sons não vozeados, a excitação consiste sempre de alguns impulsos por quadro de voz. Uma vez que todos os sons são sintetizados usando-se a mesma forma de excitação, os problemas encontrados no modelo anterior de excitação quanto à estimação do período de "PITCH" e à análise vozeada/não vozeada são eliminados.

As amplitudes dos impulsos e suas localizações são computadas quadro a quadro pelo critério do mínimo erro médio quadrático, calculado entre o sinal de voz original e a saída do filtro preditor. O número de impulsos necessários em cada quadro de voz deve satisfazer a uma relação de compromisso entre a qualidade de voz desejada e a taxa de bits requerida.

Segundo SINGHAL & ATAL[1], 6 a 8 impulsos por quadro de voz, com 10 ms de duração por quadro, são suficientes para sintetizar a voz com alta qualidade.

Embora este modelo de excitação seja simples, o problema de localização dos impulsos é computacionalmente complexo.

Um terceiro modelo de excitação consiste em utilizar-se uma biblioteca de excitações correntemente denominada na literatura por "codebook"[8-12]. A idéia central desta técnica é selecionar desta biblioteca uma excitação cuja voz gerada seja a de "melhor" qualidade, em vez de, para cada quadro, determinar-se uma nova excitação. Esta forma de codificação ficou conhecida como CELP ("Code Excited Linear Prediction"). É importante lembrar que neste caso é utilizado o índice da biblioteca para a codificação da excitação, diminuindo-se substancialmente a taxa de bits para representá-la. Para a determinação do "codebook", diversos algoritmos foram propostos na literatura, destacando-se o trabalho de KLEIJN et alii[9], os quais fizeram um estudo das vantagens e desvantagens destes diversos algoritmos.

Muitas variações destes modelos de excitação podem ser encontrados em [1-3,5-14].

2.6. PROJETO DO GERADOR DE EXCITAÇÃO MULTI-PULSO

A excitação multi-pulso consiste de alguns impulsos, estrategicamente localizados com determinadas amplitudes, para cada quadro de voz. Os impulsos são determinados (localizações e amplitudes) pela minimização do erro médio quadrático entre a voz original e a saída do PCA-AR.

Infelizmente, o resultado da minimização conduz a um problema combinatorial cuja solução global requer um esforço computacional demasiadamente grande.

Assim sendo, naturalmente deve-se buscar soluções alternativas para este problema.

Uma técnica sub-ótima, proposta por SINGHAL & ATAL[1], consiste na determinação das localizações e amplitudes dos impulsos em diversas etapas. Em cada etapa, determina-se a localização e a amplitude de apenas um impulso através de uma pesquisa exaustiva. Os impulsos obtidos nas etapas precedentes são assumidos constantes durante a pesquisa do impulso corrente.

A Figura 2.5 mostra o diagrama de blocos do procedimento de análise descrito por SINGHAL & ATAL[1].

Nesta Figura, o gerador de excitação produz uma seqüência de impulsos que é aplicada à entrada do filtro PCA-AR. O erro entre o sinal de voz original, $s(n)$, e o sinal de voz sintetizado, $\tilde{s}(n)$, é ponderado através de um filtro apropriado produzindo um sinal de erro ponderado, $e(n)$, que controlará o processo.

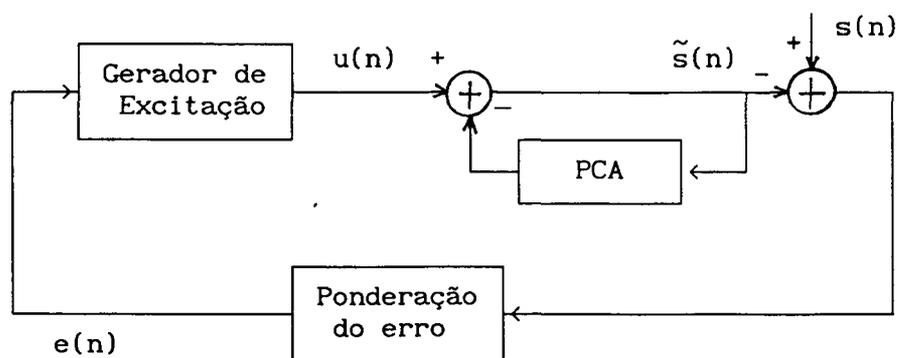


FIGURA 2.5 - Diagrama de Blocos da Etapa de Análise do Codificador Multi-Pulso

Assim, o gerador de excitação fornece as amplitudes e localizações dos impulsos tal que o erro ponderado seja de alguma forma minimizado.

O propósito do filtro de ponderação do erro é atenuar o erro nas frequências que são perceptivelmente menos importantes e enfatizá-lo nas frequências mais importantes.

Por esta razão, os componentes do erro nas regiões de frequência em torno dos formantes são menos perceptíveis que aquelas entre formantes. O mesmo pode ser dito para as regiões de alta e baixa energia dos sons não vozeados.

O filtro ponderador tem a seguinte função de transferência:

$$W(z) = [1 + P(z)] / [1 + P(z/\gamma)] \quad (2.10)$$

para

$$1 + P(z) = 1 + \sum_{k=1}^{OD} a_k \cdot z^{-k} \quad (2.11)$$

Onde $P(z)$ é a função de transferência do preditor e γ é uma constante entre zero e um que determina o grau para o qual o erro é desenfaturado[8,1,15,5].

O parâmetro γ também pode ser expresso como

$$\gamma = e^{-2\pi\delta/f_s} \quad (2.12)$$

onde δ caracteriza a redução da magnitude dos pólos de $W(z)$ e f_s é a frequência de amostragem utilizada.

Para valores de δ entre 100 e 400 obtém-se os melhores resultados em termos de tolerância ao ruído[8].

Para uma frequência de amostragem de 10 kHz, γ tem valores entre 0,78 e 0,94. Os resultados das simulações apresentados neste trabalho foram obtidos para $\gamma = 0,85$.

Visto que o filtro ponderador é linear, a minimização de $e(n)$ para a obtenção das localizações e amplitudes dos impulsos é equivalente à minimização do erro médio quadrático obtido entre o sinal de voz ponderado $y(n)$ e o correspondente sinal de voz sintetizado ponderado $\hat{y}(n)$, como mostra o diagrama da Figura 2.6.

Assim,

$$y(n) = s(n) + \sum_{k=1}^{OD} a_k \cdot \left[s(n-k) - \gamma^k \cdot y(n-k) \right] \quad (2.13)$$

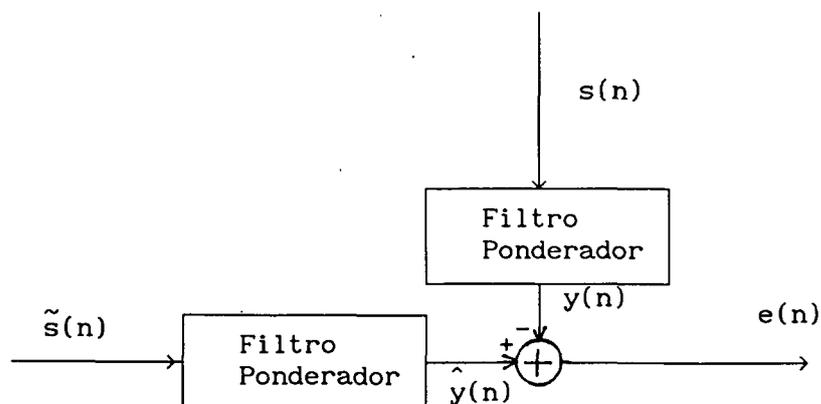


FIGURA 2.6 - Determinação do Sinal de Erro Ponderado

O sinal de excitação $u(n)$ é filtrado pelo PCA-AR e pelo filtro ponderador resultando em,

$$\hat{y}(n) = \sum_{i=0}^n h(i) \cdot u(n-i) + \hat{y}_0(n) \quad , 0 \leq n < N \quad (2.14)$$

onde N designa o número de amostras por quadro de voz, $h(n)$ e $u(n)$ são a resposta ao impulso do filtro PCA-AR ponderado e a correspondente excitação, respectivamente, e $\hat{y}_0(n)$ é a contribuição para $\hat{y}(n)$ da memória do PCA-AR.

Se a entrada do presente quadro é modelada como uma série de m impulsos de amplitudes $\beta_0, \beta_1, \dots, \beta_{m-1}$ nas localizações n_0, n_1, \dots, n_{m-1} , respectivamente, tem-se

$$u(n) = \sum_{k=0}^{m-1} \beta_k \cdot \delta(n-n_k) \quad (2.15)$$

onde $\delta(n)$ é função impulso unitário. Então

$$\hat{y}(n) = \sum_{i=0}^n h(i) \cdot \sum_{k=0}^{m-1} \beta_k \cdot \delta(n-i-n_k) + \hat{y}_0(n) \quad (2.16)$$

ou

$$\hat{y}(n) = \sum_{k=0}^{m-1} \beta_k \cdot h(n-n_k) + \hat{y}_0(n) \quad , 0 \leq n < N \quad (2.17)$$

As amplitudes dos impulsos $\beta_0, \beta_1, \dots, \beta_{m-1}$ e as correspondentes localizações n_0, n_1, \dots, n_{m-1} são obtidas pela minimização do erro quadrático ε definido por

$$\varepsilon = \sum_{n=0}^{N-1} \left[y(n) - \hat{y}(n) \right]^2 \quad (2.18)$$

O erro ε deve ser minimizado com respeito às amplitudes dos impulsos $\beta_0, \dots, \beta_{m-1}$ e às localizações n_0, \dots, n_{m-1} . A minimização com respeito às amplitudes não representa um problema maior, resultando no seguinte sistema de Equações:

$$\sum_{k=0}^{m-1} \beta_k \cdot \alpha_{n_k n_j} = C_{n_j} \quad , 0 \leq j < m \quad (2.19)$$

onde α é a função de autocorrelação da resposta ao impulso, definida por,

$$\alpha_{ij} = \sum_{n=0}^{N-1} h(n-i) \cdot h(n-j) \quad , 0 \leq i, j < N \quad (2.20)$$

e C é a função de correlação cruzada entre a resposta $h(n)$ ao impulso e o sinal

$$\bar{y}(n) = y(n) - \hat{y}_0(n), \quad (2.21)$$

assim

$$C_i = \sum_{n=0}^{N-1} \bar{y}(n) \cdot h(n-i) \quad , 0 \leq i < N \quad (2.22)$$

O erro mínimo ϵ é dado por

$$\epsilon_{\min} = \sum_{n=0}^{N-1} \bar{y}^2(n) - \sum_{k=0}^{m-1} \beta_k \cdot C_{n_k} \quad (2.23)$$

A minimização de ϵ com respeito as localizações n_0, n_1, \dots, n_{m-1} é um problema combinatorial e não tem uma forma fechada de solução. É necessário escolher m de N possíveis localizações, tal que a correspondente solução de (2.19) leve a um mínimo erro.

Uma pesquisa exaustiva das localizações dos impulsos requer $\binom{N}{m} = N! / m!(N-m)!$ soluções de (2.19), que é impraticável quando o número m de impulsos aumenta.

Como anteriormente discutido, pode-se obter uma solução sub-ótima, efetuando-se a pesquisa em etapas, onde em cada etapa, uma única localização é determinada.

Em cada etapa, a amplitude e a localização de somente um impulso são pesquisadas. Isto reduz o número de pesquisas para

$$NR = \sum_{i=0}^{m-1} (N-i) \quad (2.24)$$

Uma vez que as localizações dos impulsos são encontradas, (2.19) é usada para reotimizar todas as amplitudes dos impulsos. A seguir descreve-se brevemente este procedimento[1,6,7].

