

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA

Este exemplar corresponde à redação final da tese defendida por Fernando Oscar Runstein e aprovada pela Comissão Julgadora em 2 de Março de 1990.

F. Violaro

24/04/90

CODIFICAÇÃO DIGITAL DE VOZ EM SUB-BANDAS

POR: FERNANDO OSCAR RUNSTEIN *autor*

Engenheiro Eletricista Eletrônico

Universidad Nacional de Córdoba - ARGENTINA - 1985

ORIENTADOR: PROF. DR. FÁBIO VIOLARO *lim*

Professor do Departamento de Comunicações da
Faculdade de Engenharia Elétrica da UNICAMP

BANCA EXAMINADORA:

Fábio Violaro (UNICAMP)

José Geraldo Chiquito (UNICAMP)

Fernando Sakane (CTA-ITA)

Suplente: Michel Daoud Yacoub (UNICAMP)

Tese apresentada à Faculdade de Engenharia Elétrica da
UNICAMP, como parte dos requisitos exigidos para a
obtenção do título de Mestre em Engenharia Elétrica.

CAMPINAS, 2 de Março de 1990

Este trabalho foi realizado com o apoio da CAPES e do CPqD-TELEBRÁS, através do Convênio UNICAMP/TELEBRÁS 208/87.

AGRADECIMENTOS

Ao meu Orientador, Professor Dr. Fábio Violaro, pela sua disposição, ajuda e confiança na realização deste trabalho.

Aos membros do Departamento de Comunicações, DECOM, e do Convênio Unicamp/Telebrás pela sua valiosa colaboração.

Ao CPqD-Telebrás, na pessoa de José Sindi Yamamoto, e à CAPES, pelo suporte financeiro durante o Mestrado.

À senhora Maria Lucia C. Cardoso e ao senhor Gerson A. Marconi pela confecção de parte dos desenhos da Tese.

Aos senhores C.F.Jaccoud e Raimundo Duarte, pela cessão de material bibliográfico.

Aos meus pais e às minhas irmãs, simplesmente por serem.

A Mirta, pelo seu amor e paciência.

A Ricardo Enrique Saad e Jorge Alfredo Rocha por terem me aturado durante tanto tempo e pela sua amizade e companherismo.

A Carlos José de Campos Matos e Eliane Ribeiro Matos pela sua amizade e pelos bons momentos compartilhados.

A Carlos José M. Olguín, José Eudoxio C. de Queiroz, Mônica G. Zoppi Fontana, João Batista Rickheim Filho, Adrián S. De Witt Batista, Maria Conceição Perez Young, Edwin Mauricio Loboschi, Rogério T. Ramos e Ângelo D. Rossi pela sua amizade e camaradagem.

A Susana Marta Daniele, Maurílio Nunes Vieira, Flávia Martinho Ferreira, Raimundo B. Do Nascimento, Edson José Nagle, Juan José Lopensino, e demais amigos, amigas e colegas da UNICAMP, por tornar mais agradável minha estada no Brasil.

AOS MEUS PAIS.

PARA MIRTA.

S U M Á R I O

A codificação digital de sinais de voz a baixas taxas (16 [Kbits/s] ou menores), com qualidade comparável aos sistemas em uso atualmente (PCM a 64 [Kbits/s], ADPCM a 32 [Kbits/s]), é um dos objetivos que se pretende atingir em pouco tempo. Os codificadores de sub-bandas se apresentam como bons candidatos para alcançar este objetivo, preenchendo o vazio que existe entre sofisticados codificadores de baixas taxas que trabalham no domínio da frequência, como codificadores ATC (Adaptive Transform Coding), e codificadores simples, porém de altas taxas, que trabalham no domínio do tempo, como PCM, APCM, ADM, ADPCM, etc.

Este trabalho trata dos codificadores de sub-bandas aplicados a sinais de voz e das questões que dizem respeito a seu projeto e desempenho.

ABSTRACT

Digital coding of speech at low bit rate (16 [Kbits/s] or less), whose performance could be comparable to today's digital systems (64 [Kbits/s] PCM, 32 [Kbits/s] ADPCM), is one of the objectives to be achieved in the very near future. The sub-band coders lead themselves as good candidates for this purpose, bridging the gap between the complex low bit rate coders, working in the frequency domain (such as the Adaptive Transform Coding - ATC) and the simple, but high bit rate coders, working in the time domain (such as PCM, APCM, ADM, ADPCM, etc).

This work investigates the sub-band speech coders and the questions related to their design and performance.

ÍNDICE

CAP.1 - INTRODUÇÃO	1
1.1 - Codificação digital de voz	1
1.2 - Conteúdo do trabalho	3
CAP.2 - CODIFICADORES DE SUB-BANDAS	5
2.1 - Introdução	5
2.2 - Codificadores de sub-bandas	6
2.2.1 - Estrutura divisora e reconstutora do espectro	10
2.2.2 - Codificação dos sinais a transmitir	11
2.3 - Vantagens da codificação em sub-bandas	13
2.4 - Cálculo da taxa de transmissão, nos codificadores de sub-bandas	15
REFERÊNCIAS	17
CAP.3 - ESTRUTURA FILTRANTE NOS CODIFICADORES DE SUB-BANDAS	19
3.1 - Introdução	19
3.2 - Banco de filtros passa-banda em paralelo	20
3.2.1 - Divisão do espectro utilizando filtros passa-banda em paralelo	20
3.2.2 - Considerações sobre o projeto de filtros passa-banda em paralelo	24

3.2.2.1 - Divisão do espectro em sub-bandas de igual largura	24
3.2.2.2 - Divisão do espectro em sub-bandas de largura diferente	25
3.2.3 - Distorção no banco de filtros passa-banda	30
3.2.3.1 - Distorção na dizimação	31
3.2.3.2 - Distorção na interpolação	34
3.3 - Banco de filtros espelho em quadratura (QMF Bank)	36
3.3.1 - Estudo do esquema básico de um banco QMF	36
3.3.2 - Implementação e cálculo do banco QMF	46
3.3.3 - Divisão do espectro em 2^p sub-bandas de igual largura, utilizando-se filtros QMF	52
3.3.4 - Divisão do espectro em oitavas, utilizando-se filtros QMF	58
REFERÊNCIAS	61
 CAP.4 - CODIFICAÇÃO DOS SINAIS NOS CODIFICADORES DE SUB-BANDAS	 65
4.1 - Introdução	65
4.2 - Tipos de quantizador	66
4.2.1 - Quantizadores uniformes	67
4.2.2 - Quantizadores não uniformes	68
4.2.3 - Quantizadores uniformes e não uniformes para uma relação sinal/ruído de quantização ótima	69
4.2.4 - Quantizadores adaptativos	70
4.2.4.1 - Adaptação AQF	72
4.2.4.2 - Adaptação AQB	74
4.2.5 - Quantizadores fixos	75
4.3 - Características dos sinais de voz	75
4.4 - Codificação dos sinais nas diferentes sub-bandas	81
4.4.1 - Codificador BC-PCM	81
4.4.2 - Cálculo da taxa de informação lateral nos codificadores BC-PCM	89

4.4.3 - Codificador APCM-AQB	92
4.4.4 - Algoritmo robusto de GOODMAN e WILKINSON	96
4.5 - Atribuição de bits por banda e utilização de BC-PCM ou APCM-AQB	98
REFERÊNCIAS	100
 CAP.5 - TESTES E DESEMPENHO	 103
5.1 - Introdução	103
5.2 - Descrição dos objetivos	104
5.2.1 - Taxa final do codificador	105
5.2.2 - Número de sub-bandas em que se divide o espectro	105
5.2.3 - Sub-bandas codificadas e transmitidas	105
5.2.4 - Número de bits atribuído a cada sub-banda e método de codificação empregado	106
5.2.5 - Número de "taps" dos filtros QMF	106
5.2.6 - Taxas de erros de canal	107
5.3 - Descrição do "Software" utilizado	108
5.4 - Descrição das simulações	110
5.4.1 - Medição da relação sinal/ruído convencional e segmental	112
5.5 - Sistemas simulados e resultados obtidos	113
5.5.1 - À taxa de 8 [Kbits/s]	114
5.5.2 - À taxa de 16 [Kbits/s]	118
5.5.3 - À taxa de 24 [Kbits/s]	123
5.6 - Comparação com outros codecs	130
REFERÊNCIAS	133
 CAP.6 - CONCLUSÕES	 135
6.1 - Análise dos resultados obtidos	135

APÊNDICES	141
APÊNDICE A - DIZIMAÇÃO E INTERPOLAÇÃO DE SEQUÊNCIAS E SUA RELAÇÃO COM A LARGURA DO ESPECTRO . . .	143
APÊNDICE B - QUANTIZADORES ÓTIMOS	149
APÊNDICE C - FAMÍLIA DE FILTROS QMF UTILIZADA NAS SIMULAÇÕES	153

CAPÍTULO 1

INTRODUÇÃO

CONTEÚDO

1.1 - Codificação digital de voz	1
1.2 - Conteúdo do trabalho	2

1.1 CODIFICAÇÃO DIGITAL DE VOZ

Os estudos na área de codificação e transmissão digital de voz, tem como objetivo a busca de novas técnicas que permitam a transmissão de voz com a maior qualidade possível, à menor taxa de transmissão e com baixo custo.

Tipicamente, o custo de um sistema de codificação digital de voz está diretamente relacionado à complexidade do codificador, a qual, por sua vez, é proporcional à eficiência do codificador e à eficiência no uso do canal de comunicação.

Na medida em que a taxa total do codificador diminui, a complexidade do sistema aumenta e com ela o seu custo de implementação. Adicionalmente, a complexidade pode inviabilizar sua implementação em tempo real.

Atualmente as taxas em uso para transmissão digital de voz em telefonia, são de 64 e 32 [Kbits/s]. Seria desejável poder diminuir ainda mais estes valores, para assim aumentar a quantidade de informação a transmitir em um mesmo canal de comunicação. Desta forma, um dos objetivos a alcançar a curto prazo é a codificação de voz a taxas iguais o menores a

16 [Kbits/s] com qualidade comparável aos sistemas em uso atualmente (PCM a 64 [Kbits/s] e ADPCM a 32 [Kbits/s]).

Para conseguir este objetivo, os codificadores devem explorar ao máximo as características próprias dos sinais de voz, tais como sua "distribuição de amplitudes", "autocorrelação entre amostras" e "densidade espectral de potência". Para isso existem diversas técnicas de codificação, algumas trabalhando no domínio do tempo e outras no domínio da frequência. No domínio do tempo podem ser citados os codificadores PCM (Pulse Code Modulation), APCM (Adaptive PCM), ADPCM (Adaptive Differential PCM), ADM (Adaptive Delta Modulation), etc. No domínio da frequência podem ser citados os codificadores ATC (Adaptive Transform Coding) e SBC (Sub-Band Coding).

Para baixas taxas (menores que 20 [Kbits/s]), os codificadores no domínio da frequência tem tido, até agora, um desempenho superior aos que trabalham no domínio do tempo e, entre os dois tipos mencionados acima, o SBC apresenta menor complexidade que o ATC (em contrapartida os codificadores ATC apresentam melhor desempenho). Assim a complexidade de um codificador de sub-bandas pode-se comparar à de um ADPCM com preditor fixo. Em consequência, os codificadores de sub-bandas se apresentam como bons candidatos para diversas aplicações em baixas taxas, tais como: telefonia convencional, telefonia digital móvel, armazenamento de voz, etc.

O objetivo deste trabalho foi o estudo, projeto e simulação, em computador, de sistemas de codificação em sub-bandas aplicados a sinais de voz e a avaliação do desempenho dos mesmos com a variação dos vários parâmetros que intervêm no seu comportamento. A metodologia seguida para conseguir este objetivo foi a seguinte: num primeiro estágio foi feito um estudo da técnica de codificação em sub-bandas para posteriormente estudar as características dos sinais de voz em banda integral e em sub-bandas. Com estes elementos foi possível definir as características gerais que deveriam reunir os codecs e, em consequência, projetar e desenvolver

um "software" de trabalho, com o qual se realizaram simulações de diversos sistemas de codificação em sub-bandas trabalhando às mais variadas taxas de transmissão. As simulações foram realizadas em tempo não real, em linguagem Pascal.

Quanto aos critérios de avaliação utilizados, foram feitos testes objetivos e subjetivos. Os primeiros consistiram da medição das relações sinal/ruído convencional e segmental. Os segundos consistiram na audição, por parte de várias pessoas, das sentenças obtidas com os codificadores simulados, e a posterior escolha daqueles considerados de melhor qualidade. A ênfase foi dada nos testes subjetivos, já que é a percepção humana a que determina, em última instância, a inteligibilidade e qualidade de um codificador.

Os sistemas simulados foram testados supondo canal ideal e canal ruidoso. Neste último caso utilizaram-se taxas de erros de 10^{-4} , 10^{-3} e 10^{-2} .

Finalmente foram obtidos os melhores sistemas de codificação em sub-bandas trabalhando às taxas de 8, 16 e 24 [Kbits/s]. Adicionalmente foi obtido o codificador de sub-bandas que iguala, subjetivamente, o desempenho do PCM-Lei A de 64 [Kbits/s] e do ADPCM de 32 [Kbits/s].

1.2 CONTEÚDO DO TRABALHO

O conteúdo deste trabalho, por capítulo, é o seguinte.

No Cap.2 é feita uma descrição geral dos codificadores de sub-bandas explicando-se seu princípio de funcionamento e os refinamentos que podem se utilizados neles. São então enumeradas e explicadas as vantagens que apresenta a técnica de codificação em sub-bandas. O capítulo finaliza com o cálculo da taxa de transmissão nestes codificadores.

No Cap.3 são apresentados os problemas decorrentes da utilização de filtros passa-banda em paralelo, para a divisão e reconstrução do espectro em sub-bandas. É desenvolvida então a teoria dos bancos de filtros espelho em quadratura, que eliminam os problemas dos filtros anteriores. São apresentados depois os esquemas de divisão do espectro em 2, 4 e em geral 2^P sub-bandas de igual largura, assim como o esquema de divisão em oitavas. É calculado também, o número de multiplicações e adições necessárias para este fim durante cada período de amostragem T .

O Cap.4 trata da codificação dos sinais a transmitir. Neste capítulo são analisados diferentes tipos de quantizadores e estudadas as características dos sinais de voz em banda integral e em sub-bandas. São explicados então os codificadores escolhidos para codificar os sinais de sub-bandas, justificando-se esta escolha. O capítulo finaliza com diretrizes que auxiliam na alocação de bits entre as diferentes sub-bandas.

O Cap.5 é o capítulo de Testes e Desempenho. Neste capítulo é apresentado o "software" desenvolvido para realizar as simulações e o ambiente de trabalho. São apresentados então os principais codecs simulados e a metodologia utilizada durante os testes e simulações. São detalhados também os critérios de avaliação utilizados e os resultados obtidos.

Finalmente, no Cap.6, são apresentadas as conclusões do trabalho realizado e sugestões para futuros trabalhos na área.

CAPÍTULO 2

CODIFICADORES DE SUB-BANDAS

CONTEÚDO

2.1 - Introdução	5
2.2 - Codificadores de sub-bandas	6
2.2.1 - Estrutura divisora e reconstutora do espectro	10
2.2.2 - Codificação dos sinais a transmitir	11
2.3 - Vantagens da codificação em sub-bandas	13
2.4 - Cálculo da taxa de transmissão, nos codifica- dores de sub-bandas	15
REFERÊNCIAS	17

2.1 INTRODUÇÃO

A codificação de sinais em sub-bandas, oferece várias vantagens sobre os codificadores de banda completa tradicionais (full-band coding). Nos codificadores de sub-bandas o espectro do sinal de entrada é dividido em N sub-bandas, para depois se codificar as bandas resultantes de forma independente uma das outras.

Por este processo de se dividir o espectro do sinal de entrada em N sub-bandas, cada uma delas pode ser codificada com maior, igual ou menor precisão que as restantes, segundo sua influência na inteligibilidade e qualidade final do sinal reconstruído. Na reconstrução do sinal feita no receptor, os sinais das diferentes sub-bandas são decodificados e depois somados, obtendo-se então um sinal em banda integral, que é uma réplica aproximada do sinal de entrada original.

Neste capítulo faz-se uma descrição geral dos codificadores de sub-bandas, incluindo-se alguns possíveis refinamentos factíveis de se utilizar nestes codificadores, e se enumeram e explicam as vantagens que apresenta este tipo de codificação. O capítulo finaliza com o cálculo da Taxa Total de bits nos codificadores de sub-bandas.

A menos que se indique o contrário, a explicação será focalizada em codificadores de sub-bandas aplicados a sinais de voz.

2.2 CODIFICADORES DE SUB-BANDAS

Na Fig.2.1 é mostrado o diagrama em blocos de um codificador de sub-bandas básico. O sinal de entrada $x(t)$, limitado em faixa à frequência $F_s/2$ [Hz], e amostrado à taxa de Nyquist F_s , é submetido a um banco de filtros passa-banda em paralelo, $H_1, H_2, \dots, H_N$, que divide o espectro da sequência de entrada $x(n)$, em N sub-bandas contíguas. Estas sub-bandas podem ter igual largura, como mostra a Fig. 2.2(a), ou larguras diferentes, segundo a Fig.2.2(b).

Nas figuras, $X(e^{j\omega})$ denota o espectro da sequência de entrada $x(n)$, enquanto que $X_1(e^{j\omega}), X_2(e^{j\omega}), \dots, X_N(e^{j\omega})$, denotam os espectros das sequências $x_1(n), x_2(n), \dots, x_N(n)$, à saída do banco de filtros passa-banda $H_1, H_2, \dots, H_N$. Nas mesmas figuras, $\Omega = 2\pi f$ [rad/s] é a frequência utilizada na representação do sinal analógico $x(t)$, enquanto que $\omega = \Omega/F_s$ [rad], é a frequência normalizada utilizada na representação do sinal amostrado $x(n)$. Como a largura de faixa de cada um dos sinais $x_1(n), x_2(n), \dots, x_N(n)$ é menor que π , estes sinais possuem informação redundante (ver Apêndice A), e, em consequência, o próximo passo é reamostrá-los de forma de manter a taxa total de amostras (soma das taxas de cada sub-banda), igual a seu valor inicial de F_s [am/s]. Este reamostragem pode-se realizar de duas formas: na

primeira forma se faz uma translação do espectro de cada sub-banda para a origem do eixo de frequências, através de uma modulação, e depois se reamostram os sinais resultantes a uma taxa igual ao dobro de suas larguras de faixa (Teorema de Nyquist). A segunda forma, mostrada na Fig.2.1, requer que as diferentes sub-bandas satisfaçam a condição "Integer Band Sampling", a ser explicada logo mais. Neste caso é possível reamostrar diretamente as sub-bandas ao dobro de suas larguras de faixa, sem ter que traslada-las previamente à origem do eixo de frequências, dispensando, portanto, o uso de modulação. Em ambos os casos, a reamostragem consiste na *dizimação* dos sinais a reamostrar (blocos "Diz x" na Fig.2.1). Cada um dos espectros das seqüências dizimadas $\tilde{x}_1(n)$, $\tilde{x}_2(n)$, ..., $\tilde{x}_N(n)$, ocupa, agora, todo o eixo ω de $-\pi$ a π . Em seguida estes sinais são codificados individualmente, multiplexados e transmitidos.

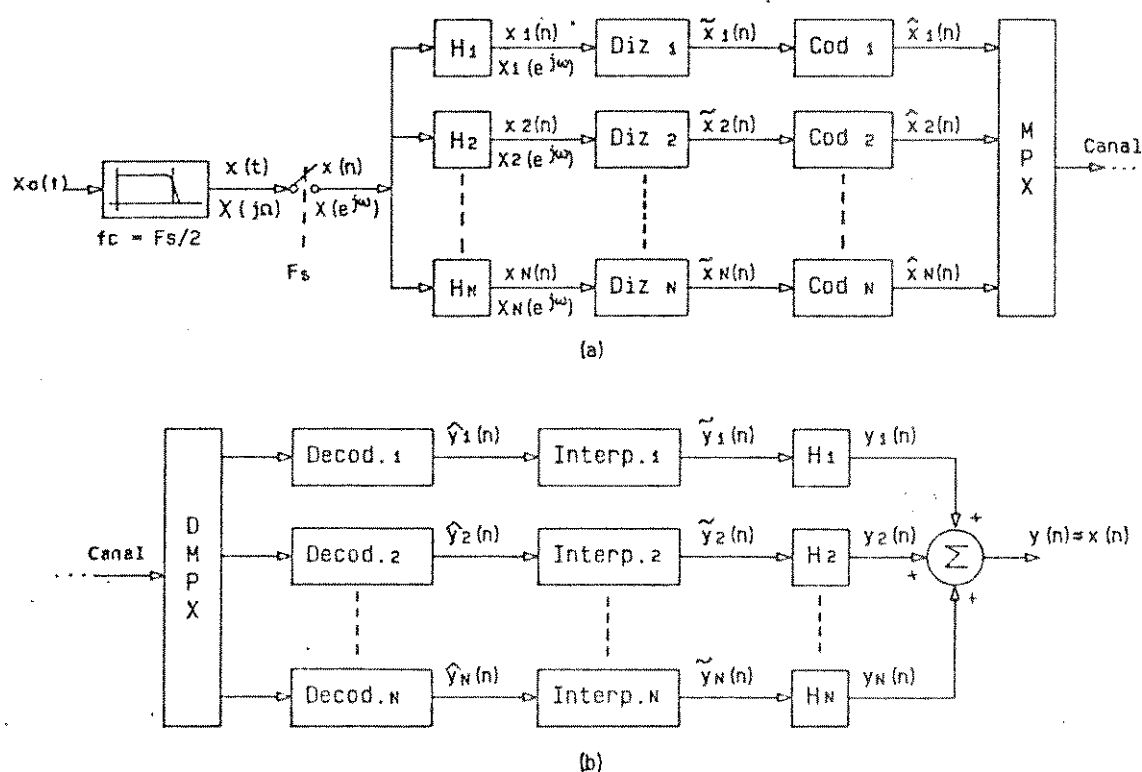


Figura 2.1: Diagrama em blocos de um codificador de sub-bandas:
(a) Transmissor; (b) Receptor.

No receptor os sinais são primeiramente demultiplexados e decodificados, obtendo-se desta forma as seqüências $\hat{y}_1(n)$, $\hat{y}_2(n)$, ..., $\hat{y}_N(n)$. Estes sinais são então interpolados, sendo que a interpolação restaura a taxa de amostragem e a largura e posição de cada espectro, a seu valor original. Por último, os sinais são filtrados por um banco de filtros passa-banda H_1 , H_2 , ..., H_N similar ao utilizado no transmissor, que permite eliminar as imagens espectrais indesejadas geradas durante a interpolação. Os sinais de saída dos filtros são então somados, obtendo-se o sinal $y(n)$, réplica aproximada do sinal de entrada original $x(n)$ (Ref.[1,2,3]).

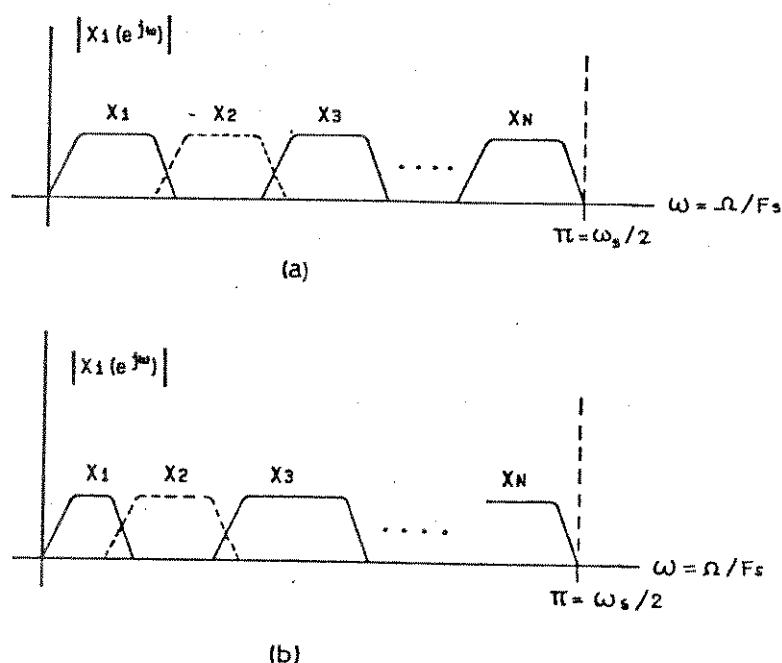


Figura 2.2: Divisão do espectro do sinal de entrada em N sub-bandas:
(a) De igual largura; (b) De largura diferente.

Como foi mencionado anteriormente, para evitar o uso de moduladores antes da dizimação (o que obrigaria a usá-los também após a interpolação), é necessário que a faixa de cada filtro e suas frequências de corte estejam inter-relacionadas segundo a condição "Integer Band Sampling"

(Ref.[1,2,3]). Esta condição exige que os filtros $H_k(e^{j\omega})$ sejam projetados de modo a ter uma faixa $\Delta\omega_k$, uma frequência de corte inferior $m_k \Delta\omega_k$ e uma frequência de corte superior igual a $(m_k + 1) \Delta\omega_k$, onde m_k é um inteiro (tipicamente 0 ou 1 para as bandas baixas e 2 ou 3 para as bandas altas). Adicionalmente a largura de faixa $\Delta\omega_k$ deve ser um sub-múltiplo de π , de modo a permitir a dizimação/interpolação das amostras por um fator inteiro. Em todas as simulações realizadas ao longo deste trabalho, os filtros foram desenvolvidos satisfazendo a condição acima.

Na Fig.2.3 é mostrado, como exemplo, o espectro do sinal $x_2(n)$ ao longo dos diferentes estágios do transmissor (Tx), e do receptor (Rx), quando se aplica a condição de "Integer Band Sampling".

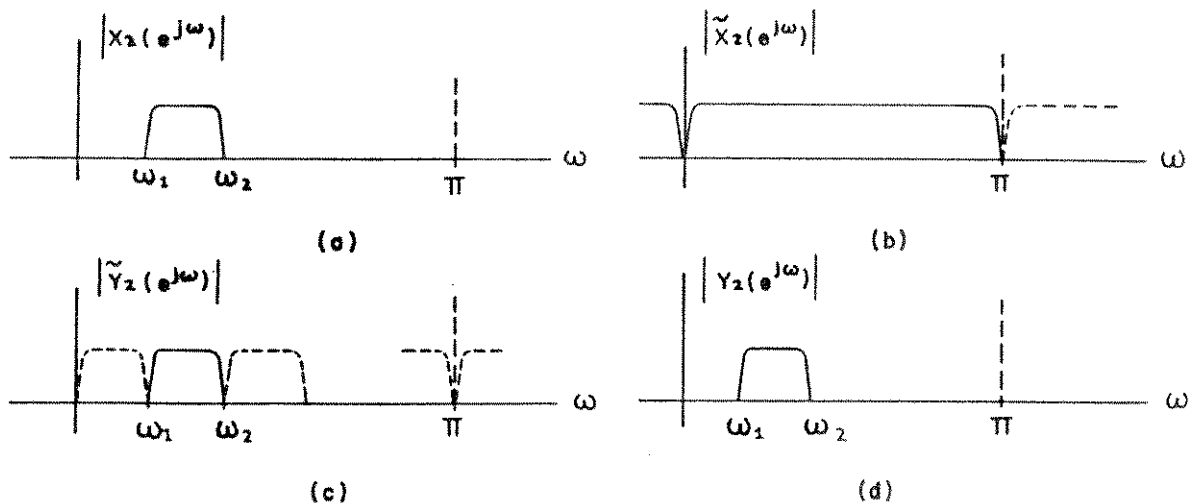


Figura 2.3: Espectro do sinal $x_2(n)$ ao longo dos diferentes estágios do Tx e Rx: (a) Após a filtragem H_2 do Tx; (b) Após a dizimação; (c) Após a interpolação; (d) Após a filtragem H_2 do Rx.

Daqui para frente, e para facilidade de apresentação, os sistemas de codificação de sub-bandas serão divididos em dois sub-sistemas: o primeiro sub-sistema abrange a "estrutura

filtrante" que divide o espectro do sinal de entrada em N sub-bandas no transmissor, e realiza a síntese (reconstrução), do sinal de saída no receptor. Incluem-se neste sub-sistema as operações de dizimação e interpolação dos sinais. O segundo sub-sistema abrange a "codificação e decodificação" dos sinais das diferentes sub-bandas, assim como a de qualquer informação lateral que deva ser transmitida. A seguir são apresentados estes 2 sub-sistemas.

2.2.1 ESTRUTURA DIVISORA E RECONSTRUTORA DO ESPECTRO

Quanto a este primeiro sub-sistema, a estrutura divisora do espectro pode consistir de um banco de filtros passa-banda em paralelo como discutido até agora (Ref. [1,2,3]), de um banco de filtros espelho em quadratura (QMF Bank, Ref. [4,5,6]), ou de um banco de filtros espelho em quadratura generalizado (GQMF Bank, Ref. [7]). O mesmo é válido para a estrutura que reconstrói o espectro no receptor.

Em caso de se utilizar um banco de filtros passa-banda em paralelo, o número de sub-bandas em que se divide o espectro, assim como a largura e localização de cada sub-banda, é arbitrário. Deve-se considerar, porém, que em caso de se utilizar a técnica Integer Band Sampling, existem restrições entre as larguras das bandas e suas posições (ver seção anterior). Normalmente, e por razões práticas, o número de sub-bandas é limitado a 4 ou 5.

Em caso de se utilizar um banco de filtros espelho em quadratura, o espectro pode ser dividido em 2^p sub-bandas de igual largura (p inteiro), ou em oitavas (neste caso a largura de uma sub-banda é o dobro da largura da sub-banda anterior). Normalmente os melhores desempenhos dos codificadores de sub-bandas utilizando bancos de filtros QMF, são obtidos para $2^p = 8$ ou 16 sub-bandas.

Por último, o banco de filtros espelho em quadratura

generalizado divide o espectro num número arbitrário de sub-bandas de igual largura e igualmente espaçadas. Este tipo de filtros não será tratado neste trabalho, enquanto que os dois primeiros serão objeto de um estudo detalhado no Cap.3.

2.2.2 CODIFICAÇÃO DOS SINAIS A TRANSMITIR

Os codificadores Cod 1, Cod 2, ..., Cod N, utilizados na codificação dos sinais das diferentes sub-bandas, utilizam em geral quantizadores adaptativos para melhor ajustar a variância de curto prazo dos sinais de entrada $x_1(n)$, $x_2(n)$, ..., $x_N(n)$, à faixa de amplitudes do quantizador. Esta adaptação pode ser feita amostra por amostra, como ocorre por exemplo nos codificadores ADM (Adaptive Delta Modulation), APCM (Adaptive PCM), ADPCM (Adaptive Differential PCM), ou a cada bloco de P amostras, como acontece no BC-PCM (Block Companded PCM). Além do mais, a adaptação pode basear-se nos valores das amostras seguintes a serem quantizadas (AQF: Adaptive Quantization Forward), ou nos valores das amostras já quantizadas (AQB: Adaptive Quantization Backward), cada uma com suas vantagens e desvantagens.

Os quantizadores podem ser i) *uniformes*, ii) *não uniformes*, ou iii) *ótimos uniformes ou não uniformes*. Deve resultar uma melhora na qualidade, caso sejam utilizados quantizadores ótimos projetados para sinais de entrada Gaussianos, uma vez que a função "Densidade de Probabilidade de Curto Prazo" dos sinais de voz, é corretamente descrita por uma função gaussiana (Ref.[8]).

Como será visto na Seção 2.3, uma das vantagens dos codificadores de sub-bandas é a de poder atribuir diferentes números de bits a cada sub-banda, com o intuito de codificar com mais precisão algumas delas. Esta atribuição de bits a cada sub-banda pode ser feita de maneira fixa ou dinâmica.

No caso de atribuição fixa de bits, é feito um estudo daquelas bandas mais importantes para a compreensão da fala e, como resultado desse estudo e de testes objetivos e subjetivos, é feita a atribuição de bits. Esta atribuição permanece inalterada, independentemente da potência de curto prazo de cada sub-banda.

No caso da atribuição dinâmica de bits a cada sub-banda, esta atribuição é baseada na potência de curto prazo das mesmas (Ref.[9]). Para isto se faz uma estimativa da potência de curto prazo de cada sub-banda, em função do valor das próximas amostras a quantizar (quando se faz atribuição dinâmica de bits, as amostras são quantizadas por blocos). A banda com maior potência recebe o primeiro bit e a sua estimativa de potência é reduzida por um fator r (normalmente $r=2$). É feita então uma nova pesquisa entre todas as bandas, recebendo o próximo bit aquela de maior potência e reduzindo posteriormente sua estimativa de potência de um fator r . Assim o procedimento se repete sucessivamente, até acabar com os bits disponíveis. Após esta atribuição, o bloco de amostras que permitiu a pesquisa é quantizado e o processo continua com um novo bloco de amostras. Os blocos normalmente tem uma duração entre 10 e 20 [ms], já que para tal intervalo de tempo os sinais de voz podem ser considerados aproximadamente estacionários. Além disto, em geral é fixado o número mínimo e máximo de bits permitido a cada sub-banda, baseando-se nos critérios mencionados quando da atribuição fixa de bits. O fato de se fazer uma atribuição dinâmica de bits às diferentes sub-bandas, implica na necessidade de se transmitir ao receptor, além das amostras quantizadas, uma informação lateral especificando o número de bits atribuído a cada sub-banda. Isto faz aumentar a taxa final de bits do codificador, além de torná-lo bem mais complexo, mas pode melhorar seu desempenho. Como alternativa à utilização da potência de curto prazo para definir a distribuição dos bits, pode-se utilizar o desvio padrão das diversas sub-bandas, ou algum outro parâmetro que esteja relacionado com a dis-

tribuição de energia entre as mesmas (Ref.[10]).

Finalmente deve-se mencionar o fato de que existem codificadores que precisam transmitir algum tipo de informação lateral sobre o quantizador. Assim, por exemplo, os codificadores BC-PCM precisam transmitir o parâmetro de adaptação do quantizador a cada bloco de P amostras, para poder adaptar dinamicamente a sua faixa de amplitudes à do bloco de amostras em questão. A necessidade de se transmitir esta informação faz com que parte dos bits seja destinada a esta finalidade, restando menos bits para a quantização das amostras. Em baixas taxas este problema é importante, devendo-se alcançar um compromisso entre os bits destinados a transmitir uma e outra informação, de forma que a qualidade final do sinal processado seja a melhor possível.

2.3 VANTAGENS DA CODIFICAÇÃO EM SUB-BANDAS

Como foi visto na Seção 2.2, nos codificadores de sub-bandas o espectro do sinal de entrada é dividido em N sub-bandas; cada uma das bandas é então codificada de forma independente das outras e os sinais resultantes são multiplexados e transmitidos.

As vantagens apresentadas por esta técnica de codificação são discutidas a seguir:

- 1) Quando o sinal de voz é dividido em sub-bandas, é possível se atribuir diferentes números de bits a cada uma delas segundo a sua importância na compreensão da fala, e com isto melhorar o desempenho do sistema. Assim as bandas baixas (até 1,2 [KHz] aproximadamente), onde está presente a informação de pitch e de formantes e a maior parte da energia dos sons vocálicos, que são os mais importantes para a compreensão da fala, são normalmente codificadas

com mais bits que as bandas altas. Nestas últimas se concentra a maior parte da energia dos sons fricativos, menos importantes para a inteligibilidade de um sinal de voz. Em resumo: as regiões mais importantes do espectro, são codificadas com maior precisão. Para se atingir taxas menores, pode-se também descartar determinadas bandas do espectro que tem pouca influência na inteligibilidade do sinal.

- 2) O fato de se poder atribuir números diferentes de bits entre as sub-bandas, faz com que se possa diminuir a taxa final de bits do codificador, conforme será explicado no final do capítulo.
- 3) O ruído de quantização total gerado é menor que numa codificação de banda completa. O motivo é que, nos codificadores de banda completa, é utilizado o mesmo passo de quantização para todo o espectro. Esse passo é escolhido em função da energia do sinal (soma das energias de todas as componentes espectrais). Se esta energia é grande, o passo de quantização também será grande e produzirá um erro de quantização grande nas regiões do espectro onde o sinal apresente baixa energia. No entanto, nos codificadores de sub-bandas é possível escolher um passo de quantização diferente para cada sub-banda, função da energia que ela possua. Assim as bandas de baixa energia poderão vir a utilizar passos de quantização menores, gerando menor ruído e contribuindo menos ao ruído de quantização total.
- 4) Em cada sub-banda o sinal útil tende a mascarar o próprio ruído de quantização (aquele gerado nessa sub-banda), tornando-o menos perceptível ao ouvido humano. Num codificador de banda completa, por outro lado, pode ocorrer de se ter um ruído de quantização elevado em regiões do espectro em que o sinal apresenta baixa energia, mascarando assim o sinal útil.

- 5) Dado que o ruído de quantização fica restrito à banda que o gerou, evita-se que ele mascare um sinal útil de baixa energia pertencente a uma outra banda de frequências.
- 6) Como cada sub-banda é codificada independentemente das outras, é possível utilizar-se codificadores diferentes em cada sub-banda de modo a melhorar o desempenho global do sistema. Assim por exemplo, poderia se utilizar, se desejado, Block-Companded PCM numa banda, APCM em outra, ADPCM em outra, ADM em outra, etc.

Como resultado de tudo isto, o ruído de quantização nos codificadores de sub-bandas é menos perceptível que nos codificadores de forma de onda tradicionais operando à mesma taxa. Isto faz que a qualidade final dos sinais processados com os codificadores de sub-bandas, seja superior à dos codificadores tradicionais.

2.4 CÁLCULO DA TAXA DE TRANSMISSÃO NOS CODIFICADORES DE SUB-BANDAS

A taxa de transmissão R , em [bits/s], num codificador de sub-bandas, é dada pela soma das taxas de bits individuais de cada sub-banda, mais a taxa da informação lateral, se ela existir, isto é:

$$\begin{aligned}
 R \text{ [bits/s]} &= \sum_{k=1}^N R_k \text{ [bits/s]} + R_L \text{ [bits/s]} \\
 &= \sum_{k=1}^N F_{sk} \text{ [am/s]} \cdot n_k \text{ [bits/am]} + R_L \text{ [bits/s]} \quad (2.1)
 \end{aligned}$$

Nesta equação, R_k representa a taxa de bits da k -ésima sub-banda, F_{sk} representa a taxa de amostras dessa sub-banda,

n_k está indicando a quantidade de bits com que se está codificando cada amostra da sub-banda k , e N representa o número de sub-bandas em que foi dividido o espectro. Logicamente que só as sub-bandas transmitidas intervêm nesta equação. R_L representa a taxa de informação lateral em [bits/s], que poderá ou não ter um valor nulo, segundo o tipo de codificador utilizado.

