

Ilusões de Áudio

<https://www.youtube.com/watch?v=kzo45hWXRWU>



Ou seja,

- São famosas as “ilusões de ótica”, mas **devido ao funcionamento do nosso sistema auditivo, muitos “ilusões de audição” podem ser criadas, que nos “enganem”** (cfr. KOHLRAUSCH2005).
- As tecnologias de Codificação de Áudio (p.ex. MPEG), ou compensação de ruído (p.ex. Dolby) utilizam os limites e restrições do nosso sistema sensorial-perceptivo auditivo, para obter os ganhos em armazenamento e/ou processamento.

Quais os princípios que motivam?

- Num Sistema PCM sem compressão a taxa de bits e a SNR são constantes (p.ex. 16 bits/96 dB para o CD).
 - Como a percepção humana varia em cada faixa de frequência (vide Modelo Auditivo), porque não se pode variar a taxa de codificação em cada sub-banda?
- Assim, o número de bits alocados em cada sub-banda pode ser adaptado para uma SNR específica de acordo com a banda crítica, diminuindo-se a taxa de bits total.
- Quanto menos bits alocados, menor a qualidade, e vice-versa.

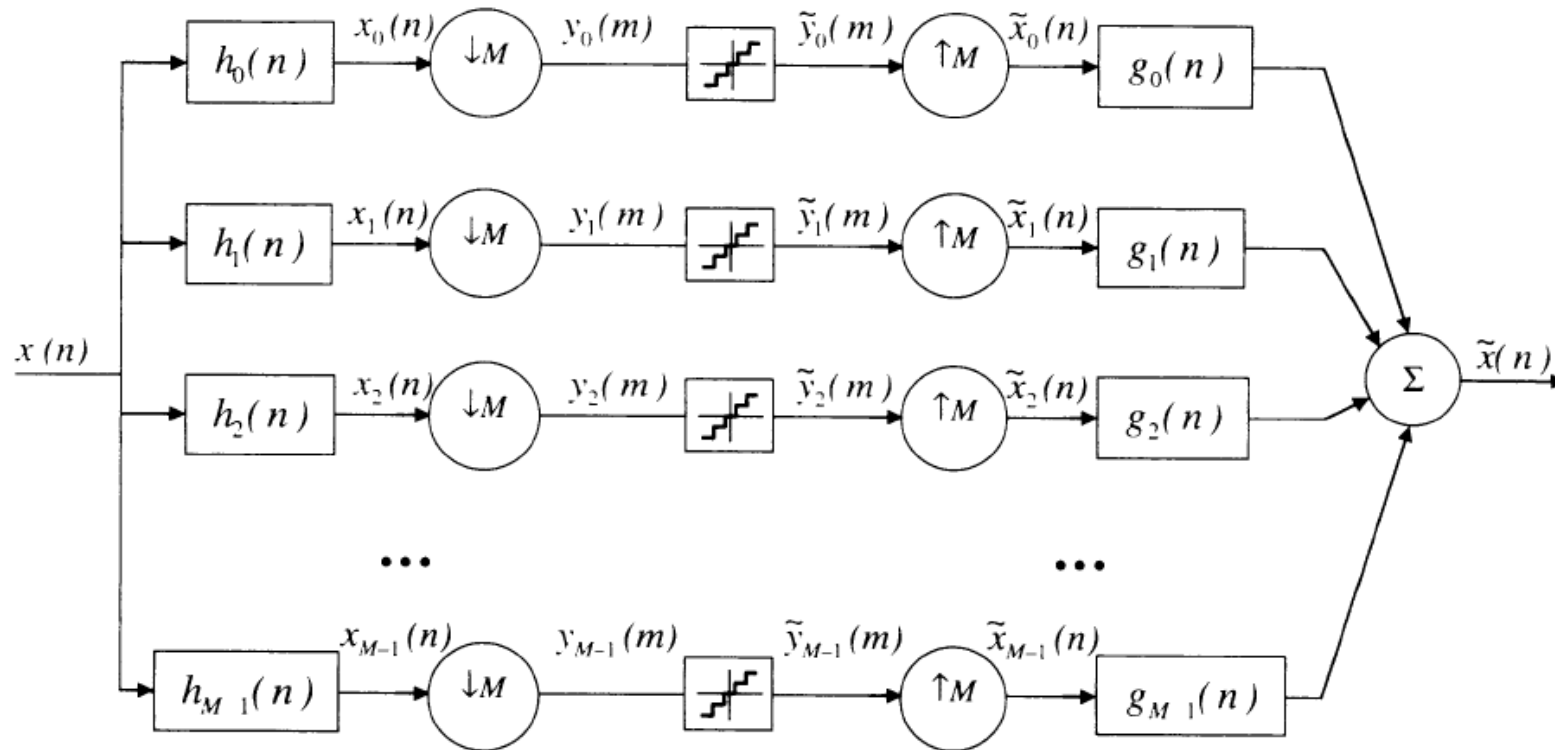
Introdução: Codificadores por Sub-Bandas x Transformada

Sub-bandas: codificadores com poucos canais de frequência como MPEG camadas I e II

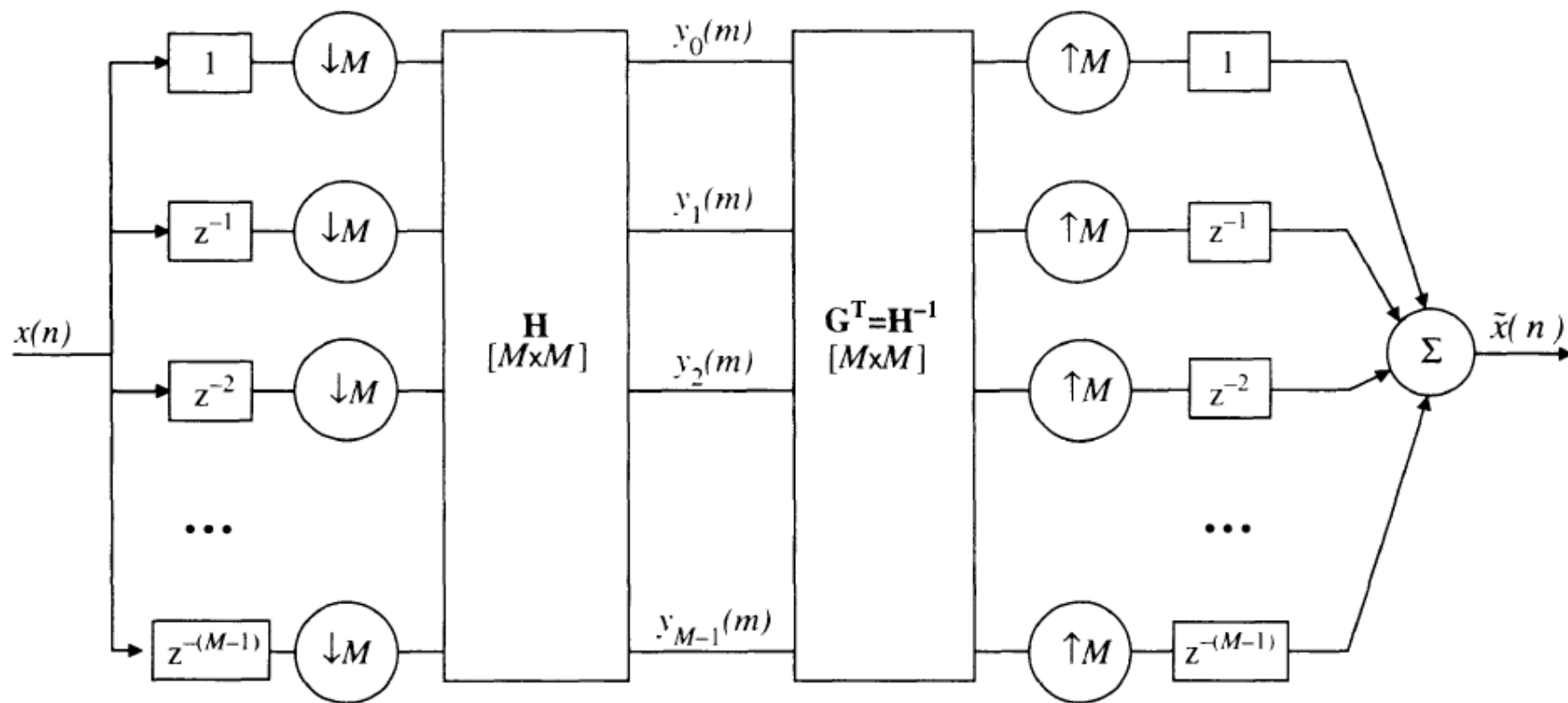
Transformada: codificadores com muitos canais de frequência como DOLBY AC2 e AC3 e MPEG AAC

Matematicamente, apesar da nomenclatura diferente, eles são equivalentes

Codificação por Sub-Bandas



Codificador por Transformada M-Canais, $N=M$



Notação Matricial Filtros de Análise e Síntese

$$\begin{bmatrix} y_0(m) \\ y_1(m) \\ \dots \\ y_{M-1}(m) \end{bmatrix} = \begin{bmatrix} h_0(M-1) & \dots & h_0(1) & h_0(0) \\ h_1(M-1) & \dots & h_1(1) & h_1(0) \\ \dots & \dots & \dots & \dots \\ h_{M-1}(M-1) & \dots & h_{M-1}(1) & h_{M-1}(0) \end{bmatrix} \begin{bmatrix} x(mM-M+1) \\ \dots \\ x(mM-1) \\ x(mM) \end{bmatrix}$$

$$\mathbf{y} = \mathbf{H}\mathbf{x}$$

$$\begin{bmatrix} x(mM-M+1) \\ \dots \\ x(mM-1) \\ x(mM) \end{bmatrix} = \begin{bmatrix} g_0(M-1) & g_1(M-1) & \dots & g_{M-1}(M-1) \\ \dots & \dots & \dots & \dots \\ g_0(1) & g_1(1) & \dots & \dots \\ g_0(0) & g_1(0) & \dots & g_{M-1}(0) \end{bmatrix} \begin{bmatrix} y_0(m) \\ y_1(m) \\ \dots \\ y_{M-1}(m) \end{bmatrix}$$

$$\mathbf{x} = \mathbf{G}^T \mathbf{y} \text{ com } \mathbf{G}^T = \mathbf{H}^{-1}$$

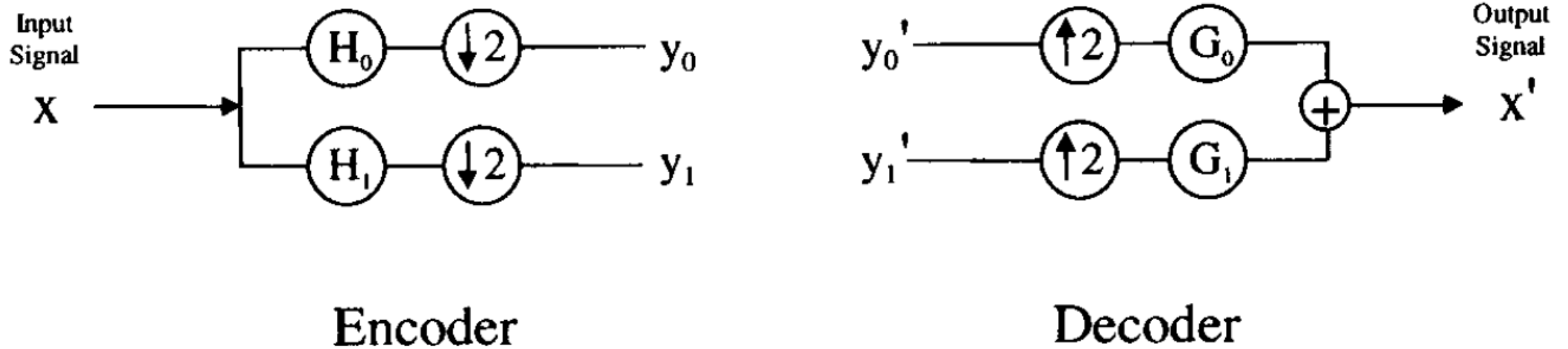
- Se o tamanho dos filtros no codificador de M canais for de M amostras, a implementação será mais eficiente ($N=M$).
- A matriz de transformação linear de codificação $\mathbf{H}_{M \times M}$ com
$$\mathbf{y} = \mathbf{H}\mathbf{x}$$
- A matriz de decodificação $\mathbf{G}_{M \times M}$, com $\mathbf{x} = \mathbf{G}^T \mathbf{y}$
- Se o tamanho N dos filtros de análise-síntese no codificador-decodificador for maior que M canais, a implementação necessitará de multiplicação das N amostras nos frames por matrizes $N \times M$ ou $M \times N$.
- O downsamplig deve ser efetuado em quadros sucessivos com overlap de $N-M$, para cada M amostras

Operação:

- O sinal de entrada $x(n)$ é decomposto M faixas de frequências:
 - ***banco de filtros de análise em sub-bandas***
- Os sinais de cada sub-banda $x_i(n)$ são decimados (redução da taxa, *downsampling*) para
 $y_i(m) = x_i(Mm)$ e depois quantizados.
- O decodificador combina os sinais das sub-bandas $\tilde{y}_i(m)$ em um sinal de saída $\tilde{x}(m)$ e
- depois interpola (*upsampling*) para $\tilde{x}_i(m)$:
$$\tilde{x}_i(mM) = \tilde{y}_i(m) \text{ e } \tilde{x}_i(mM + l) = 0 \text{ para } l = 0, \dots, M - 1$$
- Por fim, sintetiza por um Banco (Crochiere and Rabiner 1983, Vaidyanathan 1993)

Banco de Filtros de Dois Canais com Reconstrução Perfeita (RP)

- Iniciamos com um banco de dois filtros para chegar aos multi-filtros:



- Reconstrução $X'(z)$ em função dos sinais codificados, usando (1),(2) e (4):

$$X'(z) = Y_0(z^2)G_0(z) + Y_1(z^2)G_1(z) \quad (5)$$

RP: Banco de Filtros de Dois Canais

- Sinais codificados em função das entradas:

$$Y_0(z) = \frac{1}{2} [H_0(z^{1/2})X(z^{1/2}) + H_0(-z^{1/2})X(-z^{1/2})] \quad (6a)$$

$$Y_1(z) = \frac{1}{2} [H_1(z^{1/2})X(z^{1/2}) + H_1(-z^{1/2})X(-z^{1/2})] \quad (6b)$$

- Combinando (5) com (6a) e (6b) temos:

$$\begin{aligned} X'(z) = & \frac{1}{2} [H_0(z)G_0(z) + H_1(z)G_1(z)]X(z) + \\ & + \frac{1}{2} [H_0(-z)G_0(z) + H_1(-z)G_1(z)]X(-z) \end{aligned} \quad (7)$$

As componentes $X(-z)$ representam as componentes de sobreposição espectral das componentes em torno de $f - f_s/2$ nas componentes em torno de f :

- Para garantir a reconstrução perfeita precisamos cancelar estas componentes.

