

BC FD-1419
MIGUEL ARJONA RAMÍREZ

CODIFICADOR PREDITIVO DE VOZ

POR

ANÁLISE-MEDIANTE-SÍNTESE

Dissertação apresentada à Escola
Politécnica da Universidade de
São Paulo para obtenção do
título de Mestre em Engenharia.

Área de Concentração:
Engenharia de Sistemas Eletrônicos

Orientador:
Normonds Alens

São Paulo
1992

01419

DEDALUS - Acervo - EPBC



31200029227

92/110/SISTEMAS ELETRÔNICOS

fl. 109



UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA

TERMO DE JULGAMENTO
DE
DEFESA DE DISSERTAÇÃO DE MESTRADO

Aos 18 dias do mês de dezembro de 1992, às 16:00 horas,
no Departamento de Engenharia Eletrônica

da Escola Politécnica da Universidade de São Paulo, presente a Comissão Julgadora, integrada
pelos Senhores Professores Drs. Normonds Alens, Orientador do candidato,
Oswaldo Catsumi Imamura e Geraldo Lino de Campos

iniciou-se a Defesa de Dissertação de Mestrado do Senhor _____
MIGUEL ARJONA RAMÍREZ

Título da Dissertação: "Codificador Preditivo de Voz por Análise-Mediante-
Síntese"

Concluída a arguição, procedeu-se ao julgamento na forma regulamentar, tendo a Comissão
Julgadora atribuído ao candidato as seguintes notas:

<u>Prof.Dr.Normonds Alens</u>	<u>(10.0)</u>	<u>(dez)</u>
<u>Prof.Dr.Oswaldo Catsumi Imamura</u>	<u>(9.5)</u>	<u>(boa e mais)</u>
<u>Prof.Dr.Geraldo Lino de Campos</u>	<u>(10.0)</u>	<u>(dez)</u>

Para constar, é lavrado o presente termo, que vai assinado pela Comissão Julgado
ra e pelo Secretário da Seção de Pós-Graduação

São Paulo, 18 de dezembro de 1992.

Presidente: [Assinatura]

Secretário: Mara Fátima de Jesus Luz Sanchez [Assinatura]

Observações: _____

Homologada pela C.P.G. em reunião realizada a 08/02/1993

A Mariko Kawamoto, minha esposa,
pelo constante incentivo à
execução deste trabalho.

Agradecimentos

Ao amigo e orientador Prof. Dr. Normonds Alens pelas sugestões indispensáveis à manutenção da objetividade no decorrer deste trabalho e pelas indagações que tanto ajudaram na identificação de princípios fundamentais.

Ao colega Mário Minami pelas discussões sobre predição linear.

À Fundação de Amparo à Pesquisa do Estado de São Paulo pela concessão de Bolsa de Mestrado.

Sumário

Resumo	VIII
“Abstract”	IX
1 CODIFICAÇÃO DA VOZ	1
1.1 Introdução	1
1.2 Representação digital	2
1.3 Codificação	3
2 PREDIÇÃO LINEAR DO SINAL DE VOZ	4
2.1 Fundamentos espectrais da predição linear	4
2.2 O modelo LPC de síntese de voz	7
2.3 O modelo LPC de análise de voz	10
2.4 Estimação dos coeficientes do filtro LPC	11
2.5 Espectro do erro de predição	15
3 ANÁLISE MEDIANTE SÍNTESE	19
3.1 Analisadores espectrais	20
3.2 Quantização em malha fechada: ADPCM e APC	21
3.3 Filtros LPC excitados por códigos (CELP)	23
3.4 Busca da excitação CELP	25
3.5 Efeitos perceptivos na audição	27
3.6 Ponderação espectral perceptiva	27
3.7 Matriz de resposta impulsiva	30
3.8 Construção dos livros de códigos da excitação	32
3.8.1 Livro de excitações atrasadas	32
3.8.2 Livros de códigos ceifados centralmente	34
3.9 Busca em dois livros de códigos	35
3.9.1 Buscas seqüenciais simples	35
3.9.2 Buscas seqüenciais com determinação conjunta dos ganhos	37
3.9.3 Buscas seqüenciais com ortogonalidade	37
3.10 Excitação por multipulsos (MPE)	39
3.11 Excitação por soma vetorial (VSELP)	40
4 METODOLOGIA	42
4.1 Relação sinal-ruído segmentada	43
4.2 Cepstro da resposta impulsiva do filtro LPC	43

4.3	Distorção cepstral LPC	45
4.4	Medida subjetiva de qualidade da voz	48
4.5	Estimação objetiva da qualidade subjetiva	49
4.6	Orientação da pesquisa	49
5	CODIFICADOR PROPOSTO: CELP-SVD	52
5.1	Representação SVD do filtro LPC e da sua excitação	53
5.2	Representações do processo de filtragem	55
5.3	Propriedades espectrais dos vetores singulares	56
5.4	Critérios de seleção dos vetores singulares	57
5.5	Codificador proposto	61
5.6	Busca SVD da excitação	62
5.6.1	Representação SVD baseada no filtro expandido	63
5.6.2	Representação SVD baseada no filtro LPC	64
5.7	Representação da periodicidade	66
5.8	Testes de representação SVD	69
5.9	Testes de configuração	71
5.10	Testes comparativos	74
5.11	Testes de critérios de seleção	74
5.12	Testes de quantização	77
6	CONCLUSÃO	81
A	Produto interno de polinômios	83
B	Resolução vetorial das equações normais	84
C	Coefficientes LSP do filtro LPC	91
D	Expansão de faixa de filtros LPC	93

Lista de Figuras

1.1	Transmissor digital	2
1.2	Receptor digital	2
1.3	Comunicação digital	2
2.1	Espectro logarítmico de potência de um segmento de voz sonora e sua envoltória	5
2.2	Segmento de voz sonora	6
2.3	Processo de análise LPC	7
2.4	Síntese de um período fundamental de voz sonora	7
2.5	Espectro logarítmico de potência de um segmento de voz surda e sua envoltória	8
2.6	Sintetizador LPC	8
2.7	Segmento de voz surda	9
2.8	Branqueamento do segmento de voz $s(n)$ pelo seu filtro LPC inverso .	10
2.9	Filtragem do segmento de voz $s(n)$ pelo seu preditor LPC	10
2.10	Erro de predição para um segmento de voz sonora	16
2.11	Espectro logarítmico de potência do erro de predição para um seg- mento de voz sonora	17
2.12	Erro de predição para um segmento de voz surda	17
2.13	Espectro logarítmico de potência do erro de predição para um seg- mento de voz surda	18
3.1	Princípio da análise-mediante-síntese	19
3.2	Localização de formantes pelo método da análise-mediante-síntese . .	20
3.3	Codificador com quantização em malha fechada	21
3.4	Decodificador diferencial	21
3.5	Síntese CELP de um segmento de voz	24
3.6	Procedimento básico de busca da excitação	26
3.7	Busca da excitação por análise mediante síntese	26
3.8	Expansão de faixa de um filtro LPC	28
3.9	Ponderação espectral associada a um filtro LPC	29
3.10	Busca com comparação no domínio ponderado	29
3.11	Busca com filtro descarregado	31
3.12	Atualização de estados do codificador CELP	32
3.13	Codificador CELP com livro de códigos atrasados	33
3.14	Busca seqüencial com vetor-alvo u e vetor-alvo residual r	36

Lista de Tabelas

3.1	Características de codificadores ADPCM e APC	23
4.1	Escalas de qualidade do índice MOS	48
4.2	Codificadores CELP de várias técnicas de codificação	51
5.1	Codificadores para teste das representações SVD	70
5.2	Codificadores CELP-SVD identificados pela seleção de vetores	71
5.3	Codificadores CELP usados nos testes comparativos.	75
5.4	Codificadores para testes de seleção de RSVs	75
5.5	Distribuição de bits entre os parâmetros do codificador CELP-SVD	78
5.6	Codificadores CELP-SVD quantizados	78

3.15	Buscas seqüenciais com ortogonalidade	39
3.16	Decodificador de voz VSELP	41
4.1	Envoltória espectral LPC-10 logarítmica de um segmento sonoro	47
4.2	Qualidade subjetiva de voz digitalizada	49
4.3	Teste de técnicas de codificação e busca da excitação	51
5.1	Espectros de RSVs de filtro passa-baixas	57
5.2	Espectros de RSVs de filtro passa-baixas	58
5.3	Espectros de RSVs de filtro ressonante	58
5.4	Análise preditiva no CELP-SVD	61
5.5	Busca da excitação no CELP-SVD	62
5.6	Testes de representação SVD	69
5.7	Testes de representação SVD	70
5.8	Testes de representação SVD	70
5.9	Testes de configuração do codificador CELP-SVD	71
5.10	Envoltórias dos espectros logarítmicos de um segmento do sinal de voz	72
5.11	Envoltórias dos espectros logarítmicos de um segmento do sinal de voz	73
5.12	Testes comparativos do codificador CELP-SVD com os codificadores	74
5.13	Testes comparativos do codificador CELP-SVD com os codificadores	74
5.14	Testes de seleção de vetores singulares	76
5.15	Testes de seleção de vetores singulares	76
5.16	Testes de seleção de vetores singulares	76
5.17	Níveis dos quantizadores de ganho	79
5.18	Testes de quantização	79
5.19	Testes de quantização	80
A.1	Interpretação do produto interno.	83
B.1	Filtro LPC inverso de ordem $p = 2$ na estrutura em treliça.	89
B.2	Filtro LPC de ordem $p = 2$ na estrutura em treliça.	89
D.1	Diagrama de pólos de um filtro com ressonância estreita (fora de escala)	94

Abstract

Analysis-by-synthesis linear predictive speech coders are widely applied in mobile and secure telecommunications.

Linear prediction of speech signals and analysis-by-synthesis techniques are presented so that some perceptual features of human hearing may be related to signal processing techniques and parameters.

The basic operation of this class of coders is described in the framework of the code-excited predictive coder. Special coder structures such as adaptive, sparse and vector-basis codebooks are introduced as well as processing enhancements such as orthogonal searches.

A recently introduced representation of voice excitation is complemented by adaptive codebook searches to give rise to the new proposed coder, the singular-vector-decomposed excitation linear predictive coder.

The results of a study of some important coders in this class is present. The coders are compared on the basis of waveform and spectral objective distortion measures.

A further study of spectral selection of excitation features, and quantization of the whole set of parameters is performed on the proposed coder. Some interesting results are described concerning the adaptive spectral representation and the sensitivity to quantization of the excitation features.

ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO
Serviço de Bibliotecas

FOLHA GUIA PARA ELABORAÇÃO DE RESUMO

AUTOR: ARJONA RAMIREZ, *deignel*
TÍTULO: *Codificador preditivo de voz por análise - mediante ...*

COL	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	3	0	1	2	3	4	5	6	7	8	9	0	4	
L ₁	D	S	C	O	D	I	F	I	C	A	D	O	R	E	S	P	R	E	D	I	T	I	V	O	S	D	E	V	O	Z	P	O	R	A

Capítulo 1

CODIFICAÇÃO DA VOZ

1.1 Introdução

A voz manifesta-se fisicamente através de uma onda acústica. Porém, a voz constitui-se essencialmente de um conjunto de informações para as quais a onda acústica é um canal de transmissão.

A prova mais concreta da natureza informativa da voz está na sua representação escrita, que preserva a informação fonética, sintática e léxica da voz, embora não consiga registrar simultaneamente a prosódia na sua integridade nem a identificação vocal da pessoa que produziu a informação.

A representação escrita da voz é uma codificação, podendo ser facilmente associada a uma seqüência de bits ou unidades de informação¹.

Para termos uma noção quantitativa do fluxo de informação escrita numa leitura, podemos imaginar uma língua hipotética com 32 fonemas de ocorrência equiprovável e que possam ser pronunciados à taxa de 10 por segundo. Assim, cada fonema pode ser representado por 5 bits e a informação transmitida na leitura é de 50 b/s [14].

Por outro lado, o homem também procurou transmitir ou registrar a voz na sua manifestação acústica. Os processos mais práticos com esta finalidade envolvem a conversão da onda acústica em um sinal elétrico, que se manifesta em variações de pressão ou de corrente *análogas* às variações de pressão do ar. Esta é essencialmente a “especificação de projeto” do aparelho telefônico seguida por Alexander Graham Bell [14].

A capacidade de transmissão de informação de um canal telefônico depende da largura de faixa das freqüências de sinal a que ele consegue dar passagem e do seu nível de ruído, com o qual o sinal terá que “competir” para se fazer notar. Para valores típicos de relação sinal-ruído de 30 dB e de faixa de passagem de 3000 Hz, obtemos uma taxa de informação de 30 kb/s.

Na prática, prevendo canais telefônicos com melhores condições de ruído, está estabelecido um padrão de telefonia digital à taxa de 64 kb/s. Este padrão é o PCM (“Pulse Code Modulation”) [23].

¹Essa associação entre símbolos e bits é qualificada como “sem ruído” ou “sem distorção” na teoria da informação [23].



Figura 1.1: Transmissor digital



Figura 1.2: Receptor digital

1.2 Representação digital

A codificação digital básica de um sinal analógico envolve um amostrador (Fig. 1.1), caracterizado por uma frequência de amostragem f_a , que toma como amostras os valores do sinal em instantes separados por intervalos $T = 1/f_a$. Para aplicações telefônicas, $f_a = 8$ kHz.

Para não causar a distorção irreversível do recobrimento espectral (“aliasing”) [37], o sinal analógico deve ter sua frequência máxima f_c limitada à metade da frequência de amostragem, ou seja, $f_c \leq f_a/2$. Essa é a função do filtro anti-recobrimento da Fig. 1.1. Em telefonia, $f_c = 3,4$ kHz.

Se houver apenas ruído presente acima de f_c , o sinal pode ser completamente recuperado a partir da seqüência de valores amostrados pelo teorema de Nyquist [37].

Na sua realização mais simples, o codificador (Fig. 1.1) associa a cada amostra um valor escolhido dentre um conjunto finito de níveis [23]. Neste caso, ele é apenas um quantizador escalar do sinal.

Na quantização uniforme, o conjunto de níveis, após a ordenação, apresenta a mesma distância entre níveis consecutivos.

Nos codificadores digitais de que trataremos, assumimos as amostras de voz recebidas em quantização uniforme e representadas com 12 bits no mínimo. Nesses casos, podemos considerar o sinal digital praticamente de precisão infinita, pois outras características dos codificadores causam distorções maiores.

Na reconstrução do sinal analógico (Fig. 1.2), o decodificador tem que realizar analogicamente os níveis codificados, sendo seguido de um filtro de reconstrução [37].

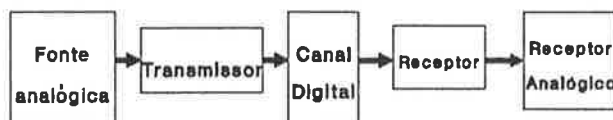


Figura 1.3: Comunicação digital

1.3 Codificação

Os codificadores manipulam a *redundância* e a *irrelevância* dos sinais [23].

A redundância de um sinal é a repetição de informação causada pelo processo de geração do sinal na fonte. Portanto, poderemos reduzir a redundância de um sinal se conhecermos leis ou restrições que regem ou condicionam a sua geração. As técnicas para a remoção da redundância são o *modelamento da fonte* e a *predição do sinal*.

Duas das principais causas de redundância do sinal de voz, que serão abordadas no Capítulo 2, são [3]:

- i) sua quase-periodicidade nos segmentos sonoros;
- ii) a característica não plana da envoltória do seu espectro de curto prazo.

A irrelevância de um sinal é a informação representada pelo excesso de precisão que ele transporta além da sensibilidade do receptor. A técnica para a redução da irrelevância é a *quantização* de amostras do sinal (seção 1.2) ou de parâmetros da fonte analógica do sinal (Fig. 1.3). A quantização pode ser escalar [23] ou vetorial [32].

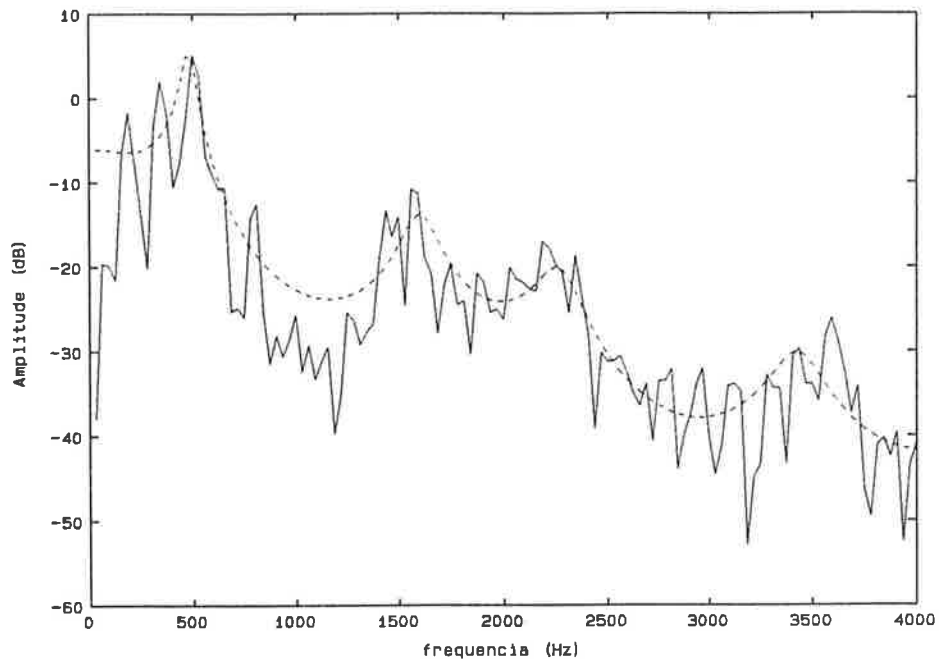


Figura 2.1: Espectro logarítmico de potência de um segmento de voz sonora e sua envoltória

- a_k , $k = 1, \dots, p$ são os coeficientes de predição;
- p é a ordem do modelo de predição linear e
- o filtro inverso é

$$A(z) = \sum_{k=1}^p a_k z^{-k}. \quad (2.2)$$

Para ajustar o espectro do modelo ao espectro do sinal modelado, devemos multiplicar a resposta em frequência $H_{LPC}(e^{j\omega})$ por

- G : ganho ou coeficiente de ajuste entre os espectro do modelo e do sinal.

As frequências onde ocorrem os picos observados na envoltória espectral (Fig. 2.1) correspondem às *formantes* ou frequências de ressonância do aparelho fonador. A densidade espectral de formantes é aproximadamente 1/kHz [38].

Um modelo LPC de ordem p tem capacidade para representar até $p/2$ formantes na faixa de frequência de 0 a $f_a/2$, onde f_a é a frequência de amostragem. Assim, para conseguir modelar a estrutura de formantes, a ordem p do modelo LPC deve ser escolhida com a condição

$$p > \frac{f_a}{1 \text{ kHz}}. \quad (2.3)$$

A estrutura fina do espectro de curto prazo de um segmento de voz sonora é harmônica e representa sua quase-periodicidade (Fig. 2.1). Um segmento periódico apresenta raias espectrais em harmônicas da frequência fundamental.

Capítulo 2

PREDIÇÃO LINEAR DO SINAL DE VOZ

A técnica de predição linear é a mais usada para análise e síntese de sinais de voz desde sua introdução nesta área em 1966.

Os primeiros pesquisadores que aplicaram a predição linear ao processamento da voz foram Saito e Itakura [40] *apud* [33] e Atal e Schroeder [2] *apud* [33].

Estas aplicações da predição linear ao modelamento da fonte de informação do sinal de voz (seção 1.3) são denominadas “linear predictive coding” (LPC).

2.1 Fundamentos espectrais da predição linear

Para análise do sinal de voz, segmentos com duração de algumas dezenas de milissegundos (tipicamente 20 ms) são considerados estacionários. Este princípio pode ser justificado fisiologicamente porque a configuração geométrica dos órgãos fonadores não sofre alterações significativas durante esse intervalo de tempo.

A voz pode, em consequência, ser caracterizada pelo seu *espectro de curto prazo*. Na Fig. 2.1 temos o espectro de potência logarítmico de um segmento de 20 ms de uma vogal.

É conveniente construir um modelo que retenha apenas as características espectrais mais marcantes.

O modelo de predição linear adapta uma envoltória ao espectro de curto prazo do sinal modelado (Fig. 2.1). Essa envoltória acompanha as tendências globais da curva espectral, com ênfase especial para os seus picos.

O bom modelamento dos picos espectrais pela predição linear decorre da sua realização através do módulo da resposta em frequência de um filtro com pólos apenas, cuja função de transferência é

$$H_{LPC}(z) = \frac{1}{1 + \sum_{k=1}^p a_k z^{-k}}, \quad (2.1)$$

onde

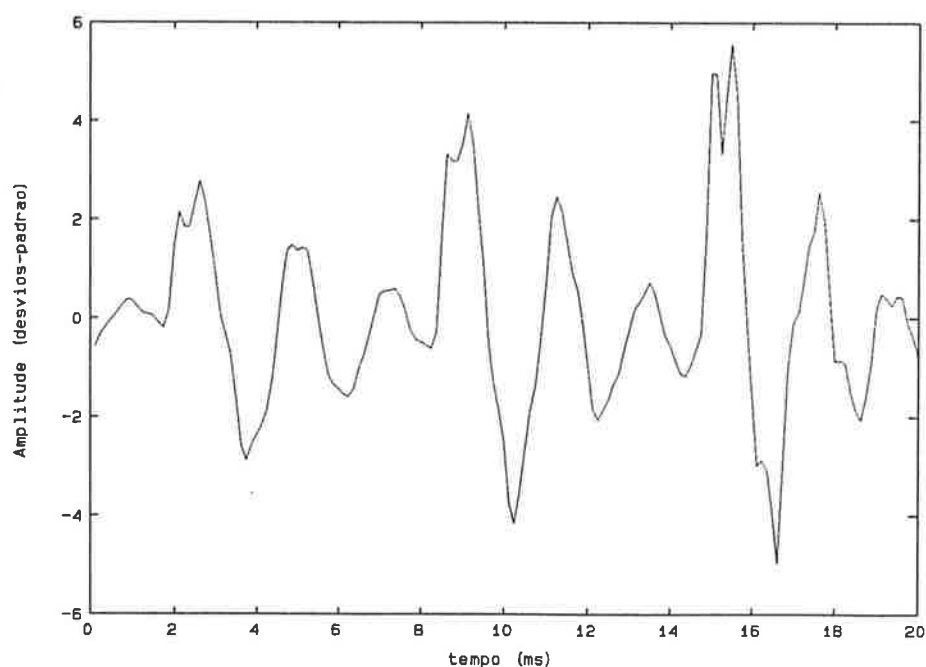


Figura 2.2: Segmento de voz sonora

A estrutura espectral harmônica corresponde à periodicidade da forma de onda do segmento de voz sonora (Fig. 2.2). O período fundamental ou período de “pitch” é

$$P = \frac{1}{F_0},$$

onde F_0 é a frequência fundamental.

O núcleo do processo de análise LPC de um sinal de voz envolve um *analisador preditivo* e um *detetor de “pitch”*, que geralmente são processos independentes (Fig. 2.3).

Os segmentos usados pelo analisador preditivo são derivados do sinal digital de voz após a sua passagem pelo *segmentador* seguida da aplicação de uma *janela* [37]. Quando a análise LPC é executada pelo método da autocorrelação (seção 2.4), costuma-se usar janelas de Hamming justapostas ou sobrepostas.

O processo de análise LPC cujo segmentador gera segmentos coincidentes com os ciclos fundamentais dos sons sonoros em duração e em fase qualifica-se de “pitch synchronous” [33]. Nestes casos, não é utilizada propriamente nenhuma janela, embora, analiticamente, a segmentação por si só corresponda à aplicação de uma janela retangular adaptativa com a duração ajustada instantaneamente a cada período fundamental.

O modelo espectral LPC representa com fidelidade o espectro total do segmento de voz sonora se duas condições forem satisfeitas:

- i) a segmentação for “pitch synchronous”;
- ii) a resposta impulsiva do aparelho fonador tiver um decaimento acentuado no decorrer de um período fundamental.

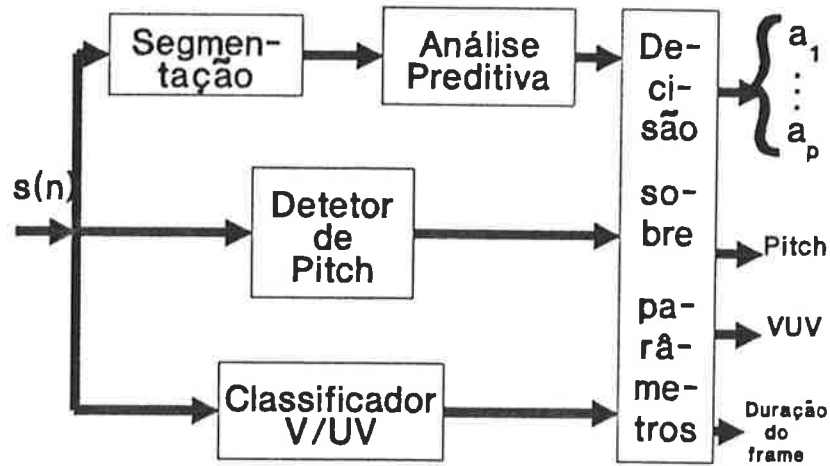


Figura 2.3: Processo de análise LPC

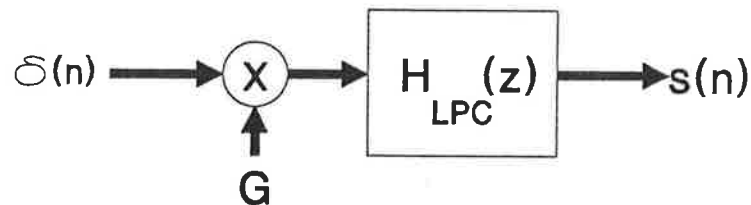


Figura 2.4: Síntese de um período fundamental de voz sonora

2.2 O modelo LPC de síntese de voz

Vimos na seção 2.1 que a envoltória de um segmento de voz

$$s(n), \quad n = 0, \dots, N - 1$$

sonora pode ser representada pela resposta em frequência do filtro $H_{LPC}(z)$ (Eq.(2.1)). Também mencionamos que a envoltória espectral é coincidente com o espectro total do segmento de voz se sua posição e sua duração forem as de um período fundamental.

Assim, para sintetizar um segmento de voz sonora que foi analisada por um processo “pitch synchronous”, devemos excitar o filtro $H_{LPC}(z)$ com um impulso unitário amplificado com ganho G (Fig. 2.4).

Mesmo quando os sons sonoros não são analisados de forma “pitch synchronous”, no processo de síntese LPC usa-se um gerador de pulsos¹ com o período fundamental.

Por outro lado, os espectros dos sons surdos não são harmônicos, embora sua envoltória espectral possa ser modelada por LPC (Fig. 2.5).

Para os sons surdos, excita-se o filtro $H_{LPC}(z)$ com um processo de ruído branco de média nula e de variância unitária amplificado com ganho G (Fig. 2.6).

¹Nas realizações práticas de sintetizadores LPC, é comum usar-se um gerador periódico com forma de onda não impulsiva. O principal inconveniente do trem de pulsos é a sua média não nula.

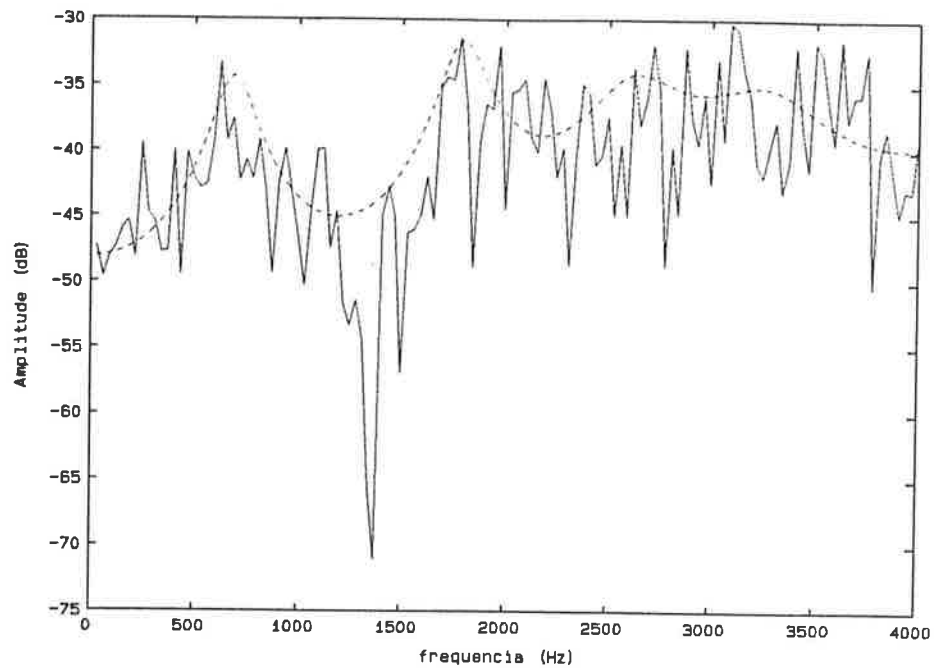


Figura 2.5: Espectro logarítmico de potência de um segmento de voz surda e sua envoltória

O modelo LPC da fonte de excitação inclui um comutador V/UV (Fig. 2.6) para selecionar o gerador conforme a natureza do som seja

- sonora ou “voiced” (V);
- surda ou “unvoiced” (UV).

Por sua vez, o analisador LPC tem que incluir um classificador dos blocos de voz em sonoros ou surdos.

Há outras possibilidades de combinar a excitação aleatória com a excitação periódica visando a uma representação mais fiel dos sons que não podem ser classificados puramente nem como sonoros nem como surdos.