Considerando-se que a taxa de amostras da sub-banda k é dada pela frequência com que ela é reamostrada (dizimada), na saída do filtro passa-banda H_k , tem-se:

$$F_{sk} \text{ [am/s]} = 2 \times \Delta F_k \text{ [Hz]} \quad (2.2)$$

onde ΔF_k é a largura da sub-banda k medida em Hz (as sub-bandas são reamostradas ao dobro de suas larguras de faixa (taxa de Nyquist)). Substituindo-se este resultado na Eq. (2.1) resulta:

$$R \text{ [bits/s]} = 2 \sum_{k=1}^N \Delta F_k \cdot n_k + R_L \quad (2.3)$$

No Cap. 4, Seção 4.4.2, é apresentada a forma de se calcular a taxa de informação lateral R_L , quando ela existe.

importante se destacar que a taxa total de amostras num codificador de sub-bandas é a mesma que num codificador de banda completa, quando todas as sub-bandas são codificadas e transmitidas. Isto pode ser visto da seguinte forma: considere-se que o sinal a ser transmitido, $x(n)$, possua uma taxa de amostras de F_s [am/s] (F_s é a frequência de amostragem do sinal $x(t)$). Se o sinal $x(n)$ é dividido em 2 sub-bandas de igual largura, então os sinais pertencentes a cada sub-banda possuirão também uma taxa inicial de F_s [am/s]. Após a dizimação destes sinais, a taxa por sub-banda cai pela metade ($F_s/2$ [am/s]), mas a taxa total, que é a soma das taxas de cada sub-banda, permanece a mesma.

O raciocínio pode ser estendido à divisão em N sub-bandas, sendo que após a divisão do espectro e a dizimação dos sinais obtidos, a soma das taxas de amostras de todas as sub-bandas resulta no valor inicial de F_s [am/s].

No caso de se transmitir todas as sub-bandas, não resulta nenhuma economia na taxa de amostras do codificador. A economia na taxa final de bits é obtida destinando-se menos bits por amostra para codificar os sinais das diferentes sub-bandas. Como será visto no Cap.5, bastam 5 ou 6 [bits/am], no máximo, para se atingir desempenhos iguais e/ou superiores a um PCM de 64 [Kbits/s] (8 [bits/am]). Isto já representa uma taxa de bits menor que 64 [Kbits/s]. Além do mais, este número de bits (5 ou 6 [bits/am]), é o destinado às bandas principais do espectro. Levando-se em conta que as bandas menos importantes são codificadas com menos bits, e que algumas delas podem não ser transmitidas, se produzirá ainda uma diminuição adicional na taxa final de bits do codificador.

REFERÊNCIAS

- [1] R.E.Crochiere, S.A.Webber and J.L.Flanagan, "Digital Coding of Speech in Sub-bands", Bell System Tech. J., Vol 55, No 8, October 1976, pp 1069-1085.
- [2] R.E.Crochiere, "On the Design of Sub-band Coders for Low-Bit-Rate Speech Communication", Bell System Tech. J., Vol 56, No 5, May-June 1977, pp 747-770.
- [3] R.E.Crochiere and J.L.Flanagan, "Sub-band Encoding of Speech", International Conference on Comm., 1977, Vol 1, pp 293-296.

- [4] D.Esteban and C.Galand, "Application of Quadrature Mirror Filters to Split Band Voice Coding Schemes", Proc. ICASSP, May 1977, pp 191-195.
- [5] C.Galand and H.Nussbaumer, "New Quadrature Mirror Filter Structures", IEEE Trans. on ASSP, Vol ASSP-32, Nº 3, June 1984, pp 522-531.
- [6] A.Crosier, D.Esteban and C.Galand, "Perfect Channel Splitting by Use of Interpolation/Decimation/Tree Decomposition Techniques", Int'l Conf. on Inform. Sciences and Systems, Patras, Greece, 1976, pp 443-446.
- [7] R.Cox et al, "New Directions in Sub-band Coding", IEEE Journal on Selected Areas in Comm., Vol 6, Nº 2, February 1988, pp 391-409.
- [8] N.S.Jayant and P.Noll, "Digital Coding of Waveforms - Principles and Applications to Speech and Video", Prentice Hall, Inc., Englewood Cliffs, N.J., 1984.
- [9] T.Ramstad, "Sub-band Coder With a Simple Adaptive Bit Allocation Algorithm, A Possible Candidate for Digital Mobile Telephony?", Conf. Rec. ICASSP, France 1982, pp 203-207.
- [10] C.F.Jaccoud, "Estudo de codificação em Sub-bandas a 16 Kb/s. Dissertação de Mestrado", Janeiro 1987, PUC, Rio de Janeiro. Cap.4.

CAPÍTULO 3

ESTRUTURA FILTRANTE NOS CODIFICADORES DE SUB-BANDAS

CONTEÚDO

3.1 - Introdução	19
3.2 - Banco de filtros passa-banda em paralelo	20
3.2.1 - Divisão do espectro utilizando filtros passa-banda em paralelo	20
3.2.2 - Considerações sobre o projeto de filtros passa-banda em paralelo	24
3.2.2.1 - Divisão do espectro em sub- bandas de igual largura	24
3.2.2.2 - Divisão do espectro em sub- bandas de largura diferente	25
3.2.3 - Distorção no banco de filtros passa-banda	30
3.2.3.1 - Distorção na dizimação	31
3.2.3.2 - Distorção na interpolação	34
3.3 - Banco de filtros espelho em quadratura (QMF Bank).	36
3.3.1 - Estudo do esquema básico de um banco QMF	36
3.3.2 - Implementação e cálculo do banco QMF	46
3.3.3 - Divisão do espectro em 2^P sub-bandas de igual largura, utilizando-se filtros QMF	52
3.3.4 - Divisão do espectro em oitavas, utilizando- se filtros QMF	58
REFERÊNCIAS	61

3.1 INTRODUÇÃO

A utilização de bancos de filtros passa-banda em paralelo nos codificadores de sub-bandas, apresenta vários inconvenientes. Entre os mais sérios estão os causados pela

superposição dos espectros individuais dos filtros, devido à largura finita das bandas de transição dos mesmos. Esta superposição faz com que durante a dizimação dos sinais no transmissor, se produza "aliasing". Pela mesma causa e após a interpolação dos sinais no receptor, as imagens do espectro principal geradas durante a interpolação não são perfeitamente eliminadas. Estas duas distorções produzem degradação perceptível no sinal de voz.

Para minimizar estes problemas os filtros devem aproximar-se o mais possível dos ideais, o que torna extremamente rigorosas as especificações de projeto dos mesmos. A solução a estes inconvenientes consiste na utilização de filtros espelho em quadratura, que permitem uma perfeita reconstrução do sinal original a partir das N sub-bandas, independentemente da ordem dos filtros utilizados.

Neste capítulo serão discutidos primeiro os bancos de filtros passa-banda em paralelo, para depois se analisar em detalhe o funcionamento dos bancos de filtros espelho em quadratura. Serão apresentadas também as estruturas que permitem a divisão do espectro em 2, 4 e em geral 2^p sub-bandas iguais, assim como a que permite a divisão do espectro em oitavas.

32 BANCO DE FILTROS PASSA-BANDA EM PARALELO

3.2.1 DIVISÃO DO ESPECTRO UTILIZANDO FILTROS PASSA-BANDA EM PARALELO

Como já foi visto no Cap.2, Seção 2.2, o sinal de voz à entrada do codificador de sub-bandas é limitado em faixa a

$F_s/2$ [Hz], gerando o sinal $x(t)$. O sinal $x(t)$ é então amostrado à taxa de Nyquist F_s [Hz], obtendo-se a sequência $x(n)$.

A transformada de Fourier do sinal analógico $x(t)$ é dada por:

$$X(j\Omega) = \int_{-\infty}^{\infty} x(t) \cdot e^{-j\Omega t} \cdot dt \quad (3.1)$$

onde $\Omega = 2\pi \cdot f$ [rad/s].

Por outro lado, a transformada de Fourier do sinal amostrado, $x(n) = x(t)|_{t=nT}$, $n=0, \pm 1, \pm 2, \dots$, onde T é o período de amostragem, é dada por:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n) \cdot e^{-j\omega n} \quad (3.2)$$

onde $\omega = \Omega/F_s$ [rad], corresponde à frequência normalizada utilizada na representação do sinal $x(n)$. Sabe-se do Apêndice A, que $X(e^{j\omega})$ é uma função contínua de ω e periódica em ω com período 2π . Além do mais, a relação entre os espectros de $x(t)$ e de $x(n)$, dada pelo teorema de amostragem (Ref.[1,2]), é a seguinte:

$$X(e^{j\omega}) = \frac{1}{T} \sum_{r=-\infty}^{\infty} X(j\Omega + j\frac{2\pi r}{T}) \Big|_{\Omega=\omega/T} \quad (3.3)$$

onde $T = 1/F_s$.

Assim, se o espectro do sinal analógico $X(j\Omega)$ é como na Fig.3.1(a), então o espectro do sinal amostrado $X(e^{j\omega})$, será como na Fig.3.1(b). A periodicidade de $X(e^{j\omega})$, como decorre da Eq.(3.3), pode ser apreciada na Fig.3.1(b).

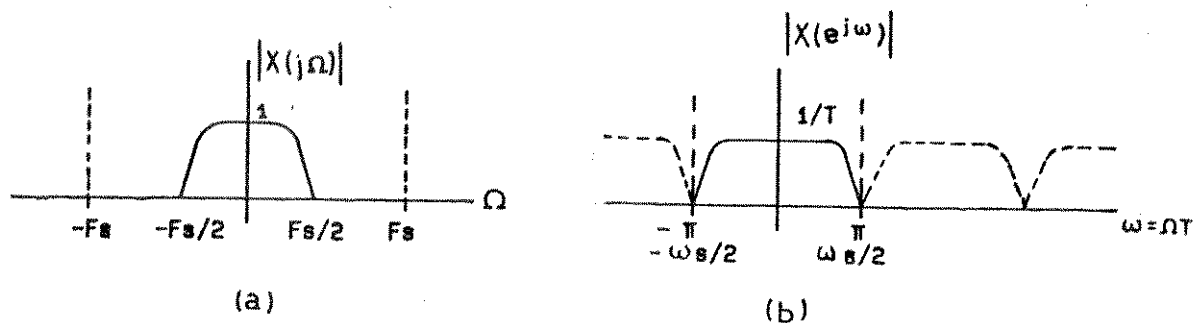


Figura 3.1: (a) Espectro do sinal $x(t)$; (b) Espectro do sinal $x(n)$.

Da Fig.3.1 pode-se ver que:

$$X(e^{j\omega}) = \frac{1}{T} X(j\Omega), \quad \text{para} \quad -\pi \leq \omega \leq \pi \quad (3.4)$$

Como indica a Fig.3.2, a sequência de entrada $x(n)$ é submetida ao banco de filtros passa-banda $H_1, H_2, \dots, H_N$, que divide o espectro $X(e^{j\omega})$ do sinal de entrada, em N sub-bandas contíguas: $X_1(e^{j\omega}), X_2(e^{j\omega}), \dots, X_N(e^{j\omega})$.

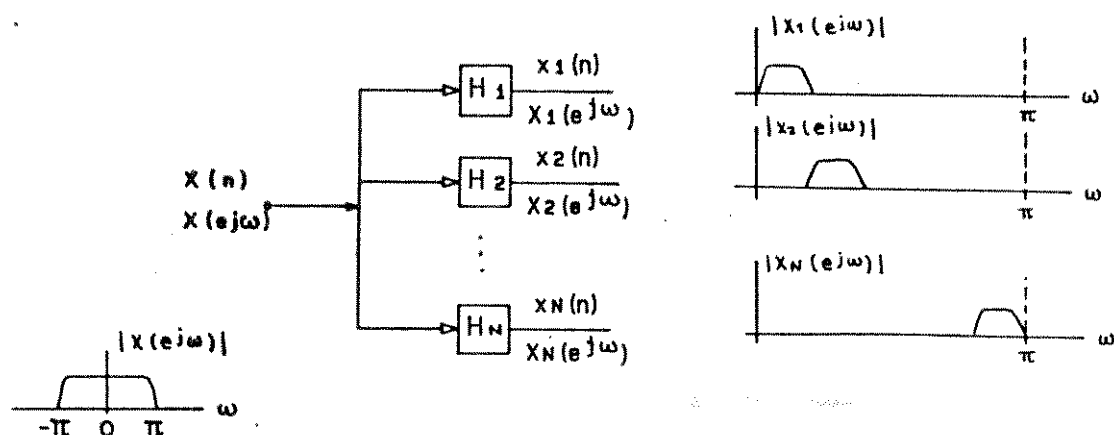


Figura 3.2: Divisão do espectro de entrada $X(e^{j\omega})$ em N sub-bandas.

Estas sub-bandas podem ser de igual largura, Fig.3.3(a), ou de largura diferente, Fig.3.3(b). Adicionalmente os espectros podem apresentar superposição nas regiões de transição, como indicam as Fig.3.3(a) e (b), ou podem ter "gaps" entre algumas sub-bandas como indica a Fig.3.3(c).

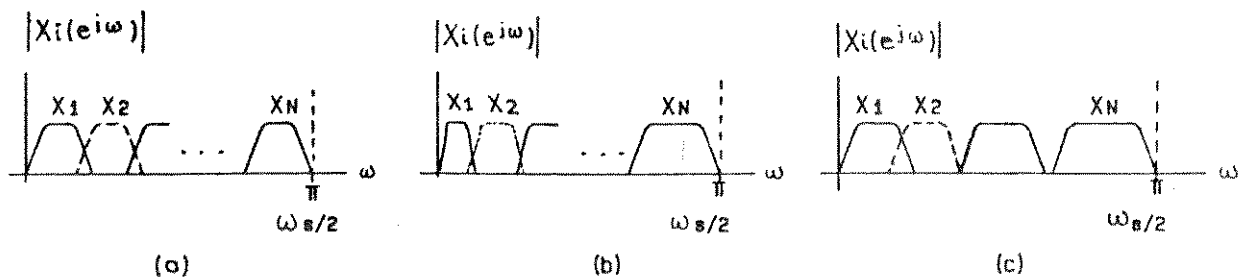


Figura 3.3: Características dos espectros das sub-bandas: (a) De igual largura e com superposição na região de transição; (b) De largura diferente e com superposição na região de transição; (c) De largura diferente e com "gaps" entre algumas sub-bandas.

No caso de sub-bandas de igual largura cobrindo todo o espectro, tem-se que a largura de cada sub-banda será:

$$\Delta\omega_k = \Delta\omega = \frac{\omega_s/2}{N} = \pi/N, \quad k = 1, 2, \dots, N. \quad (3.5)$$

onde $\Delta\omega_k$ é a largura da k -ésima sub-banda, N é o número de sub-bandas em que se divide o espectro $X(e^{j\omega})$, e $\omega_s = \Omega_s/F_s = 2\pi \cdot F_s/F_s = 2\pi$.

No caso de sub-bandas de largura diferente, é comum se utilizar uma largura maior à medida que se incrementa a frequência, por razões que serão explicadas na Seção 3.2.2. Neste caso:

$$\Delta\omega_{k+1} > \Delta\omega_k, \quad k = 1, 2, \dots, N-1. \quad (3.6)$$

3.2.2 CONSIDERAÇÕES SOBRE O PROJETO DE BANCOS DE FILTROS PASSA-BANDA EM PARALELO

A primeira questão que surge no projeto do banco de filtros passa-banda em paralelo, é o número de sub-bandas em que se dividirá o espectro. Através de simulações tem-se observado que um bom compromisso entre a qualidade final do sinal processado e a complexidade de implementação dos filtros, é obtido com $N = 4$ ou 5 sub-bandas (Ref.[3,4]). Ao se utilizar menos que 4 sub-bandas, as vantagens da técnica SBC se perdem. Por outro lado, com mais de 5 sub-bandas se perdem componentes espectrais na banda de transição dos filtros, além de se tornar mais complexa a sua implementação prática. Esta limitação no número de bandas é um dos problemas dos bancos de filtros passa-banda em paralelo, já que o ideal seria dividir o espectro em um número de sub-bandas maior.

A segunda questão que surge no projeto destes filtros, é a definição da largura e posição das diferentes sub-bandas dentro do espectro e a presença ou não de "gaps" entre algumas bandas. Na definição da largura e posição de cada banda, é desejável que se satisfaça a condição "Integer Band Sampling", apresentada na Seção 2.2 do Cap.2, para evitar o uso adicional de modulação antes da dizimação (translação passa-baixas até a origem), e depois da interpolação (translação passa-banda até a posição original da banda correspondente). Será feita, a seguir, uma análise separada para o caso de sub-bandas de igual largura e sub-bandas de larguras diferentes.

3.2.2.1 Divisão do espectro em sub-bandas de igual largura

Neste caso, a posição de cada sub-banda já fica estabelecida ao se escolher o número de sub-bandas N . Assim,

da Eq. (3.5) tem-se que, como a largura de cada sub-banda é π/N , a posição de cada uma delas será dada por:

$$\omega_k = \Delta\omega \cdot (k - 1/2) = \frac{\pi}{N} (k - 1/2) \quad ; \quad k = 1, 2, \dots, N. \quad (3.7)$$

onde ω_k corresponde à frequência central de cada sub-banda, segundo indica a Fig. 3.4.

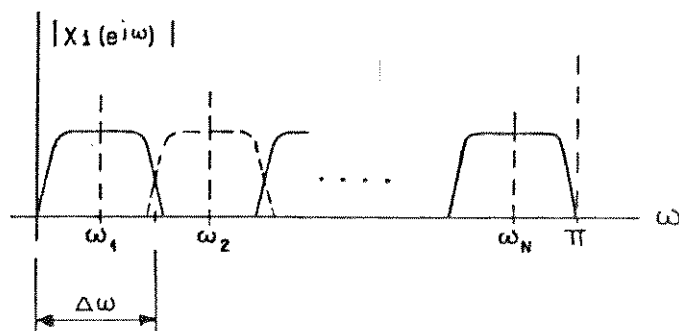


Figura 3.4: Divisão do espectro $X(e^{j\omega})$ em N sub-bandas de igual largura.

3.2.2.2 Divisão do espectro em sub-bandas de largura diferente

Para o caso de se dividir o espectro em sub-bandas de largura diferente, o problema da posição e largura de cada uma delas é bem mais complexo, intervindo vários fatores distintos como: taxa final que se pretende atingir, presença ou não de "gaps" entre sub-bandas, facilidade de implementação em "hardware", etc.

Um possível critério na eleição da largura e posição das diferentes sub-bandas, é fazer com que cada uma delas apresente aproximadamente igual contribuição ao "Índice de Articulação IA", na ausência de ruído. O índice de Articulação é um índice ligado à "inteligibilidade" de

palavras, sílabas e sentenças. É importante destacar que se está falando de "inteligibilidade" e não de "qualidade", já que um codificador pode ser inteligível mas ter uma qualidade ruim.

A Fig.3.5 reproduz os resultados de um estudo feito por Kryter em 1970 (Ref.[5]), onde são mostrados diversos tipos de testes de inteligibilidade, incluindo-se o de palavras foneticamente balanceadas (FB), sílabas e sentenças, em função do índice de Articulação. Assim um $IA = 1,0$ implica numa inteligibilidade de 100%, ao passo que um $IA = 0,8$ implica numa inteligibilidade superior a 90%.

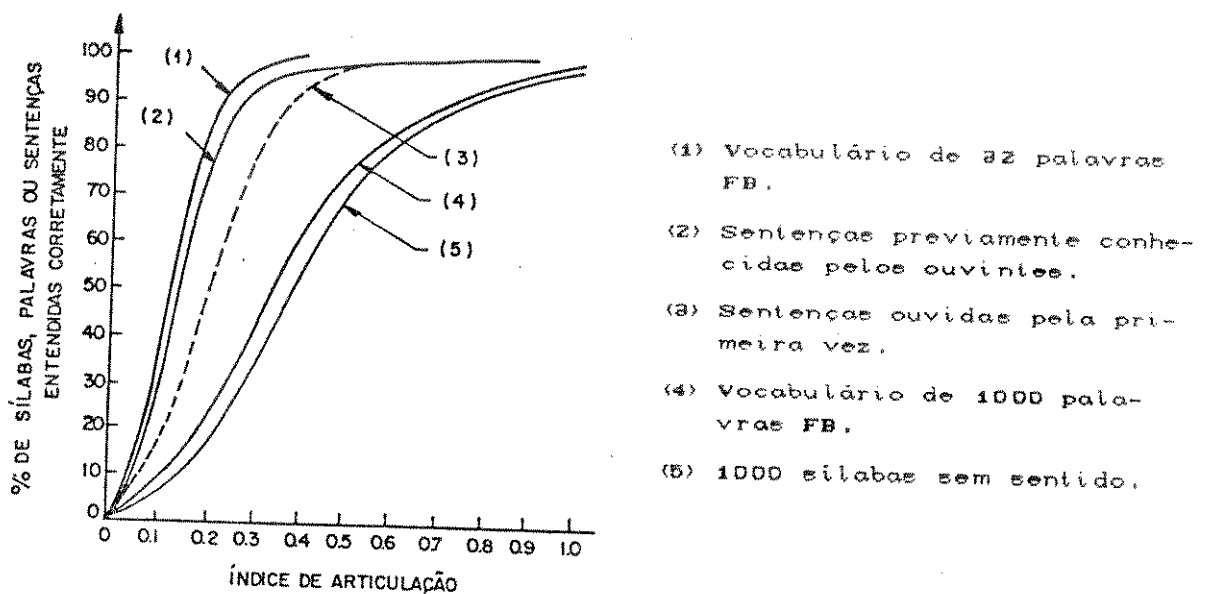


Figura 3.5: Inteligibilidade da voz versus Índice de Articulação (Kryter, 1970).

O conceito de índice de Articulação é baseado numa divisão não uniforme do espectro de frequências, até a frequência máxima de 6100 [Hz]. São obtidas assim 20 sub-bandas contíguas, cada uma contribuindo com 5% ao índice de Articulação total. A Tabela 3.1 mostra esta divisão do eixo de frequências em 20 sub-bandas.

Banda	Faixa [Hz]	Banda	Faixa [Hz]
1	200 a 330	11	1660 a 1830
2	330 a 430	12	1830 a 2020
3	430 a 560	13	2020 a 2240
4	560 a 700	14	2240 a 2500
5	700 a 840	15	2500 a 2820
6	840 a 1000	16	2820 a 3200
7	1000 a 1150	17	3200 a 3650
8	1150 a 1310	18	3650 a 4250
9	1310 a 1480	19	4250 a 5050
10	1480 a 1660	20	5050 a 6100

Tabela 3.1: Bandas de frequência com igual contribuição ao índice de Articulação.

Dado que para telefonia o espectro é limitado em 3400 [Hz], só as primeiras 16 sub-bandas são levadas em conta, o que implica num índice de Articulação de 0,80.

Baseando-se então no IA, o banco de filtros passa-banda em paralelo pode ser projetado para dividir o espectro em 4 ou 5 sub-bandas, cada uma contribuindo com a mesma percentagem ao índice de Articulação total. Assim as bandas mais baixas terão larguras menores, tornando-se progressivamente maiores à medida que a frequência aumenta, conforme indicado na Tabela 3.1 e na Eq.(3.6).

Um exemplo de divisão do espectro em 4 sub-bandas, onde cada uma delas contribui com 20% ao índice de Articulação total de 0,8, é dado na Tabela 3.2. A Fig.3.6 ilustra esta divisão, onde o sinal original é pré-filtrado na faixa de 200 a 3200 [Hz] antes de dividir seu espectro em 4 sub-bandas (Ref.[3]).

Banda	Faixa de frequências [Hz]		
1	200	a	700
2	700	a	1310
3	1310	a	2020
4	2020	a	3200

Tabela 3.2: Divisão do espectro em 4 sub-bandas, com 20% de contribuição de cada uma ao IA total de 0,8.

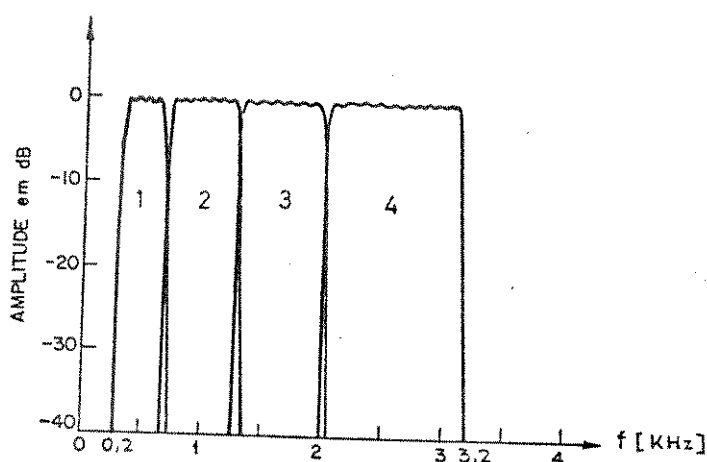


Figura 3.5: Divisão do espectro em 4 sub-bandas contíguas com igual contribuição ao IA total de 0,8.

Deve-se ressaltar aqui que esta divisão em sub-bandas não satisfaz a condição "Integer Band Sampling" e, portanto, a implementação do codificador de sub-bandas com os filtros acima exige um sistema de maior complexidade (necessidade adicional de modulação antes da dizimação e após a interpolação).

Um outro fator que intervém na escolha da largura e posição das diferentes sub-bandas, é a decisão de se

permitir, ou não, "gaps" entre algumas sub-bandas. A presença de "gaps" fará diminuir a largura das bandas envolvidas e/ou variará a posição das mesmas. Tanto neste caso, como naquele de sub-bandas de igual largura, a presença ou não de "gaps" será função da taxa final de bits que se pretende atingir. O valor desta taxa R , é dado pela Eq.(2.3) do Cap.2.

Para valores de R menores que 15 [Kbits/s], são permitidos alguns "gaps" entre as sub-bandas superiores, de modo a diminuir a largura das mesmas e com isso diminuir a taxa final de bits. Entre as bandas baixas não são permitidos "gaps", já que a perda de informação nestas bandas afeta notavelmente a qualidade final do sinal processado. Isto ocorre porque o ouvido humano é mais sensível a distorções nas bandas baixas que nas bandas altas, sendo que a divisão entre bandas baixas e altas pode estabelecer-se na faixa de 1200 a 1500 [Hz].

O efeito de um "gap" no espectro, é o de provocar reverberância (eco), no sinal reconstruído. A percepção deste efeito é função da posição e largura do "gap", sendo mais perceptível quanto mais embaixo se localize no espectro e de maior largura seja. Assim, por exemplo, um "gap" na faixa dos 2900 a 2950 [Hz], quase não é percebido em comparação com um na faixa dos 550 a 600 [Hz].

Quando são necessárias taxas ainda mais baixas (menores que 7 [Kbits/s]), é necessário reduzir-se adicionalmente a largura total do espectro e/ou o número de bits destinados a codificar as amostras. A largura total do espectro pode ser reduzida permitindo-se "gaps" de maior largura e/ou variando-se dinamicamente as frequências centrais das bandas superiores, de forma a posicionar ditas bandas nas regiões de maior energia do espectro de curto prazo da voz (Ref.[6]). Isto implica posicionar as bandas superiores ao redor das

formantes F2 e F3 ¹. Desta forma só as regiões mais significativas do espectro de voz são codificadas e transmitidas. Sob o ponto de vista de inteligibilidade, este esquema é superior a um esquema de sub-bandas fixas com a mesma largura total de espectro abrangido.

3.2.3 DISTORÇÃO NO BANCO DE FILTROS PASSA-BANDA

No transmissor dos codificadores de sub-bandas, os sinais de saída dos filtros passa-banda $H_1(e^{j\omega})$, $H_2(e^{j\omega})$, ..., $H_N(e^{j\omega})$, são reamostrados (dizimados) à taxa de Nyquist, isto é, ao dobro da largura de faixa de cada sub-banda, de forma a se evitar transmitir sinais superamostrados (Cap.2, Seção 2.2). No receptor, no entanto, estes sinais são interpolados para retorná-los à taxa de amostragem original e filtrados por um banco de filtros passa-banda idêntico ao utilizado no transmissor, de forma a poder reconstruir o sinal original a partir de suas N sub-bandas.

Idealmente, esta reconstrução do sinal original a partir de suas componentes em frequência, não deveria introduzir qualquer tipo de distorção na ausência de ruído de quantização. Entretanto, os bancos de filtros passa-banda convencionais não conseguem cumprir esta condição devido à largura finita das regiões de transição dos filtros. Se as bandas de transição dos filtros se superpõem em frequência, a reamostragem na saída dos filtros passa-banda provoca "aliasing" nos espectros dos sinais dizimados. No receptor, por outro lado, esta superposição faz com que a filtragem dos sinais interpolados não consiga eliminar totalmente as imagens espectrais geradas durante a interpolação. Estas duas

¹ Se denominam "Formantes" as frequências de ressonância do trato vocal.

distorções destroem a estrutura harmônica dos sons, provocando degradação perceptível no sinal de voz.

Se, entretanto, os espectros não se superpõem em frequência (presença de "gaps" entre bandas), o sinal não pode ser perfeitamente reconstruído devido à perda de componentes espectrais. Além do mais, os "gaps" entre as sub-bandas introduzem no sinal reconstruído o efeito de reverberância mencionado anteriormente.

3.2.3.1 Distorção na dizimação

Na Fig. 3.7(a), é mostrado, como exemplo, o diagrama em blocos parcial de um codificador de sub-bandas que divide o espectro de voz em 2 sub-bandas de igual largura. Foram suprimidos os codificadores e decodificadores de cada sub-banda, para dar ênfase ao comportamento do banco de filtros divisor do espectro no transmissor, e sintetizador do espectro no receptor. Na Fig. 3.7(b), é mostrada a resposta em frequência dos filtros respectivos, onde pode-se notar a superposição das bandas de transição dos mesmos.

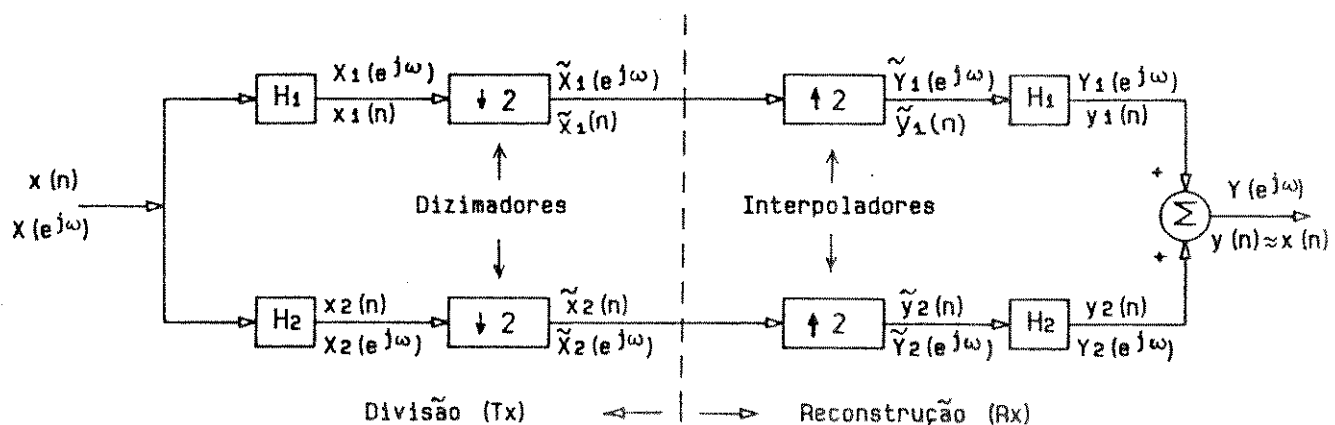


Figura 3.7(a): Esquema divisor e reconstrutor do espectro para $N=2$ sub-bandas.

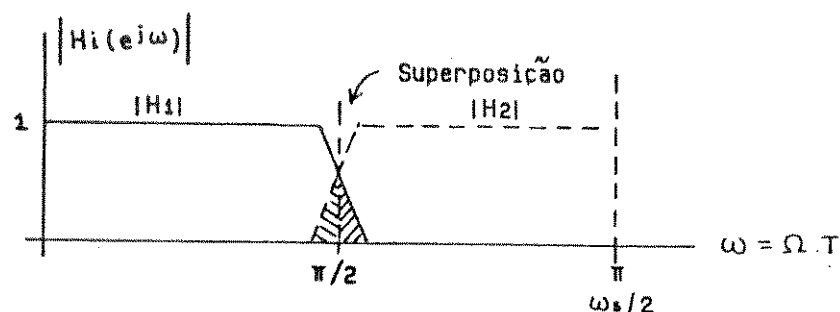


Figura 3.7(b): Resposta em frequência dos filtros.

Dado que a divisão do espectro é feita em 2 sub-bandas de igual largura (o semiespectro inferior $X_1(e^{j\omega})$ e o superior $X_2(e^{j\omega})$), a dizimação dos sinais na saída dos filtros correspondentes deve ser feita por um fator 2. Esta dizimação é indicada pelos blocos $\downarrow 2$, e é realizada descartando-se uma amostra a cada duas da sequência de entrada. Assim, para a banda inferior, tem-se:

$$\tilde{x}_1(n) = x_1(2n) \quad \therefore \quad (3.9)$$

$$\tilde{X}_1(e^{j\omega}) = \sum_{n=-\infty}^{\infty} \tilde{x}_1(n) \cdot e^{-j\omega n} = \sum_{n=-\infty}^{\infty} x_1(2n) \cdot e^{-j\omega n} \quad (3.10)$$

Fazendo-se $2n = m$ na equação acima, resulta:

$$\begin{aligned} \tilde{X}_1(e^{j\omega}) &= \sum_{\substack{m=-\infty \\ m \text{ par}}}^{\infty} x_1(m) \cdot e^{-j\omega m/2} = \sum_{m=-\infty}^{\infty} x_1(m) \cdot \frac{1 + (-1)^m}{2} \cdot e^{-j\omega m/2} \\ &= \frac{1}{2} \left[X_1(e^{j\omega/2}) + X_1(e^{j(\omega/2 - \pi)}) \right] \end{aligned} \quad (3.11)$$

Baseada na Eq.(3.11), a Fig.3.8 mostra os espectros correspondentes a $X_1(e^{j\omega})$ e a $\tilde{X}_1(e^{j\omega})$.

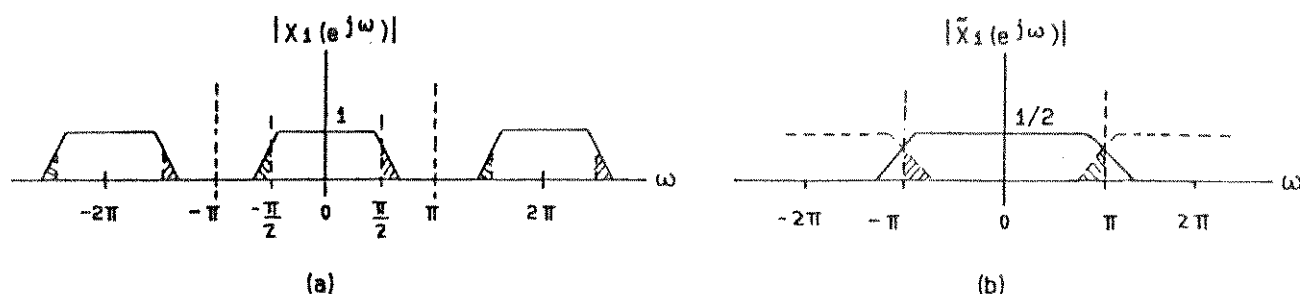


Figura 3.8: (a) Espectro de $x_1(n)$. As regiões sombreadas indicam componentes da banda alta não eliminadas por H_1 . (b) Espectro de $x_1(n)$. As regiões sombreadas indicam componentes da banda alta refletidas dentro da banda baixa durante a dizimação.

Na parte (a) da Fig.3.8, as regiões sombreadas indicam as componentes de frequência da banda alta não eliminadas pelo filtro passa-baixas $H_1(e^{j\omega})$. Durante a dizimação estas componentes são refletidas dentro da banda baixa, segundo a Fig.3.8(b), dando origem à distorção conhecida como "aliasing". Estas componentes correspondem ao segundo termo da Eq.(3.11), e ao fato dos filtros não serem ideais. No caso de filtros ideais, as componentes espectrais correspondentes ao segundo termo da Eq.(3.11) se estendem de π a 3π e de $-\pi$ a -3π , enquanto que as componentes desejadas (primeiro termo), se estendem de $-\pi$ a π . Desta forma o "aliasing" não se produz.

Para a banda alta ocorre de forma similar: as componentes de frequência abaixo de $\pi/2$, pertencentes à banda baixa (Fig.3.9(a)), são refletidas dentro da banda alta durante a dizimação, gerando "aliasing" (Fig.3.9(b)). No caso dos filtros serem ideais esta distorção não ocorre.

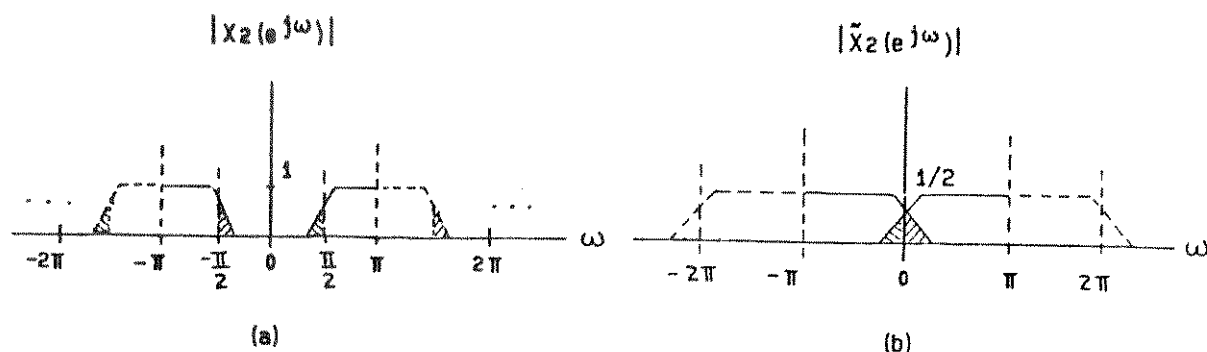


Figura 3.9: (a) Espectro de $x_2(n)$. As regiões sombreadas indicam componentes da banda baixa não eliminadas por H_2 . (b) Espectro de $x_2(n)$. As regiões sombreadas indicam componentes da banda baixa refletidas dentro da banda alta durante a dizimação.