RP: Eliminando o Aliasing (sobreposição)

$$H_0(-z)G_0(z) + H_1(-z)G_1(z) = 0 \quad (8)$$

Que pode ser obtida impondo-se:

$$\begin{cases} G_0(z) = H_1(-z) \\ G_1(z) = -H_0(-z) \end{cases} \quad (9)$$

Assim o termo em $X(-z)$ se anula e:

$$X'(z) = \frac{1}{2}[-H_0(z)H_1(-z) + H_1(z)H_0(-z)]X(z) \quad (10)$$

As expressões no DT de (9) são:

$$\begin{cases} g_0[n] = (-1)^n h_1[n] \\ g_1[n] = -(-1)^n h_0[n] \end{cases} \quad (11)$$

Reconstrução Perfeita por *Quadrature Mirror Filters* (QMF)

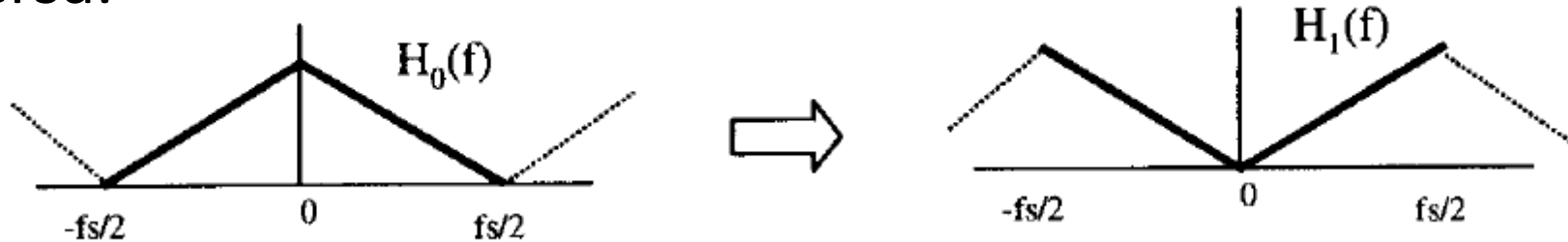
- *Aliasing* é eliminado pela escolha adequada dos filtros de síntese:

$$H_1(z) = H_0(-z) \stackrel{z}{\Leftrightarrow} h_1[n] = (-1)^n h_0[n] \quad (12)$$

- No DF teremos:

$$H_1(f) = H_0\left(f - \frac{f_s}{2}\right) \quad (13)$$

Se $H_0(f)$ for um filtro Passa-Baixas, $H_1(f)$ será um filtro Passa-Altas e vice-versa.



Filtros QMF para Reconstrução Perfeita

Definindo todos os filtros em função de $H_0(z)$:

$$\begin{cases} H_1(z) = H_0(-z) \stackrel{z}{\Leftrightarrow} h_1[n] = (-1)^n h_0[n] \\ G_0(z) = H_0(z) \stackrel{z}{\Leftrightarrow} g_0[n] = h_0[n] \\ G_1(z) = -H_0(-z) \stackrel{z}{\Leftrightarrow} g_1[n] = -(-1)^n h_0[n] \end{cases} \quad (14)$$

A saída será:

$$X'(z) = \frac{1}{2}[H_0(z)^2 - H_0(-z)^2]X(z) \quad (15)$$

O sinal de saída terá **RECONSTRUÇÃO PERFEITA** da entrada com D atrasos se:

$$H_0(z)^2 - H_0(-z)^2 = 2z^{-D} \quad (16)$$

Exemplo QMF: Filtro de Haar

O filtro de Haar $h_0[n]$ tem uma resposta ao impulso:

$$h_0[n] = \left\{ \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, \dots \right\} \xleftrightarrow{Z} H_0(z) = \frac{1}{\sqrt{2}}(1+z^{-1}) \quad (17)$$

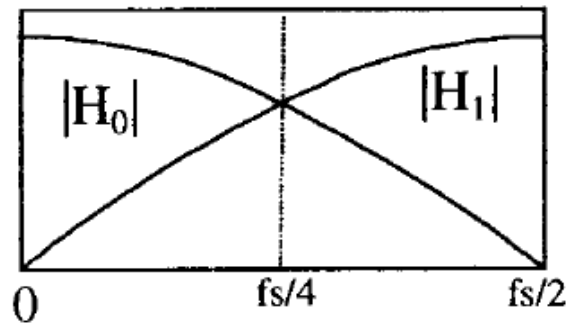
Que satisfaz a condição de Reconstrução Perfeita (16) se:

$$H_0(z)^2 - H_0(-z)^2 = 2z^{-1} \quad (18)$$

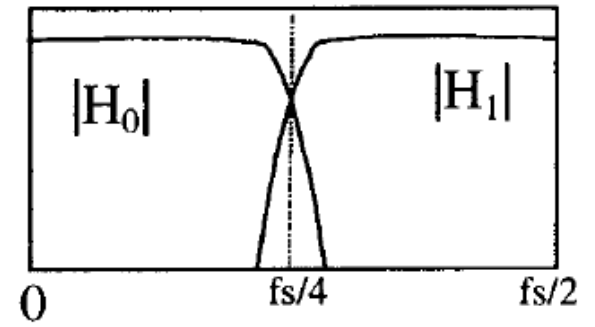
Obtendo todos os filtros:

$$\begin{cases} H_1(z) = \frac{1}{\sqrt{2}}(1-z^{-1}) \\ G_0(z) = \frac{1}{\sqrt{2}}(1+z^{-1}) \\ G_1(z) = -\frac{1}{\sqrt{2}}(1-z^{-1}) \end{cases} \quad (19)$$

Resposta dos
QMF Haar x
QMF “ideal”



Haar Filters

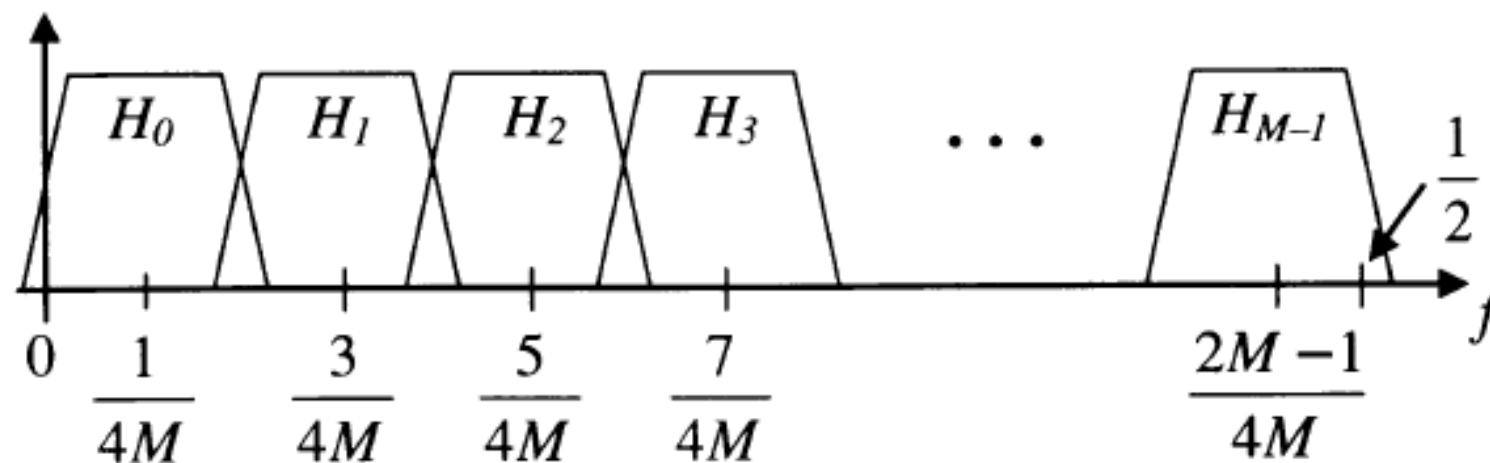


QMF

- Embora simples, a resposta impulsiva curta dos filtros Haar torna a seletividade deles muito suave
- Para transições mais seletivas outras alternativas de projetos FIR QMF “ideais” são necessárias ou outras abordagens (cfr. Vaidynathan 1993).
- Uma proposição são os FIR CQF (Conjugate Quadrature Filter)

Filtros Pseudo-QMF

(ou M Filtros Passa Faixas em Paralelo)



Filtros $H_i(z)$ são obtidos pela modulação da resposta impulsiva $h(n)$ de um protótipo PB por M portadoras f_i distribuídas uniformemente em $[0, 1/2]$:

$$h_i(n) = h(n)\cos(2\pi f_i n + \varphi_i)$$

$$g_i(n) = g(n)\cos(2\pi f_i n + \theta_i)$$

$$f_i = \frac{(2i + 1)}{2M} \frac{1}{2}$$

Condição na fase, para cancelar *aliasing*:

$$\varphi_i = -\frac{2i + 1}{4}\pi = \theta_i$$

A banda normalizada dos filtros deve ser $1/4M$ ao invés de $1/4$.

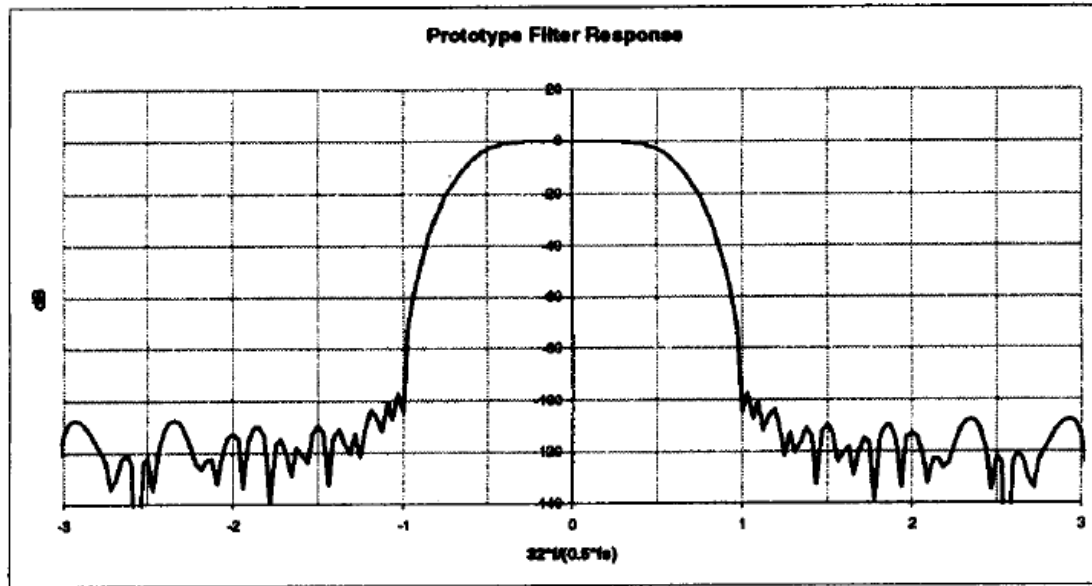
Os filtros de síntese para serem as imagens espelhadas devem ser:

$$h_i(n) = g_i(N - 1 - n),$$

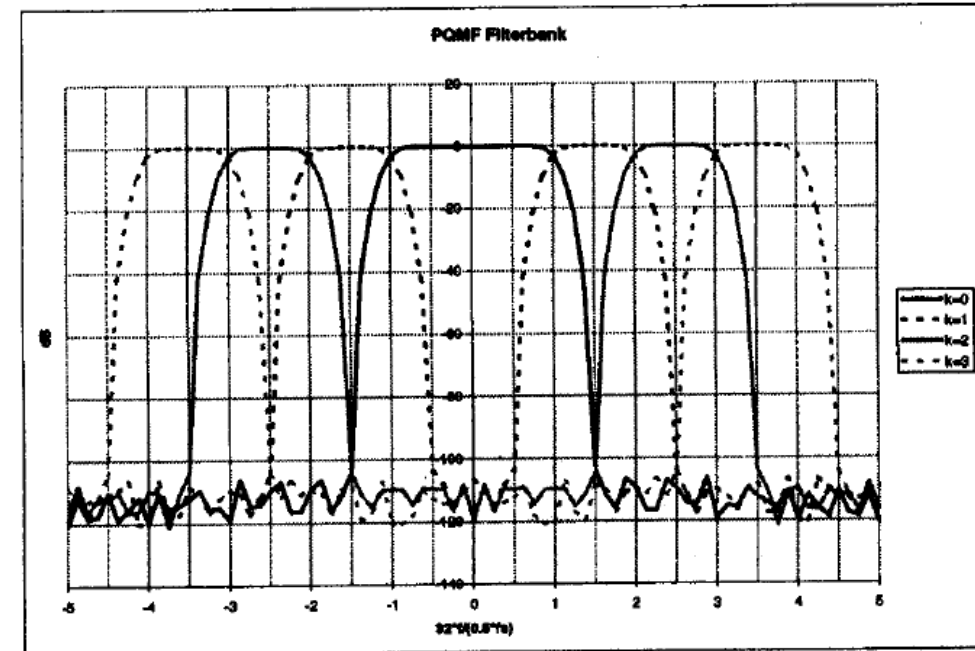
Ou seja, $H_i(z)G_i(z)$ em todas as sub-bandas deverão ter fase linear.

➤ Estes filtros PQMF são usados na norma MPEG-1, camada I de áudio.

Protótipo PQMF $f_s/64$



MPEG Audio PQMF prototype filter frequency response in units of $F_s/64$.



