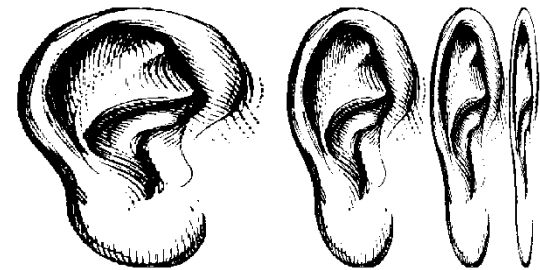


Compressão de Áudio

PTC3547 – Codificação e
Transmissão Multimídia

EPUSP – 05 / 2022

Guido Stolfi

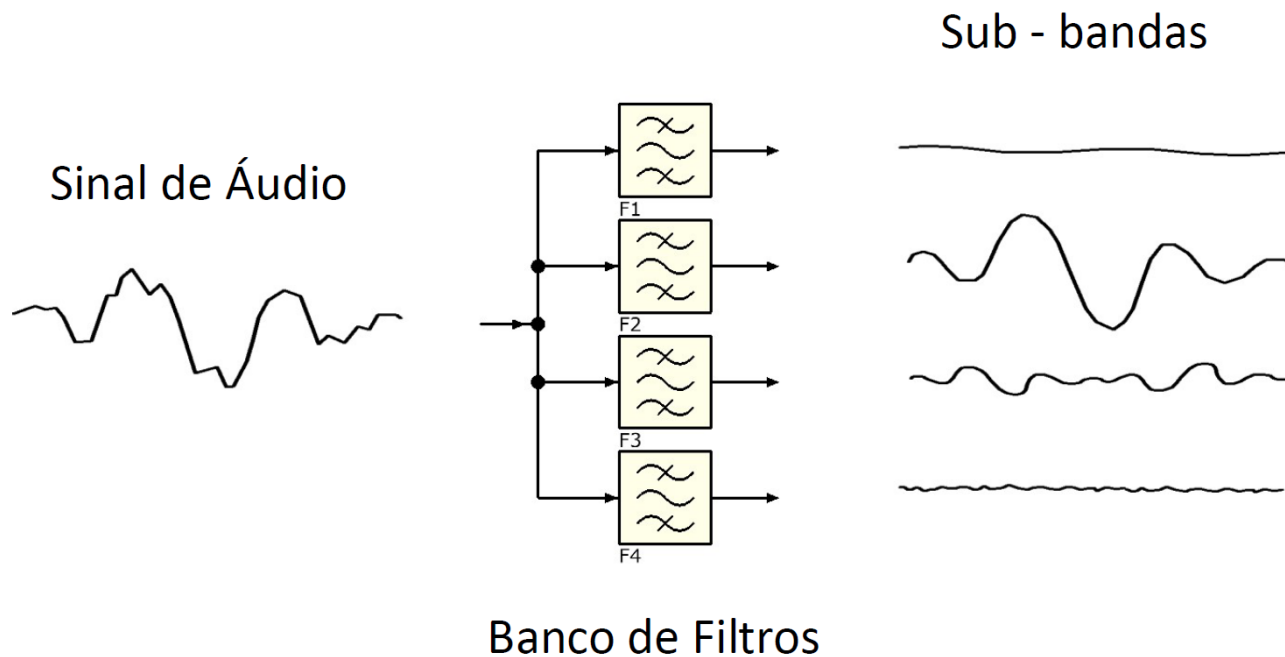


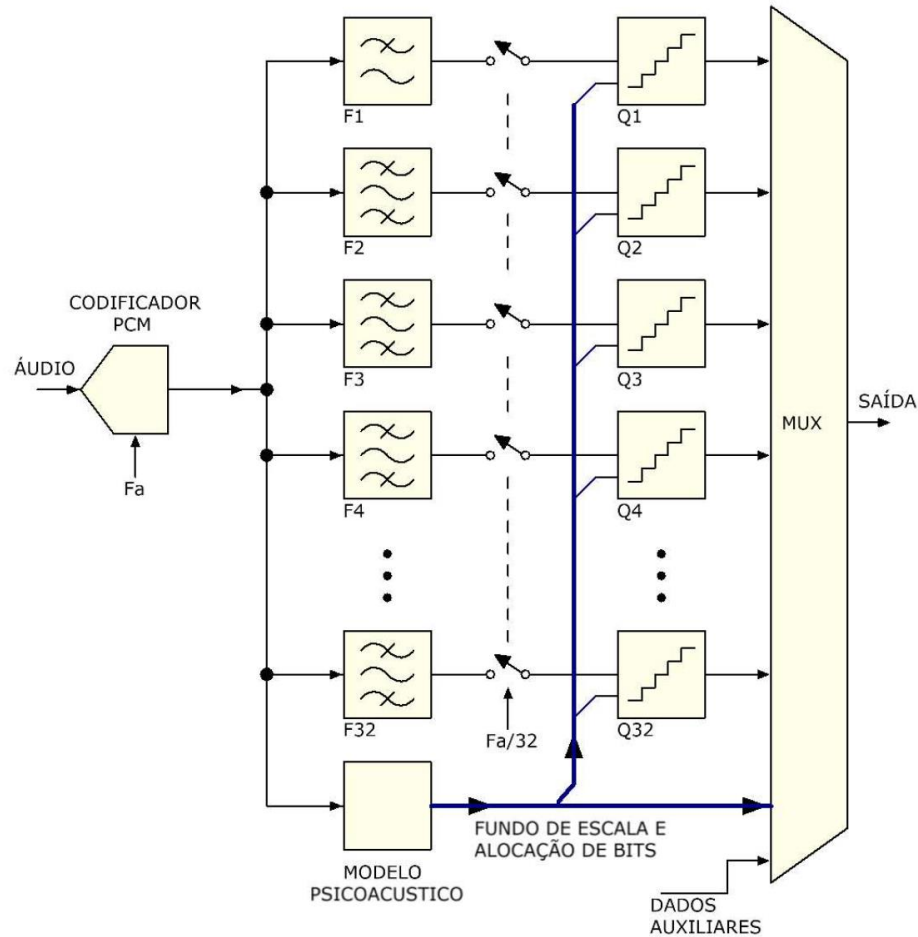
- Requisitos para transmissão de Áudio Digital
- Tecnologias chave para Compressão de Áudio
- Codificação por Sub-bandas
- Mascaramento
- Padrões MPEG
- Dolby AC-3 e outros padrões
- Comparações entre sistemas e configurações

- “Qualidade CD”
 - 2 canais de áudio (som estereofônico)
 - resposta em frequência de 20 a 20.000 Hz
 - relação Sinal/Ruído de 90 dB
- $2 \times 44.1 \text{ kHz} \times 16 \text{ bits} = 1.41 \text{ Mbits / s}$

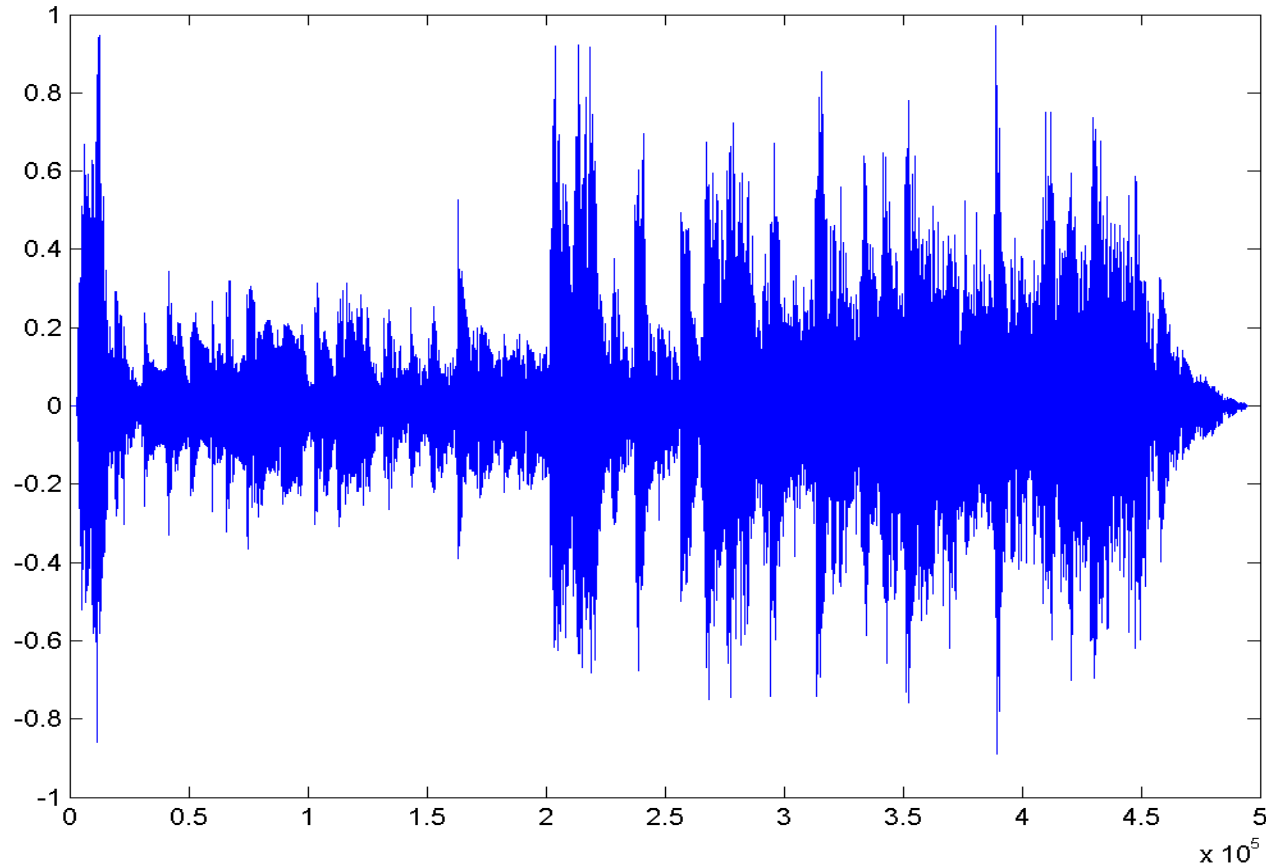
- “Qualidade Cinema”
 - 5 canais de áudio de 20 a 20.000 Hz
 - Direito, Esquerdo, Central, Traseiro Direito e Traseiro Esquerdo
 - 1 canal efeitos sonoros de 3 a 150 Hz ("Sub-Woofer", omnidirecional)
 - Faixa dinâmica de 90 dB
- $5.1 \times 48 \text{ kHz} \times 16 \text{ bits} = 3.9 \text{ Mbits / s}$

- Codificação por Sub-bandas
- Modelo psico-acústico do Mascaramento
- Codificação com Preditor
- Compactação (sem perdas)