3.2.3.2 Distorção na interpolação

No receptor, os sinais recebidos devem ser interpolados e filtrados de forma a serem realocados na posição correta do eixo de frequências ω . Para isso deve ser feita a interpolação dos sinais de entrada de um fator 2. A interpolação é implementada intercalando-se 1 amostra nula entre cada 2 amostras do sinal de entrada (bloco $\uparrow 2$ na Fig.3.7(a)). Assim, para a banda inferior tem-se:

$$\tilde{y}_1(n) = \begin{cases} \tilde{x}_1(n/2) & , \text{ para } n \text{ par.} \\ 0 & , \text{ para } n \text{ ímpar.} \end{cases} \quad (3.12)$$

$$\tilde{Y}_1(e^{j\omega}) = \sum_{n=-\infty}^{\infty} \tilde{y}_1(n) \cdot e^{-j\omega n} = \sum_{\substack{n=-\infty \\ n \text{ par}}}^{\infty} \tilde{x}_1(n/2) \cdot e^{-j\omega n} \quad (3.13)$$

Fazendo-se $n/2 = m$ e substituindo na equação acima, resulta:

$$\tilde{Y}_1(e^{j\omega}) = \sum_{m=-\infty}^{\infty} \tilde{x}_1(m) \cdot e^{-j\omega 2m} = \tilde{X}_1(e^{j2\omega}) \quad (3.14)$$

O sinal $\tilde{y}_1(n)$ é então filtrado pelo filtro passa-baixas $H_1(e^{j\omega})$, obtendo-se o sinal $y_1(n)$ segundo mostra a Fig. 3.7. Os espectros correspondentes podem ser vistos na Fig. 3.10.

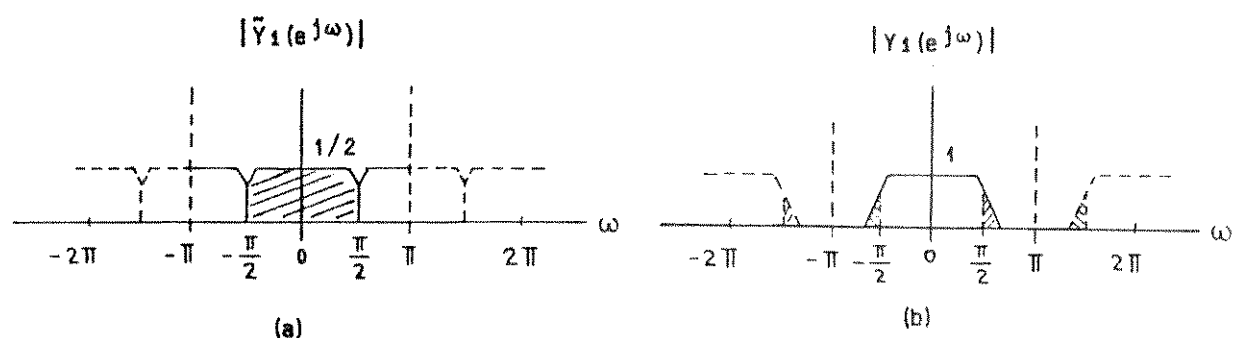


Figura 3.10: (a) Espectro de $y_1(n)$; (b) Espectro de $y_1(n)$ após filtragem. As regiões sombreadas indicam componentes espectrais não eliminadas pelo filtro H_1 .

Na Fig. 3.10(a), podem-se ver as imagens do espectro principal (aquele que vai de $-\pi/2$ a $\pi/2$), geradas ao longo do eixo ω durante a interpolação (Eq. (3.14)). Estas imagens não são totalmente eliminadas pelo filtro $H_1(e^{j\omega})$ do receptor, restando resíduos espectrais como indica a Fig. 3.10(b) (regiões sombreadas). Com a banda superior ocorre de forma similar: as imagens espectrais correspondentes à banda superior geradas durante a interpolação (Fig. 3.11(a)), não são totalmente eliminadas pelo filtro passa-altas $H_2(e^{j\omega})$ do receptor, ficando, em consequência, resíduos espectrais (Fig. 3.11(b)).

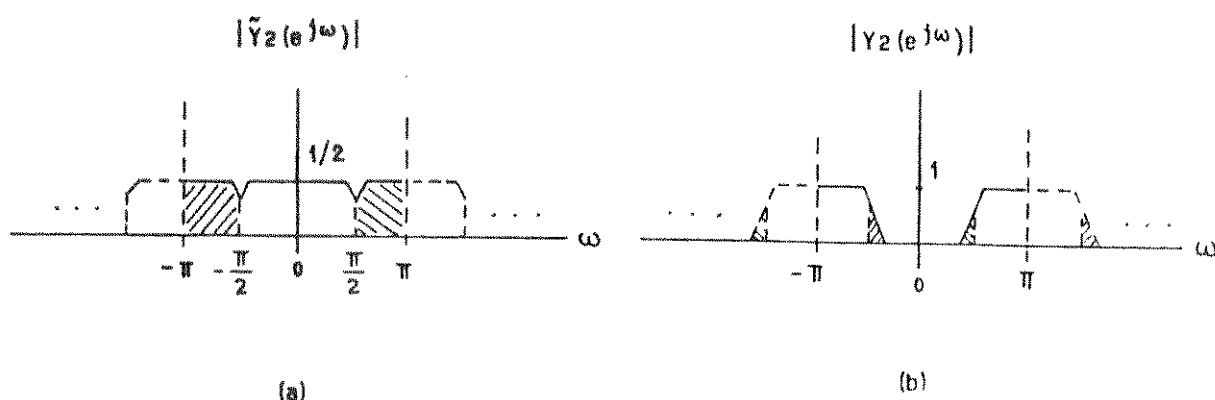


Figura 3.11: (a) Espectro de $y_2(n)$; (b) Espectro de $y_2(n)$ após filtragem. As regiões sombreadas indicam componentes espectrais não eliminadas pelo filtro H_2 .

Para reduzir ao mínimo estes problemas, as regiões de transição dos filtros devem ter a menor largura possível. Isto implica utilizar filtros quase ideais, o que torna extremamente rigorosas as especificações de projeto dos mesmos. Uma alternativa para solucionar os problemas de distorção apresentados pelos bancos de filtros passa-banda convencionais, é usar bancos de filtros espelho em quadratura. Estes filtros permitem uma reconstrução exata do sinal original a partir de suas N sub-bandas, sem a presença de "aliasing" e com uma resposta em frequência quase plana, independentemente da ordem dos filtros utilizados.

3.3 BANCO DE FILTROS ESPELHO EM QUADRATURA (QMF BANK)

3.3.1 ESTUDO DO ESQUEMA BÁSICO DE UM BANCO QMF

A técnica de divisão e reconstrução do espectro, utilizando filtros espelho em quadratura, é descrita na Fig.3.12 para um sistema de 2 sub-bandas (Ref.[7,8,9]).

A Fig.3.12(a) mostra o esquema básico divisor e reconstrutor do espectro empregando esta classe de filtros. O sinal de entrada $x(t)$, limitado em faixa a $F_s/2$ [Hz] e amostrado à taxa de Nyquist F_s [Hz], é filtrado pelos filtros de meia banda H_1 e H_2 , cuja resposta em frequência é mostrada na Fig.3.12(b).

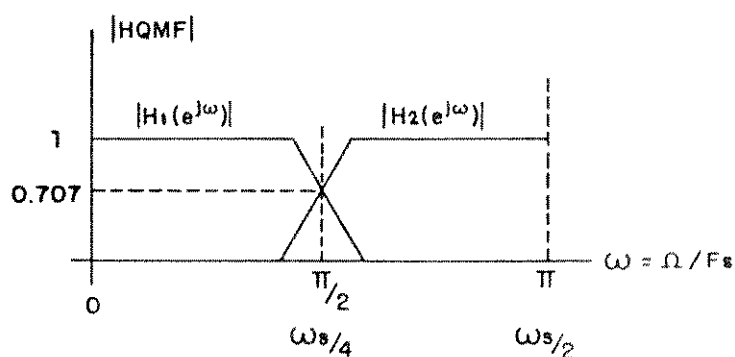
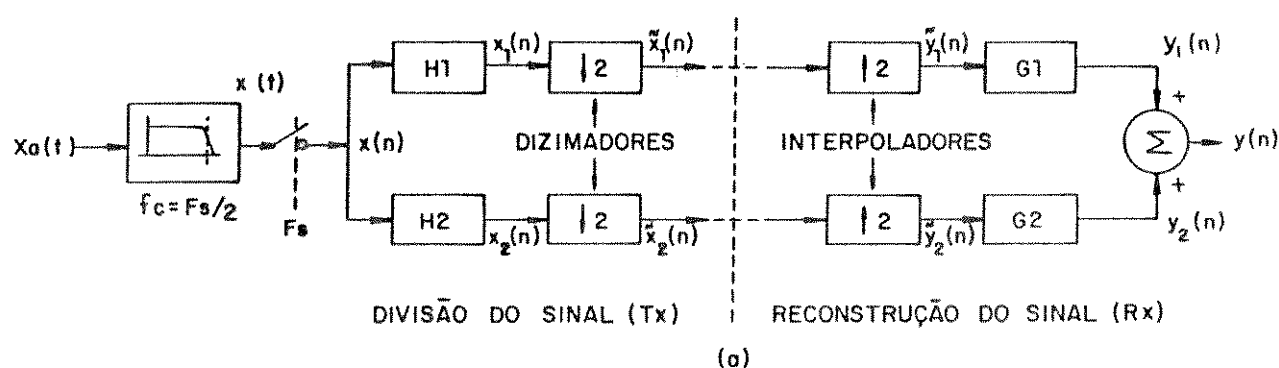


Figura 3.12: Divisão e reconstrução do espectro em 2 sub-bandas de igual largura, utilizando filtros espelho em quadratura. (a) Diagrama em blocos do sistema; (b) Resposta em frequência dos filtros respectivos.

O filtro H_1 , de resposta impulsiva $h_1(n)$ e Transformada- Z $H_1(z)$, é um filtro passa-baixas de 1/2 banda com frequência de corte $\omega_s/4 = \pi/2$, em cuja saída obtém-se o

semi-espectro inferior de $x(n)$. O filtro H_2 , de resposta impulsiva $h_2(n)$ e Transformada-Z $H_2(z)$, é um filtro passa-altas de 1/2 banda com frequência de corte $\pi/2$, em cuja saída obtém-se o semi-espectro superior de $x(n)$.

No desenvolvimento matemático que se segue, será utilizada a seguinte notação:

$$H_1(e^{j\omega}) = |H_1(e^{j\omega})| \cdot e^{j\phi_1(\omega)} \quad (3.15a)$$

$$H_2(e^{j\omega}) = |H_2(e^{j\omega})| \cdot e^{j\phi_2(\omega)} \quad (3.15b)$$

H_1 e H_2 não são filtros quaisquer. Eles constituem um par de filtros espelho, isto é, o espectro de um é a imagem refletida do espectro do outro ao redor do eixo $\omega = \omega_s/4 = \pi/2$ (Fig.3.12(b)). Matematicamente isto significa:

$$|H_2(e^{j\omega})| = |H_1(e^{j(\omega_s/2 - \omega)})| = |H_1(e^{j(\pi - \omega)})| \quad (3.16)$$

Além disto suas respostas de fase diferem de $\pi/2$ ($\phi_2(\omega) = \phi_1(\omega) \pm \pi/2$), e daí o nome de "Filtros Espelho em Quadratura".

Como $h_1(n)$ e $h_2(n)$ são filtros com coeficientes reais resulta

$$|H_2(e^{j\omega})| = |H_2(e^{-j\omega})| = |H_1(e^{j(\omega - \pi)})| \quad (3.17)$$

A Eq.(3.17) mostra que o espectro de amplitude de um filtro é o espectro de amplitude do outro deslocado de π . Para $H_1(e^{j\omega})$ se escolhe um filtro FIR simétrico (fase linear), de M taps, enquanto que $h_2(n)$ é derivado de $h_1(n)$ pela simples transformação passa-baixas $\rightarrow$ passa-altas: $h_2(n) = (-1)^n \cdot h_1(n)$

Portanto, se:

$$H_1(e^{j\omega}) = \sum_{n=0}^{M-1} h_1(n) e^{-j\omega n}, \quad \text{então} \quad (3.18)$$

$$H_2(e^{j\omega}) = \sum_{n=0}^{M-1} h_2(n) e^{-j\omega n} = \sum_{n=0}^{M-1} (-1)^n h_1(n) e^{-j\omega n}$$

Considerando que $(-1)^n = e^{j\pi n}$, resulta:

$$H_2(e^{j\omega}) = \sum_{n=0}^{M-1} h_1(n) e^{-j(\omega-\pi)n} = H_1(e^{j(\omega-\pi)}) = H_1(-e^{j\omega}) \quad (3.19)$$

Se demonstrará agora que $h_1(n)$ e $h_2(n)$ constituem um par de filtros espelho em quadratura. Por ser $h_1(n)$ um filtro FIR simétrico de M "taps", sua Transformada de Fourier pode ser expressa como (Ref[13]).

$$H_1(e^{j\omega}) = |H_1(e^{j\omega})| e^{j\phi_1(\omega)} = |H_1(e^{j\omega})| e^{-j\omega(M-1)/2} \quad (3.20)$$

Analogamente, empregando-se as Eqs. (3.19) e (3.20) tem-se:

$$\begin{aligned} H_2(e^{j\omega}) &= |H_2(e^{j\omega})| e^{j\phi_2(\omega)} = |H_1(e^{j(\omega-\pi)})| e^{j\phi_1(\omega-\pi)} \\ &= |H_1(e^{j(\omega-\pi)})| e^{-j(\omega-\pi)(M-1)/2} \\ &= |H_1(e^{j(\omega-\pi)})| e^{j\phi_1(\omega)} e^{j\pi(M-1)/2} \end{aligned} \quad (3.21)$$

Como será visto logo mais, M precisa ser par para uma perfeita reconstrução do sinal de entrada; em consequência M-1 será ímpar e portanto:

$$\begin{aligned}
 H_2(e^{j\omega}) &= |H_1(e^{j(\omega-\pi)})| \cdot e^{j\phi_1(\omega)} \cdot e^{\pm j\pi/2} \\
 &= |H_1(e^{j(\omega-\pi)})| \cdot e^{j(\phi_1(\omega) \pm \pi/2)}
 \end{aligned} \tag{3.22}$$

$$\text{De (3.22) tem-se pois: } \begin{cases} |H_2(e^{j\omega})| = |H_1(e^{j(\omega-\pi)})| \\ \phi_2(\omega) = \phi_1(\omega) \pm \pi/2 \end{cases} \tag{3.23}$$

A primeira equação se refere à condição dos filtros serem espelhados (comparar com a Eq.(3.17)), ao passo que a segunda equação se refere à condição de quadratura dos filtros.

Será analisada agora a estrutura da Fig.3.12(a). Para facilitar o tratamento matemático se trabalhará no domínio da Transformada- $\mathbb{Z}$. A relação entre as Transformadas- $\mathbb{Z}$ de $x(n)$, $h_1(n)$ e $x_1(n)$ é dada por:

$$X_1(z) = X(z) \cdot H_1(z) \tag{3.24}$$

Como $X_1(e^{j\omega})$, semi-espectro inferior de $x(n)$, ocupa a metade da largura da faixa de $X(e^{j\omega})$, o sinal $x_1(n)$ pode ser dizimado por um fator 2 sem perda de informação. A transformada- $\mathbb{Z}$ do sinal resultante $\tilde{x}_1(n)$, será dada por:

$$\tilde{X}_1(z) = \frac{1}{2} \left[X_1(z^{1/2}) + X_1(-z^{1/2}) \right] \tag{3.25}$$

No receptor, o sinal $\tilde{x}_1(n)$ é interpolado de um fator 2 para retornar à frequência de amostragem original, obtendo-se o sinal $\tilde{y}_1(n)$. Assim:

$$\tilde{Y}_1(z) = \tilde{X}_1(z^2) \tag{3.26}$$

Após a filtragem final por $G_1(z)$, o sinal $y_1(n)$ obtido terá

Transformada- Z dada por:

$$Y_1(z) = \tilde{Y}_1(z) \cdot G_1(z) \quad (3.27)$$

Combinando-se as Eqs. (3.24) a (3.27) resulta:

$$\begin{aligned} Y_1(z) &= \tilde{Y}_1(z) \cdot G_1(z) = \tilde{X}_1(z^2) \cdot G_1(z) = \frac{1}{2} [X_1(z) + X_1(-z)] \cdot G_1(z) \\ &= \frac{1}{2} [X(z) \cdot H_1(z) + X(-z) \cdot H_1(-z)] \cdot G_1(z) \end{aligned} \quad (3.28)$$

Calculando-se $Y_2(z)$ de forma similar resulta:

$$\begin{aligned} Y_2(z) &= \tilde{Y}_2(z) \cdot G_2(z) = \tilde{X}_2(z^2) \cdot G_2(z) = \frac{1}{2} [X_2(z) + X_2(-z)] \cdot G_2(z) \\ &= \frac{1}{2} [X(z) \cdot H_2(z) + X(-z) \cdot H_2(-z)] \cdot G_2(z) \end{aligned} \quad (3.29)$$

Como o sinal de saída $y(n)$ é a soma de $y_1(n)$ e $y_2(n)$, tem-se:

$$\begin{aligned} Y(z) &= Y_1(z) + Y_2(z) = \frac{1}{2} [H_1(z) \cdot G_1(z) + H_2(z) \cdot G_2(z)] \cdot X(z) + \\ &\quad \frac{1}{2} [H_1(-z) \cdot G_1(z) + H_2(-z) \cdot G_2(z)] \cdot X(-z) \end{aligned} \quad (3.30)$$

Levando-se em conta que $h_2(n) = (-1)^n \cdot h_1(n)$ e que $H_2(z) = H_1(-z)$, a equação anterior pode ser reescrita como:

$$\begin{aligned} Y(z) &= \frac{1}{2} X(z) \cdot [H_1(z) \cdot G_1(z) + H_2(z) \cdot G_2(z)] + \\ &\quad \frac{1}{2} X(-z) \cdot [H_2(z) \cdot G_1(z) + H_1(z) \cdot G_2(z)] \end{aligned} \quad (3.31)$$

O primeiro termo da Eq. (3.31) corresponde ao sinal desejado $x(n)$, ao passo que o segundo termo desta equação

representa as componentes de "aliasing" caso os filtros H_1 e H_2 não sejam ideais. Estas componentes podem ser eliminadas mesmo com os filtros H_1 e H_2 não ideais, escolhendo-se adequadamente os filtros G_1 e G_2 . Assim, se:

$$G_1(z) = H_1(z) \quad (3.32a)$$

$$G_2(z) = -H_2(z) \quad (3.32b)$$

o segundo termo da Eq.(3.31) se cancela, e com ele as possíveis componentes de "aliasing". Resulta então:

$$\begin{aligned} Y(z) &= \frac{1}{2} X(z) \cdot [H_1(z) \cdot H_1(z) - H_2(z) \cdot H_2(z)] \\ &= \frac{1}{2} X(z) \cdot [H_1^2(z) - H_2^2(z)] \end{aligned} \quad (3.33)$$

Avaliando-se esta expressão sobre a circunferência de raio unitário, obtém-se:

$$Y(e^{j\omega}) = \frac{1}{2} X(e^{j\omega}) \cdot [H_1^2(e^{j\omega}) - H_2^2(e^{j\omega})] \quad (3.34)$$

Levando-se agora em conta as Eqs.(3.20) e (3.21), (observe-se que estas equações são gerais e nelas não foi feita nenhuma hipótese a respeito de M ser par ou ímpar), obtém-se:

$$H_1^2(e^{j\omega}) = |H_1(e^{j\omega})|^2 \cdot e^{j2\phi_1(\omega)} = |H_1(e^{j\omega})|^2 \cdot e^{-j(M-1)\omega} \quad (3.35a)$$

$$H_2^2(e^{j\omega}) = |H_2(e^{j\omega})|^2 \cdot e^{j2\phi_2(\omega)} = |H_2(e^{j\omega})|^2 \cdot e^{-j(M-1)\omega} \cdot e^{j(M-1)\pi} \quad (3.35b)$$

Finalmente, substituindo-se estas equações na Eq.(3.34), obtém-se:

$$Y(e^{j\omega}) = \frac{1}{2} X(e^{j\omega}) \left[|H_1(e^{j\omega})|^2 - |H_2(e^{j\omega})|^2 \cdot e^{j(M-1)\pi} \right] \cdot e^{-j(M-1)\omega} \quad (3.36)$$

Dois casos devem ser considerados em (3.36): M par e M ímpar.

- M par:

Neste caso: $e^{j(M-1)\pi} = -1$ e a Eq. (3.36) se reduz a:

$$Y(e^{j\omega}) = \frac{1}{2} X(e^{j\omega}) \left[|H_1(e^{j\omega})|^2 + |H_2(e^{j\omega})|^2 \right] \cdot e^{-j(M-1)\omega} \quad (3.37)$$

$$\text{Se a condição: } |H_1(e^{j\omega})|^2 + |H_2(e^{j\omega})|^2 = 1 \quad \forall \omega \quad (3.38)$$

é alcançada, então:

$$Y(e^{j\omega}) = \frac{1}{2} X(e^{j\omega}) \cdot e^{-j(M-1)\omega}, \quad \text{e portanto} \quad (3.39)$$

$$y(n) = \frac{1}{2} x(n-(M-1)) \quad (3.40)$$

Isto significa que $y(n)$ é uma réplica exata do sinal de entrada $x(n)$, a menos do fator $1/2$ e do atraso de $M-1$ amostras entre o sinal reconstruído e o sinal original.

O fator $1/2$ aparece devido à perda de energia que se produz na dizimação, ao passo que o atraso de $M-1$ amostras aparece devido ao fato de se usar filtros FIR de fase linear. Esta classe de filtros provoca um atraso de $(M-1)/2$ amostras no sinal filtrado, onde M é a ordem do filtro. Considerando-se o filtro do transmissor e do receptor, o atraso total produzido no sinal é de $M-1$ amostras.

- M ímpar:

Neste caso: $e^{j(M-1)\pi} = 1$ e a Eq.(3.36) se reduz a:

$$Y(e^{j\omega}) = \frac{1}{2} X(e^{j\omega}) \cdot [|H_1(e^{j\omega})|^2 - |H_2(e^{j\omega})|^2] \cdot e^{-j(M-1)\omega} \quad (3.41)$$

Para uma perfeita reconstrução do sinal de entrada seria necessário que: $|H_1(e^{j\omega})|^2 - |H_2(e^{j\omega})|^2 = 1 \quad \forall \omega$, condição esta que é impossível de ser satisfeita já que para $\omega = \pi/2$ a expressão acima é zero.

Resumindo-se então os resultados obtidos, tem-se que as condições para uma perfeita reconstrução do sinal de entrada são:

- 1) $H_1(z)$: filtro FIR de fase linear e ordem par.
- 2) $H_2(z) = H_1(-z) \Rightarrow$ filtros espelho em quadratura. No domínio do tempo: $h_2(n) = (-1)^n \cdot h_1(n)$.
- 3) $G_1(z) = H_1(z) \Rightarrow g_1(n) = h_1(n)$
 $G_2(z) = -H_2(z) = -H_1(-z) \Rightarrow g_2(n) = (-1)^{n+1} \cdot h_1(n)$.
- 4) $|H_1(e^{j\omega})|^2 + |H_2(e^{j\omega})|^2 = |H_1(e^{j\omega})|^2 + |H_1(e^{j(\omega-\pi)})|^2 = 1 \quad \forall \omega$
 o que implica numa resposta composta plana.

Adicionalmente, o fato de se utilizar filtros FIR de fase linear na divisão e reconstrução do espectro, faz com que a função de transferência do sistema completo possua fase linear e, conseqüentemente, a única classe de distorção possível de ser introduzida seja de amplitude ($|H_1(e^{j\omega})|^2 + |H_2(e^{j\omega})|^2 \neq \text{constante}$). Embora a ausência de distorção de fase não beneficie a transmissão de voz, uma vez que o ouvido humano é quase insensível à distorção de fase, ela facilita bastante a transmissão de dados.

Conforme foi visto na Seção 3.2.3, ao se utilizar um banco de filtros passa-bandas comuns, o fato desses filtros não serem ideais produzia:

- 1) Termos de "aliasing" na dizimação.
- 2) Resíduos espectrais gerados durante a interpolação não eliminados pelos filtros do receptor.

Utilizando-se filtros espelho em quadratura estes 2 efeitos continuam existindo, mas agora eles se cancelam mutuamente em $y(n)$ após a soma de $y_1(n)$ com $y_2(n)$. Em consequência é possível uma reconstrução exata do sinal original, independentemente da superposição dos espectros dos filtros nas regiões de transição dos mesmos.

Cabe finalmente mencionar que é possível uma perfeita reconstrução do sinal de entrada ainda que utilizando M ímpar, fazendo-se

$$G_1(z) = H_1(z) \quad (3.42a)$$

$$G_2(z) = H_2(z) \quad (3.42b)$$

e introduzindo-se um atraso de 1 amostra no filtro passa-altas do transmissor e um atraso similar no filtro passa-baixas do receptor (Ref.[8]). Isto é mostrado na Fig.3.13, onde o atraso de uma amostra é representado pelo bloco z^{-1} , correspondente à Transformada- Z do atraso mencionado. Neste caso a Transformada- Z do sinal de saída é dada por:

$$Y(z) = \frac{1}{2} X(z) \cdot [H_1^2(z) + H_2^2(z)] \cdot z^{-1} \quad (3.43)$$

que avaliada sobre a circunferência de raio unitário e para M ímpar resulta em:

$$Y(e^{j\omega}) = \frac{1}{2} X(e^{j\omega}) \cdot [|H_1(e^{j\omega})|^2 + |H_2(e^{j\omega})|^2] \cdot e^{-jM\omega} \quad (3.44)$$

As Eqs. (3.43) e (3.44) devem ser comparadas com as Eqs. (3.33) e (3.37) respectivamente.

Neste trabalho, porém, utilizaram-se filtros espelho em quadratura apenas com M par, pelo que não se estenderá o estudo de bancos QMF com M ímpar.

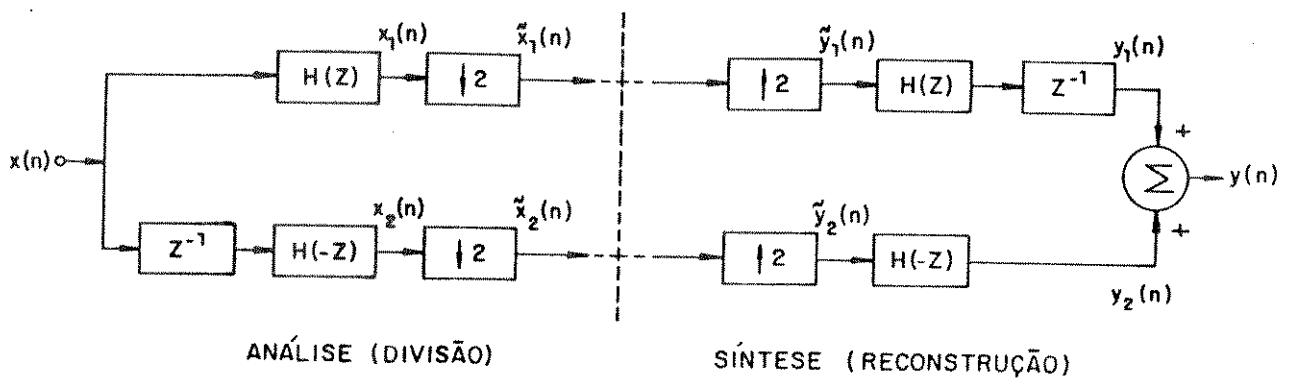


Figura 3.13: Divisão e reconstrução do espectro em 2 sub-bandas de igual largura, utilizando-se filtros espelho em quadratura de ordem ímpar.

3.3.2 IMPLEMENTAÇÃO E CÁLCULO DO BANCO QMF

A Eq. (3.38) mostra que para se reconstruir o sinal original, é preciso que:

$$|H_1(e^{j\omega})|^2 + |H_2(e^{j\omega})|^2 = |H_1(e^{j\omega})|^2 + |H_1(e^{j(\omega-\pi)})|^2 = 1 \quad \forall \omega \quad (3.45)$$

Esta condição implica que em $\omega = \omega_s/4 = \pi/2$, a resposta individual dos filtros seja 0,707 como indica a Eq. (3.46).

$$|H_1(e^{j\pi/2})| = |H_2(e^{j\pi/2})| = 0,707 \quad (3.46)$$

Não é possível se achar um filtro FIR de fase linear que verifique exatamente a Eq.(3.45), exceto em dois casos simples: o primeiro quando $N = \infty$, que é impossível de se implementar na prática, e o segundo quando $H_1(e^{j\omega})$ é de ordem 1 ($H_1(e^{j\omega}) = (1 + e^{-j\omega})/2$), que não apresenta qualquer interesse prático. O que se faz então é aproximar a Eq.(3.45), usando-se filtros com baixo "ripple" na banda passante, alta atenuação na banda de rejeição e verificando (3.45) na banda de transição. O cálculo destes filtros pode ser feito de várias formas (Ref.[8,10,11,12,13]), sendo particularmente interessante o método proposto por Johnston (Ref.[13]). Nesse trabalho, Johnston estabelece 2 parâmetros para o cálculo do banco de filtros espelho em quadratura:

1. "Ripple" na resposta composta do sistema de filtros.
2. Atenuação na banda de rejeição de cada um dos filtros.

Com estes dois parâmetros ele constrói uma função erro E , que é minimizada utilizando-se o *algoritmo de Hooke e Jeaves* (este algoritmo tenta minimizar uma função de várias variáveis (Ref[14])). A função erro E a minimizar é:

$$E = E_r + \alpha \cdot E_s(\omega_{sb}) \quad (3.47)$$

onde
$$E_r = 2 \int_{\omega=0}^{\pi/2} [|H_1(e^{j\omega})|^2 + |H_1(e^{j(\omega-\pi)})|^2 - 1]^2 d\omega \quad (3.48)$$

e
$$E_s = \int_{\omega=\omega_{sb}}^{\pi} |H_1(e^{j\omega})|^2 d\omega \quad (3.49)$$

"Er" corresponde à energia de "ripple" do sistema filtrante, ao passo que "Es" corresponde à energia da banda de rejeição de $H_1(e^{j\omega})$; " α " é um fator de ponderação de Es, e " ω_{sb} " (stop-band frequency), corresponde à frequência de início da banda de rejeição do filtro. Para achar as integrais das Eqs. (3.48) e (3.49), Johnston considera amostras de $H_1(e^{j\omega})$ em 512 pontos e realiza a somatória correspondente.

Aplicando este método, Johnston obteve uma família de filtros QMF, partindo de diferentes especificações para o "ripple" na banda passante e a atenuação na banda de rejeição. Como exemplo, na Fig.3.14 é mostrada a resposta em frequência do banco QMF obtido por Johnston para o caso de $M=32$ "taps" e banda de transição normalizada de 0,043, onde banda de transição normalizada é o quociente $(\omega_{sb}-\pi/2)/2\pi$. Na parte (a) da figura é mostrada a resposta em frequência dos filtros individuais, enquanto que na parte (b) é mostrada a resposta em frequência do sistema composto $|H_1(e^{j\omega})|^2 + |H_2(e^{j\omega})|^2$.

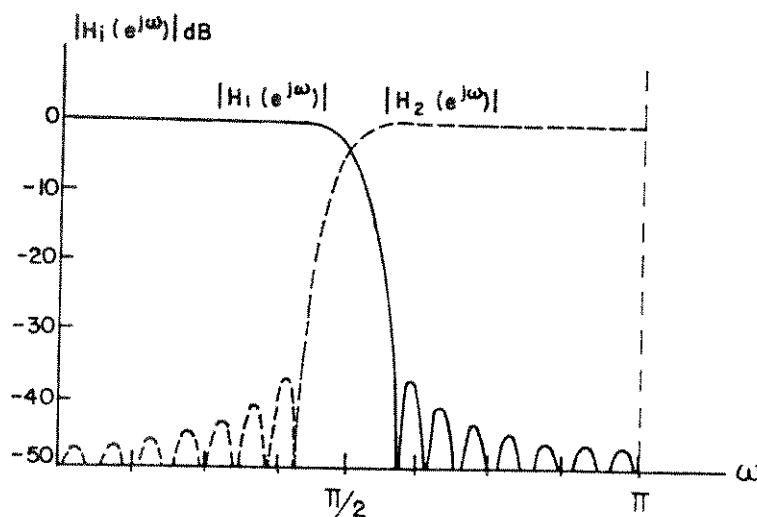


Figura 3.14(a): Resposta em frequência dos filtros individuais do banco QMF, com $M=32$ "taps" e banda de transição código D.

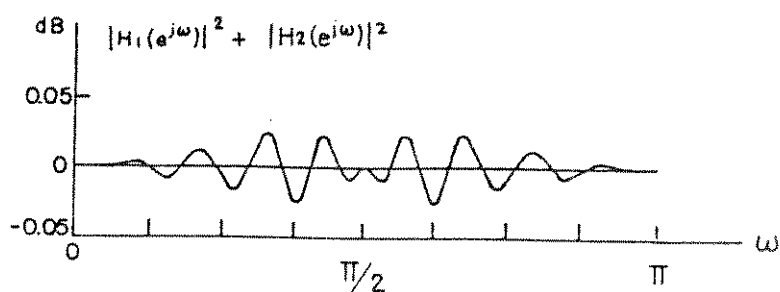


Figura 3.14(b): Resposta em frequência do sistema composto.

Como pode-se observar, a Eq.(3.45) é satisfeita com um erro de $\pm 0,025$ [dB], o qual é mais que satisfatório para um bom desempenho do sistema filtrante. A atenuação fora da banda passante varia de -38 [dB] a -48 [dB], e a largura da banda de transição do filtro é de 0,043.

Os coeficientes correspondentes a estes filtros, são apresentados no Apêndice C.

Voltando ao esquema da Fig.3.12(a), já que os filtros $h_1(n)$ e $h_2(n)$ possuem "taps" com igual magnitude ($h_2(n) = (-1)^n \cdot h_1(n)$), é possível combinar as operações aritméticas dos 2 filtros. Esta combinação, ilustrada na Fig.3.15(a) para o caso de $M=4$ "taps", é feita calculando-se separadamente as contribuições dos "taps" pares e ímpares, e depois combinando-se os resultados. De igual forma, e como mostra a Fig.3.15(b), uma técnica similar pode ser utilizada na reconstrução do sinal no receptor, uma vez que $g_1(n)$ e $g_2(n)$ também possuem "taps" com igual magnitude ($g_1(n) = h_1(n)$, $g_2(n) = (-1)^{n+1} \cdot h_1(n)$). Este procedimento reduz a carga computacional total para o processo de análise e síntese de um sinal, a M multiplicações e M adições a cada período de amostragem $T = 1/F_s$, onde M é o número de "taps" do filtro $h_1(n)$.

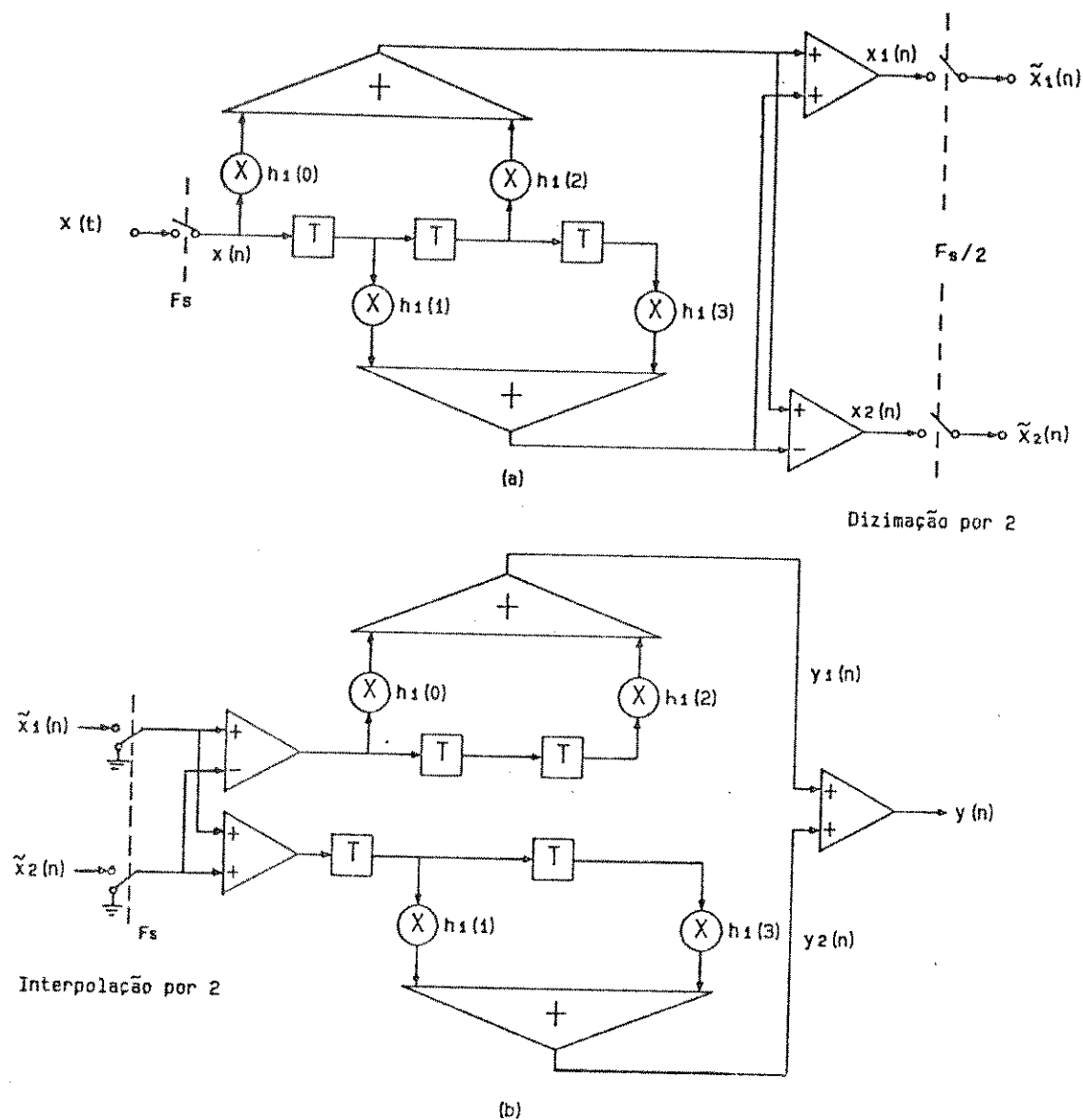


Figura 3.15: Divisão e reconstrução do espectro com filtros QMF de 4 "taps": (a) Divisão; (b) Reconstrução.