Frequency response of the first four bands of the MPEG Audio PQMF in units of $F_s/64$

Alguns filtros PQMF do MPEG1, camada 2

- A solução QMF não é a única possível solução para as equações de RP (5 a 11) para dois canais.
- Uma outra solução bem adaptada ao uso de filtros FIR é chamada de **Conjugate Quadrature Filter** (CQF):
 - Os filtros de síntese $g_0[n]$ e $g_1[n]$ são as respostas invertidas dos filtros de análise $h_0[n]$ e $h_1[n]$:

$$g_0[n] = h_0[N - 1 - n]$$

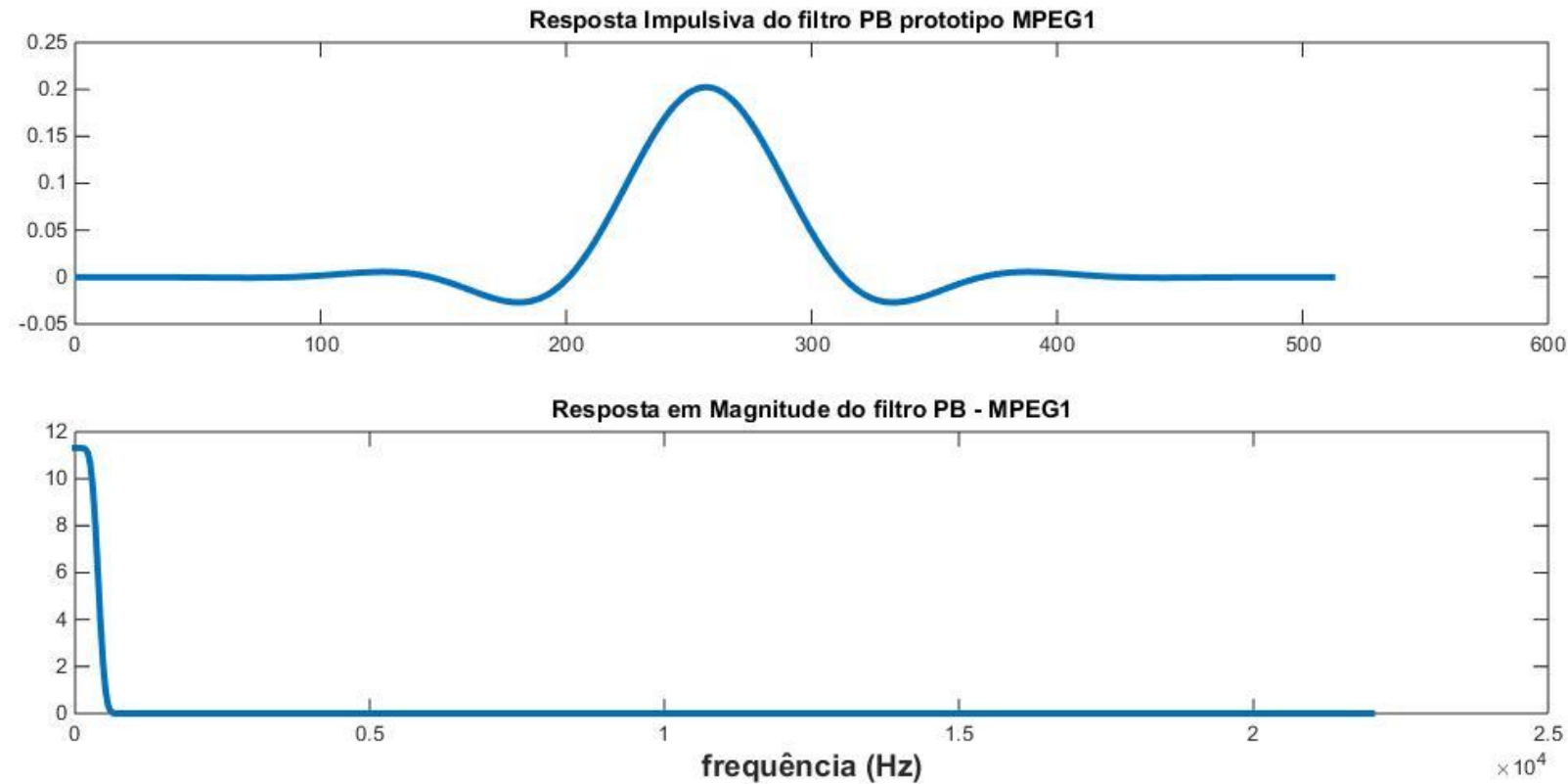
$$g_1[n] = h_1[N - 1 - n]$$
 - Assim como no QMF, o filtro de análise $h_1[n]$ é uma versão do filtro PB $h_0[n]$ modulado por $(-1)^n$, tornando-o PA.

A
Reconstrução
Perfeita (RP)
CQF

Banco de Filtros Pseudo QMF (PQMF)

- A Reconstrução Perfeita (RP) pode ser obtida por um sistema com dois filtros (Banco de Filtros de dois canais).
- Contudo, em aplicações práticas, necessitaremos de mais do que dois canais, devido à característica de análise efetuada pelo sistema auditivo humano (entre 20 a 30 canais).
- Alguns trabalhos iniciais sugeriram bancos de filtros com vários estágios QMF em cascata, para obter a análise em multicanais, assim como vimos no processamento de imagem com as Wavelets.
 - Estas estruturas não funcionaram adequadamente, devido ao aparecimento de respostas impulsivas muito longas e alta complexidade computacional.
- Uma abordagem mais eficiente é a generalização paralela multibanda da solução CQF, que foi denominada de Pseudo-QMF (PQMF).
- A ideia básica é iniciar com um filtro PB de faixa estreita, e depois modulá-lo, espalhando suas cópias no domínio da frequência.

Resposta Impulsiva e de Magnitude do FB



- O Banco de filtros PQMF consiste de K canais, cada um deles um filtro PB $h[n]$ modulado por um cosseno.
- Os filtros de análise e de síntese são:

$$h_k[n] = h[n] \cos \left\{ \pi \left(\frac{k + \frac{1}{2}}{K} \right) \left[n - \frac{(N-1)}{2} \right] + \phi_k \right\},$$

para $k=0, \dots, K-1$

$$g_k[n] = h_k[N - 1 - n]$$

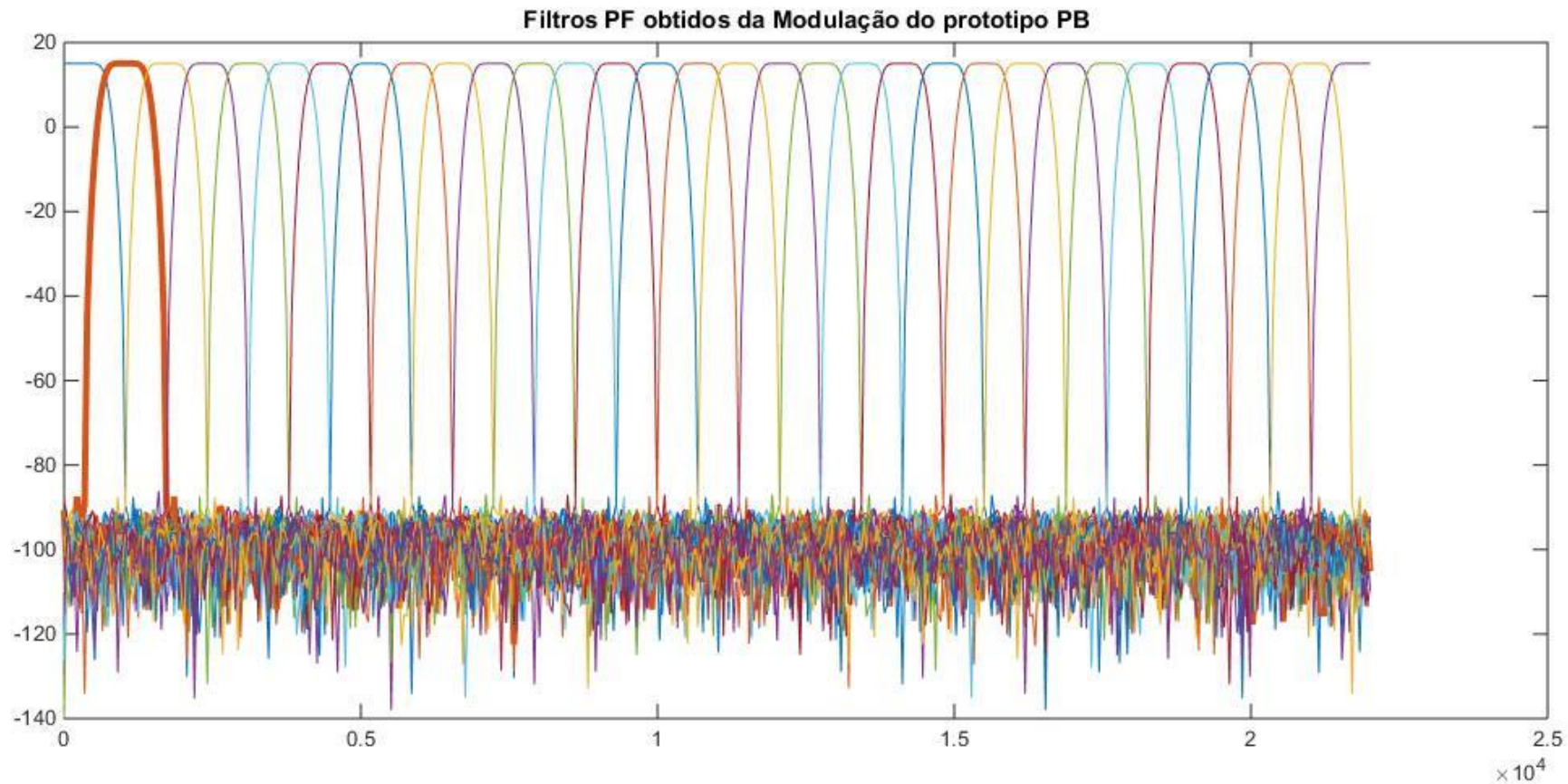
Onde N é o comprimento de $h[n]$. A fase ϕ_k é determinada pelas condições de anti-aliasing entre as bandas adjacentes:

$$\phi_k - \phi_{k-1} = \frac{\pi}{2} (2r + 1)$$

com r inteiro, sendo que o filtro de síntese, assim como na solução CQF, é a inversão temporal do filtro de análise.

Estrutura Básica PQMF

32 Filtros PF, pela Modulação do PB – 1ª modulação em destaque

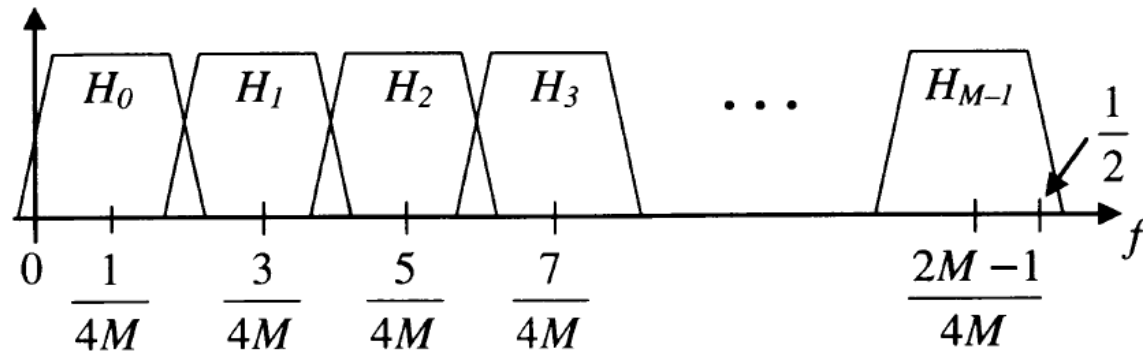


Filtros Análise e Síntese PQMF – MPEG1

$$h_i(n) = h(n) \cos(2\pi f_i n + \phi_i)$$

$$g_i(n) = h(n) \cos(2\pi f_i n + \theta_i) \quad \text{with } f_i = \frac{(2i+1)1}{2M} 2$$

M = 32 canais



$$\phi_i = -\frac{2i+1}{4}\pi = -\theta_i$$

$$h_i(n) = g_i(N-1-n)$$



Estas repostas impulsivas garantem **fase linear** em todas as bandas

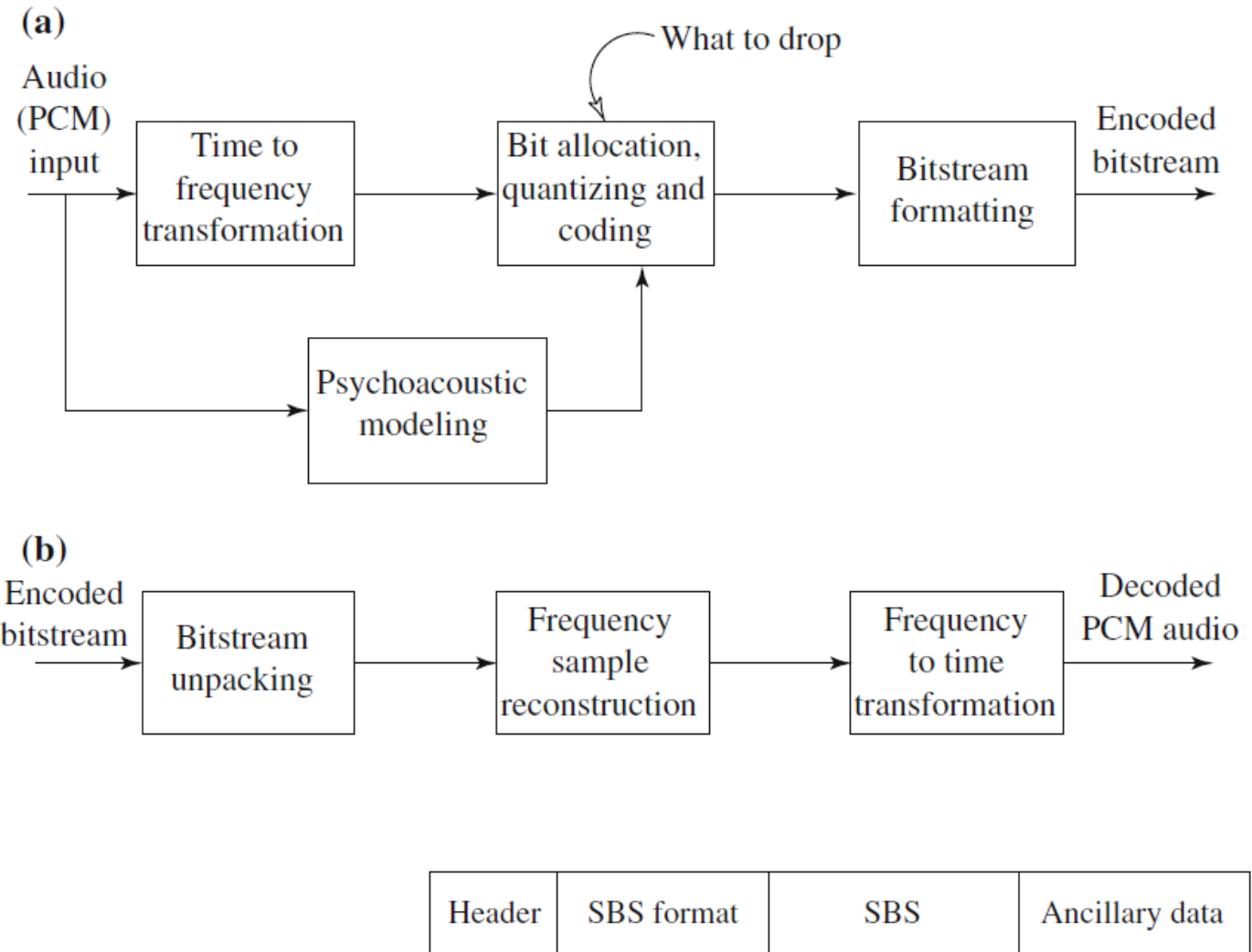
As fases são escolhidas para minimizar aliasing: **RP aproximada!**

O Mapeamento Tempo -> Frequência: MDCT

- A solução PQMF por Bancos de Filtros com RP aproximada é muito importante para qualidade, mas com pouca compressão.
- A codificação por transformadas é uma abordagem sugerida para efetuar o processamento em outro espaço de representação e assim efetuar a codificação desejada.
- Com a DFT e depois pela FFT, procura-se efetuar o mapeamento para o domínio da frequência, via processamento em blocos de dados.
- Os efeitos de borda entre os blocos dependem de como eles são obtidos, e portanto o estudo do janelamento é essencial.
- A Transformada Cossenoidal Discreta Modificada (MDCT), com janelamento variável, é usada no codificador MPEG AAC.