Seja a entrada consistindo de somente um impulso localizado na amostra i . As equações (2.19) e (2.23) reduzem-se para

$$\beta_0(i) = C_i / \alpha_{ii} \quad (2.25)$$

$$\epsilon_{\min}(i) = \sum_{n=0}^{N-1} \bar{y}^2(n) - \left[C_i^2 / \alpha_{ii} \right] \quad (2.26)$$

onde β_0 e ϵ_{\min} são funções explícitas de i , para enfatizar sua dependência com respeito à localização do impulso i durante a pesquisa.

De (2.26), vê-se que $\epsilon_{\min}(i)$ é mínimo para os valores de i onde c_i^2/α_{ii} é máximo. Isto implica que o primeiro impulso seria localizado na posição correspondente ao máximo da função correlação cruzada normalizada.

Seja esta localização n_0 e a correspondente amplitude β_0 . Assume-se agora que β_0 e n_0 são constantes conhecidas em (2.17) e minimiza-se ϵ com respeito a $\beta_1, \dots, \beta_{m-1}$ obtendo-se a seguinte Equação

$$\sum_{k=1}^{m-1} \beta_k \cdot \alpha_{n_k n_j} = C_{n_j} - \beta_0 \cdot \alpha_{n_0 n_j} \quad , 1 \leq j < m \quad (2.27)$$

Fazendo-se o número de impulsos igual a 2, (2.27) contém agora somente duas incógnitas, ou seja, β_1 e n_1 . Se o segundo impulso é localizado em i , tem-se

$$\beta_1(i) = \left[C_i - \beta_0 \cdot \alpha_{n_0 i} \right] / \alpha_{ii} \quad (2.28)$$

O erro ϵ_{\min} passa a ser

$$\epsilon_{\min} = \sum_{n=0}^{N-1} \bar{y}^2(n) - \left[\left[C_i - \beta_0 \cdot \alpha_{n_0 i} \right]^2 / \alpha_{ii} \right] \quad (2.29)$$

Como antes, faz-se $n_1 = i$ para a localização onde o segundo termo em (2.29) é máximo. A correspondente amplitude é calculada de (2.28).

Pode-se notar que (2.29) tem a mesma forma que (2.23), onde a função correlação cruzada C_i foi substituída por $[C_i - \beta_0 \cdot \alpha_{n_0 i}]$ e a localização do impulso corresponde à localização do máximo valor da função de correlação cruzada normalizada "atualizada". O procedimento é repetido para cada impulso, isto é, o $(j+1)$ -ésimo impulso é localizado no máximo de

$$\left[C_i^{(j)} \right]^2 / \alpha_{ii} = \left[C_i^{(j-1)} - \beta_{j-1} \cdot \alpha_{n_{j-1} i} \right]^2 / \alpha_{ii}, \quad 0 \leq i < N \quad (2.30)$$

e sua amplitude é

$$\beta_j = C_{n_j}^{(j)} / \alpha_{n_j n_j} \quad (2.31)$$

com

$$C_j^{(0)} = C_i \quad (2.32)$$

Nas simulações efetuadas foi assumido que $\alpha_{ij} = \alpha_{|i-j|}$, tornando a normalização da função correlação cruzada desnecessária.

Esta consideração consiste na prática em aproximar-se a função α_{ij} , que tem a forma de uma matriz simétrica, para a forma de uma matriz Toeplitz. Isto implica que pode-se utilizar os procedimentos de LEVINSON - ROBINSON[15,16] para solucionar (2.19), em vez do uso da decomposição de CHOLESKY[3,1], diminuindo-se, desta forma, a complexidade computacional requerida.

2.7. PROJETO DO PLA PARA EXCITAÇÃO MULTI-PULSO

Segundo ATAL et alii[6,7,8], 6 a 8 impulsos são necessários por período de "PITCH" na excitação multi-pulso para a produção de voz com alta qualidade. Isto implica que a voz feminina, por conter componentes de frequência mais elevada, requer mais impulsos por quadro do que a voz masculina, para uma qualidade comparável.

O modelo de excitação multi-pulso discutido, não assume nada com relação às características do sinal de voz, ou seja, todos os sons da voz foram modelados com o mesmo sinal de excitação. Para sons vozeados, a excitação multi-pulso apresenta uma significativa correlação de um período de "PITCH" para o próximo.

Utilizando-se um outro preditor para aproveitar esta correlação, pode-se diminuir o número de impulsos na excitação multi-pulso[1], mantendo-se a mesma relação sinal-ruído.

O modelo do gerador de excitação multi-pulso mostrado na Figura 2.5 é alterado para incluir um filtro PLA entre o filtro PCA e o gerador de excitação. A Figura 2.7 mostra o diagrama de bloco desta nova estrutura do codificador.

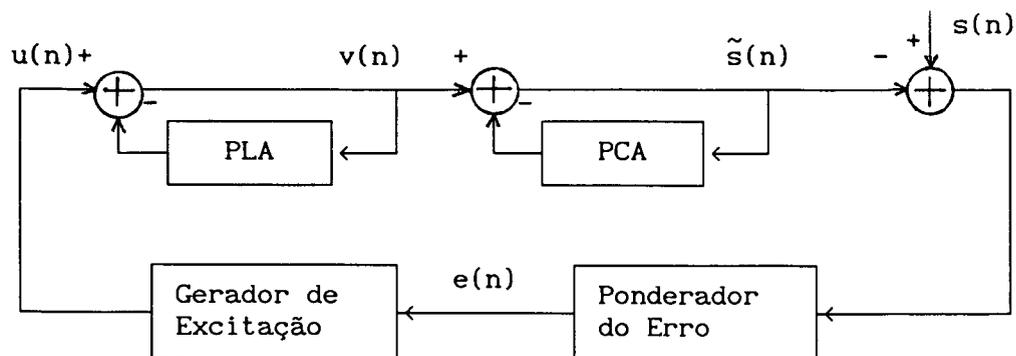


FIGURA 2.7 - Codificador Multi-pulso com PLA

Agora, a excitação do PCA será:

$$v(n) = u(n) + \sum_{i=-T}^T g_i \cdot v(n-M+i) \quad , 0 \leq n < N \quad (2.33)$$

onde $u(n)$ é a n -ésima amostra da excitação multi-pulso, g_i , $-T \leq i < T$ são os ganhos de cada derivação do filtro preditor e M é o atraso do preditor (PLA).

O atraso do preditor (M) deve ser maior do que alguns períodos de "PITCH" e, em geral, é maior do que $N-1+T$.

Agora, pode-se expressar o sinal de voz sintetizado ponderado como:

$$\hat{y}(n) = \sum_{k=0}^{m-1} \beta_k \cdot h(n-n_k) + \sum_{i=-T}^T g_i \cdot \sum_{k=0}^{N-1} h(k) \cdot v(n+i-M-k) + \hat{y}_0(n) \quad (2.34)$$

A Equação (2.34) consiste de três termos. O primeiro é relativo à excitação multi-pulso. O segundo é a excitação obtida pelo PLA representando aos quadros passados. E o terceiro é a memória do PCA-AR. Outra vez, é necessário minimizar o erro médio quadrático ϵ entre os sinais de voz original e sintetizado.

A minimização em relação a g_i resulta em um problema linear quando $M > N - 1 + T$. A minimização em relação ao atraso M outra vez resulta em um problema não linear.

Pode-se obter os impulsos de excitação e o atraso M do PLA utilizando-se para tal, pesquisa exaustiva. Porém o número de pesquisas torna impraticável esta forma de obtenção destes parâmetros.

Uma forma de reduzir a complexidade é obter o PLA e a excitação multi-pulso em dois passos.

No primeiro, assume-se que a excitação multi-pulso é zero e pesquisa-se o valor do atraso M e os ganhos g_i do PLA, tal que ϵ seja minimizado. Depois, o PLA é mantido constante e a excitação multi-pulso é determinada como anteriormente descrito.

$$\text{Seja } \beta_k = 0, \quad 0 \leq k < m$$

e

$$z_M = \sum_{k=0}^{N-1} h(k) \cdot v(n-M-k) \quad , 0 \leq n < N \quad (2.35)$$

A minimização do ϵ em relação ao ganho g_k resulta em

$$\sum_{i=-T}^T g_i \cdot \sum_{n=0}^{N-1} z_M^{(n+i)} \cdot z_M^{(n+k)} = \sum_{n=0}^{N-1} \bar{y}(n) \cdot z_M^{(n+k)} \quad , -T \leq k < T \quad (2.36)$$

e

$$\epsilon_{\min}^{(M)} = \sum_{n=0}^N \bar{y}^2(n) - \sum_{i=-T}^T g_i \cdot \sum_{n=0}^{N-1} \bar{y}(n) \cdot z_M^{(n+i)} \quad (2.37)$$

O sistema de equações (2.36) é usado para determinar os ganhos g_i do PLA para todos os possíveis valores de M , e o erro resultante é calculado por (2.37).

Com o uso do PLA consegue-se melhorar de 2 dB a 5 dB a relação sinal ruído da voz. Em [1,18] pode-se encontrar maiores detalhes a respeito dos PLA.

2.8. CONCLUSÃO

Este capítulo apresentou uma revisão bibliográfica das principais técnicas paramétricas utilizadas na síntese de sinais de voz. Foram discutidos os principais modelos e técnicas utilizados na determinação do gerador de excitação e do filtro preditor linear.

Para aproveitar a forte correlação das amostras dos sinais de voz distanciados por períodos de "PITCH", foi apresentado o PLA (Preditor de Longo Atraso), o qual foi inserido entre o gerador de excitação e o filtro PCA (Preditor de Curto Atraso).

C A P Í T U L O 3

CODIFICADORES MULTI-PULSO COM PCA - ARMA

3.1. INTRODUÇÃO

Neste capítulo, propõe-se a utilização de um filtro preditor tipo ARMA no lugar do AR, tradicionalmente utilizado no projeto do PCA e no projeto da excitação multi-pulso. São apresentadas tabelas com os resultados obtidos nas simulações onde verifica-se que o uso do filtro preditor ARMA sintetiza a voz com uma maior relação sinal-ruído do que o uso do filtro preditor AR.

Também é mostrado que procurando-se o "melhor" preditor para cada quadro de voz, consegue-se melhorar a relação sinal-ruído e reduzir a taxa média de bits para representar os parâmetros do filtro preditor.

Finalmente, é proposta uma nova técnica para a determinação das localizações dos impulso, com uma menor complexidade computacional do que a técnica apresentada por SINGHAL & ATAL[1], abordada no Capítulo 2.

3.2. ESTUDOS E CONSIDERAÇÕES PARA A DETERMINAÇÃO DO GERADOR DE EXCITAÇÃO COM PCA - ARMA

Neste item é apresentado o procedimento para a determinação da excitação multi-pulso com PCA-ARMA.

A Figura 3.1 mostra o diagrama de blocos da estrutura de análise para a determinação dos preditores lineares com filtro de predição do tipo ARMA.