Um sistema SBC completo para $N=2$ sub-bandas, utilizando-se filtros espelho em quadratura, encontra-se ilustrado na Fig.3.16. Na Fig.3.17 são mostrados os espectros dos sinais nos diferentes estágios do transmissor e do receptor.

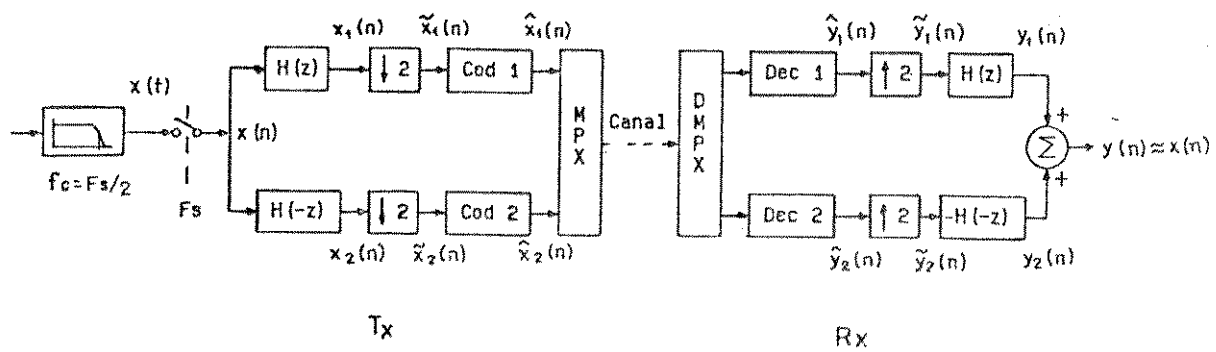


Figura 3.16: Esquema completo de um codificador de sub-bandas, utilizando-se filtros espelho em quadratura.

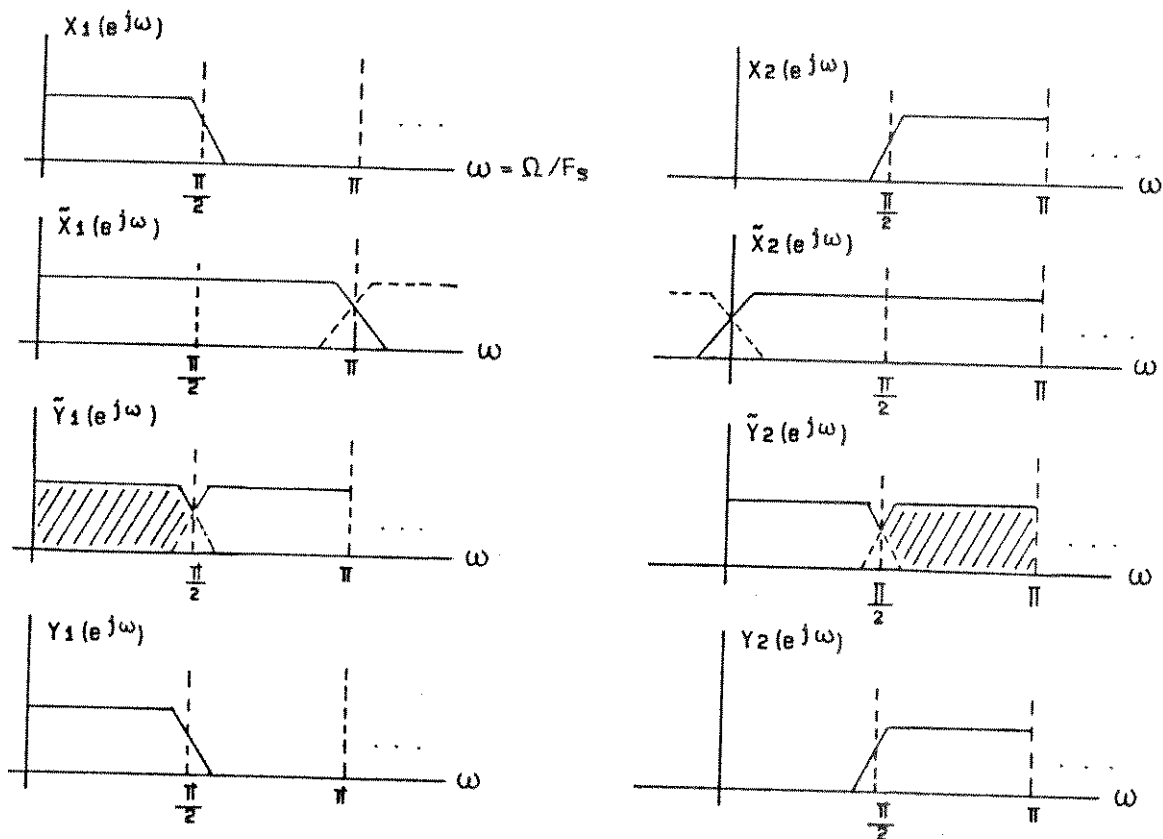


Figura 3.17: Espectro dos sinais da Fig.3.16 nos diferentes estágios do transmissor e do receptor.

3.3.3 DIVISÃO DO ESPECTRO EM 2^P SUB-BANDAS DE IGUAL LARGURA, UTILIZANDO-SE FILTROS QMF

A técnica de divisão do espectro em 2 sub-bandas utilizando-se filtros QMF, pode facilmente ser estendida à divisão em 2^P sub-bandas de igual largura ("p" inteiro), usando-se uma estrutura em árvore de filtros espelho em quadratura. A Fig.3.18 exemplifica esta técnica para $N = 2^P = 4$ sub-bandas. A parte (a) desta figura, mostra a estrutura filtrante que divide o espectro do sinal de entrada em 4 sub-bandas no transmissor; na parte (b) é mostrada a estrutura que realiza a reconstrução do espectro no receptor, e na parte (c) é mostrado o esquema da resposta em frequência do banco QMF.

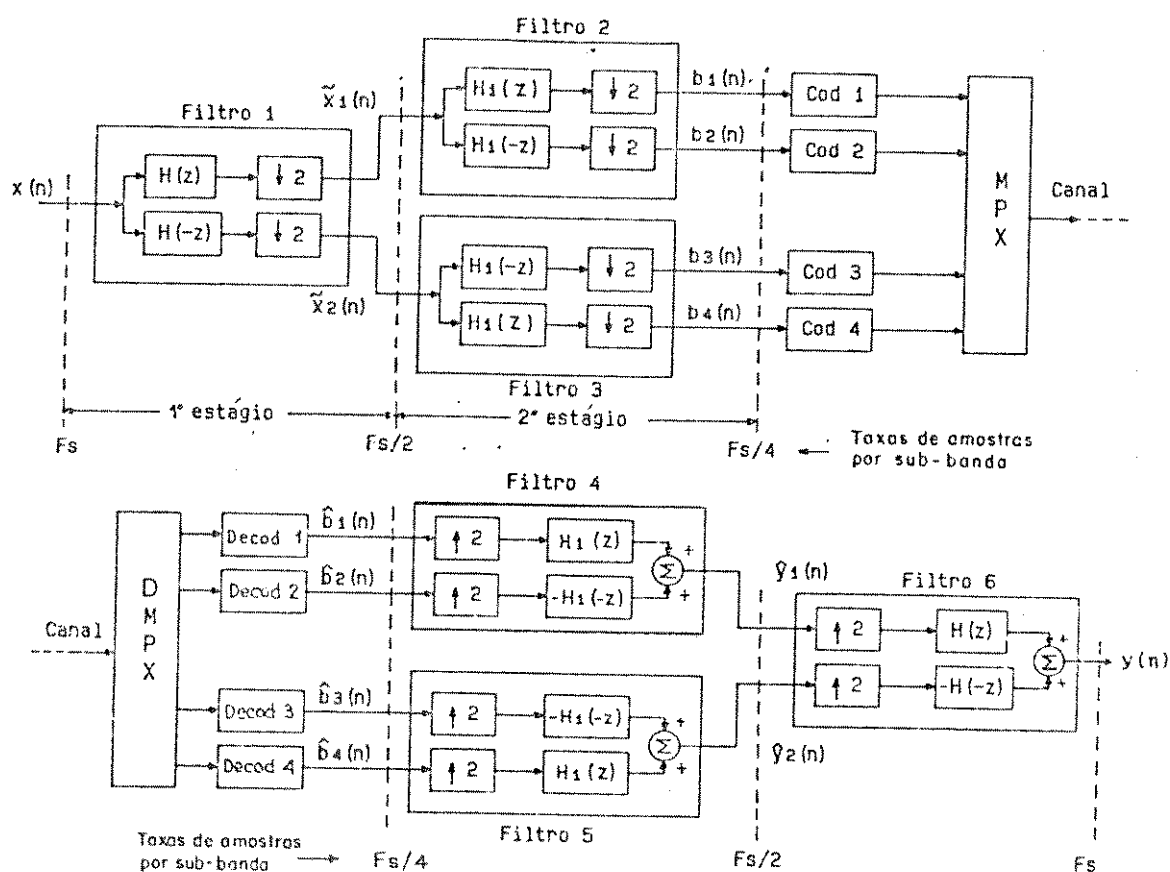


Figura 3.18: Diagrama em blocos de um sistema SBC para $N=4$ sub-bandas: (a) Tx; (b) Rx.

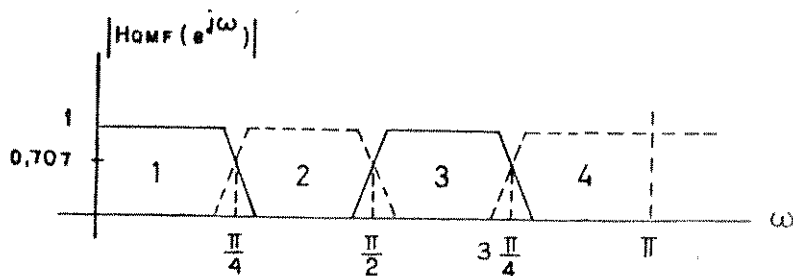


Figura 3.18(c): Resposta em frequência do banco QMF.

O par de filtros espelho em quadratura $H_k(z)$ e $H_k(-z)$ ($H_k(e^{j\omega})$ e $H_k(e^{j(\omega-\pi)})$, respectivamente), junto com seus respectivos dizimadores $\downarrow 2$, será considerado a partir de agora como um único filtro chamado "Filtro divisor do espectro" (FDE). Este filtro permite a divisão do sinal de entrada em duas sub-bandas de igual largura e a posterior dizimação dos sinais resultantes por um fator 2, de modo a manter a taxa total de amostras constante em seu valor original (Fig. 3.19(a)). De igual forma o par de filtros $H_k(z)$, $-H_k(-z)$ com seus respectivos interpoladores $\uparrow 2$, será considerado como um único filtro denominado "Filtro reconstrutor do espectro" (FRE), que permite a reconstrução do sinal original a partir de seus semi-espectros inferior e superior, tal como eles são obtidos na saída de um FDE (Fig. 3.19(b)).

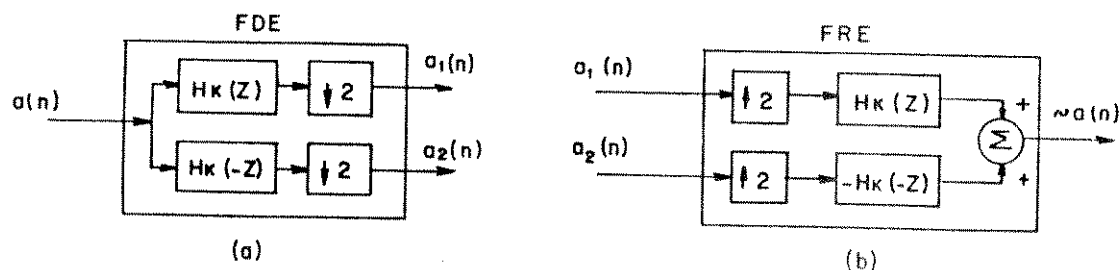


Figura 3.19: (a) Filtro Divisor do Espectro (FDE); (b) Filtro Reconstrutor do Espectro (FRE).

Da Fig.3.18(a), pode-se ver que para se obter $N = 2^p = 4$ sub-bandas, o sinal original é, num primeiro estágio, dividido em 2 sub-bandas de igual largura mediante um FDE. Num segundo estágio cada um destes sinais é por sua vez dividido em outras 2 sub-bandas, utilizando-se os FDE 2 e 3. Na saída deste segundo estágio se obtém os 4 sinais $b_1(n), \dots, b_4(n)$ tal como desejado. Foram necessários, em consequência, $p = 2$ estágios e $N-1$ filtros divisores do espectro, FDE, para se obter as 4 sub-bandas.

A taxa de amostras de cada sinal na saída do primeiro estágio é $F_s/2$ [am/s], ao passo que a taxa de amostras de cada sinal na saída do segundo e último estágio é $F_s/4$ [am/s] (F_s/N em geral).

Para se reconstruir o sinal original no receptor, é utilizada uma estrutura de filtros simétrica àquela do transmissor. Esta estrutura possuirá, em consequência, " p " estágios e $N-1$ filtros reconstrutores de espectro FRE. As taxas de amostras à saída de cada estágio estão indicadas na Fig.3.18(b).

É claro que o processo da Fig.3.18(a) pode ser reaplicado a cada sub-banda gerada. Assim, se os sinais $b_k(n)$ são divididos em 2 sub-bandas cada um, através de um terceiro estágio divisor, então na saída do sistema obteremos $N = 2^3 = 8$ sub-bandas, cada uma com uma taxa de $F_s/8$ [am/s]. Generalizando para $N = 2^p$ sub-bandas, obtém-se a Fig.3.20.

Pode-se observar que:

- Para se obter $N = 2^p$ sub-bandas, são necessários " p " estágios.
- O k -ésimo estágio converte 2^{k-1} sub-bandas, em 2^k sub-bandas, utilizando 2^{k-1} FDE. Na estrutura toda, são utilizados:

$$\sum_{k=1}^P 2^{k-1} = 2^P - 1 = N-1 \quad [\text{FDE}] \quad (3.50)$$

- A taxa de amostras do sinal de cada sub-banda à saída do k -ésimo estágio é: $F_s/2^k$ [am/s]. Portanto, na saída do último estágio será: $F_s/2^P = F_s/N$ [am/s]. A taxa total de amostras na saída de qualquer estágio, é constante e igual a F_s [am/s].
- Na reconstrução do espectro no receptor é necessária uma estrutura em árvore simétrica à do transmissor; em consequência são necessários "p" estágios com $N-1$ filtros reconstrotores do espectro.

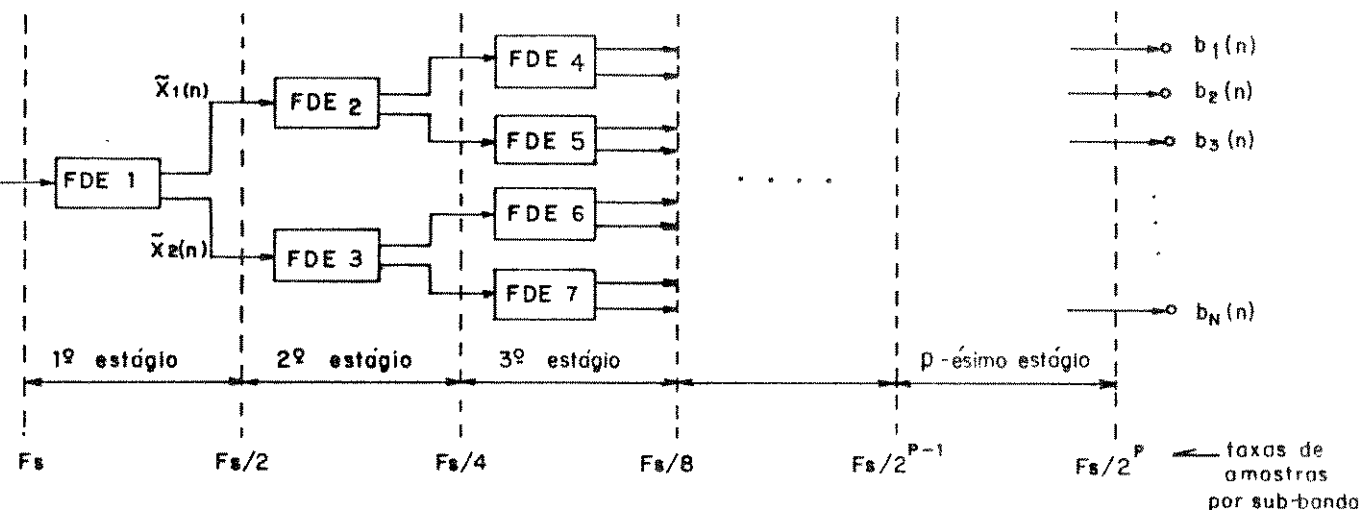


Figura 3.20: Divisão do espectro em 2^P sub-bandas de igual largura, utilizando-se uma estrutura em árvore de FDE.

Na Seção 3.3.1, Eq.(3.40), foi visto que quando $N=2$ sub-bandas, o sinal reconstruído era dado por:

$$y(n) = \frac{1}{2} x(n - (M-1)) \quad (3.51)$$

onde o fator $1/2$ era devido à perda de energia provocada pela dizimação do sinal no transmissor e $M-1$ era o atraso em amostras, entre o sinal reconstruído e o original, devido à utilização de filtros FIR de fase linear e M "taps". No caso de dividir-se o espectro do sinal de entrada em 2^p sub-bandas de igual largura, o sinal reconstruído será dado por:

$$y(n) = \frac{1}{2^p} x(n - \text{Atraso}) \quad (3.52)$$

onde

$$\text{Atraso [am]} = \sum_{k=1}^p 2^{k-1} \cdot (M_k - 1) \quad (3.53)$$

e M_k é a ordem (número de "taps"), dos filtros FIR do k -ésimo estágio.

A Eq.(3.53) surge do seguinte: o k -ésimo estágio do transmissor, assim como seu correspondente estágio simétrico no receptor, provoca um atraso de $(M_k-1)/2$ amostras sobre os sinais que entram nele. Assim, limitando-se aos atrasos produzidos no transmissor se terá:

Estágio 1: Atraso em amostras: $(M_1-1)/2$ [am].
 Atraso em segundos: $T \cdot (M_1-1)/2$ [s].

Estágio 2: Atraso em amostras: $(M_2-1)/2$ [am].
 Atraso em segundos: $2T \cdot (M_2-1)/2$ [s].

(As amostras de entrada ao segundo estágio estão separadas $2T$ [s], devido à dizimação produzida no primeiro estágio).

Estágio 3: Atraso em amostras: $(M_3-1)/2$ [am].
 Atraso em segundos: $4T \cdot (M_3-1)/2$ [s].

(As amostras de entrada ao terceiro estágio estão separadas $4T$ [s], devido à dizimação produzida no estágio anterior).

Continuando este raciocínio se obtém:

$$\begin{aligned} \text{Atraso Tx [s]} &= \frac{M_1-1}{2} T + \frac{M_2-1}{2} 2T + \frac{M_3-1}{2} 4T + \dots + \frac{M_P-1}{2} 2^{P-1} T \\ &= \sum_{k=1}^P \frac{M_k-1}{2} 2^{k-1} \cdot T \end{aligned} \quad (3.54a)$$

Fazendo um desenvolvimento similar para o receptor:

$$\begin{aligned} \text{Atraso Rx [s]} &= \frac{M_P-1}{2} 2^{P-1} T + \frac{M_{P-1}-1}{2} 2^{P-2} T + \dots + \frac{M_1-1}{2} T \\ &= \text{Atraso Tx [s]} \end{aligned} \quad (3.54b)$$

Em consequência o atraso total será:

$$\begin{aligned} \text{Atraso [s]} &= \text{Atraso Tx [s]} + \text{Atraso Rx [s]} \\ &= 2 \cdot \sum_{k=1}^P \frac{M_k-1}{2} 2^{k-1} \cdot T = \sum_{k=1}^P (M_k - 1) \cdot 2^{k-1} \cdot T \end{aligned} \quad (3.55)$$

Expressando este atraso em amostras obtém-se:

$$\text{Atraso [am]} = \sum_{k=1}^P 2^{k-1} \cdot (M_k - 1) \quad (3.56)$$

que corresponde à Eq. (3.53).

Para finalizar, serão feitas algumas considerações sobre a carga computacional necessária para a divisão e reconstrução do espectro. Ao utilizarem-se filtros com igual número de "taps" em todos os estágios, a divisão e reconstrução completa do espectro requer pM multiplicações e pM adições a cada período de amostragem $T = 1/F_s$, onde $M = M_1 = M_2 = \dots = M_p$ é o número de "taps" de cada filtro. Porém, para um dado "ripple" na banda passante e uma dada atenuação

na banda de rejeição, verifica-se que (Ref.[8]):

$$\Delta F \cdot M_k / F_s \cong \text{constante} \quad (3.57)$$

onde ΔF é a largura normalizada da banda de transição dos filtros, M_k o número de taps dos filtros do k-ésimo estágio e F_s a taxa de amostras à saída desse estágio. Dado que F_s diminui pela metade a cada estágio, é possível também se diminuir pela metade o número de "taps" dos filtros de um estágio para outro, mantendo-se aproximadamente constante a largura das bandas de transição dos filtros. Neste caso, como $M_k = M_{k-1}/2$, a carga computacional para a divisão e reconstrução completa do espectro se reduz a $2M_1 \cdot (1 - 1/N)$ multiplicações e $2M_1 \cdot (1 - 1/N)$ adições a cada intervalo de amostragem $T = 1/F_s$, onde M_1 é o número de "taps" dos filtros do primeiro estágio e N é o número de sub-bandas em que se divide o espectro.

3.3.4 DIVISÃO DO ESPECTRO EM OITAVAS UTILIZANDO-SE FILTROS QMF

Às vezes é desejável dividir-se o espectro do sinal de entrada $x(n)$ em sub-bandas de largura diferente. Neste caso, a estrutura de decomposição em árvore pode ser feita só em alguns ramos, obtendo-se assim sub-bandas com a metade da largura das outras pertencentes ao mesmo estágio. Assim por exemplo, para se dividir o espectro $X(e^{j\omega})$ em 3 sub-bandas, poderia usar-se o esquema da Fig.3.21(a). Desta figura pode-se ver que o FDE faltante no ramo inferior do segundo estágio, é substituído por um filtro $C(z)$ onde $C(z) = 1/2 \cdot [H_2^2(z) - H_2^2(-z)]$, permitindo, desta forma, uma perfeita reconstrução do espectro (Ref.[8]). Normalmente, ao invés de $C(z)$ é utilizado apenas um atraso de $M_2 - 1$ [am], que

compensa os atrasos do FDE faltante no transmissor e do FRE faltante no receptor. Isto não introduz qualquer degradação audível no sinal reconstruído. A Fig. 3.21(b) mostra o esquema da resposta em frequência do sistema mencionado.

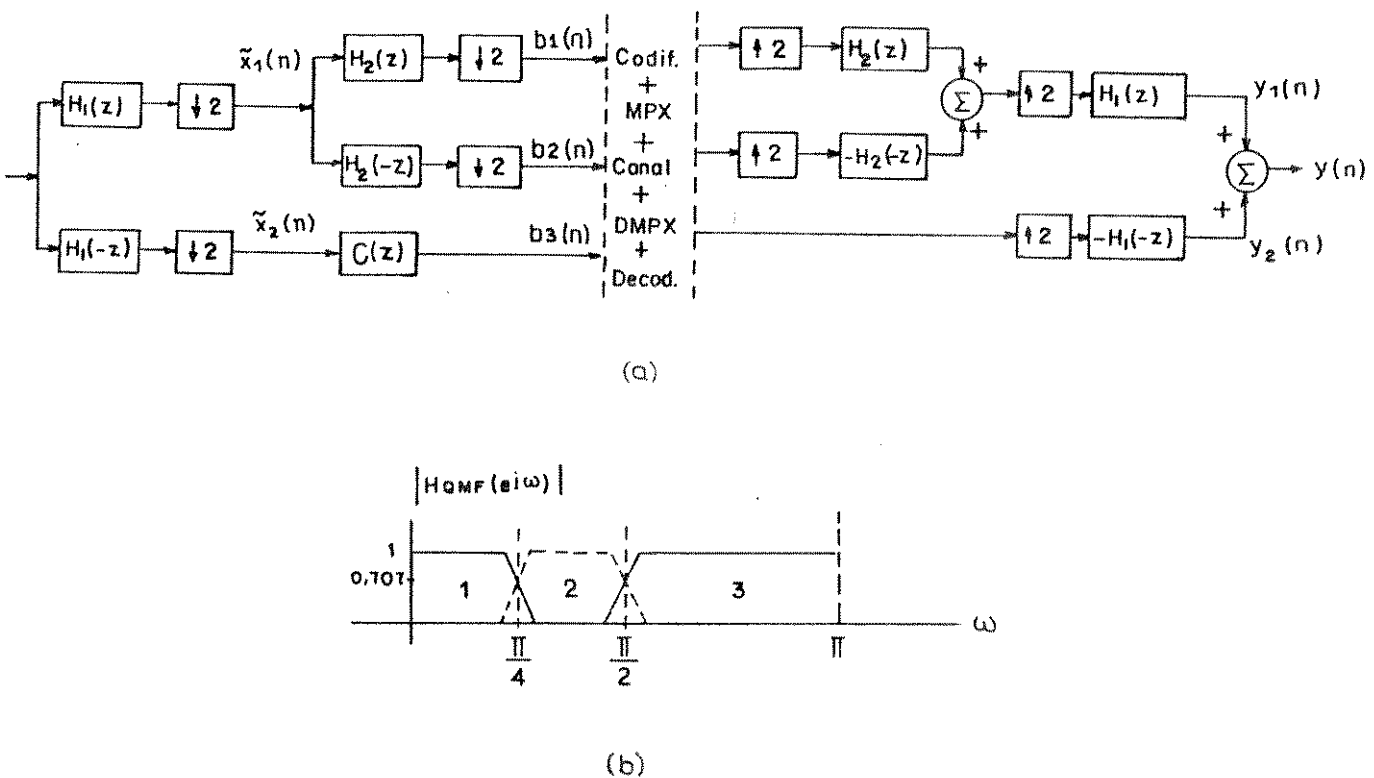


Figura 3.21: Divisão do espectro em 3 sub-bandas utilizando-se filtros espelho em quadratura: (a) Diagrama em blocos do sistema; (b) Resposta em frequência da estrutura filtrante.

Baseada na técnica explicada acima, é possível a divisão do espectro em oitavas utilizando-se FDE só nos ramos superiores da estrutura em árvore tradicional. Um exemplo desta estrutura é dado na Fig. 3.22. Segundo pode-se ver nesta figura, os filtros divisores de espectro faltantes nos ramos inferiores, são substituídos pelos atrasos LA1 e LA2, onde $LA1 = (M_2 - 1) + 2(M_3 - 1)$ [am] e $LA2 = M_3 - 1$ [am].

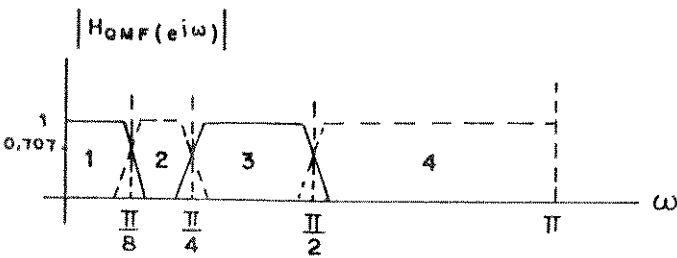
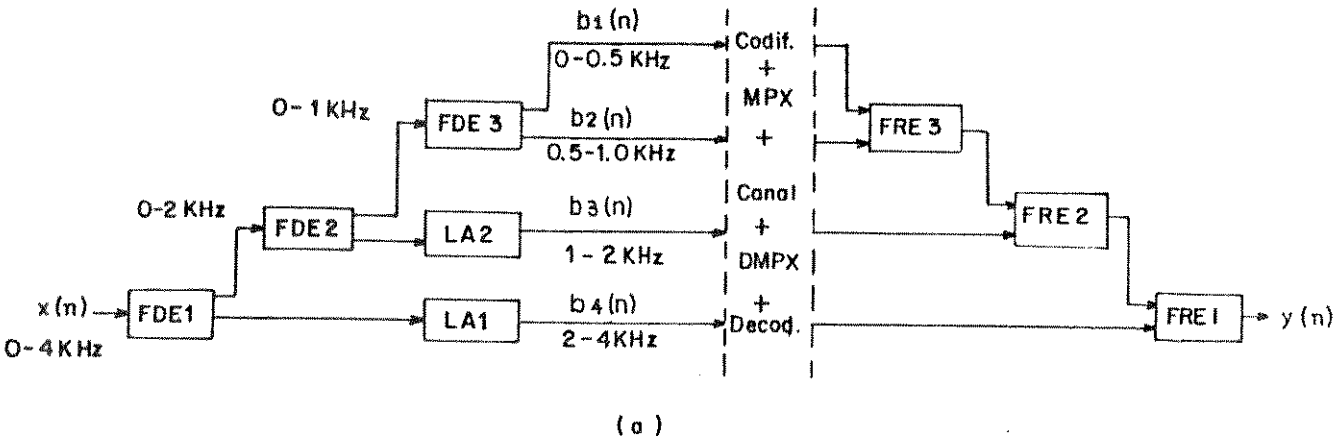


Figura 3.22: Divisão do espectro em oitavas: (a) Diagrama em blocos do sistema; (b) Resposta em frequência da estrutura filtrante.

A tabela seguinte mostra a largura das diferentes sub-bandas obtidas com o esquema da Fig.3.22:

Banda	Faixa de frequências [Hz]		
1	0	a	500
2	500	a	1000
3	1000	a	2000
4	2000	a	4000

Tabela 3.4: Largura das diferentes sub-bandas na divisão do espectro em oitavas.

Apesar da primeira e segunda sub-bandas terem igual largura, normalmente o sinal de entrada é pré-filtrado através de um filtro passa-faixa com frequência de corte inferior entre os 200 e 300 [Hz], e uma frequência de corte superior entre os 3200 e 3400 [Hz], fazendo com que a largura da primeira sub-banda se torne aproximadamente a metade da largura da segunda sub-banda. Esta pré-filtragem permite eliminar as frequências mais baixas do espectro que possuem elevada energia e portanto produzem alto ruído de quantização se não são quantizadas com precisão suficiente.

Para finalizar o capítulo, deve-se destacar que os bancos de filtros OMF satisfazem a condição "Integer Band Sampling" tanto na divisão do espectro em 2^P sub-bandas de igual largura, como na divisão do espectro em sub-bandas de largura diferente (exemplo: divisão em oitavas). Quanto à possibilidade de se criar "gaps" no espectro, isto só pode ser implementado cancelando-se completamente algumas das sub-bandas. Isto implica em não codificar nem transmitir essas sub-bandas. Neste caso, porém, surge o seguinte inconveniente: na reconstrução do espectro no receptor, não se dispõe das componentes espectrais pertencentes à sub-banda eliminada, que cancelam os termos de "aliasing" das sub-bandas contíguas (termos gerados durante a dizimação). Como consequência se produz a distorção de "aliasing" e o efeito de reverberância devido à presença do "gap". Estes efeitos serão mais ou menos perceptíveis, segundo a posição da banda eliminada.

REFERÊNCIAS

- [1] L.R. Rabiner and R.W. Schafer, "Digital Processing Of Speech Signals", Prentice Hall, Inc., Englewood Cliffs, NJ, 1978.

- [12] A.V.Oppenheim and R.W.Schafer, "Digital Signal Processing", Prentice Hall, Inc., Englewood Cliffs, NJ, 1975.
- [13] R.E.Crochiere, S.A.Webber and J.L.Flanagan, "Digital Coding of Speech in Sub-bands", Bell System Tech. J., Vol 55, Nº 8, October 1976, pp 1069-1085.
- [14] R.E.Crochiere, "On the Design of Sub-band Coders for Low-Bit-Rate Speech Communication", Bell System Tech. J., Vol 56, Nº 5, May-June 1977, pp 747-770.
- [15] K.D.Kryter, "Masking and Speech Communications in Noise", Academic Press, NY, 1970.
- [16] R.E.Crochiere and M.R.Sambur, "A Variable-Band Coding Scheme for Speech Encoding at 4.8 Kb/s", Bell System Tech. J., Vol 56, Nº 5, May-June 1977, pp 771-779.
- [17] D.Esteban and C.Galand, "Application of Quadrature Mirror Filters to Split Band Voice Coding Schemes", Proc. ICASSP, May 1977, pp 191-195.
- [18] C.Galand and H.Nussbaumer, "New Quadrature Mirror Filter Structures", IEEE Trans. on ASSP, Vol ASSP-32, Nº 3, June 1984, pp 522-531.
- [19] A.Crosier, D.Esteban and C.Galand, "Perfect Channel Splitting by Use of Interpolation/Decimation/Tree Decomposition Techniques", Int'l Conf. on Inform. Sciences and Systems, Patras, Greece, 1976, pp 443-446.
- [10] M.J.Smith and T.P.Barnwell III, "Exact Reconstruction Techniques for Tree-Structured Sub-band Coders", IEEE, Trans on ASSP, Vol ASSP-34, Nº 3, June 1986, pp 434-441.

-
- [11] P.P.Vaidyanathan and Phuong-Quan Hoang, "Lattice Structures for Optimal Design and Robust Implementation of Two Channel Perfect-Reconstruction QMF Banks", IEEE on ASSP, Vol ASSP-36, No 1, January 1988, pp 81-94.
- [12] P.Chu, "Quadrature Mirror Filter Design for an Arbitrary Number of Equal Bandwidth Channels", IEEE on ASSP, Vol ASSP-33, No 1, February 1985, pp 203-218.
- [13] J.D.Johnston, "A Filter Family Designed for Use in Quadrature Mirror Filter Banks", Conf.Rec., ICAASP, Denver, Co., 1980, pp 291-294.
- [14] R.Hooke and T.Jeaves, "Direct Search Solution of Numerical and Statistical Problems", J.Ass. Comp. Mach., Vol 8, April 1961, pp 212-229.

CAPÍTULO 4

CODIFICAÇÃO DOS SINAIS NOS CODIFICADORES DE SUB-BANDAS

CONTEÚDO

4.1 - Introdução	65
4.2 - Tipos de quantizador	66
4.2.1 - Quantizadores uniformes	67
4.2.2 - Quantizadores não uniformes	68
4.2.3 - Quantizadores uniformes e não uniformes para uma relação sinal/ruído de quantização ótima	69
4.2.4 - Quantizadores adaptativos	70
4.2.4.1 - Adaptação AQF	72
4.2.4.2 - Adaptação AQB	74
4.2.5 - Quantizadores fixos	75
4.3 - Características dos sinais de voz	75
4.4 - Codificação dos sinais nas diferentes sub-bandas	81
4.4.1 - Codificador BC-PCM	81
4.4.2 - Cálculo da taxa de informação lateral nos codificadores BC-PCM	89
4.4.3 - Codificador APCM-AQB	92
4.4.4 - Algoritmo robusto de GOODMAN e WILKINSON	96
4.5 - Atribuição de bits por banda e utilização de BC-PCM ou APCM-AQB	98
REFERÊNCIAS	100

4.1 INTRODUÇÃO

Nos codificadores de sub-bandas simulados neste trabalho, foram utilizados basicamente 2 métodos para codificar os sinais das diferentes sub-bandas: o método BC-PCM (Block Companded PCM), e o método APCM (Adaptive PCM).

Para se chegar a esta escolha foi preciso analisar os diferentes tipos de codificação possíveis, assim como as características dos sinais a codificar. Este capítulo é dedicado à análise destes pontos.

O capítulo começa com um estudo dos diferentes quantizadores usados nos codificadores tradicionais, para depois passar a analisar quais são as características dos sinais de voz em banda integral e em sub-bandas. São então explicados em detalhe os 2 tipos de codificadores escolhidos, BC-PCM e APCM, com as respectivas vantagens e desvantagens que apresentam. Para o caso de codificação APCM, são apresentados os 2 algoritmos de adaptação do passo de quantização utilizados, sendo que um deles é robusto a erros de canal.

O capítulo finaliza com algumas diretrizes que auxiliam na atribuição do número de bits a cada sub-banda, assim como na escolha do método de codificação (BC-PCM ou APCM), para uma determinada sub-banda.

4.2 TIPOS DE QUANTIZADOR

Uma primeira classificação dos quantizadores, que serão representados por $Q[]$, pode ser feita em função da distribuição de seus níveis de quantização e intervalos de decisão. Sob este ponto de vista os quantizadores se classificam em *uniformes* e *não uniformes*.

Uma outra forma de classificação pode ser feita em função das características do quantizador se alterarem ou não, em função das variações estatísticas do sinal de entrada, tentando se adaptar a elas. Sob este ponto de vista os quantizadores podem ser classificados em *adaptativos* e *fixos*.

Já que todos os tipos mencionados acima são de interesse para este trabalho, será feita uma breve descrição dos mesmos.

4.2.1 QUANTIZADORES UNIFORMES

Estes quantizadores se caracterizam por terem seus níveis de quantização e intervalos de decisão uniformemente espaçados. A Fig. 4.1 dá um exemplo de quantizador uniforme do tipo "Mid-Riser" com 8 níveis (3 [bits] para quantizar cada amostra). O tipo "Mid-Riser" se caracteriza por não possuir nível 0.

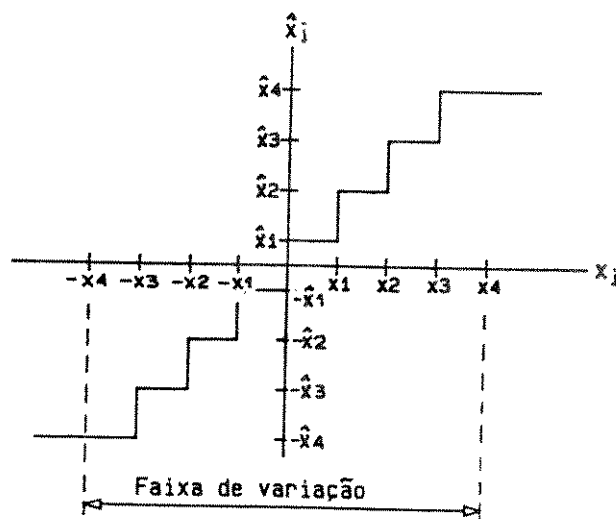


Figura 4.1: Quantizador uniforme de 8 níveis do tipo "Mid-Riser".

Nesta figura pode-se ver que: $x_j - x_{j-1} = \Delta = \text{constante}$ e $\hat{x}_j - \hat{x}_{j-1} = \Delta = \text{constante}$, onde Δ é o passo de quantização do quantizador e $\hat{x}_j = Q[x_j]$.