MPEG:

- (a) Coder
- (b) Decoder
- (c) bit Frame



Algoritmo MPEG básico

Divisão da entrada em 32 subbandas de frequência, via um banco de filtros.

É uma operação linear cujas entradas são conjuntos de 32 amostras PCM resultando na saída 32 coeficientes de frequência.

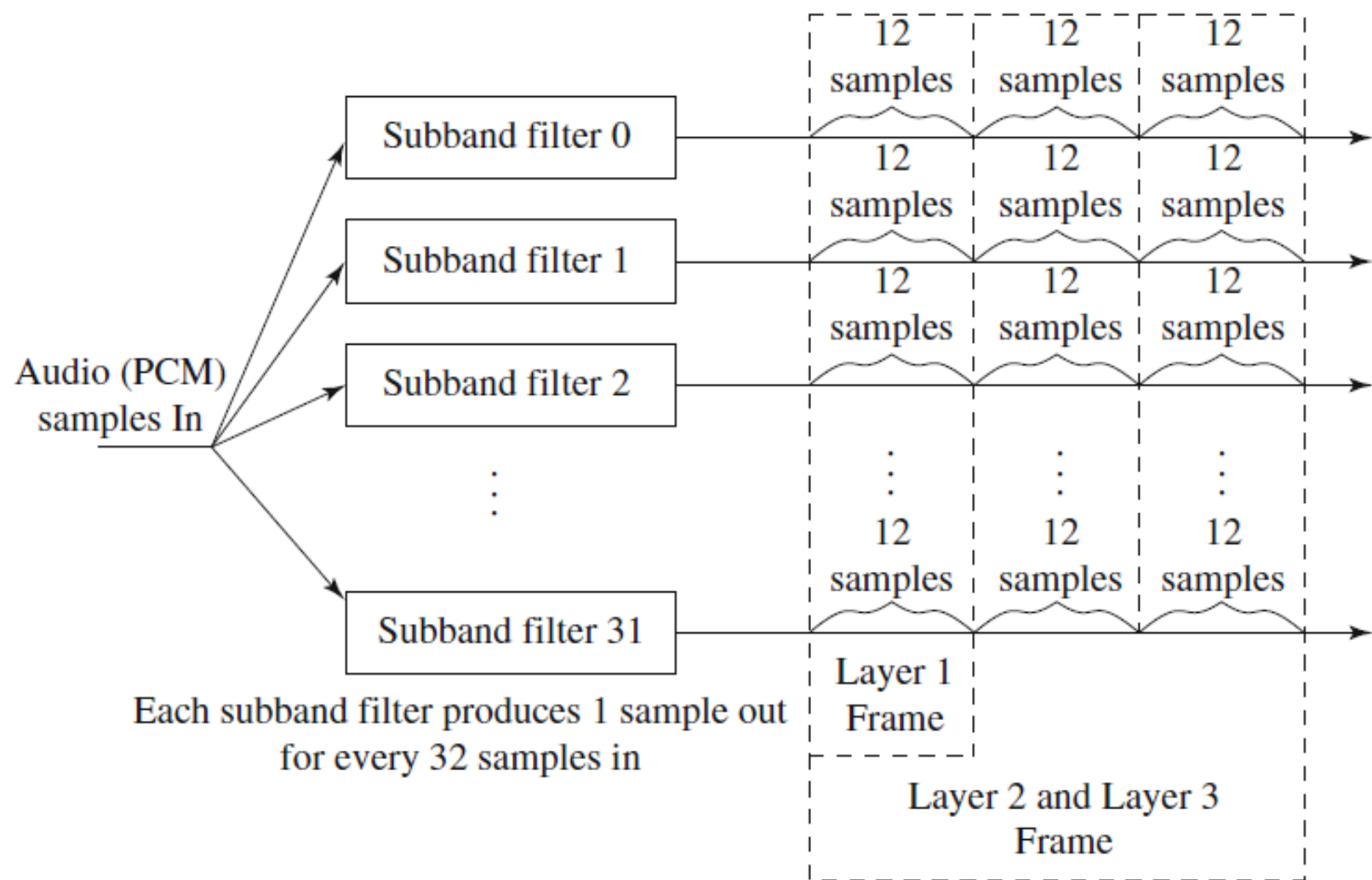
- Se a frequência de amostragem for $f_s = 48$ kamostra/s (i.e., 48 kHz), então pelo teorema de Nyquist, a máxima frequência mapeada será de $f_s/2$.

A largura de banda mapeada será dividida em 32 segmentos iguais, cada um com largura $f_s/64$ (estes segmentos serão sobrepostos - *overlapped*).

MPEG Layer 1

- Neste codificador os conjuntos de 32 amostras PCM são agrupados em 12 grupos de 32.
- O Codificador tem um time lag, igual ao tempo acumulado de 384 (i.e., 12×32) amostras.
 - P.ex., se a amostragem for de 32 kbits/s, então teremos um Δt de 12 ms para cada 32 amostras transmitidos por cada milissegundo. Estes conjuntos de 12 amostras, de tamanho 32, são denominados de *segmentos*.
- O atraso (*delay*) é na verdade maior que das 384 amostras acumuladas, devido ao cabeçalho.
- *Ancillary data*, tal como *multilingual data* e *surround-sound data*, é permitido.
- Camadas superiores também permitem mais que 384 amostras, e assim o formato das amostras de subbandas (SBS – Subband Samples) é adicionado também, como no *frame data* (slide 18).
- O cabeçalho (*header*) contém:
 - Código de sincronismo (doze 1s—111111111111),
 - A taxa de amostragem usada, bitrate,
 - Informação Estéreo.

MPEG: Frames dos Bancos de Filtros



MPEG Compressão:

- Nos canais estéreo pode-se ter um modo especial *joint-stereo* que produz um bit stream simples usando a redundância entre os dois canais. Esta é a versão de áudio de um sinal composto de vídeo. Pode-se também ter um *duplo-monofônico*—dois canais codificados independentemente. Isto é útil para tratamento paralelo de áudio — p.ex, dois streams de voz, um em Inglês e outro em Português.
- Considere o segment 32×12 como uma matriz 32×12 . O próximo estágio algoritmo é uma quantização apropriada:
 - Para cada uma das 32 subbandas, a amplitude máxima das 12 amostras na linha do vetor é determinada, o *fator de escalamento* desta subbanda.
 - Este máximo é passado para o algoritmo de alocação de bit no bloco, nesta SBS.
 - Determina-se qual a porção de bits ficará disponível na subbanda para minimizar a audição do ruído de quantização.
- O Modelo psicoacústico é razoavelmente complexo — mais que uma consulta a tabelas (e o modelo não é padronizado na especificação).
- Na Camada 1, um passo de decisão é incluído para qual banda de frequência é igual ao tom ou ao ruído. Desta decisão e do fator de escalamento, um limiar (threshold) de mascaramento é calculado para cada banda e comparada com um limiar de audição.

MPEG1: Psicoacústica

Agrupamento dos valores espectrais em bandas críticas

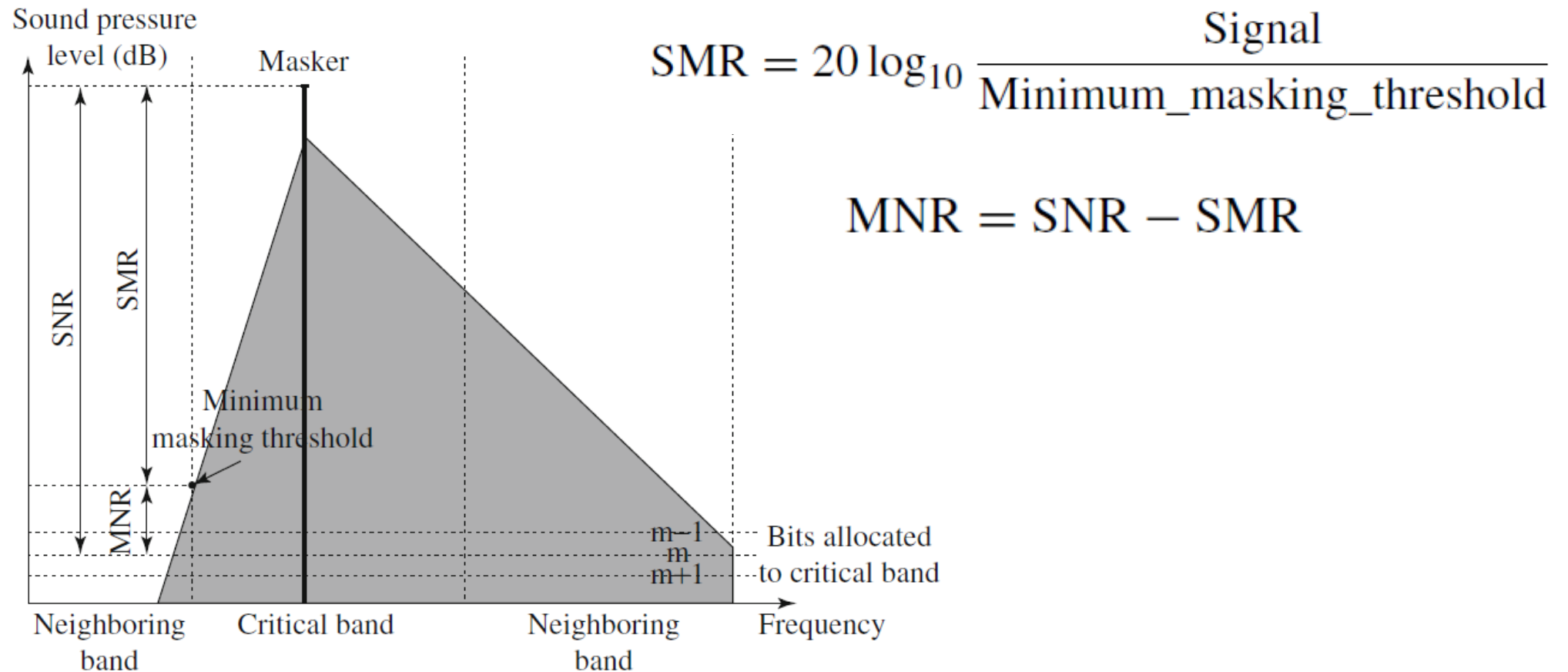
Separação dos valores espectrais em componentes tonais e não tonais com base dos picos de espectro de potência

Determinação de um limiar de mascaramento, por uma função de mascaramento empírica

Aplicação do limiar de mascaramento, para cada sub-banda

Cálculo do signal-to-mask ratio (SMR) e envia este valor para alocação de bits no codificador

Signal-to-Noise ratio (SNR), Mask-to-noise ratio (MNR), Signal-to-mask ratio (SMR)



MPEG1, camadas 1, 2 e 3:

Camada (Layer) 1: menor complexidade, maior taxa (32-224 kbit/s por canal)

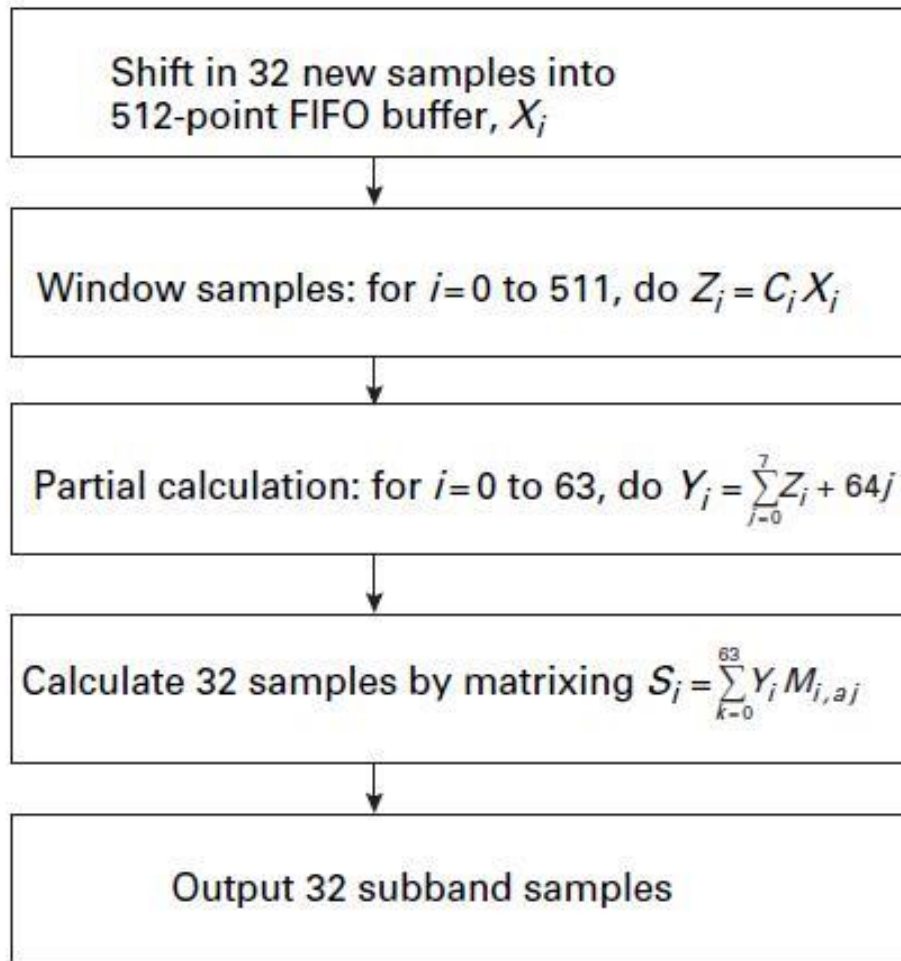
- taxa final de 192kbit/s/canal.
- Sensibilidade auditiva e mascaramento de freq.
- frame 384 amostras, 32 sub-bandas, 12 amostras