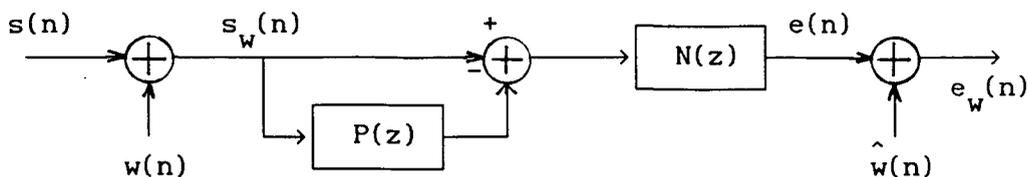


FIGURA 3.1 - Modelo de Análise para Preditores Transversais com Filtro de Predição do Tipo ARMA.

onde

$$P(z) = \sum_{k=1}^{OD} a_k \cdot z^{-k} \quad (3.1)$$

e

$$N(z) = \sum_{k=0}^{ON} b_k \cdot z^{-k} \quad (3.2)$$

Os procedimentos para determinação dos parâmetros a_k e b_k de $P(z)$ e $N(z)$, respectivamente, podem ser os mesmos descritos no capítulo 2. Porém, aquela abordagem leva a um sistema de equações não-lineares, implicando em um importante esforço computacional para resolvê-lo.

Neste trabalho foi utilizada a técnica de determinação do filtro ARMA proposta em [4], por esta levar a um sistema de equações lineares facilmente solucionáveis.

Tal técnica é baseada no critério do filtro inverso, ou seja, excitando-se o filtro inverso $1/H(z)$ com a resposta ao impulso desejada, obtém-se na saída um impulso adicionado de um erro denominado de erro do filtro inverso. Minimizando-se este erro, através de algumas considerações e aproximações convenientes, chega-se a um sistema de equações lineares que quando solucionadas fornecem os coeficientes do filtro $H(z)$, onde $H(z) = N(z) / (1+P(z))$.

Uma vez estando determinados os parâmetros do filtro preditor ARMA, resta determinar a excitação multi-pulso. O procedimento de análise é ilustrado na Figura 3.2.

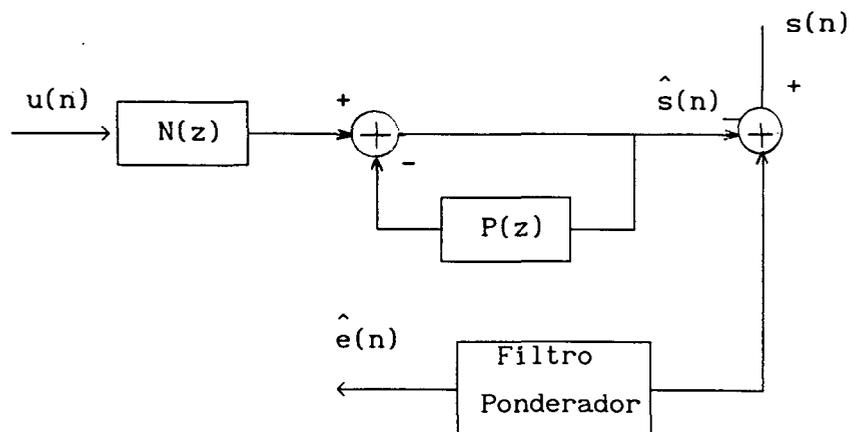


FIGURA 3.2 - Modelo de Análise para a Determinação da Excitação Multi-pulso com um Filtro Preditor do Tipo ARMA.

A excitação multi-pulso é determinada pela minimização do erro médio quadrático ϵ , onde

$$\epsilon = \sum_{n=0}^{N-1} \hat{e}^2(n) \quad (3.3)$$

Identicamente ao procedimento descrito no capítulo 2, pode-se ponderar o erro, ponderando-se ambos $s(n)$ e $\hat{s}(n)$, conforme a Figura 3.3.

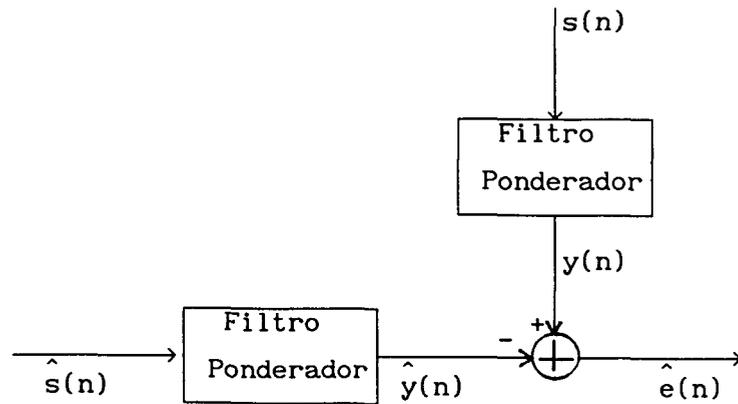


FIGURA 3.3 - Determinação do Erro Ponderado.

A finalidade do filtro ponderador ($W(z)$) é a mesma descrita no capítulo 2. Sua função de transferência é,

$$W(z) = \frac{1 + \sum_{k=1}^{OD} a_k \cdot z^{-k}}{1 + \sum_{k=1}^{OD} a_k \cdot \gamma^k \cdot z^{-k}} \quad (3.4)$$

O filtro preditor ponderado $H(z)$ será então

$$H(z) = \frac{\sum_{k=0}^{ON} b_k \cdot z^{-k}}{1 + \sum_{k=1}^{OD} a_k \cdot \gamma^k \cdot z^{-k}} \quad (3.5)$$

O sinal reconstruído será descrito por

$$\hat{s}(n) = \sum_{k=0}^{ON} b_k \cdot \sum_{j=0}^{m-1} \beta_j \cdot \delta(n-k-n_j) - \sum_{k=1}^{OD} a_k \cdot \hat{s}(n-k) \quad (3.6)$$

e $h(n)$ será descrito por

$$h(n) = \sum_{k=0}^{ON} b_k \cdot \delta(n-k) - \sum_{k=1}^{OD} a_k \cdot \gamma^k \cdot h(n-k) \quad (3.7)$$

O procedimento para a determinação da excitação multi-pulso, quando utilizada a técnica proposta por SINGHAL & ATAL[1], é o mesmo descrito no capítulo 2 incluindo-se apenas algumas alterações conforme acima descrito.

O restante das equações permanecem idênticas àquelas apresentadas no capítulo 2.

3.3. NOVA TÉCNICA PARA DETERMINAÇÃO DAS LOCALIZAÇÕES DOS IMPULSOS DO GERADOR DE EXCITAÇÃO MULTI-PULSO

Como descrito no capítulo 2, a determinação da localização dos impulsos do gerador de excitação multi-pulso é efetuada em etapas, ou seja, apenas um impulso é determinado a cada etapa.

Com isso, o número total de pesquisas é dado por:

$$NR = \sum_{i=0}^{m-1} [N - i] \quad (3.8)$$

Exemplificando-se, para $N=100$, $m=8$, tem-se $NR=772$ pesquisas. É evidente que este número é muito menor do que 186.10^9 pesquisas necessárias para a determinação ótima. Porém, ainda assim, 772 pesquisas é um número grande.

É sabido que os pontos de máximos da magnitude da função de correlação cruzada entre dois sinais, por nós definidos como pontos ou modos de máxima "pseudo-energia" cruzada, determinam as posições onde estes sinais têm a maior semelhança[23]. É também sabido que se um destes sinais representar a resposta ao impulso de um sistema linear invariante e se este sistema for excitado com impulsos cujas localizações forem as mesmas destes pontos de máximos, pode-se ter, ajustando-se adequadamente as amplitudes destes impulsos, a máxima semelhança, em termos de erro médio quadrático, entre o sinal de saída deste sistema e aquele correspondente ao outro sinal.

Assim, baseado neste princípio, propomos determinar a localização dos impulsos, numa primeira aproximação, através da escolha dos pontos de máxima "pseudo-energia" cruzada entre o sinal de voz original e o correspondente à resposta ao impulso do filtro preditor.

O princípio parece elementar, no entanto, diversas dificuldades surgem em sua imediata aplicação. Uma delas refere-se aos quadros de voz onde não existe uma forte correlação entre as amostras do mesmo. Outra, é que os elementos de C_i (Equação 2.22) são estimados da correlação cruzada entre o sinal de voz original e a resposta ao impulso do filtro preditor. Para valores de i próximos a $N-1$, esta estimativa inclui alguns poucos termos no somatório e a solução da (Equação 2.22) poderá vir a ser mal condicionada.

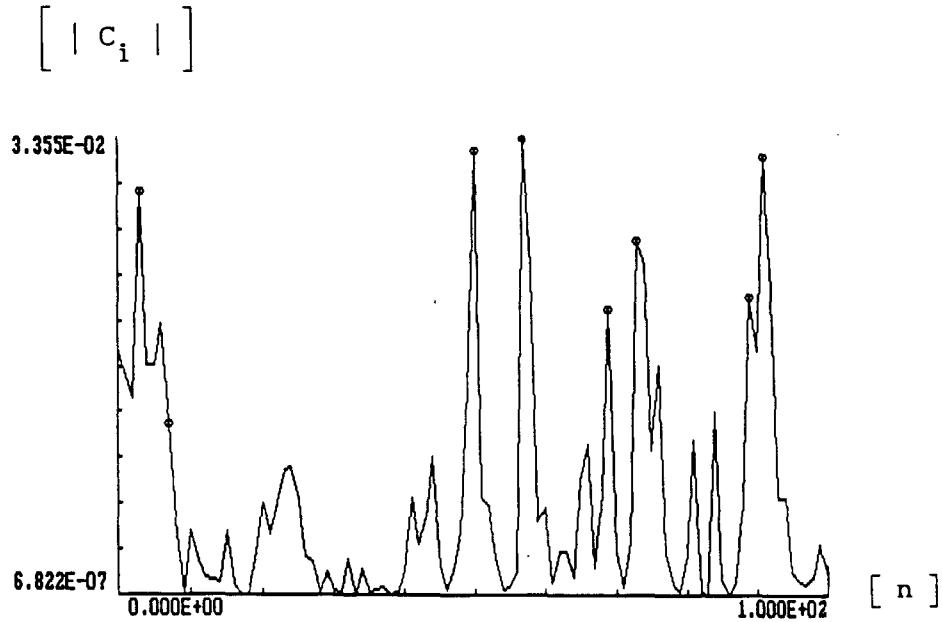
A título de ilustração, mostra-se na Figura 3.4 uma curva típica dos modos de máxima "pseudo-energia" cruzada entre o sinal de voz original e a resposta ao impulso do filtro preditor.

São também mostrados, nesta Figura, os pontos referentes às localizações determinadas pelo método apresentado em [1]. Pode-se observar a quase perfeita concordância entre os maiores modos de máxima "pseudo-energia" cruzada e as localizações obtidas pelo método da referência [1], divergindo apenas em uma localização para $m=8$. Como esta localização encontra-se numa região onde C_i é relativamente bem condicionada, é de se esperar que ao invés de se utilizar as localizações determinadas pelo método da referência [1], utilizar-se as localizações determinadas pelos maiores modos de máxima "pseudo-energia" cruzada, consiga-se um melhor resultado.

Para este exemplo em particular, foram obtidos os seguintes resultados:

A seqüência das localizações, em ordem decrescente para $m=8$ e $N=100$, utilizando o método apresentado na referência [1] é

n_0	n_1	n_2	n_3	n_4	n_5	n_6	n_7
57	50	91	3	73	69	7	89



Legenda

- o - Localização Determinada pelo Método
Apresentado na Referência [1]

FIGURA 3.4 - Curva Típica da Função de Correlação Cruzada entre o Sinal de Voz Original e a Resposta ao Impulso do Filtro Preditor Para um Quadro de Voz

resultando em $EMQ_N^{(*)} = 0,3728$ e $RSR^{(*)} = 4,29$ dB.

E para os maiores modos de máxima "pseudo-energia" cruzada obteve-se

n_0	n_1	n_2	n_3	n_4	n_5	n_6	n_7
57	50	91	3	73	89	69	6

resultando em $EMQ_N^{(*)} = 0,3593$ e $RSR^{(*)} = 4,44$ dB.