Nos quantizadores uniformes, a variância σ_e^2 do erro de quantização é constante e independente do nível do sinal de entrada, desde que este não exceda a faixa de amplitudes do quantizador. O erro de quantização, $e(n)$, é definido como sendo a diferença entre $x(n)$ e seu valor quantizado $Q[x(n)]$:

$$e(n) = x(n) - Q[x(n)] \quad (4.1)$$

Para $E[e(n)] = 0$ tem-se:

$$\sigma_e^2 = E[e^2(n)] = E[(x - Q[x])^2] \quad (4.2)$$

onde o operador $E[\]$ indica média estatística.

Uma medida de desempenho do quantizador é dada pela "relação sinal/ruído de quantização", definida como o quociente entre a variância do sinal a quantizar, σ_x^2 , e a variância do erro de quantização σ_e^2 , isto é:

$$SNR_q \text{ [dB]} = 10 \cdot \log_{10} (\sigma_x^2 / \sigma_e^2) \quad (4.3)$$

O fato dos quantizadores uniformes possuírem σ_e^2 aproximadamente constante e independente do nível do sinal de entrada, faz com que a relação sinal/ruído do quantizador diminua com a diminuição do nível do sinal a quantizar (Eq.4.3). Consequentemente a relação sinal/ruído varia com σ_x , ocorrendo o seu máximo quando o sinal de entrada ocupa toda a faixa do quantizador.

4.2.2 QUANTIZADORES NÃO UNIFORMES

Nestes quantizadores os níveis de quantização e os intervalos de decisão não estão uniformemente espaçados. O

objetivo de um quantizador não uniforme é aumentar a faixa dinâmica do mesmo, mantendo a relação sinal/ruído de quantização aproximadamente constante sobre uma ampla faixa de amplitudes do sinal de entrada (exemplo: quantizadores logarítmicos $Le\text{-}A$ ou $Le\text{-}\mu$). Para conseguir este objetivo, deve-se fazer com que a variância do erro de quantização σ_e^2 seja função do nível do sinal a quantizar. Assim nos quantizadores não uniformes a relação sinal/ruído de quantização é aproximadamente constante com a variação de σ_x . A Fig.4.2 mostra um exemplo de quantizador não uniforme "Mid-Riser" de 8 níveis.

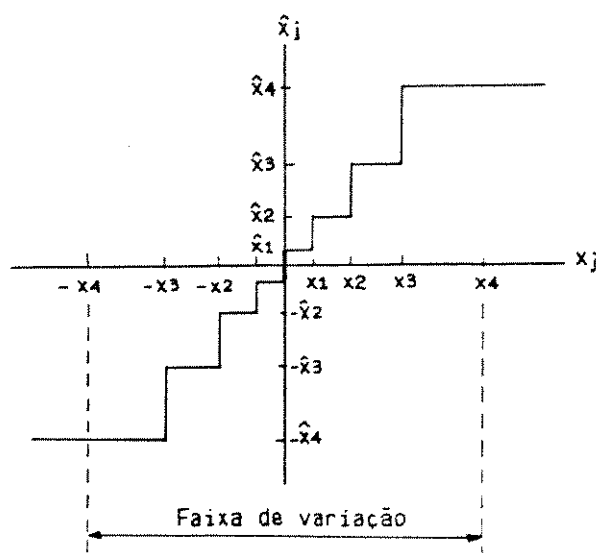


Figura 4.2: Quantizador não uniforme de 8 níveis do tipo "Mid-Riser".

4.2.3 QUANTIZADORES UNIFORMES E NÃO UNIFORMES PARA UMA RELAÇÃO SINAL/RUÍDO DE QUANTIZAÇÃO ÓTIMA

Das seções anteriores, tem-se que:

(a) Nos quantizadores uniformes a SNR_q varia em função do

nível do sinal de entrada (σ_x), já que a variância do erro de quantização é constante.

- (b) Nos quantizadores não uniformes a SNR_q é aproximadamente constante, já que a variância do erro de quantização acompanha a variação de nível do sinal de entrada.

Na hipótese de se conhecer a "*função densidade de probabilidade*" (PDF) do sinal que vai ser quantizado, é possível se projetar um quantizador uniforme ou não uniforme (calculando-se seus níveis de quantização e intervalos de decisão), de modo a se minimizar a variância do erro de quantização e, conseqüentemente, maximizar a SNR_q . Estes quantizadores se denominam "*ótimos*" e, como mencionado, podem ser uniformes ou não uniformes, sendo que os últimos possuem um desempenho superior aos primeiros. Tem que ser ressaltado, porém, que o valor máximo da SNR_q ocorre para uma determinada variância do sinal de entrada e, portanto, é necessário que estes quantizadores sejam adaptativos para se conseguir uma SNR_q aproximadamente constante sobre uma ampla faixa de amplitudes do sinal a quantizar. No Apêndice B são apresentados os resultados obtidos por Max (Ref.[1]), no projeto de quantizadores ótimos para sinais de entrada gaussianos, e por Paez e Glisson (Ref.[2]), para sinais com densidade de probabilidade laplaciana e gama.

4.2.4 QUANTIZADORES ADAPTATIVOS

Quantizadores adaptativos são aqueles que mudam suas características de modo a adaptá-las àquelas do sinal que está sendo quantizado. Desta forma consegue-se uma melhora no desempenho do quantizador ou, para um mesmo desempenho, um aumento na sua faixa dinâmica.

Das várias adaptações possíveis, nos interessa apenas aquela que permite adaptar a faixa de amplitudes do quantizador às "variações de amplitude" do sinal a quantizar. Para isso se faz uma estimativa periódica do nível do sinal que está entrando no quantizador e, em função dessa estimativa, se comprime ou se expande linearmente a função de transferência do quantizador, de modo a adaptá-la ao sinal de entrada.

A estimativa do nível do sinal a quantizar é feita calculando-se periodicamente algum parâmetro a ele relacionado, como, por exemplo, o desvio padrão σ_x , a variância σ_x^2 , a média dos valores absolutos das amostras, $E[|x(n)|]$ ou a amostra máxima correspondente $|x(n)|_{\max}$. Este parâmetro será denominado "parâmetro de adaptação do quantizador".

A Fig. 4.3(a) mostra o esquema básico dos quantizadores adaptativos que utilizam o método de adaptação AQF (Adaptive Quantization Forward), ao passo que a Fig. 4.3(b) mostra o esquema que utiliza o método de adaptação AQB (Adaptive Quantization Backward). No primeiro caso, AQF, o cálculo do parâmetro de adaptação é baseado no valor das amostras futuras ainda não quantizadas, enquanto que no segundo caso, AQB, o cálculo é baseado no valor das amostras já quantizadas.

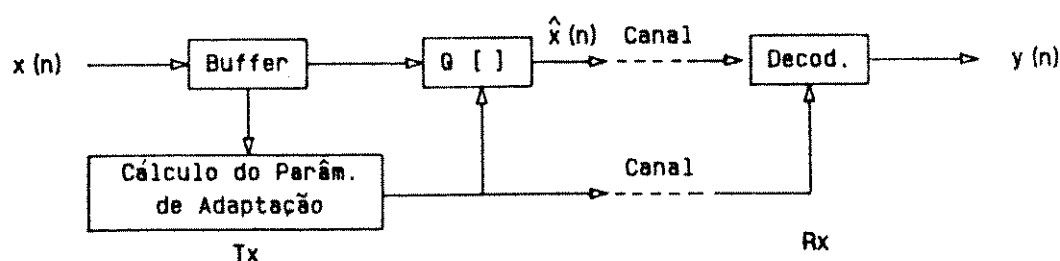


Figura 4.3(a): Quantizador adaptativo AQF (adaptive quantization forward).

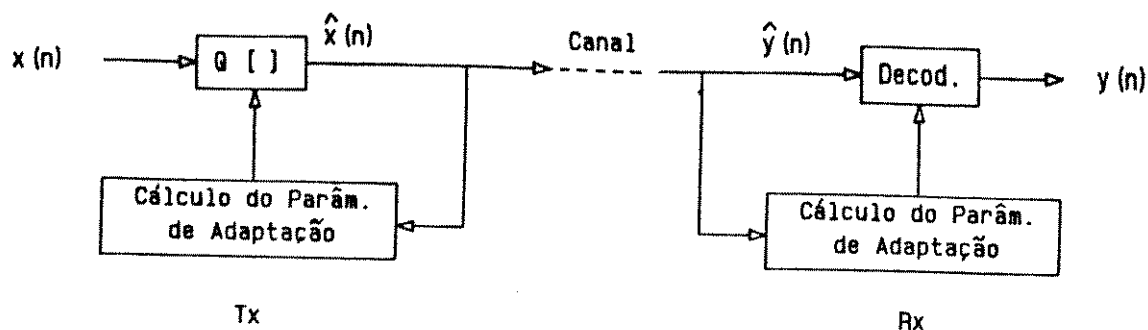


Figura 4.3(b): Quantizador adaptativo AQB (adaptive quantization backward).

4.2.4.1 Adaptação AQF

Nestes sistemas o cálculo do parâmetro de adaptação do quantizador é mais confiável, já que ao basear-se em amostras não quantizadas o cálculo não é afetado pelo ruído de quantização. Há porém uma desvantagem pois, como essas amostras não estão disponíveis no receptor para realizar o mesmo cálculo, deve-se transmitir ao receptor o parâmetro de adaptação calculado. Isto implica na transmissão de informação lateral além das amostras quantizadas. Como consequência aumenta a taxa total de bits do codificador, o que pode ser crítico quando se desejam baixas taxas. Por outro lado, protegendo-se adequadamente os bits de informação lateral, garante-se que o decodificador do receptor seja adaptado sem erros.

Um outro problema do sistema AQF é o atraso inerente que provoca no sistema. Se o cálculo do parâmetro de adaptação é baseado nas próximas P amostras a quantizar, deve-se esperar a chegada de todas elas (armazenando-as num "buffer"), para poder-se calcular o parâmetro de adaptação. Só então o quantizador é adaptado e as amostras quantizadas e

transmitidas. Este atraso pode ou não ser importante, dependendo da aplicação desejada.

O atraso pode ser fixado no projeto do sistema, definindo-se a extensão P do bloco de amostras no qual basear-se-a o cálculo do parâmetro de adaptação. Deve-se considerar que:

- 1) Escolhendo-se P grande, a taxa de informação lateral diminui já que o quantizador é adaptado menos frequentemente, mas por esta mesma razão ele não consegue seguir de perto as variações estatísticas do sinal de entrada. Além disto, o atraso do sinal é considerável.
- 2) Escolhendo-se P pequeno, o quantizador segue de perto as variações do sinal a quantizar, e portanto a quantização é mais precisa. Porém a taxa de informação lateral aumenta devido às adaptações mais frequentes do quantizador.

Nos codificadores utilizados neste trabalho, a duração do bloco de amostras foi fixada em 16 [ms], já que para tal intervalo de tempo os sinais de voz podem ser considerados aproximadamente estacionários. Para sinais amostrados a 8 [KHz], isto implica em blocos de 128 amostras ($P=128$).

A Tabela 4.1 compara os desempenhos de vários quantizadores ótimos, fixos e adaptativos (AQF) de 8 níveis, testados com sinais de voz reais (Ref. [3]). A diferença entre os valores da primeira coluna (quantizadores fixos), quando comparados com os fornecidos no Apêndice B para os mesmos quantizadores, é devida ao fato de que, neste caso, se utilizam sinais de voz reais para testá-los. No Apêndice B o teste é realizado com sinais de densidade de probabilidade gaussiana e laplaciana respectivamente.

Pode-se ver que, no caso de quantizadores projetados para sinais de entrada gaussianos, o fato de usarem-se

quantizadores adaptativos dobra a SNR_q . Também pode-se observar que o desempenho dos quantizadores adaptativos "uniformes e não uniformes" é aproximadamente o mesmo. Este resultado sugere o uso de quantizadores uniformes, já que são bem mais simples de implementar.

Quantizador não uniforme	Fixo (SNR_q em dB)	Adaptativo (SNR_q em dB)
Gaussiano	7.3	15.0
Laplaciano	9.9	13.3
Quantizador Uniforme		
Gaussiano	6.7	14.7
Laplaciano	7.4	13.4

Tabela 4.1: Desempenho de quantizadores ótimos uniformes e não uniformes, para sinais de voz reais.

4.2.4.2 Adaptação AQB

Nestes sistemas (Fig.4.3(b)), o cálculo do parâmetro de adaptação é baseado nas amostras de saída do quantizador. A grande vantagem deste método, é que não é necessário se transmitir o parâmetro de adaptação do quantizador ao receptor, já que ele pode ser determinado a partir das amostras de entrada ao mesmo. Adicionalmente não existe atraso na transmissão da informação, como ocorre no método AQF. A desvantagem, por outra parte, é que o cálculo do parâmetro de adaptação é afetado pelo ruído de quantização das amostras e, portanto, é menos preciso que no método AQF.

Além do mais, este sistema é mais sensível a erros de canal que o método AQF, já que erros em algumas das amostras que chegam ao receptor, provocam erros no cálculo do parâmetro de adaptação do decodificador. Neste caso tanto as amostras corretas como as erradas serão decodificadas erroneamente.

Normalmente o quantizador adaptativo AQB adapta sua faixa dinâmica baseando-se na variância do sinal de entrada, cálculo esse feito em função das amostras passadas e já quantizadas desse sinal.

4.2.5 QUANTIZADORES FIXOS

Quantizadores fixos são aqueles cujas características permanecem constantes, independentemente das variações estatísticas do sinal de entrada. Para sinais com grandes variações de nível (ampla faixa dinâmica), o desempenho destes quantizadores é inferior ao dos quantizadores adaptativos. Isto acontece especialmente a baixas taxas, onde a quantidade média de bits utilizados na quantização das amostras é baixa. Para sinais de voz amostrados a 8 [KHz] e com taxas de bits inferiores a 20 [Kbits/s], a quantidade média de bits por amostra é menor que 3.

4.3 CARACTERÍSTICAS DOS SINAIS DE VOZ

A escolha correta dos codificadores e dos parâmetros de codificação mais adequados para codificar os sinais de voz, é determinada, em parte, pelas características de longo prazo destes sinais e, em parte, pelas características de curto prazo. Portanto se faz necessário explicar brevemente algumas destas características.

Os sinais de voz são sinais não estacionários. Sua variância de curto prazo pode variar grandemente num intervalo de até 40 [dB] aproximadamente, em função da pessoa que fala, do estado de ânimo da mesma, do ambiente onde se realiza a comunicação e, ainda dentro de uma mesma sentença, ao se passar de sons sonoros a sons não sonoros. Por outro lado, a frequência de ocorrência das diferentes amplitudes da voz, depende de se estar considerando segmentos de voz de curta duração ou de longa duração. Assim, ao se considerar segmentos de curta duração (entre 10 e 20 [ms]), onde os sinais de voz podem ser ditos aproximadamente estacionários, a função densidade de probabilidade das amplitudes de voz é quase gaussiana (Ref.[4]). No entanto, ao se considerar segmentos de voz de duração longa (maiores a 0,5 [s]), a função densidade de probabilidade dos sinais de voz é melhor aproximada por uma função gama com a forma:

$$p(x) = \left[\frac{\sqrt{3}}{8\pi \cdot \sigma_x \cdot |x|} \right]^{1/2} \cdot e^{-\frac{\sqrt{3} \cdot |x|}{2 \cdot \sigma_x}} \quad (4.4)$$

onde σ_x corresponde à raiz quadrada da variância do sinal de voz (desvio padrão).

Uma aproximação mais simples da Eq.(4.4), pode ser obtida com a função densidade de probabilidade laplaciana dada por:

$$p(x) = 1/(\sqrt{2} \cdot \sigma_x) \cdot e^{-\frac{\sqrt{2} \cdot |x|}{\sigma_x}} \quad (4.5)$$

A Fig.4.4 mostra as distribuições correspondentes a estas duas funções de densidade de probabilidade, junto com a obtida experimentalmente para um sinal de voz real. As curvas foram normalizadas, de modo que a variância é unitária e a

média nula (Ref.[2]). Nesta figura pode-se ver que a função gama é uma excelente aproximação para a voz real. Entretanto, às vezes é conveniente se reemplazar a função gamma pela de laplace para facilitar os cálculos.

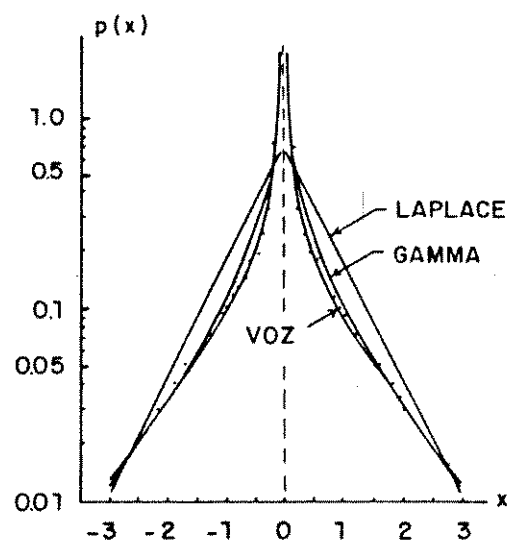


Figura 4.4: Função densidade de probabilidade de longo prazo para um sinal de voz real e para as distribuições gama e de Laplace.

Quanto à correlação de longo prazo entre amostras, a Fig.4.5 mostra os valores obtidos por Noll para sinais de voz reais amostrados a 8 [KHz] e submetidos a filtragem passa-baixas e passa-banda (Ref.[5]). No primeiro caso (curvas superiores), os sinais foram limitados em 3400 [Hz], enquanto que no segundo caso (curvas inferiores), os sinais foram filtrados por um filtro passa-banda com frequências de corte em 300 e 3400 [Hz]. As regiões sombreadas ao redor dos valores base, são devidas às variações existentes entre diferentes locutores. Ainda nesta figura, pode-se ver que a correlação entre amostras adjacentes é alta, mas diminui rapidamente para espaçamentos maiores.

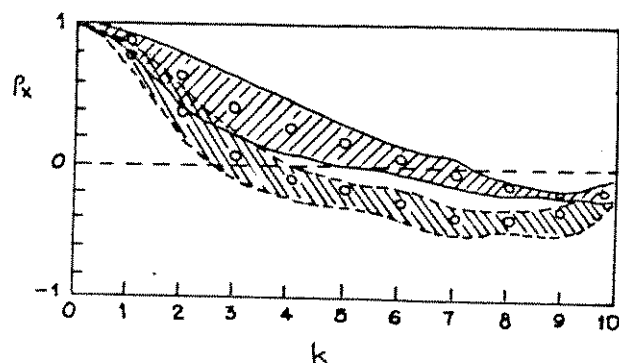


Figura 4.5: Função de autocorrelação de longo prazo para sinais de voz (valores normalizados: p_x). Curvas superiores para sinais passa-baixas e curvas inferiores para sinais passa-banda.

Finalmente para a densidade espectral de potência de longo prazo, a Fig. 4.6 mostra os resultados de estudos feitos por Beranek (Ref.[6]) e por Dunn e White (Ref.[7]), que indicam a existência de um máximo na faixa dos 400 a 600 [Hz] e uma posterior queda com inclinação de 8 a 10 [dB/oitava].

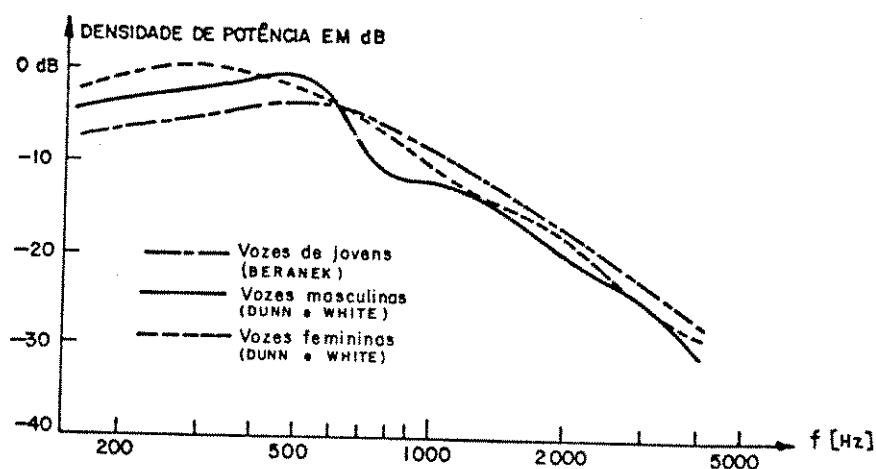


Figura 4.6: Densidade espectral de potência de longo prazo para vozes masculinas, femininas e de jovens.

Com estes dados e o que foi visto na Seção 4.2, é possível tirar algumas conclusões:

- 1) Para baixas taxas (≤ 3 [bits/am]), e devido à ampla faixa dinâmica dos sinais de voz, os quantizadores fixos (por exemplo logarítmicos), não conseguem uma boa solução de compromisso entre faixa dinâmica e relação sinal/ruído de quantização. Portanto, existe a necessidade de se utilizar quantizadores adaptativos para se alcançar este objetivo. Os quantizadores fixos só conseguem uma ampla faixa dinâmica e elevada SNR_Q, utilizando maior quantidade de bits por amostra.
- 2) Ao se usar quantizadores adaptativos, a máxima relação sinal/ruído é obtida com quantizadores ótimos projetados para sinais de entrada gaussianos. Isto porque, neste caso, a adaptação do passo de quantização deve ser feita a intervalos de tempo curtos (menores a 20 [ms]), e para tais intervalos de tempo a função densidade de probabilidade da voz é aproximadamente gaussiana.
- 3) Ao se usar quantizadores fixos, a máxima relação sinal/ruído é obtida com quantizadores ótimos projetados para sinais de entrada com densidade de probabilidade segundo a função gama. O motivo é que ao não ser feita a adaptação do quantizador, os quantizadores ótimos para sinais gama se adaptam melhor à densidade de probabilidade de longo prazo dos sinais de voz. Se o número de bits por amostra é elevado (> 6), poderiam, como alternativa, ser utilizados quantizadores logarítmicos (Lei-A ou Lei- μ).
- 4) A alta correlação entre amostras adjacentes permite que sejam utilizados com vantagens quantizadores diferenciais (aqueles que quantizam a diferença entre o valor da amostra e uma predição desse valor). Com isto se obtém um

aumento da SNR_Q ou, para um mesmo valor de SNR_Q , se consegue utilizar uma menor quantidade de bits. É conveniente se ressaltar que este resultado vale para sinais de voz em "banda integral", onde a correlação entre amostras é alta.

Ao se dividir o espectro do sinal de voz em sub-bandas, algumas destas características se mantêm, ao passo que outras se modificam. Principalmente no que se refere à correlação entre amostras, aparece uma diferença substancial. O fato de se dividir o espectro em sub-bandas e dizimar depois os sinais obtidos, faz com que o intervalo de tempo entre amostras pertencentes a uma mesma sub-banda aumente consideravelmente. Por exemplo, para um sinal de voz amostrado a 8 [KHz] e dividido em 4 sub-bandas, as amostras do sinal original estão espaçadas de $T = 125 [\mu s]$, ao passo que as pertencentes a cada sub-banda estão espaçadas em $4T$. Se N é igual a 8 sub-bandas, então o espaçamento sobe a $8T$ devido à dizimação adicional por um fator 2. Este aumento de tempo entre amostras sucessivas, faz com que a correlação entre elas desça a níveis muito baixos. Assim, a utilização de codificadores diferenciais não é vantajosa para a codificação de sub-bandas, ao contrário do que ocorre para sinais em banda integral.

O resultado acima pode também ser concluído a partir da Fig. 4.6. O espectro de densidade de potência da voz tem uma característica passa-baixas. Porém, ao longo de qualquer sub-banda a queda na densidade de potência é relativamente pequena. Como estas sub-bandas são re-amostradas à taxa de Nyquist (dizimadas), elas praticamente aparecem como tendo espectro plano e, portanto, a correlação entre amostras quase não existe.

A Fig. 4.6 também permite prever como se distribuirão entre as diferentes sub-bandas, os bits disponíveis para a

codificação das mesmas. Devido à maior energia existente nas sub-bandas baixas, um erro de quantização nessas sub-bandas será provavelmente de maior energia que um erro nas bandas altas. Portanto erros nas sub-bandas baixas contribuem mais ao ruído de quantização total. Para minimizar o valor destes erros, há necessidade de se codificarem estas bandas com maior precisão que as altas, utilizando-se para isso mais bits por amostra.

Com relação às funções densidade de probabilidade de curto e longo prazo dos sinais pertencentes às diversas sub-bandas, tem-se que, como a divisão em sub-bandas é feita por filtragem linear, estas funções não mudam e continuam sendo aproximadamente gaussianas e laplacianas respectivamente.

Resumindo, com excessão do que foi dito a respeito do uso de codificadores diferenciais, as outras conclusões antes obtidas são válidas tanto para sinais de voz em banda integral, quanto para sinais de sub-bandas.

4.4 CODIFICAÇÃO DOS SINAIS NAS DIFERENTES SUB-BANDAS

Baseados numa análise das seções anteriores, os codificadores escolhidos para codificar os sinais das diferentes sub-bandas foram o "*Block-Companded PCM*" (BC-PCM) e o "*Adaptive-PCM*" com adaptação AQB (APCM-AQB).

Deve-se ressaltar que para cada sub-banda existe um par codificador/decodificador (no transmissor e receptor respectivamente).

4.4.1 CODIFICADOR BC-PCM

Este codificador consiste, basicamente, de um

quantizador adaptativo que utiliza o método de adaptação AQF como visto na Seção 4.2.4.1 (Fig. 4.3(a)). Seu princípio de funcionamento, pode ser resumido da seguinte maneira:

- As amostras são codificadas em blocos. Para cada bloco de P amostras a quantizar, é calculado o parâmetro de adaptação do quantizador de modo que a maior das amostras do bloco caiba dentro da faixa de amplitudes do quantizador. Desta forma o parâmetro de adaptação do quantizador consiste da amostra máxima do bloco a ser quantizado, valor denominado $X_{k_{max}}$.
- O quantizador é adaptado em função desse parâmetro e as amostras são então quantizadas e transmitidas.
- O parâmetro de adaptação do quantizador é transmitido junto com as amostras quantizadas, para se poder realizar a mesma adaptação no decodificador do receptor.

As principais vantagens deste método de codificação são: ampla faixa dinâmica e ausência de "clipping" e "slope overload". Sua desvantagem é a necessidade de se transmitir o parâmetro de adaptação do quantizador como informação lateral, de forma a adaptar o decodificador do receptor. Isto diminui o número de bits destinados a quantizar as amostras, já que agora uma parte dos bits deve-se destinar à transmissão deste parâmetro. O valor da taxa de informação lateral será função da quantidade de bits destinados a codificar o parâmetro de adaptação do quantizador e da extensão P do bloco de amostras no qual se baseia o cálculo de tal parâmetro (Ref.[8]).

Para o caso de sinais de voz, a utilização de blocos de 10 a 20 [ms] se mostrou razoável, já que para tal intervalo de tempo os sinais de voz podem ser considerados estacionários. Nos codificadores de sub-bandas simulados

neste trabalho, foram utilizados blocos de 16 [ms], com o conseqüente atraso de 16 [ms] no sinal a transmitir.

A Fig.4.7 mostra o princípio do método BC-PCM para o caso de se adaptar um quantizador de 8 níveis a cada bloco de 10 amostras.

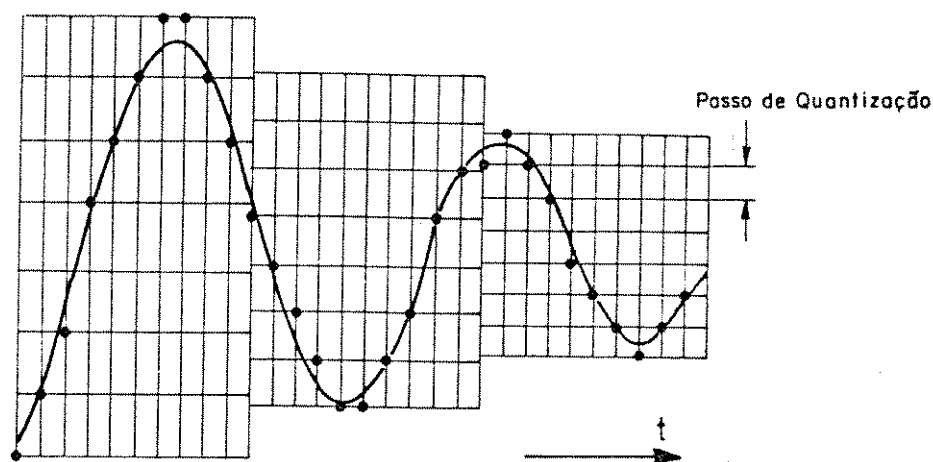


Figura 4.7: Princípio do método BC-PCM, para um quantizador de 8 níveis e blocos de 10 amostras.

Quanto ao número de bits utilizado na codificação do parâmetro de adaptação do quantizador, devem ser feitas algumas considerações. Se são destinados poucos bits para este fim (3 [bits] ou menos), a adaptação torna-se grosseira, permitindo, porém, que um maior número de bits seja destinado a quantizar as amostras. Por outro lado, destinando-se muitos bits para codificar o parâmetro de adaptação (6 [bits] ou mais), a adaptação torna-se bem mais precisa, mas às custas de um aumento considerável, e as vezes inaceitável, da taxa de informação lateral. Um valor razoável para se codificar esta informação é 4 ou 5 [bits/bloco]. No caso de se utilizar

3 [bits/bloco], existirá uma tabela de 8 valores possíveis para o parâmetro de adaptação do quantizador. Para 4 [bits/bloco] este número de valores possíveis sobe a 16 e para 5 [bits/bloco] tal valor é de 32. Um dos problemas mais difíceis de se resolver é a determinação dos valores destas tabelas. A Fig.4.8 exemplifica o problema para o caso de se utilizar 3 [bits/bloco].

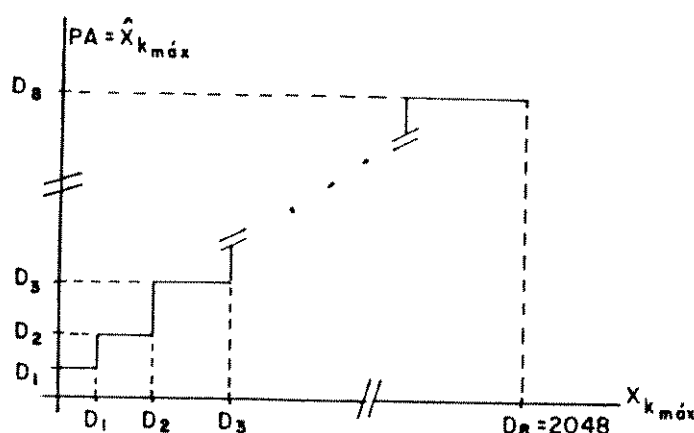


Figura 4.8: Determinação de $\hat{X}_{k_{\max}}$ em função de $X_{k_{\max}}$.

Da figura acima pode-se ver que o problema consiste na determinação das faixas de $X_{k_{\max}}$, que definem os valores do parâmetro de adaptação codificado $\hat{X}_{k_{\max}}$. Assim, se $X_{k_{\max}}$ cai no intervalo $[D_i, D_{i+1}]$ o parâmetro de adaptação escolhido será D_{i+1} . Neste trabalho, para a determinação dos valores das tabelas foram testadas diversas alternativas:

- divisão em faixas de igual largura;

- divisão supondo uma lei gaussiana para o parâmetro de adaptação (as faixas são determinadas de modo a assegurar igual probabilidade de ocorrência para cada valor do parâmetro de adaptação);
- divisão em faixas segundo uma lei logarítmica.

Os resultados obtidos com os valores assim determinados não foram, entretanto, satisfatórios, motivo pelo qual se fez uma segunda determinação das faixas, esta vez empírica, baseada na frequência com que aparecem os diferentes valores $X_{k_{max}}$. Intuitivamente as faixas deveriam ser mais estreitas perto da origem, tornando-se progressivamente maiores com o aumento de $X_{k_{max}}$. Assim o eixo $X_{k_{max}}$ foi dividido em faixas de largura crescente de forma que, em cada faixa, os $X_{k_{max}}$ ocorressem aproximadamente com igual probabilidade. Para determinar estas faixas foram utilizados vários arquivos de voz diferentes, chegando-se aos seguintes valores:

1) Codificação do parâmetro de adaptação com 3 [bits/bloco]:

D ₁ = 9	D ₅ = 340
D ₂ = 31	D ₆ = 470
D ₃ = 64	D ₇ = 660
D ₄ = 185	D ₈ = 2048

2) Codificação do parâmetro de adaptação com 4 [bits/bloco]:

D ₁ = 3	D ₅ = 42	D ₉ = 256	D ₁₃ = 560
D ₂ = 9	D ₆ = 64	D ₁₀ = 342	D ₁₄ = 660
D ₃ = 21	D ₇ = 105	D ₁₁ = 415	D ₁₅ = 882
D ₄ = 31	D ₈ = 185	D ₁₂ = 470	D ₁₆ = 2048

3) Codificação do parâmetro de adaptação com 5 [bits/bloco]:

$D_1 = 2$	$D_9 = 40$	$D_{17} = 350$	$D_{25} = 710$
$D_2 = 6$	$D_{10} = 45$	$D_{18} = 400$	$D_{26} = 790$
$D_3 = 8$	$D_{11} = 80$	$D_{19} = 450$	$D_{27} = 850$
$D_4 = 12$	$D_{12} = 100$	$D_{20} = 490$	$D_{28} = 950$
$D_5 = 16$	$D_{13} = 150$	$D_{21} = 540$	$D_{29} = 1015$
$D_6 = 20$	$D_{14} = 200$	$D_{22} = 570$	$D_{30} = 1140$
$D_7 = 26$	$D_{15} = 250$	$D_{23} = 620$	$D_{31} = 1370$
$D_8 = 32$	$D_{16} = 300$	$D_{24} = 660$	$D_{32} = 2048$

Desta forma o esquema final de um codificador de sub-bandas utilizando filtros QMF para a divisão e reconstrução do espectro e codificadores BC-PCM para codificar as sub-bandas, é ilustrado na Fig.4.9.

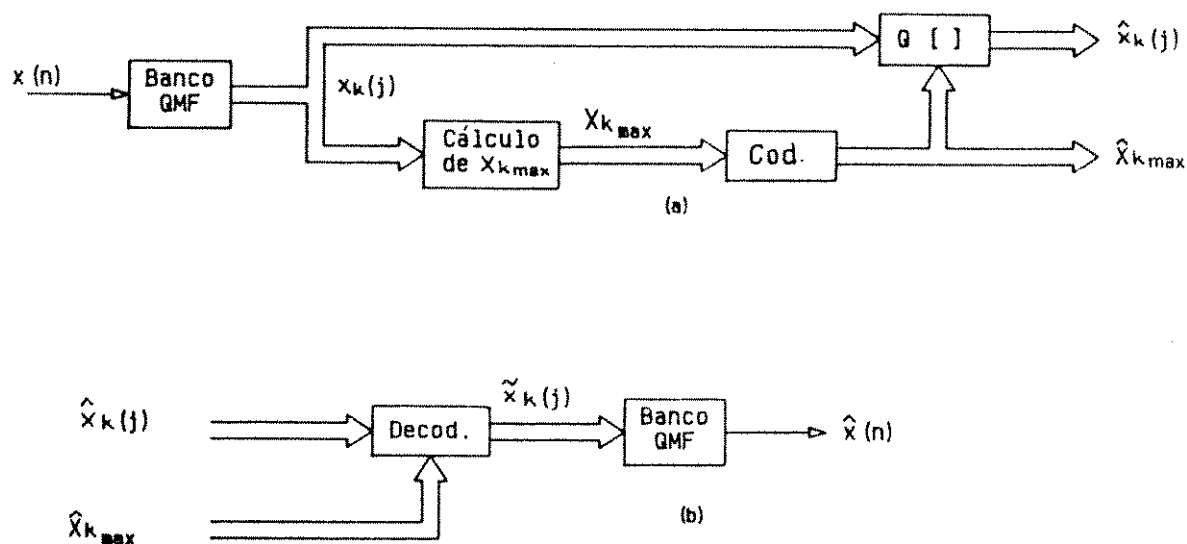


Figura 4.9: Diagrama em blocos de um codificador de sub-bandas que utiliza filtros espelho em quadratura e BC-PCM: (a) Tx; (b) Rx.

Nesta figura, $x(n)$ corresponde ao sinal de entrada limitado em faixa a 3,4 [KHz] e amostrado a 8 [KHz]. O espectro deste sinal é então dividido em N sub-bandas, utilizando-se um banco de filtros espelho em quadratura. A largura de cada sub-banda é dada pelo quociente $F_s/2N$ [Hz]. Assim, para $N=4$ sub-bandas, a largura de cada sub-banda será de 1 [KHz]; para $N=8$ sub-bandas será de 500 [Hz] e para $N=16$ sub-bandas esta largura será de 250 [Hz]. $x_k(j)$ denota a j -ésima amostra da k -ésima sub-banda com $j=1,2,\dots,P$ (extensão do bloco), e $k=1,2,\dots,N$ (número de sub-bandas).

Para cada sub-banda existe um buffer que armazena as P amostras do bloco no qual se baseará o cálculo de $X_{k_{\max}}$. Calculado o $X_{k_{\max}}$ o valor é codificado e transmitido ($\hat{X}_{k_{\max}}$), além de ser utilizado na adaptação da faixa de amplitudes do quantizador. As amostras $x_k(j)$ são então quantizadas ($\hat{x}_k(j)$) e transmitidas junto com $\hat{X}_{k_{\max}}$.

No receptor as amostras são decodificadas, obtendo-se os sinais $\tilde{x}_k(j)$ que, após a filtragem pelo banco QMF, permitem a reconstrução do sinal $\tilde{x}(n)$.

O cálculo da amostra máxima do bloco a quantizar, $X_{k_{\max}}$, deve realizar-se a cada P amostras e em cada sub-banda. Desta forma:

$$X_{k_{\max}} = \text{Max } |x_k(j)| \quad ; \quad j=1,\dots,P \quad ; \quad k=1,\dots,N. \quad (4.6)$$

Este valor, após codificado, controla as características do quantizador pertencente à sub-banda "k", aumentando ou diminuindo o seu passo a cada P amostras, segundo a equação:

$$\Delta k = \frac{\hat{X}_{k_{\max}}}{2^{n_k} - 1} \quad ; \quad k=1,\dots,N. \quad (4.7)$$

Na Eq.(4.7) Δ_k denota o passo de quantização da sub-banda "k" e n_k o número de bits por amostra atribuídos à k-ésima sub-banda. Esta equação só é válida, logicamente, para quantizadores uniformes.