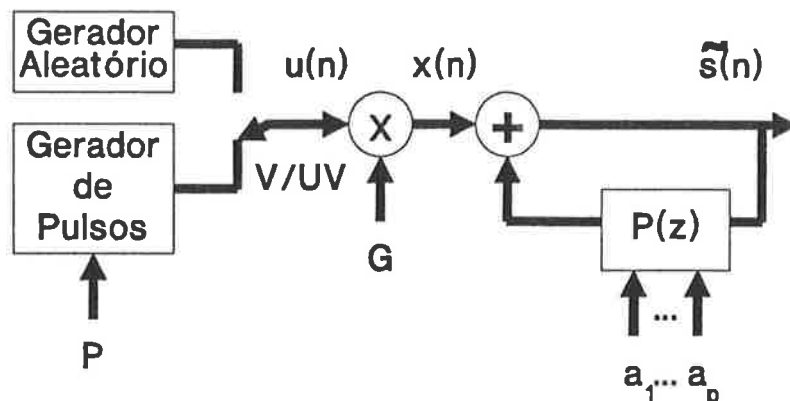


Figura 2.6: Síntese LPC

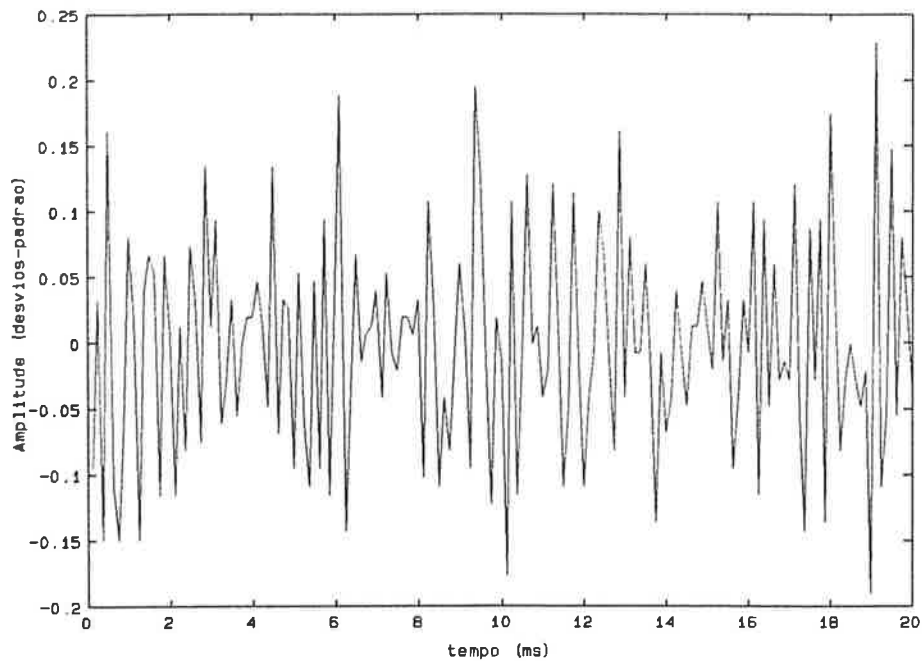


Figura 2.7: Segmento de voz surda

Para poder representar com mais fidelidade sons com características sonoras e surdas, Itakura e Saito em 1968 [33] no primeiro “vocoder”² LPC usaram um somador no lugar do comutador V/UV . Além disso, os seus geradores possuem amplificadores independentes com ganhos U e UV . É imposta a condição

$$V + UV = 1,$$

de forma que esta fonte de excitação reduz-se àquela que apresentamos nos casos de sons sonoros puros ou surdos puros.

O filtro $H_{LPC}(z)$, excitado pelo sinal $x(n) = Gu(n)$ (Fig. 2.6), produz

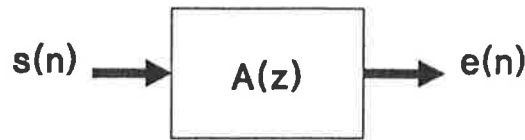
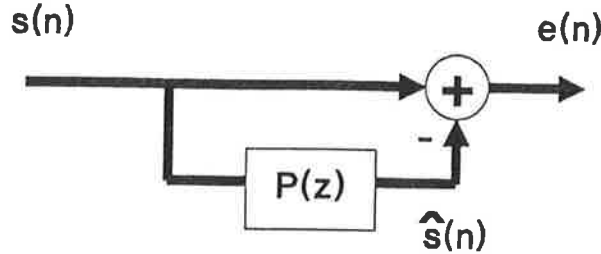
$$\tilde{s}(n) = Gu(n) - \sum_{k=1}^p a_k \tilde{s}(n-k), \quad (2.4)$$

onde as amostras

$$\tilde{s}(n-k), \quad k = 0, \dots, p$$

são do sinal reconstruído $\tilde{s}(n)$.

²“Vocoder” foi o nome dado ao primeiro codificador da fonte de informação do sinal de voz (seção 1.3), concebido por Homer Dudley em 1928 [41]. Esse codificador foi assim chamado porque “opera sob o princípio de derivar códigos de voz para recriar a voz que ele analisa” [13]. A partir de então, todos os codificadores paramétricos da fonte do sinal de voz, em particular aqueles que usam técnicas LPC, passaram a ser denominados “vocoders”.

Figura 2.8: Branqueamento do segmento de voz $s(n)$ pelo seu filtro LPC inversoFigura 2.9: Filtragem do segmento de voz $s(n)$ pelo seu preditor LPC

2.3 O modelo LPC de análise de voz

Aplicando o filtro LPC inverso (Eq.(2.2)) ao segmento $s(n)$, $n = 0, \dots, N - 1$, do sinal original (Fig. 2.8), obtemos na saída o segmento

$$e(n) = s(n) + \sum_{k=1}^p a_k s(n - k), \quad (2.5)$$

que será caracterizado em seguida.

Rearranjando os termos da Eq.(2.5), obtemos uma equação de diferenças para o sinal de entrada

$$s(n) = \hat{s}(n) + e(n), \quad (2.6)$$

onde

$$\hat{s}(n) = - \sum_{k=1}^p a_k s(n - k), \quad (2.7)$$

que é o sinal predito (Fig. 2.9).

O nome de *predição linear* dado a esta técnica vem da aproximação, mediante a Eq.(2.6), de $s(n)$ por $\hat{s}(n)$, que é uma combinação linear de amostras anteriores do sinal original $s(n)$ conforme a Eq.(2.7).

O vetor

$$e(n), \quad n = 0, \dots, N - 1$$

é o erro de predição do segmento $s(n)$ pelo preditor

$$P(z) = - \sum_{k=1}^p a_k z^{-k}. \quad (2.8)$$

O erro de predição será usado para o estabelecimento de condições para a análise LPC na seção 2.4. Uma correspondência entre o erro de predição $e(n)$ (Eq.(2.6)) e a

excitação $x(n) = Gu(n)$ (Eq.(2.4) será estabelecida na seção 2.5, onde vincularemos os modelos LPC de análise e de síntese.

Expandindo a Eq.(2.5) no intervalo de valores de n onde o erro $e(n)$ pode ser não nulo, obtemos

$$e(0) = s(0) \quad (2.9)$$

$$e(1) = s(1) + a_1s(0) \quad (2.10)$$

$$\vdots$$

$$e(p) = s(p) + a_1s(p-1) + \dots + a_ps(0) \quad (2.11)$$

$$\vdots$$

$$e(N-1) = s(N-1) + a_1s(N-2) + \dots + a_ps(N-p-1) \quad (2.12)$$

$$e(N) = a_1s(N-1) + \dots + a_ps(N-p) \quad (2.13)$$

$$\vdots$$

$$e(N+p-1) = a_ps(N-1). \quad (2.14)$$

2.4 Estimação dos coeficientes do filtro LPC

Nesta seção pretendemos obter um sistema de equações lineares cuja solução $\{\alpha_k, k = 1, \dots, p\}$ sejam valores passíveis de atribuição aos coeficientes $\{a_k, k = 1, \dots, p\}$ do filtro LPC $H_{LPC}(z)$ (Eq.(2.1)) associado ao segmento de voz

$$s(n), \quad n = 0, \dots, N-1.$$

Uma *predição perfeita* é caracterizada pela coincidência

$$\hat{s}(n) = s(n), \quad n = p, \dots, N-1$$

entre a parte “estacionária” do vetor predito e a parte correspondente do segmento original ou, equivalentemente pela Eq.(2.6), por

$$e(n) = 0, \quad n = p, \dots, N-1, \quad (2.15)$$

um vetor de erro nulo durante a parte “estacionária” da predição.

A imposição das condições (2.15) de predição perfeita às Eqs.(2.11) a (2.12) resultaria, com 100% de probabilidade, num sistema linear sem solução de $N-p$ equações a p incógnitas.

Em vez de impor $N-p$ condições ao vetor de erro, trataremos de obter um sistema linear determinado de equações mediante o estabelecimento de p condições sobre o *erro quadrático total*, que pode ser genericamente definido como [33]

$$\varepsilon = \sum_{n=n_0}^{n_1} e^2(n), \quad (2.16)$$

onde $e(n)$, $n = n_0, \dots, n_1$ são amostras do erro de predição, determinadas em princípio pela Eq.(2.5).

Na análise LPC, adota-se o critério da minimização do erro quadrático total para a determinação dos p coeficientes de predição.

Os limites da somatória na Eq.(2.16) podem ser definidos segundo dois métodos diferentes:

autocorrelação usam-se todas as amostras potencialmente não nulas do erro de predição. Inspeccionando as Eqs. (2.9) a (2.14), concluímos pelos limites

$$n_0 = 0$$

e

$$n_1 = N + p - 1;$$

covariância usam-se apenas as amostras $e(n)$ do erro de predição para as quais o sinal predito $\hat{s}(n)$ foi determinado com o número máximo p de amostras potencialmente não nulas. Inspeccionando as Eqs. (2.9) a (2.14), concluímos pelos limites³

$$n_0 = p$$

e

$$n_1 = N - 1.$$

O método da autocorrelação dá origem a um sistema de equações lineares que permite considerar o segmento de voz $s(n)$ como um sinal estacionário. Em consequência, podemos elaborar uma interpretação espectral da análise LPC (seção 2.1) [33].

O método da covariância, aplicado à voz sonora com segmentação “pitch synchronous”, permite determinar exatamente o filtro LPC quando este tem resposta impulsiva de grande decaimento num período de “pitch”.

Neste trabalho adotamos o método da autocorrelação e, a partir deste ponto, estaremos nos referindo a ele, salvo declaração expressa em contrário.

Vamos determinar os coeficientes

$$\vec{\alpha} = [\alpha_1 \quad \cdots \quad \alpha_k \quad \cdots \quad \alpha_p]^T \quad (2.17)$$

do filtro LPC que minimiza o erro quadrático total ε (Eq.(2.16)). Para tanto, $\vec{\alpha}$ tem que satisfazer duas condições [1]:

i) o gradiente de ε no ponto descrito pela solução $\vec{\alpha}$ tem que ser nulo, isto é,

$$\vec{\nabla}\varepsilon(\vec{\alpha}) = \vec{0}; \quad (2.18)$$

ii) a matriz Hessiana \mathcal{H} de ε no ponto descrito pela solução $\vec{\alpha}$ tem que ter todos os seus autovalores positivos, ou seja,

$$\lambda_i(\mathcal{H}(\vec{\alpha})) > 0, \quad i = 1, \dots, p. \quad (2.19)$$

³Note-se a utilização das mesmas amostras do vetor de erro tanto para a predição perfeita quanto para a predição pela covariância. Porém, os problemas diferem um do outro porque esta estabelece apenas p condições contra $N - p$ da outra.

Define-se a matriz Hessiana de ε no ponto $\vec{\alpha}$ como aquela que contém o valor de todas as derivadas parciais de segunda ordem daquele funcional neste ponto assim dispostas

$$\mathcal{H}_{ij}(\vec{\alpha}) = \frac{\partial^2 \varepsilon}{\partial a_i \partial a_j}(\vec{\alpha}), \quad 1 \leq i \leq p, \quad 1 \leq j \leq p. \quad (2.20)$$

Explicitando a dependência entre ε e a_k pela substituição da Eq.(2.5) na Eq.(2.16), temos

$$\begin{aligned} \varepsilon &= \sum_{n=0}^{N+p-1} \left(s(n) + a_k \sum_{k=1}^p s(n-k) \right)^2 \\ &= \sum_{n=0}^{N+p-1} s^2(n) + 2 \sum_{k=1}^p a_k \sum_{n=0}^{N+p-1} s(n)s(n-k) + \sum_{i=1}^p \sum_{j=1}^p a_i a_j \sum_{n=0}^{N+p-1} s(n-i)s(n-j), \end{aligned} \quad (2.21)$$

onde estamos expandindo o segmento de voz $s(n)$, $n = 0, \dots, N-1$ com a extensão nula

$$s(n) = 0, \quad n = N, \dots, N+p-1.$$

Definimos os coeficientes de autocorrelação do segmento $s(n)$ como

$$R_{i-j} = \sum_{n=0}^{N+p-1} s(n-i)s(n-j). \quad (2.22)$$

Ademais, como

$$R_{i-j} = R_{j-i}, \quad (2.23)$$

vamos escrever unicamente $R_{|i-j|}$.

Aplicando (2.22) a (2.21), temos

$$\varepsilon = R_0 + 2 \sum_{k=1}^p a_k R_k + \sum_{i=1}^p \sum_{j=1}^p a_i a_j R_{|i-j|}. \quad (2.24)$$

As coordenadas do $\nabla \varepsilon$ da condição (2.18) são as derivadas parciais de ε em relação aos coeficientes a_i . A partir de (2.24), a i -ésima coordenada do gradiente é dada por

$$\frac{\partial \varepsilon}{\partial a_i} = 2R_i + 2 \sum_{j=1}^p a_j R_{|i-j|}. \quad (2.25)$$

A condição vetorial (2.18) dá origem ao sistema de equações

$$\frac{\partial \varepsilon}{\partial a_i}(\vec{\alpha}) = 0, \quad i = 1, \dots, p, \quad (2.26)$$

que, a partir de (2.25), pode ser escrito como

$$\sum_{j=1}^p \alpha_j R_{|i-j|} = -R_i, \quad i = 1, \dots, p. \quad (2.27)$$

As equações do sistema (2.27) são denominadas *equações normais* [31] ou *equações normais de Yule-Walker* [34].

O sistema de equações lineares (2.27) pode ser escrito em forma matricial como

$$R\vec{\alpha} = -\vec{r}, \quad (2.28)$$

onde

- $\vec{\alpha}$ é o vetor solução (Eq. (2.17));
- $R, p \times p$, é a matriz de autocorrelação do segmento de sinal $s(n)$, dada por

$$R = \begin{bmatrix} R_0 & R_1 & R_2 & \cdots & \cdots & R_{p-3} & R_{p-2} & R_{p-1} \\ R_1 & R_0 & R_1 & R_2 & \cdots & \cdots & R_{p-3} & R_{p-2} \\ R_2 & R_1 & R_0 & R_1 & R_2 & \cdots & \cdots & R_{p-3} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ R_{p-3} & \ddots & \ddots & R_2 & R_1 & R_0 & R_1 & R_2 \\ R_{p-2} & R_{p-3} & \ddots & \cdots & R_2 & R_1 & R_0 & R_1 \\ R_{p-1} & R_{p-2} & R_{p-3} & \cdots & \cdots & R_2 & R_1 & R_0 \end{bmatrix}; \quad (2.29)$$

- \vec{r} é o vetor dos p primeiros deslocamentos não nulos da seqüência de autocorrelação de $s(n)$, sendo expresso como

$$\vec{r} = [R_1 \quad R_2 \quad R_3 \quad \cdots \quad \cdots \quad R_{p-2} \quad R_{p-1} \quad R_p]^T. \quad (2.30)$$

A matriz de autocorrelação R é definida positiva [23] [18], ou seja,

$$\vec{x}^T R \vec{x} > 0, \quad \forall \vec{x} \in \mathbb{R}^p. \quad (2.31)$$

Adicionalmente, $\vec{\alpha}$, para ser a solução, ainda tem que satisfazer a condição (2.19).

A partir da definição (2.20) de \mathcal{H} , derivando as derivadas parciais de primeira ordem (2.25) e comparando a matriz obtida com R , concluímos

$$\mathcal{H}(\vec{\alpha}) = 2R. \quad (2.32)$$

Assim, os autovetores e autovalores de $\mathcal{H}(\vec{\alpha})$ são idênticos aos de R .

Seja \vec{v} um autovetor de R e λ o autovalor correspondente, isto é,

$$R\vec{v} = \lambda\vec{v} \quad (2.33)$$

Usando a propriedade de definição positiva de R , temos

$$\vec{v}^T R \vec{v} > 0. \quad (2.34)$$

Substituindo (2.33) em (2.34), temos

$$\vec{v}^T \lambda \vec{v} > 0 \quad (2.35)$$

ou

$$\lambda (\vec{v}^T \vec{v}) > 0. \quad (2.36)$$

Portanto,

$$\lambda > 0 \quad (2.37)$$

e a condição (2.19) está automaticamente satisfeita se escolhermos o filtro de coeficientes $\vec{\alpha}$ dados pela solução das equações (2.28).

2.5 Espectro do erro de predição

A análise LPC adapta uma envoltória ao espectro de um segmento do sinal de voz (seção 2.5).

Podemos decompor o espectro de curto prazo da voz em duas componentes: a envoltória e a estrutura fina [41].

O modelo LPC de síntese (Eq.(2.4) e Fig. 2.6) pode ser representado pela transformada de Fourier de curto prazo [38] [41] como

$$\hat{S}(e^{j\omega}) = G H_{LPC}(e^{j\omega}) U(e^{j\omega}), \quad (2.38)$$

onde

- $H_{LPC}(e^{j\omega})$ é a resposta em frequência do filtro LPC e
- $\tilde{S}(e^{j\omega})$ e $X(e^{j\omega})$ são as transformadas de Fourier de curto prazo do segmento $\tilde{s}(n)$ do sinal sintético e do segmento $x(n)$ do sinal de excitação, respectivamente.

O espectro de curto prazo do sinal sintético (Eq.(2.38)) é dado logaritmicamente por

$$\tilde{S}(e^{j\omega})|_{\text{dB}} = G_{\text{dB}} + H_{LPC}(e^{j\omega})|_{\text{dB}} + X(e^{j\omega})|_{\text{dB}}, \quad (2.39)$$

onde

- $G_{\text{dB}} = 20 \log_{10} G$;
- $W(e^{j\omega})|_{\text{dB}} = 20 \log_{10} |W(e^{j\omega})|$, substituindo W por \tilde{S} , H_{LPC} e X , alternadamente.

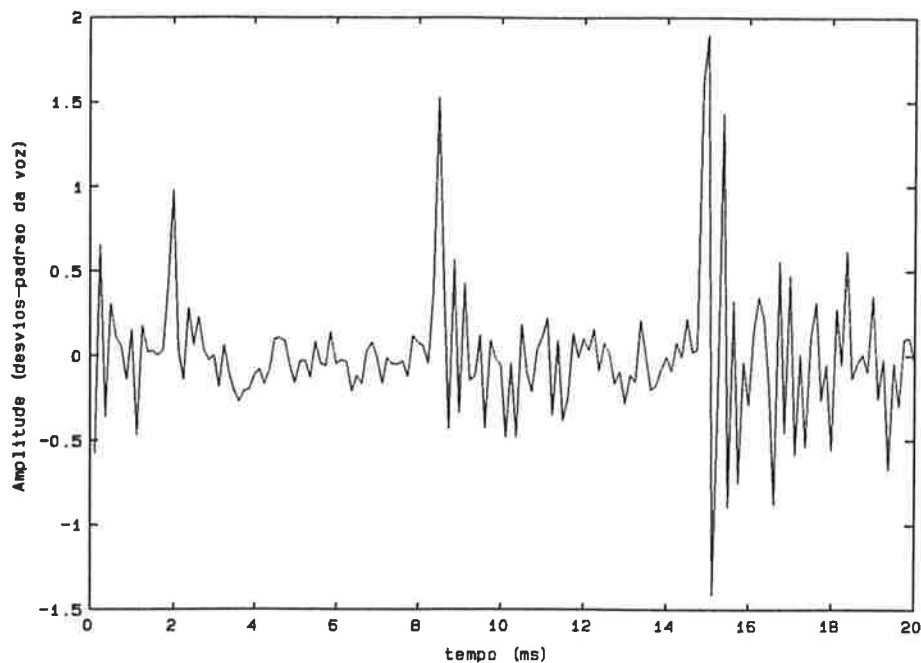


Figura 2.10: Erro de predição para um segmento de voz sonora

O espectro logarítmico do sinal sintético é composto, por construção, por duas componentes aditivas, que são a envoltória $G_{dB} + H_{LPC}(e^{j\omega})|_{dB}$ e a estrutura fina $X(e^{j\omega})|_{dB}$.

Por outro lado, o modelo de análise (Eq.(2.5) fornece a seguinte relação

$$S(e^{j\omega}) = H_{LPC}(e^{j\omega})E(e^{j\omega}) \quad (2.40)$$

entre os espectros do sinal e do erro.

Se a excitação $x(n) = Gu(n)$ fosse idêntica ao erro de predição $e(n)$, a síntese seria uma perfeita reconstrução do sinal original. Entretanto, o sintetizador LPC dispõe apenas do ganho G , que controla a proximidade média entre os espectros.

O modelamento LPC tem um desempenho melhor com sinais sonoros, conforme transparece nas características dos erros de predição mostrados nas figuras 2.10 e 2.12.

Mesmo assim, a representação do erro sonoro por um sinal impulsivo periódico (seção 2.2) não é exata e a determinação do seu período está sujeita a alguma arbitrariedade no processamento.

O problema da detecção de "pitch" se agrava porque o ouvido humano é uma ordem de grandeza mais sensível a variações relativas na frequência fundamental do que em qualquer outro parâmetro da voz [20].

A importância perceptiva do "pitch", a dificuldade de estimá-lo bem e constantemente e a grande simplicidade do modelo de excitação LPC motivaram a experimentação com outras formas de codificação e de construção da excitação, conforme veremos no próximo capítulo.

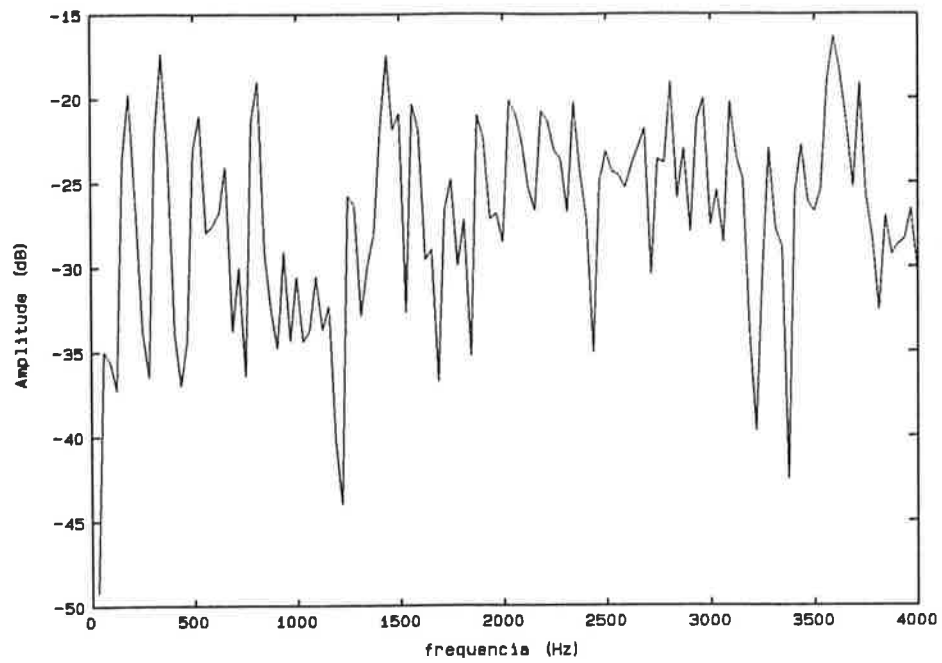


Figura 2.11: Espectro logarítmico de potência do erro de predição para um segmento de voz sonora

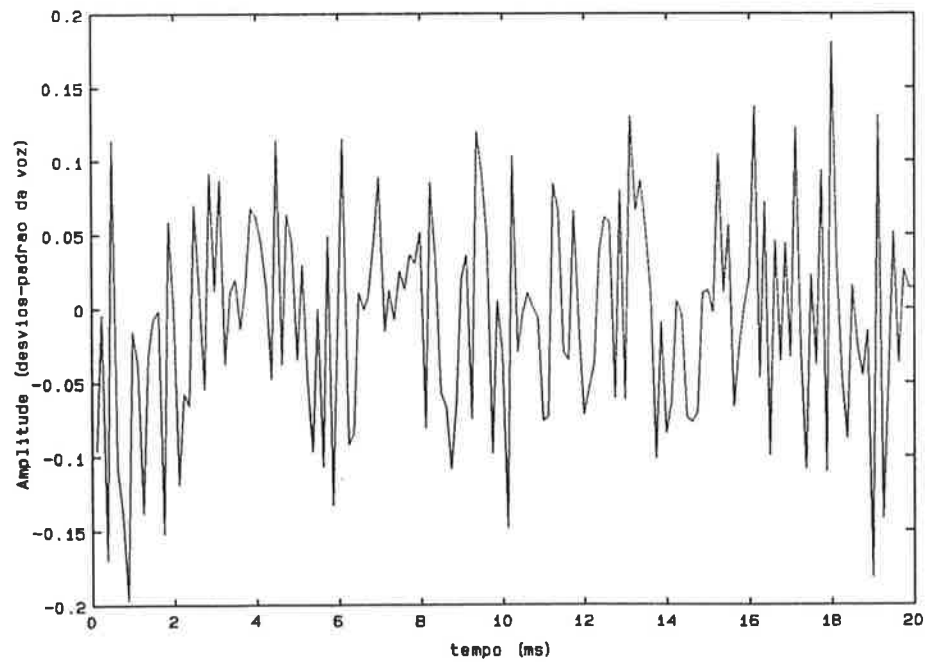


Figura 2.12: Erro de predição para um segmento de voz surda

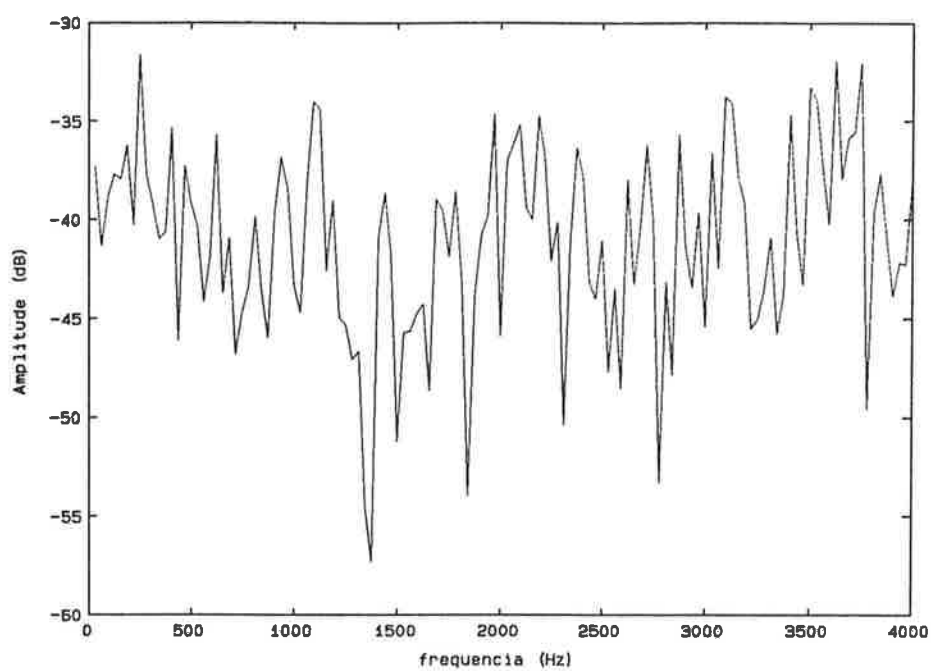


Figura 2.13: Espectro logarítmico de potência do erro de predição para um segmento de voz surda

Capítulo 3

ANÁLISE MEDIANTE SÍNTESE

O princípio da análise mediante síntese em processamento de voz é aplicado em todos os analisadores que incluem em sua constituição um sintetizador que controlam para reconstruir com a máxima fidelidade determinada característica $y(l)$ do sinal de voz $s(n)$ (Fig. 3.1). O resultado $c(m)$ da análise codifica o valor da melhor combinação dos parâmetros de controle do sintetizador.

A designação de “*analysis-by-synthesis*” (*análise-mediante-síntese*) foi dada pela primeira vez a um detetor de formantes (seção 3.1).

O objetivo deste capítulo é a análise da codificação da excitação de filtros LPC (Cap. 2) pelo método da análise-mediante-síntese (seção 3.4).

A primeira apresentação mais genérica de um codificador desta classe foi feita por Schroeder e Atal [42] *apud* Kroon [28], que denominaram seu processo de codificação de “*Code-excited linear prediction*” (CELP). Usaremos “CELP” para designar qualquer codificador da classe.

Os codificadores da excitação de filtros LPC que vamos considerar são vetoriais e admitem uma interpretação espectral para o seu critério de comparação.

Como os codificadores CELP incorporam uma versão vetorial de *quantização em malha fechada*, apresentaremos algumas das anteriores aplicações escalares deste conceito na seção 3.2. Ademais, a quantização em malha fechada também pode ser enquadrada no conceito mais geral de análise-mediante-síntese.

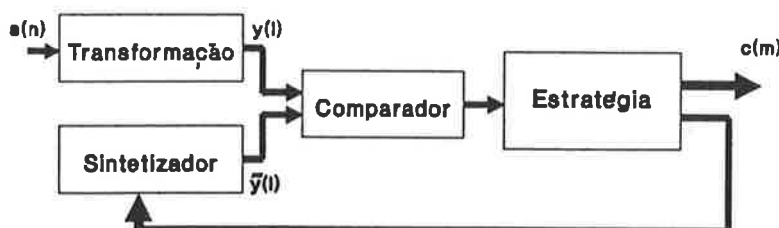


Figura 3.1: Princípio da análise-mediante-síntese

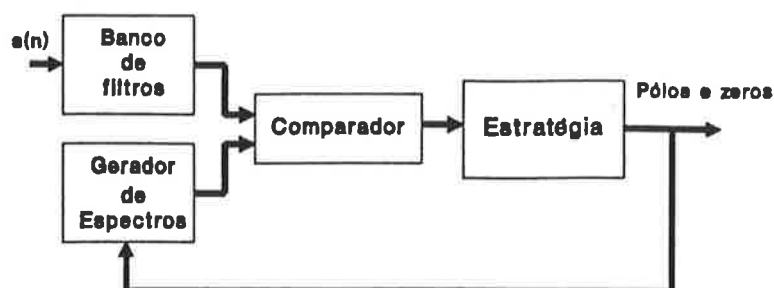


Figura 3.2: Localização de formantes pelo método da análise-mediante-síntese

3.1 Analisadores espectrais

A expressão “*analysis-by-synthesis*” (*análise-mediante-síntese*) foi usada pela primeira vez para designar uma técnica de ajuste de espectros para a localização das formantes [8] *apud* [14].

O espectro de curto prazo da voz é obtido através de um banco de filtros (Fig. 3.2).

Um gerador de espectros opera a partir de um modelo de pólos e zeros do aparelho fonador e da sua excitação.

Os espectros real e sintético são comparados com base num erro quadrático ponderado. Uma estratégia de controle varia o valor dos pólos e zeros do modelo até encontrar a combinação que minimize o erro ponderado, que é então armazenada na memória.

Outro extrator de formantes contemporâneo àquele descrito acima é o que foi apresentado em [35] *apud* [14], baseado numa análise “pitch synchronous” dos sons sonoros.

Calcula-se o espectro pela determinação da série de Fourier do sinal periódico montado a partir de cada período fundamental individual do sinal de voz. O sinal de voz, durante os seus intervalos sonoros, é segmentado em períodos fundamentais através de uma inspeção visual do seu oscilograma.

Um espectro sintético é gerado a partir de um modelo de pólos e zeros do aparelho fonador e da excitação glotal.

As frequências e os amortecimentos dos pólos e dos zeros são variados para ajustar, pelo critério de minimização de um erro quadrático ponderado, o espectro sintético à envoltória do espectro calculado, ambos representados em escala logarítmica. Por isso, este processo também pode ser classificado como sendo de “análise-mediante-síntese”.

Obtém-se bons ajustes em processamentos envolvendo de 10 a 20 ciclos de comparação, com erros típicos de aproximação variando entre 1,5 e 2,5 dB.