Camada (Layer) 2: complexidade média

- 128 kbit/s
- Mascaramento temporal e em frequência
- Frame 1152 amostras, 32 sub-bandas, 36 amostras

Camada (Layer 3): maior complexidade e “boa” qualidade

- Taxas de até 64kbit/s
- Instituto Fraunhofer e a Universidade de Erlangen
- Mascaramento temporal e em frequência e codificação por entropia (tamanho variável) Huffman
- 1152 amostras em 32 sub-bandas, 36 amostras cada

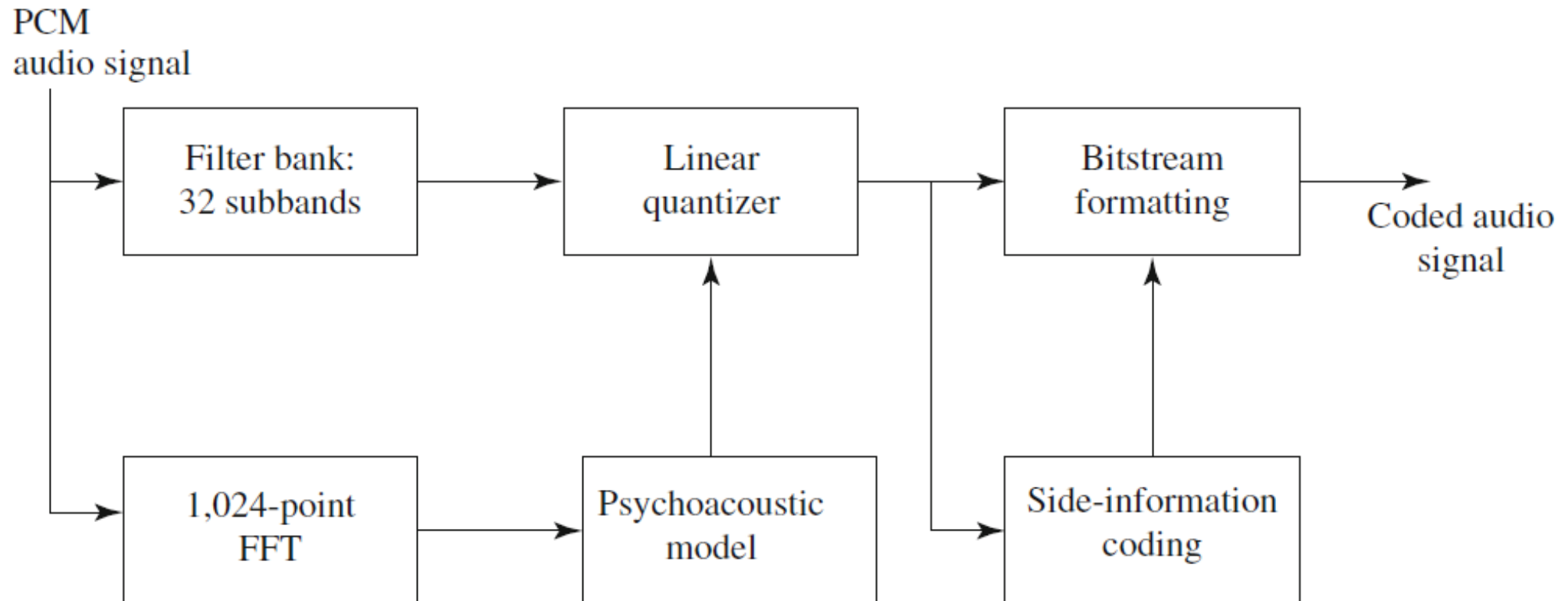


$$M(i, k) = \cos \left[\frac{(2i + 1)(k - 16)\pi}{64} \right]$$

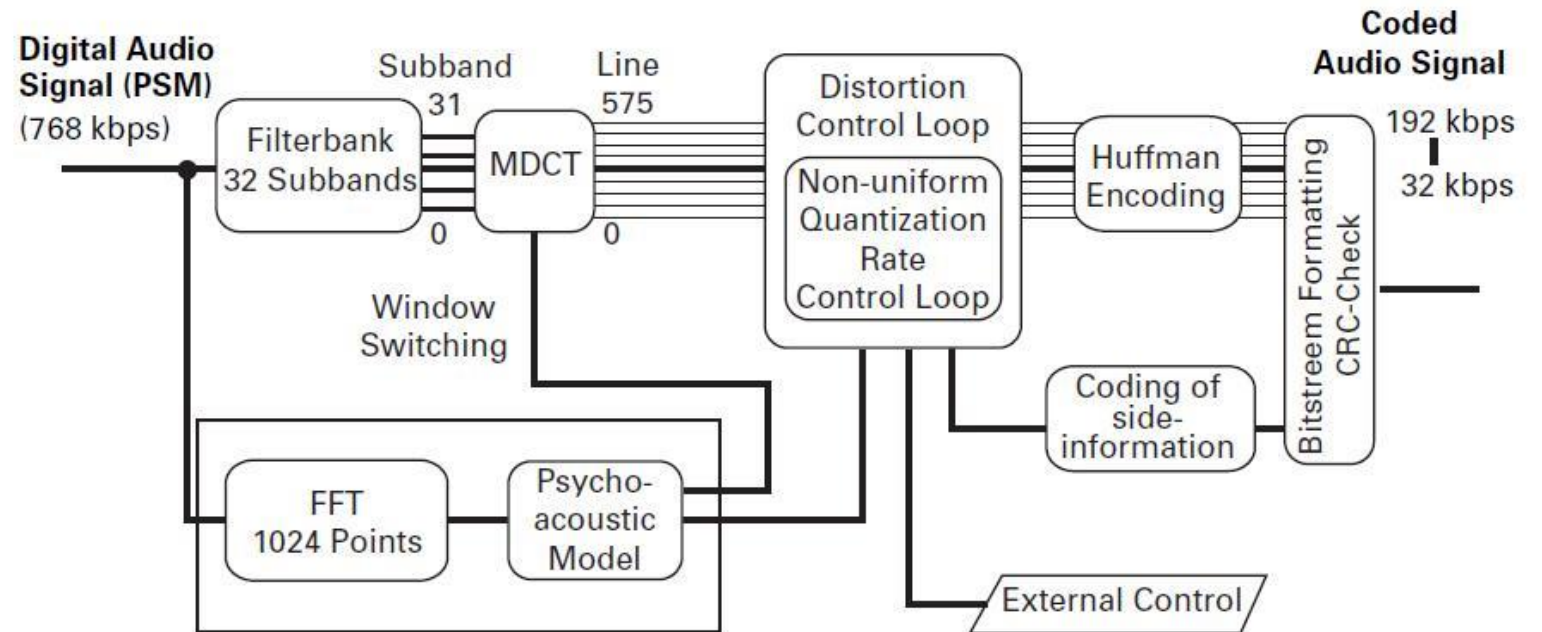
Banco de Filtros Polifásicos MPEG1

$$S_t(i) = \sum_{k=0}^{63} \sum_{j=0}^7 M(i, k) [C(k + 64j)x(k + 64j)]$$

MPEG 1 camadas 1 e 2



MPEG1, camada 3

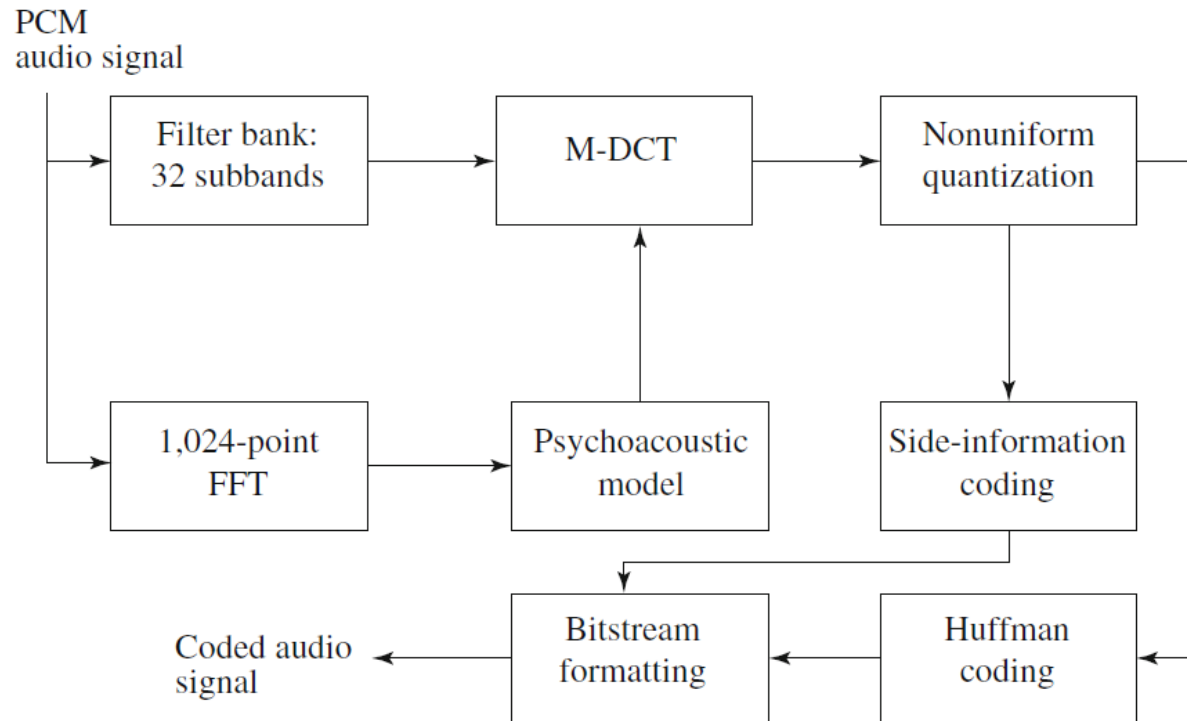


$$X(k) = \sum_{n=0}^{2M-1} x(n) \cos \left[\frac{\pi}{M} \left(n + \frac{1}{2} + \frac{M}{2} \right) \left(k + \frac{1}{2} \right) \right], k = 0, 1, 2, \dots, M-1$$

Características Gerais

- Codec da TV digital do SBTVD
- Aceita sinais MPEG-1
- MPEG-2 no modo compatível (BC - Backward Compatible) usa mesma tecnologia que o MPEG-1 (layers 1, 2 e 3)
- Multicanais, dito 5.1: 3/2+1, 3 frontais, 2 surround +1 sub-woofer LFE(opcional, até 120Hz)
- Previsão de “transparência” em 64kbit/s/canal
- Amostragens de 32, 44.1, 48 e 96kHz
- Padrões ISO/IEC 13818-3 e -7 (1994 e 1997)

MPEG 1 Camada 3 (MP3)



Sound quality	Bitrate	Mode	Compression ratio
Telephony	8 kbps	Mono	96:1
Better than shortwave	16 kbps	Mono	48:1
Better than AM radio	32 kbps	Mono	24:1
Similar to FM radio	56–64 kbps	Stereo	26:1–24:1
Near-CD	96 kbps	Stereo	16:1
CD	112–128 kbps	Stereo	14:1–12:1

MP3 – Tabela de Compressão

MDCT

$$F(u) = \sum_{i=0}^{N-1} f(i) \cos \left[\frac{2\pi}{N} \left(i + \frac{N/2 + 1}{2} \right) (u + 1/2) \right], u = 0, \dots, N/2 - 1$$

Procedimento de Sobreposição-e-Adição (*Overlap-And-Add*)



Para a reconstrução (síntese), no retorno do mapeamento



A transformação para o DF permite a separação do sinal (seu espectro) em blocos de bandas de frequência com redundância de informação.



Utiliza-se o efeito de mascaramento em frequência no Sistema Auditivo Humano para remoção de componentes “irrelevantes”

Descarte de componentes de frequência inaudíveis
Alocação de bits variável por banda, penalizando faixas espectrais encobertas pelo ruído de quantização

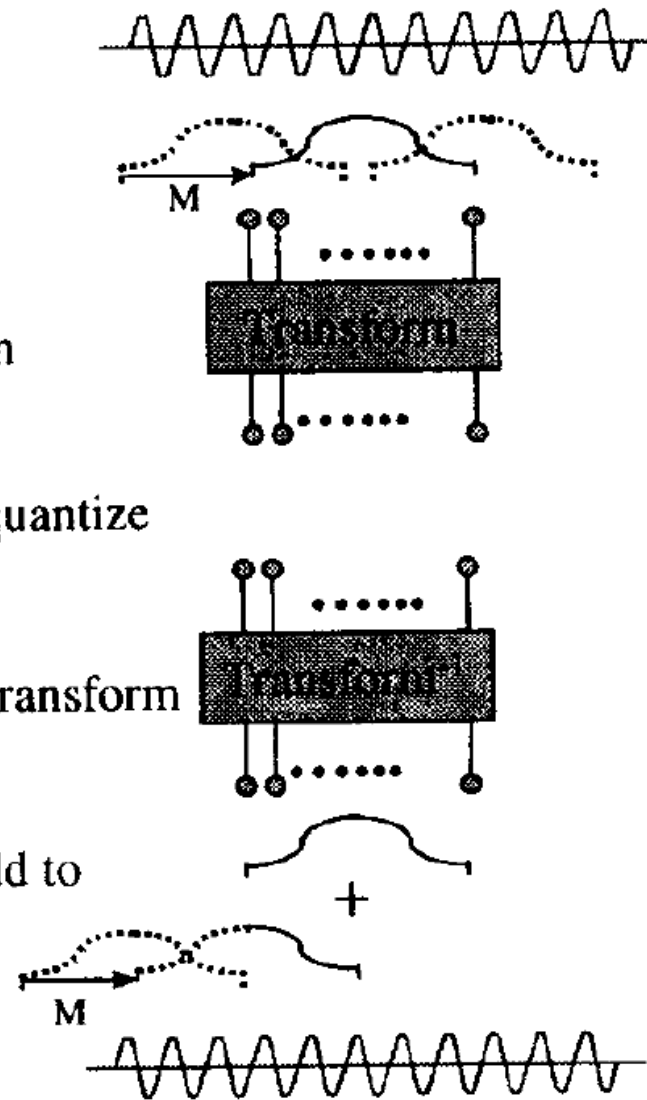
1) Slide M samples and Window
(window length = N)