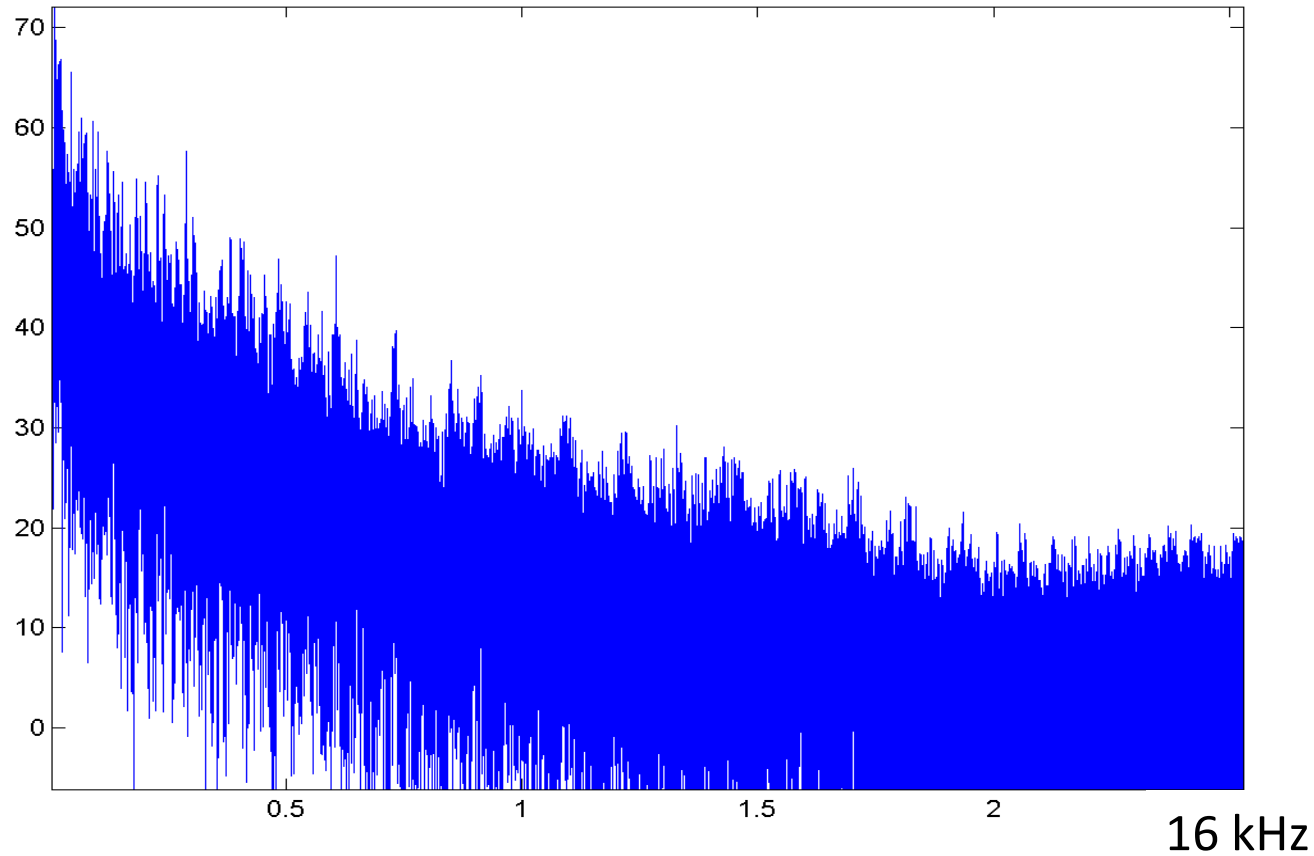


Filtragem em Sub-Bandas

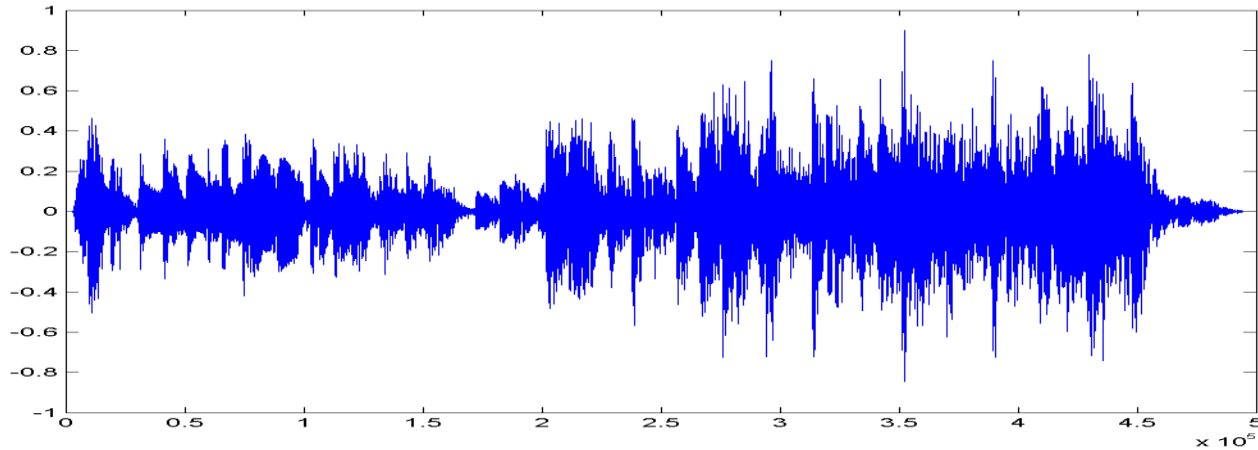


“Original”
 $f_a = 32 \text{ kHz}$

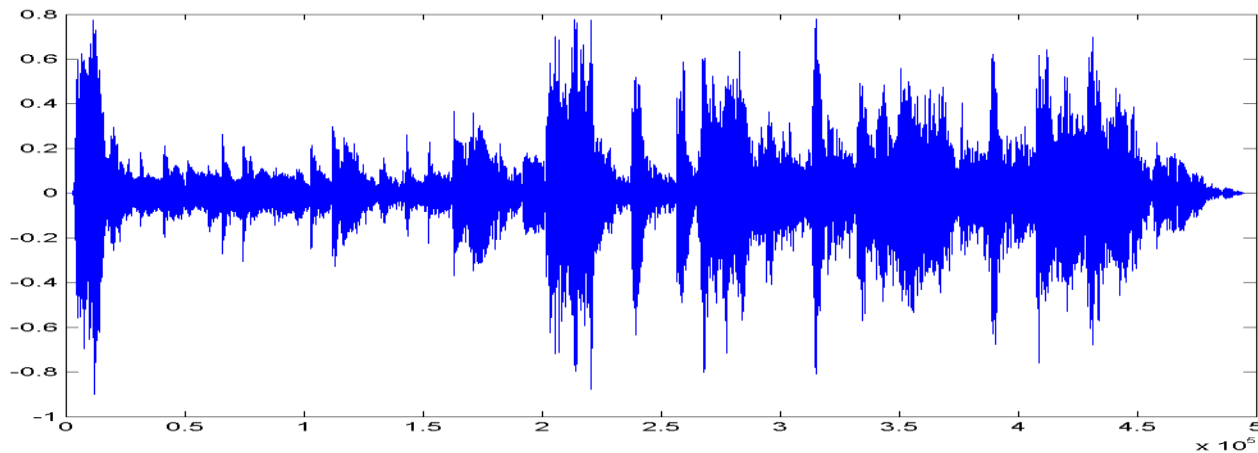
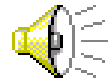
Espectro do Sinal de Teste (dB)



Separação em Sub-Bandas (1-2)



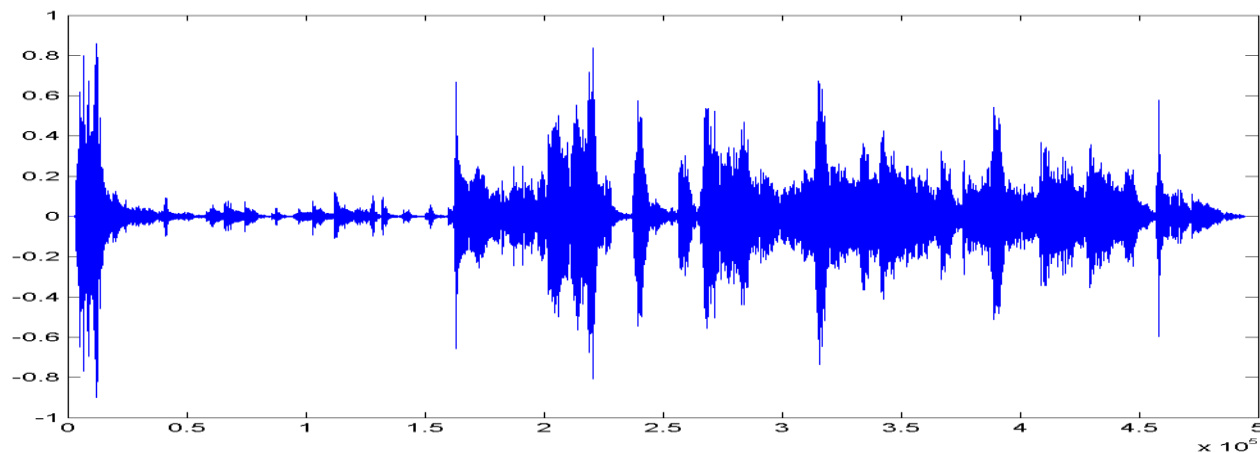
0-500 Hz



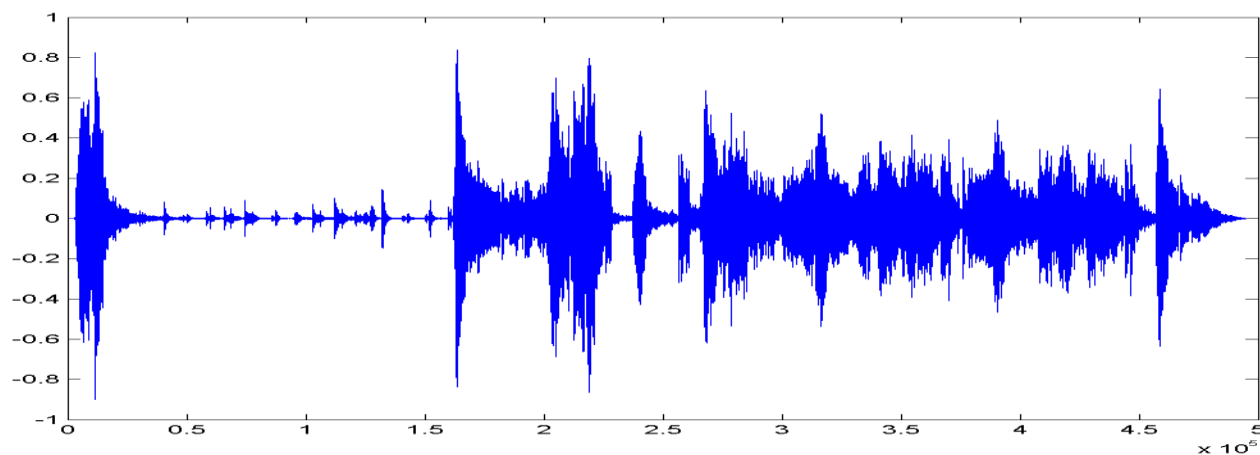
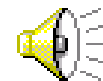
500-1000 Hz



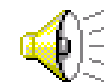
Separação em Sub-Bandas (3-4)



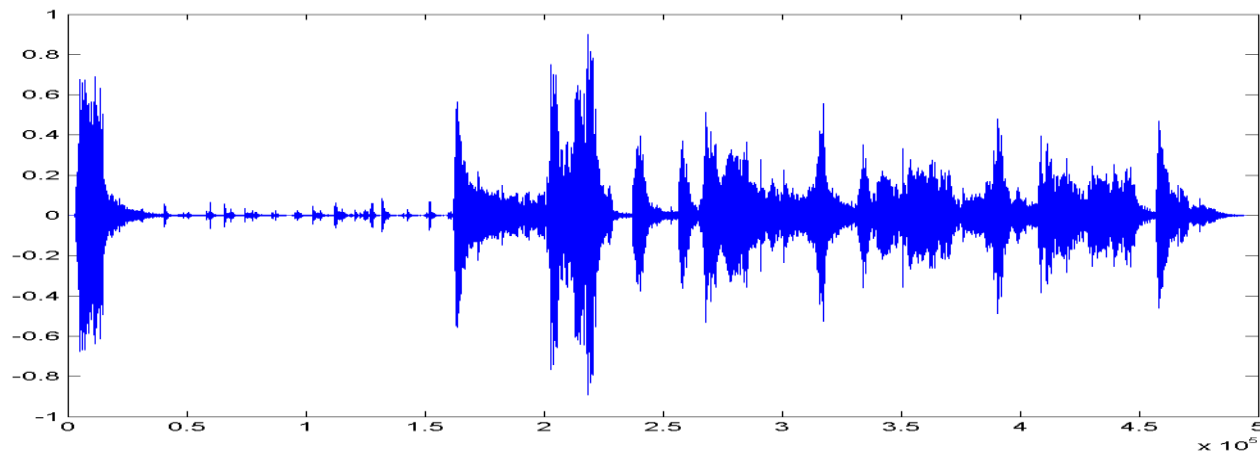
1000-1500 Hz



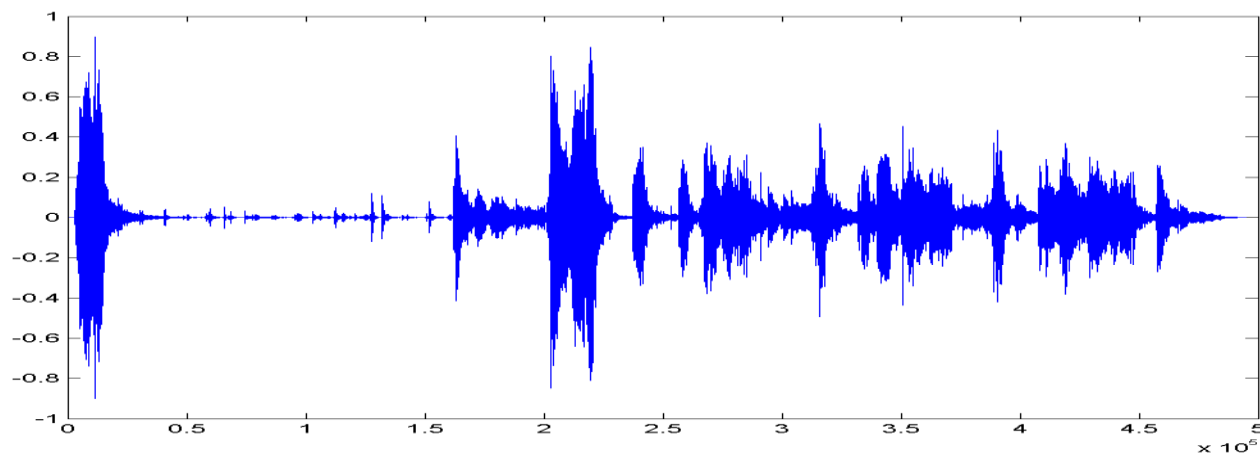
1500-2000 Hz



Separação em Sub-Bandas (5-6)



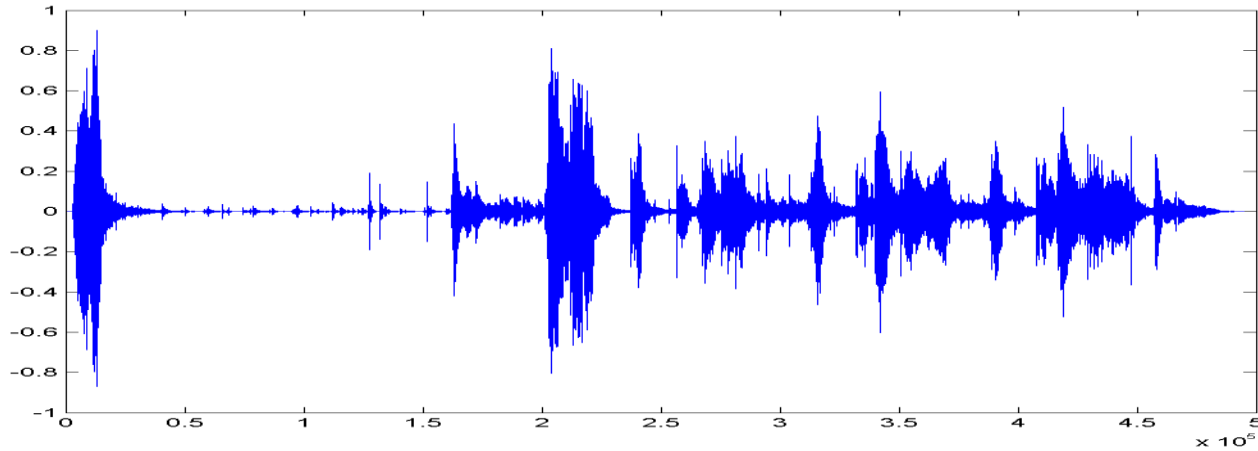
2000-2500 Hz



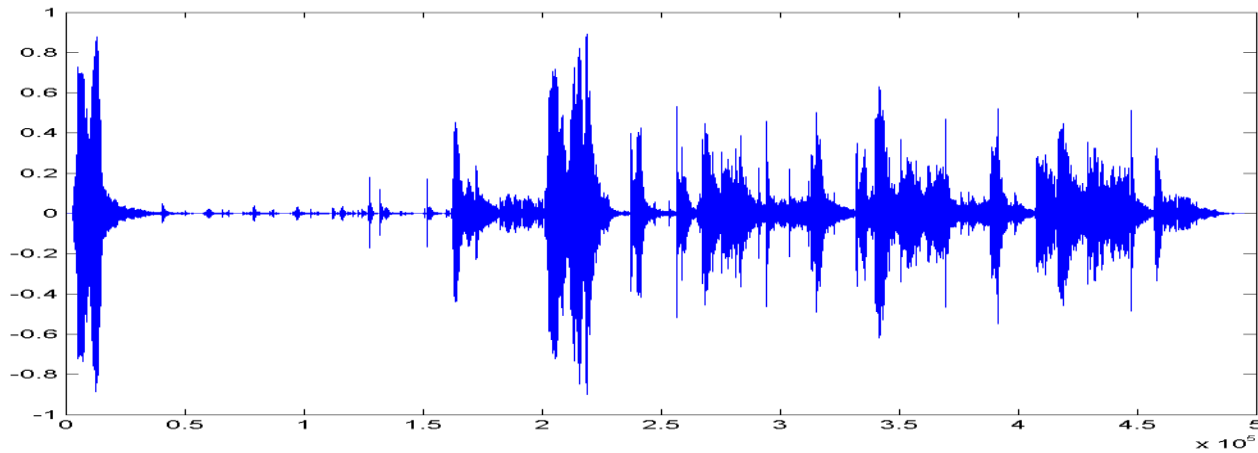
2500-3000 Hz



Separação em Sub-Bandas (7-8)