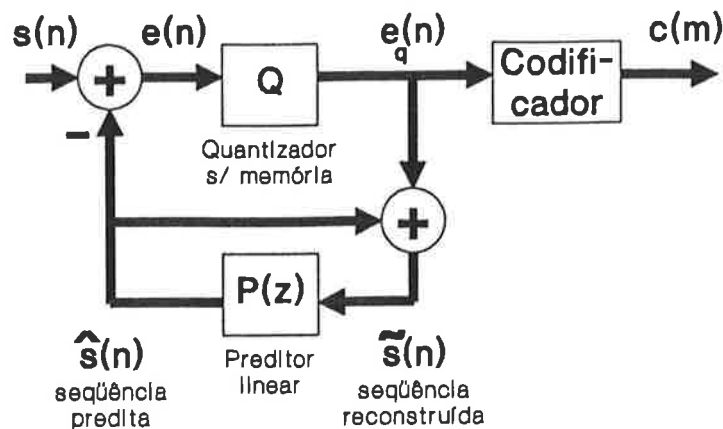


Figura 3.3: Codificador com quantização em malha fechada

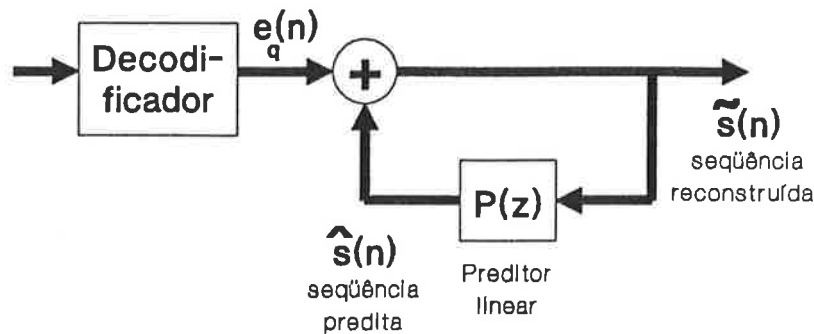


Figura 3.4: Decodificador diferencial

3.2 Quantização em malha fechada: ADPCM e APC

A quantização em malha fechada foi introduzida com base no reconhecimento de que a distorção em sistemas com quantizadores é inevitável e é melhor considerá-la no próprio codificador, pois ela será percebida, após a decodificação, no sinal de voz reconstruído.

Os codificadores com quantizadores realimentados possuem um preditor $P(z)$ na malha (Fig. 3.3) e o quantizador Q opera sobre o sinal de diferença

$$e(n) = s(n) - \hat{s}(n), \quad (3.1)$$

onde

- $s(n)$ é uma amostra do sinal de voz;
- $\hat{s}(n)$ é o valor predito para a amostra $s(n)$, baseado *indiretamente* em amostras de voz anteriores $s(n-k)$, $k = 1, \dots, p$, sendo p a ordem do preditor $P(z)$ (Eq.(2.8)).

O preditor LPC da seção 2.3 foi aplicado às amostras originais $s(n-k)$, $k = 1, \dots, p$ do sinal de voz. Nesta malha fechada de quantização (Fig. 3.3), o preditor

é suprido com amostras do *sinal reconstruído* $\tilde{s}(n - k)$, $k = 1, \dots, p$ [17], que são obtidas como

$$\tilde{s}(n - k) = \hat{s}(n - k) + e_q(n - k), \quad k = 1, \dots, p, \quad (3.2)$$

onde $e_q(n - k)$ são amostras quantizadas do sinal diferença $e(n - k)$, ou seja,

$$e_q(n - k) = Q[e(n - k)] \quad k = 1, \dots, p. \quad (3.3)$$

O codificador tem disponíveis as amostras originais $s(n - k)$, $k = 1, \dots, p$, mas o decodificador recebe apenas as amostras $e_q(n - k)$, $k = 1, \dots, p$ codificadas na seqüência de bits $c(m)$ (Fig. 3.4) e, a partir dela, também tem condições de determinar as amostras $\hat{s}(n - k)$, $k = 1, \dots, p$ do sinal predito.

A aplicação do princípio da análise-mediante-síntese reside na restrição do codificador a usar apenas a informação que estará disponível ao decodificador no instante da reconstrução, na ausência de erros na transmissão ou no armazenamento da informação codificada (seção 1.3).

O quantizador dos codificadores diferenciais é adaptativo. O preditor pode ser fixo ou adaptativo.

Os codificadores diferenciais podem adotar uma entre duas estratégias de adaptação dos parâmetros dos seus preditores:

“**forward**” agrupa as amostras de voz originais em blocos com duração de alguns milissegundos e estima o preditor pela técnica LPC (seção 2.4);

“**backward**” atualiza os parâmetros do preditor para cada instante de amostragem do sinal de voz adicionando-lhes incrementos que são ponderações do *erro de predição quantizado*¹, que é estimado a partir de amostras passadas do sinal de diferença quantizado $e_q(n)$.

Na adaptação “forward” do preditor, os seus parâmetros precisam ser codificados e enviados como informação lateral, complementar à seqüência de bits $c(m)$, porque foram calculados a partir da voz original, não disponível ao decodificador. Esta estratégia de adaptação do preditor é empregada pelos codificadores “*Adaptive Predictive Coders*” (APC) em geral [3] e, em particular, por aquele proposto pelo Departamento de Defesa dos EUA (Tabela 3.1) [46].

Para a adaptação “backward” do preditor, usam-se algoritmos para filtros adaptativos, como o *Least Mean Squares* (LMS), baseado no gradiente do erro [23] ou outros baseados em filtros de Kalman. Estes codificadores não precisam transmitir nenhuma informação adicional além da seqüência de bits $c(m)$ que codifica o sinal de diferença quantizado $e_q(n)$. Uma classe de codificadores que emprega esta estratégia é a dos “*Adaptive Differential Pulse Code Modulators*” (ADPCM) [23], e, em particular, aquele proposto pela CCITT (Tabela 3.1) [11].

¹Note-se que este erro é conceitualmente equivalente ao *erro quadrático total* da seção 2.4, a menos do método de cálculo.

Tabela 3.1: Características de codificadores ADPCM e APC

		ADPCM	APC
Quantizador	tipo	adaptativo	adaptativo
	R (bit/amostra)	4	1
Preditor	tipo	adaptativo	adaptativo
	ordem	4	4
Taxa total (kbit/s)		32	9,6

3.3 Filtros LPC excitados por códigos (CELP)

Os codificadores preditivos de voz por análise mediante síntese constituem uma vasta classe. A primeira apresentação formal de um codificador típico desta classe foi feita por Schroeder e Atal [42] *apud* Kroon [28], que denominaram seu processo de codificação de “*Code-excited linear prediction*” (CELP).

Os codificadores CELP compõem-se de dois subsistemas (Fig. 3.5):

- um filtro $H_{LPC}(z)$, que modela a envoltória do espectro de curto prazo do sinal de voz;
- um gerador da excitação $x(n) = gc(n)$, que modela a estrutura fina do espectro de curto prazo do sinal de voz.

Os parâmetros do filtro $H_{LPC}(z)$ são determinados a partir da *análise preditiva* ou “*linear predictive coding*” (LPC) [33] de segmentos de dezenas de milissegundos de duração, representados por blocos (“frames”) de N amostras do sinal de voz (Cap. 2).

A seqüência de excitação é gerada em sub-blocos (“subframes”) de comprimento L menor que os blocos para análise preditiva.

Nas nossas simulações, usamos $N = 160$ e $L = 40$ com freqüência de amostragem $f_a = 8$ kHz.

Os segmentos da seqüência de excitação são especificados numa representação forma-ganho [28], onde a forma é dada por vetores-código

$$c_i(n), \quad n = 1, \dots, L$$

e o ganho g é um fator de escala (Fig. 3.5).

Os vetores-código são escolhidos de um conjunto

$$\{c_i(n), \quad i = 1, \dots, M\},$$

denominado livro de códigos (“codebook”) através de *análise mediante síntese*.

Para a caracterização completa do processo de síntese CELP (Fig. 3.5), convém considerar

- a seqüência de excitação $c(n)$, $n = 1, 2, \dots$, formada pela concatenação dos vetores-código $c_{i(m)}$ selecionados a cada sub-bloco m de excitação;

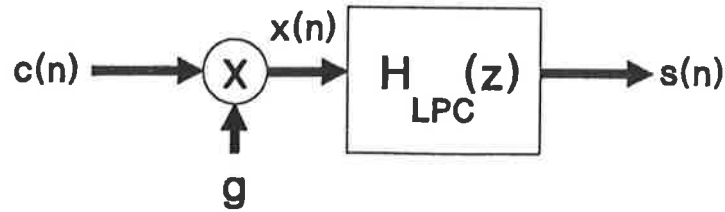


Figura 3.5: Síntese CELP de um segmento de voz

- a seqüência de modulação da excitação $g(n)$, $n = 1, 2, \dots$, definida a cada instante n pelo ganho g_m determinado para o sub-bloco m atual;
- o filtro linear e variante no tempo

$$H_{LPC,l}(z) = \frac{1}{1 + \sum_{k=1}^p a_{kl} z^{-k}}, \quad (3.4)$$

onde a_{kl} , $k = 1, \dots, p$ são os coeficientes de predição linear para o bloco l de análise LPC;

- a seqüência de voz sintética $\tilde{s}(n)$, $n = 1, 2, \dots$, formada pela concatenação dos vetores \tilde{s}_m reconstruídos a cada sub-bloco m de excitação.

Há dois tipos de transições na síntese CELP:

- transições de bloco com a mudança dos coeficientes do filtro;
- transições de sub-bloco com a mudança do ganho e do vetor de excitação.

Para assegurar uma certa suavidade ao processo de filtragem, na passagem pelos dois tipos de transições mantém-se o estado do filtro $H_{LPC,l}(z)$.

Seja $z^m(n)$ o vetor variável que contém os p estados do filtro $z_k^m(n)$, $k = 1, \dots, p$ no instante n do sub-bloco m . Estaremos interessados em três fases do sub-bloco m , onde os estados serão representados por

$$\begin{cases} z^m(0) & \text{as condições iniciais no sub-bloco } m, \\ z^m(n), \quad n = 1, \dots, L-1 & \text{o estado ao longo do sub-bloco } m, \\ z^m(L) & \text{o estado ao fim do sub-bloco } m, \end{cases}$$

impomos a cada transição de sub-bloco a condição

$$z_{m,0} = z_{m-1,L}. \quad (3.5)$$

Em seguida, vamos apresentar uma das possíveis representações de estados do filtro LPC, que descreve completamente o processo de filtragem.

Denotando a seqüência de vetores de estado do filtro $H_{LPC,l}(z)$ por $z(n)$, $n = 0, 1, \dots$, podemos, a cada instante n , obter o estado seguinte $z(n)$ a partir do estado atual $z(n-1)$ e da entrada atual $g(n)c(n)$ através de

$$z(n) = A_l z(n-1) + b g(n) c(n), \quad (3.6)$$

onde

- a matriz de transição de estados do filtro $H_{LPC,l}(z)$ durante o bloco² $l = n \div N + 1$ é

$$A_l = \begin{bmatrix} -a_{1l} & -a_{2l} & \cdots & -a_{p-1,l} & -a_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}; \quad (3.7)$$

- o vetor de coeficientes da entrada é b , dado por

$$b(k) = \begin{cases} 1 & k = 1 \\ 0 & k = 2, \dots, p \end{cases}.$$

Completamos a representação de estados do filtro LPC com a equação amostral da seqüência de saída

$$\tilde{s}(n) = z_1(n), \quad (3.8)$$

onde $z_1(n)$ é o valor da 1^a das variáveis de estado $z_k(n)$, $k = 1, \dots, p$, do filtro LPC no instante n .

3.4 Busca da excitação CELP

Nesta seção serão delineados os procedimentos gerais que são aplicados nos codificadores CELP para a construção da seqüência de excitação $c(n)$.

A seleção da excitação é um processo de busca, onde, conceitualmente³, a cada sub-bloco m , cada vetor-código c_i , amplificado adequadamente,

$$x_i = g_i c_i \quad (3.9)$$

é passado pelo filtro LPC (Fig. 3.6) com condições iniciais idênticas ao estado final do sub-bloco anterior $m - 1$, resultando numa reconstrução \hat{s}_i .

Cada sub-bloco reconstruído \hat{s}_i é comparado com o sub-bloco original s , produzindo o vetor de erro e_i

$$e_i = s - \hat{s}_i. \quad (3.10)$$

Calcula-se o erro quadrático acumulado no sub-bloco

$$\varepsilon_i = \sum_{n=0}^{L-1} e_i^2(n). \quad (3.11)$$

Escolhe-se o índice do vetor-código de menor erro quadrático

$$i = \arg \left(\min_{j=1, \dots, M} \varepsilon_j \right). \quad (3.12)$$

² $p \div q$ indica o quociente inteiro da divisão de p por q

³Há codificadores cujo conjunto de possíveis vetores-código não é definido explicitamente, como o LPC com excitação por multipulsos (MPE) (seção 3.10).

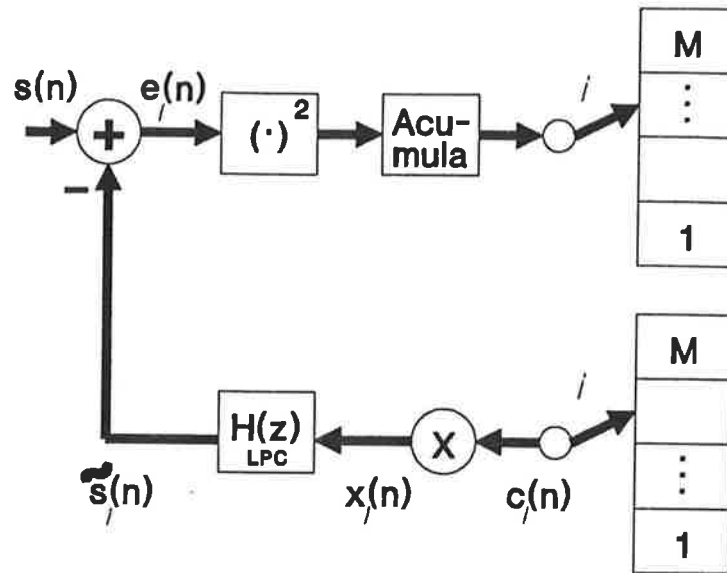


Figura 3.6: Procedimento básico de busca da excitação

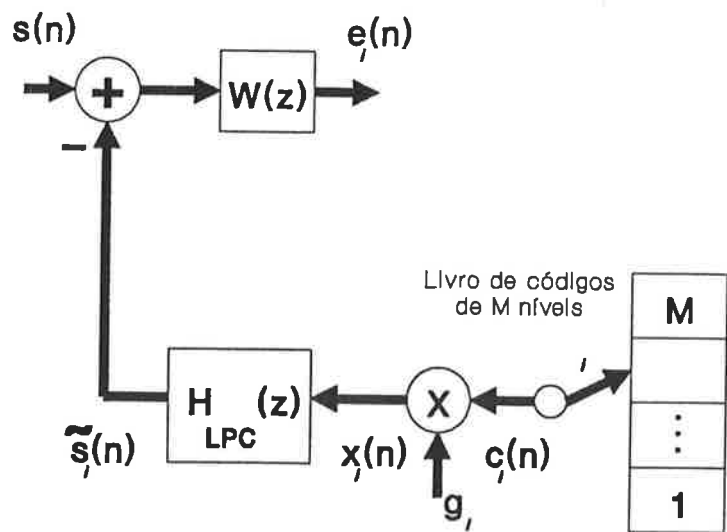


Figura 3.7: Busca da excitação por análise mediante síntese

O critério usado na análise LPC é a minimização do erro quadrático total de predição (seção 2.4).

Na codificação CELP, minimiza-se o erro de predição na etapa inicial de análise LPC. O critério da etapa de busca da excitação é a minimização do erro quadrático total de reconstrução. Podemos dizer que é feita a minimização da distorção de quantização vetorial da excitação condicionada à minimização do erro de predição.

3.5 Efeitos perceptivos na audição

Em testes de percepção, foram caracterizados os seguintes efeitos de mascaramento auditivo [4]:

- nas faixas espectrais das formantes, o ruído de quantização é parcial ou totalmente encoberto pelo sinal de voz;
- o ouvido tolera mais ruído durante os sons de transição quando a largura de faixa das formantes é maior.

Mais quantitativamente, para sons vocálicos, o ruído de quantização é audível em todas as faixas de frequência em que a sua envoltória espectral estiver menos do que 3 dB abaixo da envoltória espectral do sinal de voz [43].

Conforme vimos para a análise LPC (seção 2.5), o critério de minimização do erro quadrático total tende a produzir um sinal de erro com espectro uniforme. De uma forma geral, o erro de quantização de outros preditores que adotam um critério de minimização quadrático também possui espectro plano [4].

Para codificadores CELP, podemos manter a análise LPC com o seu erro de quantização de espectro plano e procurar compensar o perfil espectral do erro de reconstrução.

Na seleção para excitação entre dois vetores-código que acarretam o mesmo erro total de reconstrução, para obter melhores resultados perceptivos, deve-se escolher aquele cujo ruído nas faixas das formantes seja maior, contanto que não se situe a menos de 3 dB do sinal.

3.6 Ponderação espectral perceptiva

Devemos efetuarmos uma remodelagem do espectro do ruído de quantização (“quantization noise shaping”) conforme sugerido na seção 3.5.

Pretendemos alterar o espectro da seqüência de erro $e(n)$ pela inserção de um filtro antes da segmentação do vetor de erro e_i e do cálculo do erro quadrático (Fig. 3.6). Seja $W(z)$ a função de transferência desse filtro de ponderação espectral perceptiva (“perceptual weighting filter”).

Para impedir a passagem do erro de reconstrução na faixa das formantes, $W(e^{j\omega})$ deve ter atenuação máxima nessas regiões espectrais. Portanto, o filtro LPC inverso é um bom candidato a filtro de ponderação espectral.

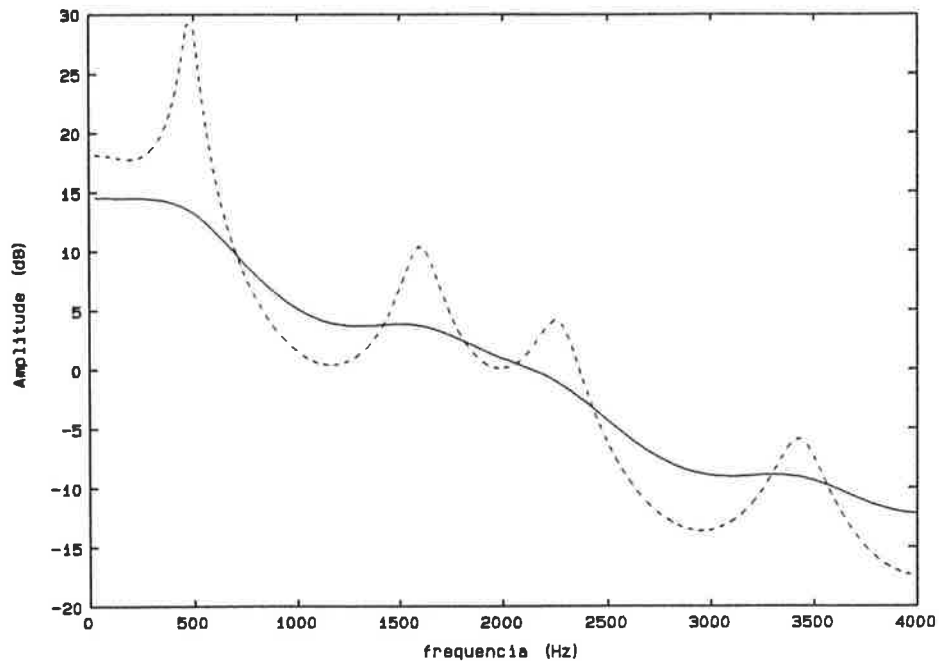


Figura 3.8: Expansão de faixa de um filtro LPC

Entretanto, seria desejável que $W(z)$ tivesse algum parâmetro para acentuar a passagem das formantes caso o filtro LPC inverso as atenuasse em excesso.

Considerando os dois últimos parágrafos e o efeito de atenuação de ressonâncias pela expansão de faixa (Apêndice D), propomos que o filtro de ponderação espectral ajustável seja

$$W(z) = \frac{H_{LPC}(z/\gamma)}{H_{LPC}(z)}. \quad (3.13)$$

A partir da Eq.(D.24), temos a expansão de faixa

$$\Delta B = -(f_a/\pi) \ln \gamma, \quad (3.14)$$

onde γ é o coeficiente de expansão de faixa e f_a é a frequência de amostragem do sinal de voz.

Tomamos $\gamma = 0,8$, que acarreta uma expansão $\Delta B = 568$ Hz, conforme o módulo da resposta em frequência do filtro expandido à Fig. 3.8. O filtro de ponderação espectral está representado na Fig. 3.9, tornando clara a tolerância de erros maiores nas formantes.

Retornando à aplicação do filtro de ponderação, o esquema de busca da Fig. 3.7 demanda a execução de duas filtragens por vetor-código. Podemos prescindir de uma delas com o deslocamento do filtro $W(z)$ para a entrada do somador de comparação, como indicado na Fig. 3.10, onde $y(n)$ é o segmento de voz ponderada.

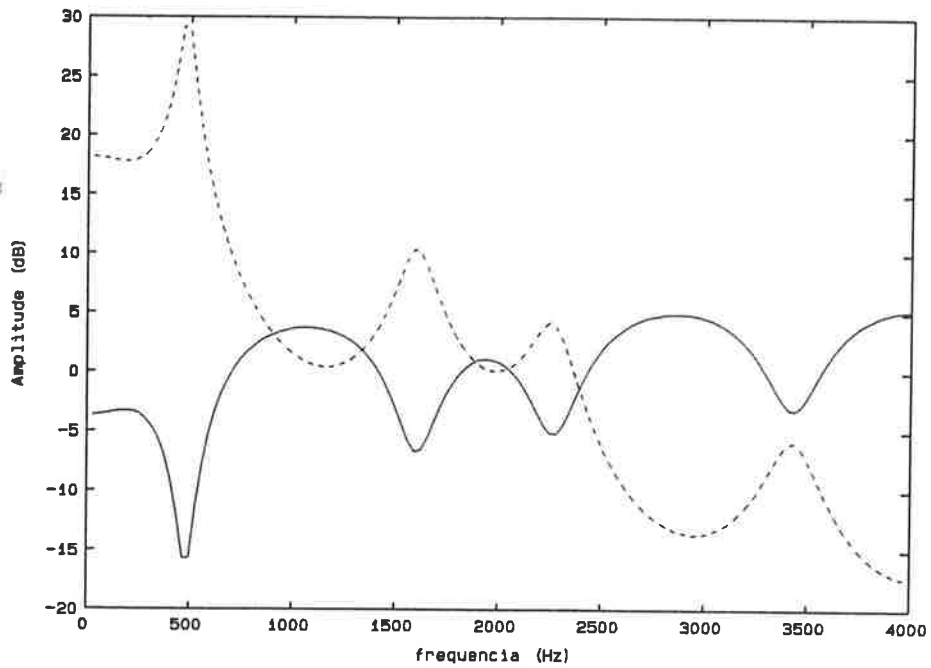


Figura 3.9: Ponderação espectral associada a um filtro LPC

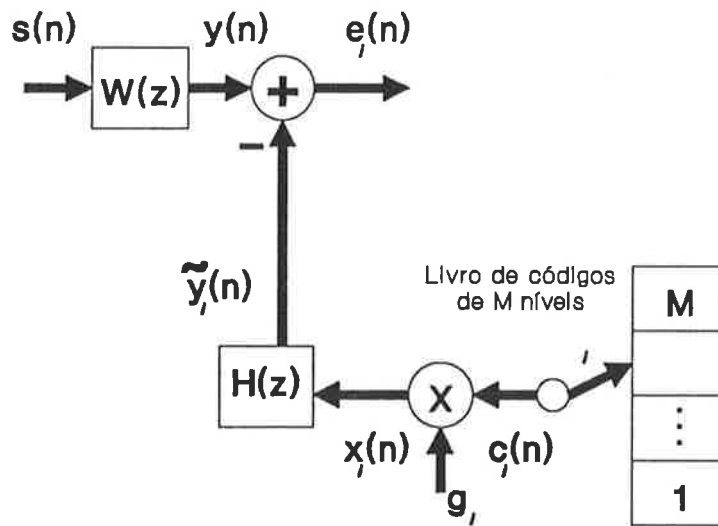


Figura 3.10: Busca com comparação no domínio ponderado

3.7 Matriz de resposta impulsiva

Dado o filtro $H(z)$, seja $\mathbf{h}(n)$ sua seqüência de resposta impulsiva com condições iniciais nulas, que é dada por

$$\mathbf{h}(n) = \mathcal{Z}^{-1}\{H(z)\}, \quad n = 0, 1, \dots, \quad (3.15)$$

onde \mathcal{Z}^{-1} , a transformada z inversa, expressa-se analiticamente por

$$\mathbf{h}(n) = \frac{1}{2\pi j} \oint_C H(z) z^{n-1} dz,$$

onde C é qualquer curva fechada situada na região de convergência de $H(z)$ e cujo interior contenha a origem do plano complexo z [37].

Define-se o vetor de resposta impulsiva do filtro $H(z)$ no sub-bloco de duração L como

$$\mathbf{h} = [\mathbf{h}(0) \ \mathbf{h}(1) \ \dots \ \mathbf{h}(L-1)]^T, \quad (3.16)$$

sua seqüência de resposta impulsiva truncada às primeiras L amostras.

A matriz de resposta impulsiva do filtro $H(z)$ é uma matriz Toeplitz triangular inferior que contém o seu vetor de resposta impulsiva na 1ª coluna:

$$H = \begin{bmatrix} \mathbf{h}(0) & 0 & \dots & 0 & 0 \\ \mathbf{h}(1) & \mathbf{h}(0) & 0 & \dots & 0 \\ \vdots & \dots & \dots & \dots & \vdots \\ \mathbf{h}(L-2) & \dots & \mathbf{h}(1) & \mathbf{h}(0) & 0 \\ \mathbf{h}(L-1) & \mathbf{h}(L-2) & \dots & \mathbf{h}(1) & \mathbf{h}(0) \end{bmatrix}. \quad (3.17)$$

Adicionalmente, quando $H(z)$ é um filtro LPC na forma

$$H(z) = \frac{1}{A(z)},$$

com

$$A(z) = 1 + \sum_{k=1}^p a_k z^{-k},$$

temos

$$\mathbf{h}(0) = 1.$$

A matriz de resposta impulsiva permite escrever a resposta total do filtro expandido $H(z)$ num sub-bloco como

$$\hat{y}_i = Hx_i + y_0, \quad (3.18)$$

onde

- H é a matriz de resposta impulsiva do filtro $H(z)$ com $z_m(0) = 0$;

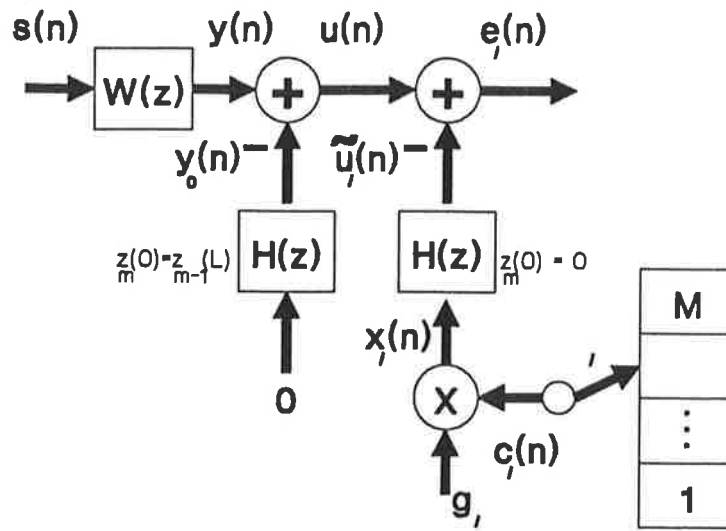


Figura 3.11: Busca com filtro descarregado

- y_0 é a resposta à excitação nula de $H(z)$ truncada a um sub-bloco com $z_m(0) = z_{m-1}(L)$;

sendo

$$\begin{cases} z_m(0) & \text{as condições iniciais no sub-bloco } m, \\ z_m(n), \quad n = 1, \dots, L-1 & \text{o estado ao longo do sub-bloco } m, \\ z_m(L) & \text{o estado ao fim do sub-bloco } m. \end{cases}$$

Durante o processo de busca (Fig. 3.10), o erro de reconstrução é dado por

$$e_i = y - \tilde{y}_i. \quad (3.19)$$

Das equações (3.18) e (3.19), temos

$$e_i = y - y_0 - Hx_i. \quad (3.20)$$

Definindo o vetor-alvo da busca como

$$u = y - y_0 \quad (3.21)$$

e a sua reconstrução como

$$\tilde{u}_i = Hx_i, \quad (3.22)$$

podemos escrever trivialmente o erro de reconstrução como

$$e_i = u - \tilde{u}_i. \quad (3.23)$$

A busca do vetor-alvo u , obtido após a subtração da descarga y_0 do sub-bloco de voz ponderada y , está ilustrada na Fig. 3.11.

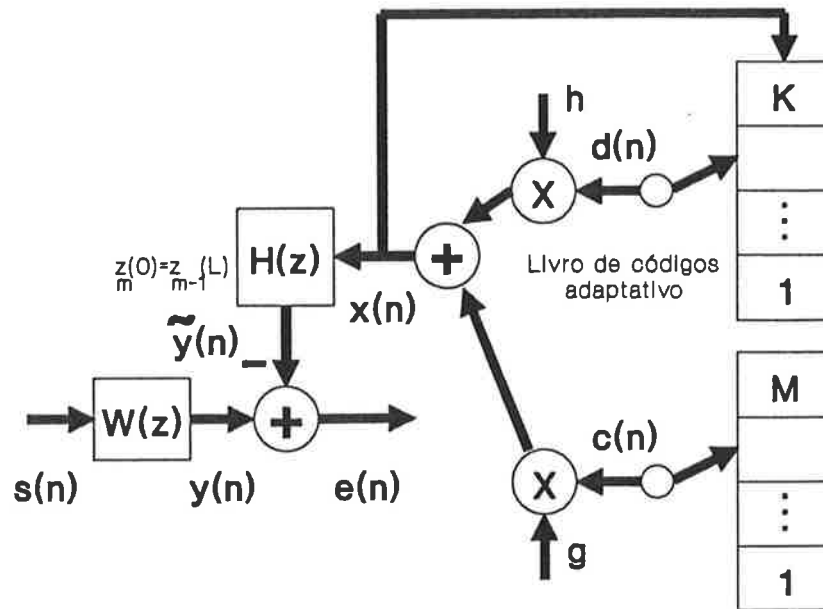


Figura 3.12: Atualização de estados do codificador CELP

3.8 Construção dos livros de códigos da excitação

Vamos tratar nesta seção do importante livro de códigos adaptativo de excitações atrasadas e dos livros de códigos estocásticos com vetores ceifados centralmente.

Um outro tipo de livros de códigos é aquele definido por base vetorial (seção 3.11), que vem encontrando aplicações em telefonia celular.

Ainda outra sub-classe de codificadores que admite a definição conceitual de um livro de códigos é a dos que são excitados por multipulsos (seção 3.10).

3.8.1 Livro de excitações atrasadas

Temos tratado o livro de códigos implicitamente como uma entidade fixa (seção 3.4). Entretanto, para reconstruir segmentos quase-periódicos de voz, recorreremos a uma estrutura adaptativa para poder redefinir os vetores-código à medida que o sinal varia.