2) Perform an N-point Transform

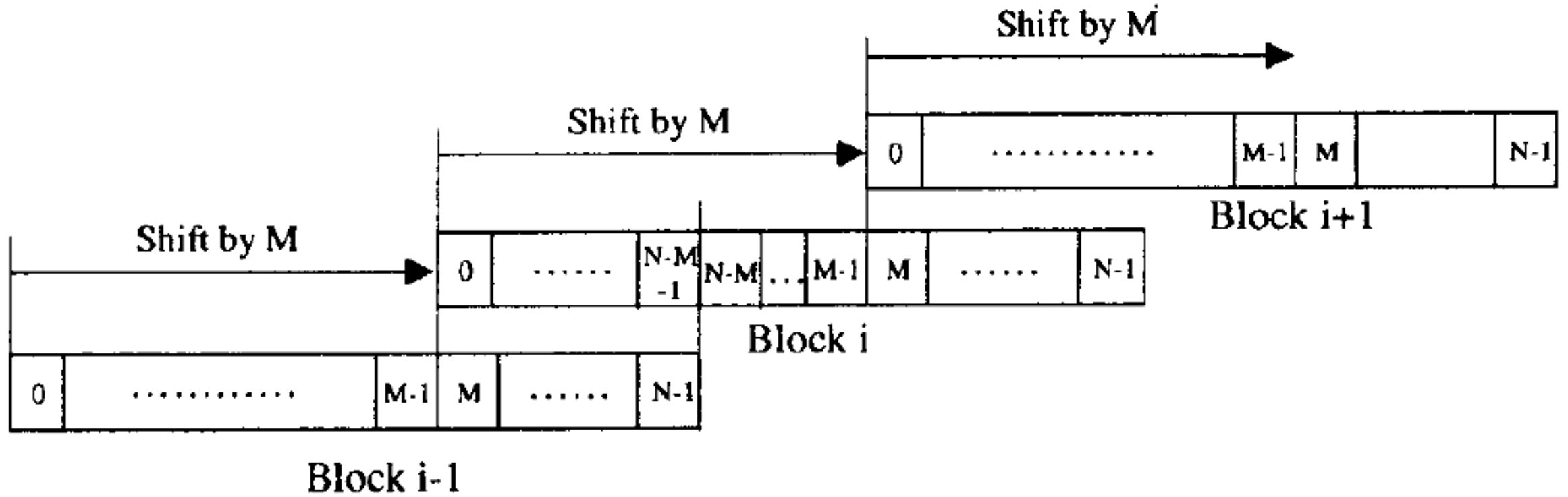
3) Quantize, Store/Transmit, Dequantize

4) Perform an N-point Inverse Transform

5) Window then Overlap and Add to
last N-M Outputs



Overlap-And-Add

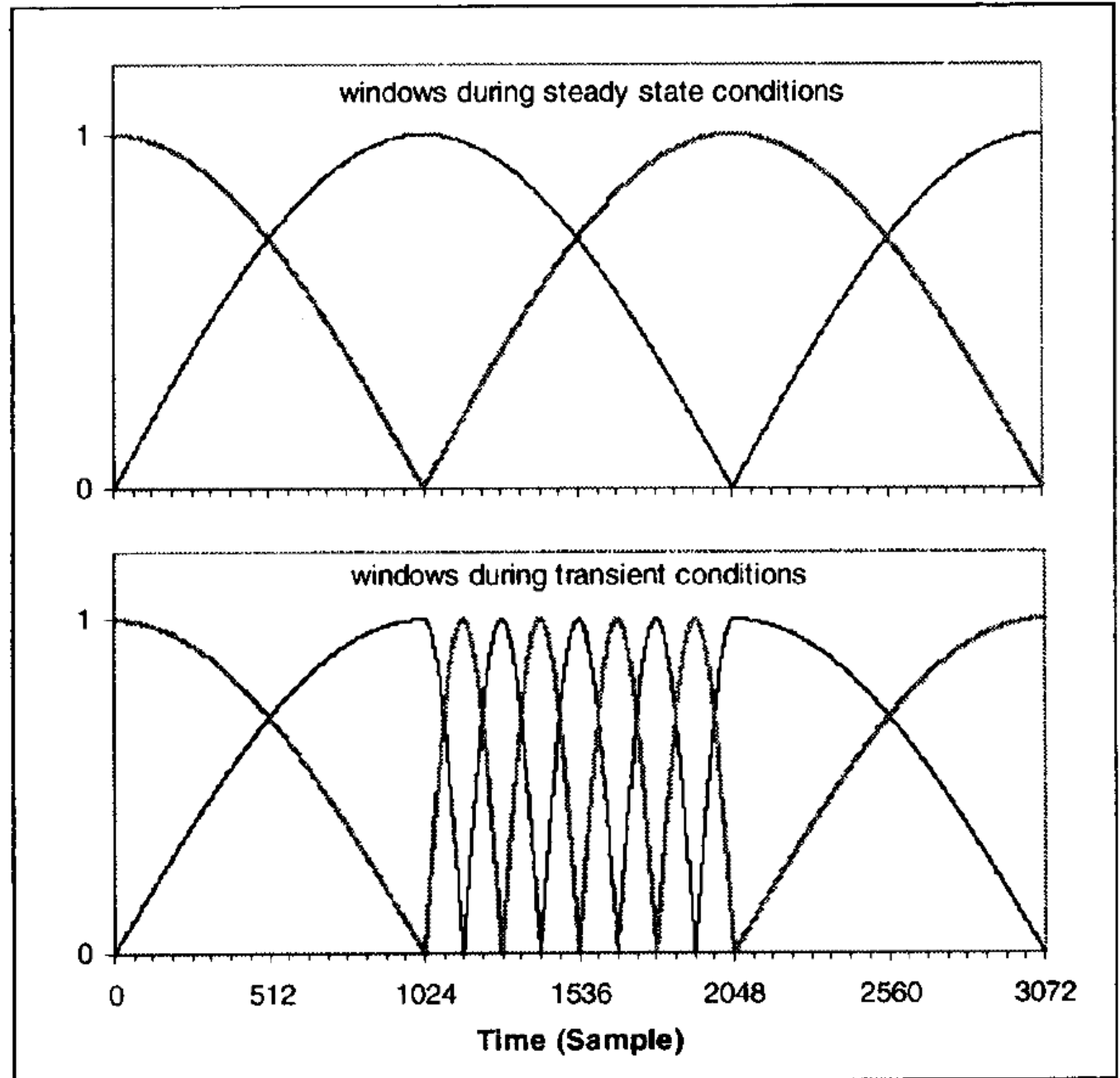


Overlap-And-Add

Block Switching

Ajustar a seletividade do Banco de Filtros:

- Mudar o tamanho da janela mantendo o critério de Reconstrução Perfeita
- Para efeitos transitórios, janelas estreitas
- Para sinais mais estacionários, janelas mais largas
- Não é necessário manter simetria nas janelas



FILTROS MDCT (usados no MP3 e no AAC)

- São Filtros PQMF generalizados para bancos de filtros modulados por cossenos com Reconstrução Perfeita, ou ***Modified Discrete Cosine Transform (MDCT) filter*** (Princen e Bradley 1986).
- Respostas impulsivas dos filtros de análise:

$$h_i(n) = h(n) \sqrt{\frac{2}{M}} \cos(2\pi f_i n + \varphi_i) \quad \text{com} \quad f_i = \frac{(2i + 1)}{2M} \frac{1}{2}$$

$$g_i(n) = h_i(n)(2M - 1 - n) \quad \varphi_i = \frac{(2i + 1)(M + 1)}{4M}$$

- Condições PR:

$$h^2(n) - h^2(n + M) = 1$$

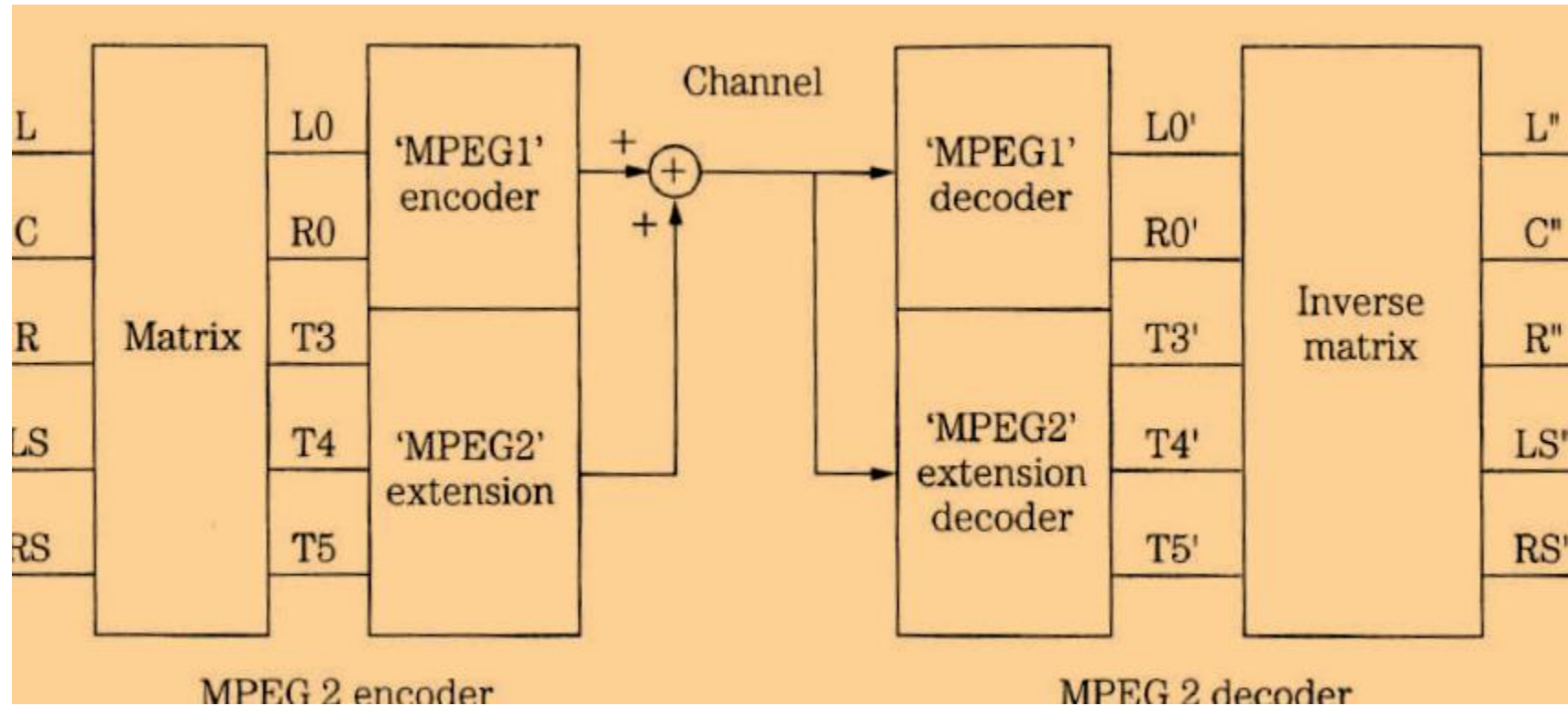
$$h(n) = h(2M - 1 - n) \quad \text{para } n = 0, \dots, M-1$$

$$h(n) = \sin \left[\left(n + \frac{1}{2} \right) \frac{\pi}{2M} \right], \text{ para } n=0, \dots, 2M-1$$

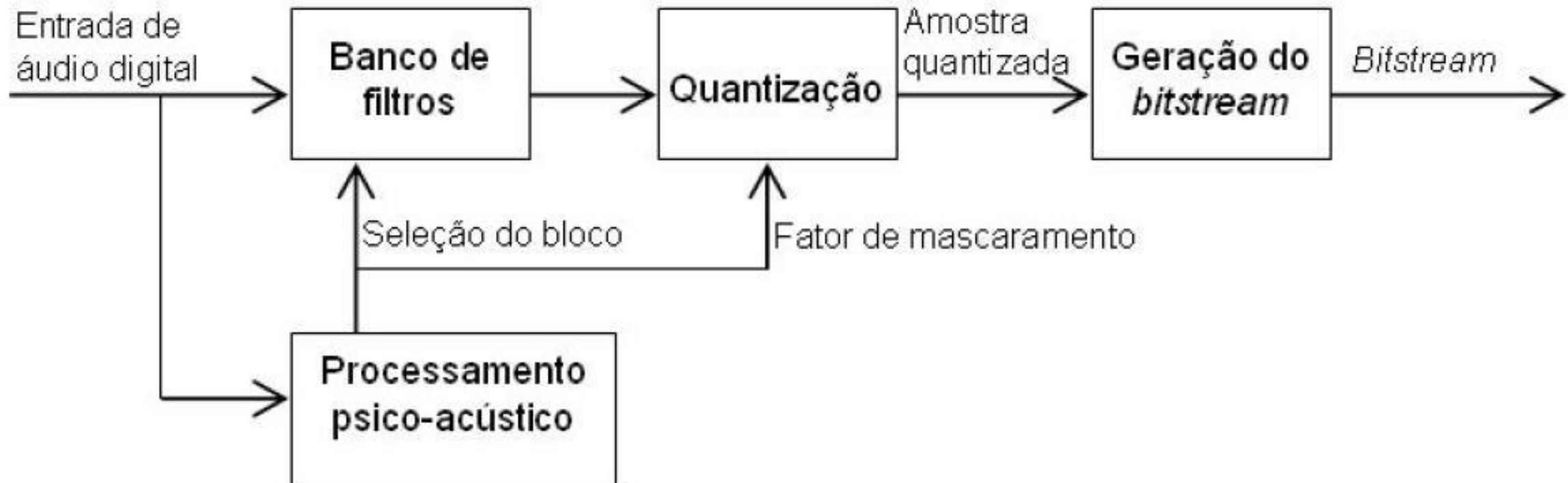
- Utilizados nas normas:
 - MPEG-1, camada 3 (MP3)
 - MPEG-2 AAC
 - MPEG4