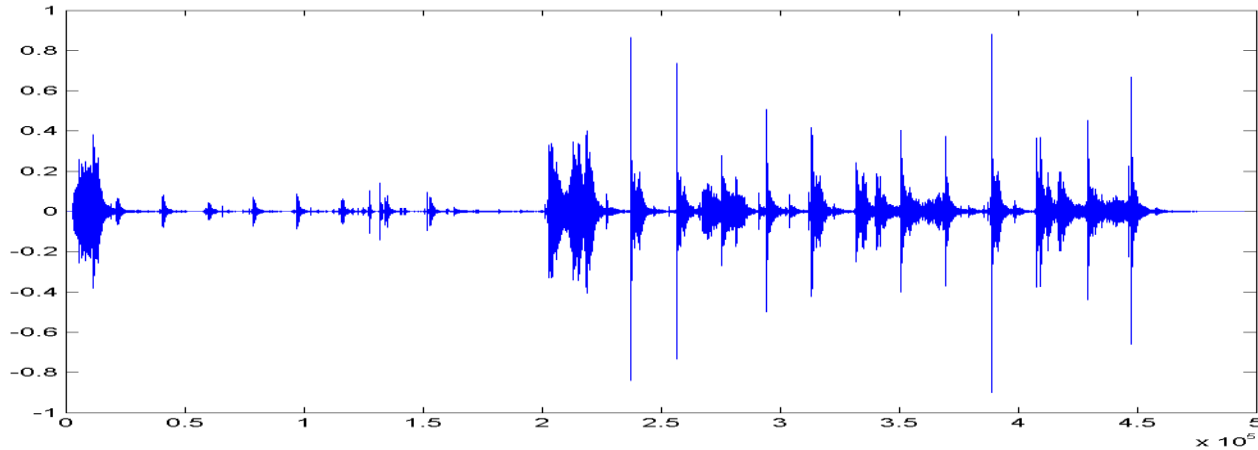
3000-3500 Hz



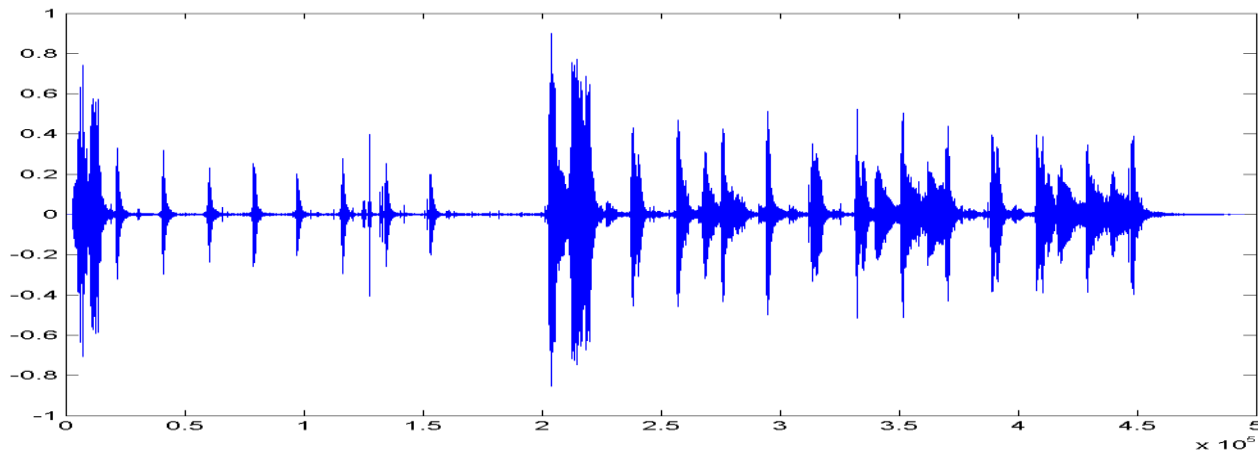
3500-4000 Hz



Separação em Sub-Bandas (16, 24)



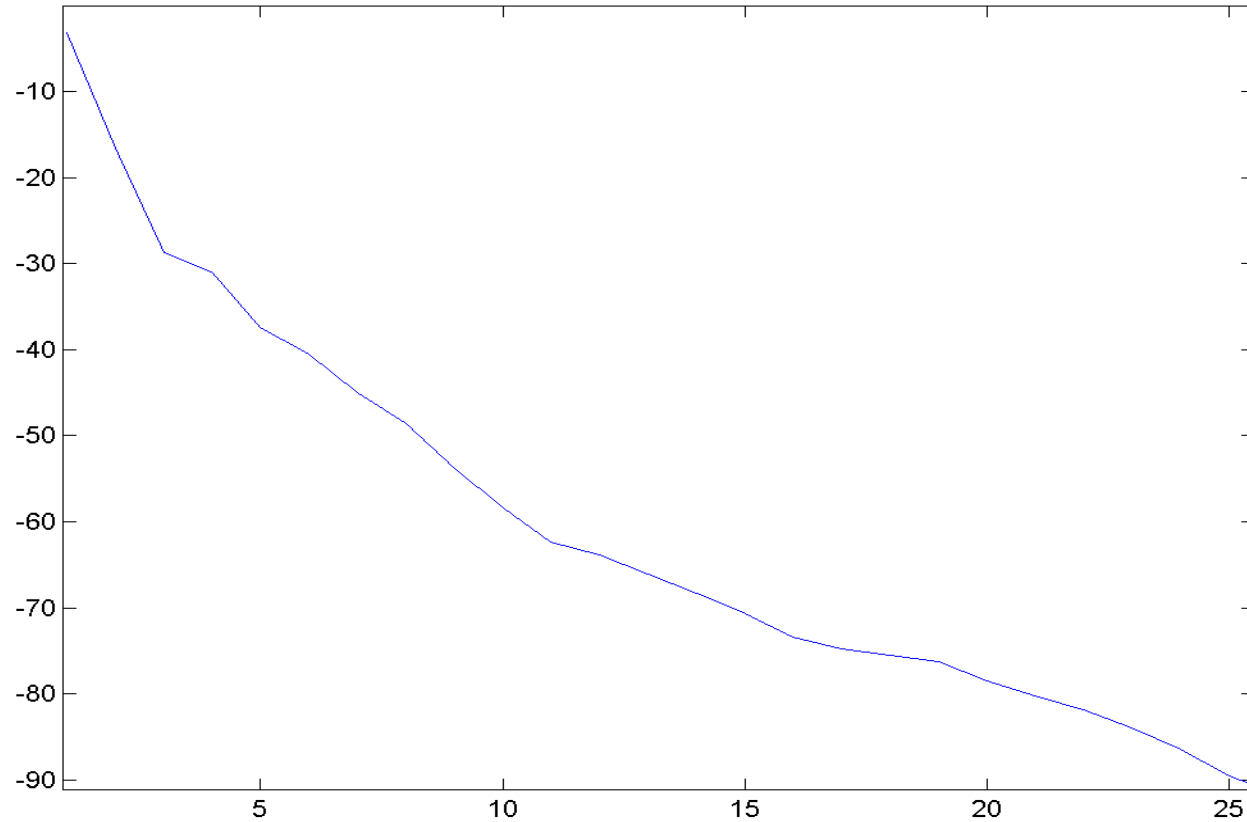
7500-8000 Hz



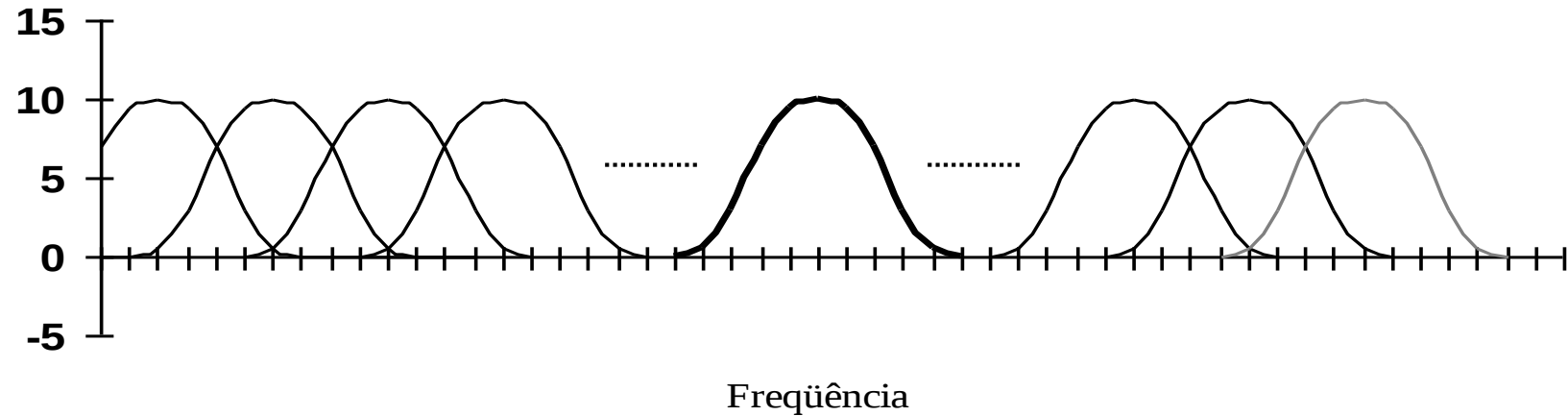
11500-12000 Hz



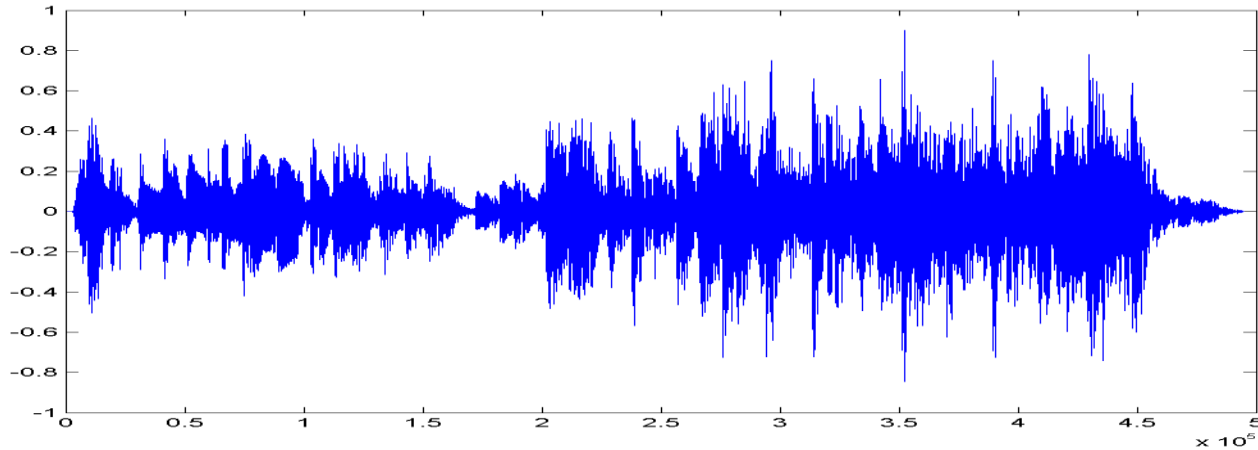
Energia Relativa das Sub-Bandas (dB)



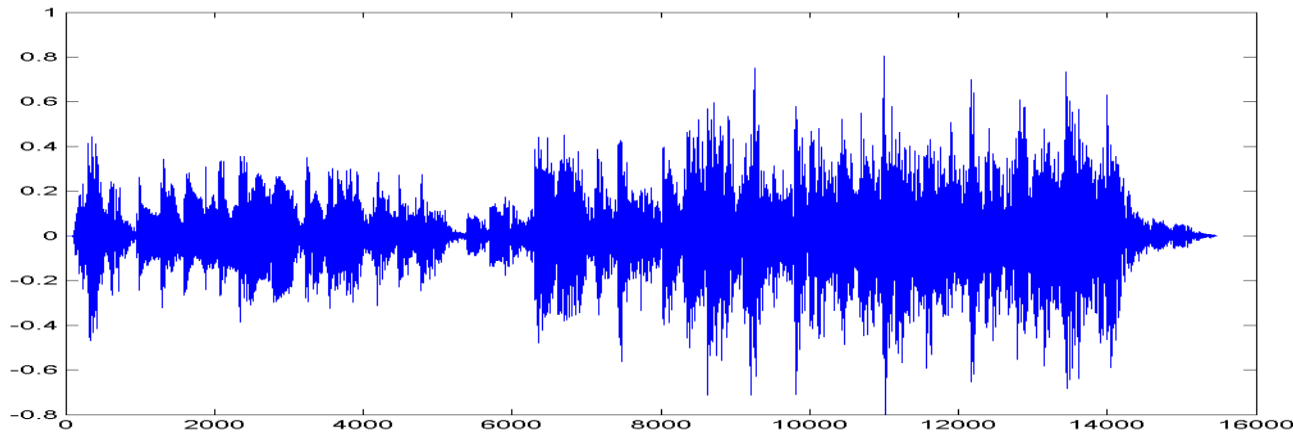
- Cada sub-banda ocupa $1/32$ da banda total de frequências
- Pode ser reamostrada com frequência de amostragem igual a $1/32$ do original
- Vai ocorrer rebatimento espectral, pois os filtros não são ideais
- Efeitos de rebatimento espectral serão cancelados na reconstrução devido à simetria das respostas dos filtros (propriedade dos filtros QMF)



Sub-Bandas Reamostradas (1)



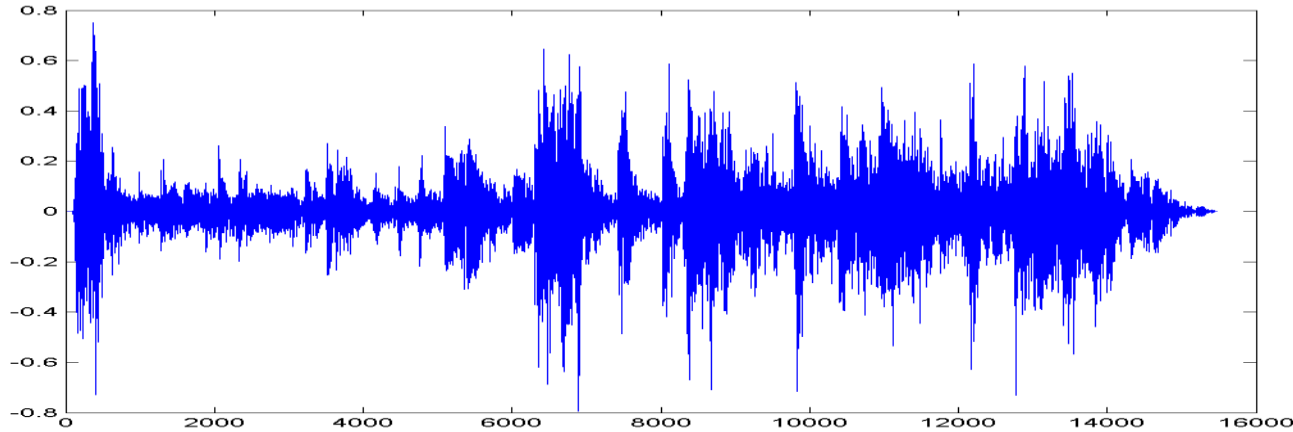
0-500 Hz
 $f_a = 32 \text{ kHz}$



0-500 Hz
 $f_a = 1 \text{ kHz}$

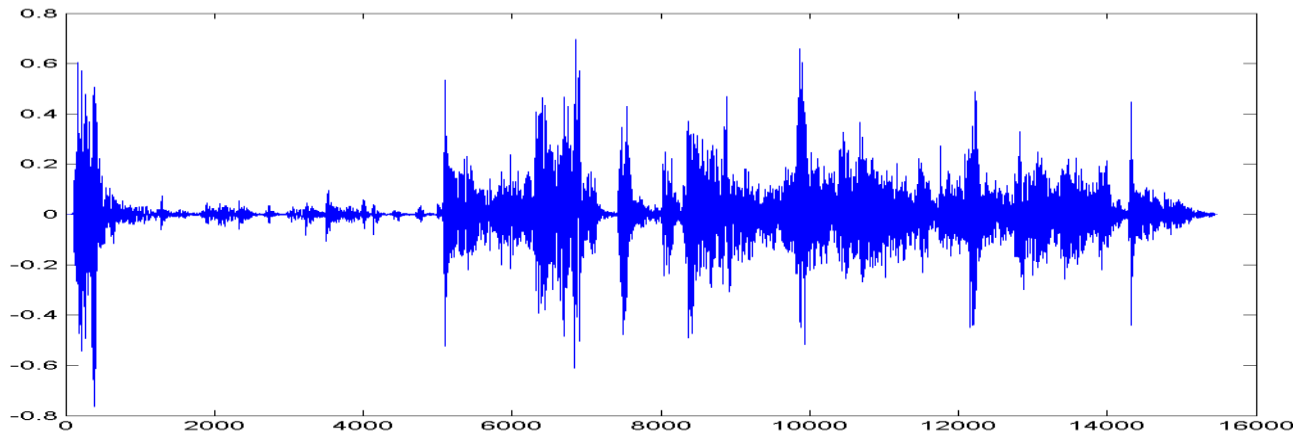


Sub-Bandas Reamostradas (2 - 3)