Este livro de códigos adaptativo têm sua constituição básica definida pelo vetor b de excitações atrasadas, que, no sub-bloco m , encontra-se no estado

$$b_m = [\mathbf{x}(mL - K + 1) \quad \mathbf{x}(mL - K + 2) \quad \cdots \quad \mathbf{x}(mL)], \quad (3.24)$$

onde $\mathbf{x}(n)$, $n = \dots, mL$ é a seqüência de excitações passadas e K é o comprimento do vetor de excitações atrasadas (Fig. 3.12).

O vetor b é atualizado ao fim de cada sub-bloco de busca da excitação mediante os dois procedimentos seguintes

- i) esquecimento das L amostras mais antigas;

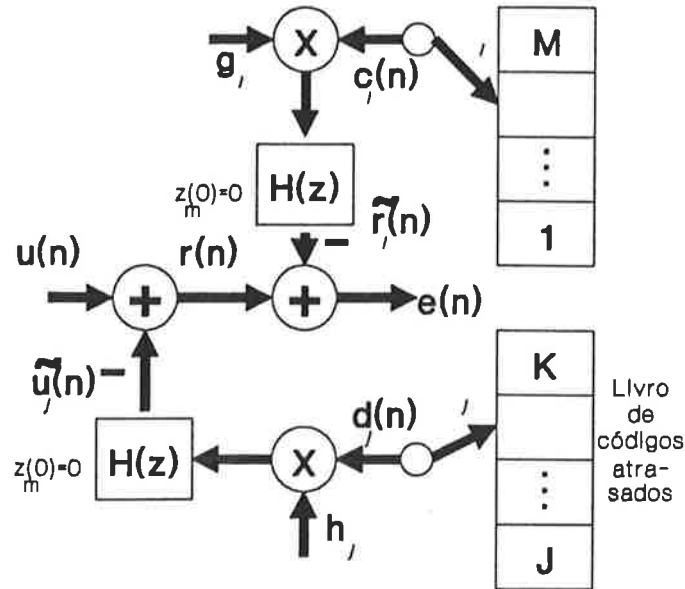


Figura 3.13: Codificador CELP com livro de códigos atrasados

ii) anexação do vetor x de excitação mais recente.

O vetor atualizado é

$$b_{m+1}(k) = \begin{cases} b_m(k+L) & 1 \leq k \leq K-L \\ x(k-K+L) & K-L+1 \leq k \leq K \end{cases} \quad (3.25)$$

Os vetores-código são segmentos de comprimento L retirados do vetor b com sobreposição de $L-1$ amostras entre vetores adjacentes. Isto é,

$$d_j(n) = b(K-j+n) \quad \begin{matrix} n = 1, \dots, L \\ j = L, \dots, K. \end{matrix} \quad (3.26)$$

Nos codificadores simulados, usamos sub-blocos de comprimento $L = 40$ e frequência de amostragem $f_a = 8$ kHz. Isto significa que temos índice mínimo $j = 40$, correspondente a uma frequência fundamental de 200 Hz.

Há falantes com frequência fundamental acima de 200 Hz. Assim, seguindo [10], redefinimos o índice mínimo de atraso como $J = 20$.

Adotando um livro de códigos de 128 níveis, devemos projetar o comprimento do vetor b das excitações atrasadas em $K = 147$. Assim, conseguimos cobrir uma faixa de períodos fundamentais de 2,500 ms a 16,375 ms com passo de 0,125 ms ou uma faixa de frequências fundamentais equivalente de 61 Hz a 400 Hz.

Ao estendermos a definição (3.26) para índices j tais que $J \leq j \leq L-1$, constatamos que faltam amostras antigas de b para completar as amostras dos vetores-código d_j correspondentes. Por este motivo, denominaremos estes vetores-código de *incompletos*, enquanto aqueles definidos por (3.26) serão chamados de *completos*.

Complementamos a definição (3.26) para os vetores-código incompletos assumindo que eles sejam periódicos de períodos iguais aos seus respectivos índices. Resultam,

pois, definidos como

$$d_j(n) = \begin{cases} b(K - j + n) & n = 1, \dots, j \\ b(K - 2j + n) & n = j + 1, \dots, L \end{cases} \quad (3.27)$$

A busca da excitação é executada pelo mesmo processo descrito na seção 3.4 para o livro de códigos fixo. Isto é, seleciona-se o vetor-código $d = d_P$ tal que

$$P = \arg \left(\min_{j=J, \dots, K} \varepsilon_j \right), \quad (3.28)$$

sendo ε_j o erro quadrático ponderado de cada reconstrução (Eq.(3.11)).

Finalizando, o livro de códigos atrasados é tão importante que um codificador, o “vocoder” auto-excitado (SEV) [39], funciona somente com este tipo de estrutura.

3.8.2 Livros de códigos ceifados centralmente

Os livros de códigos com ceifagem central são gerados a partir de um livro estocástico gaussiano de média nula e variância unitária.

Gera-se cada vetor-código ceifado f_i pela aplicação ao seu vetor-código estocástico associado c_i da função de ceifagem central (“center-clipping”)

$$f_i(n) = \begin{cases} c_i(n) & \text{se } |c_i(n)| > n_{cc} \\ 0 & \text{caso contrário} \end{cases} \quad (3.29)$$

onde

- n_{cc} é o nível de ceifagem central;
- L é o comprimento do sub-bloco de excitação e
- M é o número de vetores em cada livro de códigos.

Para a geração de um livro de códigos ternário, quantizam-se os vetores-código ceifados centralmente com a função signum, ou seja,

$$t_i(n) = \begin{cases} \text{sign}(f_i(n)) & n = 1, \dots, L \\ & i = 1, \dots, M \end{cases} \quad (3.30)$$

Um valor usual para o nível de ceifagem central é $n_{cc} = 1,2$, gerando vetores-código com 76% de esparsidade [10].

Os vetores-código ceifados assim como os ternários têm espectros mais brancos do que os gaussianos originais.

3.9 Busca em dois livros de códigos

Os codificadores com dois livros de códigos requerem uma análise muito mais complexa do que aquela dos codificadores de livro único se o objetivo for uma busca ótima.

Na busca ótima em dois livros de códigos, para cada par de vetores-código

$$(c_i, d_j), \quad 1 \leq i \leq M, \quad J \leq j \leq K,$$

onde M e $K - J + 1$ são, respectivamente, o número de níveis dos livros de códigos 1 e 2. Estamos assumindo que o livro 2 seja adaptativo de excitações atrasadas, como é usual.

Como um processo desta complexidade é praticamente irrealizável, recorre-se a buscas seqüenciais. Às vezes, implementam-se algumas correções, tentando aproximar-se da solução ótima.

Examinaremos três procedimentos de busca em 2 livros de códigos:

- Buscas seqüenciais simples;
- Buscas seqüenciais com determinação conjunta dos ganhos;
- Buscas seqüenciais com ortogonalidade.

3.9.1 Buscas seqüenciais simples

Uma busca seqüencial simples em dois livros de códigos compõe-se apenas de duas buscas, sendo uma em cada livro.

A primeira busca usa o vetor-alvo u (Fig. 3.13 e Fig. 3.14), definido pela Eq.(3.21).

A segunda busca usa como vetor-alvo o erro residual da 1ª busca (Fig. 3.14), que é dado por

$$r = u - \hat{u}. \quad (3.31)$$

O erro quadrático ponderado ao final da 2ª busca (Fig. 3.14) é

$$e = r - \tilde{r}. \quad (3.32)$$

As buscas da excitação seguem o processo descrito na seção 3.4.

Primeiro, busca-se o vetor-código atrasado $d = d_P$ (seção 3.8.1).

O vetor de voz ponderada gerado pelo vetor atrasado d é

$$q = Hd. \quad (3.33)$$

O vetor reconstruído \hat{u} é a projeção euclidiana do vetor-alvo u sobre o vetor ponderado q , dada por

$$\hat{u} = \frac{u^T q}{q^T q} q. \quad (3.34)$$

O ganho g_P é definido pela coordenada da projeção \hat{u} na direção de projeção q , ou seja,

$$\hat{u} = g_P q. \quad (3.35)$$

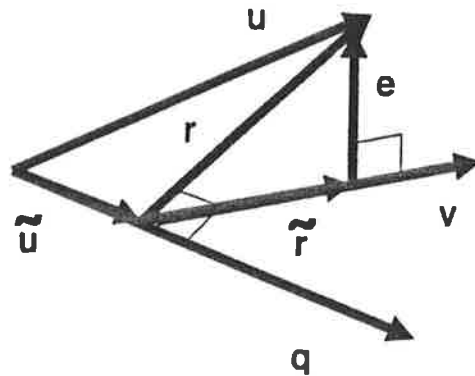


Figura 3.14: Busca seqüencial com vetor-alvo u e vetor-alvo residual r

Comparando a Eq.(3.35) com a Eq.(3.34), deduzimos que o ganho é

$$g_P = \frac{u^T q}{q^T q}. \quad (3.36)$$

Para a busca no livro de códigos fixo, executamos os mesmos procedimentos:

- busca do vetor-código $c = c_i$ (Fig. 3.13);
- filtragem de c , resultando no vetor de voz ponderada

$$v = Hc; \quad (3.37)$$

- reconstrução do vetor-alvo r baseada em v , dada pelo vetor

$$\tilde{r} = \frac{r^T v}{\|v\|^2} v; \quad (3.38)$$

- determinação do ganho g tal que

$$\tilde{r} = gv, \quad (3.39)$$

que resulta

$$g = \frac{r^T v}{\|v\|^2}. \quad (3.40)$$

Após as duas buscas, podemos reconstruir o vetor-alvo u pela filtragem da excitação total

$$x_0 = g_P d + g c, \quad (3.41)$$

resultando na reconstrução

$$\hat{u}_0 = g_P q + g v. \quad (3.42)$$

3.9.2 Buscas seqüenciais com determinação conjunta dos ganhos

No método de buscas seqüenciais simples, devido às projeções realizadas (Eq. (3.34) e Eq.(3.38)), temos

$$r \perp q \iff r^T q = 0 \quad (3.43)$$

$$e \perp v \iff e^T v = 0. \quad (3.44)$$

Porém, com as condições (3.43) e (3.44) apenas, não é possível garantir que o erro final e seja ortogonal ao plano $[q \ v]$ (Fig. 3.14) porque não sabemos se a ortogonalidade “ $e \perp q$ ” se verifica ou não.

Para garantir a ortogonalidade de e a $[q \ v]$, dada a Eq.(3.44), basta impor a condição

$$e \perp q \iff e^T q = 0. \quad (3.45)$$

Revedo o procedimento descrito na seção 3.9.1, verificamos que podemos melhorar a reconstrução (3.42) se, em vez de tomar ganhos g_P e g pelos seus valores dados em (3.36) e (3.40), respectivamente, usarmos suas expressões gerais (3.35) e (3.39).

Da Fig. 3.14, extraímos a condição de decomposição do vetor-alvo

$$u = \hat{u} + \hat{r} + e, \quad (3.46)$$

onde, introduzindo os ganhos através das Eqs.(3.35) e (3.39), temos

$$u = g_P q + g v + e. \quad (3.47)$$

Tomando os produtos internos membro a membro de (3.47) por q e por v separadamente, obtemos

$$u^T q = g_P (q^T q) + g (v^T q) + (e^T q) \quad (3.48)$$

$$u^T v = g_P (q^T v) + g (v^T v) + (e^T v), \quad (3.49)$$

respectivamente.

Impondo as condições (3.44) e (3.45) às equações (3.48) e (3.49), obtemos o sistema de equações lineares

$$\begin{bmatrix} q^T q & q^T v \\ q^T v & v^T v \end{bmatrix} \begin{bmatrix} g_P \\ g \end{bmatrix} = \begin{bmatrix} u^T q \\ u^T v \end{bmatrix}, \quad (3.50)$$

cujas resoluções fornece os novos valores dos ganhos.

3.9.3 Buscas seqüenciais com ortogonalidade

Na busca com ganhos determinados conjuntamente da seção anterior, a ortogonalidade do vetor-erro ao plano $[q \ v]$ dos vetores ponderados é imposta *a posteriori*.

Podemos aproximar um pouco mais as buscas seqüenciais da busca conjunta ótima se a 2ª busca já for realizada com a imposição da ortogonalidade.

A busca adaptativa com vetor-alvo u , resultando no vetor ponderado q e no ganho g_P , é executada inicialmente como na seção 3.9.2.

A busca no livro de códigos fixo envolve as seguintes etapas (Fig. 3.15):

- cada vetor-código c_j é filtrado, produzindo o vetor ponderado

$$v_j = Hc_j; \quad (3.51)$$

- o plano $[q \ v_j]$ é representado por uma base ortogonal através da determinação da componente w_j de v_j ortogonal a q , que é

$$w_j = v_j - \frac{q^T v_j}{q^T q} q; \quad (3.52)$$

- projetando-se o vetor-alvo u no vetor ortogonal w_j , obtém-se o vetor

$$t_j = \frac{w_j^T u}{w_j^T w_j} w_j, \quad (3.53)$$

cujas norma quadrada é

$$\|t_j\|^2 = \frac{(w_j^T u)^2}{w_j^T w_j}; \quad (3.54)$$

- seleciona-se o vetor-código c_i de índice i tal que

$$i = \arg \left(\max_{j=1, \dots, M} \|t_j\|^2 \right); \quad (3.55)$$

- determinação conjunta dos ganhos g_P e g pelo sistema (3.50), com $v = Hc_j$, resultando a excitação total

$$x = g_P q + g v.$$

O critério de busca (3.55) aqui utilizado é equivalente àquele do CELP original, enunciado pela Eq.(3.12) da seção 3.4, conforme mostramos em seguida.

Tomando o ganho g_P da busca adaptativa inicial e a projeção t_j do vetor-alvo u em cada vetor ortogonal w_j dada pela Eq.(3.53), podemos expressar a correspondente projeção de u em cada plano $[q \ v_j]$ como

$$\hat{u}_j = g_P q + t_j. \quad (3.56)$$

A cada projeção, o vetor-alvo u pode ser recomposto (Fig. 3.15) como

$$u = \hat{u}_j + e_j, \quad (3.57)$$

onde e_j é o vetor-erro correspondente.

Considerando que, por construção,

$$e_j \perp \hat{u}_j,$$

tomando o produto interno de cada membro da Eq.(3.57) por si próprio, obtemos

$$\|u\|^2 = \|\hat{u}_j\|^2 + \varepsilon_j, \quad (3.58)$$

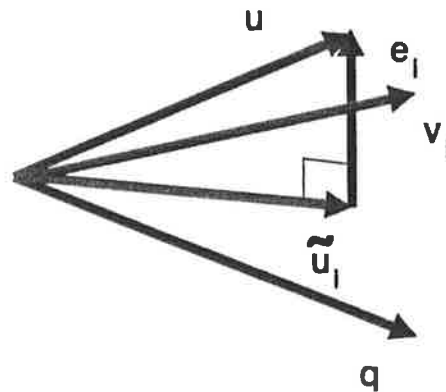


Figura 3.15: Buscas sequenciais com ortogonalidade

onde

$$\varepsilon_j = \|e_j\|^2$$

é o erro quadrático total associado a cada vetor-código c_j .

Como durante o processo de busca $\|u\|^2$ é constante, decorre da condição (3.58) que a ocorrência do mínimo de ε_j coincide com a ocorrência do máximo de $\|\tilde{u}_j\|^2$, ou seja,

$$\arg \left(\min_{j=1, \dots, M} \varepsilon_j \right) = \arg \left(\max_{j=1, \dots, M} \|t_j\|^2 \right). \quad (3.59)$$

Portanto, o critério de busca pelo menor erro quadrático ponderado total (Eq. (3.12) da seção 3.4) equivale ao critério de busca pela maior projeção (Eq.(3.55)).

3.10 Excitação por multipulsos (MPE)

A excitação por multipulsos parte da consideração de que a excitação LPC com um pulso por período fundamental não é suficiente para representar o sinal de erro de predição (seção 2.5).

A excitação é composta por vários pulsos de localização e de amplitude variáveis, ocorrendo tipicamente à taxa de 8 por 10 ms [6].

Esta é uma definição muito mais genérica da excitação. Por outro lado, a sua determinação tem que ser simplificada para torná-la viável.

Uma busca sub-ótima da excitação por multipulsos consiste na determinação da localização e da amplitude de um pulso por vez [6].

A excitação por pulsos regulares (RPE) [27] também simplifica a busca. Ela consiste em pulsos uniformemente espaçados, tipicamente á taxa de 10 por 5 ms, onde o problema se reduz à localização do primeiro pulso e à determinação das amplitudes de todos os pulsos.

O codificador RPE foi adotado como padrão na telefonia móvel digital da Europa, operando á taxa de 13 kb/s [47].

Numa outra estruturação da busca, os pulsos podem ser decompostos numa parte previsível e noutra imprevisível como no codificador proposto em [21].

3.11 Excitação por soma vetorial (VSELP)

O codificador preditivo de voz excitado por soma vetorial (VSELP) contém um livro de códigos atrasados adaptativo e dois livros de códigos fixos [16].

O vetor de excitação total num sub-bloco (Fig. 3.16) é dado por

$$x = \beta d + \gamma_1 c_1 + \gamma_2 c_2, \quad (3.60)$$

onde

- d é o vetor-código atrasado e β é o seu ganho;
- c_1 e c_2 são vetores-código selecionados de cada um dos dois livros de códigos fixos, com ganhos γ_1 e γ_2 , respectivamente.

Ademais, os livros de códigos fixos 1 e 2 são definidos pelos conjuntos de vetores-base

$$\{b_{1k}, \quad k = 1, \dots, B\}$$

e

$$\{b_{2k}, \quad k = 1, \dots, B\}$$

respectivamente.

Os vetores-código c_1 e c_2 são produzidos como

$$c_1 = \sum_{k=1}^B \theta_{1k} b_{1k} \quad (3.61)$$

$$c_2 = \sum_{k=1}^B \theta_{2k} b_{2k} \quad (3.62)$$

respectivamente.

As coordenadas θ_{lk} assumem apenas valores binários

$$\theta_{lk} \in \{-1, 1\}, \quad l = 1, 2.$$

Assim, cada coordenada pode ser representada por um bit. Definimos

$$\nu_{lk} = 1 \leftrightarrow \theta_{lk} = 1; \quad (3.63)$$

$$\nu_{lk} = 0 \leftrightarrow \theta_{lk} = -1. \quad (3.64)$$

Os bits ν_{1k} e ν_{2k} são os dígitos binários dos índices i_1 e i_2 , respectivamente, ou seja, estes índices são representados binariamente pelas seqüências de bits

$$i_1 = \sum_{k=1}^B \nu_{1k} 2^{k-1} \quad (3.65)$$

$$i_2 = \sum_{k=1}^B \nu_{2k} 2^{k-1}. \quad (3.66)$$

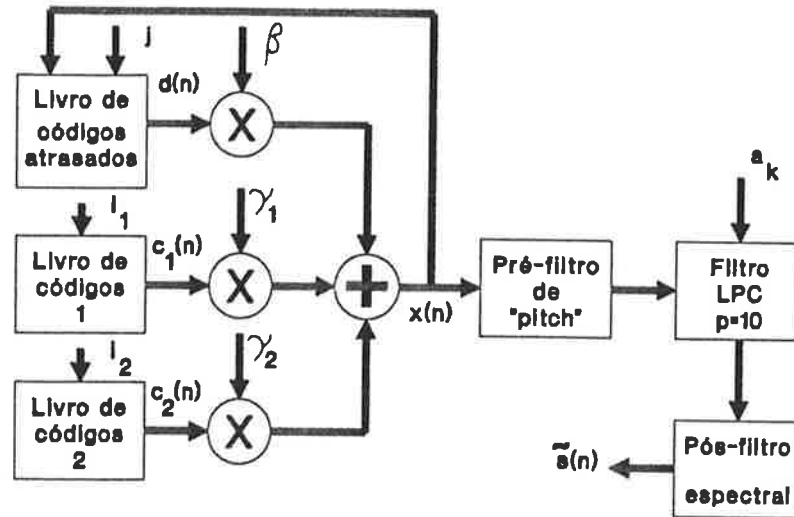


Figura 3.16: Decodificador de voz VSELP
(adaptada de [16]).

O pré-filtro de “pitch” (Fig. 3.16) acentua a periodicidade da excitação gerada e o pós-filtro espectral realça a voz [16]. Nós não usamos nenhum desses dois filtros nas nossas simulações.

No VSELP original [16], a dimensão de cada livro de códigos fixo é

$$B = 7.$$

Capítulo 4

METODOLOGIA

Os codificadores de voz testados foram simulados no ambiente de programação MATLAB.

Para avaliação de desempenho, escolhemos sinais de voz típicos e duas medidas de distorção.

Para treinamento e teste dos codificadores pesquisados, usamos três sinais:

- v** – é um trecho de voz sonora masculina de 320 ms de duração, amostrado à taxa de 8 kHz com com quantização PCM lei μ de 8 bits linearizada em 13 bits [23], que corresponde às duas primeiras sílabas da frase “*um galo*”;
- c** – é um sinal artificial de voz masculina de 10 s de duração, gerado conforme recomendação do CCITT [12] à taxa de 16 kHz e reamostrado à frequência de 8 kHz;
- d** – são os dígitos de 0 a 9, pronunciados em português por 3 locutores masculinos e 2 femininos e amostrados à taxa de 8 kHz com precisão de 12 bits.

O sinal **v** serve para detetar dificuldades dos codificadores com voz sonora mais rapidamente porque é um sinal breve.

O sinal **c** foi usado na expectativa de se obter uma cobertura razoável da variabilidade estatística da voz com uma quantidade de dados bem restrita.

O sinal **d** permite uma exposição dos codificadores a variações mais naturais do sinal de voz.

Escolhemos duas medidas de distorção: a relação sinal-ruído segmentada (SNRSEG) (seção 4.1) e a distância cepstral alisada (CD) (seção 4.3).

A relação sinal-ruído segmentada foi escolhida pela sua popularidade como medida objetiva de qualidade. Ela mede o ajuste entre a forma de onda original e a reconstruída (seção 4.1). Não é, portanto, a mais apropriada para os codificadores CELP, que se baseiam em critérios de percepção auditiva.

A distorção cepstral, uma medida objetiva de distorção espectral, foi escolhida para estimativa da qualidade subjetiva.

Os testes subjetivos são indispensáveis para a avaliação definitiva de codificadores de voz. Entretanto, os testes objetivos viabilizam testes com muitas combinações de variáveis.

Há também outras medidas de distorção espectral que estão sendo propostas como estimativas da qualidade subjetiva [49].

4.1 Relação sinal-ruído segmentada

A relação sinal-ruído de reconstrução do sinal $\{s(n), n = 0, 1, \dots, NM - 1\}$ pelo sinal $\{\hat{s}(n), n = 0, 1, \dots, NM - 1\}$ é uma medida quadrática da fidelidade de reprodução da forma de onda, dada em forma logarítmica por

$$\text{SNR} = 10 \log_{10} \frac{\sum_{n=0}^{NM-1} s^2(n)}{\sum_{n=0}^{NM-1} (\hat{s}(n) - s(n))^2}. \quad (4.1)$$

A relação sinal-ruído segmentada (SNRSEG) de comprimento K é a média aritmética das relações sinal-ruído $\{\text{SNR}(m), m = 0, \dots, M - 1\}$ (4.1) entre todos os M segmentos correspondentes dos dois sinais. Isto é,

$$\text{SNRSEG} = \frac{1}{M} \sum_{m=0}^{M-1} \text{SNR}(m). \quad (4.2)$$

A vantagem de usar-se a SNRSEG em vez da SNR simplesmente consiste no realce que a ponderação logarítmica da SNRSEG dá aos segmentos de baixa energia. Assim, ela está mais de acordo com a percepção subjetiva de qualidade.

Seguindo sugestão dada em [23], estabelecemos segmentos de duração $M = 128$, ou seja, 16 ms.

4.2 Cepstro da resposta impulsiva do filtro LPC

O cepstro de um sinal é a transformada de Fourier inversa do logaritmo do seu espectro de potência. Adaptando [38], definimos o cepstro da seqüência $\{x(n)\}$ como a seqüência

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |X(e^{j\omega})|^2 e^{j\omega n} d\omega, \quad (4.3)$$

onde

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n) e^{-j\omega n}. \quad (4.4)$$

A transformada de Fourier $C(e^{j\omega})$ do cepstro $\{c_n\}$ é real pela definição (4.3) e, assim, pode ser expressa como

$$C(e^{j\omega}) = \sum_{n=-\infty}^{\infty} c_n e^{-j\omega n}, \quad (4.5)$$

com

$$c_{-n} = c_n, \quad n = 1, 2, \dots \quad (4.6)$$

Seja $\{c_n\}$ o cepstro da resposta impulsiva $\{h(n)\}$ do filtro LPC $H(z)$ (Cap. 2)

$$H(z) = \frac{1}{A(z)}, \quad (4.7)$$

com

$$A(z) = 1 + \sum_{k=1}^p a_k z^{-k}. \quad (4.8)$$

Então, a transformada de Fourier $C(e^{j\omega})$ do cepstro $\{c_n\}$ é

$$C(e^{j\omega}) = \ln |H(e^{j\omega})|^2 \quad (4.9)$$

$$= \ln (H(e^{j\omega})H^*(e^{j\omega})), \quad (4.10)$$

onde $H(e^{j\omega})$ é a resposta em frequência de $H(z)$.

Extendendo a Eq.(4.10) no plano complexo z [34], obtemos

$$C(z) = \ln (H(z)H^*(1/z^*)). \quad (4.11)$$

O filtro LPC $H(z)$ (4.7) tem todos os seus zeros em $z = 0$, podendo ter sua estabilidade garantida na geração (Apêndice B). Portanto, ele é de mínima fase, isto é, seus pólos e seus zeros estão todos no interior do círculo unitário. Assim, na separação membro a membro de (4.11) nas suas componentes causal e anticausal, $H(z)$ está associada a $C_0(z)$ através de

$$C_0(z) = \ln H(z), \quad (4.12)$$

onde $C_0(z)$ é a transformada z da série unilateral causal $\{c_n, n = 0, 1, \dots\}$ ([36] *apud* [19]) que entra na composição da série cepstral bilateral $\{c_n\}$, e é explicitamente

$$C_0(z) = \sum_{k=0}^{\infty} c_k z^{-k}. \quad (4.13)$$

Aplicando a definição (4.7) a (4.12), temos

$$C_0(z) = -\ln A(z), \quad (4.14)$$

que é uma relação muito importante. Conforme veremos em seguida, ela permite determinar o cepstro da resposta impulsiva do filtro LPC sem recorrer às transformadas de Fourier requeridas pela definição nas Eqs.(4.3) e (4.4).

Tomando $z^{-1} = 0$ em (4.14), temos

$$C_0(\infty) = -\ln A(\infty), \quad (4.15)$$

que, dadas as definições (4.13) e (4.8), fornece

$$c_0 = 0, \quad (4.16)$$

permitindo reescrever a transformada z (4.11) do cepstro LPC causal como

$$C(z) = \sum_{k=1}^{\infty} c_k z^{-k}, \quad (4.17)$$

Tomando as derivadas de cada membro da Eq.(4.14) em relação a z^{-1} , temos

$$\frac{dC(z)}{d(z^{-1})} = -\frac{1}{A(z)} \frac{dA(z)}{d(z^{-1})}, \quad (4.18)$$

que, pela substituição de (4.8) e de (4.17), produz, com um rearranjo de termos,

$$-\sum_{k=1}^p k a_k z^{-(k-1)} = \left(\sum_{k=1}^{\infty} k c_k z^{-(k-1)} \right) \left(1 + \sum_{k=1}^p a_k z^{-k} \right). \quad (4.19)$$

Igualando membro a membro os coeficientes das potências

$$z^0, z^{-1}, z^{-2}, \dots, z^{-p}, z^{-(p+1)}, \dots,$$

em (4.19), obtemos

$$\begin{aligned} -a_1 &= c_1 \\ -2a_2 &= c_1 a_1 + 2c_2 \\ -3a_3 &= c_1 a_2 + 2c_2 a_1 + 3c_3 \\ &\vdots \\ 0 &= c_1 a_p + 2c_2 a_{p-1} + \dots + p c_p a_1 + (p+1)c_{p+1} \\ 0 &= 2c_2 a_p + 3c_3 a_{p-1} + \dots + (p+1)c_{p+1} a_1 + (p+2)c_{p+2} \end{aligned} \quad (4.20)$$

$$\vdots \quad (4.21)$$

Podemos resumir as equações (4.20) no seguinte algoritmo recursivo para conversão dos coeficientes preditivos nos coeficientes cepstrais:

$$\begin{aligned} c_1 &= -a_1 \\ c_l &= -a_l - \frac{1}{l} \sum_{k=1}^{l-1} k c_k a_{l-k} \quad l = 2, 3, \dots, p \\ c_l &= -\frac{1}{l} \sum_{k=1}^p (l-k) c_{l-k} a_k \quad l = p+1, p+2, \dots \end{aligned} \quad (4.22)$$

4.3 Distorção cepstral LPC

Nesta seção, estamos interessados na medida da distorção existente entre dois modelos espectrais LPC.

Sejam dois modelos espectrais LPC dados por

$$\begin{aligned} H(e^{j\omega}) &= \frac{G}{1 + \sum_{k=1}^p a_k e^{-j\omega k}} \\ H'(e^{j\omega}) &= \frac{G'}{1 + \sum_{k=1}^p a'_k e^{-j\omega k}} \end{aligned} \quad (4.23)$$

A norma L_2 aplicada à diferença entre os espectros logarítmicos de potência dos modelos acima [19] fornece a medida de distorção d_2 dada por

$$d_2^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\ln |H(e^{j\omega})|^2 - \ln |H'(e^{j\omega})|^2|^2 d\omega. \quad (4.24)$$

Substituindo (4.9) e (4.5) em (4.24), temos

$$d_2^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{k=-\infty}^{\infty} (c_k - c'_k) e^{-j\omega k} \right|^2 d\omega, \quad (4.25)$$

onde

$$c_0 = \ln G \quad \text{e} \quad c'_0 = \ln G'. \quad (4.26)$$

A Eq.(4.25), mediante a aplicação do teorema de Parseval, se transforma em

$$d_2^2 = \sum_{k=-\infty}^{\infty} (c_k - c'_k)^2, \quad (4.27)$$

que, considerando a propriedade (4.6), resulta em

$$d_2^2 = (c_0 - c'_0)^2 + 2 \sum_{k=1}^{\infty} (c_k - c'_k)^2. \quad (4.28)$$

Podemos simplificar a norma logarítmica quadrática d_2 pelo truncamento do cepstro, que, além de permitir o uso prático da Eq.(4.28), causa um alisamento dos espectros comparados (Fig. 4.1).

Pelas relações de conversão (4.22), os coeficientes cepstrais c_1, c_2, \dots, c_p determinam completamente os coeficientes de predição e, portanto, o filtro LPC. Assim, eles constituem o menor subconjunto de coeficientes cepstrais que tem que ser mantido para uma medida de distorção entre modelos LPC de ordem p .

Ademais, estamos mais interessados em medir distorções de forma espectral do que de nível de potência. Por isso, propomos a eliminação do termo $c_0 - c'_0$, que representa a diferença de ganho logarítmico entre os modelos (Eq.(4.26)).