Pode-se constatar, como anteriormente afirmado, que a determinação das localizações através dos maiores modos de máxima "pseudo-energia" cruzada, neste caso, leva a um melhor resultado. No entanto, nem sempre isto é verdade conforme já mencionado.

Assim, numa primeira etapa, que levaria a uma redução do número total de pesquisas, poder-se-ia utilizar o método da referência [1] para pesquisar as localizações dos impulsos em N_p localizações, correspondendo aos N_p maiores modos de máxima "pseudo-energia" cruzada, ao invés de se fazê-las em todo o comprimento do quadro de voz.

Assim sendo, o número total de pesquisas diminuiria para

$$NR = \sum_{i=0}^{m-1} [N_p - i] \quad (3.9)$$

Exemplificando-se, com $N=100$, $N_p=60$ e $m=8$, temos $NR=452$ pesquisas, ou seja, 41% a menos do que o obtido pela Equação (3.8).

(*) Obtidos pelas equações (3.11) e (3.12).

Um outro fato interessante é que, na maioria das vezes, as primeiras localizações encontradas, coincidem com aquelas dos primeiros maiores modos de máxima "pseudo-energia" cruzada do procedimento proposto. Com esta consideração, pode-se desprezar as pesquisas para estas primeiras localizações.

Definindo-se N_f como o número de localizações destas primeiras posições, constata-se, assim, que o número total de pesquisas diminui, ainda mais, para

$$NR = \sum_{i=0}^{m-N_f-1} [(N_p - N_f) - i] \quad (3.10)$$

Exemplificando-se, para $N=100$, $N_p=60$, $m=8$ e $N_f=3$, temos $NR=275$ pesquisas, ou seja, 64% a menos do que o número de pesquisas necessárias para o algoritmo descrito no capítulo 2.

3.4. RESULTADOS

A seguir são apresentados os principais resultados obtidos nas simulações das várias técnicas propostas para a realização do codificador multi-pulso. Como figura de mérito para medir a qualidade do sinal de voz sintetizado utilizou-se o critério do erro médio quadrático normalizado, EMQ_N , entre $s(n)$ e $\tilde{s}(n)$, sinais de voz original e sintetizado, respectivamente. Assim tem-se:

$$EMQ_N = \frac{\sum_{n=0}^{N-1} [s(n) - \tilde{s}(n)]^2}{\sum_{n=0}^{N-1} s^2(n)} \quad (3.11)$$

Define-se a relação sinal-ruído, RSR, por:

$$RSR = 10 \log \frac{1}{EMQ_N} \quad (3.12)$$

3.4.1. PARÂMETROS PARA AS SIMULAÇÕES

Os seguintes parâmetros foram utilizados para todas as simulações:

$$\gamma = 0,85$$

$$m = 8 \text{ impulsos}$$

$$N = 100 \text{ amostras}$$

$$fs = 10 \text{ kHz}$$

$$nq = 100 \text{ quadros}$$

Onde γ é o coeficiente de desênfase do filtro de ponderação do erro, m é o número de impulsos utilizados no gerador de excitação, N é o número de amostras por quadro de voz, fs é a frequência de amostragem, e nq o número total de quadros utilizados.

O sinal de voz utilizado em todas as avaliações consiste de uma gravação de um segmento do programa "Voz do Brasil".

3.4.2. RESULTADOS OBTIDOS COM UM FILTRO PCA - AR

Para um preditor com OD=10 e utilizando-se o algoritmo descrito no capítulo 2 referente à técnica apresentada na referência [1], obteve-se:

$$EMQ_N = 0,253021$$

$$RSR = 5,97 \text{ dB}$$

O projeto do preditor foi realizado utilizando-se a formulação da autocorrelação e o algoritmo de DURBIN[17] para obter a solução do sistema de equações lineares.

O número de pesquisas necessárias para a determinação da excitação sub-ótima, para cada quadro de voz, utilizando-se a supracitada técnica, é NR = 772 pesquisas.

A seguir será apresentada uma série de resultados onde pode-se observar o desempenho das técnicas por nós propostas, em termos da relação sinal-ruído e taxa média de bits do sinal de voz codificado.

Como primeira proposta para a melhoria da relação sinal- ruído, em vez de se manter constante a ordem OD do filtro PCA-AR, buscou-se para cada quadro o "melhor" preditor, variando sua ordem, de forma a obter-se a maior relação sinal-ruído. Com isso, obteve-se os seguintes resultados:

$$EMQ_N = 0,229244$$

$$RSR = 6,40 \text{ dB}$$

$$\text{Ordem média} = 8,16$$

A Tabela 3.1 mostra a estatística da ordem do preditor encontrada.

OD	Número de Ocorrências
1	2
2	1
4	2
5	3
6	10
7	10
8	19
9	23
10	30
Total	100

TABELA 3.1 - Estatística da Ordem do Preditor para $OD \leq 10$.

Pode-se notar na Tabela acima, que somente para 30% dos quadros, o "melhor" filtro preditor é o PCA-AR com $OD=10$.

A relação sinal-ruído aumentou de 5,97 dB para 6,40 dB, um ganho de 0,43 dB. Também é importante o resultado em relação à diminuição média da ordem do preditor. Se considerarmos que cada coeficiente do preditor fosse quantizado com o mesmo número de bits, pode-se dizer que houve uma economia de 18,4% em média no número de bits total para representar o filtro PCA-AR.

Esta economia na taxa média de bits para representar o filtro PCA-AR é significativa na contribuição para a diminuição da taxa média de bits do sinal de voz codificado.

Por isso, com o objetivo de diminuir ainda mais a ordem média do preditor, procurou-se não a melhor relação sinal-ruído, mas o filtro preditor de menor ordem de cada quadro com pelo menos a mesma relação sinal-ruído do filtro preditor obtido com a técnica descrita na referência [1], ou seja, buscou-se um filtro PCA-AR com $OD \leq 10$ e $RSR \geq RSR$ para $OD=10$.

Obteve-se os seguintes resultados:

$$EMQ_N = 0,237920$$

$$RSR = 6,24$$

$$\text{Ordem média} = 7,12$$

A Tabela 3.2 apresenta a estatística da ordem do preditor encontrada.

OD	Número de Ocorrências
1	3
2	1
3	3
4	18
5	6
6	9
7	10
8	8
9	12
10	30
Total	100

TABELA 3.2 - Estatística da Ordem do Preditor para $OD \leq 10$ com $RSR \geq RSR$ para PCA-AR com $OD = 10$

Estes resultados mostram que o número médio de bits para representar o filtro PCA-AR reduziu-se 28,8% em relação ao PCA-AR calculado conforme a referência [1]. A relação sinal-ruído manteve-se maior para todos os quadros. Pode-se também notar que o número de quadros com ordem $OD=10$ para o preditor permanece igual ao da simulação anterior.

Para a próxima simulação, buscou-se aumentar a relação sinal-ruído mantendo-se a taxa média de bits para representar o filtro PCA-AR igual à da referência [1]. Esta simulação foi realizada para $OD \leq 13$, obtendo-se os seguintes resultados:

$$EMQ_N = 0,223366$$

$$RSR = 6,51 \text{ dB}$$

$$\text{Ordem média} = 9,65$$

A Tabela 3.3 apresenta a estatística da ordem do preditor encontrada.

OD	Número de Ocorrências
1	2
4	2
5	2
6	8
7	6
8	14
9	12
10	11
11	9
12	16
13	18
Total	100

TABELA 3.3 - Estatística da Ordem do Preditor para $OD \leq 13$.

Pode-se notar que a ordem média ficou menor do que para o PCA-AR com $OD=10$. Contudo, obteve-se uma relação sinal-ruído de 6,51 dB, portanto 0,54 dB maior do que a do PCA-AR com $OD=10$ proposto na referência [1].

3.4.3. RESULTADOS OBTIDOS COM UM FILTRO PCA - ARMA

Substituindo-se a estrutura do filtro preditor AR por um filtro do tipo ARMA[4], obteve-se ainda melhores resultados, conforme serão apresentados.

A primeira simulação corresponde a procurar-se o "melhor" filtro PCA-ARMA cuja ordem de complexidade em termos de número de operações necessárias na síntese do sinal de voz seja menor ou igual à do filtro PCA-AR com OD=10, ou seja, procurou-se um filtro PCA-ARMA com $OD + ON \leq 10$. Obteve-se os seguintes resultados:

$$EMQ_N = 0,207424$$

$$RSR = 6,83 \text{ dB}$$

$$\text{Ordem média} = 8,90$$

A Tabela 3.4 apresenta a estatística da ordem do preditor encontrada.

ON	OD	Número de Ocorrências
0	4	1
0	6	4
0	8	8
0	9	11
0	10	9
1	3	1
1	6	2
1	7	2
1	8	1
1	9	5
2	5	1
2	6	6
2	7	4
2	8	5
3	5	3
3	6	5
3	7	4
4	1	1
4	5	1
4	6	10
5	3	1
5	5	1
6	4	4
7	1	1
7	3	1
8	1	1
8	2	2
9	0	3
9	1	1
10	0	1
Total		100

TABELA 3.4 - Estatística da Ordem do Preditor para $ON + OD \leq 10$.

Estes resultados mostram que obteve-se uma relação sinal-ruído de 6,83 dB, 0,86 dB a mais do que a obtida para o filtro PCA-AR com OD=10. Além disso, a taxa média de bits para representar o filtro PCA-ARMA foi 11% menor do que o filtro PCA-AR com OD=10.

Pode-se também notar a diversidade de filtro PCA-ARMA, ou seja, um grande número de preditores diferentes ao longo de 100 quadros de voz.

Fazendo-se da mesma forma que no PCA-AR, foi procurado manter a ordem média de complexidade da síntese, melhorando-se a relação sinal-ruído, ou seja, para $ON + OD \leq 12$. Os resultados obtidos foram:

$$EMO_N = 0,196721$$

$$RSR = 7,06 \text{ dB}$$

$$\text{Ordem média} = 10,31$$

A Tabela 3.5 apresenta a estatística da ordem do preditor encontrada para $ON + OD \leq 12$.

ON	OD	Número de Ocorrências
0	4	1
0	6	3
0	8	5
0	9	8
0	10	1
0	11	2
0	12	7
1	3	1
1	7	2
1	9	2
1	10	3
2	5	1
2	6	2
2	7	1
2	9	4
2	10	6
3	5	2
3	6	4
3	7	2
3	8	2
3	9	5
4	1	1
4	5	1
4	7	3
4	8	3
5	6	1
5	7	3
6	4	1
6	6	5
7	4	3
7	5	2
8	3	1
8	4	2
9	2	2
10	1	3
10	2	3
11	0	1
11	1	1
Total		100

TABELA 3.5 - Estatística da Ordem do Preditor para $ON + OD \leq 12$.

A ordem de complexidade permaneceu aproximadamente a mesma, contudo a relação sinal-ruído aumentou de 1,09 dB.

Na próxima simulação, procurou-se manter a mesma relação sinal-ruído do PCA-AR com OD=10, ou seja, procurou-se um PCA-ARMA para cada quadro onde a soma de OD com ON fosse a menor possível e cuja relação sinal-ruído fosse pelo menos igual a do PCA-AR com OD=10. Com isso, obteve-se os seguintes resultados:

$$EMQ_N = 0,233974$$

$$RSR = 6,31 \text{ dB}$$

$$\text{Ordem média} = 6,40$$

A Tabela 3.6 apresenta a estatística da ordem do preditor encontrada.