500-1000 Hz

$f_a = 1$ kHz

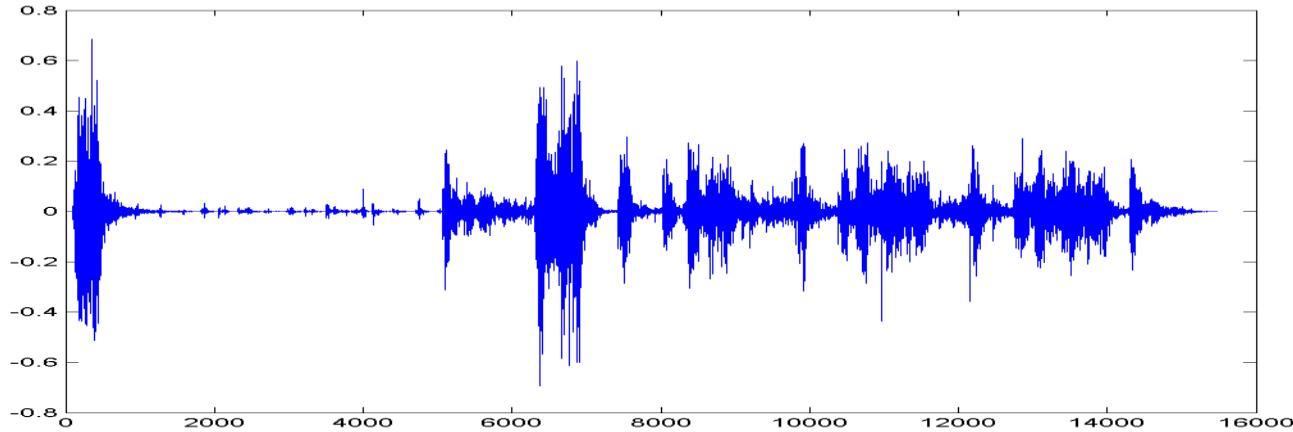


1000-1500 Hz

$f_a = 1$ kHz

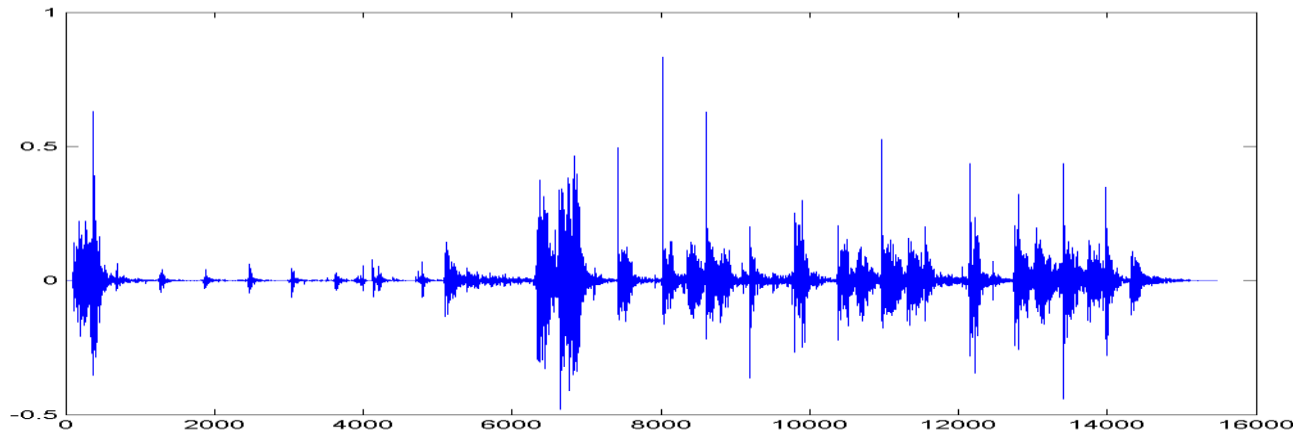


Sub-Bandas Reamostradas (6, 12)



2500-3000 Hz

$f_a = 1$ kHz

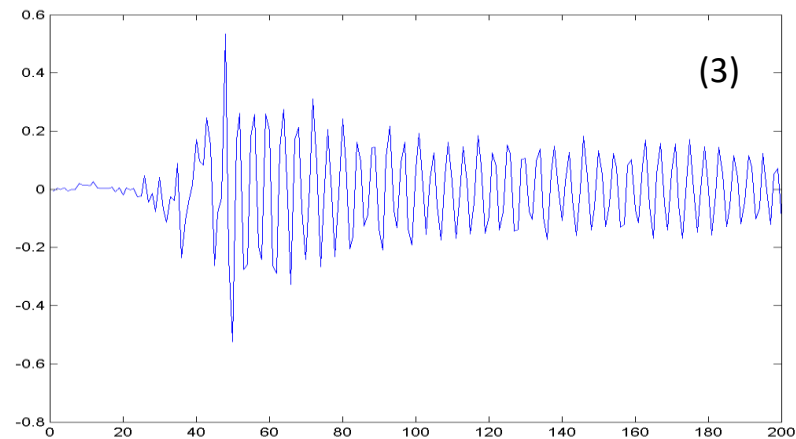
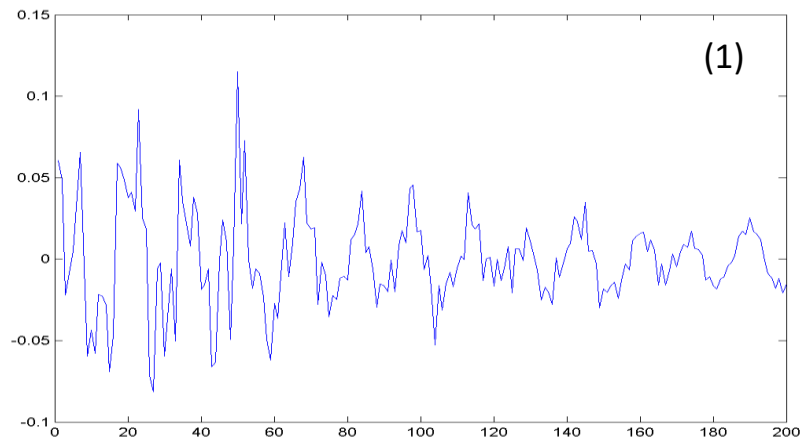
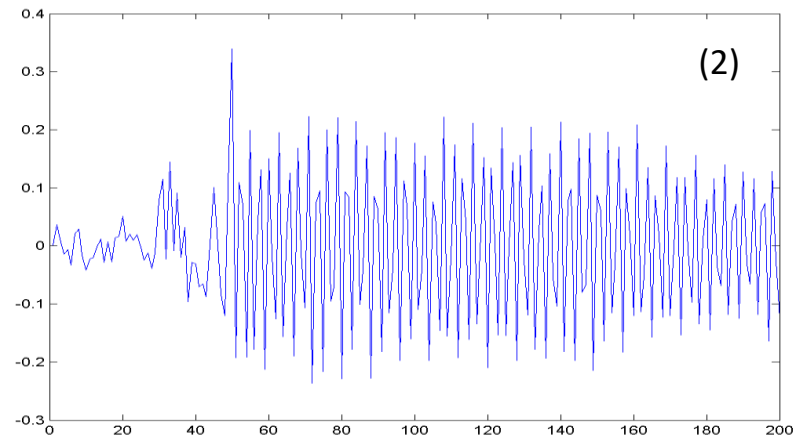
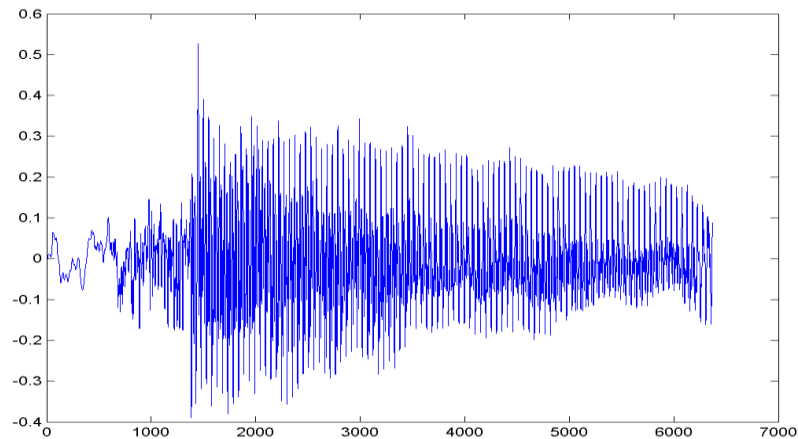


5500-6000 Hz

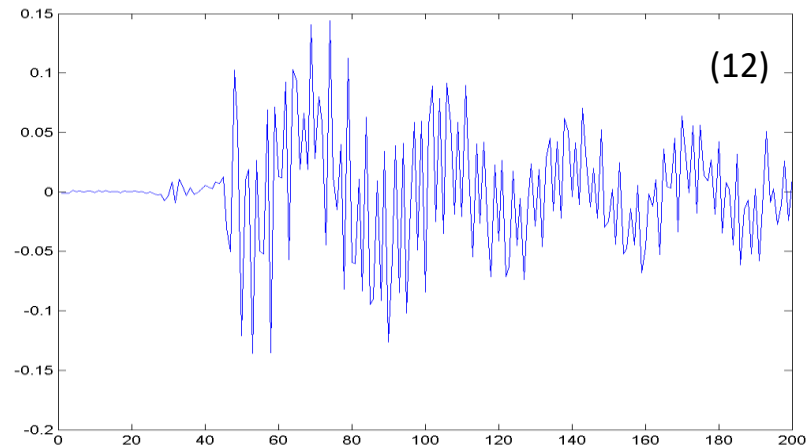
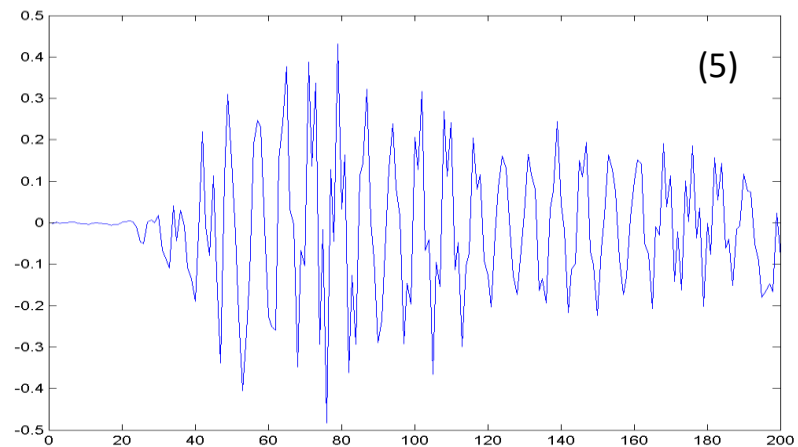
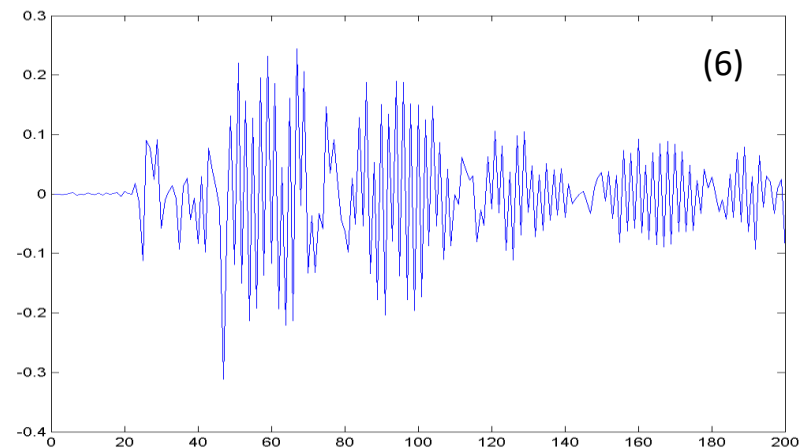
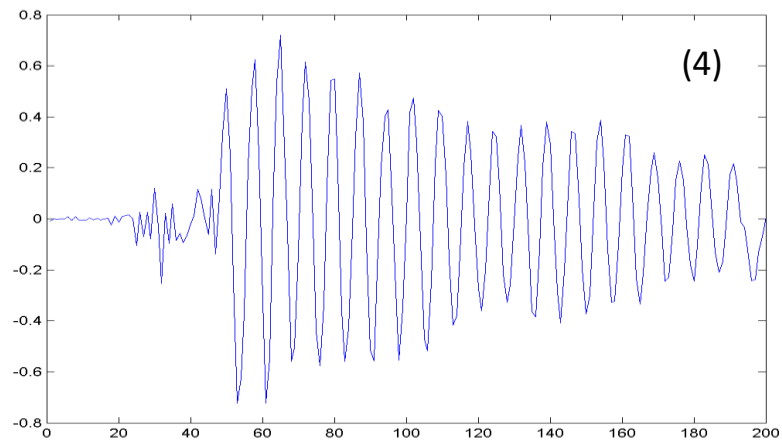
$f_a = 1$ kHz