Implementando as duas medidas simplificadoras acima na Eq.(4.28), obtemos a seguinte aproximação para d_2^2

$$\tilde{d}_2^2 = 2 \sum_{k=1}^p (c_k - c'_k)^2. \quad (4.29)$$

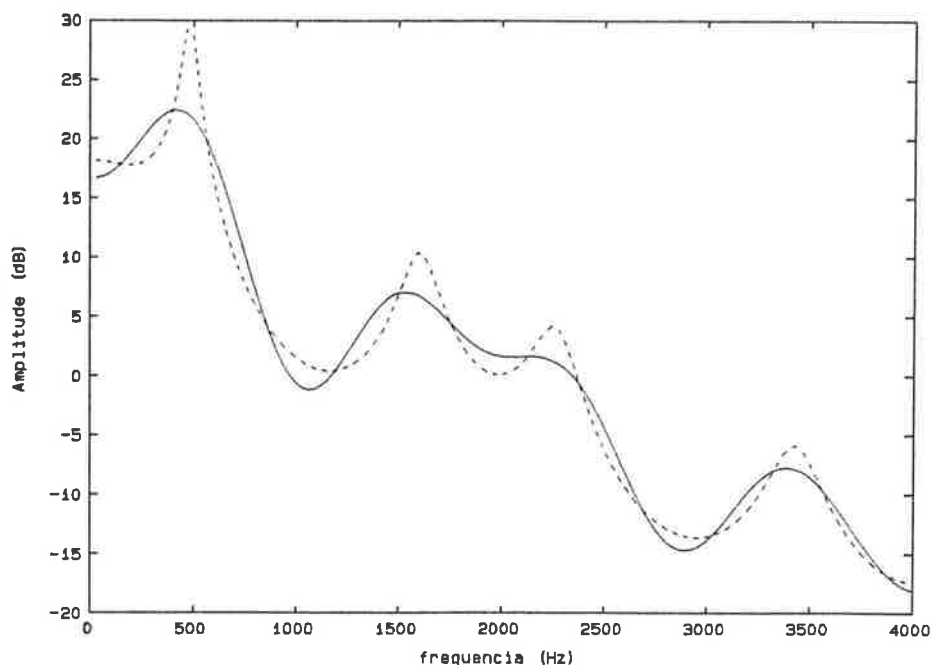


Figura 4.1: Envoltória espectral LPC-10 logarítmica de um segmento sonoro do sinal v e sua reconstrução pelo cepstro truncado $\{c_n, n = -10, -9, \dots, 10\}$.

Finalmente, é conveniente expressar a distorção em dB. Considerando a definição (4.24) de d_2 , devemos aplicar o fator multiplicativo $10/\ln 10$ à Eq.(4.29). Resulta, assim, a seguinte expressão [25] para a distorção cepstral alisada

$$CD = \frac{10}{\ln 10} \sqrt{2 \sum_{k=1}^p (c_k - c'_k)^2}. \quad (4.30)$$

Nas medidas realizadas neste trabalho para determinar a distorção cepstral entre os sinais $\{s(n), n = 0, 1, \dots, NM - 1\}$ e $\{s'(n), n = 0, 1, \dots, NM - 1\}$, seguiu-se o seguinte procedimento:

- i) segmentação dos dois sinais em vetores justapostos de dimensão $N = 128$;
- ii) análise LPC de cada par de segmentos correspondentes dos dois sinais, determinando-se os vetores de coeficientes de predição

$$[a_1 \ a_2 \ \dots \ a_p] \text{ e } [a'_1 \ a'_2 \ \dots \ a'_p],$$

que definem os modelos (4.23), respectivamente;

- iii) cálculo dos vetores de coeficientes cepstrais

$$[c_1 \ c_2 \ \dots \ c_p] \text{ e } [c'_1 \ c'_2 \ \dots \ c'_p]$$

de cada par de segmentos correspondentes dos dois sinais usando as relações (4.22);

Tabela 4.1: Escalas de qualidade do índice MOS
(extraída de [23]).

Conceitos numéricos	Escala de qualidade	Escala de degradação
5	Excelente	Imperceptível
4	Bom	(Apenas) perceptível mas não incômoda
3	Regular	(Perceptível e) ligeiramente incômoda
2	Fraco	Incômoda
1	Insatisfatório	Muito incômoda

- iv) determinação da distorção cepstral $CD(m)$ entre os m -ésimos segmentos dos sinais $\{s(n)\}$ e $\{s'(n)\}$ pela Eq.(4.30);
- v) atribuição do valor médio à distorção cepstral entre os sinais

$$\{s(n), n = 0, 1, \dots, NM - 1\} \text{ e } \{s'(n), n = 0, 1, \dots, NM - 1\}$$

conforme

$$CD = \frac{1}{M} \sum_{m=1}^M CD(m). \quad (4.31)$$

4.4 Medida subjetiva de qualidade da voz

Para o cálculo de índices estatísticos de qualidade subjetiva da voz é conveniente associar uma escala de conceitos numéricos a descrições adjetivais [23] (Tabela 4.1).

A média aritmética dos conceitos numéricos atribuídos por um grupo selecionado de ouvintes a um sinal de voz é o seu “Mean Opinion Score” (MOS). O índice MOS também pode ser estimado a partir de testes cuja tarefa básica seja uma comparação entre dois sinais de voz.

A Figura 4.2 apresenta a variação da qualidade subjetiva de voz produzida a várias taxas de transmissão por codificadores de três níveis de complexidade e pelos “vocoders” (seção 2.2).

Este trabalho trata principalmente de codificadores de alta complexidade operando a taxas de transmissão em torno de 8 kHz.

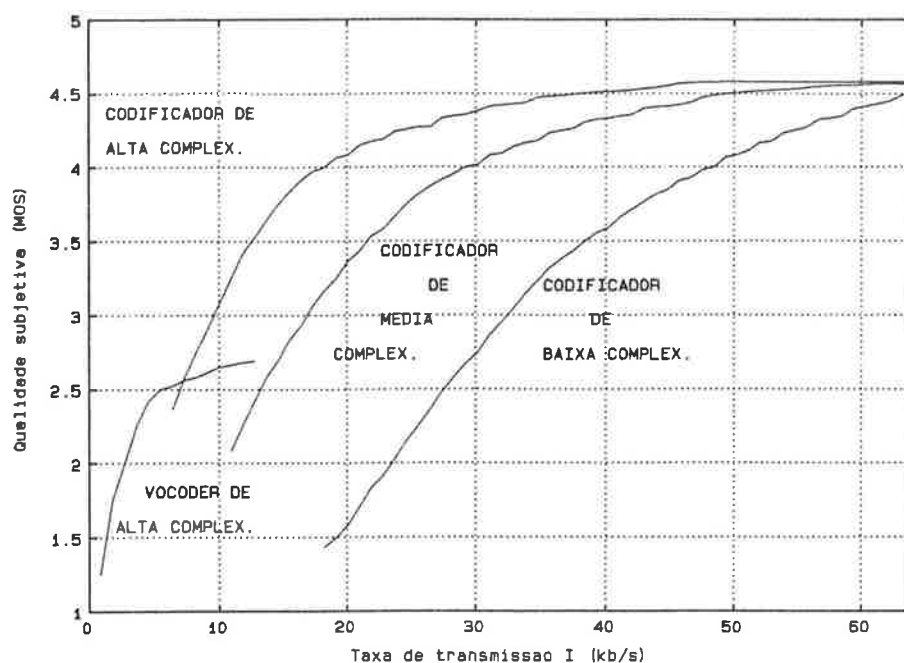


Figura 4.2: Qualidade subjetiva de voz digitalizada (extraída de [23])

4.5 Estimação objetiva da qualidade subjetiva

Em estudos comparativos [24] *apud* [25] com vários tipos de medidas objetivas de qualidade da voz, concluiu-se que

- i) as medidas espectrais de qualidade têm melhor correspondência com as medidas subjetivas do que as medidas temporais;
- ii) dentre as medidas espectrais, aquelas baseadas na envoltória do espectro guardam mais correspondência com as medidas subjetivas e
- iii) a medida de distorção espectral baseada no cepstro LPC (Eq.(4.30)) têm a melhor correspondência com as medidas subjetivas dentre todos os métodos de cálculo da envoltória espectral baseados em técnicas LPC.

Confrontando o índice MOS (seção 4.4) obtido em testes subjetivos com a medida objetiva de distorção CD (seção 4.3), os autores de [25] determinaram a seguinte curva de ajuste

$$\widehat{\text{MOS}} = 0,04\text{CD}^2 - 0,80\text{CD} + 3,56, \quad (4.32)$$

que usamos para calcular o $\widehat{\text{MOS}}$, estimação do MOS a partir da CD, dada pela Eq.(4.31).

4.6 Orientação da pesquisa

Na etapa inicial de simulações, testamos codificadores de voz baseados em quatro técnicas usuais:

- i) livros de códigos estocásticos;
- ii) livros de códigos esparsos;
- iii) buscas ortogonais;
- iv) livros de códigos definidos por bases de vetores.

Os codificadores preditivos por análise-mediante-síntese implementados possuem dois ou três livros de códigos, sendo o primeiro adaptativo e constituído por vetores-código atrasados (seção 3.8.1). A natureza do segundo livro de códigos caracteriza cada codificador:

Estocástico usa o livro de códigos proposto originalmente por Atal e Schroeder (seção 3.8), constituído de $M = 1024$ vetores-código aleatórios gaussianos de média nula e de variância unitária;

Estocástico ortogonal contém o mesmo livro de códigos fixo do codificador acima, porém, a sua busca é feita com ortogonalidade à busca adaptativa (seção 3.9.3);

Esparso usa o livro de vetores-código ceifados centralmente definido na seção 3.8.2;

Esparso ortogonal contém o mesmo livro de códigos fixo do codificador acima, porém, a sua busca é feita com ortogonalidade à busca adaptativa (seção 3.9.3);

Vetores-base usa dois livros de códigos, sendo cada um definido por uma base de $B = 7$ vetores (seção 3.11) aleatórios distribuídos normalmente, com média nula e variância unitária.

Com exceção do codificador “esparso”, todos os codificadores executam a segunda busca ortogonalmente à primeira (seção 3.9.3).

O codificador com excitação por “vetores-base” emprega uma janela de Hamming de largura $N = 160$ sem sobreposição para alisamento espectral. Os demais codificadores efetuam uma expansão de faixa de $\Delta B = 15$ Hz.

A partir da Eq.(D.24), temos o coeficiente de expansão de faixa

$$\gamma = \exp(-\pi\Delta B/f_a), \quad (4.33)$$

resultando em $\gamma = 0,994$.

Em testes de codificação do sinal de voz v , verificamos que o codificador CELP esparso ortogonal teve o melhor desempenho. O VSELP foi o segundo melhor, ultrapassando o CELP estocástico em desempenho subjetivo (Fig. 4.3).

Este resultado confirma o paradoxo dos codificadores esparsos citado por [26]: eles são menos complexos e alcançam uma qualidade subjetiva superior aos estocásticos.

Os codificadores esparsos são passíveis de implementação com menor complexidade computacional do que os CELPs estocásticos. Mesmo assim, apresentam um desempenho suficientemente bom para que um codificador usando esta estrutura tenha sido adotado como padrão de telefone seguro a 4,8 kbit/s pelo Departamento de Defesa dos EUA [10].

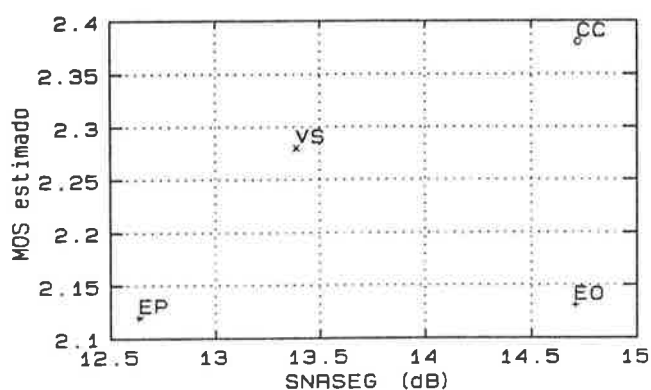


Figura 4.3: Teste de técnicas de codificação e busca da excitação com o sinal de voz v .

Tabela 4.2: Codificadores CELP de várias técnicas de codificação e de busca da excitação.

Codificador	Descrição
EO	Estocástico ortogonal
EP	Esparso (ceifagem central)
CC	Esparso (ceifagem central) ortogonal
VS	Base vetorial (VSELP)

Confirma-se a importância das buscas com ortogonalidade pela grande diferença de desempenho dos codificadores “esparso” e “esparso ortogonal”.

O “codificador definido por base vetorial” apresenta um desempenho intermediário entre o “esparso ortogonal” e o “estocástico”. Ele ainda possui a vantagem de ter uma estrutura mais organizada, que possibilita buscas de menor complexidade. Um codificador deste tipo, o VSELP (seção 3.11), é usado em telefonia celular digital na América do Norte.

Porém, como os vetores-base do VSELP não são ortogonais entre si e as representações ortogonais são menos sensíveis à codificação se escolhidas apropriadamente [23], resolvemos pesquisar propostas de representações ortogonais da excitação.

Capítulo 5

CODIFICADOR PROPOSTO: CELP-SVD

A enorme complexidade do codificador CELP na sua definição original (capítulo 3) motivou a pesquisa de métodos mais simples de realização da busca de vetores-código para a excitação do filtro LPC.

Pode-se tornar a busca mais eficiente tanto pela redução da quantidade de vetores-código usados como pela simplificação do processo de filtragem.

A representação da excitação por base de vetores possibilita a redução do número de vetores-código pela conjugação de dois processos:

- i) redução da dimensão do vetor de excitação;
- ii) quantização eficiente das coordenadas associadas aos vetores-base da representação.

Um codificador de destaque com excitação definida por vetores-base é o VSELP (seção 3.11), que reduz a dimensão do vetor de excitação de $L = 40$ para $2B = 14$ e quantiza cada uma das $2B$ coordenadas com apenas um bit.

O VSELP com vetores-base aleatórios, como vimos na seção 4.6, sofre apenas uma pequena degradação de desempenho em comparação com o CELP. Projetando-se os vetores-base por treinamento, o VSELP supera o CELP [16].

Entretanto, o VSELP não simplifica a filtragem e os seus vetores-base são fixos e não ortogonais entre si.

Propomos a utilização de uma base vetorial definida pelos parâmetros do filtro LPC e que é redefinida adaptativamente a cada bloco de análise LPC.

O filtro LPC é usado para a determinação da base vetorial através da sua matriz de resposta impulsiva H (seção 3.7). A base vetorial é composta por vetores singulares à direita (RSVs) da matriz H .

Os RSVs são determinados pela matriz H através da sua *decomposição em valores singulares* (SVD) [18]. A estrutura decorrente desta decomposição matricial é interessante por três razões:

Completeza – a base ortogonal constituída pelo conjunto completo dos L RSVs de H

$$V = \{v_i, \quad i = 1, \dots, L\} \quad (5.1)$$

comporta a representação de qualquer vetor de excitação de comprimento L

$$x = \sum_{i=1}^L \xi_i v_i, \quad (5.2)$$

desde que as coordenadas ξ_i , $i = 1, \dots, L$ sejam expressas com precisão infinita.

Precisão arbitrária – as coordenadas ξ_i , $i = 1, \dots, L$ que caracterizam a excitação x (Eq.(5.2)) podem ser quantizadas com um número finito de bits, bem como podemos selecionar um subconjunto

$$V_M \subset V \quad (5.3)$$

de M RSVs para representar a excitação com um número reduzido de dimensões, admitindo uma distorção finita;

Adaptabilidade – a cada variação dos parâmetros do filtro LPC varia conjuntamente a matriz H e, para representações com precisão finita e/ou de dimensão reduzida, isto acarreta uma variação efetiva do livro de códigos de excitação.

Esta representação completa da excitação do filtro LPC foi proposta por Atal [7], que estudou o efeito da quantização das coordenadas singulares (CSs) ξ_i em codificadores que empregam a dimensão total do espaço das excitações.

Vamos estudar algumas propriedades de codificadores que operam em subespaços V_M de dimensões reduzidas, contidos no espaço V dos segmentos de excitação. Designaremos codificadores desta classe por CELP-SVD.

O codificador proposto também emprega outro livro de códigos adaptativo de excitações atrasadas.

Finalmente, serão abordadas algumas pesquisas com versões quantizadas do codificador CELP-SVD proposto.

5.1 Representação SVD do filtro LPC e da sua excitação

A decomposição em valores singulares (SVD) da matriz de resposta impulsiva H (seção 3.7) do filtro LPC expressa-se por

$$H = UDV^T, \quad (5.4)$$

onde

- os vetores singulares à direita (RSVs) são as colunas da matriz V , $L \times L$,

$$V = [v_1 \ v_2 \ \dots \ v_L] \quad (5.5)$$

e constituem uma base ortonormal do espaço dos segmentos de excitação;

- os vetores singulares à esquerda (LSVs) são as colunas da matriz U , $L \times L$,

$$U = [u_1 \quad u_2 \quad \cdots \quad u_L] \quad (5.6)$$

e constituem uma base ortonormal do espaço dos segmentos de voz e

- a matriz diagonal D , $L \times L$, de diagonal não negativa composta pelo conjunto de valores singulares de H

$$\text{diag}(D) = [d_1 \quad d_2 \quad \cdots \quad d_L], \quad (5.7)$$

dispostos em ordem decrescente

$$d_i \geq d_{i+1} \geq 0, \quad i = 1, \dots, L - 1. \quad (5.8)$$

Sendo H uma matriz triangular inferior com todos os elementos unitários na diagonal principal (seção 3.7), temos

$$\det(H) = 1. \quad (5.9)$$

Como as matrizes U e V contêm bases ortonormais (Eq.(5.6) e Eq.(5.5)), temos

$$U^T U = I \quad (5.10)$$

$$V^T V = I. \quad (5.11)$$

Considerando que, para qualquer matriz quadrada A ,

$$\det(A^T) = \det(A), \quad (5.12)$$

extraímos de (5.10) e (5.11) as propriedades

$$\det(U) = \det(V) = 1. \quad (5.13)$$

A partir de (5.4), (5.9) e (5.13), temos

$$\det(D) = 1. \quad (5.14)$$

Expandindo a Eq.(5.14), chegamos à seguinte condição global sobre os valores singulares da matriz de resposta impulsiva do filtro LPC com coeficiente de ganho unitário (Eq.(2.1))

$$\prod_{i=1}^L d_i = 1. \quad (5.15)$$

5.2 Representações do processo de filtragem

Apliquemos o segmento de excitação

$$\mathbf{x} = [x(1) \ x(2) \ \cdots \ x(L)]^T$$

ao filtro LPC $H(z)$ com condições iniciais nulas. Sendo $\mathbf{h}(n)$, $n = 0, 1, \dots$, a seqüência de resposta impulsiva de $H(z)$, obtemos a seqüência de saída $\mathbf{y}(n)$, dada pela convolução

$$\mathbf{y}(n) = \mathbf{h}(n-1) * \mathbf{x}(n), \quad n = 1, 2, \dots, \quad (5.16)$$

onde definimos

$$\mathbf{x}(n) = \begin{cases} x(n) & n = 1, \dots, L \\ 0 & \text{caso contrário} \end{cases}$$

como a seqüência de excitação.

Tomando-se apenas o vetor de saída

$$\mathbf{y} = [\mathbf{y}(1) \ \mathbf{y}(2) \ \cdots \ \mathbf{y}(L)]^T,$$

podemos escrever cada uma de suas componentes, a partir de (5.16), como

$$y(n) = \sum_{k=1}^n \mathbf{h}(n-k)x(k), \quad n = 1, \dots, L. \quad (5.17)$$

Podemos representar as Eq.(5.17) em forma vetorial usando a matriz de resposta impulsiva H (seção 3.7) do filtro $H(z)$, resultando em

$$\mathbf{y} = H\mathbf{x}. \quad (5.18)$$

Decompondo H em valores singulares conforme (5.4), temos

$$\mathbf{y} = UDV^T\mathbf{x}. \quad (5.19)$$

Tomando-se as coordenadas de \mathbf{y} na base dos vetores singulares á esquerda (LSVs) (Eq.(5.6)), temos

$$U^T\mathbf{y} = U^TUDV^T\mathbf{x}. \quad (5.20)$$

Considerando-se a ortonormalidade dos LSVs (Eq.(5.10)), podemos simplificar a Eq.(5.20) para

$$\theta = D\xi, \quad (5.21)$$

onde

$$\xi = V^T\mathbf{x} \quad (5.22)$$

$$\theta = U^T\mathbf{y} \quad (5.23)$$

são, respectivamente, as coordenadas da excitação e da saída expressas nas bases dos RSVs e dos LSVs.

Ainda podemos escrever a Eq.(5.21) escalarmente como

$$\theta_i = d_i \xi_i \quad i = 1, \dots, L, \quad (5.24)$$

em função dos valores singulares d_n (Eq.(5.7)).

Portanto, a convolução (5.17) simplificou-se na multiplicação ponto a ponto (5.24) através da decomposição de H em valores singulares e através da representação de x e de y nas bases dos vetores singulares.

A saída y pode ser construída a partir da Eq.(5.23) como

$$y = U\theta. \quad (5.25)$$

A norma quadrática do vetor de saída y é dada por

$$\|y\| = \sqrt{y^T y}, \quad (5.26)$$

que, com a substituição de (5.25), torna-se

$$\|y\| = \sqrt{\theta^T U^T U \theta}, \quad (5.27)$$

resultando da ortonormalidade (5.10) de U que

$$\|y\| = \sqrt{\theta^T \theta}. \quad (5.28)$$

A norma do vetor y pode ser relacionada com os valores singulares pela substituição de (5.24) em (5.28), que produz

$$\|y\| = \sqrt{\sum_{i=1}^L d_i^2 \xi_i^2}. \quad (5.29)$$

A Eq.(5.29) salienta que o efeito da filtragem é o aumento do peso relativo na construção de $\|y\|$ dos índices i de maiores valores singulares, que, adotadas as convenções (5.8), são aqueles cujos índices são menores.

5.3 Propriedades espectrais dos vetores singulares

Os espectros dos vetores singulares à direita (RSVs) de filtros LPC, em geral, têm aspecto passa-faixas [7]. Ademais, o conjunto completo dos RSVs cobre todo o espectro da frequência nula à metade da frequência de amostragem ($f_a/2$).

A forma como os RSVs cobrem o espectro pode ser ilustrada pelo comportamento dos RSVs associados a dois tipos de filtros:

Filtros sem ressonância em freq. intermediária – O filtro passa-baixas de 1^a ordem

$$H(z) = \frac{1}{1 - z^{-1}}$$

possui um conjunto de RSVs cujos espectros representam um banco de filtros de largura de faixa constante e com frequências centrais uniformemente distribuídas (Fig. 5.1 e Fig. 5.2);

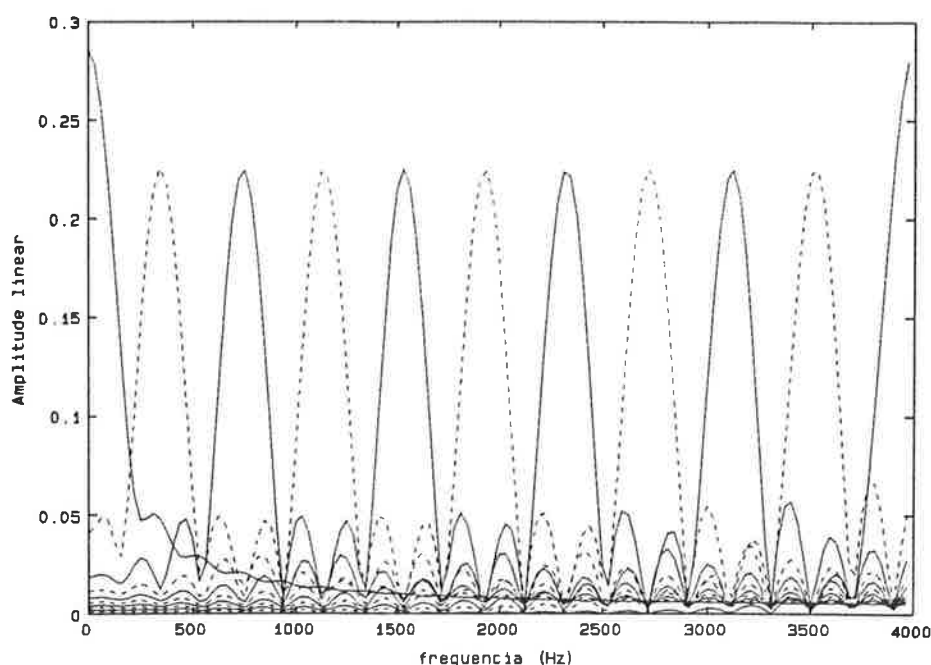


Figura 5.1: Espectros de RSVs de filtro passa-baixas índices 1 e $4n$, $n = 1, \dots, 10$.

Filtros ressonantes – O filtro ressonante em $f_a/4$

$$H(z) = \frac{1}{1 + z^{-2}}$$

possui RSVs cujos espectros apresentam duas faixas de passagem equidistantes da frequência de ressonância de $H(z)$, a menos dos dois primeiros RSVs, que têm ressonância em $f_a/4$ (Fig. 5.3).

5.4 Critérios de seleção dos vetores singulares

Após a decomposição em valores singulares da matriz H , temos o espaço completo das excitações definido pelos L RSVs que estão dispostos nas colunas da matriz V (Eq.(5.5)).

A etapa seguinte do processo de codificação é a definição de um subespaço de excitações V_M através de um subconjunto de M vetores-base (Fig. 5.4).

O vetor de seleção, que contém os índices dos RSVs selecionados a cada bloco de análise preditiva, é dado por

$$l = [l(1) \ \dots \ l(M)], \quad (5.30)$$

sendo

$$l(i) \in \{1, \dots, L\}, \quad i = 1, \dots, M.$$

O subespaço V_M (Fig. 5.4) de excitação definido pelo vetor de seleção l é dado por

$$V_M = [v_{l(1)} \ \dots \ v_{l(M)}], \quad (5.31)$$

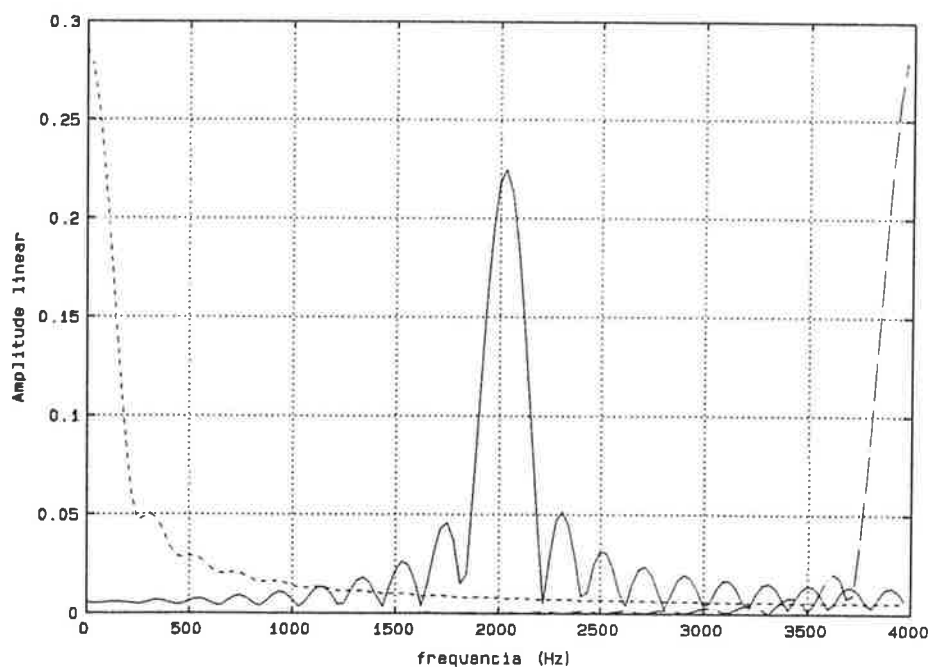


Figura 5.2: Espectros de RSVs de filtro passa-baixas índice 21 (linha contínua), 1 (linha tracejada) e 40 (linha traço-e-ponto.)

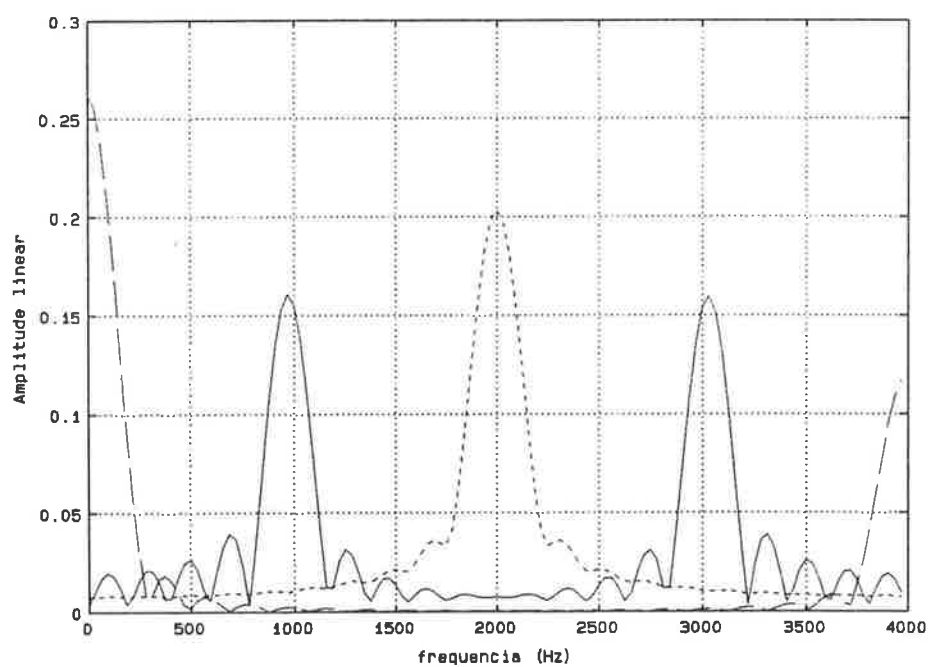


Figura 5.3: Espectros de RSVs de filtro ressonante índice 21 (linha contínua), 1 (linha tracejada) e 40 (linha traço-e-ponto)

onde v_i são os RSVs definidos pela Eq.(5.5).

Definimos a matriz esparsa S , $M \times L$, de redução de dimensão como

$$S = V_M^T V. \quad (5.32)$$

Aplicando a definição (5.31) do subespaço V_M e a propriedade (5.11) de ortonormalidade da base de RSVs à definição (5.32) da matriz S , podemos defini-la explicitamente como

$$S_{ij} = \begin{cases} 1 & j = l(i) \\ 0 & \text{caso contrário} \end{cases}, \quad (5.33)$$

isto é, S é uma matriz composta das linhas da matriz identidade I , $L \times L$, selecionadas através do vetor l .

O vetor l pode ser estabelecido por algum critério fixo ou determinado a cada bloco de análise por um critério adaptativo.

Experimentamos com dois tipos de critérios de seleção de RSVs, respectivamente baseados em:

i) valores singulares;

ii) propriedades espectrais dos RSVs – testamos três critérios:

Fixo – com seleção de índices espectralmente distribuídos

$$l = [1 \quad \dots \quad 5 \quad 11 \quad \dots \quad 13 \quad 21 \quad \dots \quad 23 \quad 31 \quad \dots \quad 33] \quad (5.34)$$

SP – Picos espectrais

SW Pares de raias espectrais.

O critério de seleção de RSVs baseado nos *valores singulares* tem origem na análise do papel desempenhado pelos valores singulares no processo de filtragem (seção 5.2).