Após adaptar o quantizador, cada amostra é codificada segundo a equação:

$$\hat{x}_k(j) = \text{Trunc} [x_k(j)/\Delta_k] \quad ; \quad j = 1, \dots, P \quad ; \quad k = 1, \dots, N. \quad (4.8)$$

onde Trunc significa "truncamento". No receptor, as amostras são decodificadas segundo a regra seguinte:

$$\tilde{x}_k(j) = \Delta_k [\hat{x}_k(j) + 0.5] \quad ; \quad j = 1, \dots, P \quad ; \quad k = 1, \dots, N. \quad (4.9)$$

Uma alternativa ao uso de $X_{k_{\max}}$ como parâmetro de adaptação do quantizador, é a utilização do desvio padrão σ_x do bloco a ser quantizado (Ref.[9]), isto é:

$$\sigma_k = \left[\frac{1}{P} \sum_{j=1}^P x_k^2(j) \right]^{1/2} \quad ; \quad k = 1, 2, \dots, N. \quad (4.10)$$

onde P é o número de amostras do bloco. σ_k seria então codificado e transmitido e utilizado para adaptar o quantizador no transmissor e o decodificador no receptor. Deve-se ressaltar, no entanto, que o uso de σ_x como parâmetro de adaptação do quantizador implica na existência de "clipping", o que pode ser detectável se ocorre nas baixas frequências. Assim, uma modificação proposta por C.Jaccoud e A.Alcaim (Ref.[10]), sugere utilizar como parâmetro de adaptação do quantizador para as sub-bandas inferiores (até 1,5 [KHz] aproximadamente) $X_{k_{\max}}$, e para as bandas superiores (acima de 1,5 [KHz]), σ_k .

4.4.2 CÁLCULO DA TAXA DE INFORMAÇÃO LATERAL NOS CODIFICADORES BC-PCM

Como já foi explicado, esta taxa é função da duração do bloco de amostras e da quantidade de bits destinados a codificar o parâmetro de adaptação do quantizador.

Para o cálculo da taxa lateral de bits, deve-se considerar que a contribuição da k -ésima sub-banda a esta taxa será R_{Lk} [bits/s], onde:

$$R_{Lk} \text{ [bits/s]} = 1/\tau_k \text{ [blocos/s]} \times n_{Lk} \text{ [bits/bloco]} ; k = 1, \dots, N. \quad (4.11)$$

Nesta equação τ_k representa a duração do bloco de amostras (em segundos) da sub-banda k e, portanto, $1/\tau_k$ é o número de blocos por segundo, pertencentes a essa sub-banda, que devem ser codificados e transmitidos. O parâmetro $1/\tau_k$ também representa o número de vezes que o k -ésimo quantizador é adaptado em 1 segundo. n_{Lk} corresponde ao número de bits com que se codificará e transmitirá o parâmetro de adaptação do quantizador da sub-banda k .

Desta forma a taxa total de informação lateral será:

$$R_L \text{ [bits/s]} = \sum_{k=1}^N R_{Lk} \text{ [bits/s]} = \sum_{k=1}^N 1/\tau_k \times n_{Lk} \quad (4.12)$$

onde N corresponde ao número de sub-bandas em que tem-se dividido o espectro (nesta equação só intervêm as sub-bandas transmitidas). O valor R_L deve ser substituído na Eq. (2.3) do Cap. 2, para se obter a taxa total de bits do codificador de sub-bandas no caso de se utilizar BC-PCM.

Se a duração τ_k do bloco de amostras é igual para todas as sub-bandas e se escolhe $\tau_k = 16$ [ms], então tem-se 62,5 [blocos/s] por sub-banda. Nesta situação, a tabela seguinte fornece a taxa lateral por sub-banda para 3, 4 e 5 [bits/bloco], destinados à codificação e transmissão do parâmetro de adaptação do quantizador.

Bits por bloco: n_{Lk}	Taxa Lateral por S-B : R_{Lk}
3 [bits/bloco]	187,5 [bits/s]
4 [bits/bloco]	250,0 [bits/s]
5 [bits/bloco]	312,5 [bits/s]

Tabela 4.2: Taxa lateral por sub-banda, para codificação do PA do quantizador com 3, 4 e 5 [bits/bloco].

Por outra parte, a Tabela 4.3 fornece o valor da taxa de informação lateral R_L para o caso de se dividir o espectro em 4, 8 e 16 sub-bandas. Nessa tabela supos-se que todas as sub-bandas são codificadas e transmitidas e que em todas elas se utiliza igual número de bits/bloco para se codificar o parâmetro de adaptação do quantizador.

Taxa Lateral Total R_L [bits/s]			
	N= 4 [SB]	N= 8 [SB]	N= 16 [SB]
3 [bits/bloco]	750	1500	3000
4 [bits/bloco]	1000	2000	4000
5 [bits/bloco]	1250	2500	5000

Tabela 4.3: Taxa de informação lateral na codificação do parâmetro de adaptação do quantizador.

Da Tabela 4.3 podem-se tirar algumas conclusões interessantes:

- 1) Para codificadores com taxa final de 16 [Kbits/s] ou menor, a divisão em 16 sub-bandas provoca valores demasiado elevados de taxa lateral, mesmo utilizando-se adaptação grosseira de 3 [bits/bloco]. Portanto, para essas taxas e com codificadores BC-PCM, a divisão em 16 sub-bandas não é recomendável.
- 2) Ao se duplicar o número de sub-bandas em que se divide o espectro, se duplica a taxa de informação lateral.
- 3) Para $N = 4$ sub-bandas, passar de 3 a 4 [bits/bloco] ou de 4 a 5 [bits/bloco], não provoca um aumento significativo no valor absoluto da taxa de informação lateral. Portanto se deveria utilizar neste caso 5 [bits/bloco], para codificar a informação lateral.

A Tabela 4.3 fornece os resultados máximos (de pior caso), da taxa lateral. Normalmente algumas bandas não são transmitidas e, portanto, o valor da taxa lateral pode diminuir. Além disso, se nas bandas altas do codificador são destinados menos bits para a codificação do parâmetro de adaptação do quantizador, a taxa lateral sofre uma diminuição adicional. O mesmo ocorre se algumas das bandas são codificadas com um outro método de codificação que não precise transmitir qualquer informação lateral. Nesses casos pode-se negociar a quantidade de bits destinados à transmissão do parâmetro de adaptação e os destinados à transmissão das amostras, de modo a se otimizar o desempenho do codificador.

4.4.3 CODIFICADOR APCM-AQB

Esta classe de codificadores já foi apresentada nas Seções 4.2.4 e 4.2.4.2. A estratégia de adaptação do passo de quantização utilizada nas simulações, foi baseada no "Método de Adaptação com Memória de Uma Palavra", proposto por Jayant, Flanagan e Cummiskey (Ref.[11]). Na figura seguinte tem-se o esquema de como trabalha este método de adaptação.

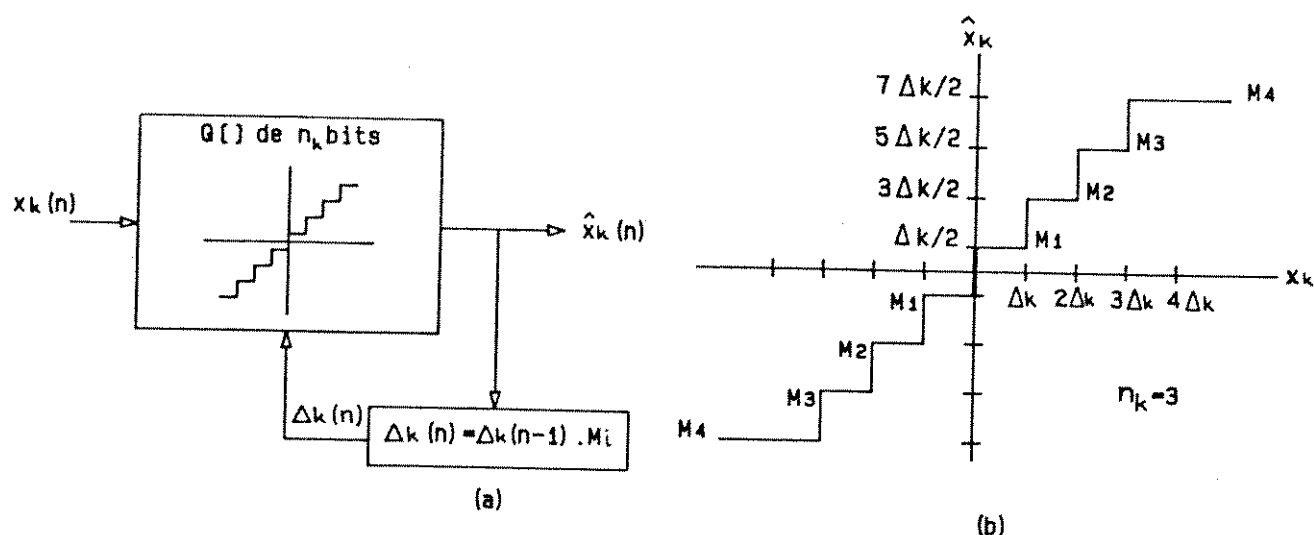


Figura 4.10: Quantizador adaptativo AQB: (a) Algoritmo utilizado para adaptação do passo de quantização; (b) Características do quantizador utilizado.

O sinal da k -ésima sub-banda, $x_k(n)$, é quantizado com um quantizador de 2^{n_k} níveis, obtendo-se o sinal $\hat{x}_k(n) = Q[x_k(n)]$. O quantizador adapta seu passo de amostra para amostra, segundo a relação:

$$\Delta_k(n) = \Delta_k(n-1) \cdot M(|\hat{x}_k(n-1)|) \quad (4.13)$$

$$\text{com} \quad \Delta k_{\min} \leq \Delta k \leq \Delta k_{\max} \quad (4.14)$$

Nestas equações $\Delta k(n)$ corresponde ao passo de quantização para a amostra n , $\Delta k(n-1)$ é o passo de quantização utilizado na amostra anterior e M é um fator cujo valor depende da magnitude da última amostra quantizada. $\Delta k(n-1)$ contém informação condensada da história recente do quantizador, ao passo que $M(|\hat{x}_k(n-1)|)$, que só depende da palavra digital $\hat{x}_k(n-1)$, permite adaptar o quantizador ao nível estimado da próxima amostra a quantizar; daí o nome de "Método de Adaptação com Memória de Uma Palavra".

O fator M pode tomar $2^{n_k-1} = B$ valores como indica a Fig.4.10(b). Se na quantização da amostra $n-1$ é utilizado o menor nível de saída do quantizador, então um valor $M_1 < 1$ é utilizado no cálculo do próximo passo de quantização, de modo a reduzir o mesmo. Se o maior nível de saída é utilizado na quantização da amostra $n-1$, então um valor $M_B > 1$ é usado no cálculo do próximo passo de quantização, de forma de aumentar o tamanho do mesmo. Desta forma o quantizador adapta continuamente seu passo, tentando seguir as variações da variância de curto prazo do sinal de entrada. Por razões práticas, o passo de quantização é condicionado a variar entre um valor mínimo e máximo ($\Delta k_{\min}$ e $\Delta k_{\max}$ respectivamente). Um projeto correto dos multiplicadores M_i , implica em:

$$1 > M_1 \leq M_2 \leq \dots \leq M_B > 1 \quad (4.15)$$

Jayant calculou os valores do multiplicador M_i para o caso de se codificar sinais de voz em banda integral, com codificadores APCM e DPCM (Ref.[12]). Para a codificação de sinais de voz pertencentes a sub-bandas, os valores por ele achados não representam necessariamente os valores ótimos. Por essa razão, Crochiere determinou experimentalmente um novo conjunto de multiplicadores adequados à codificação em

sub-bandas. A Tabela 4.4 mostra estes resultados para codificadores uniformes APCM-AQB de 4, 3, 2, $1\frac{1}{2}$ e $1\frac{1}{3}$ [bits/am], junto com os valores típicos de SNR_a obtidos na codificação de sinais de sub-bandas (Ref.[13]). Pequenas alterações nos valores destes multiplicadores, não afetam significativamente o desempenho do codificador.

Multiplicadores M_i para codificadores APCM-AQB					
n_k [bits] :	4	3	2	$1\frac{1}{2}$	$1\frac{1}{3}$
M_1	0,90	0,85	0,85	0,92	0,92
M_2	0,90	1,00	1,90	1,40	1,40
M_3	0,90	1,00			
M_4	0,90	1,50			
M_5	1,20				
M_6	1,60				
M_7	2,00				
M_8	2,40				
SNR_a [dB] :	18,0	11,5	7,0	4,0	3,3

Tabela 4.4: Multiplicadores M_i para codificadores APCM-AQB uniformes.

Voltando ao algoritmo de codificação proposto, deve-se notar que ele é aplicável só a quantizadores uniformes com 4 ou mais níveis (2 [bits/am]). Caso contrário mais de uma palavra de memória seria necessária para uma adaptação AQB. Porém, uma estratégia proposta por Goodman (Ref.[14]), permite a codificação de sinais a uma taxa média de $1+1/r$ [bits/am], onde r é um inteiro. Neste caso o "sinal" de $x_k(n)$ é codificado a cada amostra, ao passo que a "magnitude" de $x_k(n)$ é codificada uma vez a cada r amostras. O algoritmo

de adaptação permanece o mesmo, empregando o mesmo multiplicador M_k e o mesmo nível de saída $\hat{x}_k(n)$ durante r amostras. Assim, por exemplo, para $r=2$, nas amostras ímpares se transmite o bit de sinal e o de magnitude, enquanto que para as amostras pares só se transmite o bit de sinal. No decodificador, por outra parte, se assume que a magnitude das amostras pares (não transmitida), é igual à magnitude da amostra imediatamente anterior. Isto permite a utilização de poucos bits na codificação das bandas mais altas, onde uma perda de precisão não é tão percebida pelo ouvido humano. As últimas colunas da Tabela 4.4 fornecem o valor dos multiplicadores M_k para a codificação com $11/2$ e $11/3$ [bits/am].

A faixa dinâmica do quantizador da k -ésima sub-banda, é dada pelos valores de $\Delta k_{\min}$ e $\Delta k_{\max}$ (Eq.(4.14)). Este valor, em dB, é expresso por:

$$\text{Faixa em dB} = 20 \log_{10} \frac{\Delta k_{\max}}{\Delta k_{\min}} \quad ; \quad k = 1, \dots, N. \quad (4.16)$$

O valor de $\Delta k_{\min}$ controla o nível de ruído de canal vazio (ruído granular com entrada nula ou quase nula). O valor de $\Delta k_{\max}$, controla o nível de distorção de "clipping".

Nas simulações realizadas neste trabalho, foram testadas relações $\Delta k_{\max} / \Delta k_{\min}$ com valores de 100 e 128 (aproximadamente 40 [dB] de faixa dinâmica). Como não se percebeu nenhuma alteração significativa de desempenho ao se variar a relação de 100 a 128, finalmente foi adotado o valor de 100. Quanto ao $\Delta k_{\min}$ utilizado, seu valor foi definido, em princípio, por meio da seguinte equação:

$$\Delta k_{\min} = 5 - n_k \quad , \quad k = 1, 2, \dots, N. \quad (4.17)$$

e sobre este valor se fizeram alterações até obter o melhor

desempenho do codificador (n_k é o número de bits destinados à quantização das amostras da k -ésima sub-banda e em caso desse número ser $11/2$ ou $11/3$ foi adotado $n_k = 2$).

Para finalizar esta seção é preciso se estabelecer as condições iniciais de trabalho do algoritmo utilizado. Ao começar a quantização de um sinal, deve-se escolher o valor do passo de quantização inicial para cada sub-banda, isto é, $\Delta k(0)$. Este valor poderia ser qualquer um dentro do intervalo $[\Delta k_{\min}, \Delta k_{\max}]$. Porém, quando se espera que a parte inicial do sinal a ser quantizado seja de baixa amplitude, como nos períodos de silêncio que precedem o começo da fala, é apropriado se escolher $\Delta k(0) = \Delta k_{\min}$, $\forall k$. Por esta razão, nas simulações feitas neste trabalho, foi escolhido $\Delta k(0)$ igual ao $\Delta k_{\min}$ da banda correspondente.

4.4.4 ALGORITMO ROBUSTO DE GOODMAN E WILKINSON

Uma desvantagem do algoritmo de Jayant, Flanagan e Cummiskey, é que ao ocorrer um erro no cálculo de $\Delta k(n)$, o sistema permanece em erro até $\Delta k(n)$ atingir o valor $\Delta k_{\min}$ ou $\Delta k_{\max}$. Se este erro ocorre na adaptação do decodificador do receptor, devido por exemplo a erros de canal, então, ainda que as palavras código sejam transmitidas sem erros, elas serão decodificadas erroneamente.

Uma modificação deste algoritmo, proposta por Goodman e Wilkinson, permite que o cálculo de $\Delta k(n)$ seja menos sensível a erros passados (Ref.[14]). A forma do algoritmo é a seguinte:

$$\Delta k(n) = [\Delta k(n-1)]^{\beta} \cdot M(|\hat{x}_k(n-1)|) \quad (4.18)$$

com
$$\Delta k_{\min} \leq \Delta k \leq \Delta k_{\max} \quad (4.19)$$

Comparando este algoritmo com o anterior, Eq.(4.13), se percebe que a única diferença é a presença do parâmetro β . Este parâmetro é levemente menor que 1 e faz com que os efeitos dos erros passados sejam dissipados após algumas amostras. Os valores típicos de β são: 63/64, 31/32 e 15/16. No limite, quando $\beta = 1$, o algoritmo se reduz ao de Jayant, Flanagan e Cumiskey. Na escolha de β deve-se levar em conta que quanto menor for seu valor, mais rapidamente o sistema dissipa os efeitos de erros passados, mas por outro lado cai o desempenho do sistema na ausência de erros.

Os multiplicadores M_i devem ser corrigidos de modo a se levar em conta a presença de β no cálculo do passo de quantização. Crochiere determinou experimentalmente um fator de correção G em função de β e dos passos $\Delta k(n)$ esperados, de forma a otimizar os valores M_i (Ref.[15]). Este fator de correção é dado na Fig.4.11.

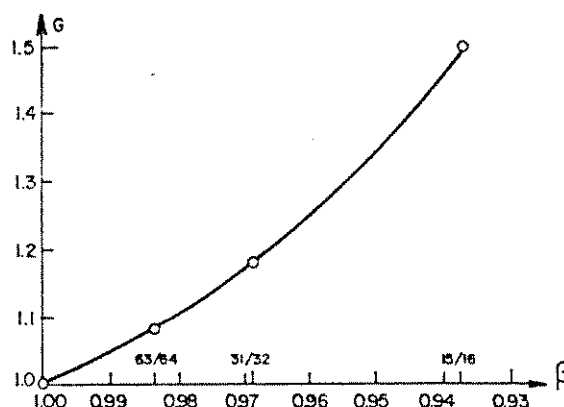


Figura 4.11: Fator de correção G dos multiplicadores M_i , em função do β escolhido.

Desta forma:

$$M_i = G \cdot \hat{M}_i, \quad i = 1, 2, \dots, B. \quad (4.20)$$

onde $\hat{M}_i$ corresponde aos valores dos multiplicadores quando $\beta = 1$, M_i aos valores otimizados para $\beta < 1$ e G ao fator de correção.

4.5 ATRIBUIÇÃO DE BITS POR BANDA E UTILIZAÇÃO DE BC-PCM OU APCM-AQB

Na distribuição dos bits para codificar as diferentes sub-bandas, e na escolha do codificador específico para codificar cada uma delas (BC-PCM ou APCM-AQB), devem ser consideradas as seguintes questões:

- A maior parte da energia dos sons vocálicos, que são os mais importantes para a compreensão da fala, se concentra na parte baixa do espectro (até os 1200 [Hz] aproximadamente). Acima dessa frequência começa a aparecer a energia correspondente aos sons surdos e fricativos, que possuem um peso menor na inteligibilidade da voz. Consequentemente a parte alta do espectro (acima dos 1200 [Hz]), pode ser codificada mais grosseiramente que a região das frequências baixas. Isto permite, também, preservar melhor a informação de pitch e de formantes que está presente na região baixa do espectro.
- Nas sub-bandas inferiores há uma maior concentração de energia que nas sub-bandas superiores (Fig.4.6). Portanto, para uma mesma quantidade de bits em todas as bandas, os

erros de quantização das sub-bandas inferiores possuirão maior energia que os das sub-bandas superiores. Levando-se em conta o formato da curva de densidade espectral de potência (Fig. 4.6), é possível minimizar o ruído de quantização total tirando-se bits das bandas altas e transferindo-os às bandas baixas.

- O ouvido humano é mais sensível a distorções nas bandas baixas que nas bandas altas.

Por estes motivos é necessário se atribuir uma maior quantidade de bits às sub-bandas inferiores que às superiores, durante a alocação dos bits.

Quanto à quantidade específica de bits por sub-banda, esta será função da taxa final de bits que se pretende atingir e do tipo de codificador utilizado (BC-PCM ou APCM-AQB).

A atribuição de bits a cada sub-banda pode ser feita em forma fixa ou dinâmica (Cap. 2, Seção 2.2.2). Neste trabalho foi utilizada atribuição fixa de bits, para todos os codificadores simulados. A utilização de atribuição dinâmica de bits pode melhorar o desempenho do codificador de sub-bandas, caso a transmissão de informação se realize através de canais não ruidosos. Na presença de taxas de erro elevadas, entretanto, ocorre uma degradação de desempenho. Isto porque na atribuição dinâmica de bits, é necessário transmitir-se ao receptor informação adicional sobre o número de bits que está sendo utilizada nos codificadores do transmissor. Se esta informação é afetada pelo ruído do canal e chega errada ao receptor, os decodificadores possuirão uma referência errada e, portanto, a decodificação da informação de chegada será errada. Logicamente, é possível a proteção desta informação mediante a utilização de códigos imunes a erros, mas isto acarreta um aumento na taxa total de bits, o que a baixas taxas pode ser inaceitável. Adicionalmente a

atribuição dinâmica de bits torna o codificador bem mais complexo.

Com respeito aos codificadores utilizados nas diferentes sub-bandas, as simulações realizadas permitiram determinar a conveniência de se utilizar codificadores BC-PCM nas bandas inferiores, com $X_{k_{max}}$ como parâmetro de adaptação do quantizador. Com isto se impede a ocorrência de "clipping" e "slope overload" nessas bandas, preservando melhor a informação contida nelas. Para as bandas superiores, entretanto, foi testada a utilização de ambas classes de codificadores (BC-PCM e APCM-AQB).

No Cap. 5 serão apresentados os resultados obtidos com as diferentes combinações de codificadores utilizados.

REFERÊNCIAS

- [1] J. Max, "Quantizing for Minimum Distortion", IRE Trans. Inform. Theory, Vol IT 6, March 1960, pp 7-12.
- [2] M.D. Paez and T.H. Glisson, "Minimum Mean-Squared-Error Quantization in Speech PCM and DPCM Systems", IEEE Trans. Comm., Vol COM-20, Apr 1972, pp 225-230.
- [3] P. Noll, "Adaptive Quantizing in Speech Coding Systems", Proc. 1974 Zürich Seminar on Digital Communications, Zürich, March 1974, pp B3(1)-B3(6).
- [4] N.S. Jayant and P. Noll, "Digital Coding of Waveforms - Principles and Applications to Speech and Video", Prentice Hall, Inc., Englewood Cliffs, N.J., 1984.

- [5] P.Noll, "Non-adaptive and Adaptive DPCM of Speech Signals", Polytech. Tijdschr. Ed. Elektrotech/Elektron (The Netherlands), Nº 19, 1972.
- [6] L.L.Beranek, "The Design of Communication Systems", Proc. IRE 35, September 1949, pp 880-890.
- [7] H.K.Dunn and S.D.White, "Statistical Measurements on Conversational Speech", J.Acoust.Soc.Amer.,11, 1940, pp 278-288.
- [8] D.Esteban and C.Galand, "Application of Quadrature Mirror Filters to Split Band Voice Coding Schemes", Proc. ICASSP, May 1977, pp 191-195.
- [9] T.Ramstad, "Sub-band Coder With a Simple Adaptive Bit Allocation Algorithm, A Possible Candidate for Digital Mobile Telephony?", Conf. Rec. ICASSP, France 1982, pp 203-207.
- [10] C.F.Jaccoud and A.Alcain, "Improvement of the Coding Structure in Sub-band Encoding of Speech Signals", Conf.Rec, Int l Conf. on Digital Signal Processing, Florence, Italy, 1987, pp 315-318.
- [11] P.Cummiskey, N.S.Jayant and J.L.Flanagan, "Adaptive Quantization in Differential PCM Coding of Speech", Bell System Tech. J., Vol 52, Nº 7, Sept. 1973,, pp 1105-1118.
- [12] N.S.Jayant, "Digital Coding of Speech Waveform: PCM, DPCM and DM Quantizers", Proc. IEEE, 62, May 1974, pp 611-632.

- [13] R.E.Crochiere, "On the Design of Sub-band Coders for Low-Bit-Rate Speech Communication", Bell System Tech. J., Vol 56, Nº 5, May-June 1977, pp 747-770.
- [14] D.J.Goodman and R.M.Wilkinson, "A Robust Adaptive Quantizer", IEEE Trans.Comm., Nov 1975, pp 1362-1365.
- [15] R.E.Crochiere, "An Analysis of 16 Kb/s Sub-band Coder Performance: Dynamic Range, Tandem Connections and Channel Errors", Bell System Tech. J., Vol 57, Nº 8, Oct 1978, pp 2927-2952.

CAPÍTULO 5

TESTES E DESEMPENHO

CONTEÚDO

5.1 - Introdução	103
5.2 - Descrição dos objetivos	104
5.2.1 - Taxa final do codificador	105
5.2.2 - Número de sub-bandas em que se divide o espectro	105
5.2.3 - Sub-bandas codificadas e transmitidas	105
5.2.4 - Número de bits atribuído a cada sub-banda e método de codificação empregado	106
5.2.5 - Número de "taps" dos filtros QMF	106
5.2.6 - Taxas de erros de canal	107
5.3 - Descrição do "Software" utilizado	108
5.4 - Descrição das simulações	110
5.4.1 - Medição da relação sinal/ruído convencional e segmental	1123
5.5 - Sistemas simulados e resultados obtidos	113
5.5.1 - À taxa de 8 [Kbits/s]	114
5.5.2 - À taxa de 16 [Kbits/s]	119
5.5.3 - À taxa de 24 [Kbits/s]	123
5.6 - Comparação com outros codecs	130
REFERÊNCIAS	133

5.1 INTRODUÇÃO

Neste capítulo são descritas as simulações realizadas ao longo deste trabalho, assim como os principais resultados obtidos.

O capítulo começa com a descrição dos diferentes parâmetros que foram variados durante os testes e que

intervém no desempenho dos codificadores de sub-bandas. Em seguida é descrito o "software" utilizado para realizar as simulações e cujo projeto foi um dos objetivos do presente trabalho. É descrito também o ambiente de trabalho, o mecanismo das simulações e os critérios de avaliação utilizados.

Foram feitas simulações de cerca de 190 sistemas de codificação em sub-bandas, operando às taxas de 8, 16 e 24 [Kbits/s]. São apresentadas as especificações e resultados de desempenho dos melhores codificadores obtidos, operando em cada uma das taxas acima.

Finalmente é feita uma breve comparação de desempenho entre os codificadores de sub-bandas e alguns outros sistemas de codificação. Neste ponto é apresentando o codificador de sub-bandas cujo desempenho se iguala, subjetivamente, ao do PCM-Lei A, de 64 [Kbits/s], e ao do ADPCM¹ de 32 [Kbits/s].

52 DESCRIÇÃO DOS OBJETIVOS

O objetivo deste trabalho foi o estudo, projeto e simulação em computador, de um sistema completo de codificação em sub-bandas, e a posterior avaliação de seu desempenho com a variação dos seguintes parâmetros:

- 1) Taxa final do codificador.
- 2) Número de sub-bandas em que se divide o espectro.
- 3) Sub-bandas codificadas e transmitidas.
- 4) Número de bits atribuído a cada sub-banda e método de codificação empregado.
- 5) Número de "taps" dos filtros QMF.
- 6) Taxa de erros do canal.

¹ Codificador padronizado pela CCITT na recomendação G.721.

A seguir será feita uma descrição sucinta de cada um destes items.

5.2.1 TAXA FINAL DO CODIFICADOR

Foram obtidos os melhores sistemas de codificação em sub-bandas, trabalhando à taxa de 8, 16 e 24 [Kbits/s]. Também foi obtida a taxa do codificador de sub-bandas cuja qualidade e inteligibilidade é comparavel à do PCM de 64 [Kbits/s] e à do ADPCM de 32 [Kbits/s]. Na Seção 5.3 são dados os critérios de avaliação utilizados na comparação dos diferentes codecs (codificadores-decodificadores), simulados.

5.2.2 NÚMERO DE SUB-BANDAS EM QUE SE DIVIDE O ESPECTRO

Quanto ao número de sub-bandas em que se divide o espectro, foram testadas divisões em 4, 8 e 16 sub-bandas. Em todos os casos a divisão e reconstrução do espectro foi realizada com bancos de filtros QMF. A divisão em 2 sub-bandas não foi utilizada, devido ao pouco ganho que introduz nesta técnica de codificação.

5.2.3 SUB-BANDAS CODIFICADAS E TRANSMITIDAS

O fato de se dividir o espectro em N sub-bandas, não implica que todas elas devam ser necessariamente codificadas e transmitidas. Em função das baixas taxas finais dos codecs simulados e da pouca contribuição à qualidade final do sistema, as componentes de frequência acima de 3000 [Hz] não foram transmitidas.

5.2.4 NÚMERO DE BITS ATRIBUÍDO A CADA SUB-BANDA E MÉTODO DE CODIFICAÇÃO EMPREGADO

Em cada codificador simulado, a atribuição de bits por banda foi fixa. A distribuição de bits entre as sub-bandas, foi baseada nos critérios mencionados na Seção 4.5 do Cap.4. Quanto ao método de codificação empregado nelas, os testes determinaram a conveniência de se codificar as bandas baixas com BC-PCM, com $X_{k_{max}}$ como parâmetro de adaptação do quantizador, e as bandas altas com BC-PCM ou APCM-AQB. Devido ao uso de $X_{k_{max}}$ como parâmetro de adaptação do quantizador nos codificadores BC-PCM, os quantizadores utilizados foram uniformes não ótimos. Para se utilizar quantizadores ótimos, é preciso usar σ_k como parâmetro de adaptação do quantizador. Nos codificadores APCM-AQB também foram utilizados quantizadores uniformes não ótimos.

5.2.5 NÚMERO DE "TAPS" DOS FILTROS QMF

O número de "taps" utilizado nos filtros do banco QMF reflete-se na qualidade final do codificador, já que desse número depende, de forma geral, o "ripple" na banda passante, a atenuação na banda de rejeição e a largura da banda de transição dos filtros. Assim, ao aumentar o número de "taps" dos filtros a qualidade do codec melhora, mas também provoca um atraso maior no sinal processado, além de incrementar a complexidade computacional (tempo de processamento) e o "hardware" do sistema. Nas simulações realizadas neste trabalho, foram testados filtros com 8, 12, 16, 24, 32, 48, e 64 "taps" em cada estágio da estrutura QMF. O resultado das simulações permitiu concluir que o ganho obtido com o aumento do número de "taps" não justifica, em geral, o uso de filtros com mais de 32 ou, no máximo, 48 "taps". A utilização de filtros com, por exemplo, 64 "taps", produz uma melhora

inferior a 1 [dB] no valor da SNR_{seg}^1 , com a desvantagem do maior atraso e complexidade do sistema.

Considerando-se que o número de "taps" pode diminuir pela metade de estágio em estágio (Cap.3, Seção 3.3.3), após testes se obtiveram as seguintes combinações úteis:

N= 4 32 e 16 "taps" ou 48 e 24 "taps" (1º e 2º estágio respectivamente).

N= 8 32, 16 e 8 "taps" ou 48, 24 e 12 "taps" (1º, 2º e 3º estágio respectivamente).

N= 16 32, 16, 8 e 8 "taps" ou 48, 24, 12 e 8 "taps" (1º, 2º, 3º e 4º estágio respectivamente).

Não é conveniente se utilizarem filtros com menos de 16 "taps" no segundo estágio da estrutura filtrante, já que são esses filtros os que determinam a rejeição das componentes de frequência acima dos 3000 [Hz] (componentes que não serão transmitidas). Os coeficientes dos filtros utilizados, são dados no Apêndice C.

5.2.6 TAXAS DE ERROS DE CANAL

Os codificadores simulados foram testados supondo canal ideal e canal ruidoso. Para o segundo caso utilizaram-se taxas de erros de canal de 10^{-4} , 10^{-3} e 10^{-2} , e utilizou-se o algoritmo robusto de Goodman e Wilkinson na adaptação do passo de quantização dos codificadores APCM-AQB. Para fixar o valor de β empregado no algoritmo, foi testado o uso de 31/32 e 63/64. O valor finalmente adotado foi 63/64, devido à pouca degradação que este valor provoca no caso de transmissão em canal ideal (degradação menor a 0,4 [dB]).

¹ SNR_{seg} : relação sinal/ruído segmental. Sua definição é dada na Seção 5.4.1.

Para os codificadores BC-PCM, onde é necessário se transmitir periodicamente ao receptor o parâmetro de adaptação do quantizador, foi suposto que esta informação esteve adequadamente protegida, de forma de não ser afetada pelos erros do canal. Porém, não foi feita nenhuma suposição sobre a taxa extra de bits necessária para protegê-la. Quanto ao código utilizado na transmissão das amostras, foi empregado apenas o binário em complemento de dois.

5.3 DESCRIÇÃO DO "SOFTWARE" UTILIZADO

Para realizar as simulações em computador dos diferentes sistemas de codificação em sub-bandas, foi necessário projetar e desenvolver um "software" específico, que permitisse a variação dos diferentes parâmetros que intervêm no desempenho destes codificadores. Assim, o primeiro passo foi o desenvolvimento deste "software" de trabalho, o que foi realizado num ambiente Turbo-Pascal (V 5.0).

O programa desenvolvido, SBC2 (Sub-band Coding 2), na sua versão final, tem as seguintes características:

1. Possibilidade de dividir o espectro do sinal de entrada em 2, 4, 8 ou 16 sub-bandas.
2. Para cada um dos estágios da estrutura divisora do espectro (banco QMF), tem-se a possibilidade de se escolherem filtros de 8, 12, 16, 24, 32, 48 ou 64 "taps". Adicionalmente, e para cada número de "taps", é possível se escolher diferentes classes de filtros, caracterizados pela largura da banda de transição dos mesmos.
3. Uma vez feita a divisão do espectro em N sub-bandas, existe a possibilidade de se reconstruir o sinal de entrada sem realizar a codificação das amostras, ou a alternativa de se codificar as mesmas antes da re-

construção do sinal original. Em ambos os casos, só as bandas escolhidas pelo usuário serão utilizadas na reconstrução do sinal de entrada. As restantes serão descartadas.

4. Na área de codificação, tem-se a possibilidade de se escolher o número de bits destinados a codificar as amostras pertencentes a cada sub-banda. Neste caso as opções são: 1, 3/2, 2, 3, 4 ou 5 [bits/am].
5. Alocado este número de bits/amostra, o usuário deve definir o modo de codificação escolhido para cada sub-banda. Neste caso as possibilidades são duas: BC-PCM ou APCM-AQB (esta última opção não é válida em caso de se utilizar 1 [bit/am]).
6. No caso de se optar por BC-PCM, tem-se a possibilidade de se definir o número de bits que será destinado à codificação do parâmetro de adaptação do quantizador. As opções são: 3, 4 ou 5 [bits/bloco]. No caso de se escolher APCM-AQB, o usuário pode definir o valor do passo de quantização mínimo desejado para estes codificadores. Aqui as opções são: $\Delta k_{\min} = 1, 2$ ou 3 . Em ambos os casos, a escolha é feita para cada uma das bandas a transmitir.
7. Posteriormente o usuário escolhe se o processamento deve-se realizar em canal ideal ou em canal ruidoso. Para o caso de se escolher canal ruidoso, as taxas de erros disponíveis são: 10^{-4} , 10^{-3} e 10^{-2} .
8. Após a entrada destes dados, o programa fornece a taxa final do codificador de sub-bandas, com a possibilidade de se reprogramar os dados de entrada.
9. Finalmente o usuário define qual é o arquivo de voz a processar e qual o nome do arquivo que será gerado. Desta forma o programa processa o arquivo de voz desejado, obtendo-se um novo arquivo para posterior avaliação de desempenho do codificador de sub-bandas simulado.

A grande variedade de opções oferecidas pelo programa, permite uma grande flexibilidade de trabalho, sendo possível simular codificadores de sub-bandas com taxas finais variando de praticamente 1,5 até 45 [Kbits/s].

Para complementar a informação sobre este "software", deve-se mencionar que o programa fonte possui 3344 linhas, que compiladas geram um programa executável de 114 [Kbytes].

5.4 DESCRIÇÃO DAS SIMULAÇÕES

As simulações foram realizadas num microcomputador PC-AT com processador 80286, co-processador 80287 e relógio de 8 [MHz], ligado ao Sistema de Análise e Processamento Digital de Voz SAPDV-A (Ref.[1]), pertencente ao Laboratório de Comunicações Digitais do DECOM/FEE/UNICAMP¹.

O SAPDV-A é um sistema que permite a geração e gravação direta, em disco Winchester, de arquivos de voz de longa duração, e sua posterior reprodução através de alto-falante ou telefone. A duração dos arquivos de voz é função apenas da capacidade de armazenamento do Winchester; assim para um disco de 20 [Mbytes], o arquivo gerado pode ter uma duração máxima de 28 minutos. Os arquivos gerados podem, posteriormente, ser processados no microcomputador em tempo não real e em linguagens de alto nível, e em seguida ser reproduzidos em alto-falante ou telefone. Isto permite uma avaliação subjetiva de desempenho do sinal processado. Assim, o SAPDV-A facilita bastante o estudo e a avaliação de algoritmos de processamento de voz.

Neste trabalho, os sinais de entrada aos codificadores de sub-bandas simulados consistiram de 3 sentenças, 2 com voz

¹ Departamento de Comunicações da Faculdade de Engenharia Elétrica da UNICAMP.

masculina e 1 com voz feminina, totalizando 26 segundos de voz.