Modulated
Lapped
Transform
Filter Bank
(Malvar 1990),
caso especial
da MDCT

Compatibilidade entre MPEG-1 e 2



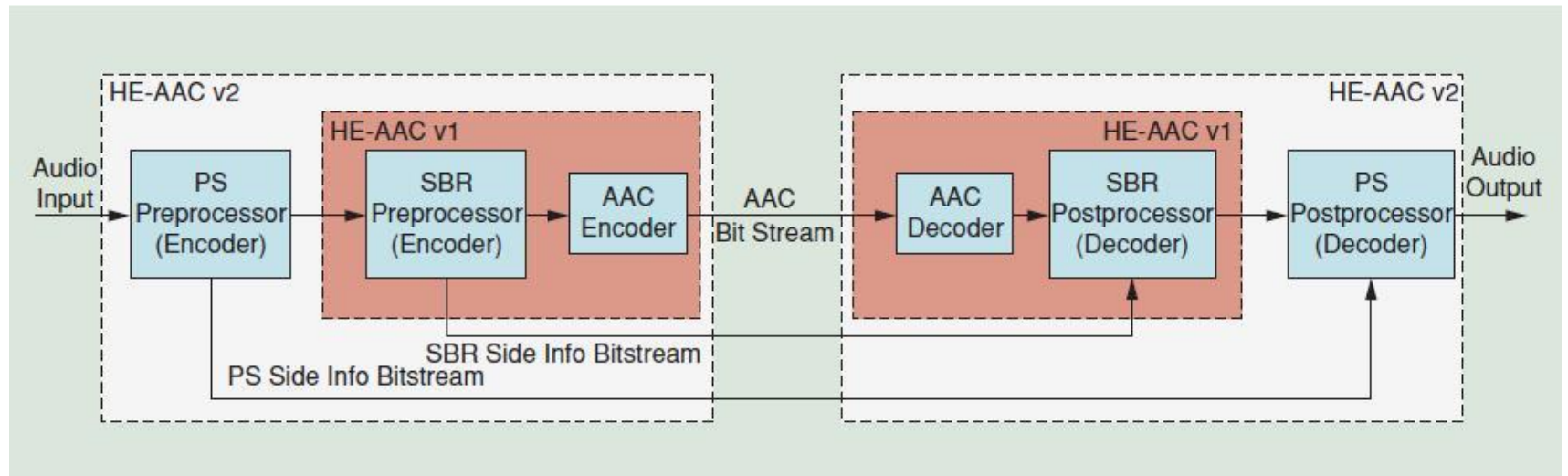
Nomenclatura da Norma do SBTVD
ABNT NBR 15602-2:2007
MPEG2 AAC



Padrões e Níveis do MPEG-2 AAC no SBTVD

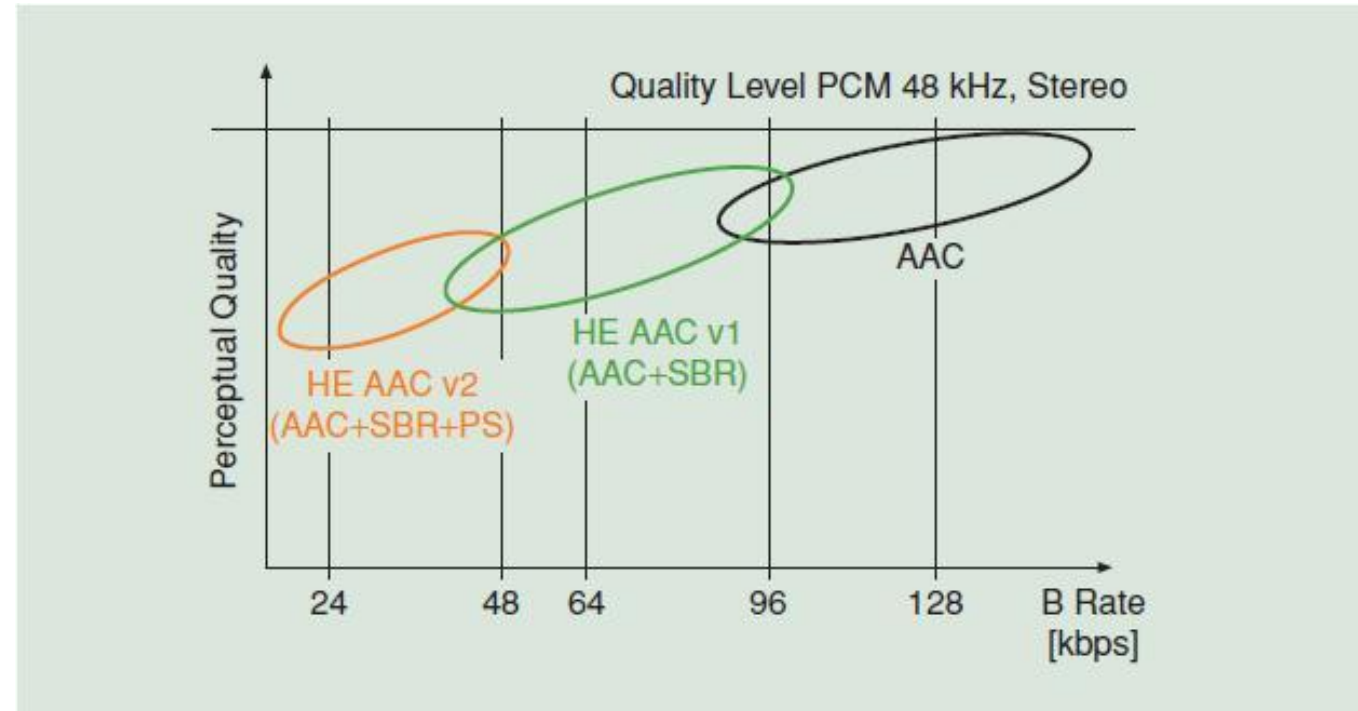
- **LC (Low Complexity)**, Perfil Básico do padrão AAC, níveis L2 e L4
- **HE (High Efficiency)**, Perfil Avançado de Alta Eficiência, combinando o perfil LC com o uso da ferramenta SBR (Spectral Band Replication), para versão 1 deste perfil, níveis L2 e L4
- **HE combinado à ferramenta PS (parametric stereo)** para a versão 2 deste perfil, nível L2

Estrutura do MPEG-2 HE- AAC



- AAC Encoder/Decoder (CORE)
- “Melhora” Espectral:
 - SBR – Spectral Band Replication Tool
 - PS – Parametric Stereo Tool

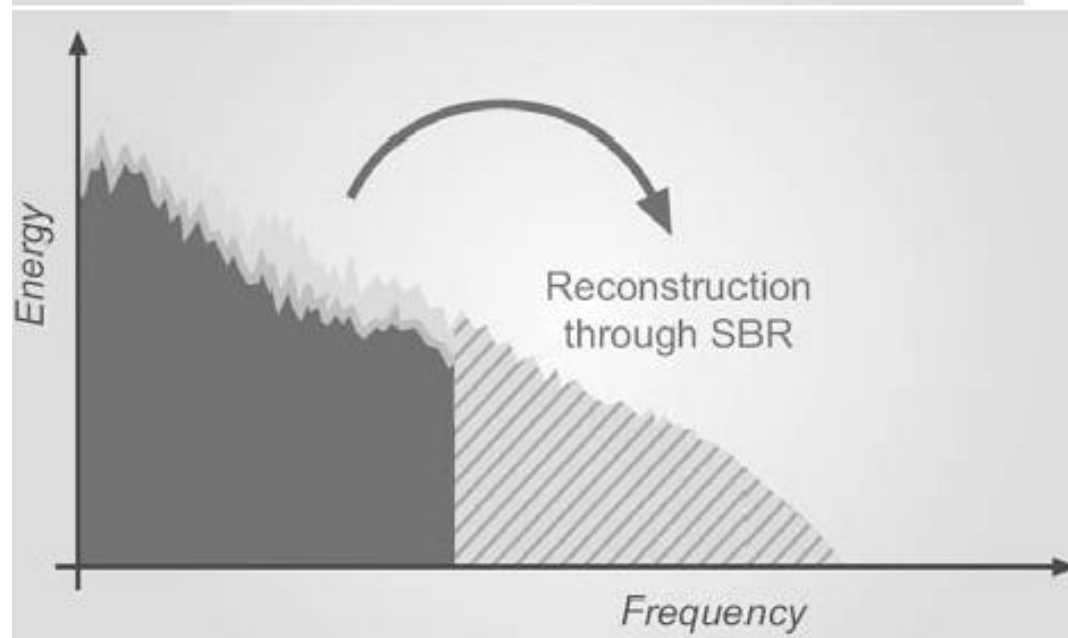
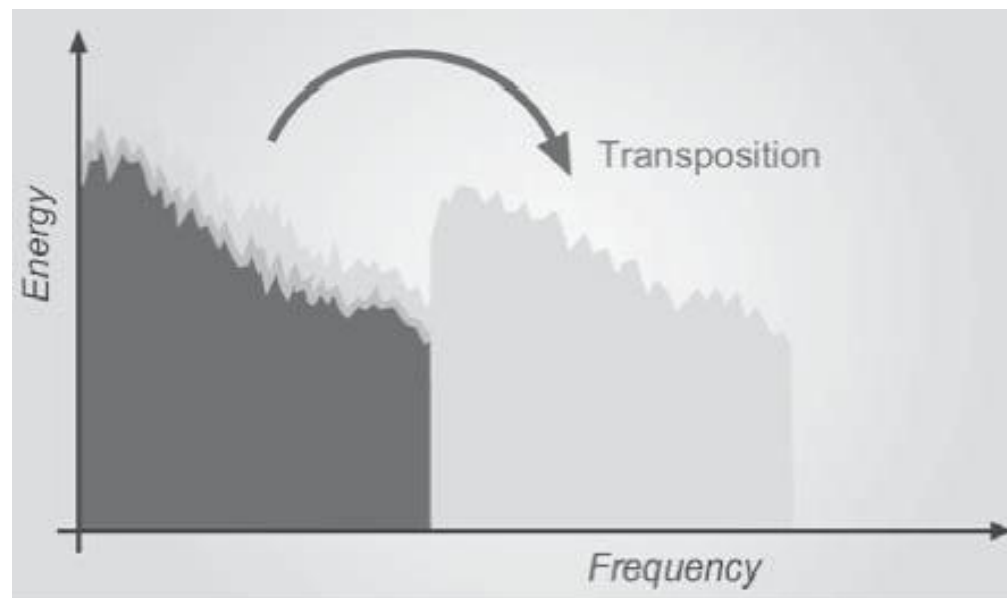
Desempenho HE-AAC



Spectral Band Replication (SBR):

- Aproveita menor sensibilidade do ouvido humano para altas frequências
- “Regenera” as altas frequências a partir das baixas, com QMF
 - High-Frequency Reconstruction
 - Envelope Adjustment

SBR: Transposição e Ajuste



Parametric Stereo

- Joint Stereo Coding: explora efeitos psicoacústicos binaurais
 - M/S (Middle/Side Channels) Stereo Coding.
 - Codifica os canais R e L conjuntamente, pois sinais quase monofônicos economizam bits.
 - Sinais até 2kHz possuem 20dB de níveis de mascaramento
 - Intensity Stereo Coding
 - Dynamic Crosstalk ou Channel coupling
 - É possível para certos tipos de sinais multifônicos a transmissão de um conjunto de valores espectrais compartilhados, sem perda significativa de qualidade