Na composição SVD da norma do vetor reconstruído \hat{u} , usando a Eq.(5.28),

$$\|\hat{u}\| = \sqrt{\sum_{i=1}^L d_i^2 \xi_i^2} \quad (5.35)$$

com

$$\xi = V^T x, \quad (5.36)$$

nota-se que os valores singulares d_i maiores acentuam a influência das respectivas coordenadas ξ_i no valor da norma $\|\hat{u}\|$.

Com a convenção adotada de ordem decrescente dos valores singulares d_i (5.8), os maiores são aqueles de índices menores.

Considerando-se a grandeza dos valores singulares, a seleção mais razoável é a dos primeiros índices, isto é

$$l(i) = i, \quad i = 1, \dots, M. \quad (5.37)$$

Foram feitos testes com $M=14$ (seção 5.9).

As seleções espectrais baseiam-se nas propriedades espectrais dos vetores singulares (seção 2.1).

A expansão de faixa torna o filtro LPC mais próximo de um filtro passa-baixas do que de um filtro ressonante (seção 3.6). Assim, para a busca SVD baseada no filtro expandido (seção 5.6.1), podemos considerar que as frequências centrais dos RSVs amostram o espectro uniformemente.

A seleção *spectral fixa* procura representar os aspectos mais marcantes da forma de onda pelos 5 índices inferiores e escolhe 3 conjuntos de 3 índices centrados nos valores neutros de formantes de 1,2; 2,2 e 3,2 kHz, respectivamente.

A seleção de picos espectrais (SP) procura as frequências Ω de ocorrência dos 4 primeiros picos da função de transferência $H_{LPC}(j\omega)$ do filtro LPC.

O vetor F contém as frequências de ocorrência dos picos efetivamente encontrados, mapeadas para a faixa de índices de 1 a L , isto é

$$F = \lfloor \frac{L}{\pi} \Omega \rfloor, \quad (5.38)$$

ou, em caso de não encontrar nenhum pico, são atribuídos quatro índices fixos

$$F = [2 \quad 12 \quad 22 \quad 32], \quad (5.39)$$

onde $\lfloor x \rfloor$ indica o maior número inteiro contido em x .

O vetor de seleção l é definido pelos índices contidos no vetor de picos F , pelos seus antecessores e pelos seus sucessores, ou seja,

$$l(i) = \begin{cases} F(i \div 3 + 1) & \text{se } i \bmod 3 = 2 \\ F(i \div 3 + 1) - 1 & \text{se } i \bmod 3 = 1 \\ F(i \div 3 + 1) + 1 & \text{se } i \bmod 3 = 0 \\ i = 1, 2, \dots, 3n_F \end{cases}, \quad (5.40)$$

onde

- $n_F = \dim(F)$;
- $m \div n$ indica o quociente inteiro da divisão de m por n , ambos inteiros;
- $m \bmod n$ indica o resto da divisão de m por n , ambos inteiros.

Notamos que o vetor de picos F e, em consequência, o vetor de seleção l terão dimensão variável a cada sub-bloco de excitação.

A seleção de pares de raias espectrais (SW) utiliza os quatro seguintes coeficientes LSP pares (Apêndice C)

$$\Omega = [\omega_2 \quad \omega_4 \quad \omega_8 \quad \omega_{10}], \quad (5.41)$$

que são as fases de raízes não reais do polinômio anti-simétrico

$$P(z) = (A(z) - z^{-(p+1)}A(1/z))$$

associado ao filtro LPC inverso

$$A(z) = \frac{1}{H_{LPC}(z)}.$$

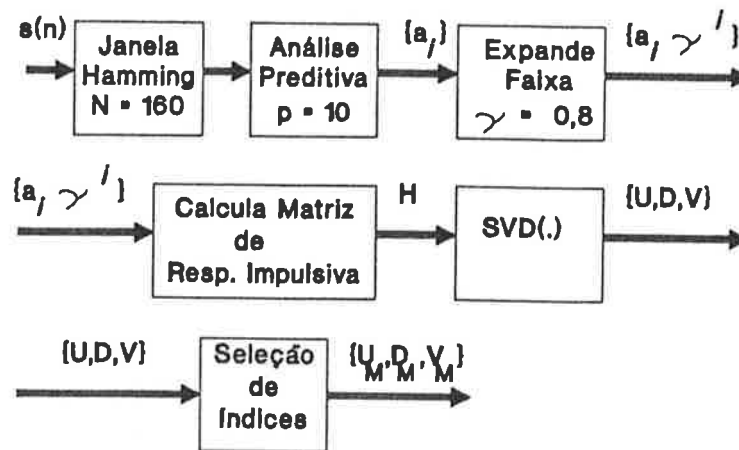


Figura 5.4: Análise preditiva no CELP-SVD

O vetor Φ contém as frequências Ω mapeadas para a faixa de índices de 1 a L , isto é

$$\Phi = \left\lfloor \frac{L}{\pi} \Omega \right\rfloor. \quad (5.42)$$

O vetor de seleção l é definido pelos índices contidos no vetor de picos Φ , pelos seus antecessores e pelos seus sucessores, ou seja,

$$l(i) = \begin{cases} \Phi(i \div 3 + 1) & \text{se } i \bmod 3 = 2 \\ \Phi(i \div 3 + 1) - 1 & \text{se } i \bmod 3 = 1 \\ \Phi(i \div 3 + 1) + 1 & \text{se } i \bmod 3 = 0 \\ & i = 1, 2, \dots, 12 \end{cases} \quad (5.43)$$

O vetor l tem dimensão fixa e igual a 12 na seleção SW. Como vantagens adicionais do critério SW sobre o SP temos

- a determinação dos parâmetros LSP é mais confiável do que a extração de picos espectrais e
- os parâmetros LSP são convenientes para a codificação do próprio filtro LPC pela sua baixa sensibilidade à quantização e pela sua facilidade de interpolação [15].

5.5 Codificador proposto

O codificador proposto (CELP-SVD) executa dois processos principais: a *análise preditiva* e a *busca da excitação*.

A *análise preditiva* define os coeficientes do filtro LPC para um bloco de análise de $N = 160$ amostras de voz (Fig. 5.4). Por sua vez, o filtro LPC, mediante a sua matriz de resposta impulsiva H no bloco de análise, delimita o conjunto de vetores-base que podem ser usados para a busca do vetor de excitação de comprimento $L = 40$ nos quatro sub-blocos de busca que compõem o bloco de análise.

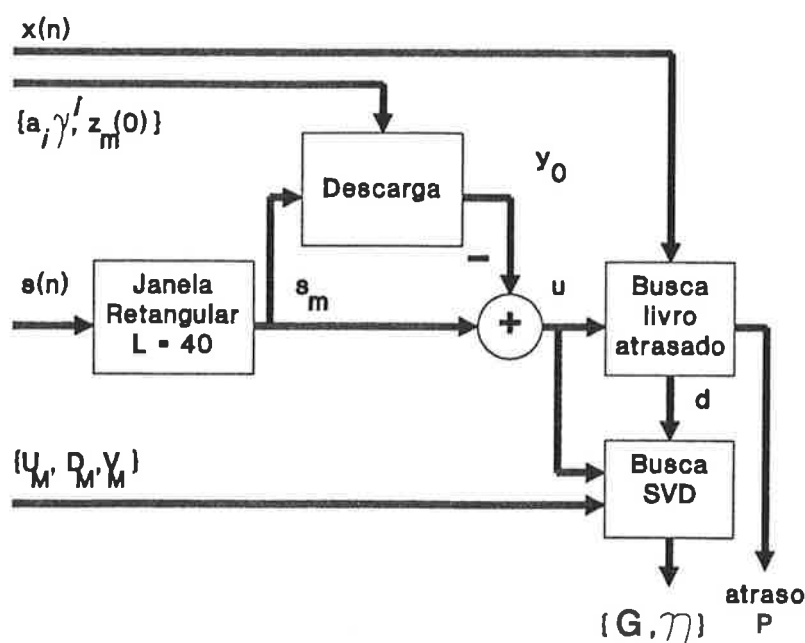


Figura 5.5: Busca da excitação no CELP-SVD

Para cada bloco de análise preditiva, selecionam-se M vetores-base do conjunto de L vetores singulares à direita (RSVs) de H mediante um dado critério de seleção (seção 5.4). Esses RSVs geram o subespaço V_M usado na busca SVD (seção 5.6).

Busca-se a excitação em dois livros de códigos: o SVD e o adaptativo de vetores atrasados.

Testamos mais intensivamente o codificador com representação SVD baseada no filtro expandido (seção 5.6.1).

A busca atrasada é realizada de forma convencional (seção 3.8.1), sendo retida na representação final apenas a componente ortogonal ao subespaço V_M do vetor atrasado selecionado (seção 5.7).

5.6 Busca SVD da excitação

Após a seleção de índices, temos o subespaço V_M das excitações de dimensão reduzida (Eq.(5.3) e Fig. 5.4).

A busca SVD usa o vetor-alvo integral u (Fig. 5.5) e não o vetor-alvo r , residual da busca atrasada, como no CELP convencional com dois livros de códigos (Fig. 3.13).

A busca SVD da excitação não envolve a filtragem de todos os vetores-código como ilustrado na Fig. 3.11 para a busca CELP convencional.

A introdução da decomposição matricial em valores singulares aumenta consideravelmente a complexidade computacional. Em compensação, a busca SVD da excitação não envolve processos iterativos, mas apenas dois processos diretos, que são

Filtragem inversa do vetor-alvo u resultando na excitação x_c ideal de dimensão

completa

$$x_c = H_e^{-1}u, \quad (5.44)$$

onde H_e é a matriz de resposta impulsiva do filtro LPC expandido;

Projeção do vetor de excitação completa no subespaço de entrada V_M determinado pelos índices selecionados (Fig. 5.4), resultando na excitação SVD de dimensão reduzida x_M

$$\xi_M = V_M^T x_c \quad (5.45)$$

$$x_M = V_M \xi_M. \quad (5.46)$$

Se a matriz de resposta impulsiva do filtro expandido for escolhida para a representação SVD em vez daquela do filtro LPC, a filtragem inversa (Eq.(5.44)) e a projeção (Eq.(5.45)) podem ser implementadas conjuntamente. Isto é, a redução de dimensão já pode ser efetuada a partir do vetor-alvo (seção 5.6.1).

Devido a esta facilidade, adotamos o filtro expandido para a representação SVD na maioria das simulações. Entretanto, a representação SVD baseada no filtro LPC conseguiu melhores resultados em vários testes (seção 5.8).

5.6.1 Representação SVD baseada no filtro expandido

Vamos tomar a matriz de resposta impulsiva H_e do filtro LPC expandido como matriz de representação SVD, isto é,

$$H = H_e. \quad (5.47)$$

Substituindo H_e , com a condição (5.47), pela sua decomposição SVD (5.4) na Eq.(5.44), temos

$$x_c = (UDV^T)^{-1}u. \quad (5.48)$$

Como U e V são matrizes ortogonais (Eqs.(5.10) e (5.11)), a inversão matricial em (5.48) resolve-se facilmente, resultando

$$x_c = VD^{-1}U^T u. \quad (5.49)$$

Substituindo (5.49) em (5.45), temos

$$\xi_M = V_M^T VD^{-1}U^T u. \quad (5.50)$$

Usando a especificação (5.32) da matriz S de redução de dimensão, podemos reescrever a Eq.(5.50) como

$$\xi_M = SD^{-1}U^T u. \quad (5.51)$$

Dada uma matriz quadrada A , $L \times L$, aplicando-lhe a matriz de redução de dimensão S , resulta a matriz reduzida

$$A^r = SA, \quad (5.52)$$

que difere de A pela eliminação de todas as $L - M$ linhas cujos índices não pertencem ao vetor de seleção l .

A matriz reduzida de A através da seleção l é dada explicitamente por

$$A^r = [A_{l(1)}^T \quad A_{l(2)}^T \quad \cdots \quad A_{l(M)}^T]^T, \quad (5.53)$$

onde A_i , $i = 1, \dots, L$, são as linhas de A .

Analogamente à explicitação de (5.52) por (5.53), lembrando (5.7) e (5.6), obtemos a explicitação de (5.51) como

$$\xi_M(i) = \frac{u_{l(i)}^T u}{d_{l(i)}}, \quad i = 1, \dots, M. \quad (5.54)$$

Portanto, para obter as coordenadas SVD ξ_M da excitação reduzida, basta obter o vetor-alvo reduzido e dividir cada uma de suas coordenadas $u_i^T u$ pelo valor singular correspondente.

5.6.2 Representação SVD baseada no filtro LPC

Optando-se pela matriz de resposta impulsiva H do filtro LPC para matriz de representação SVD, temos dois processos independentes constituindo a busca SVD: a filtragem inversa e a projeção (seção 5.6).

Como neste caso não podemos nos valer imediatamente da decomposição SVD, vamos deduzir uma expressão para a filtragem inversa (5.44) a partir da representação recorrente (“recursive”) da filtragem direta.

A filtragem direta usada no processo de busca (Fig. 3.11) pode ser realizada como

$$\hat{u}_i = H_\epsilon x_i, \quad (5.55)$$

onde

$$H_\epsilon = \begin{bmatrix} h(0) & 0 & \cdots & 0 & 0 \\ h(1) & h(0) & 0 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ h(L-2) & \ddots & h(1) & h(0) & 0 \\ h(L-1) & h(L-2) & \cdots & h(1) & h(0) \end{bmatrix} \quad (5.56)$$

é a matriz de resposta impulsiva (seção 3.7) de duração L do filtro expandido.

A função de transferência do filtro expandido é

$$\begin{aligned} H(z) &= H_{LPC}(z/\gamma) \\ &= \frac{1}{1 + \sum_{k=1}^p a_k \gamma^k z^{-k}} \\ &= \frac{1}{1 + \sum_{k=1}^p \alpha_k z^{-k}}, \end{aligned} \quad (5.57)$$

onde

5.7 Representação da periodicidade

Sabemos de antemão que os blocos de voz sonoros demandam da representação da excitação precisão crescente com o seu grau de periodicidade [7]. Assim, resolvemos representar com maior precisão a periodicidade.

Por outro lado, a representação SVD quantizada de segmentos sonoros requer aumento da precisão com a periodicidade para manter a distorção em níveis inaudíveis [7].

Para representar a periodicidade, vamos buscar um segmento de excitação atrasada usando o mesmo livro de códigos adaptativo descrito na seção 3.8.1 e depois incorporá-lo na nossa representação SVD reduzida.

Nesta seção, assumimos que a busca SVD está baseada no filtro expandido (seção 5.6.1).

Seja P o índice selecionado na busca no livro de excitações atrasadas (Fig. 3.13), correspondente ao vetor-código

$$d = d_P, \quad (5.62)$$

que gera o vetor de voz ponderada

$$q = Hd, \quad (5.63)$$

onde H é a matriz de resposta impulsiva do filtro expandido.

Conforme vimos na seção 5.6.1, a excitação reduzida x_M é determinada pela filtragem inversa

$$x_M = H^{-1}u_M \quad (5.64)$$

da projeção u_M do vetor-alvo no subespaço U_M . Portanto, representando em precisão infinita as coordenadas ξ_M da excitação reduzida (Eq.(5.46)), consegue-se reconstruir perfeitamente o vetor-alvo reduzido u_M .

A componente do vetor-alvo u na direção do vetor atrasado ponderado q é

$$u_P = \frac{q^T u}{q^T q} q. \quad (5.65)$$

Propomos as duas especificações seguintes para a excitação total x .

Especificação 1 *A excitação total x deve reconstruir perfeitamente a componente u_P de periodicidade do vetor-alvo quando representada em precisão infinita.*

Especificação 2 *A excitação total x é composta aditivamente pelas componentes ortogonais x_M e x_P , isto é,*

$$x = x_M + x_P,$$

onde x_M é a excitação SVD de dimensão reduzida e x_P é a excitação complementar de periodicidade.

Para atender à Especificação 1, temos que expandir o subespaço de projeção U_M pela inclusão do vetor q , isto é,

$$U_\epsilon = [U_M \quad q]. \quad (5.66)$$

A componente x_M da excitação está contida no subespaço U_M , que é dado pela representação ortonormal

$$U_M = [u_{l(1)} \quad u_{l(2)} \quad \cdots \quad u_{l(M)}], \quad (5.67)$$

extraída de (5.54).

Além disso, como x_M é definida perfeitamente por u_M (Eq.(5.64)), o seu complemento ortogonal x_P têm que ser determinado por um vetor normal ao subespaço U_M .

Num processo de ortogonalização Gram-Schmidt de U_e , basta tomar a componente q_\perp do vetor atrasado ponderado ortogonal ao subespaço U_M , dada por

$$q_\perp = q - U_M(U_M^T q). \quad (5.68)$$

Conseqüentemente, U_e pode ser expresso por

$$U_e = [U_M \quad q_\perp]. \quad (5.69)$$

Por outro lado, nós temos a representação do espaço U completo e podemos escrever (5.68) alternativamente como

$$q_\perp = \overline{U_M}(\overline{U_M}^T q), \quad (5.70)$$

onde $\overline{U_M}$, o subespaço complementar de U_M , é definido pela base de $L - M$ vetores

$$u_i, \quad \begin{cases} i \in \{1, \dots, L\} \\ i \notin l \end{cases},$$

sendo l o vetor de seleção (seção 5.6.1).

Analogamente à Eq.(5.64), considerando (5.63), temos que q_\perp é determinado por

$$d_\perp = H^{-1} q_\perp, \quad (5.71)$$

que é a componente ortogonal ao subespaço V_M do vetor atrasado d .

Dado o vetor atrasado d , podemos determinar sua componente d_\perp por projeção de d seguida de reconstrução, ou seja,

$$\delta = \overline{V_M}^T d \quad (5.72)$$

$$d_\perp = \overline{V_M} \delta, \quad (5.73)$$

onde $\overline{V_M}$, o subespaço complementar de V_M , é definido pela base de $L - M$ vetores

$$v_i, \quad \begin{cases} i \in \{1, \dots, L\} \\ i \notin l \end{cases},$$

Assim, a nossa excitação total x está contida no subespaço expandido

$$V_e = [V_M \quad d_\perp]. \quad (5.74)$$

Como V_M está definido por vetores unitários, vamos normalizar d_\perp , obtendo o vetor-base atrasado

$$v_P = \frac{d_\perp}{\|d_\perp\|}, \quad (5.75)$$

onde $\|d_\perp\|$ denota a norma quadrática de d_\perp .

Em base ortonormal, o subespaço expandido de excitação é dado por

$$V_e = [V_M \ v_P]. \quad (5.76)$$

Agora podemos expressar a excitação total x através de

$$\zeta = V_e^T x \quad (5.77)$$

$$x = V_e \zeta. \quad (5.78)$$

As M primeiras coordenadas ζ da excitação, lembrando (5.45), são dadas por

$$\zeta(i) = \xi_M(i), \quad i = 1, \dots, M. \quad (5.79)$$

Vamos determinar a última coordenada $\zeta(M+1)$ da excitação total pela projeção do vetor-alvo u no vetor-base atrasado ponderado q_P , dado pela filtragem do vetor-base atrasado. Isto é,

$$\zeta(M+1) = \frac{q_P^T u}{q_P^T q_P} q_P, \quad (5.80)$$

onde

$$q_P = H v_P. \quad (5.81)$$

De passagem, considerando-se (5.71) e (5.75), temos a relação

$$q_P = \frac{q_\perp}{\|d_\perp\|}. \quad (5.82)$$

O vetor-base atrasado ponderado q_P (5.81), usando-se a representação SVD (5.72), pode ser calculado mais facilmente como

$$\rho(i) = \delta(i) d_i, \begin{cases} i \in \{1, \dots, L\} \\ i \notin l \end{cases} \quad (5.83)$$

$$q_P = \overline{U_M} \rho, \quad (5.84)$$

onde d_i são os valores singulares de H (Eq.(5.7)).

Finalmente, vamos normalizar o conjunto ζ (Eq.(5.77)) de coordenadas da excitação total x , resultando as coordenadas normallizadas

$$\eta = \zeta/G, \quad (5.85)$$

onde

$$G = \|\zeta\|. \quad (5.86)$$

Assim, a excitação total é dada por

$$x = G V_e \eta \quad (5.87)$$

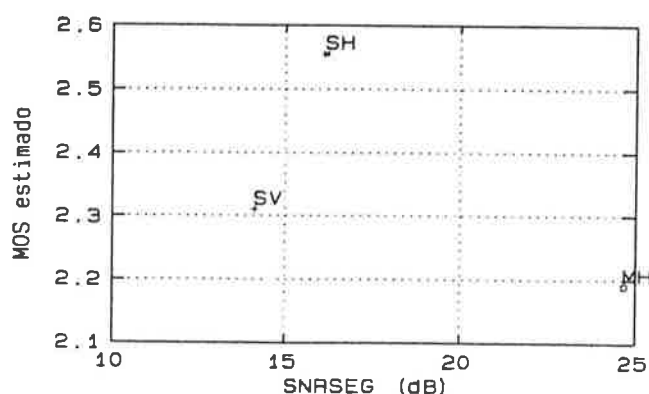


Figura 5.6: Testes de representação SVD com o sinal de voz v .

e a excitação complementar de periodicidade x_P (Especificação 2) é realizada por

$$x_P = G\eta(M+1)v_P. \quad (5.88)$$

Resumindo, o algoritmo para representação da periodicidade e montagem da excitação total x inclui os seguintes procedimentos:

- i*) a busca do vetor-código atrasado (5.62);
- ii*) a projeção do vetor atrasado (5.72);
- iii*) a filtragem simplificada (5.83);
- iv*) a construção do vetor filtrado (5.84);
- v*) a determinação da coordenada atrasada (5.80) e
- vi*) o cálculo do ganho (5.86) e a normalização das coordenadas da excitação (5.85).

5.8 Testes de representação SVD

As duas representações SVD apresentadas na seção 5.6 foram testadas com os três sinais de prova (capítulo 4).

Foram simulados os três codificadores descritos na Tabela 5.1 com medidas de qualidade indicadas nas Figs. 5.6, 5.7 e 5.8 para as vozes v , c e d , respectivamente.

Notamos que o codificador SH reconstrói os sinais com a melhor combinação de índices de qualidade.

O codificador MH sempre apresenta a melhor SNRSEG, mas o \widehat{MOS} baixo indica algum problema com o vetor de seleção, como verificaremos nos testes da próxima seção.

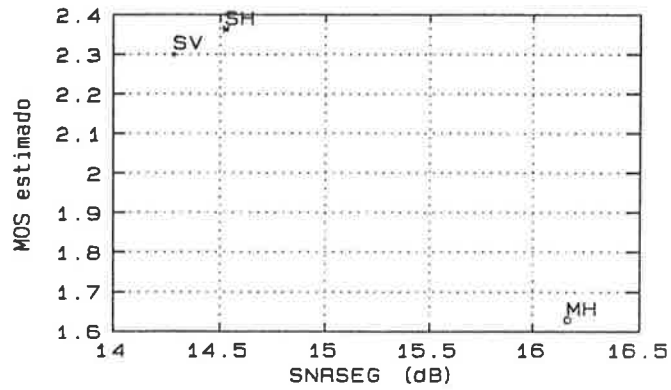


Figura 5.7: Testes de representação SVD com o sinal de voz *c*.

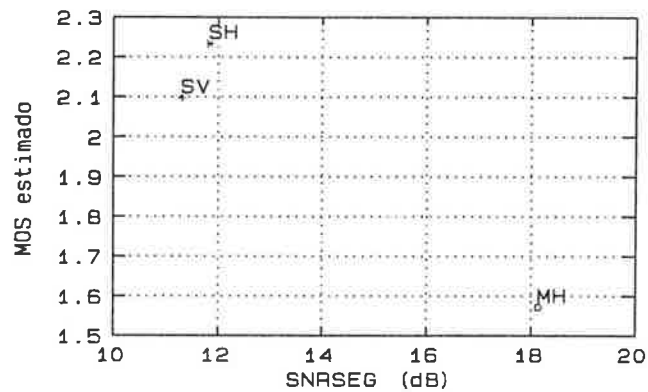


Figura 5.8: Testes de representação SVD com o sinal de voz *d*.

Tabela 5.1: Codificadores para teste das representações SVD

Codificador	Vetor de seleção	Rep. SVD
MH	1 a 14	LPC
SH	1 a 5, 11 a 13, 21 a 23, 31 a 33	LPC
SV	1 a 5, 11 a 13, 21 a 23, 31 a 33	Exp.

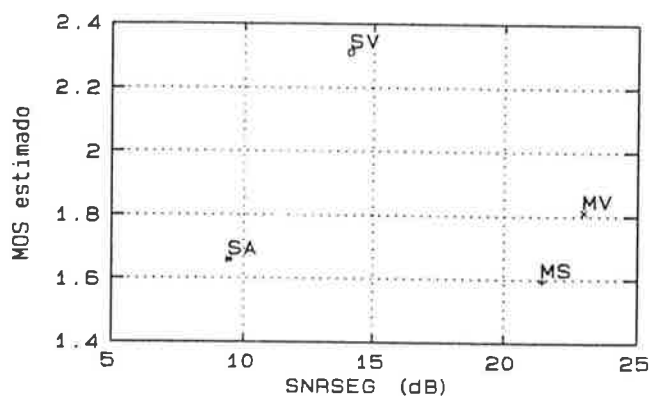


Figura 5.9: Testes de configuração do codificador CELP-SVD com o sinal de voz v .

Tabela 5.2: Codificadores CELP-SVD identificados pela seleção de vetores singulares e pelo uso do vetor atrasado.

Codificador	Vetor de seleção	Vetor atrasado
MS	1 a 14	Não
SA	1 a 5, 11 a 13, 21 a 23, 31 a 33	Não
SV	1 a 5, 11 a 13, 21 a 23, 31 a 33	Sim
MV	1 a 14	Sim

5.9 Testes de configuração

Estes testes foram executados na fase inicial de avaliação dos codificadores CELP-SVD. Os resultados obtidos originaram a distinção maior que fizemos entre critérios baseados em valores singulares e baseados em propriedades espectrais na seção 5.4.

Na reconstrução parcial da forma de onda de um sub-bloco do sinal, os vetores singulares à esquerda (RSVs) mais significativos são os que têm maiores valores singulares.

Embora a qualidade subjetiva não possa ser avaliada apenas pela reconstrução da forma de onda, consideramos importante testar a importância dos valores singulares como critério de seleção dos RSVs.

Assim, fizemos uma simulação da reconstrução do sinal de voz v (capítulo 4) por um codificador CELP-SVD com excitações restritas ao subespaço definido pelos 14 RSVs de valores singulares mais significativos, isto é, o vetor de seleção é

$$l = [1 \ 2 \ \dots \ 14] .$$

Esse codificador, o MS (Tabela 5.2), aplicado à voz v , apresentou uma SNRSEG = 21,46 dB, digna de codificadores de média complexidade (Fig. 4.2) que operam à taxa de 32 kb/s, como o ADPCM [23] [11].

Conseqüentemente, esperávamos um índice MOS ≈ 4 , mas encontramos $\widehat{MOS} = 1,60$ (Fig. 5.9).

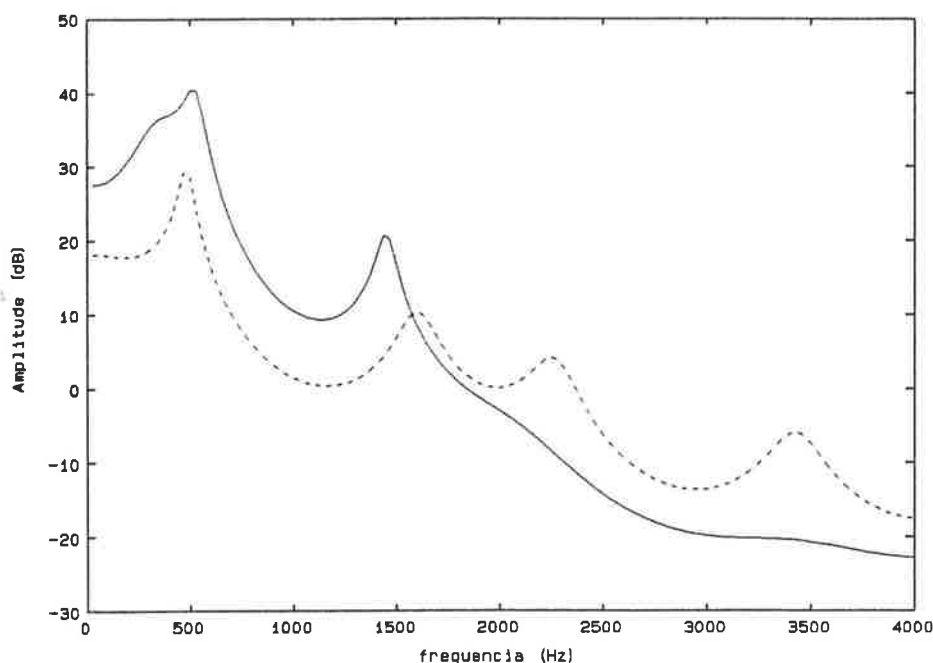


Figura 5.10: Envoltórias dos espectros logarítmicos de um segmento do sinal de voz v (linha tracejada) e da sua reconstrução pelo codificador MS (linha contínua).

Fomos em busca da razão para esses resultados contraditórios considerando, em primeiro lugar, que o nosso estimador \widehat{MOS} está baseado na diferença entre as envoltórias dos espectros logarítmicos de curto prazo dos sinais original e reconstruído.

Comparando as envoltórias dos espectros logarítmicos de um segmento do sinal v e da sua reconstrução pelo codificador MS (Fig. 5.10), notamos que a envoltória desta não acompanha a do sinal original nas faixas da 3ª e da 4ª formantes.

Neste teste usamos a representação SVD baseada no filtro expandido (seção 5.6.1). Como o filtro expandido tem ressonâncias pouco pronunciadas (seção 3.6), o espectro do seu conjunto de RSVs comporta-se de forma semelhante àquele do filtro passa-baixas (seção 2.1).

Assim, os 14 primeiros RSVs cobrem uma faixa de 0 a 1400 Hz, não é conseguindo cobrir a 3ª e a 4ª formantes.

O codificador SA (Tabela 5.2) emprega uma seleção espectralmente mais bem distribuída de vetores singulares, mantendo uma concentração maior nos baixos índices.

O codificador SA apresentou uma pequena melhora em \widehat{MOS} (Fig. 5.9), porém, com uma queda acentuada da SNRSEG em relação ao desempenho de MS.

Apesar de ser pequena a melhora subjetiva prevista de SA sobre MS, decidimos pesquisar com mais cuidado a sua seleção de vetores singulares.

Como a periodicidade é muito importante para caracterizar a voz sonora, implementamos o codificador SV, que acrescenta ao espaço das excitações de SA mais uma dimensão, que é determinada a partir da busca de um vetor de excitação atrasado (seção 3.8.1).