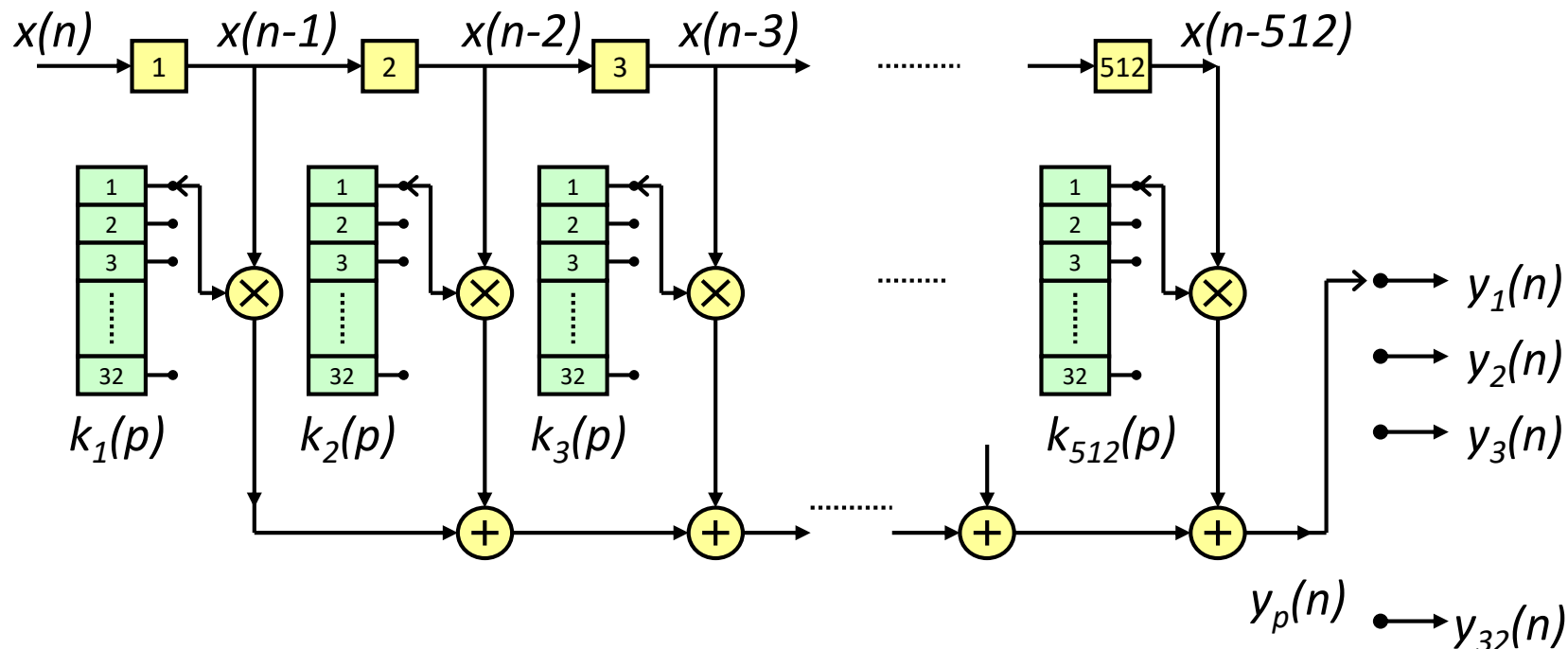


Detalhes da Forma de Onda

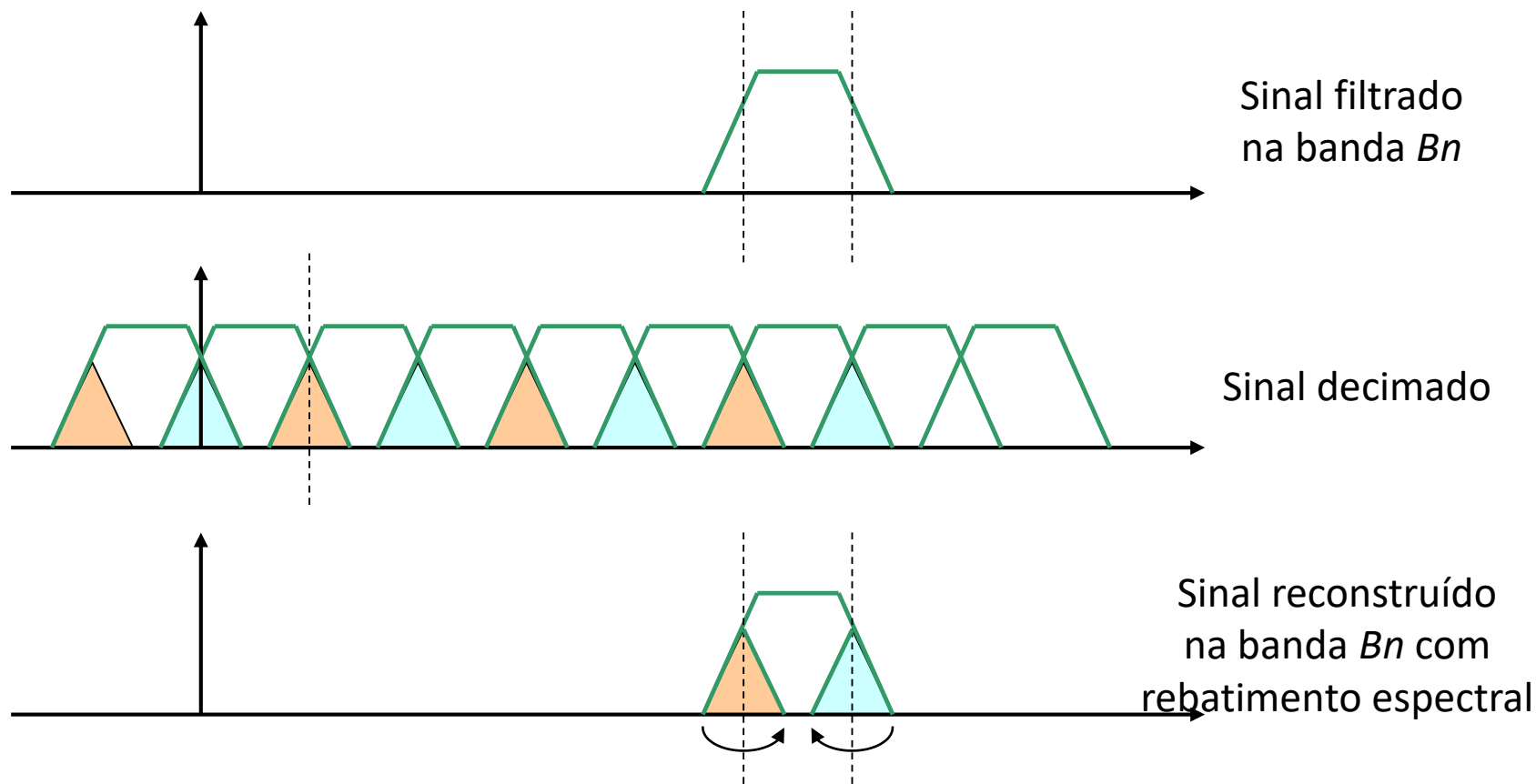


Detalhes da Forma de Onda

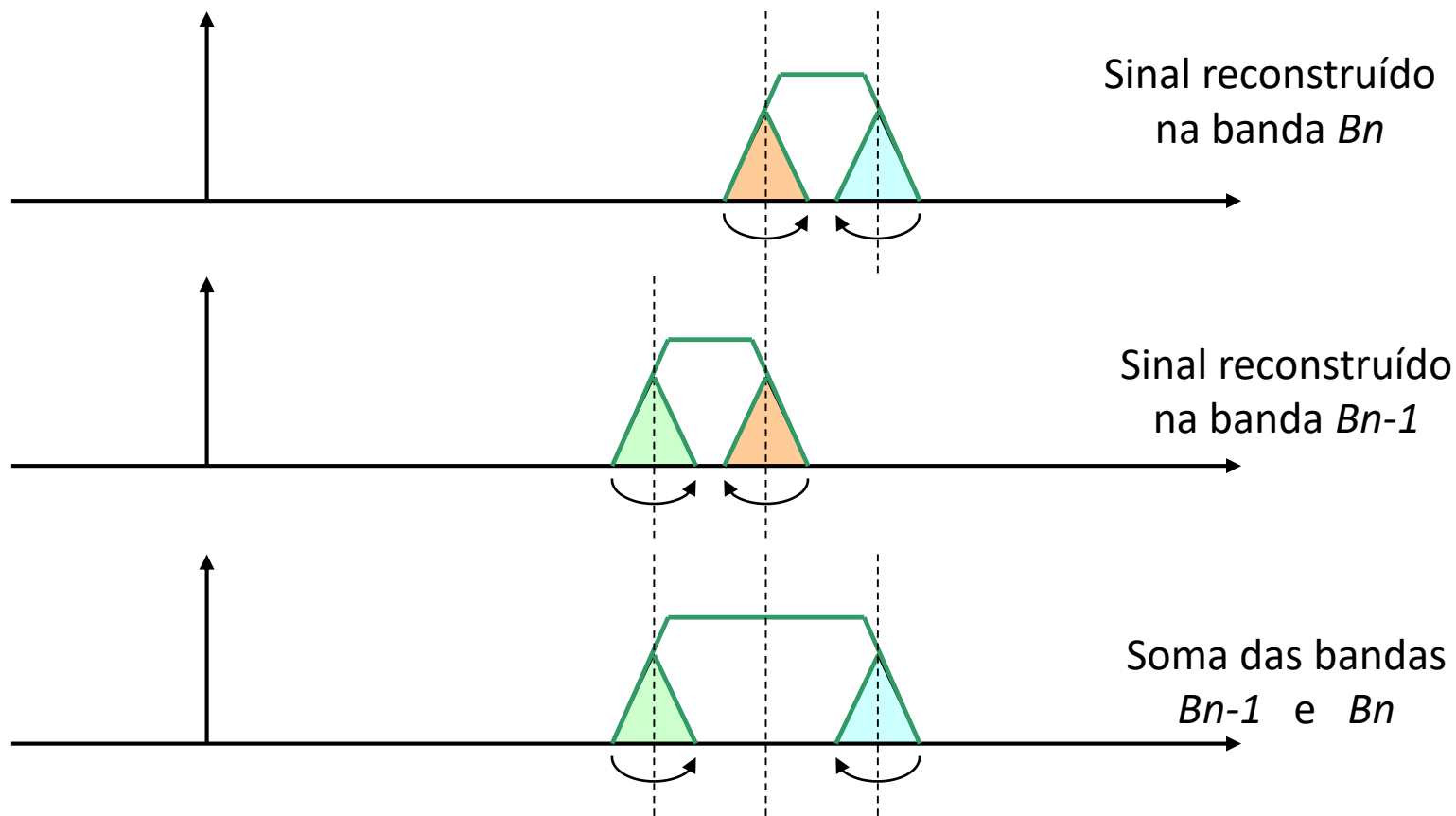




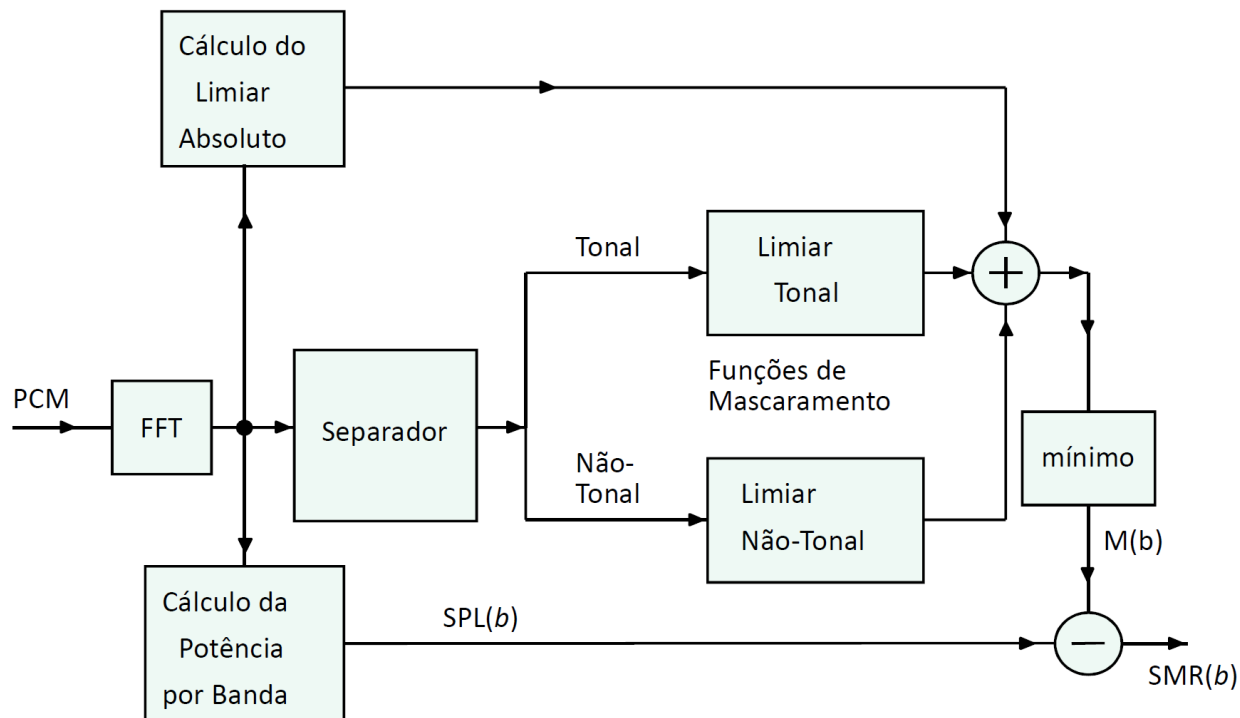
Banco de filtros polifásicos, 32 x 512 , $y_p(n)$ calculado cada 32 amostras