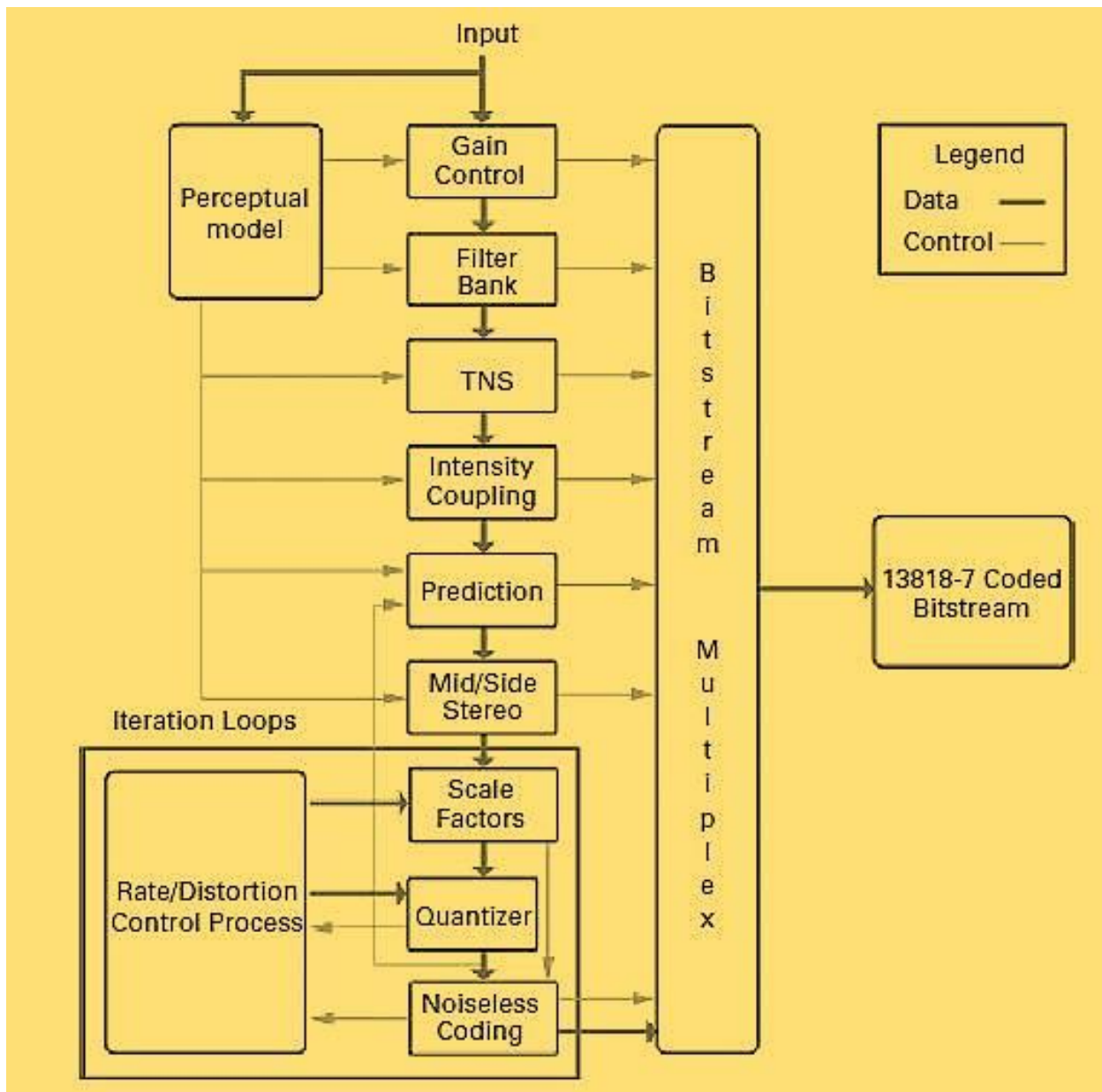
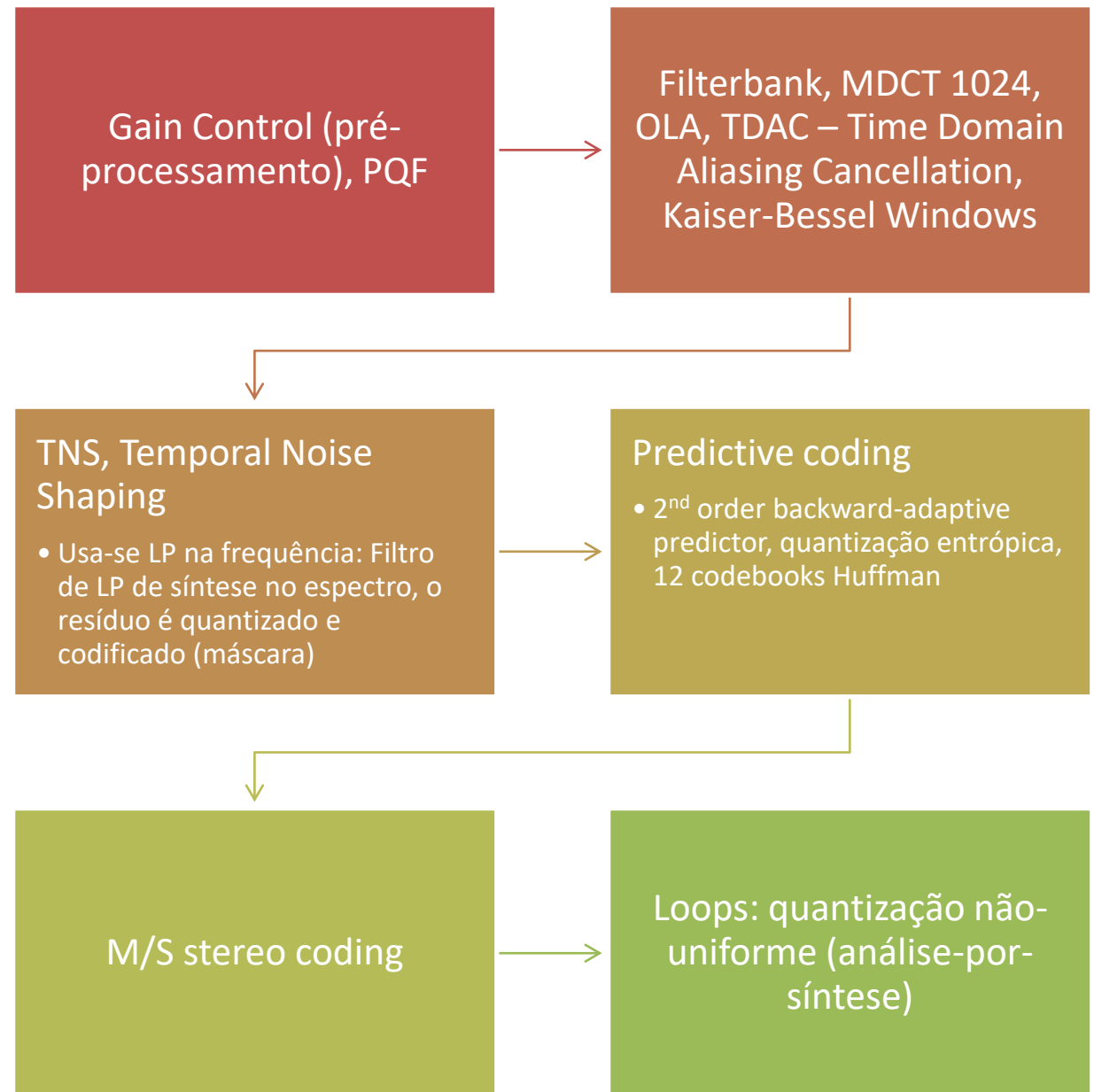
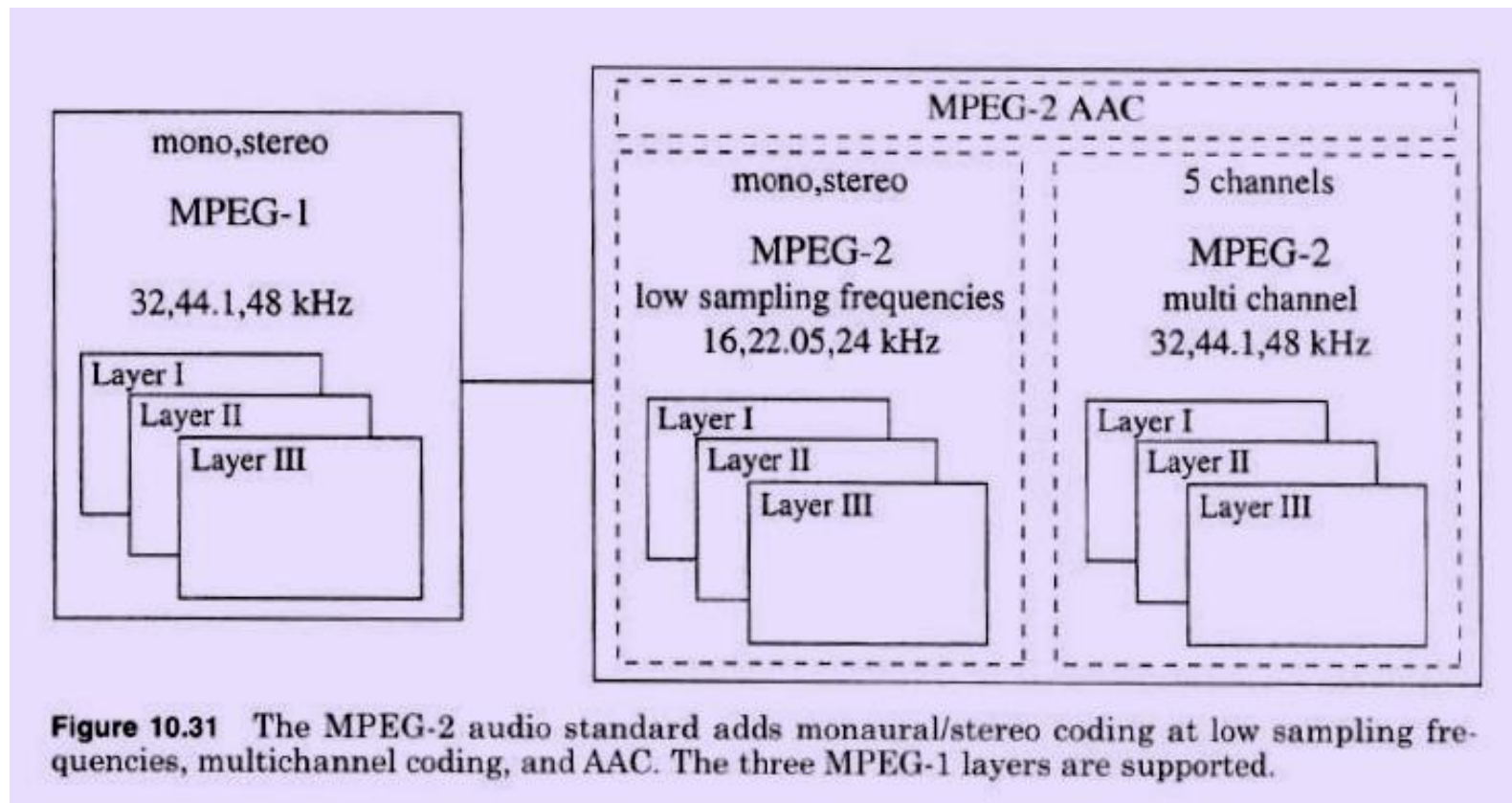


Diagrama em Blocos do Codificador MPEG-2 AAC

Diagrama em Blocos MPEG-2 AAC



Comparação MPEG-1 e MPEG-2



MPEG-4: Introdução

Não padroniza métodos de codificação áudio

Conceitos diferentes de outros MPEG

Suporta codificação em baixas taxas

Padrão de codificação em objetos, múltiplas ferramentas

Suporta som sintetizado

Prevê robustez a erros

Suporta Escalabilidade

Ferramentas de Codificação de Voz

Voz Natural

- Compressão, Transmissão e Decodificação de Voz

Voz Sintetizada

- Interface TTS (Text-to-Speech)
- Baixas taxas, para conexão com aplicações de animação facial e videoconferência em baixas taxas

Voz Natural

HVXC – Harmonic Vector eXcitation Coding

- “near toll quality”, 100-3800Hz, 8kHz amostragem
- 2 a 4 kbit/s fixas
- 1.2 a 1.7 kbit/s variável

Baseado no CELP – Code Excited Linear Prediction

- 8 e 16kHz de amostragem
- 100-3800Hz ou 50-7000Hz banda
- VAD (Voice Activity Detector), DTX (Discontinuous Transmission), CNG (Comfort Noise Generator)

TTS – Text-to-Speech Interface

Pode-se criar narrativas sem gravações

Comandos para uma única interface TTS

Informações transmitidas via IPA (International Phonetic Alphabet) ou texto (qualquer língua)

Hybrid/Multilevel Scalable TTS Interface – extensão com prosódia, FA – Facial Animation, AP – Animation Parameters. ISO/IEC 14496-2 Annex C.

Ferramentas
de
Codificação
de Áudio

Taxas de 6kbit/s até centenas de kbit/s, mono, estéreo e multi-canaís

MPEG-2 AAC com melhorias: MPEG-4 AAC

GA (General Audio) tools mais Quantização e Codificação adicionais

MPEG-4 low delay

MPEG-4 BSAC (Bit Sliced Arithmetic Coding), para melhora escalável da granularidade fina na quantização

MPEG-4 SBR

Questões, P2c (duas semanas)

1. No que consiste o critério de Reconstrução Perfeita? Como ele se caracteriza num banco de dois filtros QMF?
2. O que é o procedimento Overlap-And-Add (OLA) no TTS e no MPEG?
3. Explique resumidamente como é obtida a compressão no MPEG1? Qual é o modelo Psicoacústico empregado?
4. O que distingue o MP3 do MPEG-2?
5. Quais são as funcionalidades do Parametric Stereo no MPEG2 AAC?
6. O que é o SBR (Spectral Band Replication) no MPEG2 AAC e no MPEG4?
7. O que caracteriza um banco de filtros PQMF?

Questões Prova P2c (1 semana)

Sejam $r_1 r_2 r_3 r_4$ os quatro último números do seu RA, p.ex. RA = 11062416, $r_1 = 2$, $r_2 = 4$, $r_3 = 1$ e $r_4 = 6$

1) Projete um banco de filtros de reconstrução (síntese) de três canais (por transformada) onde os filtros de análise são:

$$\begin{bmatrix} H_0(z) \\ H_1(z) \\ H_2(z) \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_2 & r_4 & r_1 \\ r_4 & r_3 & r_2 \end{bmatrix} \begin{bmatrix} 1 \\ z^{-1} \\ z^{-2} \end{bmatrix}$$

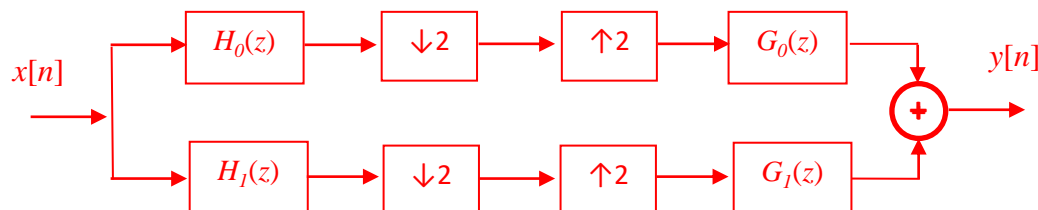
- a) (0,5 pontos) Forneça a expressão matemática que implementa os filtros de síntese e
- b) (0,5 pontos) Sugira uma realização eficiente deste banco de filtros.

2) Sejam as respostas impulsivas de um banco de filtros de dois canais:

$$h_0[n] = \left\{ \frac{1}{2}, SZ(mod(r_3 r_4, 3)), \frac{1}{2}, 0, 0, 0, 0, 0, \dots \right\}$$

$$h_1[n] = \left\{ -\frac{1}{2}, SZ(mod(r_3 r_4, 3)), \frac{1}{2}, 0, 0, 0, 0, 0, \dots \right\}$$

- onde $SZ(mod(r_3 r_4, 3))$ é a **sequência de zeros** da operação módulo do número $r_3 r_4$ por 3, no RA exemplo, $r_3 r_4 = 16$, logo $mod(16, 3) = 1$, $SZ(1) = \{0\}$, assim para o exemplo, $h_0[n] = \left\{ \frac{1}{2}, 0, \frac{1}{2}, 0, 0, 0, 0, 0, \dots \right\}$ e $h_1[n] = \left\{ -\frac{1}{2}, 0, \frac{1}{2}, 0, 0, 0, 0, 0, \dots \right\}$
- $SZ(0) = \emptyset$, conjunto vazio.



- a) (0,5 pontos) Obtenha a forma polinomial no domínio z dos respectivos filtros de análise;
- b) (0,5 pontos) Verifique se obedecem a condição de reconstrução perfeita para filtros QMF;
- c) (0,5 pontos) Obtenha a forma polinomial no domínio z dos filtros de síntese, para eliminação de *aliasing* com RP.

Referências

- DUTOIT, T. and MOREAU, N., **How is sound processed in an MP3 player?**, in: Applied Signal Processing, Ch.3, Springer, New York, 2009.
- BOSI, M. and GOLDBERG, R. E., **Introduction to Digital Audio Coding and Standards**, Kluwer Academic Publishers, Dordrecht, 2003.
- KOHLRAUSCH, A. and PAR, S., **Audio-Visual Interaction in the Context of Multi-Media Applications**, in: Communication Acoustics, Edited by Blauert, J., Springer, Berlin, 2005.
- VAIDYNATHAN, P. P., **Multirate Systems and Filter Banks**, Prentice-Hall PTR, New Jersey, 1993.
- PAN, D., **A Tutorial on MPEG/Audio Compression**, IEEE Multimedia Magazine, pp. 60-74, 1995.
- MINAMI, M., **Codificação de Áudio**, Slides da disciplina EN2616, 2013.
- VETTERLI, M. and KOVACEVIC, J., **Wavelets and Subband Coding**, Prentice Hall, Englewood Cliffs, 1995.
- PEREIRA, F. and EBRAHIMI, T. (editores), **The MPEG-4 Book**, Prentice-Hall, USA, 2002, cap.10 e 11, 451-544.