ON	OD	Número de Ocorrências
0	1	3
0	2	1
0	3	2
0	4	16
0	5	5
0	6	9
0	7	8
0	8	6
0	9	6
0	10	9
1	1	1
1	3	1
1	4	2
1	6	2
1	7	1
1	8	1
1	9	1
2	1	1
2	4	4
2	5	2
2	6	2
2	7	1
3	0	1
3	1	1
3	3	1
3	4	1
3	5	2
3	6	1
4	0	1
4	1	2
4	6	2
6	4	1
7	0	1
8	1	1
9	1	1
Total		100

TABELA 3.6 - Estatística da Ordem do Preditor para $ON + OD \leq 10$ com $RSR \geq RSR$ do PCA - AR com $OD=10$.

Nesta situação a relação sinal-ruído foi de 6,31 dB, mas a ordem média foi de 6,40, ou seja, 36% a menos do que a proposta pela técnica da referência [1]. Isto implica numa economia de 36% na taxa média de bits para representar o filtro PCA-ARMA em relação ao que seria necessário para o filtro PCA-AR com OD=10, considerando-se que cada coeficiente do preditor fosse quantizado com igual número de bits.

3.4.4. RESULTADOS OBTIDOS COM UM FILTRO PCA - AR, UTILIZANDO-SE A NOVA TÉCNICA PROPOSTA NA DETERMINAÇÃO DA LOCALIZAÇÃO DOS IMPULSOS.

Neste item, as simulações foram efetuadas com a nova técnica de procura das localizações dos impulsos proposta neste trabalho.

No item A, fez-se a pesquisa com $N_f=3$ e $N_p=60$, ou seja, uma economia de 64% em termos de computação necessária para a localização dos impulsos.

No item B, fez-se a pesquisa com $N_f=2$ e $N_p=80$, ou seja, uma economia de 41% em termos de computação necessária para a localização dos impulsos.

A) Determinação da localização dos impulsos para $N_f = 3$ e $N_p = 60$.

Os seguintes resultados foram obtidos para um preditor PCA-AR com OD=10.

$$EMQ_N = 0,264752$$

$$RSR = 5,77 \text{ dB}$$

Comparando-se este resultado com o obtido com o PCA-AR com OD=10 utilizando-se o algoritmo da referência [1], obteve-se 0,20 dB a menos em termos da relação sinal-ruído, porém com um ganho de computação de 64%.

A seguir, apresenta-se o resultado de simulação para um preditor PCA-AR com $OD \leq 10$, ou seja, para cada quadro de voz procurou-se o preditor que tivesse a maior relação sinal-ruído.

Os resultados obtidos foram:

$$EMQ_N = 0,234922$$

$$RSR = 6,29 \text{ dB}$$

$$\text{Ordem média} = 7,86$$

A Tabela 3.7 apresenta a estatística da ordem do preditor encontrada.

OD	Número de Ocorrências
1	3
2	1
4	4
5	4
6	15
7	8
8	16
9	19
10	30
Total	100

TABELA 3.7 - Estatística da Ordem de Preditor para $OD \leq 10$, com $N_f=3$ e $N_p=60$.

A relação sinal-ruído obtida foi de 6,29 dB, 0,32 dB a mais do que a conseguida pelo algoritmo descrito no capítulo 2 referente à técnica apresentada em [1]. Obteve-se também uma menor taxa média de bits para representar o filtro PCA-AR. Esta economia foi de 21,4%, considerando-se igual número de bits para quantizar cada um dos coeficientes do preditor.

A seguir, procurou-se manter a mesma relação sinal-ruído do PCA-AR com OD=10, mas com a menor ordem média. Os resultados obtidos foram:

$$EMQ_N = 0,247877$$

$$RSR = 6,06 \text{ dB}$$

$$\text{Ordem média} = 6,85$$

A Tabela 3.8 apresenta a estatística da ordem do preditor encontrada.

OD	Número de Ocorrências
1	5
2	1
3	6
4	18
5	6
6	11
7	4
8	7
9	12
10	30
Total	100

TABELA 3.8 - Estatística da Ordem do Preditor para OD \leq 10 com a RSR \geq RSR do PCA-AR, OD=10, Nf=3 e Np=60.

Nota-se que a relação sinal-ruído ficou praticamente a mesma do filtro PCA-AR com OD=10, porém houve uma redução de 31,5% na taxa média de bits necessária para representar o filtro PCA-AR, considerando-se que os parâmetros do filtro preditor são quantizados com igual número de bits.

Para a próxima simulação, buscou-se aumentar a relação sinal-ruído mantendo-se a taxa média de bits para representar o filtro PCA-AR igual a da referência [1]. Esta simulação foi realizada para $OD \leq 13$, obtendo-se os seguintes resultados:

$$EMQ_N = 0,228875$$

$$RSR = 6,40 \text{ dB}$$

$$\text{Ordem média} = 9,21$$

A Tabela 3.9 apresenta a estatística da ordem do preditor encontrada.

OD	Número de Ocorrências
1	3
4	4
5	4
6	13
7	6
8	12
9	9
10	5
11	11
12	15
13	18
Total	100

TABELA 3.9 - Estatística da Ordem de Preditor para $OD \leq 13$, $N_f=3$ e $N_p=60$.

B) Determinação das localizações dos impulsos para $N_f = 2$ e $N_p = 80$.

Neste item, são refeitas todas as simulações apresentadas no item A, porém para $N_f=2$ e $N_p=80$. Como anteriormente mencionado, o número de pesquisas realizadas é 41% a menos do que o necessário em relação à técnica apresentada na referência [1], mantendo-se os resultados praticamente inalterados.

Para um preditor PCA-AR com OD=10, obteve-se:

$$EMQ_N = 0,256697$$

$$RSR = 5,91 \text{ dB}$$

Para o "melhor" filtro PCA-AR com OD \leq 10, obteve-se:

$$EMQ_N = 0,230708$$

$$RSR = 6,37 \text{ dB}$$

$$\text{Ordem média} = 7,98$$

A Tabela 3.10 apresenta a estatística da ordem do preditor encontrada.

OD	Número de Ocorrências
1	3
2	1
4	4
5	2
6	13
7	11
8	13
9	22
10	31
Total	100

TABELA 3.10 - Estatística da Ordem do Preditor para OD \leq 10, Nf=2 e Np=80.

Para o "melhor" filtro PCA-AR com OD \leq 10, procurando minimizar a ordem do preditor de tal forma que a relação sinal-ruído fosse pelo menos igual à obtida com o PCA-AR com OD=10, obteve-se os seguintes resultados:

$$EMQ_N = 0,240342$$

$$RSR = 6,19 \text{ dB}$$

$$\text{Ordem média} = 7,07$$

A Tabela 3.11 apresenta a estatística da ordem do preditor encontrada.

OD	Número de Ocorrências
1	3
2	1
4	22
5	4
6	10
7	5
8	9
9	12
10	31
Total	100

TABELA 3.11 - Estatística da Ordem do Preditor para $OD \leq 10$, $N_f=2$ e $N_p=10$ com $RSR = RSR$ para $OD=10$.

Para o "melhor" filtro preditor PCA-AR com $OD \leq 13$, obteve-se:

$$EMQ_N = 0,223552$$

$$RSR = 6,51 \text{ dB}$$

$$\text{Ordem média} = 9,43$$

A Tabela 3.12 apresenta a estatística da ordem do preditor encontrada.

OD	Número de Ocorrências
1	3
4	4
5	2
6	11
7	7
8	9
9	12
10	9
11	9
12	12
13	22
Total	100

TABELA 3.12 - Estatística da Ordem do Preditor para $OD \leq 13$, $N_f=2$ e $N_p=80$.

3.5. CONCLUSÕES

As Tabelas 3.13 a 3.15 apresentam um sumário dos resultados mais relevantes. Pode-se notar claramente as vantagens da utilização do PCA-ARMA no lugar do PCA-AR em termos de relação sinal-ruído e economia de bits para representar o PCA.

Da Tabela 3.13, verifica-se que utilizando-se um filtro PCA-ARMA com $OD + ON \leq 12$, ou seja, praticamente a mesma ordem média do PCA-AR com $OD=10$ consegue-se uma melhoria de 1,09 dB na relação sinal-ruído.

Também é interessante observar que procurando-se manter a mesma relação sinal-ruído com o filtro PCA-ARMA com $OD + ON \leq 10$, conseguiu-se uma economia de 36% em termos de taxa média de bits necessária para representar o filtro preditor.

As Tabelas 3.14 e 3.15 mostram que se conseguiu praticamente os mesmos resultados com uma economia de 64% e 41%, em termos de computação na determinação da localização dos impulsos da excitação multi-pulso.

PCA	EMQ_N	RSR	ORDEM MÉDIA
AR OD = 10	0,2530	5,97	10,00
AR OD \leq 10	0,2292	6,40	8,16
AR OD \leq 13	0,2234	6,51	9,65
AR OD \leq 10 e com RSR p/AR OD = 10	0,2379	6,24	7,12
ARMA OD + ON \leq 10	0,2074	6,83	8,90
ARMA OD + ON \leq 12	0,1967	7,06	10,31
ARMA OD + ON \leq 10 com RSR p/AR OD = 10	0,2340	6,31	6,40

TABELA 3.13 - Resultados do Método para a Determinação das Localizações Proposto por SINGHAL & ATAL[1].

PCA	EMQ_N	RSR	ORDEM MÉDIA
AR OD = 10	0,2648	5,77	10,00
AR OD \leq 10	0,2349	6,29	7,86
AR OD \leq 13	0,2289	6,40	9,21
AR OD \leq 10 e com RSR p/AR OD = 10	0,2479	6,06	6,85

TABELA 3.14 - Resultados do Novo Método para a Determinação das Localizações com $N_f=3$ e $N_p=60$.

PCA	EMQ _N	RSR	ORDEM MÉDIA
AR OD = 10	0,2567	5,91	10,00
AR OD ≤ 10	0,2307	6,37	7,98
AR OD ≤ 13	0,2236	6,51	9,43
AR OD ≤ 10 e com RSR p/AR OD = 10	0,2403	6,19	7,07

TABELA 3.15 - Resultados do Novo Método para a Determinação das Localizações com Nf=2 e Np=80.

C A P Í T U L O 4

CONCLUSÕES

Foi apresentado neste trabalho um estudo e algumas considerações a respeito da implementação de codificadores paramétricos de sinais de voz com excitação multi-pulso ("MPELPC").

A estrutura geral do codificador consiste de dois preditores cascadeados excitados por uma seqüência de impulsos.

O primeiro preditor denominado PLA (Preditor de Longo Atraso), tira proveito da forte correlação existente entre as amostras separadas por um período de "PITCH" do sinal de voz.

O segundo preditor, denominado PCA (Preditor de Curto Atraso), explora a importante correlação existente entre as amostras adjacentes.

A excitação consiste de uma seqüência de impulsos estrategicamente localizados de forma a minimizar o erro médio quadrático entre os sinais de voz original o sintetizado.

No Capítulo 2 foi apresentada uma revisão bibliográfica das principais técnicas de determinação dos referidos preditores e da excitação multi-pulso.

No Capítulo 3 foi apresentada uma nova estrutura para o filtro preditor, consistindo basicamente na inserção de zeros na sua função de transferência. Também foi proposta a busca do melhor preditor, em termos de sua ordem, para cada quadro de voz. Obteve-se com isso uma melhoria de até 1 dB na relação sinal-ruído em relação à obtida pela técnica apresentada na referência [1], mantendo-se praticamente a mesma complexidade computacional.

Também foi proposta uma nova técnica para a determinação das localizações dos impulsos da excitação multi-pulso. Com essa nova técnica, conseguiu-se diminuir em até 64% o número de pesquisas necessárias à determinação das localizações em relação à proposta na referência [1], mantendo-se uma relação sinal-ruído comparável.