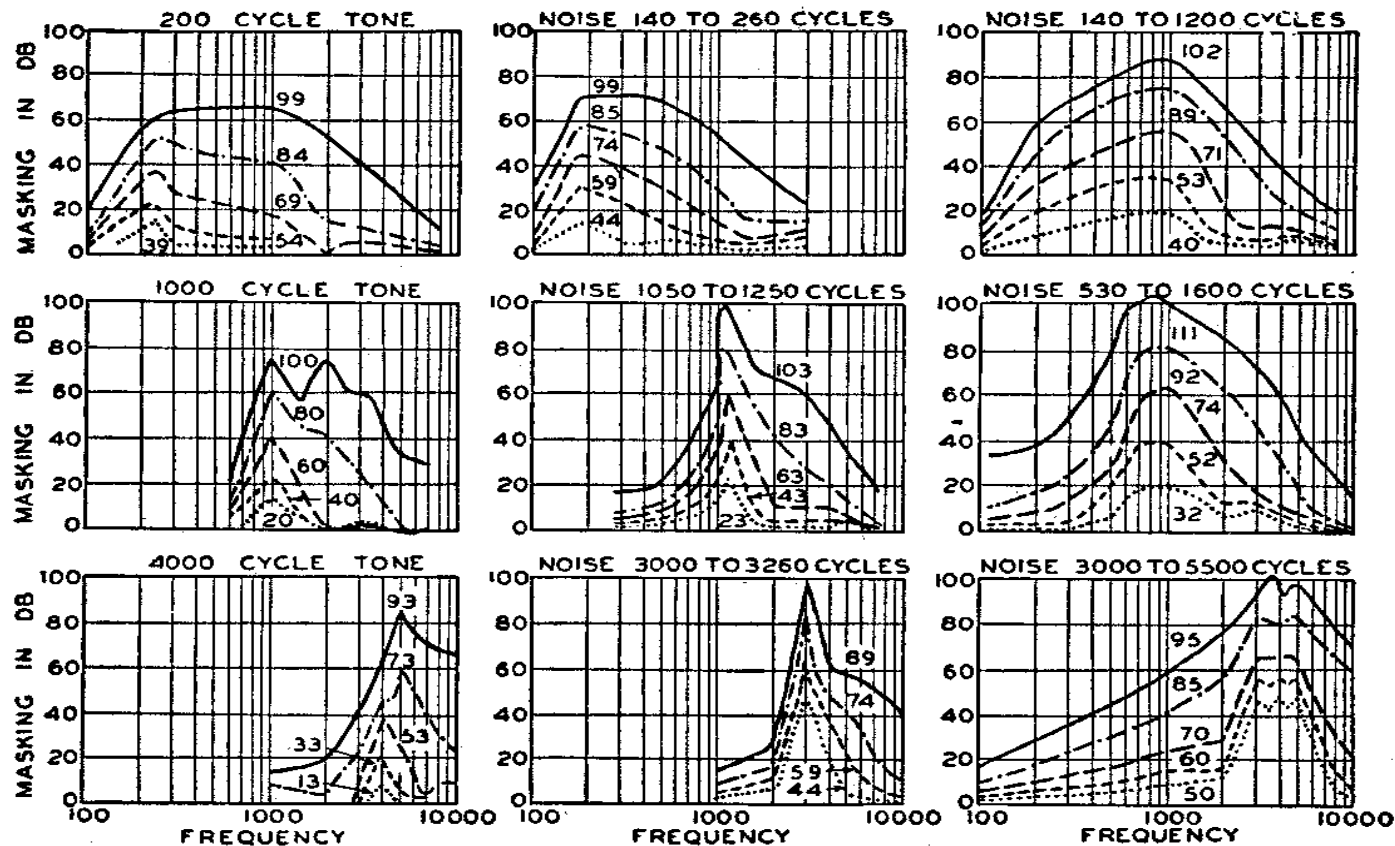
Cancelamento de Rebatimento Espectral



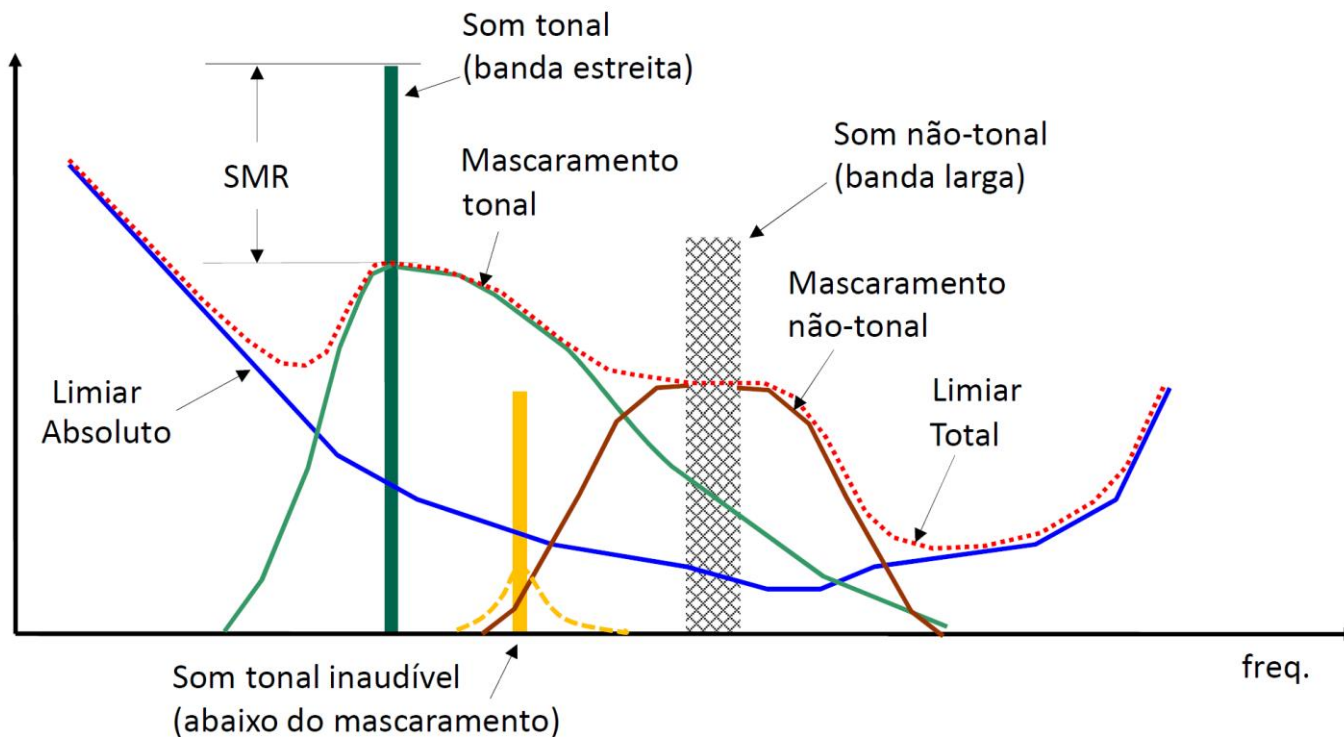
Mascaramento

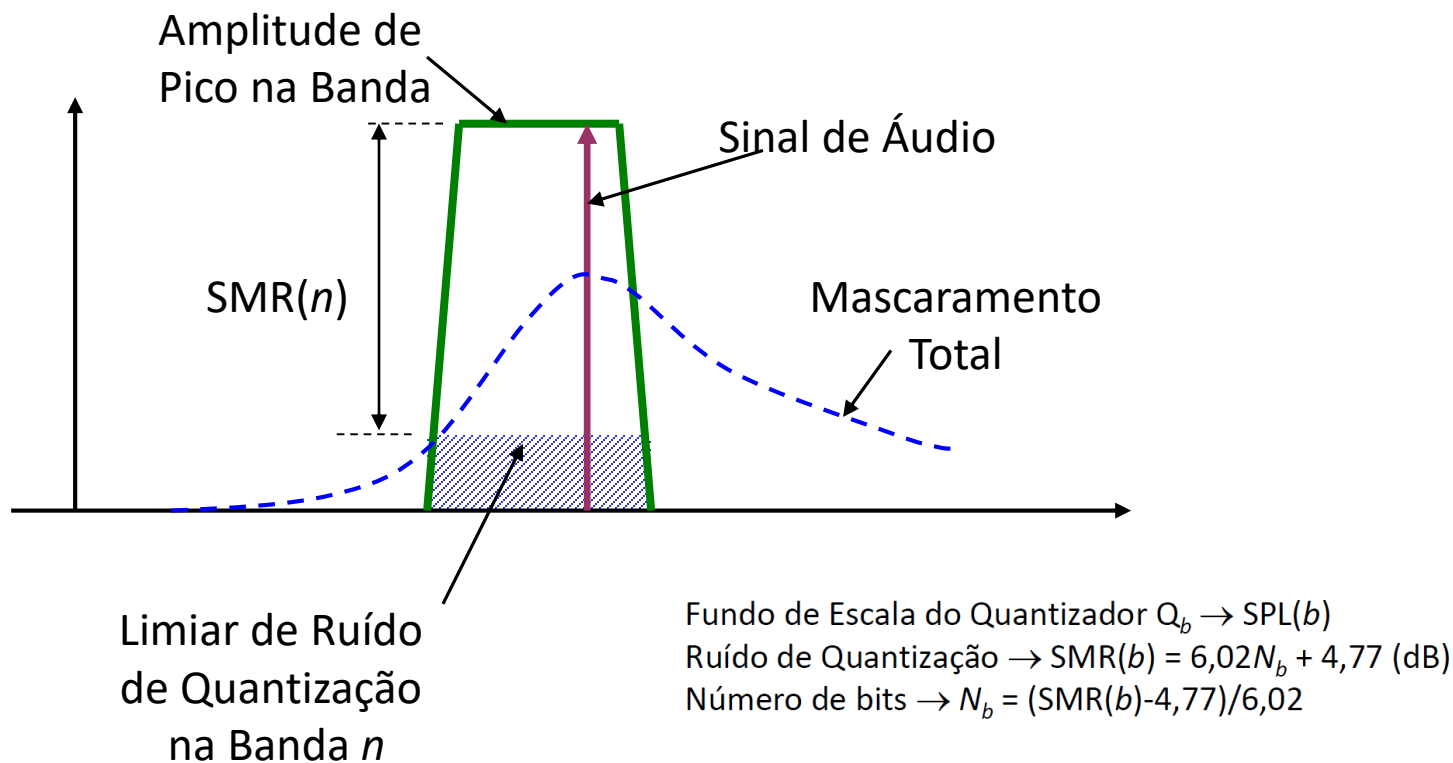


Curvas de Mascaramento (ex.)



Exemplo de Cálculo de Limiar de Mascaramento Total





MPEG-1 Nível I:

Codificação feita em Quadros de 12 amostras por sub-banda (384 amostras e 32 sub-bandas);

Para cada quadro permanecem constantes os fatores de escala e alocações de bits por amostra por banda;

Aceita taxas de amostragem de 32kHz, 44.1 kHz ou 48 kHz para um ou dois canais independentes ou associados;

MPEG-1 Nível II (MUSICAM):

Codificação em Quadros de 3×12 amostras por sub-banda (1152 amostras) proporcionando menor "overhead" na transmissão;

A análise de espectro para o modelo psico-acústico é feita com FFT de 1024 pontos;

Admite codificação estatística agrupando 3 amostras consecutivas em um único código;

MPEG-1 Nível III (MP3):

Codificação em Quadros de 3×12 amostras por sub-banda (1152 amostras);

Usa um modelo psico-acústico mais elaborado;

Utiliza transformada de cossenos (DCT) para subdividir as Sub-bandas em 6 ou 18 sub-sub-bandas, proporcionando redução de "Aliasing" entre sub-bandas;

Efetua aplicação de janelas temporais curtas (6 amostras) ou longas (18 amostras) antes da DCT;

Pode usar Quantizador não-linear : $Y=Q(X^{0.75})$;

Adota fator de escala válido por bandas;

Aplica codificação de Huffman nas amostras quantizadas;

Utiliza Memória Elástica ("Buffer") p/ regularizar taxa de bits;

Implementa um controle de taxa de compressão baseado na ocupação do "Buffer";

MPEG-2 (Níveis I, II, III):

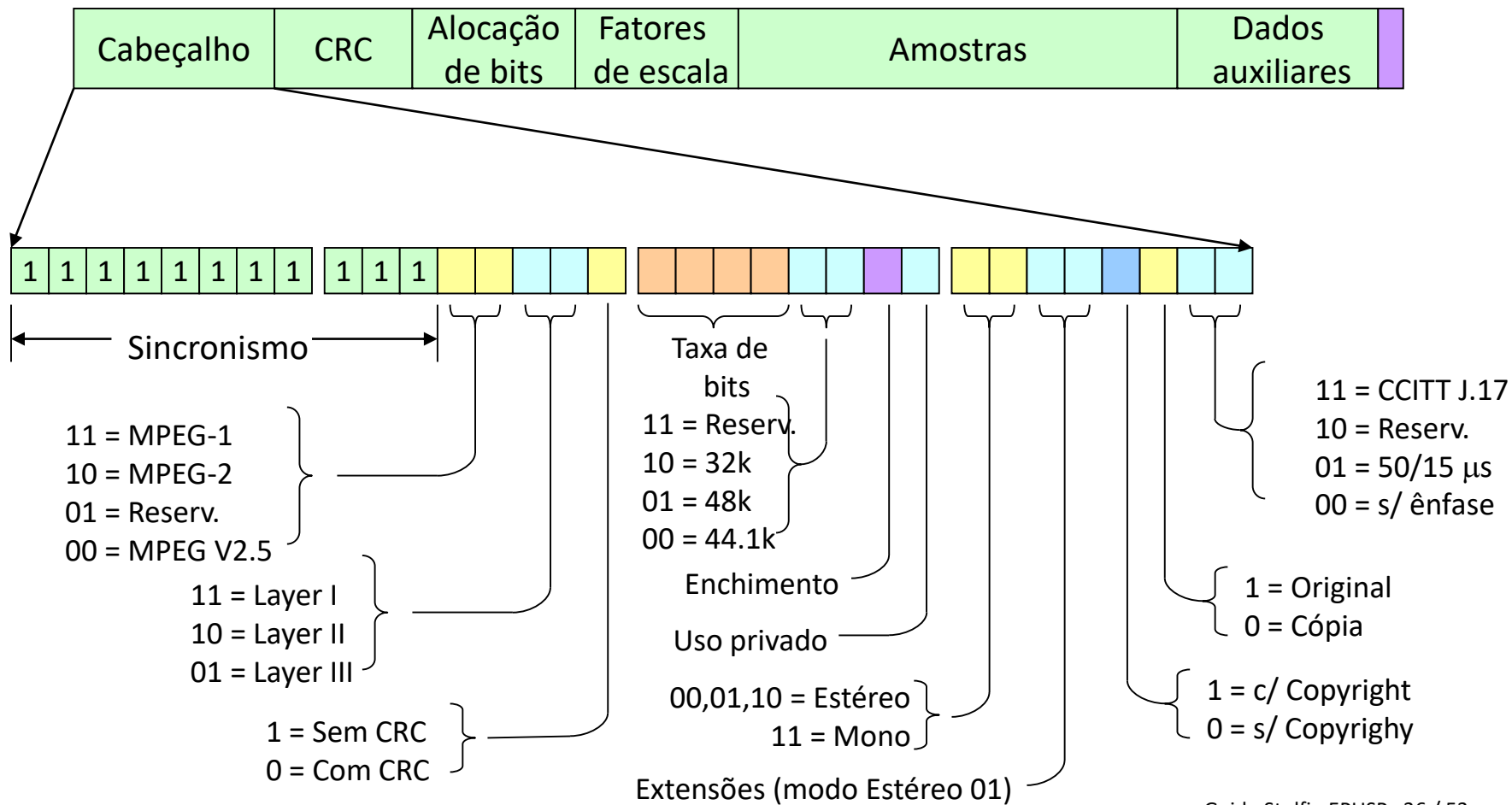
Suporte para 5.1 canais (HDTV), com predição adaptativa dos canais adicionais baseada nos canais L e R;

Prevê suporte para até 7 canais adicionais de voz (multilingual);

Codificação e decodificação compatível com MPEG-1 (sintaxe inclui um fluxo de dados compatível com MPEG-1 carregando os canais L e R);

Aceita outras taxas de amostragem (16, 22.05 e 24 kHz).