Procurou-se utilizar sentenças foneticamente equilibradas, isto é, contendo sons fricativos sonoros (v,z,j,...), fricativos surdos (f,s,sh...), oclusivos (b,p,d,...) e vogais (a,e,...), de modo a se poder realizar uma avaliação correta dos codificadores simulados. As sentenças utilizadas foram:

- 1) *"Festas surpresa são difíceis de realizar. Muitas coisas podem sair erradas. Algumas vezes os erros são realmente engraçados."* (Voz masculina).
- 2) *"A profissão de aeromoça parece ser excitante mas é também muito cansativa. Lidar com pessoas faz parte do trabalho."* (Voz feminina).
- 3) *"Bom dia garçonetes de Piracicaba. Vocês já prepararam as torradas, meu café e nosso chá?"* (Voz masculina).

Os sinais analógicos pertencentes a estas sentenças foram inicialmente limitados em faixa a 3400 [Hz], amostrados a 8 [KHz], quantizados com 12 [bits/am], e armazenados em disco, obtendo-se assim os sinais de entrada aos codecs de sub-bandas simulados. As amostras do sinal de voz excursionam, pois, no intervalo de -2048 a 2047.

Estes sinais, uma vez processados pelos algoritmos de sub-bandas, foram comparados com os sinais originais de entrada para se avaliar o desempenho dos sistemas simulados.

Como medida de desempenho fizeram-se testes objetivos e subjetivos. A ênfase foi dada aos testes subjetivos, já que em baixas taxas os testes objetivos não refletem de maneira confiável a qualidade e inteligibilidade dos sistemas de codificação.

Os testes objetivos consistiram na medição da relação sinal/ruído convencional e segmental. Porém, só esta última foi utilizada como critério de comparação, por aproximar-se

mais das medidas subjetivas de desempenho.

Os testes subjetivos consistiram na audição, por parte de 4 pessoas, dos sinais de voz processados pelos diferentes codificadores simulados, e a posterior seleção daqueles considerados de melhor qualidade. As sentenças foram reproduzidas várias vezes e em ordem aleatória, através de alto-falante e de telefone.

5.4.1 MEDIÇÃO DA RELAÇÃO SINAL/RUÍDO CONVENCIONAL E SEGMENTAL

Relação sinal/ruído convencional

A relação sinal/ruído convencional é definida como:

$$SNR [dB] = 10 \log_{10} \frac{P_S}{P_N} \quad (5.1)$$

onde P_S é a potência do sinal e P_N é a potência de ruído, ambas medidas sobre a totalidade da sentença. Assim:

$$P_S = \frac{1}{U} \sum_n x^2(n) \quad (5.2)$$

$$P_N = \frac{1}{U} \sum_n [x(n) - y(n)]^2 \quad (5.3)$$

onde U é o número total de amostras da sentença, $x(n)$ é o sinal de entrada ao codificador de sub-bandas no transmissor e $y(n)$ o sinal de saída do decodificador de sub-bandas no receptor. Antes da utilização de $y(n)$ na Eq.(5.3), deve-se descontar o atraso e a atenuação que o sistema provoca sobre este sinal (Eqs.3.52 e 3.53). As somatórias incluem todas as amostras da sentença processada.

Relação sinal/ruído segmental

Para esta medida a sentença é dividida em segmentos de curta duração (tipicamente 15 a 30 [ms]), e sobre cada segmento é obtida a relação sinal/ruído convencional em dB. Em seguida se faz uma média de todas as SNR computadas, obtendo-se desta forma a relação sinal/ruído segmental. Assim:

$$\text{SNRseg [dB]} = \frac{1}{L} \sum_{i=1}^L \text{SNR}_i \text{ [dB]} \quad (5.4)$$

onde

$$\text{SNR}_i \text{ [dB]} = 10 \log_{10} \frac{P_s(i)}{P_n(i)} \quad (5.5)$$

Nestas equações L é o número de segmentos que contém a sentença processada e $P_s(i)$ e $P_n(i)$ são as potências de sinal e de ruído, respectivamente, correspondentes ao segmento "i". Neste trabalho foram adotados segmentos de 16 [ms], isto é, segmentos de 128 amostras.

Para evitar que segmentos onde predomina o silêncio afetem indevidamente a medida dada pela Eq.(5.4), os segmentos onde a potência do sinal é inferior a um determinado limiar não são incluídos no cálculo da SNRseg (Ref.[2]). Assim o algoritmo utilizado na medição da SNRseg desconsidera os segmentos onde a potência do sinal é menor que 4,5 [dB] (este valor não é arbitrário e surge de diversas considerações).

5.5 SISTEMAS SIMULADOS E RESULTADOS OBTIDOS

Ao longo deste trabalho foram simulados aproximadamente

190 codecs diferentes, trabalhando às taxas de 8, 16 e 24 [Kbits/s]. Dentre esses sistemas somente serão apresentados aqueles de melhor desempenho.

5.5.1 À TAXA DE 8 [KBITS/S]

(a) N= 4 sub-bandas: B[1] a B[4]

Largura de cada sub-banda: 1000 [Hz].

Bandas transmitidas: B[1] e B[2] (0 a 2000 [Hz]).

Banco QMF utilizado: 1o estágio: 32 "taps" - Código C

2o estágio: 16 "taps" - Código C

CODEC 1

Sub-banda	Bits/am	Codif.	Bits I.L. ¹	$\Delta k_{\min}$ ²
1	2	BC-PCM	5	—
2	1,5	BC-PCM	5	—

Taxa de Informação Lateral = 625 [bits/s]

Taxa Total = 7625 [bits/s]

SNR = 8,84 [dB]

SNRseg = 7,43 [dB]

¹ Bite de informação lateral (número de bite utilizados para transmitir o parâmetro de adaptação do quantizador nos codificadores BC-PCM).

² Valor do passo de quantização mínimo nos codificadores APCM-AQB.

CODEC 2

Sub-banda	Bits/am	Codif.	Bits I.L.	$\Delta k_{\min}$
1	2	APCM-AQB	—	2
2	2	APCM-AQB	—	2

Taxa de Informação Lateral = 0

Taxa Total = 8000 [bits/s]

SNR = 4,72 [dB]

SNRseg = 6,09 [dB]

(b) N= 8 sub-bandas: B[1] a B[8]

Largura de cada sub-banda: 500 [Hz].

Bandas transmitidas: B[1] a B[5] (0 a 2500 [Hz]).

Banco QMF utilizado: 1o estágio: 32 "taps" - Código C

2o estágio: 16 "taps" - Código C

3o estágio: 8 "taps"

CODEC 3

Sub-banda	Bits/am	Codif.	Bits I.L.	$\Delta k_{\min}$
1	2	BC-PCM	5	—
2	1,5	BC-PCM	5	—
3	1	BC-PCM	5	—
4	1	BC-PCM	5	—
5	1	BC-PCM	4	—

Taxa de Informação Lateral = 1500 [bits/s]

Taxa Total = 8000 [bits/s]

SNR = 7,62 [dB]

SNRseg = 5,94 [dB]

CODEC 4

Sub-banda	Bits/am	Codif.	Bits I.L.	Δk_{min}
1	2	BC-PCM	4	—
2	1,5	APCM-AQB	—	2
3	1,5	APCM-AQB	—	2
4	1,5	APCM-AQB	—	1
5	1	BC-PCM	4	—

Taxa de Informação Lateral = 500 [bits/s]

Taxa Total = 8000 [bits/s]

SNR = 5,43 [dB]

SNRseg = 5,06 [dB]

(c) N= 16 Sub-bandas: B[1] a B[16]

A divisão em 16 sub-bandas para taxa final de 8 [Kbits/s], resultou numa SNRseg inferior a 3 [dB]; portanto não será apresentado nenhum dos codecs simulados.

Ainda que nos testes objetivos o Codec 1 tenha obtido o melhor desempenho, os testes subjetivos mostraram que o Codec 3 apresenta uma qualidade superior. Portanto este último foi o escolhido para avaliação de comportamento em canal ruidoso (taxas de erros de 10^{-4} , 10^{-3} e 10^{-2}). Os resultados obtidos foram:

Taxa de erro= 0 : $\begin{cases} \text{SNR} = 7,62 \text{ [dB]} \\ \text{SNRseg} = 5,94 \text{ [dB]} \end{cases}$

Taxa de erro= 10^{-4} : $\begin{cases} \text{SNR} = 7,61 \text{ [dB]} \\ \text{SNRseg} = 5,91 \text{ [dB]} \end{cases}$

$$\text{Taxa de erro} = 10^{-3} : \begin{cases} \text{SNR} = 7,15 \text{ [dB]} \\ \text{SNRseg} = 5,50 \text{ [dB]} \end{cases}$$

$$\text{Taxa de erro} = 10^{-2} : \begin{cases} \text{SNR} = 3,50 \text{ [dB]} \\ \text{SNRseg} = 2,61 \text{ [dB]} \end{cases}$$

A figura seguinte sumariza estes resultados:

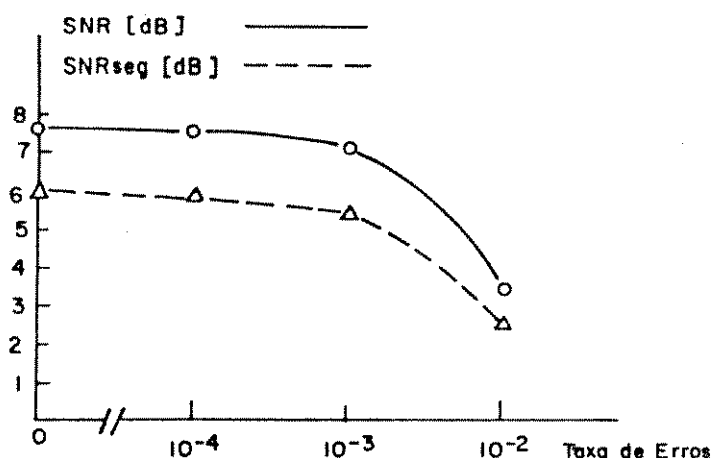


Figura 5.1: Comportamento do Codec 3 com taxa de erros.

Com taxa de erros de 10^{-4} o sistema é totalmente inteligível e de qualidade similar à do codec com taxa de erros nula. Com taxa de erros de 10^{-3} , o sistema é ainda totalmente inteligível e de boa qualidade. Finalmente para taxa de erros de 10^{-2} , tanto a inteligibilidade como a qualidade do codec caem bastante, percebendo-se ruído durante a fala que dificulta a identificação do locutor.

5.52 À TAXA DE 16 [KBITS/s]

(a) N= 4 sub-bandas: B[1] a B[4]

Largura de cada sub-banda: 1000 [Hz].

Bandas transmitidas: B[1] a B[3] (0 a 3000 [Hz]).

Banco QMF utilizado: 1o estágio: 32 "taps" - Código C

2o estágio: 16 "taps" - Código C

CODEC 5

Sub-banda	Bits/am	Codif.	Bits I.L.	Δk_{min}
1	3	BC-PCM	5	—
2	3	BC-PCM	5	—
3	1,5	BC-PCM	5	—

Taxa de Informação Lateral = 937,5 [bits/s]

Taxa Total = 15937,5 [bits/s]

SNR = 14,91 [dB]

SNRseg = 12,95 [dB]

(b) N= 8 sub-bandas: B[1] a B[8]

Largura de cada sub-banda: 500 [Hz].

Bandas transmitidas: B[1] a B[6] (0 a 3000 [Hz]).

Banco QMF utilizado: 1o estágio: 32 "taps" - Código C

2o estágio: 16 "taps" - Código C

3o estágio: 8 "taps"

CODEC 6

Sub-banda	Bits/am	Codif.	Bits I.L.	$\Delta k_{\min}$
1	4	BC-PCM	5	—
2	3	BC-PCM	5	—
3	3	BC-PCM	5	—
4	2	BC-PCM	5	—
5	1	BC-PCM	5	—
6	1	BC-PCM	5	—

Taxa de Informação Lateral = 1875 [bits/s]

Taxa Total = 15875 [bits/s]

SNR = 17,02 [dB]

SNRseg = 14,02 [dB]

CODEC 7

Sub-banda	Bits/am	Codif.	Bits I.L.	$\Delta k_{\min}$
1	4	BC-PCM	5	—
2	3	BC-PCM	5	—
3	3	BC-PCM	5	—
4	2	APCM-AQB	—	2
5	1,5	APCM-AQB	—	1
6	1,5	APCM-AQB	—	1

Taxa de Informação Lateral = 937,5 [bits/s]

Taxa Total = 15937,5 [bits/s]

SNR = 17,16 [dB]

SNRseg = 15,18 [dB]

(c) N= 16 Sub-bandas: B[1] a B[16]

Largura de cada sub-banda: 250 [Hz].

Bandas transmitidas: B[1] a B[12] (0 a 3000 [Hz]).

Banco QMF utilizado: 1º estágio: 32 "taps" - Código C

2º estágio: 16 "taps" - Código C

3º estágio: 8 "taps"

4º estágio: 8 "taps"

CODEC 8

Sub-banda	Bits/am	Codif.	Bits I.L.	$\Delta k_{\min}$
1	3	BC-PCM	4	—
2	4	BC-PCM	4	—
3	4	BC-PCM	4	—
4	3	BC-PCM	4	—
5	3	BC-PCM	4	—
6	2	BC-PCM	4	—
7	1,5	BC-PCM	4	—
8	1,5	BC-PCM	4	—
9	1	BC-PCM	4	—
10	1	BC-PCM	4	—
11	1	BC-PCM	4	—
12	1	BC-PCM	4	—

Taxa de Informação Lateral = 3000 [bits/s]

Taxa Total = 16000 [bits/s]

SNR = 17,39 [dB]

SNRseg = 12,73 [dB]

CODEC 9

Sub-banda	Bits/am	Codif.	Bits I.L.	$\Delta k_{\min}$
1	4	BC-PCM	4	—
2	4	BC-PCM	4	—
3	4	BC-PCM	4	—
4	3	BC-PCM	4	—
5	3	BC-PCM	4	—
6	2	BC-PCM	4	—
7	1,5	APCM-AQB	—	2
8	1,5	APCM-AQB	—	1
9	1,5	APCM-AQB	—	1
10	1,5	APCM-AQB	—	1
11	1,5	APCM-AQB	—	1
12	1,5	APCM-AQB	—	1

Taxa de Informação Lateral = 1500 [bits/s]

Taxa Total = 16000 [bits/s]

SNR = 17,71 [dB]

SNRseg = 15,16 [dB]

Nas medições objetivas de desempenho (SNRseg), o Codec 7 foi o melhor. Na avaliação subjetiva, tanto o Codec 7 como o Codec 9 obtiveram um desempenho similar. Porém, o Codec 9 precisa de um tempo de processamento maior já que nele a divisão é em 16 sub-bandas. Isto implica, também, numa maior complexidade de implementação. Por estas duas razões não se justifica a sua utilização e, portanto, foi escolhido o Codec 7 para testes em canal ruidoso. Os resultados obtidos foram os seguintes:

Taxa de erro= 0 : $\begin{cases} \text{SNR} = 17,16 \text{ [dB]} \\ \text{SNRseg} = 15,18 \text{ [dB]} \end{cases}$

$$\text{Taxa de erro} = 10^{-4} : \begin{cases} \text{SNR} = 16,20 \text{ [dB]} \\ \text{SNRseg} = 14,74 \text{ [dB]} \end{cases}$$

$$\text{Taxa de erro} = 10^{-3} : \begin{cases} \text{SNR} = 10,83 \text{ [dB]} \\ \text{SNRseg} = 12,23 \text{ [dB]} \end{cases}$$

$$\text{Taxa de erro} = 10^{-2} : \begin{cases} \text{SNR} = 2,46 \text{ [dB]} \\ \text{SNRseg} = 2,89 \text{ [dB]} \end{cases}$$

Em forma gráfica:

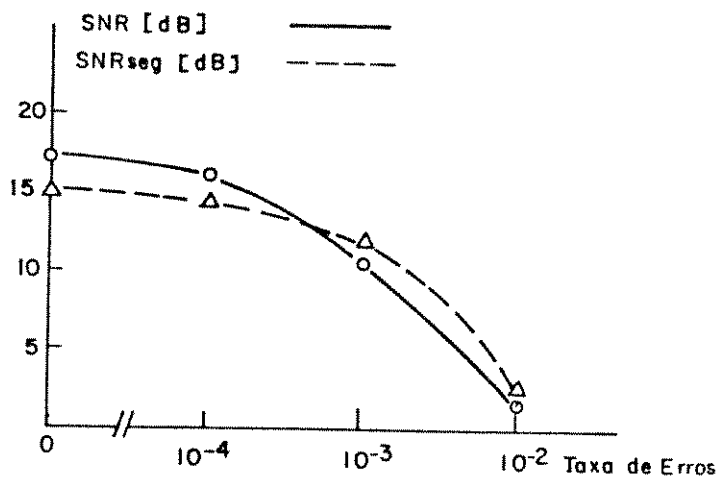


Figura 5.2: Comportamento do Codec 7 com taxa de erros.

Quanto à inteligibilidade do codec, para taxas de erros de 10^{-4} e 10^{-3} é totalmente inteligível e a qualidade permanece aproximadamente a mesma que com taxa de erros nula. Entretanto, para taxa de erros de 10^{-2} o codec é ainda inteligível na maior parte das palavras, mas aparece um ruído bastante molesto durante a fala que dificulta o reconhecimento do locutor.

5.5.3 À TAXA DE 24 [Kbits/s]

(a) N= 4 sub-bandas: B[1] a B[4]

Largura de cada sub-banda: 1000 [Hz].

Bandas transmitidas: B[1] a B[3] (0 a 3000 [Hz]).

Banco QMF utilizado: 1o estágio: 32 "taps" - Código C

2o estágio: 16 "taps" - Código C

CODEC 10

Sub-banda	Bits/am	Codif.	Bits I.L.	$\Delta k_{\min}$
1	4	BC-PCM	5	—
2	4	BC-PCM	5	—
3	3	BC-PCM	5	—

Taxa de Informação Lateral = 137,5 [bits/s]

Taxa Total = 22937,5 [bits/s]

SNR = 21,16 [dB]

SNRseg = 18,97 [dB]

(b) N= 8 sub-bandas: B[1] a B[8]

Largura de cada sub-banda: 500 [Hz].

Bandas transmitidas: B[1] a B[6] (0 a 3000 [Hz]).

Banco QMF utilizado: 1o estágio: 32 "taps" - Código C

2o estágio: 16 "taps" - Código C

3o estágio: 8 "taps"

CODEC 11

Sub-banda	Bits/am	Codif.	Bits I.L.	$\Delta k_{\min}$
1	5	BC-PCM	5	—
2	4	BC-PCM	5	—
3	4	BC-PCM	5	—
4	3	BC-PCM	5	—
5	3	BC-PCM	5	—
6	3	BC-PCM	5	—

Taxa de Informação Lateral = 1875 [bits/s]

Taxa Total = 23785 [bits/s]

SNR = 24,22 [dB]

SNRseg = 21,80 [dB]

CODEC 12

Sub-banda	Bits/am	Codif.	Bits I.L.	$\Delta k_{\min}$
1	5	BC-PCM	5	—
2	5	BC-PCM	5	—
3	4	BC-PCM	5	—
4	3	BC-PCM	5	—
5	3	BC-PCM	5	—
6	2	BC-PCM	5	—

Taxa de Informação Lateral = 1875 [bits/s]

Taxa Total = 23875 [bits/s]

SNR = 25,83 [dB]

SNRseg = 22,14 [dB]

(c) N= 16 Sub-bandas: B[1] a B[16]

Largura de cada sub-banda: 250 [Hz].

Bandas transmitidas: B[1] a B[12] (0 a 3000 [Hz]).

Banco QMF utilizado: 1º estágio: 32 "taps" - Código C

2º estágio: 16 "taps" - Código C

3º estágio: 8 "taps"

4º estágio: 8 "taps"

CODEC 13

Sub-banda	Bits/am	Codif.	Bits I.L.	Δk_{min}
1	5	BC-PCM	5	—
2	5	BC-PCM	5	—
3	5	BC-PCM	5	—
4	4	BC-PCM	5	—
5	4	BC-PCM	5	—
6	4	BC-PCM	5	—
7	3	BC-PCM	5	—
8	3	BC-PCM	5	—
9	2	BC-PCM	5	—
10	2	BC-PCM	5	—
11	2	BC-PCM	5	—
12	1,5	BC-PCM	5	—

Taxa de Informação Lateral = 3750 [bits/s]

Taxa Total = 24000 [bits/s]

SNR = 25,72 [dB]

SNRseg = 21,49 [dB]

CODEC 14 (ao dorso)

Sub-banda	Bits/am	Codif.	Bits I.L.	$\Delta k_{\min}$
1	5	BC-PCM	5	—
2	5	BC-PCM	5	—
3	5	BC-PCM	5	—
4	5	BC-PCM	5	—
5	5	BC-PCM	5	—
6	4	BC-PCM	5	—
7	3	APCM-AQB	—	2
8	3	APCM-AQB	—	2
9	3	APCM-AQB	—	1
10	2	APCM-AQB	—	1
11	2	APCM-AQB	—	1
12	2	APCM-AQB	—	1

Taxa de Informação Lateral = 1875 [bits/s]

Taxa Total = 23875 [bits/s]

SNR = 22,04 [dB]

SNRseg = 19,98 [dB]

Nos testes objetivos o Codec 12 obteve o melhor desempenho. Nos testes subjetivos, todos os codecs apresentados tiveram excelente desempenho, destacando-se os Codecs 12 e 13. Devido ao fato de que o Codec 13 implica numa divisão em 16 sub-bandas, com o conseqüente aumento em complexidade, o Codec 12 foi o escolhido para testes em canal ruidoso. Os resultados obtidos foram os seguintes:

Taxa de erro = 0 : $\begin{cases} \text{SNR} = 25,83 \text{ [dB]} \\ \text{SNRseg} = 22,14 \text{ [dB]} \end{cases}$

Taxa de erro = 10^{-4} : $\begin{cases} \text{SNR} = 23,87 \text{ [dB]} \\ \text{SNRseg} = 21,64 \text{ [dB]} \end{cases}$

$$\text{Taxa de erro} = 10^{-2} : \begin{cases} \text{SNR} = 11,69 \text{ [dB]} \\ \text{SNRseg} = 16,84 \text{ [dB]} \end{cases}$$

$$\text{Taxa de erro} = 10^{-3} : \begin{cases} \text{SNR} = 2,77 \text{ [dB]} \\ \text{SNRseg} = 4,25 \text{ [dB]} \end{cases}$$

Na figura seguinte estão sumarizados estes resultados:

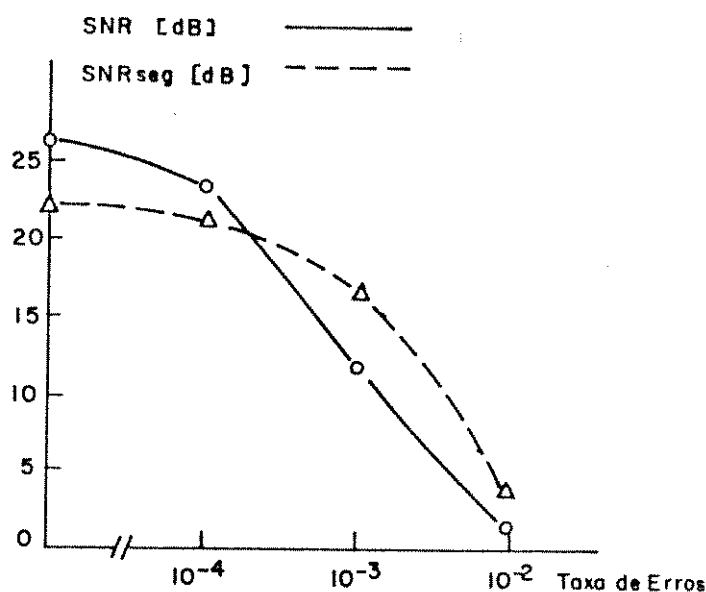


Figura 5.3: Comportamento do Codec 12 com taxa de erros.

Na avaliação subjetiva do Codec 12, quando submetido a taxas de erros, verifica-se que para valores de 10^{-4} e 10^{-3} o codec continua inteligível e de boa qualidade, sendo que para taxa de erros de 10^{-4} , não se percebe diferença com o codec original. Para taxa de erros com valor de 10^{-2} , o codec torna-se muito ruidoso sendo ainda inteligível na maior parte das palavras. Porém, é detectável alguma dificuldade no reconhecimento do locutor.

Como pode-se observar das páginas anteriores, em todos os codecs apresentados foram utilizados bancos de filtros QMF

de 32, 16, 8 e 8 "taps" no 1º, 2º, 3º e 4º estágios respectivamente. O ganho médio obtido ao se utilizar filtros de 48, 24, 12 e 8 "taps" ao invés dos mencionados, foi de 0,3 [dB] e com filtros de 64, 32, 16 e 8 "taps" foi de 0,6 [dB].

Nas figuras seguintes se resumem os principais resultados obtidos. Na Fig.5.4 é mostrado o valor da *relação sinal/ruído convencional*, em função do número de sub-bandas em que se divide o espectro, para taxas de transmissão de 8, 16 e 24 [Kbits/s]. Na Fig.5.5 é mostrado o valor da *relação sinal/ruído segmental*, em função destes parâmetros.

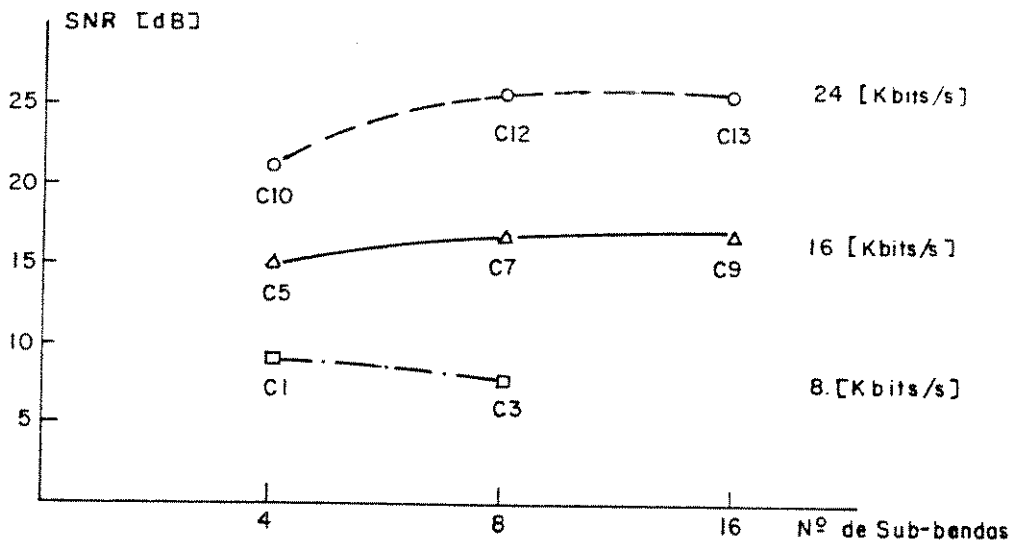


Figura 5.4: Valor da SNR convencional em função do número de sub-bandas em que se divide o espectro e da taxa de transmissão (Ci indica o código).

SNR C1 = 8,84 [dB]	;	SNRseg C1 = 7,43 [dB]
SNR C3 = 7,62 [dB]	;	SNRseg C3 = 5,94 [dB]
SNR C5 = 14,91 [dB]	;	SNRseg C5 = 12,95 [dB]
SNR C7 = 17,16 [dB]	;	SNRseg C7 = 15,18 [dB]
SNR C9 = 17,71 [dB]	;	SNRseg C9 = 15,16 [dB]
SNR C10 = 21,16 [dB]	;	SNRseg C10 = 18,97 [dB]
SNR C12 = 25,83 [dB]	;	SNRseg C12 = 22,14 [dB]
SNR C13 = 25,72 [dB]	;	SNRseg C13 = 21,49 [dB]

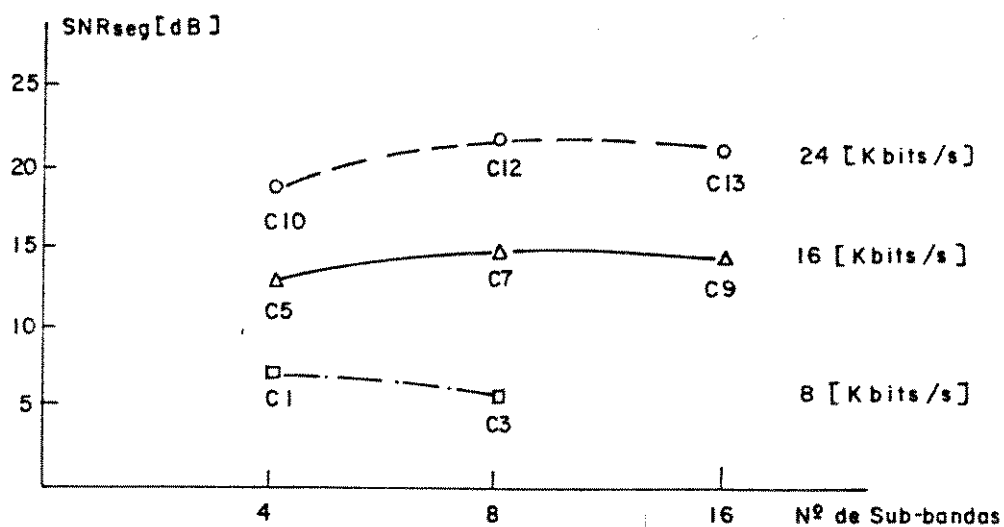


Figura 5.5: Valor da SNR segmental em função do número de sub-bandas em que se divide o espectro e da taxa de transmissão (Ci indica o codec).

Na figura seguinte, 5.6, é mostrado o valor da relação sinal/ruído segmental obtida pelos melhores codecs trabalhando a cada taxa (Codec 3 a 8 [Kbits/s], Codec 7 a 16 [Kbits/s] e Codec 12 a 24 [Kbits/s]), para taxas de erros de 10^{-4} , 10^{-3} e 10^{-2} .

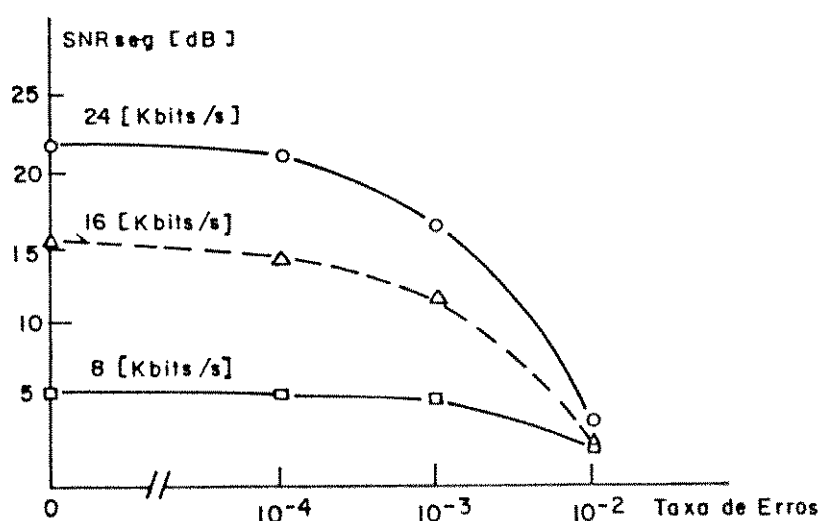


Figura 5.6: Valor da SNR segmental dos melhores codecs, em função da taxa de transmissão e da taxa de erros do canal.

Finalmente, na Fig. 5.7, é mostrado o valor da relação sinal/ruído segmental para as diferentes taxas de erros, em função da taxa total do codificador.

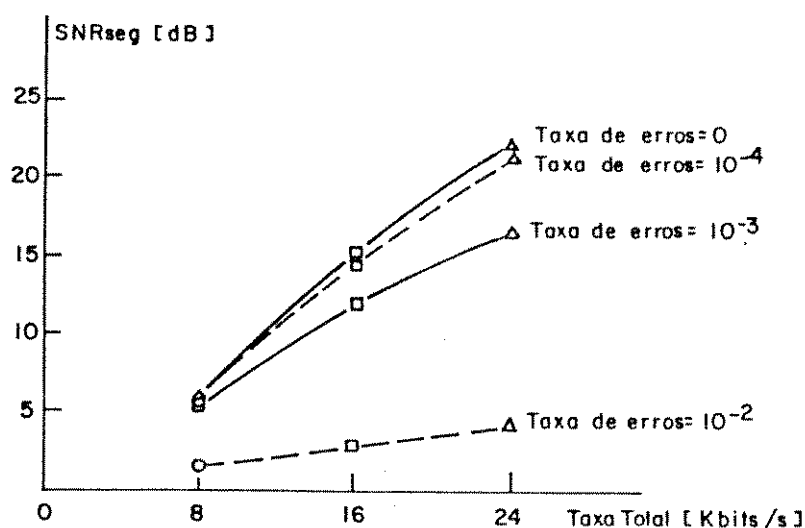


Figura 5.7: Valor da SNR segmental dos codecs a diferentes taxas de erros, em função da taxa total do codificador.

Quanto ao tempo de processamento gasto pelo programa tem-se que, para um codificador com as características do Codec 7, isto é: $N=8$ sub-bandas, taxa final $\cong 16$ [Kbits/s], banco QMF de 32, 16 e 8 "taps" respectivamente e codificação BC-PCM nas bandas baixas e APCM-AQB nas bandas altas, o programa utiliza 20,3 minutos para processar um arquivo de 10,5 segundos (aproximadamente 120 segundos de processamento para cada segundo do arquivo original). Deste tempo, aproximadamente 98% corresponde à execução da filtragem QMF, e 2% à codificação das amostras.

5.6 COMPARAÇÃO COM OUTROS CODECS

Nas referências 3,4,5 e 6, podem-se encontrar resultados de comparações feitas entre codificadores de sub-ban-

das trabalhando a diferentes taxas (9,6 , 16 e 32 [Kbits/s]), e codificadores ADM, AD-PCM e BC-PCM "full-band". Esses estudos indicam uma vantagem dos codificadores de sub-bandas, de aproximadamente 10 [Kbits/s], quando são comparados com os outros codificadores mencionados. Assim um codificador de sub-bandas trabalhando à taxa de 9,6 [Kbits/s], é equivalente a um ADM de 19 [Kbits/s]. De igual modo um codificador de sub-bandas de 16 [Kbits/s], é aproximadamente equivalente a um ADPCM de 26 [Kbits/s].

Para complementar os estudos anteriores, neste trabalho se realizaram comparações dos codificadores de sub-bandas com os codificadores PCM Lei A de 64 [Kbits/s] e ADPCM de 32 [Kbits/s]. O primeiro deles com uma $SNR_{seg} = 35,14$ [dB] e o segundo com uma $SNR_{seg} = 27,09$ [dB]. O objetivo da comparação foi achar o codificador de sub-bandas que nos testes subjetivos tivesse uma qualidade equivalente aos codificadores PCM e ADPCM mencionados¹. Desta forma determinou-se que o codificador de sub-bandas seguinte (Codec 15), trabalhando à taxa de 28 [Kbits/s], iguala, subjetivamente, o desempenho do PCM-Lei A de 64 [Kbits/s] e do ADPCM de 32 [Kbits/s]. A avaliação foi feita utilizando-se um arquivo de voz com 3 diferentes níveis de potência: arquivo original, arquivo original com ganho de amplitude igual a 1/2 (variação de -6 [dB] na potência), e arquivo original com ganho de amplitude igual a 1/4 (variação de -12 [dB] na potência).

CODEC 15

N= 8 sub-bandas: B[1] a B[8].

Largura de cada sub-banda: 500 [Hz].

Bandas transmitidas: B[1] a B[6] (0 a 3000 [Hz]).

¹ Subjetivamente estes dois codificadores apresentam a mesma qualidade e inteligibilidade, ainda que, objetivamente, o PCM-Lei A seja superior ao ADPCM em aproximadamente 8 [dB].

Banco QMF utilizado: 1o estágio: 32 "taps" - Código C
 2o estágio: 16 "taps" - Código C
 3o estágio: 8 "taps"

Sub-banda	Bits/am	Codif.	Bits I.L.	$\Delta k_{\min}$
1	5	BC-PCM	5	—
2	5	BC-PCM	5	—
3	5	BC-PCM	5	—
4	4	BC-PCM	5	—
5	4	BC-PCM	5	—
6	3	BC-PCM	5	—

Taxa de Informação Lateral = 1875 [bits/s]

Taxa Total = 27875 [bits/s]

SNR = 27,42 [dB]

SNRseg = 24,79 [dB]

Pode observar-se que, ainda que este codificador possua um desempenho subjetivo equivalente aos codificadores antes mencionados, as medições objetivas dão aproximadamente 10,5 [dB] de vantagem para o PCM-Lei A e 2,3 [dB] para o ADPCM. A explicação deste fato é que à medida que a taxa de bits de um codificador aumenta, variações significativas da relação sinal/ruído convencional ou segmental não se traduzem, necessariamente, numa variação subjetiva equivalente. Isto ocorre porque acima de um certo limiar de qualidade, o ouvido humano não percebe de forma tão evidente melhoras no desempenho de um codificador.

Desta forma a utilização de codificação de voz em sub-bandas permitiria alcançar, em princípio, qualidade suficiente para aplicações em telefonia, com menos da metade da taxa de bits do PCM em uso atualmente.

REFERÊNCIAS

- [1] Fábio Violaro, "Nova Versão do Sistema de Análise e Processamento Digital de Voz SAPDV-A", 7º Simpósio Brasileiro de Telecomunicações, Florianópolis, SC, Brasil, 1989, pp 50-53.
- [2] R.E.Crochiere, "An Analysis of 16 Kb/s Sub-Band Coder Performance: Dynamic Range, Tandem Connections and Channel Errors", Bell System Tech.J., Vol 57, Nº 8, Oct 1978, pp 2927-2952.
- [3] R.E.Crochiere, S.A.Webber and J.L.Flanagan, "Digital Coding of Speech in Sub-bands", Bell System Tech. J., Vol 55, Nº 8, October 1976, pp 1069-1085.
- [4] R.E.Crochiere, "On the Design of Sub-band Coders for Low-Bit-Rate Speech Communication", Bell System Tech. J., Vol 56, Nº 5, May-June 1977, pp 747-770.
- [5] R.E.Crochiere and J.L.Flanagan, "Sub-band Encoding of Speech", International Conference on Comm., 1977, Vol 1, pp 293-296.
- [6] D.Esteban and C.Galand, "Application of Quadrature Mirror Filters to Split Band Voice Coding Schemes", Proc. ICASSP, May 1977, pp 191-195.