Um maior número de simulações faz-se necessário para uma avaliação mais precisa dos resultados ora apresentados. Estas simulações não foram realizadas devido à necessidade de um computador de maior porte, não disponível durante o trabalho.

Para se ter uma idéia, são necessárias 200 horas de CPU em um microcomputador tipo IBM PC-AT, dispoendo de coprocessador aritmético e "clock" de 13 MHz, para analisar um segmento de somente 30 segundos de voz.

Com a experiência adquirida no desenvolvimento deste trabalho, surgiram diversas idéias interessantes no que diz respeito à melhoria do codificador de voz com excitação multi-pulso. A seguir serão relacionadas algumas dessas idéias, que poderão fazer parte de futuras investigações.

Como primeira idéia, ao invés de somente na primeira correlação cruzada, como proposto no Capítulo 3, definir-se um número N_f de localizações fixas, propõe-se "atualizar" a correlação cruzada e nesta etapa novamente fixar outras N_f localizações. Com isso a complexidade computacional diminuirá ainda mais em relação ao próprio algoritmo proposto neste trabalho.

Uma segunda idéia consiste em formular-se uma nova técnica de projeto dos PCA, considerando-se não a resposta do impulso, mas a resposta ao multi-pulso.

O procedimento poderia ser o seguinte: determina-se o PCA por um dos métodos apresentados neste trabalho; determina-se a excitação multi-pulso; e reprojeta-se o PCA, considerando a excitação multi-pulso encontrada. Pode-se fazer algumas iterações até que o erro atinja um valor pré-determinado.

Acredita-se que o algoritmo, assim proposto, convergiria levando, talvez, a muito bons resultados.

Também é necessário levar-se em consideração na concepção de novos codificadores de sinais de voz, o mecanismo humano de percepção da voz. Isso poderá melhorar significativamente os codificadores, tanto em termos de qualidade, quanto em redução da taxa de bits.

Por último, cabe colocar que uma análise da quantização dos coeficientes dos codificadores paramétricos de sinal de voz com excitação multi-pulso, apresentados neste trabalho, seria de grande valia.

Finalmente cabe colocar que um dado de grande importância, que serviria como referência para inúmeros trabalhos nesta área, seria a determinação da excitação ótima. No entanto, isto não foi possível realizar. Para se ter uma idéia da complexidade computacional requerida, estima-se que seriam necessários 231 dias de operação ininterrupta de um microcomputador do tipo IBM-PC-AT, dispendo de um coprocessador aritmético e "clock" 13 MHz, para a determinação da excitação ótima de um quadro de apenas 10ms de voz.

A excitação ótima, por sua vez, seria utilizada como parâmetro de comparação dos diversos métodos sub-ótimos propostos.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] SINGHAL, Sharad & ATAL, Bishnu S. Amplitude optimization and Pitch prediction in multipulse coders. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 37(3):317-27, mar. 1989.
- [2] JAYANT, N. S. High-quality coding of telephone speech and wideband audio. IEEE Communications Magazine, E.U.A., 28(1):10-20, Jan. 1990.
- [3] RABINER, Lawrence R. & SCHAFER, Ronald W. Digital Processing of Speech Signals. New Jersey, Prentice-Hall, 1978. 512p.
- [4] BAÚ, Norberto & SEARA, Rui & BERMUDEZ, José Carlos M. Design of IIR digital filters from time-domain specifications. Florianópolis, UFSC/LINSE, 1990. 7p. Artigo submetido ao IEEE International Symposium on Circuits and Systems, Singapore, June. 1991.
- [5] SUKKAR, Rafid & LoCICERO, Joseph L. & PICONE, Joseph W. Decomposition of the LPC excitation using the zincbasis functions. IEEE Transactions on Acoustics, Speech, and Signal Processing, E.U.A., 37(9):1329-41, set. 1989.

- [6] ATAL, B. S. & REMDE, J. R. "A new model of LPC excitation for producing natural-sounding speech at low bit rates. "In Proc. Int. Acoust., Speech, Signal Processing, Paris, 1980. p.614-17 apud SINGHAL, Sharad & ATAL, Bishnu S. Amplitude optimization and Pitch prediction in multipulse coders. IEEE Transaction on Acoustics, Speech, and Signal Processing, E.U.A, 37(3):317-27, mar. 1989.
- [7] ATAL, B. S. Linear predictive coding of speech, In Computer Speech Processing, Fallside and W. A. Woods, Eds. Englewood Cliffs, NJ: Prentice-Hall, 1985. apud SINGHAL, Sharad & ATAL, Bishnu S. Amplitude optimization and Pitch prediction in multipulse coders. IEEE Transaction on Acoustics, Speech, and Signal Processing, E.U.A, 37(3):317-27, mar. 1989.
- [8] TRONCOSO, Isabel M. & ATAL, Bishnu S. Efficient search procedures for selecting the optimum innovation in stochastic coders. IEEE Transaction on Acoustics, Speech, and Signal Processing, E.U.A., 38(3):385-96, mar. 1990.
- [9] KLEIJN, W. Bastiaan & KRASINSKI, Daniel & KETCHUM, Richard H. Fast methods for the CELP speech coding algorithm. IEEE Transaction on Acoustics, Speech, and Signal Processing, E.U.A., 38(8):1330-342, agos. 1990.
- [10] GALAND, C. & MENEZ, J. & ROSSO, M. Complexity reduction of CELP coders. In: INTERNATIONAL TELECOMMUNICATIONS SYMPOSIUM, Rio de Janeiro, 1990. Anais... Rio de Janeiro, SBT/IEEE, 1990. S.9, p.181-85.

- [11] YAMAMOTO, José S. & VIOLARO, Fábio. Investigation on vector quantized multipulse excitation with fixed pattern of positions. In: INTERNATIONAL TELECOMMUNICATIONS SYMPOSIUM, Rio de Janeiro, 1990. Anais... Rio de Janeiro, SBT/IEEE, 1990. S. 9, p.191-95.
- [12] FERRETTI, Marc. & CINARE, François. Synthese, Reconnaissance de la Parole. França, Éditests, 1983. 282p.
- [13] VARY, Peter. Speech Codec for the European Mobile Radio System. Revista Telecommunications, EUA, 51-4, set. 1988.
- [14] KROON, Peter & DEPRETTEERE, Ed F. & SLUYTER, Rob J. Regular-pulse excitation - A novel approach to effective and efficient multipulse coding of speech. IEEE Transaction on Acoustics, Speech, and Signal Processing, E.U.A., ASSP-34(5): 1054-63, out. 1986.
- [15] LEVINSON, N.: "The Wiener RMS error criterion in filter design and prediction", J. Math. Phys., Vol. 25, n^o4, pp. 261-278, 1947. apud ROCHA, Luis F. Prediccion lineal aplicada a se~nales de voz. Revista Telegrafica Electronica, Argentina, :962-69, set. 1979.
- [16] ROBINSON, E. A.: "Estatistical communication and detection". Ed. Hainer, N. Y., 1967. apud ROCHA, Luis F. Prediccion lineal aplicada a se~nales de voz. Revista Telegrafica Electronica, Argentina, :962-69, set. 1979.

- [17] DURBIN, J.: "The fitting of time series models", Rev. Inst. Int. Statist., Vol. 28, N^o3, pp. 233-243, 1960. apud ROCHA, Luis F. Prediccion lineal aplicada a señales de voz. Revista Telegrafica Electronica, Argentina, :962-69, set. 1979.
- [18] RAMACHANDRAN, Ravi P. & KABAL, Peter. Pitch prediction filters in speech coding. IEEE Transaction on Acoustics, Speech, and Signal Processing, E.U.A., 37(4):467-78, abr. 1989.
- [19] MORIKAWA, Hiroyoshi. Adaptive estimation of time-varying model order in the ARMA speech analysis. IEEE Transaction on Acoustics, Speech, and Signal Processing, E.U.A., 38(7): 1073-83, jul. 1990.
- [20] KABAL, Peter & RAMACHANDRAN, Ravi P. Joint optimization of linear predictors in speech coders. IEEE Transaction on Acoustics, speech, and Signal Processing, E.U.A., 37(5): 642-50, maio 1989.
- [21] NG, Chun Sum & MILENKOVIC, Paul H. Unstable covariance LPC solutions from nonstationary speech waveforms. IEEE Transaction on Acoustics, Speech, and Signal Processing, E.U.A., 37(5):651-54, maio 1989.

- [22] ATAL, B. S. & SCHROEDER, M. R. Predictive coding of speech signals and subjective error criteria. IEEE Transaction on Acoustics, Speech, Signal Processing, ASSP-27:247-54, jun. 1979. apud RAMACHANDRAN, Ravi P. & KABAL, Peter. Pitch prediction filters in speech coding. IEEE Transaction on Acoustics, Speech, and Signal Processing, E.U.A., 37(4):467-78, abr. 1989.
- [23] KUNT, M. Traitement numérique des signaux. 3.ed. Paris, Dunod, 1981.
- [24] JACKSON, Leland. Digital filters and signal processing. 2.ed. Boston, Kluwer Academic Publishers, 1989. 410p.
- [25] CHAPMAN, C. Vocoders - o que são, como funcionam, suas possibilidades. Revista Elektor, São Paulo, 41-46, 1989. 1a. parte.

BIBLIOGRAFIAS CONSULTADAS

- [26] OPPENHEIM, Alan V. & SCHAFER, Ronald W. Digital signal processing. New Jersey, Prentice-Hall, 1975. 585p.
- [27] COULON, Frederic de. Signal theory and processing. Washington, Artech House, 1986. 618p.
- [28] BABB, Michael. First DSP - Based Microcontroller Makes Debut. Revista Control Engennering, EUA :91, out. 1988.
- [29] BLUESTEIN, L. I. A linear filter approach to the computation of the discrete Fourier transform, 1968, Nerem Rec., pp. 218-19 apud BRIGHAM, E. O. The Fast Fourier Transform, New Jersey, Prentice-Hall, 1974, 195p.
- [30] BORLAND. Turbo Pascal 5.0. EUA, Borland, 1988.
- [31] CARLSON, Athol B. Sistemas de Comunicação. São Paulo, McGraw-Hill, 1981. 487p.
- [32] COX, Richard U. & CROCHIERE, Ronald E. & JOHNSTON, James. Time Domain Harmonic Scaling of Speech for Rate Modification and Coding. IEEE Transactions on Acoustics, Speech, and Signals Processing, EUA, ASSP-31(1) :258-72, fev. 1983.

- [33] DIGITAL SIGNAL PROCESSING COMMITTEE. Programs for Digital Signal Processing. New York, IEEE Press, 1979.
- [34] FELDMAN, Joel A. & HOFSTETTER, Edward M. & MALPASS, Marilyn. A Compact, Flexible LPC Vocoder Based on a Commercial Signal Processing Microcomputer. IEEE Transactions on Acoustics, Speech, and Signals Processing, EUA, ASSP-31 (1) :252-57, fev. 1983.
- [35] FELLDMAN, Ronald D. & BRODERSEN, Robert W. A Switched -Capacitor Adaptive Lattice Filter. IEEE Transaction on Acoustics, Speech and Signals Processing, EUA, ASSP-31 (1) :294-304, fev. 1983.
- [36] FETTE, Bruce & HARRISON, Dave & OLSON, Don & ALLEN, Steven. A Family of Special Purpose Micro programable Digital Signal Processor IC'S in an LPC Vocoder System. IEEE Transaction on Acoustics, Speech and Signals Processing, EUA, ASSP-31(1):273-81, fev. 1983.
- [37] GREGORIAN, Roubick. & AMIR, Gideon. A Single Chip Speech Synthesizer Using a Switched - Capacitor Multiplier. IEEE Transaction on Acoustics, Speech and Signals Processing, EUA, ASSP-31(1):313-23, fev. 1983.
- [38] HAGIWARA, Yoshimune & KITA, Yuzo & MIYAMOTO, Takanori & TOBA, Yoshitomi & HARA, Hideo & AKAZAWA, Takashi. A Single Chip Digital Signal Processor and its Application to Real-Time Speech Analysis. IEEE Transaction on Acoustics, Speech and Signals Processing, EUA, ASSP-31(1): 339-46, fev. 1983.