Estrutura do Pacote Áudio MPEG-1

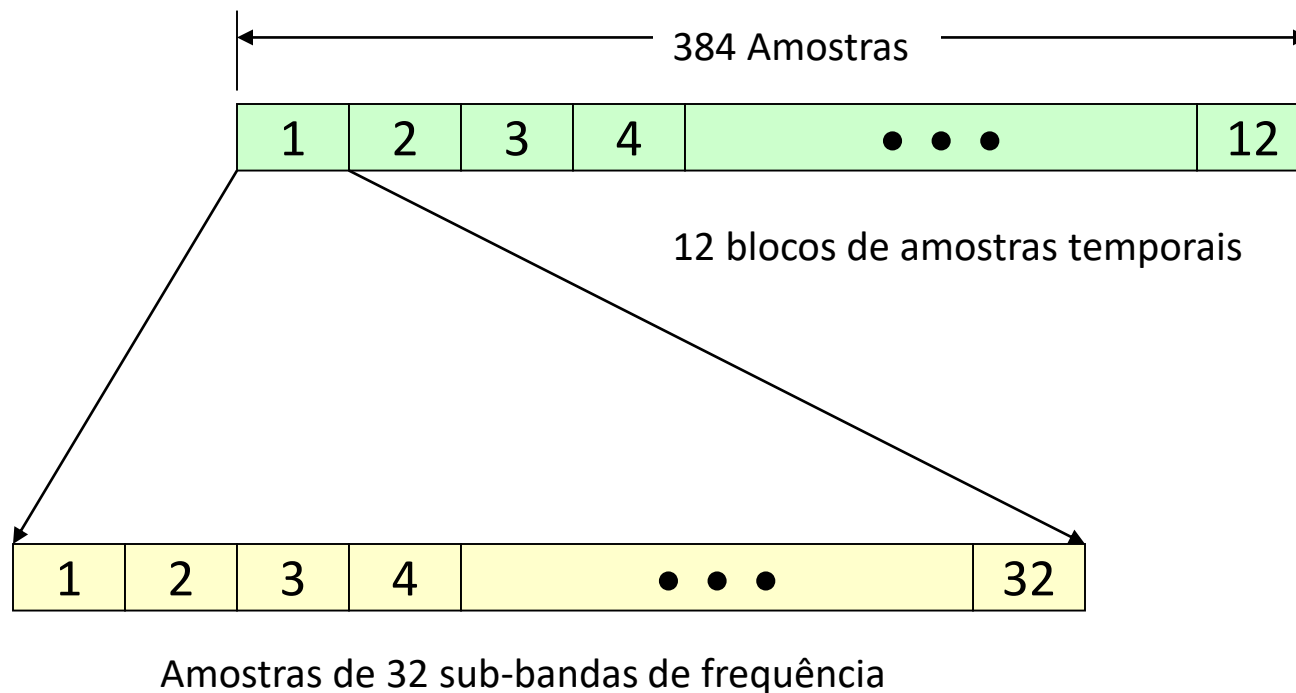


	MPEG-1	MPEG-1	MPEG-1	MPEG-2	MPEG-2
	Layer I	Layer II	Layer III	Layer I	Layers II e III
0000	livre	livre	livre	livre	livre
0001	32	32	32	32	8
0010	64	48	40	48	16
0011	96	56	48	56	24
0100	128	64	56	64	32
0101	160	80	64	80	40
0110	192	96	80	96	48
0111	224	112	96	112	56
1000	256	128	112	128	64
1001	288	160	128	144	80
1010	320	192	160	160	96
1011	352	224	192	176	112
1100	384	256	224	192	128
1101	416	320	256	224	144
1110	448	384	320	256	160
1111	N. A.	N. A.	N. A.	N. A.	N. A.

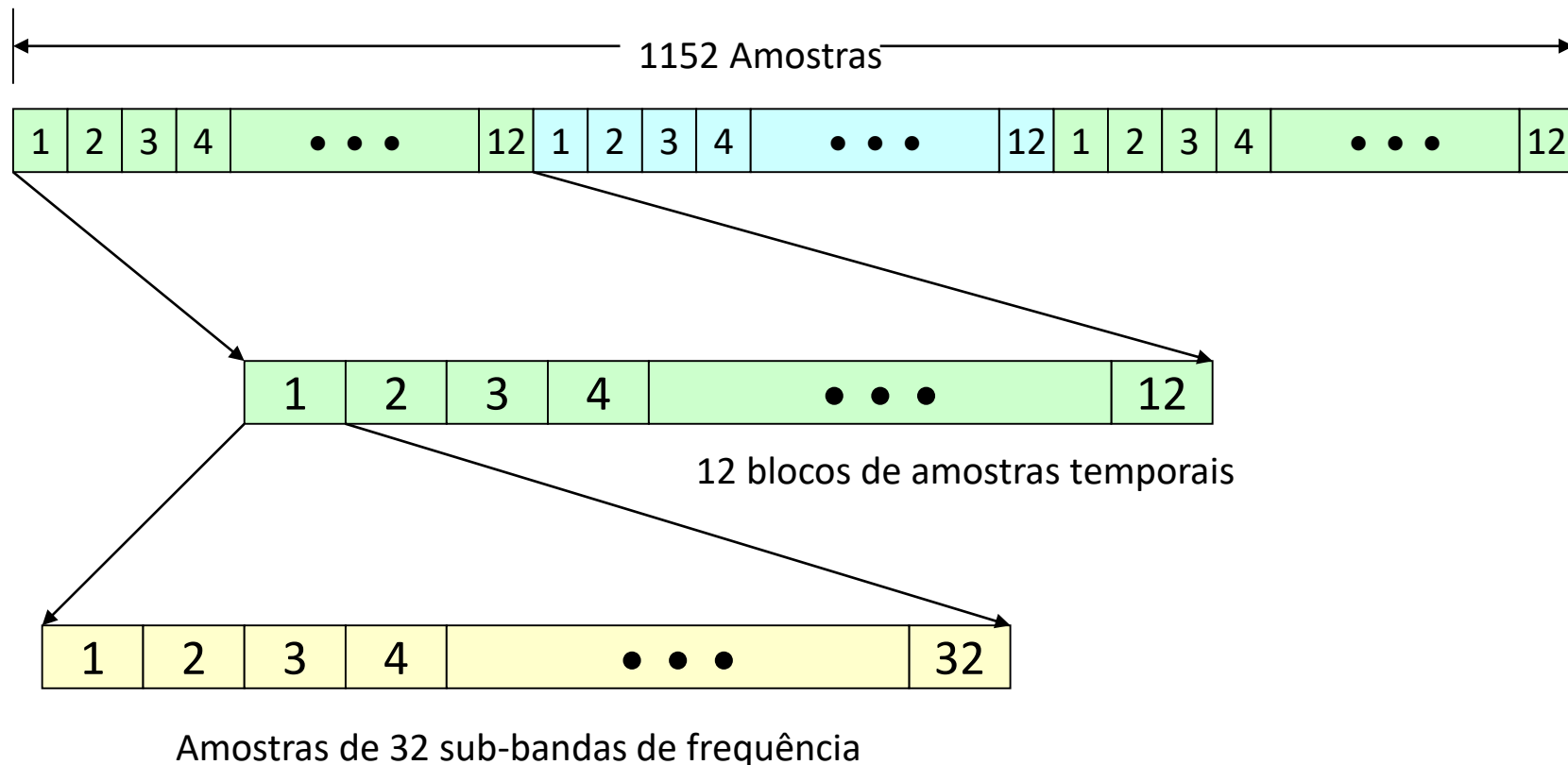
- Sub-bandas divididas em 3 grupos (Low, Mid, High)
- Bandas 'Low': 4 bits por banda selecionam 1 entre 15 quantizadores, com 3 a 65535 níveis de quantização
- Bandas 'Mid': 3 bits por banda, selecionam um de 7 quantizadores
- Bandas 'High': 2 bits por banda, selecionam um de 3 quantizadores possíveis

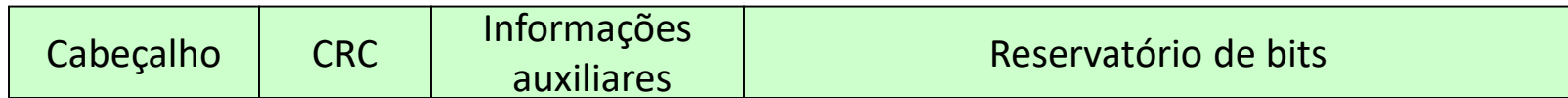
- 6 bits para cada banda, selecionam fatores de escala (ganhos) com faixa dinâmica de 120 dB.
- Layer II: Campo de 2 bits por banda (*SCFSI – Scale Factor Select Information*) indica se o fator de escala vale para um, dois ou três blocos consecutivos de 12 amostras.

Blocos de Amostras (MPEG-1, Layer 1)

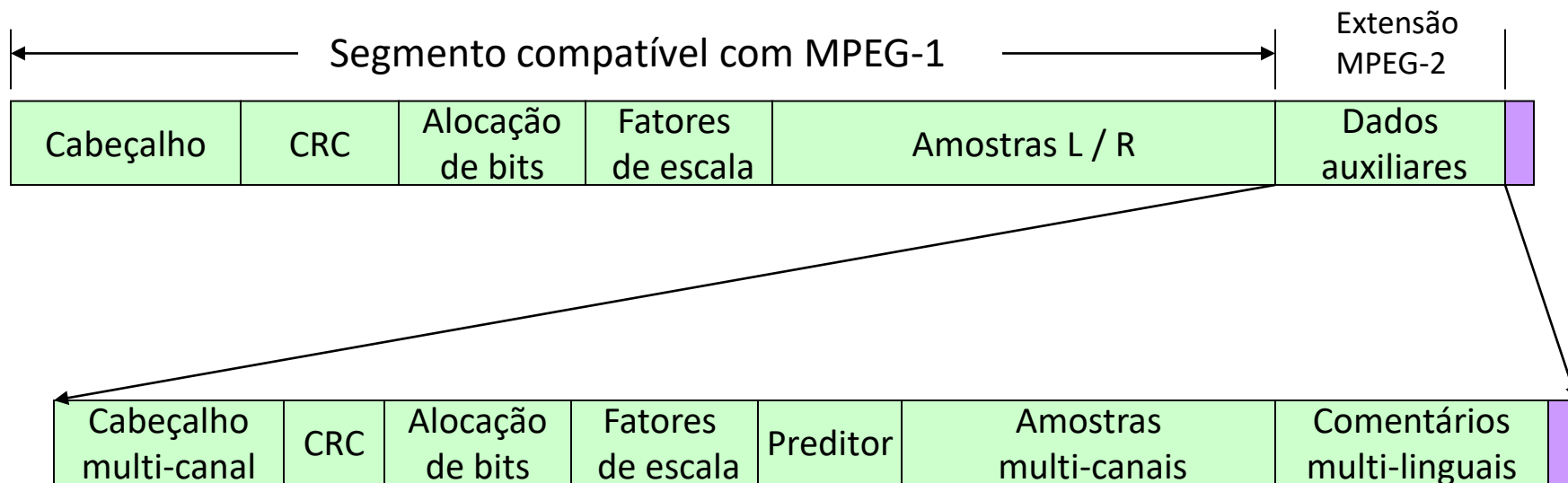


Blocos de Amostras (MPEG-1, Layer 2)





Amostras de até 2 quadros anteriores



Áudio Codificado (layer 3, $f_A = 48$ kHz)

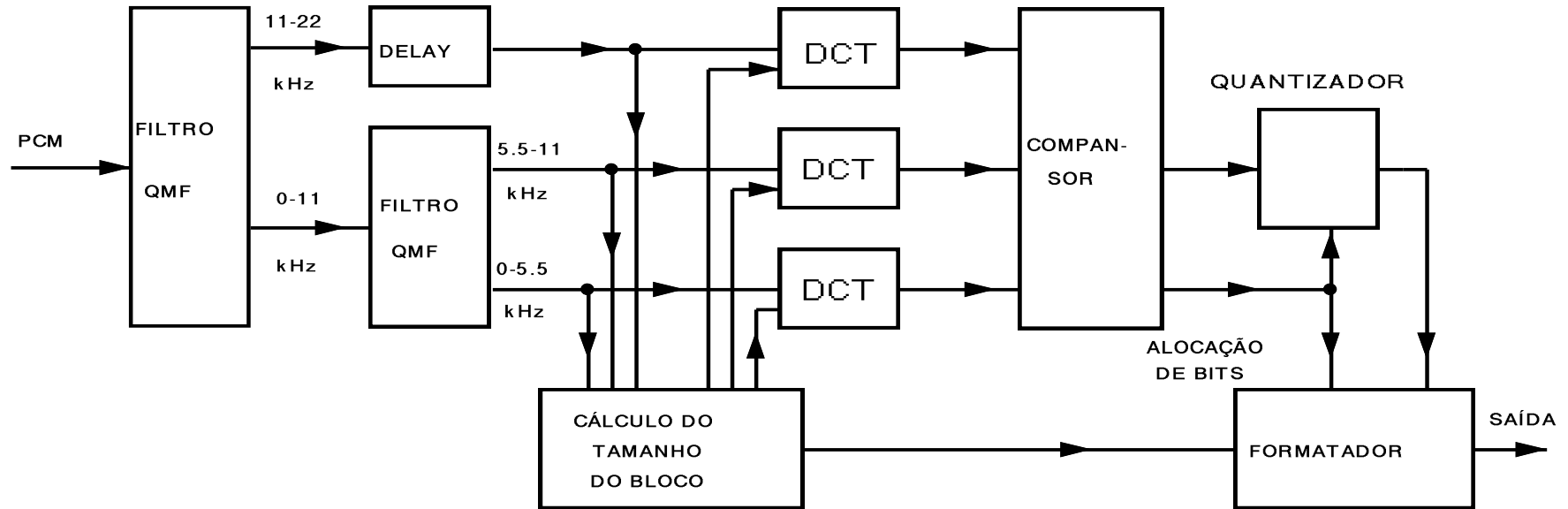
```
00004b80h: FF FB 80 64 EB 80 02 D2 38 5A 53 0F 2A D8 4F 24 ; üûdë.ò8ZS.*00$
00004b90h: 8B 4D 61 E5 4B 08 60 E1 79 AC 1C 4B 61 7E AB EF ; MaãK.`áy~.Ka~«i
00004ba0h: 28 F2 95 B7 07 44 4B B5 29 D2 21 34 E6 2A F5 B0 ; (ò.-.DKµ)ò!4æ*õ°
00004bb0h: 80 BA 65 04 55 0C 3C 5C B1 79 CA 07 CB 09 FD F6 ; ñe.U.;\±yĚ.Ě.ýö
00004bc0h: 53 F5 E6 D0 00 00 9C 96 D9 65 BE 36 66 18 EF BB ; Sðæð..üæ%6f.ï»
00004bd0h: E9 52 F1 D2 4C C9 EB 75 B4 66 03 0F 88 D6 7C B9 ; éRñÒLÉëu'f..ü|'
00004be0h: E5 E4 7D 16 61 FE 3F 8C 00 26 15 6E 10 13 59 52 ; ää}.ap?&.n..YR
00004bf0h: AE EC B4 31 8C EA FC 30 05 6A A2 84 32 7A 86 2F ; @ì'1üü0.jç2z/
00004c00h: 74 04 55 3B F1 30 3F B6 53 1F B9 D9 2A 92 20 10 ; t.U;ñ0?¶S.'ù*'.
00004c10h: 14 73 5D B6 D0 44 C1 B5 88 84 96 D6 33 F1 8D 77 ; .s]¶DÁµüü3ñw
00004c20h: 7D DB 48 2C 0F E0 DA F1 AB 5E 94 C1 D0 8A 57 51 ; }ÜH,.äüñ^üÁüWQ
00004c30h: 12 65 6F F0 10 50 05 3D CE 55 58 10 91 77 2E AE ; .eoð.P.=ÏUX.'w.0
00004c40h: DF 39 5A 9C A2 99 8C AE 60 82 56 EE C8 81 F6 79 ; 09ZñçüüüüVÏÈöy
00004c50h: 27 53 3E 8E EC 8E 8E 8E 10 3D D3 CB B1 20 29 63 ; 'S>üüüü.=óĚ±)c
00004c60h: 21 B3 21 14 72 8B BC 1C 4A 1B E9 80 44 00 04 E1 ; ?!'.r%J.éüD..ä
00004c70h: D1 2D 21 B4 04 8F 45 77 40 34 51 6A D4 04 A7 7A ; Ñ-!'.üEw@4Qjô.şz
00004c80h: 13 89 B3 C7 D1 E6 52 FB CD B5 56 4F 0F 52 C8 69 ; .ü³ÇÑæRÜíµV0.REi
00004c90h: B9 75 3A C0 94 BF 9B 71 C1 E8 87 85 1D 16 7D 0D ; 'u:Áü.üqáèüü..}.
00004ca0h: 93 3F 46 C7 78 FD 8F FF F6 23 95 02 6A 3B BC 53 ; ü?FÇxýüö#ü.j;%S
00004cb0h: FA 7F 1E 2E CC A8 70 01 00 0C 14 A4 64 CA E0 4E ; úü..î'p...sðĚàN
00004cc0h: 4E 75 3C 34 DC 5C 11 57 3C D2 E7 6A 2F 33 69 B0 ; Nu<4Ü\W<0çj/3i°
00004cd0h: B8 F2 88 F9 A3 96 90 0B 0B 13 AB DC 78 3E 72 50 ; ðüüĚüü...«Üx>rP
00004ce0h: B7 A2 40 57 71 7F 73 7F B1 46 B2 83 BA FF FB 82 ; -ç@WqüüüüF²üüüü
```