O codificador SV alcançou uma SNRSEG intermediária entre aquelas dos codificadores MS e SA. Mas, a incorporação da dimensão do vetor atrasado teve um

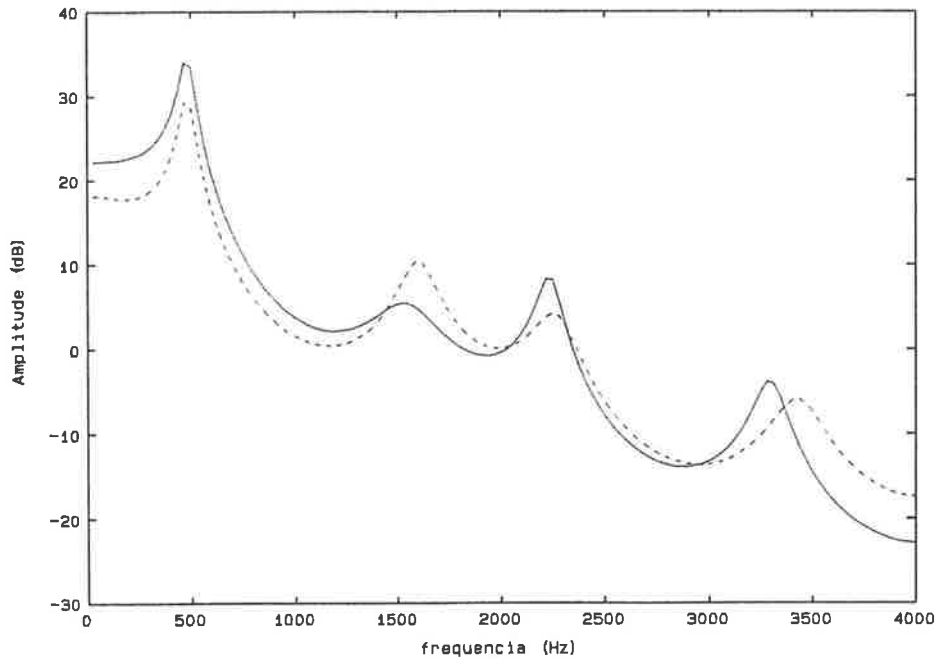


Figura 5.11: Envoltórias dos espectros logarítmicos de um segmento do sinal de voz v (linha tracejada) e da sua reconstrução pelo codificador SV (linha contínua).

efeito muito mais significativo sobre o \widehat{MOS} , como evidenciado pela comparação dos desempenhos dos codificadores SA e SV (Fig. 5.9).

Como a diferença em \widehat{MOS} entre os codificadores MS e SA foi muito pequena e MS teve um desempenho muito melhor em SNRSEG, perguntamo-nos se a mesma transformação

$$SA \xrightarrow{\dim(\text{vet.atras.})} SV$$

que levou SA a SV, quando aplicada a MS

$$MS \xrightarrow{\dim(\text{vet.atras.})} MV,$$

não produziria um codificador MV com melhor qualidade subjetiva estimada do que SV.

Contrariando essa expectativa, o codificador MV apresentou um \widehat{MOS} muito inferior àquele produzido por SV (Fig. 5.9). Isto revela a superioridade da seleção espectralmente distribuída de vetores singulares.

A aproximação da envoltória do espectro logarítmico de um segmento do sinal v está ilustrada na Fig. 5.11. O segmento é o mesmo já apresentado na Fig. 5.10.

Assim, concluímos que o codificador CELP-SVD com uma seleção espectralmente balanceada de vetores singulares reconstrói a voz sonora com mais fidelidade do que com uma seleção “tradicional”, baseada na ordem decrescente de valores singulares.

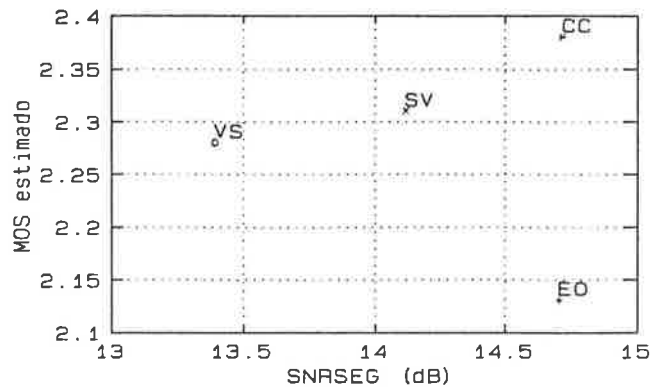


Figura 5.12: Testes comparativos do codificador CELP-SVD com os codificadores CELP descritos na Tabela 5.3, usando o sinal de voz *v*.

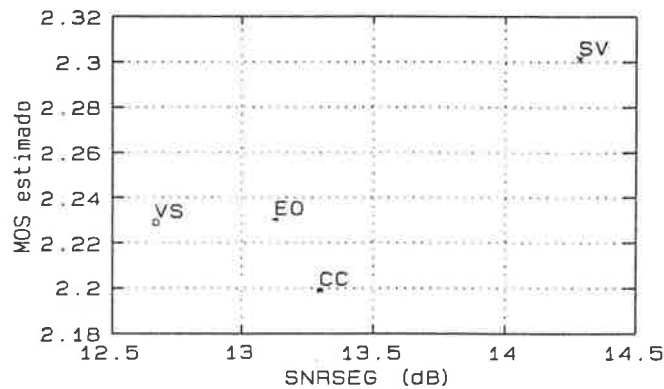


Figura 5.13: Testes comparativos do codificador CELP-SVD com os codificadores CELP descritos na Tabela 5.3, usando o sinal de voz *c*.

5.10 Testes comparativos

Estes testes objetivam avaliar o desempenho da melhor configuração do codificador CELP-SVD do teste anterior (seção 5.9) em relação a alguns dos codificadores CELP mais conhecidos. Esses codificadores estão designados na Tabela 5.3 e já foram descritos na seção 4.6.

Nos testes com a voz sonora *v*, o codificador SV superou o VSELP. Entretanto, o melhor desempenho foi conseguido pelo codificador esparsos com ceifagem central CC.

Nos testes com a voz CCITT masculina *c*, o codificador SV apresentou o melhor desempenho entre todos os codificadores testados.

5.11 Testes de critérios de seleção

O codificador CELP-SVD com um critério arbitrário de seleção espectral dos vetores singulares foi testado com a voz sonora de teste *v*, apresentando um bom desempenho frente a outros codificadores CELP (seção 5.10). Esses resultados motivaram a pesquisa de critérios mais elaborados de seleção espectral dos vetores

Tabela 5.3: Codificadores CELP usados nos testes comparativos.

Codificador	Descrição
EO	Estocástico ortogonal
CC	Esparso (ceifagem central) ortogonal
VS	Excitação por soma vetorial (VSELP)
SV	CELP-SVD com seleção distribuída de índices e com vet. atras.

Tabela 5.4: Codificadores para testes de seleção de RSVs

Codificador	Vetor de seleção	Rep. SVD
SV	1 a 5, 11 a 13, 21 a 23, 31 a 33	Exp.
SP	4 picos, antecessores e sucessores	Exp.
SW	4 LSPs pares, antecessores e sucessores	Exp.
WH	4 LSPs pares, antecessores e sucessores	LPC

singulares.

Propusemos dois critérios adaptativos de seleção de vetores singulares com espectros concentrados nas regiões das ressonâncias do filtro LPC (seção 5.4):

SP Picos espectrais

SW Pares de raias espectrais

O critério SW consiste na seleção de um subconjunto de RSVs, determinado a partir dos LSPs pares (Apêndice C)

$$\{ \omega_2, \omega_4, \dots, \omega_{10} \}$$

Foram simulados os três codificadores descritos na Tabela 5.4 com medidas de qualidade indicadas nas Figs. 5.14, 5.15 e 5.16 para as vozes **v**, **c** e **d**, respectivamente.

Notamos que o codificador SW apresentou melhor desempenho subjetivo estimado do que SV para a voz **c**. Apesar de SW ter tido as duas medidas de desempenho piores para as vozes **v** e **d**, SW ainda pode ser conveniente se os coeficientes LSP forem usados para codificar o filtro LPC.

O codificador WH, com a representação SVD baseada no filtro LPC, sempre obteve o melhor desempenho subjetivo estimado, com exceção da voz **v**.

Para a voz **v**, o melhor desempenho subjetivo estimado foi conseguido pelo codificador SP. Porém, o critério de seleção baseado em picos espectrais forneceu os piores resultados com as outras duas vozes.

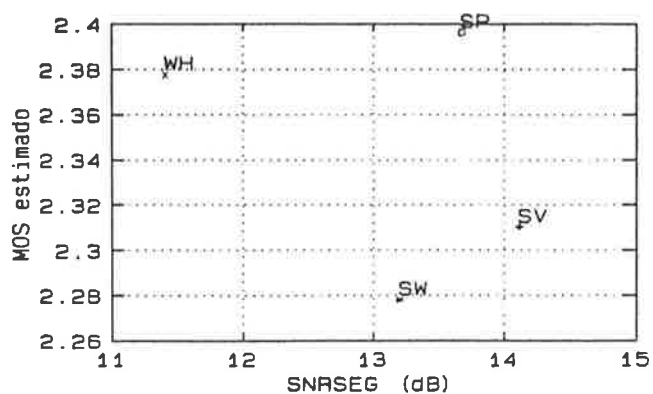


Figura 5.14: Testes de seleção de vetores singulares CELP descritos na Tabela 5.4, usando o sinal de voz *v*.

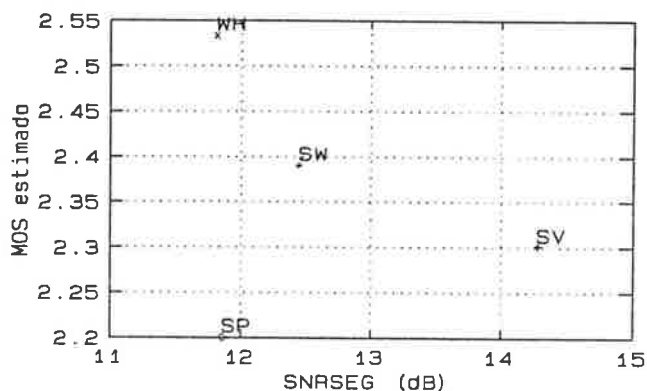


Figura 5.15: Testes de seleção de vetores singulares CELP descritos na Tabela 5.4, usando o sinal de voz *c*.

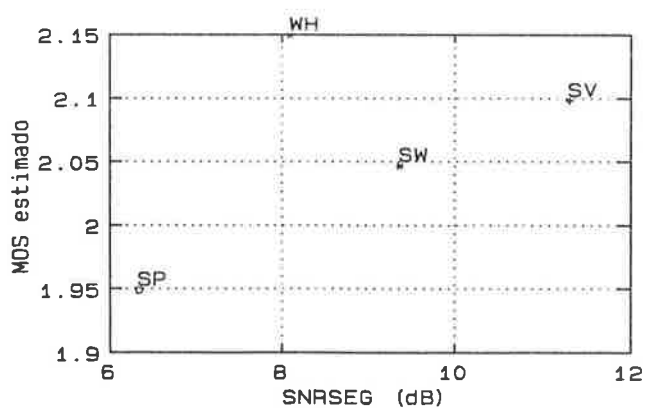


Figura 5.16: Testes de seleção de vetores singulares CELP descritos na Tabela 5.4, usando o sinal de voz *d*.

5.12 Testes de quantização

Os testes de quantização procuraram fornecer uma noção do desempenho do codificador CELP-SVD com o critério de seleção SW (seção 5.4) para taxas de codificação próximas de 8 kHz.

Os parâmetros do codificador foram classificados em quatro grupos:

ganho (G) é a norma quadrática do vetor de coordenadas selecionado (seção 5.7);

coordenadas normalizadas (Cs) η da excitação na base composta pelos RSVs selecionados e pelo vetor unitário ortogonal atrasado v_P (seção 5.7);

índice do vetor atrasado (P) é determinado pela busca no livro adaptativo de excitações atrasadas (seção 5.7) e

pares de raias espectrais (LSPs) que definem a função de transferência do filtro LPC (Apêndice C).

Exceto o índice do vetor atrasado (P), todos os parâmetros são determinados com precisão ponto flutuante dupla e têm que ser quantizados.

As cinco primeiras coordenadas singulares, η_i , $i = 1, \dots, 5$, e a coordenada de periodicidade η_{13} são representadas com precisão maior (Tabela 5.5) devido à sua importância na reconstrução da forma de onda e pela sua importância perceptiva, respectivamente.

Procedeu-se ao projeto de dois quantizadores escalares de cada parâmetro com uma versão unidimensional do algoritmo LBG [29], usando-se como sinal de treinamento as vozes **c** e **d** separadamente.

Os dois quantizadores de cada coordenada singular normalizada η_i , $i = 1, \dots, 12$, tiveram níveis coincidentes. Ou seja, para duas vozes de origens bem diferentes, os quantizadores das coordenadas singulares normalizadas convergiram para os mesmos níveis.

O algoritmo LBG não conseguiu concluir o projeto do quantizador da coordenada de periodicidade η_{13} com a voz **d** porque na última fragmentação surgiu uma célula vazia.

Assim, usamos para todos os testes o quantizador das coordenadas projetado para a voz **c**.

Para o fator de ganho G, testamos também um quantizador logarítmico uniforme com passo de 2 dB. Os níveis mínimo e máximo foram fixados em -45 dB e 15 dB relativos ao desvio padrão do sinal de voz, além de um código que foi reservado para o nível 0. A excursão de ganho foi tomada da distribuição de potência do sinal de voz especificada pela recomendação CCITT [12]. A Figura 5.17 apresenta as curvas dos três quantizadores de ganho usados nos testes.

Para a voz **c**, notamos que há uma queda significativa (≈ 3 dB) entre o desempenho medido pela SNRSEG do codificador SW e aquele de qualquer uma das suas versões quantizadas. Além disso, a SNRSEG não se altera quando trocamos de codificador quantizado.

Tabela 5.5: Distribuição de bits entre os parâmetros do codificador CELP-SVD

Grupo de parâmetros	Precisão (bits)	Taxa de codificação (b/s)
Ganho (G)	5	1000
Coordenadas (Cs)	3,3,2,2,2,1,1,1,1,1,1,3	4400
Índice de atraso (P)	7	1400
Filtro (LSPs)	4,4,4,4,4,4,4,4,3	1950
Taxa de codificação total (b/s)		8750

Tabela 5.6: Codificadores CELP-SVD quantizados

Codificador	Quantização
SW	sem quantização
QQ	signal de prova
XQ	signal de referência
QE	logarítmica
GQ	LSPs: signal de prova, G: signal de referência
CQ	LSPs: signal de referência, G: signal de prova

Constata-se que o uso de um quantizador com espaçamento logarítmico uniforme para o ganho, QE, não altera significativamente nenhuma medida de desempenho em relação ao QQ.

Para a voz *d*, notamos uma certa ortogonalidade no comportamento do ganho e dos coeficientes LSP em relação à quantização:

- quando o ganho é treinado externamente (GQ), há queda apenas da SNRSEG em relação ao treinamento com o próprio signal de prova e,
- quando os LSPs são treinados externamente (CQ), degrada-se apenas o \widehat{MOS} em relação ao treinamento com o próprio signal de prova.

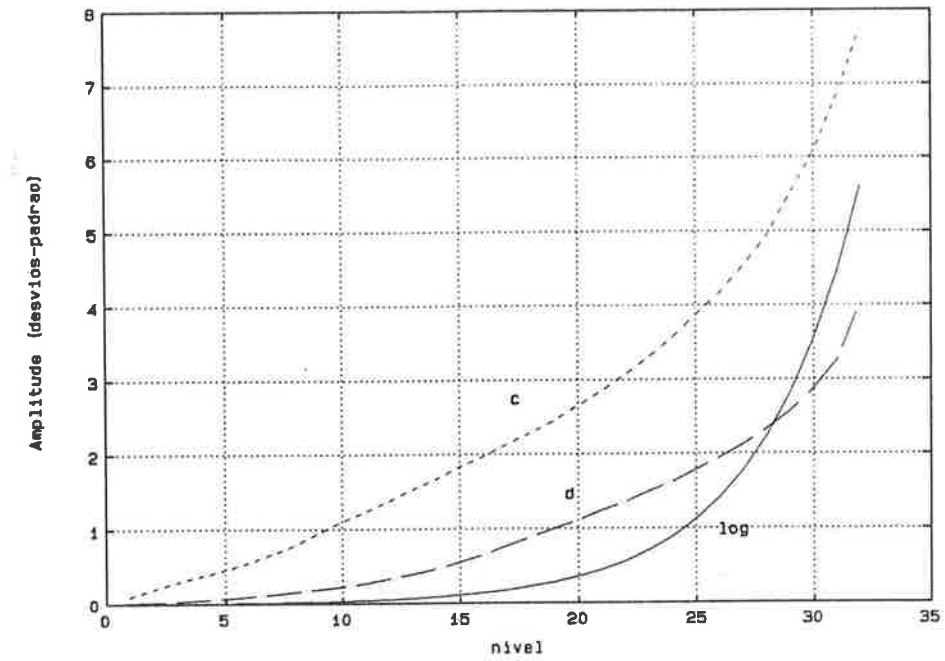


Figura 5.17: Níveis dos quantizadores de ganho logarítmico (curva contínua), treinado com o sinal *c* (curva tracejada) e treinado com o sinal *d* (curva traço-e-ponto).

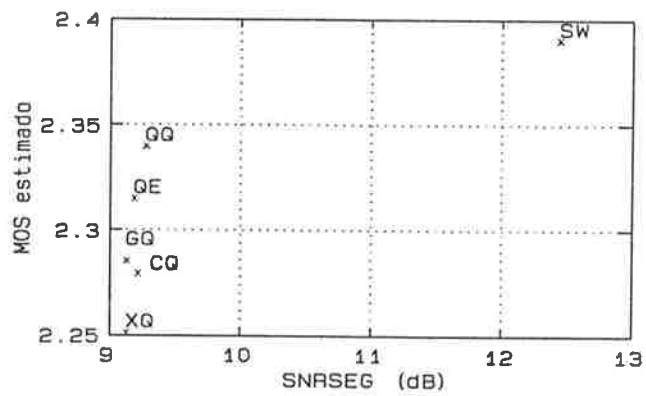


Figura 5.18: Testes de quantização com o sinal de prova *c* e sinal de referência *d*.

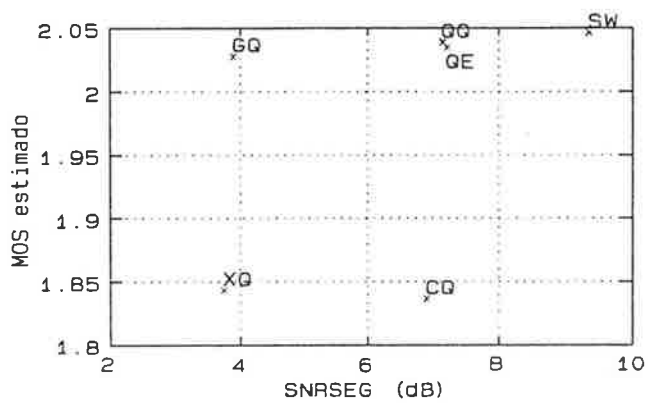


Figura 5.19: Testes de quantização com o sinal de prova d e sinal de referência c .

Capítulo 6

CONCLUSÃO

Foram testados alguns codificadores preditivos de voz por análise mediante síntese selecionados:

- o codificador CELP original com livro de códigos estocástico;
- um codificador com livro de códigos esparsos, o CELP estocástico ternário com ceifagem central (CC);
- um codificador com livros de códigos definidos por bases vetoriais (VSELP).

Comprovamos a superioridade do codificador esparsos com voz sonora. Porém, ela só ocorreu quando a busca da excitação foi feita com ortogonalidade.

O codificador VSELP, apesar de apresentar um desempenho em SNRSEG mais fraco, sempre manteve um bom desempenho em pela medida de distorção espectral. Pode ser que esse bom desempenho espectral conjugado com a sua menor complexidade computacional tenha contribuído para a sua adoção como padrão para telefonia móvel na América do Norte e no Japão [9]. Além disso, deve-se considerar o seu pioneirismo no uso das buscas ortogonais.

O codificador proposto, CELP-SVD, está baseado na representação da excitação por base vetorial que se adapta à variação espectral do sinal. Ainda, permite representar a excitação com qualquer grau de precisão.

A representação completa da excitação requer vários parâmetros, sendo que a seleção criteriosa de apenas alguns deles permite que o codificador trabalhe a uma taxa variável de bits, que é de interesse para as aplicações multimídia [21].

Em testes com uma seleção fixa de parâmetros baseada no desempenho espectral, o CELP-SVD superou o VSELP tanto em qualidade espectral quanto temporal, estando ambos os codificadores não quantizados.

Para reduzir o número de parâmetros do CELP-SVD a uma quantidade viável para a sua quantização a uma taxa próxima de 8 kb/s, recorreremos a uma seleção adaptativa baseada em parâmetros comuns para a codificação do filtro. O desempenho temporal sofreu uma queda, mas o espectral chegou a apresentar considerável melhora em alguns casos.

Os testes de quantização revelaram que

- i)* o ganho do codificador CELP-SVD pode ser quantizado numa escala logarítmica com distorção muito próxima daquela obtida por treinamento;
- ii)* a quantização dos parâmetros da excitação converge para níveis muito próximos com vozes de origem e características muito diversas;
- iii)* os parâmetros do filtro são os mais sensíveis à quantização.

Os resultados *i)* e *ii)* asseguram a robustez dos parâmetros correspondentes, além de simplificar o projeto dos seus respectivos quantizadores.

O resultado *iii)* poderia ser antecipado uma vez que a representação da excitação no CELP-SVD depende dos parâmetros do filtro.

O desempenho do codificador CELP-SVD a taxas mais baixas de transmissão depende de uma distribuição criteriosa de bits entre os parâmetros da excitação. Para tanto, pode ser imprescindível a sua quantização vetorial.

Apêndice A

Produto interno de polinômios

Pode-se utilizar o formalismo dos espaços vetoriais euclidianos para definir um espaço vetorial de polinômios.

Restringiremos nossa discussão a polinômios na variável complexa z^{-1} e apresentaremos o formalismo desenvolvido em [33].

Sejam $F(z)$ e $G(z)$ dois polinômios de coeficientes reais na variável complexa z^{-1} .

Consideremos $F(z)$ e $G(z)$ as funções de transferência de dois filtros (Fig. A.1). Ambos filtros são excitados pela seqüência $\mathbf{x}(n)$, produzindo os sinais $\mathbf{u}(n)$ e $\mathbf{v}(n)$, respectivamente.

Define-se o produto interno entre os polinômios $F(z)$ e $G(z)$, para o intervalo $n = n_0, \dots, n_1$, com o sinal de referência $\mathbf{x}(n)$, como [33]

$$\langle F(z), G(z) \rangle = \sum_{n=n_0}^{n_1} \mathbf{u}(n)\mathbf{v}(n). \quad (\text{A.1})$$

O produto interno é conveniente para a representação dos coeficientes de autocorrelação. As Eqs.(2.22) e (2.23) da seção 2.4, pela aplicação do produto interno com os limites de somatória $n_0 = 0$ e $n_1 = N + p - 1$ e com o sinal de referência $s(n)$, tornam-se

$$R_{i-j} = R_{j-i} = \langle z^{-i}, z^{-j} \rangle. \quad (\text{A.2})$$

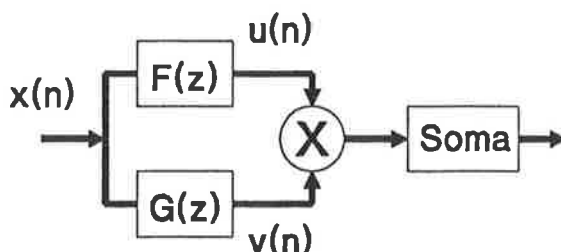


Figura A.1: Interpretação do produto interno.

Apêndice B

Resolução vetorial das equações normais

Usando o formalismo de espaço vetorial de polinômios (Apêndice A), as equações normais (2.28) da seção 2.4 podem ser escritas como

$$\langle A(z), z^{-j} \rangle = 0, \quad j = 1, 2, \dots, p, \quad (\text{B.1})$$

onde

$$A(z) = 1 + \sum_{i=1}^p \alpha_i z^{-i}. \quad (\text{B.2})$$

Esta versão das equações normais afirma que o polinômio solução $A(z)$ tem que ser ortogonal a todos os polinômios de atraso z^{-1}, \dots, z^{-p} .

Apresentaremos o algoritmo de Levinson-Durbin em forma vetorial [33] para a solução das equações normais do método da autocorrelação.

O algoritmo é recursivo e, no curso da resolução das equações normais (B.1) de ordem p , são resolvidos todos os sistemas associados de equações normais de ordens inferiores:

$$\langle A_m(z), z^{-j} \rangle = 0, \quad j = 1, 2, \dots, m \\ m = 1, \dots, p. \quad (\text{B.3})$$

Como a matriz de autocorrelação dada pela Eq.(2.29) da seção 2.4 tem estrutura Toeplitz em decorrência da relação (A.2), a predição progressiva (“forward”) é equivalente à predição regressiva (“backward”).

Define-se o polinômio de predição progressiva como

$$A_m(z) = \sum_{i=0}^m a_{mi} z^{-i} \quad \text{com } a_{m,0} = 1 \quad (\text{B.4})$$

e o polinômio de predição regressiva é dado por

$$B_m(z) = \sum_{i=1}^{m+1} b_{mi} z^{-i} \quad \text{com } b_{m,m+1} = 1. \quad (\text{B.5})$$

Pela natureza Toeplitz da matriz de autocorrelação, os polinômios $A_m(z)$ e $B_m(z)$ guardam a relação

$$B_m(z) = z^{-(m+1)}A_m(1/z), \quad (\text{B.6})$$

ou seja, os coeficientes dos filtros têm simetria especular

$$b_{mi} = a_{m,m+1-i} \quad i = 1, \dots, m. \quad (\text{B.7})$$

Ao final da iteração m , temos o polinômio $A_m(z)$ que satisfaz o conjunto de equações normais (B.3) de índice m . Tomando a j -ésima dessas equações, temos

$$\langle A_m(z), z^{-j} \rangle = 0, \quad (\text{B.8})$$

que, pela definição (B.4) do polinômio progressivo, torna-se

$$\langle 1, z^{-j} \rangle + \sum_{i=1}^m a_{mi} \langle z^{-i}, z^{-j} \rangle = 0 \quad (\text{B.9})$$

ou, usando os coeficientes de autocorrelação conforme a Eq.(A.2), retornamos à forma inicial (seção 2.4) da equação normal

$$R_j + \sum_{i=1}^m a_{mi} R_{|i-j|} = 0. \quad (\text{B.10})$$

Em correspondência especular com a equação normal (B.8), para o polinômio regressivo associado (B.5), temos

$$\langle B_m(z), z^{-(m+1-j)} \rangle = \langle z^{-(m+1)}, z^{-(m+1-j)} \rangle + \sum_{i=1}^m b_{mi} \langle z^{-i}, z^{-(m+1-j)} \rangle \quad (\text{B.11})$$

ou, usando a simetria especular (B.7),

$$\langle B_m(z), z^{-(m+1-j)} \rangle = \langle z^{-(m+1)}, z^{-(m+1-j)} \rangle + \sum_{i=1}^m a_{m,m+1-i} \langle z^{-i}, z^{-(m+1-j)} \rangle, \quad (\text{B.12})$$

que, pela mudança do índice da somatória de i para $l = m + 1 - i$, se transforma em

$$\langle B_m(z), z^{-(m+1-j)} \rangle = \langle z^{-(m+1)}, z^{-(m+1-j)} \rangle + \sum_{l=1}^m a_{m,l} \langle z^{-(m+1-l)}, z^{-(m+1-j)} \rangle. \quad (\text{B.13})$$

Inserindo na Eq.(B.13) os coeficientes de autocorrelação segundo a Eq.(A.2), encontramos

$$\langle B_m(z), z^{-(m+1-j)} \rangle = R_j + \sum_{l=1}^m a_{m,l} R_{|l-j|}, \quad (\text{B.14})$$

cujo 2^o membro coincide com o 1^o membro da Eq.(B.10), conduzindo-nos, assim, à revelação da ortogonalidade do polinômio $B_m(z)$ à potência $z^{-(m+1-j)}$. Ademais,

pela generalidade da escolha da equação j , temos de fato a revelação de um sistema completo de equações normais, que é

$$\langle B_m(z), z^{-j} \rangle = 0, \quad j = 1, 2, \dots, m \\ m = 1, \dots, p. \quad (\text{B.15})$$

Pela definição(B.5), temos o polinômio regressivo da iteração $m - 1$ dado por

$$B_{m-1}(z) = \sum_{i=1}^m b_{m-1,i} z^{-i} \quad \text{com } b_{m-1,m} = 1, \quad (\text{B.16})$$

que, sendo uma combinação linear de todas as potências z^{-j} que comparecem como segundo vetor nas equações ortogonais (B.15), verifica a seguinte relação de ortogonalidade, que, por iguais motivos, também é válida para todas as iterações anteriores e é

$$\langle B_m(z), B_j(z) \rangle = 0, \quad j = 1, \dots, m - 1. \quad (\text{B.17})$$

Consideraremos agora os resíduos progressivo e regressivo da predição de ordem m , que são dados, respectivamente, pelas normas quadradas

$$\|A_m(z)\|^2 = \langle A_m(z), A_m(z) \rangle \\ \|B_m(z)\|^2 = \langle B_m(z), B_m(z) \rangle, \quad (\text{B.18})$$

que, mediante a substituição de apenas uma das ocorrências de cada polinômio pela sua definição, Eq.(B.4) e Eq.(B.5), respectivamente, tornam-se

$$\|A_m(z)\|^2 = \langle A_m(z), 1 \rangle + \sum_{i=1}^m a_{mi} \langle A_m(z), z^{-i} \rangle \\ \|B_m(z)\|^2 = \langle B_m(z), z^{-(m+1)} \rangle + \sum_{i=1}^m a_{mi} \langle A_m(z), z^{-i} \rangle. \quad (\text{B.19})$$

Levando em conta a validade dos sistemas (B.3) e (B.15) de equações ortogonais, as somatórias nas Eqs.(B.19) se anulam, resultando

$$\|A_m(z)\|^2 = \langle A_m(z), 1 \rangle \\ \|B_m(z)\|^2 = \langle B_m(z), z^{-(m+1)} \rangle, \quad (\text{B.20})$$

que, pelas definições (B.4) e (B.5) e usando as relações (B.7), podem ser expandidas para

$$\|A_m(z)\|^2 = \langle 1, 1 \rangle + \sum_{i=1}^m a_{mi} \langle z^{-i}, 1 \rangle \quad (\text{B.21})$$

$$\|B_m(z)\|^2 = \langle z^{-(m+1)}, z^{-(m+1)} \rangle + \sum_{i=1}^m a_{m,m+1-i} \langle z^{-i}, z^{-(m+1)} \rangle. \quad (\text{B.22})$$

Mudando o índice da somatória na Eq.(B.22) de i para $l = m + 1 - i$, obtém-se

$$\|B_m(z)\|^2 = \langle z^{-(m+1)}, z^{-(m+1)} \rangle + \sum_{l=1}^m a_{m,l} \langle z^{-(m+1-l)}, z^{-(m+1)} \rangle. \quad (\text{B.23})$$

Considerando que, pela Eq.(A.2),

$$\langle z^{-(m+1-l)}, z^{-(m+1)} \rangle = \langle z^{-l}, z^0 \rangle = \langle z^{-l}, 1 \rangle,$$

temos o resíduo da predição de ordem m dado por

$$\sigma_m^2 = \|A_m(z)\|^2 = \|B_m(z)\|^2. \quad (\text{B.24})$$

Passamos agora à apresentação do algoritmo de Levinson-Durbin.