- [39] HESTER, Richard K. & TAN, Khen-Sang & HEWES, C. Robert. A monolithic data acquisition channel. IEEE Transaction on Acoustics, Speech and Signals Processing, EUA, ASSP-31 (1):305-13, fev. 1983.
- [40] INOUE, Kazuo & WAKABAYASHI, Kenji & YOSHIKAWA, Yoshinobu & MASUZAWA, Shigeaki & SANO, Kenji & KIMURA, Seiji. A Single CMOS Speech Synthesis Chip and New Synthesis Techniques. IEEE Transaction on Acoustics, Speech and Signals Processing, EUA, ASSP-31(1):335-38, fev. 1983.
- [41] IRIE, Kazunari & UNO, Takehiko & UCHIMURA, Kuniharu & IWATA, Atsushi. A single-Chip ADM LSI CODED. IEEE Transaction on Acoustics, Speech and Signals Processing, EUA, ASSP-31 (1):281-87, fev. 1983.
- [42] ISTRE, Giles L. Fonologia Transformacional e Natural. Florianópolis, UFSC, 1980. 266p.
- [43] JÚNIOR, Vidal P. S. Microcontroladores. São Paulo, Érica, 1988. 187p.
- [44] LIN, Sayon T. A Monolithic Audio Spectrum Analyser. IEEE Transaction on Acoustics, Speech and Signals Processing, EUA, ASSP-31(1):288-93, fev. 1983.
- [45] MAIA, Eleonora M. No Reino da Fala. São Paulo, Ática, 1985. 128p.
- [46] MALMBERG, Bertil. A Fonética. Lisboa, Presses Universitaires de France, 1954. 195p.

- [47] MONBARON, Jean J. & BUI, Ngoc & MICHEL, Jean G. An Integrated Voice Recognition System. IEEE Transaction on Acoustics, Speech and Signals Processing, EUA, ASSP-31 (1):323-29, fev. 1983.
- [48] O'NEILL, Mark A. Faster than Fast Fourier. Revista Byte, EUA, :293-300, abr. 1988.
- [49] RABINER, L. R & GOLD, B.. Parallel Processing Techniques for Estimating Pitch Periods of Speech in the Time Domain. The Journal of the Acoustical Society of America. V. 46, 442-48, agos. 1969.
- [50] RABINER, L. R. et alii. A comparative Performance Study of Several Pitch Detection Algorithms. IEEE Transaction on Acoustics, Speech and Signals Processing, EUA, ASSP-24 (10):399-417, out. 1976.
- [51] RAMIREZ, Boris E. DSP Front - end for Speech Recognition. Holanda, Philips, 1987. 149p. Tese de Mestrado.
- [52] ROCHA, Luiz F. Predicción Lineal Aplicada a señales de Voz. Revista Telegráfica Eletrônica, Argentina, 962-69, set. 1979. 1a. parte.
- [53] ————. Predicción Lineal Aplicada a señales deVoz. Revista Telegráfica Eletrônica, Argentina :1139-148, out. 1979. 2a. parte.
- [54] ————. Processamento de Voz. Argentina, Editorial Kapelusz, 1987. 85p.

- [55] SIEMENS. Microcomputer Components, SAB8051, Macro Assembler. EUA, Siemens, 1981.
- [56] SIEMENS. Microcomputer Components, SAB8051, Family of Single Chip Microcomputer. EUA, Siemens, 1981.
- [57] SILVEIRA, Regina C. P. Estudos de Fonética do Idioma Português. São Paulo, Cortez Editora, 1982. 115p.
- [58] TANAKA, Fuminari & SHIGEHARA, Hiroshi & SUZUKI, Hiroaki & SUZUKI, Yasoji & TAKIMOTO, Ichizoh. C2MOS Speech Synthesis Systems. IEEE Transaction on Acoustics, Speech and Signals Processing, EUA, ASSP-31(1):329-34, fev. 1983.
- [59] VIOLARO, F. Nova Versão do Sistema de Análise e Processamento Digital de Voz. In: SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES, 7, Florianópolis, 1989. Anais... Florianópolis, UFSC/LINSE, 1989. p.50-6.
- [60] WALDMAN, Helio. Processamento de Sinais Digitais. Argentina, La Editorial Kapelusz, 1987. 183p.
- [61] WILLEMS, L. F & HART, J. & NOOTEBOOM, S. G. & VOGTEN, L. L. M. Manipulação de Sons da Fala. Ciência e Tecnologia, 2(6):179-91, 1983.

A P Ê N D I C E 1

ALGORÍTMO DE LEVINSON-DURBIN PARA SOLUÇÃO DO SISTEMA DE EQUAÇÕES GERADAS PELO MÉTODO DA AUTOCORRELAÇÃO

A seguir, apresenta-se o algoritmo de LEVINSON-DURBIN e um exemplo mostrando sua aplicação. Este método é um procedimento recursivo para a partir das equações de autocorrelação, determinar os coeficientes do preditor linear. Sua vantagem em relação aos métodos tradicionais de solução de equações lineares é o aproveitamento da propriedade TOEPLITZ da matriz gerada, para diminuir o número de operações aritméticas na resolução das equações. O procedimento é o seguinte:

$$1. E^{(0)} = R(0)$$

$$2. k_i = \frac{1}{E^{(i-1)}} \cdot \left[R(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} \cdot R(i-j) \right] \quad \text{para } 1 \leq i \leq P \quad (A1.1)$$

$$3. a_i^{(i)} = k_i$$

$$4. a_j^{(i)} = a_j^{(i-1)} - k_i \cdot a_{i-j}^{(i-1)}, \quad j = 1.. i-1$$

$$5. E^{(i)} = (1 - k_i^2) \cdot E^{(i-1)}$$

E se repete o procedimento dos itens 2 a 5 para $i = 1, 2, \dots, OD$, com o final da solução dado por:

$$a_j = a_j(OD), \quad \text{para } j = 1, \dots, OD$$

Nota-se, também, que ao final da aplicação do procedimento, tem-se calculado os valores das constantes k . Estas constantes são conhecidas como coeficientes de reflexão dos tubos acústicos equivalentes das cavidades bucal e nasal.

A seguir apresentaremos um exemplo.

Seja um sinal com oito amostras:

$$s(n) = \{10,0; 3,5; 0,0; -1,8; -2,0; -1,2; 0,0; 0,9\}$$

onde deseja-se obter um filtro preditor AR com ordem $OD = 3$.

a) Cálculo da função autocorrelação

$$R_n(k) = \sum_{m=0}^{N-1-k} s(m) \cdot s(m+k) \quad \text{para } k = 0, 1, 2, \dots, OD - 1$$

$$R_n(0) = 121,74$$

$$R_n(1) = 41,00$$

$$R_n(2) = -5,22$$

$$R_n(3) = -26,80$$

Colocando-se na forma matricial, temos:

$$\begin{bmatrix} 121,74 & 41,00 & -5,22 \\ 41,00 & 121,74 & 41,00 \\ -5,22 & 41,00 & 121,74 \end{bmatrix} \times \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 41,00 \\ -5,22 \\ -26,80 \end{bmatrix}$$

Para $i = 1$, temos

$$E^{(0)} = R(0) = 121.74$$

$$k_1 = \frac{1}{E^{(0)}} \cdot [R(1) - 0] = 0.3367833087$$

$$a_1^{(1)} = k_1 = 0.3367833087$$

$$E^{(1)} = (1 - k_1^2) \cdot E^{(0)} = 107.9318843$$

Para $i = 2$, temos:

$$k_2 = \frac{1}{E^{(1)}} \cdot [R(2) - a_1^{(1)} \cdot R(1)] = -0.1762974471$$

$$a_2^{(2)} = k_2 = -0.1762974471$$

$$a_1^{(2)} = a_1^{(1)} - k_2 \cdot a_1^{(1)} = 0.3961573462$$

$$E^{(2)} = (1 - k_2^2) \cdot E^{(1)} = 104.5772761$$

Para $i = 3$, temos:

$$k_3 = \frac{1}{E^{(2)}} \cdot [R(3) - a_1^{(2)} \cdot R(2) - a_2^{(2)} \cdot R(1)] = -0.1673773115$$

$$a_3^{(3)} = k_3 = -0.1673773115$$

$$a_1^{(3)} = a_1^{(2)} - k_3 \cdot a_2^{(2)} = 0.3666491535$$

$$a_2^{(3)} = a_2^{(2)} - k_3 \cdot a_1^{(2)} = -0.1099896956$$

A P Ê N D I C E 2

PROGRAMA PARA DETERMINAÇÃO SUB-ÓTIMA DO GERADOR DE EXCITAÇÃO MULTI-PULSO

```

Program SINGHAL & ATAL;
{ Este programa determina as localizações e amplitudes sub-ótimas
do gerador de excitação multi-pulso
Parâmetros de entrada: Coeficientes a e b do PCA
Parâmetros de saída : neta e beta
}
Const
  Na          = 100;    { Número de amostras do quadro }
  Gama : Real  = 0.85   { Parâmetro Gama do filtro ponderador }
  OD         = 10      { Ordem do denominador do PCA }
  ON         = 0       { Orden do numerador do PCA }
  MP         = 10      { Máximo entre OD e ON }

Type
  tipo_quadro_aux      : array[-MP..Na-1] of Real;
  tipo_y               : tipo_quadro_aux;
  tipo_h               : tipo_quadro_aux;
  tipo_s               : tipo_quadro;
  tipo_a               : tipo_coeficientes;
  tipo_b               : tipo_coeficientes;
  tipo_quadro_real     : array[0..Na-1] of Real;
  tipo_alfa            : tipo_quadro_real;
  tipo_quadro_real_aux : array[1..Na] of Real;
  tipo_beta            : tipo_quadro_real_aux;
  tipo_c               : tipo_quadro_real_aux;
  tipo_quadro_inteiro  : array[1..Na] of integer;
  tipo_neta            : tipo_quadro_inteiro;

Var
  aux      : Real    { Variável auxiliar }
  max      : Real    { Variável auxiliar }
  soma     : Real    { Variável auxiliar }
  n,i,k,j,l : Integer { Variável inteira }
  y        : tipo_y  { Sinal de voz ponderado }
  h        : tipo_h  { Resposta ao impulso do PCA-AR ponderado }
  s        : tipo_s  { Sinal de voz original }
  a        : tipo_a  { Coeficientes do denominador }
  b        : tipo_b  { Coeficientes do numerador }
  alfa     : tipo_alfa { Vetor de autocorrelação }
  c        : tipo_c  { Vetor de correlação cruzada }
  c_aux    : tipo_c  { Cópia de c }
  neta     : tipo_neta { Vetor com a localização dos impulsos }
  beta     : tipo_beta { Amplitudes dos impulsos }
  beta_ot  : tipo_beta { Auxiliar para solução das equações }