CAPÍTULO 6

CONCLUSÕES

CONTEÚDO

6.1 - Análise dos resultados obtidos	126
--	-----

6.1 ANÁLISE DOS RESULTADOS OBTIDOS

Na descrição dos objetivos deste trabalho (Cap.1 e Cap.5, Seção 5.2), foram enumerados os vários parâmetros que podem ser alterados nos codificadores de sub-bandas, para se avaliar em que medida eles influenciam no desempenho do codec. Através da análise dos resultados obtidos durante as simulações, e de resultados prévios enumerados ao longo dos capítulos restantes, é possível se elaborar as seguintes conclusões:

- 1) Para taxas entre 8 e 24 [Kbits/s], os codificadores de sub-bandas possuem um desempenho superior a outros sistemas de codificação de menor ou igual complexidade (PCM, ADM, ADPCM). Isto é, para um mesmo nível de qualidade subjetiva, os sistemas SBC possuem uma vantagem de pelo menos 10 [Kbits/s] sobre os codificadores acima exemplificados. Consequentemente para uma mesma taxa de transmissão, os sinais obtidos com os codificadores de sub-bandas apresentam uma qualidade superior.

- 2) Acima de 24 [Kbits/s], e levando-se em conta que superado certo limiar de qualidade o ouvido não percebe claramente diferenças de desempenho, a distância entre os codificadores de sub-bandas e outros sistemas de menor ou igual complexidade, pode superar os 10 [Kbits/s]. Assim, o Codec 15 trabalhando a 28 [Kbits/s] equivale ao PCM-Lei A de 64 [Kbits/s].
- 3) Quanto ao número de sub-bandas que deve ser utilizado na divisão do espectro, as Figs. 5.4 e 5.5 mostram que, para taxas de 16 a 24 [Kbits/s], ao se passar de 4 para 8 sub-bandas de igual largura se produz um ganho objetivo de aproximadamente 2,5 [dB] na SNRseg. Subjetivamente a qualidade do sinal reconstruído também melhora. Porém, ao se passar de 8 para 16 sub-bandas de igual largura o desempenho objetivo praticamente não se altera. Quanto à avaliação subjetiva, não se percebe também nenhuma vantagem na divisão em 16 sub-bandas, motivo pelo qual não é recomendável se realizar esta divisão, já que ela exige um tempo de processamento maior e aumenta a complexidade do sistema. A divisão em apenas 2 sub-bandas também não traz benefício significativo comparado com a codificação "full-band", e portanto não é também recomendada. Assim, se conclue que a divisão em 8 sub-bandas é a que apresenta um desempenho melhor, pelo menos nas taxas de transmissão simuladas. Por outro lado, não foi testada a divisão do espectro em bandas de largura diferente. Fica então como sugestão para trabalhos futuros, a simulação de sistemas onde o espectro seja dividido em bandas de largura diferente.
- 4) Para a escolha do número de "taps" dos filtros QMF, os testes revelaram que a utilização de mais de 32 "taps" no primeiro estágio da estrutura filtrante, não introduz nenhum ganho apreciável. Assim, a substituição de um banco de filtros QMF de 32/16/8/8 "taps" respectivamente para um

outro de 64/32/16/8 "taps", produz uma melhora de apenas 0,6 [dB], às custas de um aumento considerável no tempo de processamento. Adicionalmente, deve-se levar em consideração que o número de "taps" dos filtros QMF determina o atraso do sistema (Eqs. (3.52) e (3.56), Cap.3). Portanto não é aconselhável o uso de mais de 32 ou, como máximo, 48 "taps" para estes filtros. Deve-se lembrar também que o número de "taps" dos filtros pode ser reduzido pela metade ao se passar de um estágio para outro, sem provocar degradação aparente no sinal reconstruído.

- 5) Quanto ao método de codificação empregado nas diferentes sub-bandas e ao número de bits atribuído a cada uma delas, deve-se tentar codificar as primeiras bandas (até os 1500 [Hz] aproximadamente), com a maior precisão possível. Desta forma se recomenda utilizar BC-PCM para as primeiras bandas e BC-PCM ou algum outro método de codificação, para as bandas restantes. Não se recomenda, em princípio, o uso de ADPCM, já que a correlação entre amostras pertencentes a uma dada sub-banda é baixa. No uso de BC-PCM, se recomenda o uso da amostra máxima do bloco como parâmetro de adaptação do quantizador, evitando assim a ocorrência de distorção nas bandas baixas ("clipping"). Para as bandas altas, ao se utilizar BC-PCM, não é indispensável o uso da amostra máxima do bloco como parâmetro de adaptação do quantizador. Pode-se, em seu lugar, utilizar o desvio padrão σ_x ou a variância σ_x^2 . Considerando-se que nos codificadores BC-PCM uma parte dos bits deve ser destinada à codificação e transmissão do parâmetro de adaptação do quantizador, e que deste número de bits depende: i) a taxa de informação lateral, ii) os bits que restam para codificar as amostras e iii) a qualidade do sinal reconstruído, chegou-se à conclusão que um valor ideal para a codificação deste parâmetro é 5 [bits/bloco], pelo menos para as taxas que foram simuladas neste trabalho. Quanto à tabela de valores do parâmetro de adaptação do

quantizador, ela foi obtida empiricamente. É portanto recomendável um estudo mais profundo deste problema, visando otimizar os valores desta tabela.

- 6) Na distribuição de bits entre as sub-bandas, como foi dito no item 3, deve-se beneficiar as frequências até 1,5 [KHz]. O número específico de bits/banda dependerá da taxa final que se pretende atingir. Na descrição dos parâmetros dos Codecs 3, 7 e 12 do Cap.5, são fornecidos os valores que produziram os melhores resultados para taxas finais de 8, 16 e 24 [Kbits/s] respectivamente. De igual modo, os parâmetros do Codec 15 resultaram num sistema SBC com desempenho equivalente ao ADPCM de 32 [Kbits/s] e ao PCM-Lei A de 64 [Kbits/s]. Segundo a literatura existente, (ver referências dos capítulos anteriores), ao se utilizar uma atribuição dinâmica de bits podem-se obter codificadores com qualidade equivalente e com uma diminuição de 2,5 [Kbits/s] na taxa final. Portanto, para trabalhos futuros na área, se sugere o estudo desta possibilidade.
- 7) Quanto ao comportamento dos sistemas em canal não ideal, com taxas de erros de 10^{-4} , 10^{-3} e 10^{-2} , devem-se destacar alguns pontos:
 - i) Os sistemas foram testados sem a utilização, na codificação das amostras, de qualquer código detector ou corretor de erros. O simples uso do código de Gray, poderia melhorar o desempenho do sistema sem aumentar a taxa final de bits.
 - ii) Nas bandas codificadas com BC-PCM, só as amostras foram submetidas à taxas de erros. Foi suposto que a informação lateral estava adequadamente protegida e, portanto, isenta de erros.
 - iii) Nas bandas codificadas com APCM-AQB, foi utilizado o algoritmo robusto de Goodman e Wilkinson na adaptação

do passo de quantização do quantizador. O valor de β utilizado no algoritmo foi de 63/64, que causa uma degradação menor que 0,4 [dB] na SNRseg para canais com taxa de erros nula. Foi também testado o valor de 31/32, mas foi descartado devido à maior degradação produzida em canais com taxa de erros nula (maior que 2 [dB] na SNRseg).

Os resultados obtidos, em relação a este comportamento em canal não ideal, foram aproximadamente iguais para todos os codificadores, em particular:

À taxa de 8 [Kbits/s] (Codec 3): o comportamento do codec com taxa de erros de 10^{-4} , não se diferencia, subjetivamente, do comportamento do codec em canal ideal. O desempenho objetivo cai, entretanto, 0,03 [dB]. Para taxa de erros de 10^{-3} , o desempenho objetivo cai apenas 0,44 [dB] com relação ao canal ideal, sendo, subjetivamente, quase indistinguível do original. Para taxa de erros de 10^{-2} , o desempenho cai consideravelmente, tanto objetiva como subjetivamente. O codec continua, entretanto, bastante inteligível mas percebe-se muito ruído durante a fala e dificuldade na identificação do locutor.

À taxa de 16 [Kbits/s] (Codec 7): o comportamento do codec com taxa de erros de 10^{-4} é subjetivamente o mesmo que com taxa de erros nula, embora objetivamente apareça alguma degradação (inferior a 0,5 [dB]). No caso de taxa de erros de 10^{-3} , o desempenho objetivo cai aproximadamente 3 [dB], sendo totalmente inteligível e de boa qualidade. Para taxa de erros de 10^{-2} , o sistema continua ainda inteligível mas percebe-se bastante ruído durante a fala. Como no codec anterior (8 [Kbits/s]), existe também dificuldade no reconhecimento do locutor. Comparando as Figs. 5.1 e 5.2, percebe-se que o Codec 7 (16 [Kbits/s]), com taxa de erros de 10^{-3} , tem ainda um desem-

penho superior ao Codec 3 (8 [Kbits/s]) com taxa de erros nula. Os testes subjetivos confirmaram este comportamento.

À taxa de 24 [Kbits/s] (Codec 12): neste caso, ao se passar de canal ideal para canal com taxa de erros de 10^{-4} , o desempenho objetivo cai apenas 0,5 [dB]. Subjetivamente o desempenho permanece inalterado. Para taxa de erros de 10^{-3} , o desempenho objetivo cai bastante (4,3 [dB]), mas, subjetivamente, o codec continua totalmente inteligível e de boa qualidade. Para taxa de erros de 10^{-2} o desempenho cai para 4,25 [dB]. Subjetivamente torna-se altamente ruidoso durante a fala, sendo ainda inteligível a maior parte das palavras. Percebe-se, entretanto, certa dificuldade na identificação do locutor. Comparando as Figs.5.2 e 5.3, nota-se que o Codec 12 (24 [Kbits/s]) com taxa de erros de 10^{-3} , tem um desempenho superior em 1,66 [dB] ao Codec 7 (16 [Kbits/s]) em canal ideal. Na avaliação subjetiva, entretanto, o Codec 7 apresenta uma melhor qualidade. Comparando agora as Figs.5.1 e 5.3, percebe-se que o Codec 3 (8 [Kbits/s]), em canal ideal, tem um desempenho superior em aproximadamente 1,7 [dB] ao desempenho do Codec 12 (24 [Kbits/s]), com taxa de erros de 10^{-2} . Subjetivamente, no entanto, a diferença no desempenho é bem maior a favor do Codec 3, já que o Codec 12 a essa taxa de erros, é de qualidade mínima.

Concluindo, os 3 codificadores tem um bom comportamento com taxas de erros de 10^{-4} e 10^{-3} . Para taxa de erros de 10^{-2} o desempenho cai bastante nos três sistemas, aumentando consideravelmente o ruído durante a fala e aparecendo dificuldade na identificação do locutor.

APÊNDICES

APÊNDICE A 143

APÊNDICE B 149

APÊNDICE C 153

APÊNDICE A

DIZIMAÇÃO E INTERPOLAÇÃO DE SEQUÊNCIAS E SUA RELAÇÃO COM

A LARGURA DO ESPECTRO

A finalidade deste apêndice é explicar resumidamente o que ocorre com o espectro de um sinal durante sua dizimação e interpolação (Ref.[1,2,3]).

Seja um sinal real $b(t)$ cujo espectro analógico, $B(j\Omega)$, é limitado em frequência a $\Omega_{\max} = 2\pi \cdot F_{\max}$, segundo mostram as Figs.A.1(a) e (b). Se este sinal é amostrado a taxa de Nyquist, isto é, ao dobro de sua frequência máxima, $F_s = 2 \cdot F_{\max}$ [Hz], a sequência $b(n)$ obtida e seu espectro $B(e^{j\omega})$, serão como nas Fig.A.1(c) e A.1(d) respectivamente.

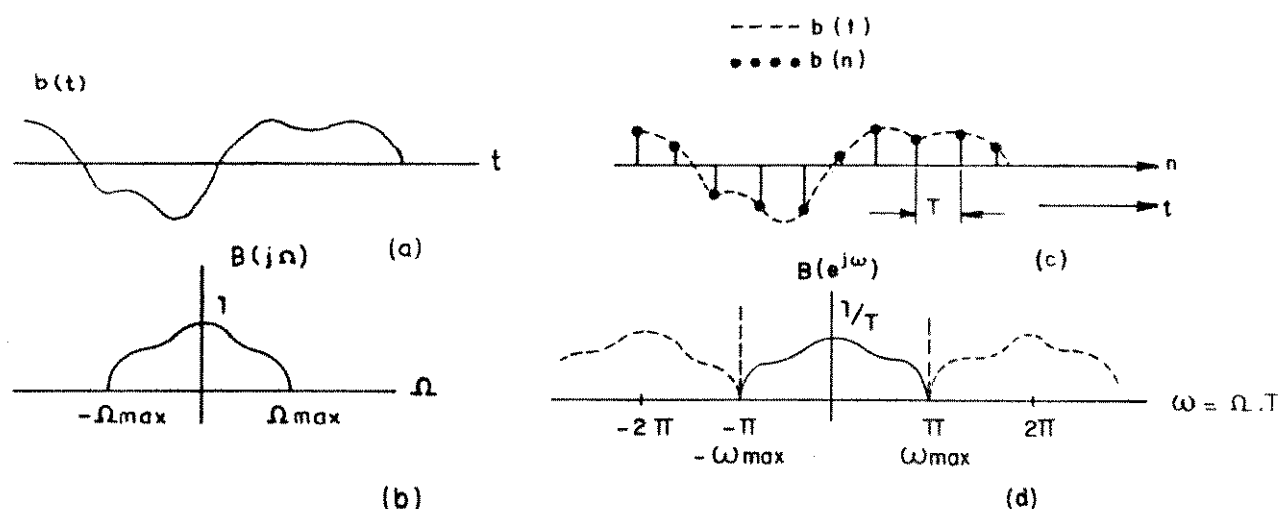


Figura A.1: (a) Sinal $b(t)$; (b) Espectro do sinal $b(t)$; (c) Sequência $b(n)$; (d) Espectro da sequência $b(n)$.

Nestas figuras, $\Omega = 2\pi \cdot f$ [rad/s], é a frequência utilizada na representação do sinal analógico $b(t)$, ao passo que $\omega = \Omega/F_s$ [rad], é a frequência utilizada na representação do sinal amostrado $b(n) = b(nT)$. Da Fig. A.1(d), pode-se ver que o espectro $B(e^{j\omega})$ ocupa todo o eixo de frequências normalizadas ω de $-\pi$ a π , repetindo-se periodicamente ao longo desse eixo com período 2π . A relação entre os espectros $B(j\Omega)$ e $B(e^{j\omega})$, dada pelo Teorema de Amostragem (Ref.[1,2]), é a seguinte:

$$B(e^{j\omega}) = \frac{1}{T} \sum_{r=-\infty}^{\infty} B(j\Omega + j \frac{2\pi r}{T}) \Big|_{\Omega=\omega/T} \quad (a.1)$$

onde T corresponde ao período de amostragem $1/F_s$. Esta equação explicita a periodicidade de $B(e^{j\omega})$ em função de ω .

O fato do espectro $B(e^{j\omega})$ ocupar todo o eixo de frequências normalizadas ω , de $-\pi$ a π , pode ser visto levando-se em conta que:

$$\omega_{\max} = \frac{\Omega_{\max}}{F_s} = \frac{2\pi \cdot F_{\max}}{F_s} = \frac{2\pi \cdot F_{\max}}{2 \cdot F_{\max}} = \pi \quad (a.2)$$

Já que o sinal $b(n)$ é real, $|B(e^{j\omega})|$ possui simetria par e em consequência ocupará a faixa de $-\pi$ a π como mencionado.

Se agora o sinal real $b(t)$ é amostrado a uma taxa superior à de Nyquist, por exemplo ao dobro dessa taxa ($F_s' = 4 \cdot F_{\max}$), então a sequência obtida $c(n)$ e o seu espectro $C(e^{j\omega})$, serão como na Fig.A.2.

Vemos que neste caso a largura do espectro $C(e^{j\omega})$ é a metade do caso anterior, ocupando a faixa de $-\pi/2$ a $\pi/2$. Como antes, o fato do espectro $C(e^{j\omega})$ ocupar o eixo ω de $-\pi/2$ a $\pi/2$, pode ser visto considerando-se que neste caso:

$$\omega_{\max} = \frac{\Omega_{\max}}{F_s'} = \frac{2\pi \cdot F_{\max}}{F_s'} = \frac{2\pi \cdot F_{\max}}{4 \cdot F_{\max}} = \pi/2 \quad (a.3)$$

Como $|C(e^{j\omega})|$ possui simetria par, o espectro ocupará a faixa de $-\pi/2$ a $\pi/2$ como dito.

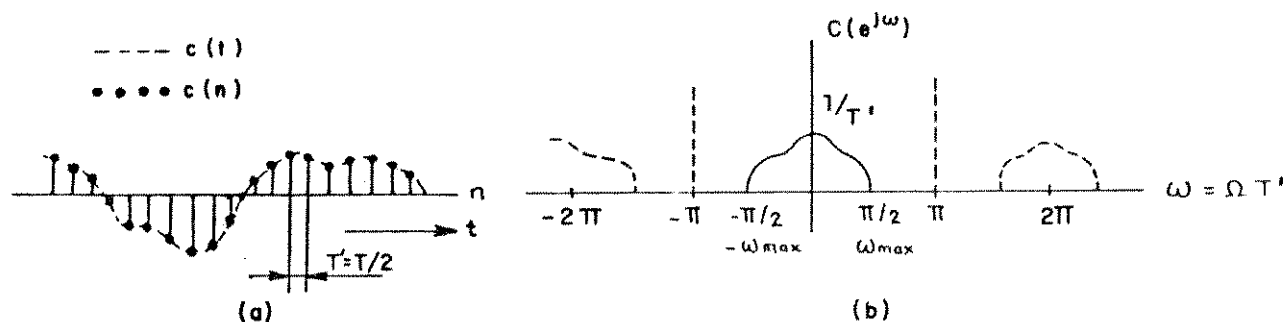


Figura A.2: (a) Sequência $c(n)$; (b) Espectro da sequência $c(n)$.

Deste exemplo pode-se concluir que "amostrar um sinal a uma taxa superior à de Nyquist, faz com que o espectro do sinal resultante possua uma largura de faixa inferior a 2π ". O recíproco também é válido: "se o espectro de um sinal ocupa uma largura de faixa inferior a 2π , é porque este sinal está superamostrado" e, conseqüentemente, possui informação redundante.

Tal sinal pode ser dizimado "sem perda de informação", até que o seu espectro se expanda ocupando a largura de 2π . No exemplo da Fig.A.2, bastaria se dizimar o sinal por um fator 2 para se atingir este objetivo (a dizimação por um fator 2 é realizada descartando-se uma amostra a cada duas da sequência $c(n)$). Após esta dizimação, a sequência obtida e o seu espectro seriam aqueles das Figs.A.1(c) e (d). Se fosse feita a dizimação de $c(n)$ por um fator maior que 2, se produziria superposição dos espectros, gerando a distorção

conhecida como "aliasing" e impossibilitando assim a recuperação do sinal original $b(t)$, sem distorção.

Considere agora que se possui a sequência $b(n)$, de espectro $B(e^{j\omega})$, proveniente da amostragem do sinal $b(t)$ à taxa de Nyquist $F_s = 2 \cdot F_{\max}$ (Figs.A.1(c) e (d)), e deseja-se interpolar esta sequência por um fator 2, de modo a se obter a sequência $c(n)$ de espectro $C(e^{j\omega})$, como na Fig.A.2. Deseja-se então passar de uma sequência $b(n)$ que ocupa toda a faixa de $-\pi$ a π , a uma outra que ocupa a metade dessa faixa e que terá, portanto, informação redundante. Para conseguir este objetivo, gera-se a sequência $d(n)$ agregando-se 1 amostra nula entre cada 2 amostras da sequência original $b(n)$, isto é:

$$d(n) = \begin{cases} b(n/2) & , \text{ para } n \text{ par.} \\ 0 & , \text{ para } n \text{ ímpar.} \end{cases} \quad (\text{a.4})$$

A sequência $d(n)$ obtida e seu espectro $D(e^{j\omega})$, serão aqueles da Fig.A.3.

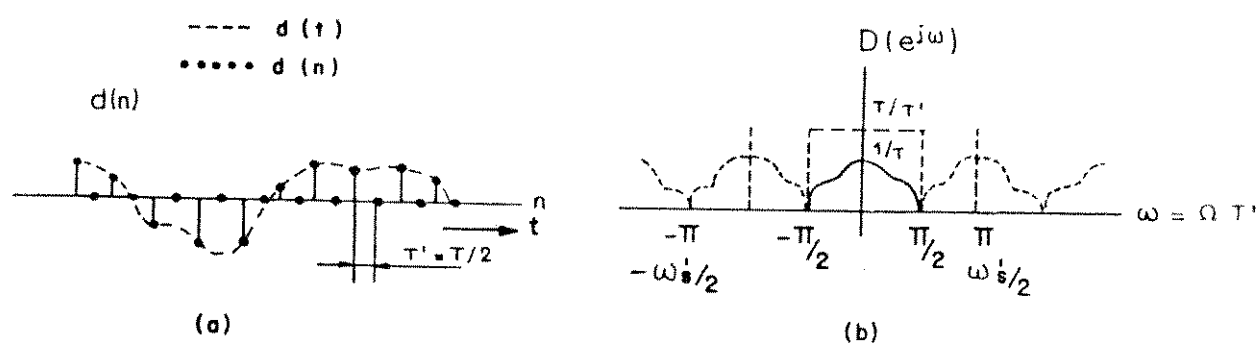


Figura A.3: (a) Sequência $d(n)$; (b) Espectro da sequência $d(n)$.

Nesta figura pode-se ver que o espectro do sinal original $B(e^{j\omega})$, se repete duas vezes no intervalo $-\pi$ a π .

São geradas, pois, imagens do espectro original no intervalo $-\pi$ a π . Filtrando-se a sequência $d(n)$ com um filtro passa-baixas de ganho T/T' ($T = 1/F_s$; $T' = 1/F_s' = 1/2F_s$), e frequência de corte $\omega_s'/4 = \pi/2$, as imagens são eliminadas, obtendo-se a sequência $c(n)$, de espectro $C(e^{j\omega})$, como desejado (Ref.[1]). Esta filtragem é indicada pela linha tracejada na Fig.A.3(b).

É claro que também poderia ser feita a interpolação e/ou a dizimação do sinal por um fator diferente de 2, de forma a conseguir que seu espectro adquira uma largura de faixa determinada. Para o caso de se dizimar o sinal original por um fator M , " *$M-1$ amostras de cada M amostras da sequência original devem ser descartadas*", isto é:

$$y(n) = x(Mn) \quad (a.5)$$

onde $x(n)$ corresponde à sequência original de taxa F_s [am/s], e $y(n)$ à sequência já dizimada, de taxa $F_s' = F_s/M$ [am/s]. Neste caso pode-se demonstrar que a relação entre os espectros de ambas sequências é dado por (Ref.[1]):

$$Y(e^{j\omega}) = \frac{1}{M} \sum_{k=0}^{M-1} X(e^{j(\omega - 2k\pi)/M}) \quad (a.6)$$

com $M = F_s/F_s'$, $F_s = 1/T$ e $F_s' = 1/T'$.

Nesta situação, se $F_s' [\text{Hz}] = F_s/M [\text{Hz}] > 2 \cdot F_{\text{max}} [\text{Hz}]$ (2 vezes a frequência máxima do sinal original $x(t)$), a dizimação por um fator M não produz "aliasing" e o sinal original, $x(t)$, pode ser recuperado a partir da sequência dizimada $y(n)$.

Por outro lado, para o caso de se interpolar a sequência original por um fator M , " *$M-1$ amostras nulas devem ser intercaladas entre cada duas amostras da sequência original*",

isto é:

$$v(n) = \begin{cases} x(n/M) & , n = 0, \pm M, \pm 2M, \pm 3M, \dots \\ 0 & , \text{demais valores de } n. \end{cases} \quad (\text{a.7})$$

onde $x(n)$ corresponde à sequência original de taxa F_s [am/s], e $v(n)$ à sequência com zeros intercalados e taxa F_s' [am/s]. Adicionalmente deve-se submeter a sequência resultante a um filtro de ganho M e frequência de corte $\pi/M = \omega_s'/2M$. Demonstra-se (Ref. 1)], que os espectros destas sequências estão relacionados através da expressão:

$$V(e^{j\omega}) = X(e^{j\omega M}) \quad (\text{a.8})$$

com $M = F_s'/F_s$, $F_s' = 1/T'$ e $F_s = 1/T$. Portanto, filtrando-se $v(n)$ com o filtro mencionado, obtém-se a sequência $y(n) = x(nT)$ desejada.

REFERÊNCIAS

- [1] L.R.Rabiner and R.W.Schafer, "Digital Processing Of Speech Signals", Prentice Hall, Inc., Englewood Cliffs, NJ, 1978.
- [2] A.V.Oppenheim and R.W.Schafer, "Digital Signal Processing", Prentice Hall, Inc., Englewood Cliffs, NJ, 1975.
- [3] R.E.Crochiere and L.R.Rabiner, "Interpolation and Decimation of Digital Signals - A Tutorial Review", Proceedings of IEEE, Vol 69, No 3, March 1981, pp 300-331.

APÊNDICE B

QUANTIZADORES ÓTIMOS

A Tabela B.1 fornece parte dos resultados obtidos por Max (Ref.[1]) e por Paez e Glisson (Ref.[2]), no projeto de quantizadores *uniformes* ótimos para sinais de entrada gaussianos (G), laplacianos (L) e gama (Γ). Na mesma tabela são fornecidos os valores máximos de SNR_Q, obtidos com estes quantizadores. Deve-se ressaltar que a tabela foi construída considerando-se sinais com média nula e variância unitária. Para se obter os valores correspondentes a um sinal com variância diferente de 1, deve-se multiplicar os valores apresentados na tabela pelo desvio padrão σ_x do sinal a quantizar (Fig.B.1). Dado que na codificação de sinais a baixas taxas, dificilmente são utilizados mais de 4 [its/am], a tabela só fornece os parâmetros do quantizador ótimo para $n_k = 1, 2, 3$ e 4 [bits/am].

n_k (bits/am)	$\Delta_{\text{ótimo}} / \sigma_x$			(SNR _Q) _{max} [dB]		
	Distribuição			Distribuição		
	G	L	Γ	G	L	Γ
1	1,5956	1,4142	1,1547	4,40	3,01	1,76
2	0,9957	1,0874	1,0660	9,26	7,07	4,96
3	0,6860	0,7309	0,7867	14,27	11,44	8,78
4	0,3352	0,4610	0,5400	19,38	15,96	13,00

Tabela B.1: Passo de quantização e máxima SNR, para quantizadores uniformes ótimos.

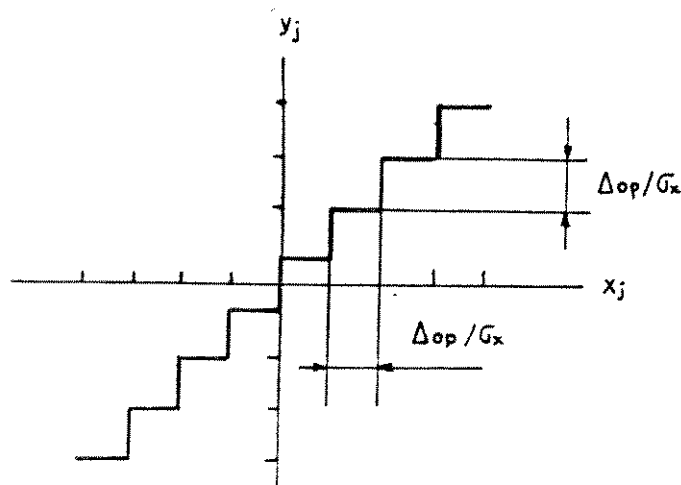


Figura B.1: Interpretação da Tabela B.1.

Nas tabelas seguintes, B.2 e B.3, são fornecidos os níveis de saída ótimos e os intervalos de decisão ótimos, dos quantizadores *não uniformes* projetados para sinais de entrada gaussianos, laplacianos e gama. Os valores destas tabelas, devem ser interpretados com o auxílio da Fig.B.2.

n_k (bits/am)		1		2	
Distrib.	j	x_j	y_j	x_j	y_j
G	1	0,000	0,798	0,000	0,459
	2			0,992	1,610
L	1	0,000	0,707	0,000	0,420
	2			1,127	1,894
Γ	1	0,000	0,577	0,000	0,313
	2			1,268	2,223

Tabela B.2: Níveis de saída e intervalos de decisão ótimos para quantizadores não uniformes de 2 e 4 níveis.

n_k (bits/am)		3		4	
Distrib.	j	x_j	y_j	x_j	y_j
G	1	0,000	0,246	0,000	0,128
	2	0,501	0,755	0,258	0,388
	3	1,050	1,344	0,522	0,657
	4	1,748	2,152	0,800	0,942
	5			1,099	1,255
	6			1,437	1,619
	7			1,844	2,059
	8			2,401	2,733
L	1	0,000	0,233	0,000	0,124
	2	0,533	0,833	0,254	0,405
	3	1,253	1,673	0,567	0,729
	4	2,380	3,087	0,920	1,111
	5			1,345	1,578
	6			1,878	2,178
	7			2,597	3,017
	8			3,725	4,432
Γ	1	0,000	0,155	0,000	0,073
	2	0,527	0,899	0,230	0,387
	3	1,478	2,057	0,591	0,795
	4	3,089	4,121	1,051	1,307
	5			1,533	1,959
	6			2,390	2,822
	7			3,422	4,051
	8			5,128	6,195

Tabela B.3: Níveis de saída e intervalos de decisão ótimos para quantizadores não uniformes de 8 e 16 níveis.

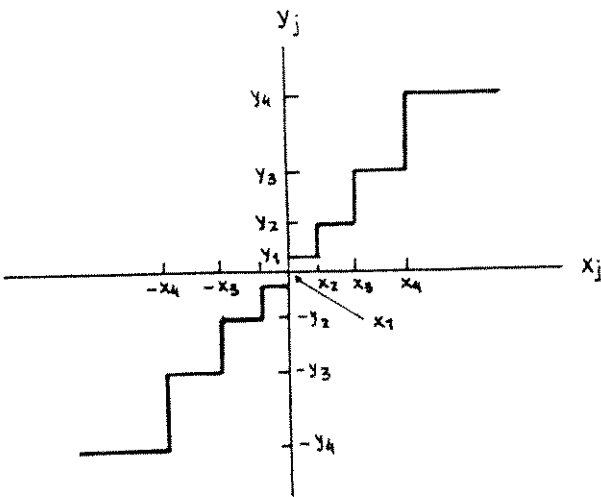


Figura B.2: Interpretação das Tabelas B.2 e B.3.

As tabelas acima só fornecem os valores para os semi-eixos positivos, já que o quantizador possui simetria em relação à origem (Fig.B.2). De maneira análoga à Tabela B.1, os valores fornecidos são para sinais com média nula e variância unitária. Para sinais com variância diferente de 1, os valores dados nas tabelas devem ser multiplicados pelo desvio padrão σ_x do sinal a quantizar.

Finalmente na Tabela B.4 são fornecidos os valores máximos da SNR_q, em dB, para os quantizadores das Tabelas B.2 e B.3 (Ref.[3]).

n_k (bits/am)	SNR _q [dB]		
	G	L	Γ
1	4,40	3,01	1,76
2	9,30	7,54	5,35
3	12,62	12,54	11,52
4	20,22	18,13	17,07

Tabela B.4: Valores da SNR para os quantizadores não uniformes ótimos.

REFERÊNCIAS

- [1] J.Max, "Quantizing for Minimum Distortion", IRE Trans. Inform. Theory, Vol IT 6, March 1960, pp 7-12.
- [2] M.D.Paez and T.H.Glisson, "Minimum Mean-Squared-Error Quantization in Speech PCM and DPCM Systems", IEEE Trans. Comm., Vol COM-20, Apr 1972, pp 225-230.
- [3] N.S.Jayant and P.Noll, "Digital Coding of Waveforms - Principles and Applications to Speech and Video", Prentice Hall Inc., Englewood Cliffs, N.J., 1984.

APÊNDICE C

FAMÍLIA DE FILTROS QMF UTILIZADA NAS SIMULAÇÕES

Os filtros QMF utilizados nas simulações dos codificadores de sub-bandas, foram os seguintes (Ref.[1]):

Tape	Banda de transição Δf	Ripple na banda passante	Atenuação min. na banda de rejeição	Atenuação máxima
8 12 16	A	$\pm 0,060$ [dB]	31 [dB]	31 [dB]
		$\pm 0,025$ [dB]	48 [dB]	50 [dB]
		$\pm 0,008$ [dB]	50 [dB]	75 [dB]
12 16 24	B	$\pm 0,040$ [dB]	33 [dB]	36 [dB]
		$\pm 0,020$ [dB]	44 [dB]	48 [dB]
		$\pm 0,008$ [dB]	50 [dB]	78 [dB]
16 24 32 48	C	$\pm 0,070$ [dB]	30 [dB]	36 [dB]
		$\pm 0,020$ [dB]	44 [dB]	49 [dB]
		$\pm 0,009$ [dB]	51 [dB]	60 [dB]
		$\pm 0,002$ [dB]	53 [dB]	80 [dB]
24 32 48 64	D	$\pm 0,100$ [dB]	30 [dB]	38 [dB]
		$\pm 0,025$ [dB]	38 [dB]	48 [dB]
		$\pm 0,006$ [dB]	50 [dB]	66 [dB]
		$\pm 0,002$ [dB]	56 [dB]	80 [dB]
32 48 64	E	$\pm 0,140$ [dB]	26 [dB]	36 [dB]
		$\pm 0,070$ [dB]	32 [dB]	46 [dB]
		$\pm 0,025$ [dB]	40 [dB]	51 [dB]

Tabela C.1: Família de filtros QMF utilizada nas simulações.

A banda de transição normalizada Δf , é definida como: $\Delta f = (\omega_{sb} - \pi/2)/2\pi$, onde ω_{sb} corresponde à frequência de início da banda de rejeição do filtro ("stop-band frequency"). Na figura seguinte é mostrado o significado dos parâmetros da Tabela C.1.

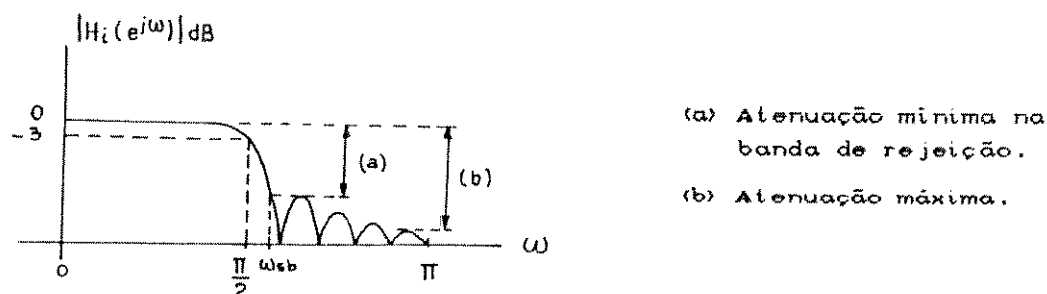


Figura C.1: Significado dos parâmetros da Tabela C.1.

A tabela seguinte, C.2, fornece os valores de Δf em função da letra código da Tabela C.1.

Código da banda de transição	Banda de transição normalizada	ω_{sb} [rad/e]
A	0,14	0,780
B	0,10	0,700
C	0,0625	0,625
D	0,043	0,586
E	0,023	0,546

Tabela C.2: Banda de transição em função da letra código da Tabela C.1.

No exemplo dado na Seção 3.3.2 do Cap.3, foi utilizado o filtro de 32 "taps" - Código D. Nos codificadores apresentados no Cap.5, foram utilizados os filtros de 32 "taps" - Código C, de 16 "taps" - Código C e de 8 "taps" - Código A. Os coeficientes correspondentes a estes filtros são dados a seguir:

8 "taps" - Código A
$h(0)=h(7)= 0,93871500E-02$ $h(1)=h(6)=-0,70651830E-01$ $h(2)=h(5)= 0,69428270E-01$ $h(3)=h(4)= 0,48998080E+00$

Tabela C.3: Coeficientes do filtro de 8 "taps" código A.

16 "taps" - Código C	
$h(0)=h(15)= 0,65256660E-02$	$h(4)=h(11)=-0,26275600E-01$
$h(1)=h(14)=-0,20487510E-01$	$h(5)=h(10)=-0,99295500E-01$
$h(2)=h(13)= 0,19911500E-02$	$h(6)=h(9) = 0,11786660E+00$
$h(3)=h(12)= 0,46476840E-01$	$h(7)=h(8) = 0,47211220E+00$

Tabela C.4: Coeficientes do filtro de 16 "taps" código C.

32 "taps" - Código C	
$h(0)=h(31)= 0,69105790E-03$	$h(8)=h(23)= -0,41874830E-02$
$h(1)=h(30)=-0,14037930E-02$	$h(9)=h(22)= -0,31238620E-01$
$h(2)=h(29)=-0,12683030E-02$	$h(10)=h(21)= 0,14568440E-01$
$h(3)=h(28)= 0,42341950E-02$	$h(11)=h(20)= 0,52947450E-01$
$h(4)=h(27)= 0,14142460E-02$	$h(12)=h(19)=-0,39348780E-01$
$h(5)=h(26)=-0,94583180E-02$	$h(13)=h(18)=-0,99802430E-01$
$h(6)=h(25)=-0,13038590E-03$	$h(14)=h(17)= 0,12855790E+00$
$h(7)=h(24)= 0,17981450E-01$	$h(15)=h(16)= 0,46640530E+00$

Tabela C.5: Coeficientes do filtro de 32 "taps" código C.

32 "taps" - Código D	
$h(0)=h(31)= 2,2451390E-03$	$h(8)=h(23)= -7,9617310E-03$
$h(1)=h(30)=-3,9711520E-03$	$h(9)=h(22)= -3,4964400E-02$
$h(2)=h(29)=-1,9696720E-03$	$h(10)=h(21)= 1,9472180E-02$
$h(3)=h(28)= 8,1819410E-03$	$h(11)=h(20)= 5,4812130E-02$
$h(4)=h(27)= 8,4268330E-04$	$h(12)=h(19)=-4,4524230E-02$
$h(5)=h(26)=-1,4228990E-02$	$h(13)=h(18)=-9,9338590E-02$
$h(6)=h(25)= 2,0694700E-03$	$h(14)=h(17)= 1,3297250E-01$
$h(7)=h(24)= 2,2704150E-02$	$h(15)=h(16)= 4,6367410E-01$

Tabela C.6: Coeficientes do filtro de 32 "taps" código D.

REFERÊNCIAS

- [1] J.D. Johnston, "A Filter Family Designed for Use in Quadrature Mirror Filter Banks", Conf. Rec., ICAASP, Denver, Co., 1980, pp 291-294.

UNIDADE	Be
PROC.	
DESCRIÇÃO, PREÇO ES.	
ATIVIDADE	
DATA	29/10/90