Inicialização – Estabelecem-se os polinômios iniciais

$$A_0(z) = 1 \quad \text{e} \quad B_0(z) = z^{-1}. \quad (\text{B.25})$$

Polinômios de grau superior – Na iteração m , especifica-se o polinômio progressivo como uma combinação linear dos polinômios da iteração anterior da forma

$$A_m(z) = A_{m-1}(z) + k_m B_{m-1}(z), \quad (\text{B.26})$$

onde k_m é o coeficiente de reflexão da iteração m . Esta escolha baseia-se na ortogonalidade de $A_{m-1}(z)$ e de $B_{m-1}(z)$ a todas as potências anteriores $z^{-1}, \dots, z^{-(m-1)}$, garantida pela validade das relações normais (B.3) e (B.15) de ordem $m - 1$. Basta-nos, assim, apenas a imposição da última equação normal (B.3)

$$\langle A_m(z), z^{-m} \rangle = 0 \quad (\text{B.27})$$

$$\langle A_{m-1}(z), z^{-m} \rangle + k_m \langle B_{m-1}(z), z^{-m} \rangle = 0, \quad (\text{B.28})$$

que, com a introdução do resíduo da ordem $m - 1$, dado simultaneamente pelas Eqs.(B.24) e (B.20) adaptadas para $m - 1$, determina

$$k_m = -\langle A_{m-1}(z), z^{-m} \rangle / \sigma_{m-1}^2. \quad (\text{B.29})$$

Aplicando (B.6) a (B.26), obtém-se a expressão

$$B_m(z) = z^{-1} (B_{m-1}(z) + k_m A_{m-1}(z)), \quad (\text{B.30})$$

correspondente a (B.26) para o polinômio regressivo.

Resíduo de ordem superior – Como partimos de $A_0(z) = 1$ e os polinômios progressivos de graus superiores foram montados por aplicações sucessivas da Eq.(B.26), temos

$$A_m(z) = 1 + \sum_{i=1}^m k_i B_{i-1}(z). \quad (\text{B.31})$$

Considerando que o conjunto de polinômios $\{B_0(z), \dots, B_{m-1}(z)\}$ é ortogonal (Eq.(B.17)), temos

$$\|A_m(z) - 1\|^2 = \sum_{i=1}^m k_i^2 \|B_{i-1}(z)\|^2, \quad (\text{B.32})$$

que, usando a generalização da relação (B.24) entre os resíduos para outras ordens de predição, torna-se

$$\sigma_m^2 - 2\langle A_m(z), 1 \rangle + \langle 1, 1 \rangle = \sum_{i=1}^m k_i^2 \sigma_{i-1}^2. \quad (\text{B.33})$$

Aplicando a 1^ª das Eqs.(B.20) e a Eq.(A.2) com $i = j = 0$ em (B.33), obtemos

$$\sigma_m^2 = R_0 - \sum_{i=1}^m k_i^2 \sigma_{i-1}^2. \quad (\text{B.34})$$

A partir de (B.34), podemos expressar o resíduo da iteração $m - 1$ por

$$\sigma_{m-1}^2 = R_0 - \sum_{i=1}^{m-1} k_i^2 \sigma_{i-1}^2. \quad (\text{B.35})$$

Subtraindo (B.35) de (B.34), obtemos a expressão iterativa do resíduo

$$\sigma_m^2 = (1 - k_m^2) \sigma_{m-1}^2. \quad (\text{B.36})$$

Os coeficientes de reflexão k_i , $i = 1, \dots, p$, ou coeficientes de correlação parcial (PARCOR), que surgiram como parâmetros auxiliares no algoritmo de Levinson-Durbin, permitem testar facilmente a estabilidade do filtro LPC resultante. A estabilidade se verifica se as condições

$$-1 \leq k_i \leq 1, \quad i = 1, \dots, p. \quad (\text{B.37})$$

forem verdadeiras.

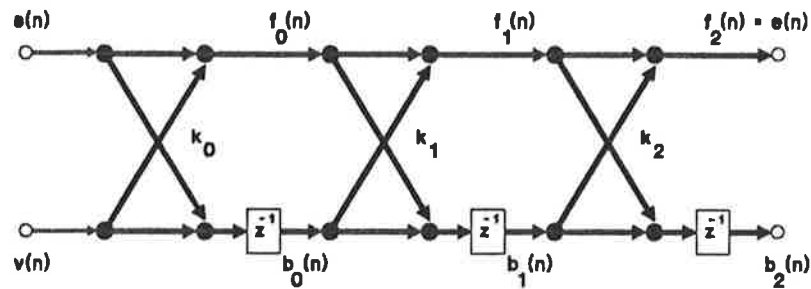
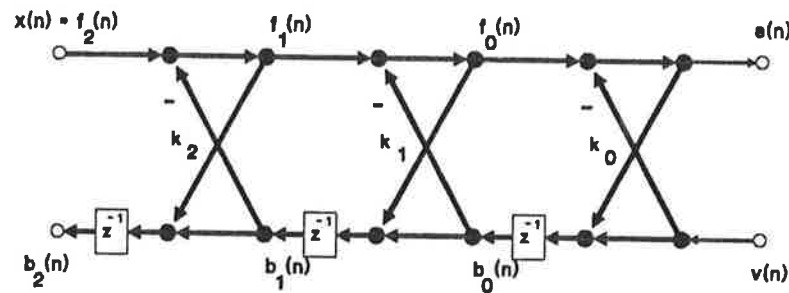
Além disso, os coeficientes de reflexão são suficientes para a representação do filtro inverso $A(z) = A_p(z)$, como mostraremos em seguida.

Numa representação amostral das Eqs.(B.26) e (B.30), sendo $f_{m-1}(n)$ e $b_{m-1}(n)$ os sinais de saída dos filtros $A_{m-1}(z)$ e $B_{m-1}(z)$, respectivamente, (Fig. B.1), temos nas saídas de $A_m(z)$ e de $B_m(z)$ os sinais

$$\begin{aligned} f_m(n) &= f_{m-1}(n) + k_m b_{m-1}(n) \\ b_m(n) &= b_{m-1}(n-1) + k_m f_{m-1}(n-1), \end{aligned} \quad (\text{B.38})$$

respectivamente.

A estrutura modular do algoritmo é mapeada na estrutura do filtro. Até a inicialização (B.25) do algoritmo pode fazer uso da célula básica (B.38) [15]. Para isso,


 Figura B.1: Filtro LPC inverso de ordem $p = 2$ na estrutura em treliça.

 Figura B.2: Filtro LPC de ordem $p = 2$ na estrutura em treliça.

sejam $s(n)$ e $v(n)$ os sinais nas entradas dos filtros $A_0(z)$ e $B_0(z)$, respectivamente, que produzem nas suas saídas os sinais

$$\begin{aligned} f_0(n) &= s(n) + k_0 v(n) \\ b_0(n) &= v(n-1) + k_0 s(n-1) \end{aligned} \quad (\text{B.39})$$

Para reconstruir o filtro inverso $A(z) = A_p(z)$ com sinal de entrada $s(n)$ e sinal de saída $e(n)$, temos que fazer

$$k_0 = 1 \quad (\text{B.40})$$

$$v(n) = 0 \quad n = 0, 1, \dots \quad (\text{B.41})$$

e temos que tomar

$$e(n) = f_p(n). \quad (\text{B.42})$$

Os sinais $f_m(n)$, $m = 0, 1, \dots, p$ são os sinais de erro de predição progressiva (“forward prediction error”) para todos os filtros LPC inversos de ordens menores ou iguais a p . Analogamente, os sinais $b_m(n)$, $m = 0, 1, \dots, p$ são os erros de predição regressiva (“backward prediction error”).

O próprio filtro LPC $H_{LPC}(z) = 1/A(z)$ pode ser representado pelos coeficientes de reflexão na estrutura em treliça (“lattice”) como ilustrado na Fig. B.2.

Para inverter o fluxo do sinal no filtro $A(z)$, vamos, a partir das Eqs.(B.38), expressar o erro progressivo de ordem inferior $m-1$ em função dos erros progressivo de ordem superior m e regressivo de mesma ordem $m-1$, ou seja,

$$\begin{aligned} f_{m-1}(n) &= f_m(n) - k_m b_{m-1}(n) \\ b_m(n) &= b_{m-1}(n-1) + k_m f_{m-1}(n-1). \end{aligned} \quad (\text{B.43})$$

Complementando, temos que

- impor a condição (B.40) de reflexão na saída do sinal de voz e a condição (B.41) de entrada regressiva nula e
- excitar o filtro $H_{LPC}(z)$ com o sinal $x(n) = Gu(n)$ (seção 2.2) pela entrada $f_p(n)$, ou seja, faz-se

$$f_p(n) = x(n). \quad (\text{B.44})$$

Pode-se fazer uma analogia do filtro LPC na estrutura em treliça (Fig. B.2) com um modelo acústico do aparelho fonador por p tubos sem perda de áreas uniformes e distintas [48] [33]. Em termos deste modelo acústico, podemos descrever alguns parâmetros variáveis e condições do filtro LPC como segue

Onda progressiva – os sinais de erro progressivo $\{f_m(n), m = p, p-1, \dots, 0\}$ são amostras da onda acústica progressiva, que parte da glote em direção aos lábios;

Onda regressiva – os sinais de erro regressivo $\{b_m(n), m = 0, 1, \dots, p\}$ são amostras da onda acústica regressiva, que parte dos lábios em direção à glote;

Reflexão – os coeficientes de reflexão $\{k_m, m = 1, 2, \dots, p\}$ fornecem a cada descontinuidade de área na interface entre dois tubos a razão de reflexão das ondas

$$k_m = \frac{A_m - A_{m+1}}{A_m + A_{m+1}}, \quad m = 1, 2, \dots, p, \quad (\text{B.45})$$

onde A_0 e A_{p+1} representam as áreas de abertura dos lábios e da glote, respectivamente, e $\{A_m, m = 1, 2, \dots, p\}$ são as áreas uniformes de cada um dos p tubos;

Modo de excitação e terminação – a linha de transmissão composta pelo conjunto de tubos sofre injeção de sinal na glote (Eq.(B.44)) e apresenta reflexão total do sinal nos lábios (Eq.(B.40), onde não há incidência de sinal externo (Eq.(B.41)).

Apêndice C

Coeficientes LSP do filtro LPC

O filtro LPC foi definido na representação através dos coeficientes de predição linear $\{a_i, i = 1, \dots, p\}$ no Capítulo 2.

Na resolução do sistema de equações normais apresentada no Apêndice B foram usados os coeficientes de reflexão $\{k_i, i = 1, \dots, p\}$ como intermediários no processo de cálculo, durante o qual possibilitam um simples teste (Eq.(B.37)) da estabilidade do filtro LPC.

Mostrou-se também que os coeficientes de reflexão podem representar o filtro independentemente, dando origem à estrutura em treliça do filtro LPC (Apêndice B). Esta estrutura admite analogias com um modelo acústico do aparelho fonador.

Ademais, os coeficientes de reflexão são muito importantes para a codificação do filtro LPC porque ele é bem menos sensível a erros na sua quantização do que a erros na quantização dos coeficientes de predição [22]. Inclusive, foram elaborados algoritmos para o cálculo dos coeficientes de reflexão que não recorrem aos coeficientes de predição nem como parâmetros auxiliares, como o algoritmo de Le Roux-Gueguen [30].

Porém, os coeficientes de reflexão não são convenientes para a interpolação do filtro LPC [15].

Por outro lado, a representação do filtro LPC por *pares de raias espectrais* ou “line spectrum pairs” (LSP) apresenta boas características de quantização e de interpolação [15] [44].

Os LSPs podem ser introduzidos imaginando que, após a determinação conforme o Apêndice B dos polinômios $A_p(z)$ e $B_p(z)$ para a predição de ordem p , queiramos determinar o polinômio progressivo de ordem $p + 1$ (Eq.(B.26)), que seria dado por

$$A_{p+1}(z) = A_p(z) + k_{p+1}B_p(z). \quad (C.1)$$

Entretanto, nossa intenção não é prosseguir com o cálculo do coeficiente de reflexão k_{p+1} em (C.1), mas sim arbitrar valores a k_{p+1} para “produzir” polinômios com características interessantes.

Notamos no Apêndice B que o filtro LPC normalmente apresenta uma reflexão perfeita na terminação labial. Se lhe for introduzida outra reflexão perfeita na terminação glotal, o filtro LPC se tornará ressonante perfeito, isto é, sem perdas. Nessas

condições, o filtro inverso correspondente terá os zeros na circunferência unitária do plano z .

Há duas possibilidades de introduzir uma reflexão glotal [48], definidas por $k_{p+1} = -1$ e $k_{p+1} = 1$. Os polinômios de grau $p+1$ assim especificados são, respectivamente:

$$P(z) = A_p(z) - B_p(z); \quad (\text{C.2})$$

$$Q(z) = A_p(z) + B_p(z). \quad (\text{C.3})$$

Sendo o polinômio progressivo do filtro LPC dado por

$$A_p(z) = 1 + a_1 z^{-1} + \dots + a_p z^{-p}, \quad (\text{C.4})$$

pelas relações especulares (B.7), o polinômio regressivo é

$$B_p(z) = a_p z^{-1} + \dots + a_1 z^{-p} + z^{-(p+1)}. \quad (\text{C.5})$$

Substituindo as expressões (C.4) e (C.5) nas definições (C.2) e (C.3), temos as seguintes expressões para os polinômios auxiliares ao cálculo dos coeficientes LSP:

$$P(z) = 1 + (a_1 - a_p)z^{-1} + \dots + (a_p - a_1)z^{-p} - z^{-(p+1)} \quad (\text{C.6})$$

$$Q(z) = 1 + (a_1 + a_p)z^{-1} + \dots + (a_p + a_1)z^{-p} + z^{-(p+1)}, \quad (\text{C.7})$$

mostrando, assim, que o polinômio $P(z)$ tem vetor de coeficientes anti-simétrico e o polinômio $Q(z)$ tem vetor de coeficientes simétrico.

Para ordem p par de análise LPC, os polinômios $P(z)$ e $Q(z)$ podem ser fatorados [15] como

$$P(z) = (1 - z^{-1}) \prod_{i=1}^{p/2} (1 - (2 \cos \omega_{2i})z^{-1} + z^{-2}) \quad (\text{C.8})$$

$$Q(z) = (1 + z^{-1}) \prod_{i=1}^{p/2} (1 - (2 \cos \omega_{2i-1})z^{-1} + z^{-2}), \quad (\text{C.9})$$

com as relações de ordem

$$\begin{aligned} 0 < \omega_2 < \omega_4 < \dots < \omega_p < \pi & \text{ e} \\ 0 < \omega_1 < \omega_3 < \dots < \omega_{p-1} < \pi, & \text{ respectivamente.} \end{aligned}$$

As raízes do polinômio anti-simétrico, por inspeção de (C.8), são

$$\{1, e^{j\omega_{2i}}, e^{-j\omega_{2i}}, i = 1, 2, \dots, p/2\}$$

e as raízes do polinômio simétrico, por inspeção de (C.9), são

$$\{-1, e^{j\omega_{2i-1}}, e^{-j\omega_{2i-1}}, i = 1, 2, \dots, p/2\}.$$

Os *pares de raias espectrais* ou “line spectrum pairs” (LSP) são definidos como os argumentos $\{\omega_i, i = 1, 2, \dots, p\}$ situados entre 0 e π das raízes dos polinômios $P(z)$ e $Q(z)$.

As raízes dos polinômios simétrico e anti-simétrico ocorrem intercaladas, ou seja, verifica-se a seguinte relação geral de ordem

$$0 < \omega_1 < \omega_2 < \dots < \omega_{p-1} < \omega_p < \pi. \quad (\text{C.10})$$

Apêndice D

Expansão de faixa de filtros LPC

O objetivo deste apêndice é a determinação de uma regra de transformação dos coeficientes de um filtro LPC

$$H(z) = \frac{1}{1 + \sum_{k=1}^p a_k z^{-k}} \quad (\text{D.1})$$

que resulte num filtro tendo essencialmente as mesmas freqüências de ressonância, porém, com as suas larguras de faixa maiores.

A derivação vai ser iniciada com a expansão de faixa de um filtro de tempo contínuo $H_c(s)$, associado ao filtro $H(z)$ pela transformação de variáveis

$$z = \exp(s/f_a) \iff s = f_a \ln z, \quad (\text{D.2})$$

onde f_a é freqüência de amostragem.

Supomos uma única ressonância estreita em $H_c(j\omega)$, isto é,

$$B \ll \omega_0, \quad (\text{D.3})$$

onde B é a largura de faixa da ressonância e ω_0 é a sua freqüência central.

Nessa condição, temos a função de transferência

$$H_c(s) = \frac{1}{(s - p_0)(s - p_0^*)}, \quad (\text{D.4})$$

onde

$$p_0 = -\sigma_0 + j\omega_0, \quad (\text{D.5})$$

com $\sigma_0 > 0$, e p_0^* são os pólos de $H_c(s)$ (Fig. D.1).

Sejam p_1 e p_1^* os pólos de $H_{ec}(s)$, o filtro expandido em tempo contínuo.

A condição (D.3) implica que os pólos estão muito próximos ao eixo $j\omega$ e, por (D.5), podemos escrever

$$\sigma_0 \ll \omega_0. \quad (\text{D.6})$$

Denotando o módulo da resposta em freqüência

$$M(\omega) = |H_c(j\omega)|, \quad (\text{D.7})$$

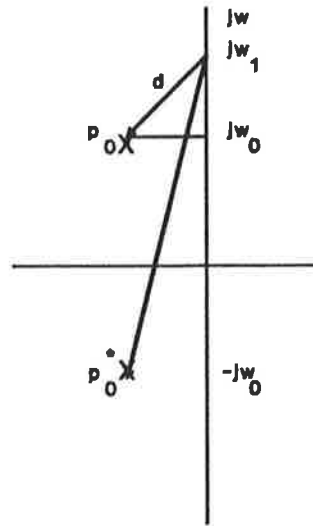


Figura D.1: Diagrama de pólos de um filtro com ressonância estreita (fora de escala)

temos, considerando (D.5), que

$$M(\omega_0) = \frac{1}{\sigma_0 |j\omega_0 - p_0^*|}. \quad (\text{D.8})$$

Adicionalmente, levando em conta (D.6), temos

$$M(\omega_0) = \frac{1}{\sigma_0 \cdot 2\omega_0}. \quad (\text{D.9})$$

Em ω_1 , o limite superior da faixa de passagem de 3 dB (Fig. D.1), que é

$$\omega_1 = \omega_0 + \pi B, \quad (\text{D.10})$$

temos, por definição,

$$M(\omega_1) = \frac{1}{\sqrt{2}} M(\omega_0) \quad (\text{D.11})$$

e, da Fig. D.1, levando em conta (D.6), temos

$$M(\omega_1) = \frac{1}{d \cdot 2\omega_0}. \quad (\text{D.12})$$

Substituindo (D.9) em (D.11) e comparando (D.11) com (D.12), temos

$$d = \sqrt{2}\sigma_0. \quad (\text{D.13})$$

Considerando (D.13) e (D.5), extraímos do triângulo retângulo que

$$\omega_1 - \omega_0 = \sigma_0 \quad (\text{D.14})$$

e, lembrando (D.10), temos

$$B = \frac{\sigma_0}{\pi}. \quad (\text{D.15})$$

Aumentar a largura de faixa de ΔB , segundo (D.15), significa

$$B + \Delta B = \frac{(\sigma_0 + \Delta\sigma_0)}{\pi}. \quad (\text{D.16})$$

Para tanto, temos que deslocar o pólo p_0 de $H_c(s)$ para o pólo

$$p_1 = -(\sigma_0 + \Delta\sigma_0) + j\omega_0 \quad (\text{D.17})$$

ou, substituindo $\Delta\sigma_0$ de (D.16), temos

$$p_1 = -\sigma_0 - \pi\Delta B + j\omega_0. \quad (\text{D.18})$$

Comparando (D.18) e (D.5), temos

$$p_1 = p_0 - \pi\Delta B. \quad (\text{D.19})$$

A transformação de pólos expressa por (D.19) pode ser efetuada mediante a transformação de variáveis de Laplace

$$s' = s + \pi\Delta B. \quad (\text{D.20})$$

Substituindo a equação direita de (D.2) em (D.20), temos

$$s' = f_a \ln z + \pi\Delta B \quad (\text{D.21})$$

ou

$$s' = f_a \ln \frac{z}{\exp(-\pi\Delta B/f_a)}. \quad (\text{D.22})$$

Definindo

$$z' = z/\gamma \quad (\text{D.23})$$

com

$$\gamma = \exp(-\pi\Delta B/f_a), \quad (\text{D.24})$$

podemos reescrever (D.22) como

$$s' = f_a \ln z'. \quad (\text{D.25})$$

Comparando (D.25) com a equação direita de (D.2), vemos que a variável s' que representa o plano da ressonância alargada está associada por amostragem com z' . Isto é, o filtro expandido que procuramos é dado por $H(z')$.

Assim, voltando à variável z através do mapeamento (D.23), obtemos

$$H_c(z) = H(z/\gamma) \quad (\text{D.26})$$

$$= \frac{1}{A(z/\gamma)}. \quad (\text{D.27})$$

Considerando que

$$A(z) = \sum_{k=0}^p a_k z^{-k} \text{ com } a_0 = 1, \quad (\text{D.28})$$

temos

$$A(z/\gamma) = \sum_{k=0}^p \alpha_k z^{-k} \text{ com } \alpha_0 = 1, \quad (\text{D.29})$$

onde

$$\alpha_k = a_k \gamma^k. \quad (\text{D.30})$$

Resumindo, dado o filtro LPC $H(z)$ especificado pelos coeficientes de predição linear $\{a_k\}$, para obter o filtro $H_e(z)$ com expansão de faixa ΔB , determinamos o coeficiente de expansão de faixa γ por (D.24) e, em seguida, determinamos $\{\alpha_k\}$, que é o conjunto de coeficientes de predição que define o filtro $H_e(z)$.

Bibliografía

- [1] APOSTOL, T.M. *Calculus: cálculo con funciones de varias variables y álgebra lineal, con aplicaciones a las ecuaciones diferenciales y a las probabilidades*. Trad. de Francisco Vélez Cantarell 2.ed. Barcelona, Reverté, 1975, v.2.
- [2] ATAL, B.S.; SCHROEDER, M.R. Predictive coding of speech signals. In: CONF. COMMUN. AND PROCESS., 1967. *Proceedings.* , p.360-1, s.l., IEEE, 1967.
- [3] ATAL, B.S.; SCHROEDER, M.R. Adaptive predictive coding of speech signals. *Bell Syst. Tech. J.*, v. 49, n. 8, p.1973-86, Oct. 1970.
- [4] ATAL, B.S.; SCHROEDER, M.R. Predictive coding of speech signals and subjective error criteria. *IEEE Trans. Acoust., Speech, Signal Processing*, v.27, n.3, p.247-54, June 1979.
- [5] ATAL, B.S. Predictive coding of speech at low bit rates. *IEEE Trans. Commun.*, v.30, n.4, p.600-14, Apr. 1982.
- [6] ATAL, B.S.; REMDE, J.R. A new model of LPC excitation for producing natural-sounding speech at low bit rates. In: IEEE INT. CONF. ACOUST., SPEECH, SIGNAL PROCESSING, Paris, 1982. *Proceedings.* v.1, p.614-7, s.l., IEEE, 1982.
- [7] ATAL, B.S. A model of LPC excitation in terms of the autocorrelation matrix of the impulse response of the LPC filter. In: IEEE INT. CONF. ACOUST., SPEECH, SIGNAL PROCESSING, Glasgow, 1989. *Proceedings.* v.1, p.45-8, s.l., IEEE, 1989.
- [8] BELL, C.G. et al. Reduction of speech spectra by analysis-by-synthesis techniques. *J. Acoust. Soc. Am.*, v.33, n.12, p.1925-36, Dec. 1961.
- [9] BOISSON DE MARCA, J.R.; DE MARTINO, E. Codificadores de voz para telefonía móvil celular. In: SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES, Brasília, 1991. *Anais.*, p.283-8, Brasília, UnB, 1991.
- [10] CAMPBELL, J.P. et al. The proposed federal standard 1016 4800 bps voice coder: CELP. *Speech Technology*, v.5, n.2, p.58-64, Apr. May 1990.
- [11] CCITT 32 kbit/s adaptive differential pulse code modulation (ADPCM). Recommendation III.3, Málaga-Torremolinos, 1984.
- [12] CCITT Artificial voices. Recommendation P.50, Blue Book, v.V, Melbourne, IXth Plenary Assembly, 1988.
- [13] DUDLEY, H. The vocoder. *Bell Labs Rec.*, v.18, p.122-6, Dec. 1939.

- [14] FLANAGAN, J.L. *Speech analysis, synthesis and perception*. 2.ed. Berlin, Springer, 1972.
- [15] FURUI, S. *Digital speech processing, synthesis, and recognition*. New York, Marcel Dekker, 1985.
- [16] GERSON, I.A.; JASIUK, M.A. Vector sum excited linear prediction (VSELP) speech coding at 8 kbps. In: IEEE INT. CONF. ACOUST., SPEECH, SIGNAL PROCESSING, Albuquerque, 1990. *Proceedings*. v.1, p.461-4, s.l., IEEE, 1990.
- [17] GIBSON, J.D. Adaptive prediction for speech encoding. *IEEE ASSP Magazine*, v.1, n.3, p.12-26, July 1984.
- [18] GOLUB, G.H.; VAN LOAN, C.F. *Matrix computations*. 2.ed. Baltimore, The Johns Hopkins, 1989.
- [19] GRAY, JR., A.H.; MARKEL, J.D. Distance measures for speech processing. *IEEE Trans. Acoust., Speech, Signal Processing*, v.24, n.5, p.380-91, Oct. 1976.
- [20] HESS, W. *Pitch determination of speech signals*. Berlin, Springer, 1983.
- [21] IMAMURA, O.C. *Multi-pulse linear predictive coding of speech based on peak discrimination method*. Tokyo, 1987. Doctor of Engineering Thesis - The University of Tokyo, Graduate School of Engineering.
- [22] ITAKURA, F.; SAITO, S. On the optimum quantization of feature parameters in the PARCOR speech synthesizer. In: IEEE CONF. ON SPEECH COMMUN. AND PROCESS., 1972. *Proceedings*., p.434-7, s.l., IEEE, 1972.
- [23] JAYANT, N.S.; NOLL, P. *Digital coding of waveforms*. Englewood Cliffs, Prentice-Hall, 1984.
- [24] KITAWAKI, N. et al. Comparison of objective speech quality measures for voice band coders. In: IEEE INT. CONF. ACOUST., SPEECH, SIGNAL PROCESSING, Paris, 1982. *Proceedings*. v.2, p.1000-2, s.l., IEEE, 1982.
- [25] KITAWAKI, N. et al. Objective quality evaluation for low-bit-rate speech coding systems. *IEEE J. Select. Areas Commun.*, v.6, n.2, p.242-8, Feb. 1988.
- [26] KLEIJN, W.B. et al. Fast methods for the CELP speech coding algorithm. *IEEE Trans. Acoust., Speech, Signal Processing*, v.38, n.8, p.1330-42, Aug. 1990.
- [27] KROON, E.F. et al. Regular-pulse excitation - a novel approach to effective and efficient multipulse coding of speech. *IEEE Trans. Acoust., Speech, Signal Processing*, v.34, n.5, p.1054-63, Oct. 1986.
- [28] KROON, P.; DEPRETTERE, E.F. A class of analysis-by-synthesis predictive coders for high quality speech synthesis at rates between 4.8 and 16 kbits/s. *IEEE J. Select. Areas Commun.*, v.6, n.2, p.353-63, Feb. 1988.
- [29] LINDE, Y. et al. An algorithm for vector quantizer design. *IEEE Trans. Commun.*, v.28, n.1, p.84-95, Jan. 1982.

- [30] LE ROUX, J.; GUEGUEN, C. A fixed point computation of partial correlation coefficients. *IEEE Trans. Acoust., Speech, Signal Processing*, v.25, n.3, p.257-9, June 1977.
- [31] MAKHOUL, J. Linear prediction: a tutorial review. *Proc. IEEE*, v.63, n.4, p.561-80, Apr. 1975.
- [32] MAKHOUL, J. et al. Vector quantization in speech coding. *Proc. IEEE*, v.73, n.11, p.155-88, Nov. 1985.
- [33] MARKEL, J.D.; GRAY, A.H. *Linear prediction of speech*. Berlin, Springer, 1976.
- [34] MARPLE, Jr., S.L. *Digital spectral analysis*. Englewood Cliffs, Prentice-Hall, 1987.
- [35] MATHEWS, M.V. et al. Pitch synchronous analysis of voiced sounds. *J. Acoust. Soc. Am.*, v. 33, n.2, p.179-86, Feb. 1961.
- [36] OPPENHEIM, A.V. et al. Non-linear filtering of multiplied and convolved signals. *Proc. IEEE*, v.56, n.8, p.1264-91, Aug. 1968.
- [37] OPPENHEIM, A.V.; SCHAFER, R.W. *Discrete-time signal processing*. Englewood Cliffs, Prentice-Hall, 1989.
- [38] RABINER, L.R.; SCHAFER, R.W. *Digital processing of speech signals*. Englewood Cliffs, Prentice-Hall, 1978.
- [39] ROSE, R.C.; BARNWELL, T.P. Design and performance of an analysis-by-synthesis class of predictive speech coders. *IEEE Trans. Acoust., Speech, Signal Processing*, v.38, n.9, p.1489-1503, Sept. 1990.
- [40] SAITO, S.; ITAKURA, F. *The theoretical consideration of statistically optimum methods for speech spectral density*. Tokyo, Electrical Communication Laboratory/N.T.T., 1966. (Report No. 3107).
- [41] SCHROEDER, M.R. Vocoders: analysis and synthesis of speech. *Proc. IEEE*, v.54, p.720-34, May 1966.
- [42] SCHROEDER, M.R.; ATAL, B.S. Code-excited linear prediction (CELP): High quality speech at very low bit rates. In: IEEE INT. CONF. ACOUST., SPEECH, SIGNAL PROCESSING, Tampa, 1985. *Proceedings*. v.2, p.437-40, s.l., IEEE, 1985.
- [43] SCHROEDER, M.R. Predictive coding of speech: historical review and directions for future research. In: IEEE INT. CONF. ACOUST., SPEECH, SIGNAL PROCESSING, Tokyo, 1986. *Proceedings*. v.4, p.3157-64, s.l., IEEE, 1986.
- [44] SUGAMURA, N.; FARVARDIN, N. Quantizer design in LSP analysis-synthesis. *IEEE J. Select. Areas Commun.*, v.6, n.2, p.432-40, Feb. 1988.
- [45] TOHKURA, Y. et al. Spectral smoothing techniques in PARCOR speech analysis-synthesis. *IEEE Trans. Acoust., Speech, Signal Processing*, v.26, n.6, p.587-96, Dec. 1978.
- [46] TREMAIN, T.E. The government standard linear predictive coding algorithm: LPC-10. *Speech Technology*, v.1, n.2, p.40-9, Apr. 1982.

- [47] VARY, P. et al. Speech codec for the European mobile radio system. In: IEEE INT. CONF. ACOUST., SPEECH, SIGNAL PROCESSING, 1988. *Proceedings*. v.1, p.227-30, s.l., IEEE, 1988.
- [48] WAKITA, H. Linear prediction voice synthesizers: Line-spectrum pairs (LSP) is the newest of several techniques. *Speech Technology*, v.1, n.1, p.17-22, Fall 1981.
- [49] WANG, S. et al. Auditory distortion measure for speech coding. In: IEEE INT. CONF. ACOUST., SPEECH, SIGNAL PROCESSING, Toronto, 1991. *Proceedings*. v.1, p.493-6, s.l., IEEE, 1991.