```

```

function delta(a:integer):integer;
begin
  if a <> 0 then
    delta := 0
  else
    delta := 1;
end;

function potencia(a : real; b : integer):real;
var
  aux : real;
begin
  aux := exp(b*ln(abs(a)));
  if (a < 0) and ((b mod 2) = 0) then
    aux := - aux;
  potencia := aux;
end;

procedure gauss(
  ne : integer;    { número de equações }
  mc : tipo_alfa; { coeficientes dependentes }
  mi : tipo_c;     { coeficientes independentes }
  r  : tipo_c;     { solução do sistema de equações }
  e  : boolean     { Erro na solução do sistema de equações }
);
{ Este procedimento resolve o sistema de equações lineares pelo
método de Gauss }
end;

begin
{ **** zero o passado **** }
for -MP to 0 do
begin
  y[n] := 0;
  h[n] := 0;
  s[n] := 0;
end;

{ **** cálculo de y[n] **** }
for n := 0 to Na-1 do
begin
  aux := s[n];
  for k := 1 to od do
    aux := aux + a[k]*( s[n-k] - potencia(gama,k)*y[n-k] );
  y[n] := aux;
end;

```

```

{ **** cálculo de h[n] **** }
for n := 0 to Na-1 do
  begin
    aux := 0;
    for k := 1 to od do
      aux := aux - a[k] * potencia(gama,k)*h[n-k];
    for k := 0 to on do
      aux := aux + b[k] * delta(n-k);
    h[n] := aux;
  end;

{ **** Otimização em amplitude **** }

{ **** cálculo da matriz alfa **** }
for i := 0 to Na-1 do
  begin
    soma := 0;
    for n := i to Na-1 do
      soma := soma + h[n-i]*h[n];
    alfa[i] := soma;
  end;

{ **** cálculo do vetor c ***** }
for i := 0 to Na-1 do
  begin
    soma := 0;
    for n := i to Na-1 do
      soma := soma + y[n]*h[n-i];
    c[i+1] := soma;
  end;
c_aux := c;

{ procura dos impulsos }

for i := 1 to Na do
  neta[i] := i;

{ primeiro impulso }
i := 1;
j := i;
max := sqr ( c[1] ) / alfa[0];
for i := 2 to Na do
  begin
    aux := sqr ( c[i] ) / alfa[0];
    if max < aux then
      begin
        max := aux;
        j := i;
      end;
  end;
neta[j] := 1;
neta[1] := j;

```

```

beta[j] := c[j] / alfa[0];

{ outros impulsos }

for l := 2 to Número_impulsos do
  begin
    k := neta[l-1];
    for i := 1 to Na do
      begin
        j := neta[i];
        c[j] := c[j] - beta[k]*alfa[abs(k-j)];
      end;
    k := neta[l];
    j := 1;
    max := sqr(c[k])/alfa[0];
    for i := l+1 to Na do
      begin
        k := neta[i];
        aux := sqr(c[k])/alfa[0];
        if max < aux then
          begin
            max := aux;
            j := i;
          end;
      end;
    i := neta[l];
    neta[l] := neta[j];
    neta[j] := i;
    beta[neta[l]] := c[neta[l]] / alfa[0];
  end;

{ **** reotimização das amplitudes **** }

for i := 1 to número_impulsos do
  for j := 1 to número_impulsos do
    alfa_ot[i,j] := alfa[abs( neta[i] - neta[j] ) ];

for i := 1 to número_impulsos do
  c_ot[i] := c_aux[neta[i]];

gauss(número_impulsos, alfa_ot, c_ot, beta_ot, erro);

for i := 1 to número_impulsos do
  beta[i] := beta_ot[i];

end;

```

A P Ê N D I C E 3

PROGRAMA PARA DETERMINAÇÃO SUB-ÓTIMA DO GERADOR DE EXCITAÇÃO MULTI-PULSO COM O NOVO ALGORÍTMO DE DETERMINAÇÃO DAS LOCALIZAÇÕES DOS IMPULSOS

```

Program Determina_impulsos_por_Correlacao_Cruzada;
{ Este programa determina as localizações e amplitudes sub-ótimas
do gerador de excitação multi-pulso com o novo algoritmo proposto
Parâmetros de entrada: Coeficientes a e b do PCA
Parâmetros de saída : neta e beta
}
Const
  Na          = 100;      { Número de amostras do quadro }
  Gama : Real  = 0.85     { Parâmetro Gama do filtro ponderador }
  OD          = 10       { Ordem do denominador do PCA }
  ON          = 0        { Orden do numerador do PCA }
  MP          = 10       { Máximo entre OD e ON }

Type
  tipo_quadro_aux      : array[-MP..Na-1] of Real;
  tipo_y               : tipo_quadro_aux;
  tipo_h               : tipo_quadro_aux;
  tipo_s               : tipo_quadro;
  tipo_a               : tipo_coeficientes;
  tipo_b               : tipo_coeficientes;
  tipo_quadro_real     : array[0..Na-1] of Real;
  tipo_alfa            : tipo_quadro_real;
  tipo_quadro_real_aux : array[1..Na] of Real;
  tipo_beta            : tipo_quadro_real_aux;
  tipo_c               : tipo_quadro_real_aux;
  tipo_quadro_inteiro  : array[1..Na] of integer;
  tipo_neta            : tipo_quadro_inteiro;
  tipo_magnitude       : tipo_quadro_real_aux;

Var
  aux      : Real   { Variável auxiliar }
  max      : Real   { Variável auxiliar }
  soma     : Real   { Variável auxiliar }
  n,i,k,j,l : Integer { Variável inteira }
  y        : tipo_y { Sinal de voz ponderado }
  h        : tipo_h { Resposta ao impulso do PCA-AR ponderado }
  s        : tipo_s { Sinal de voz original }
  a        : tipo_a { Coeficientes do denominador }
  b        : tipo_b { Coeficientes do numerador }
  alfa     : tipo_alfa { Vetor de autocorrelação }
  c        : tipo_c { Vetor de correlação cruzada }
  c_aux    : tipo_c { Cópia de c }

```

```

neta      : tipo_neta  { Vetor com a localização dos impulsos }
beta      : tipo_beta  { Amplitudes dos impulsos }
beta_ot   : tipo_beta  { Auxiliar para solução das equações }
Numero_fixos : integer; { Numero de localizações fixas - Nf }
Numero_pesquisas : integer; { Numero de localizações para
                             pesquisas - Np }
Magnitude  : tipo_magnitude { Magnitude da fução de correla-
                             ção cruzada }
Magnitude_aux : tipo_magnitude { Auxiliar de Magnitude }

function delta(a:integer):integer;
begin
  if a <> 0 then
    delta := 0
  else
    delta := 1;
end;

function potencia(a : real; b : integer):real;
var
  aux : real;
begin
  aux := exp(b*ln(abs(a)));
  if (a < 0) and ((b mod 2) = 0) then
    aux := - aux;
  potencia := aux;
end;

procedure gauss(
  ne : integer;    { número de equações }
  mc : tipo_alfa; { coeficientes dependentes }
  mi : tipo_c;    { coeficientes independentes }
  r  : tipo_c;    { solução do sistema de equações }
  e  : boolean    { Erro na solução do sistema de equações }
);
{ Este procedimento resolve o sistema de equações lineares pelo
método de Gauss }
end;

begin
{ **** "zerar" o passado **** }
for -MP to 0 do
begin
  y[n] := 0;
  h[n] := 0;
  s[n] := 0;
end;

```

```

{ **** cálculo de y[n] **** }
for n := 0 to Na-1 do
  begin
    aux := s[n];
    for k := 1 to od do
      aux := aux + a[k]*( s[n-k] - potencia(gama,k)*y[n-k] );
    y[n] := aux;
  end;

{ **** cálculo de h[n] **** }
for n := 0 to Na-1 do
  begin
    aux := 0;
    for k := 1 to od do
      aux := aux - a[k] * potencia(gama,k)*h[n-k];
    for k := 0 to on do
      aux := aux + b[k] * delta(n-k);
    h[n] := aux;
  end;

{ **** Otimizacao em amplitude **** }

{ **** cálculo da matriz alfa **** }
for i := 0 to Na-1 do
  begin
    soma := 0;
    for n := i to Na-1 do
      soma := soma + h[n-i]*h[n];
    alfa[i] := soma;
  end;

{ **** cálculo do vetor c ***** }
for i := 0 to Na-1 do
  begin
    soma := 0;
    for n := i to Na-1 do
      soma := soma + y[n]*h[n-i];
    c[i+1] := soma;
  end;
c_aux := c;

{ procura do impulsos }

{ cálculo da magnitude }

for i := 1 to Na do
  magnitude[i] := abs(c_aux[i]);
magnitude_aux := magnitude;

```

```

{ acha os Np maiores }

for i := 1 to numero_pesquisas do
  begin
    max := magnitude[1];
    k := 1;
    for j := 2 to Na do
      begin
        aux := magnitude[j];
        if max < aux then
          begin
            max := magnitude[j];
            k := j;
          end;
        end;
      magnitude[k] := 0;
      neta[i] := k;
    end;

{ determinar os Nf primeiros contemplados }

l := numero_fixos + 1;
i := 1;
repeat
  j := neta[i];
  if j = 1 then
    if magnitude_aux[1] < magnitude_aux[2] then
      begin
        k := neta[1];
        neta[1] := neta[1];
        neta[1] := k;
        inc(l);
        dec(i);
      end;
    if j = Na-1 then
      if magnitude_aux[Na-1] < magnitude_aux[Na-2] then
        begin
          k := neta[1];
          neta[1] := neta[Na-1];
          neta[Na-1] := k;
          inc(l);
          dec(i);
        end;
    if (j > 1) and (j < Na-1) then
      begin
        aux := magnitude_aux[j];
        if (magnitude_aux[j-1] > aux) or
          (aux < magnitude_aux[j+1]) then
          begin
            k := neta[1];
            neta[1] := neta[i];
            neta[i] := k;
            inc(l);
          end;
      end;
  end;
until i = numero_pesquisas;

```

```

        dec(i);
    end;
end;
inc(i);
until (i > numero_fixos) or (l > numero_pesquisas);

{ **** }
{ primeiro impulso }

j := neta[1];
beta[j] := c[j] / alfa[0];

{ outros impulsos }

for l := 2 to numero_fixos do
    begin
        k := neta[l-1];
        for i := 1 to Numero_pesquisas do
            begin
                j := neta[i];
                c[j] := c[j] - beta[k]*alfa[abs(k-j)];
            end;
        beta[neta[l]] := c[neta[l]] / alfa[0];
    end;
for l := numero_fixos+1 to Numero_impulsos do
    begin
        k := neta[l-1];
        for i := 1 to Numero_pesquisas do
            begin
                j := neta[i];
                c[j] := c[j] - beta[k]*alfa[abs(k-j)];
            end;
        k := neta[l];
        j := 1;
        max := sqr(c[k])/alfa[0];
        for i := l+1 to Numero_pesquisas do
            begin
                k := neta[i];
                aux := sqr(c[k])/alfa[0];
                if max < aux then
                    begin
                        max := aux;
                        j := i;
                    end;
            end;
        i := neta[l];
        neta[l] := neta[j];
        neta[j] := i;
        beta[neta[l]] := c[neta[l]] / alfa[0];
    end;
end;

```

```
{ **** reotimizacao das amplitudes **** }  
  
for i := 1 to número_impulsos do  
  for j := 1 to número_impulsos do  
    alfa_ot[i,j] := alfa[abs( neta[i] - neta[j] ) ];  
  
  for i := 1 to número_impulsos do  
    c_ot[i] := c_aux[neta[i]];  
  
  gauss(número_impulsos, alfa_ot, c_ot, beta_ot, erro);  
  
  for i := 1 to número_impulsos do  
    beta[i] := beta_ot[i];  
  
end.
```