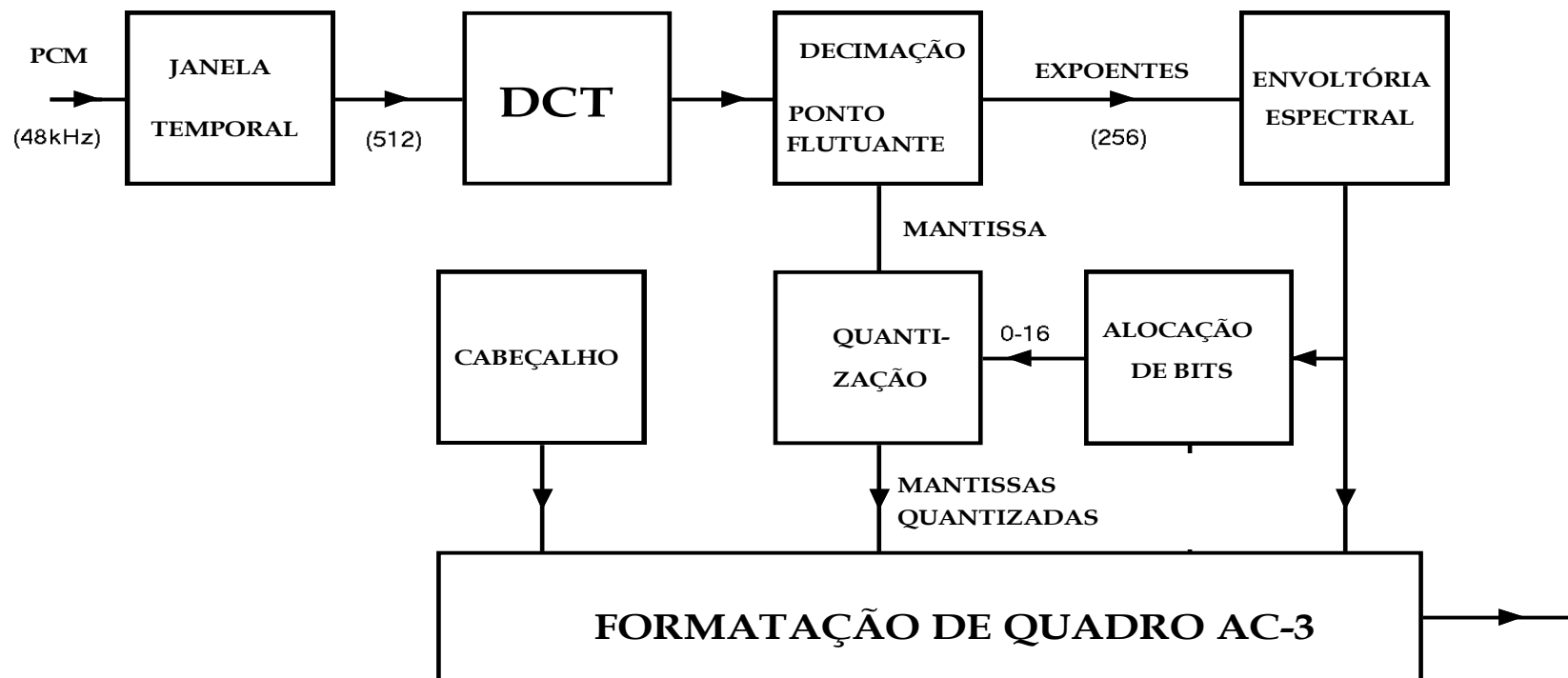
MPEG-1
112 kbps
333 bytes/frame

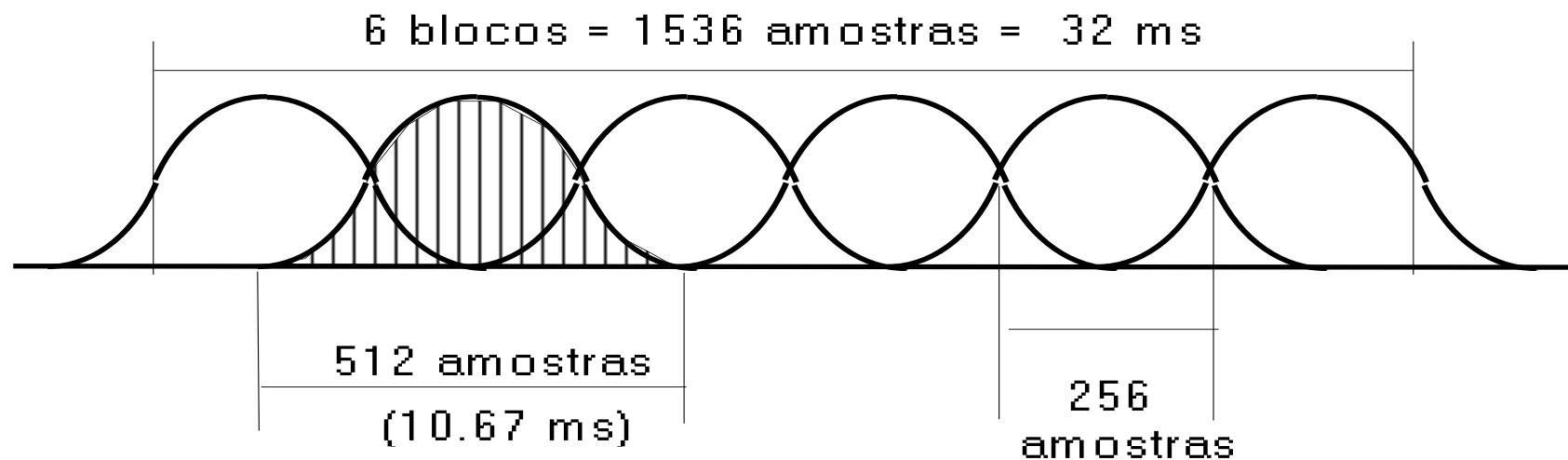
```
00008bf0h: FF F3 40 C4 CF 15 82 56 AE 06 36 26 68 8D C8 25 ; üó@ÄÏ.üV@.6&hüĚ%
00008c00h: 5A 94 D5 3A 4B 44 9E 14 28 04 CB 0B 15 48 F1 F6 ; Züü:KDü.(.Ě..Hñö
00008c10h: B7 D9 4B 09 83 1E D6 A4 D7 F7 D3 40 59 E0 D8 C9 ; -ÜK.ü.üüx÷ó@Yð0É
00008c20h: A2 2A FD DB B9 34 FF F4 3C 8D 7F FF 5C 7F 21 6A ; ç*ýÜ'4ýö<üüüü!j
00008c30h: EF FA E6 03 B5 DF 95 E1 03 D2 70 DA 3F 7B A0 00 ; ÿüæ.µ0üá.òpÜ?{.
00008c40h: 80 C3 29 DA 48 72 F3 FF 0F B6 05 34 B7 10 72 25 ; üü)ÜHröü.¶.4-.r%
00008c50h: 0E C3 10 5B 60 35 43 E9 FF F3 40 C4 D4 17 42 4E ; .Ä.[`5Céüó@ÄÜ.BN
```

MPEG-2
32 kbps
104 bytes/frame

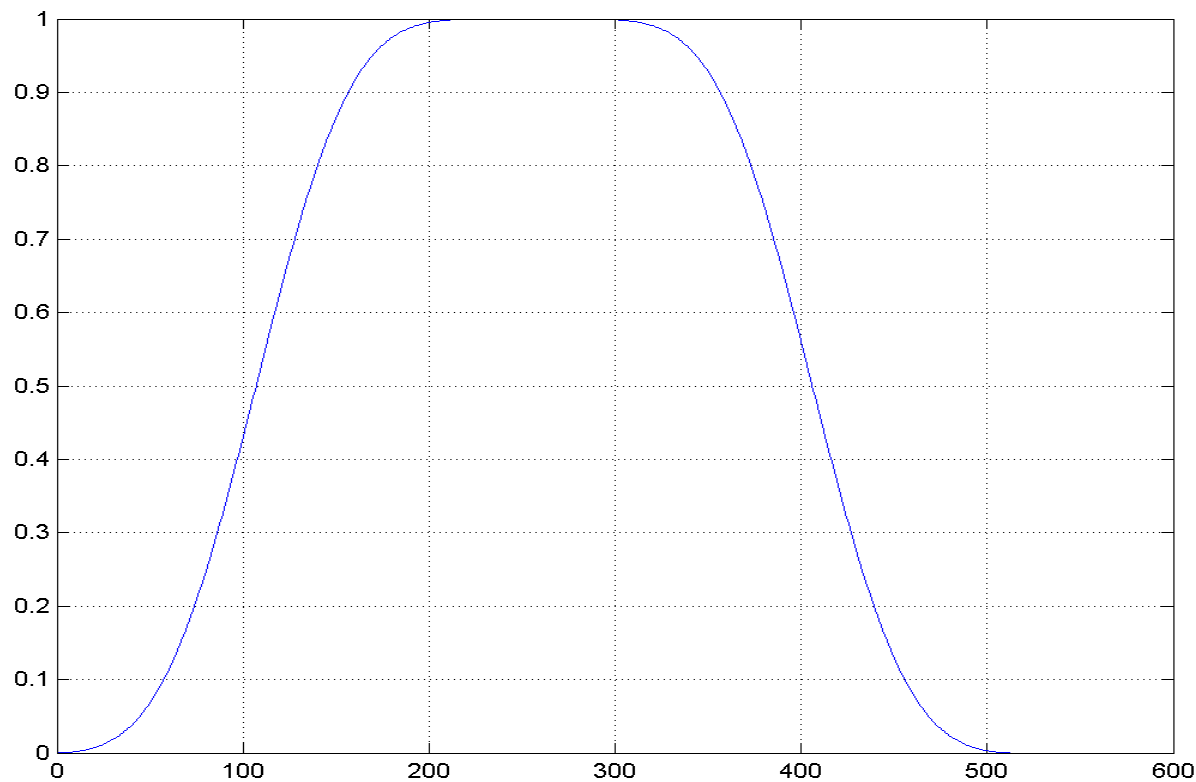
SONY "ATRAC" (*Adaptive Transform Acoustic Coder*)







$$X(k) = \frac{-1}{256} \sum_{n=0}^{511} x(n) J(n) \cos\left(\frac{2\pi}{2048} (2n+1)(2k+1) + \frac{\pi}{4} (2k+1)\right)$$



Complete Main Audio Service (CM) : Diálogo, Música e Efeitos ; 1 a 5.1 Canais, podendo ser codificado em múltiplas línguas;

Main Audio Service, Music and Effects (ME) : Música e Efeitos sem Diálogo, podendo ser combinado com áudio proveniente do Serviço D;

Visually Impaired (VI) : Descrição narrativa da informação visual, destinada a atender deficientes visuais; 1 Canal de áudio que pode ser combinado com o Serviço CM;

Hearing Impaired (HI) : Apenas canal de Diálogo, processado p/ maior inteligibilidade, destinado a deficientes auditivos;

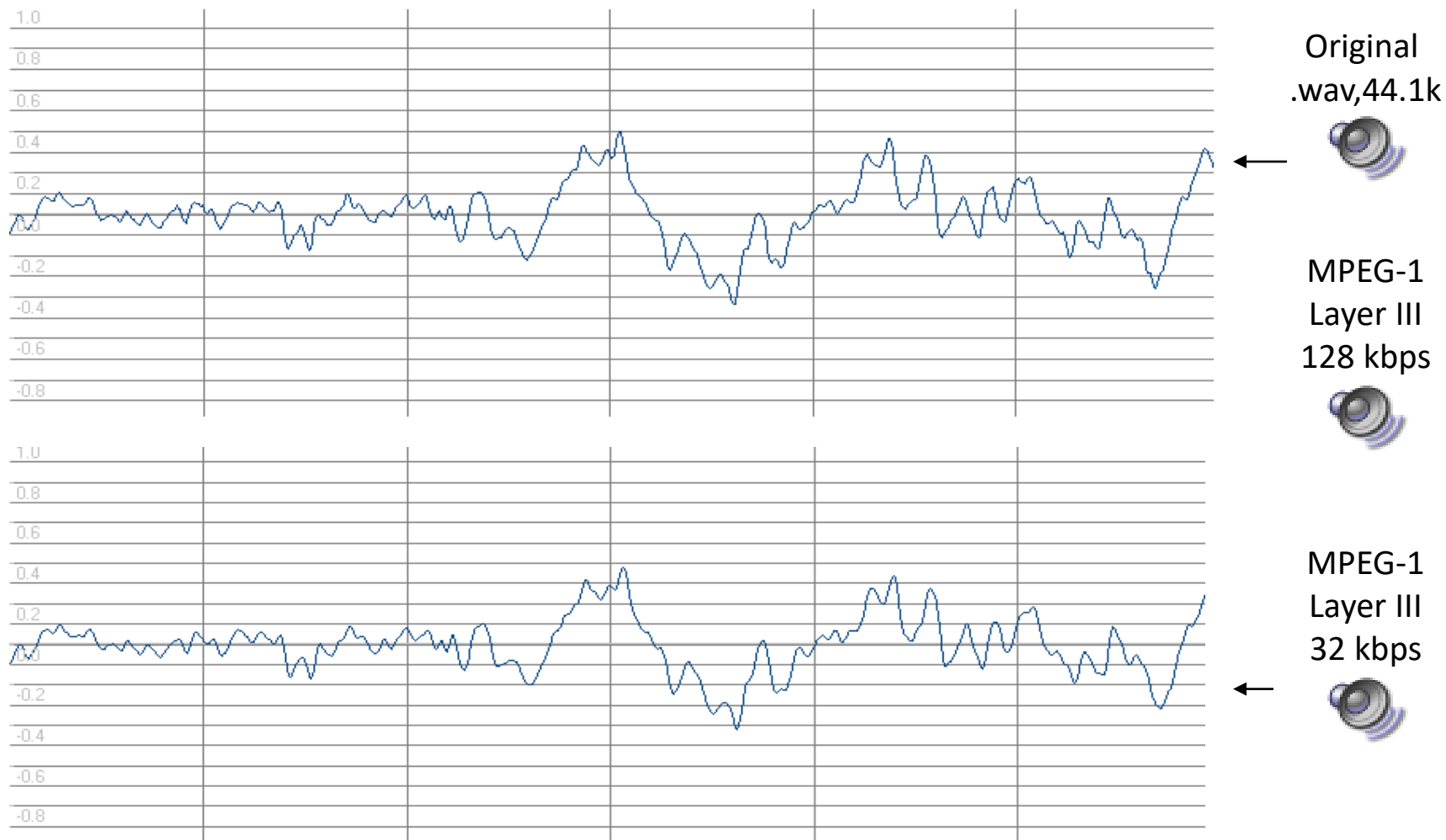
Dialogue (D) : Diálogo a ser combinado com áudio proveniente do Serviço ME; pode ser fornecido em várias línguas e codificado em 1 canal (Central) ou 2 canais;

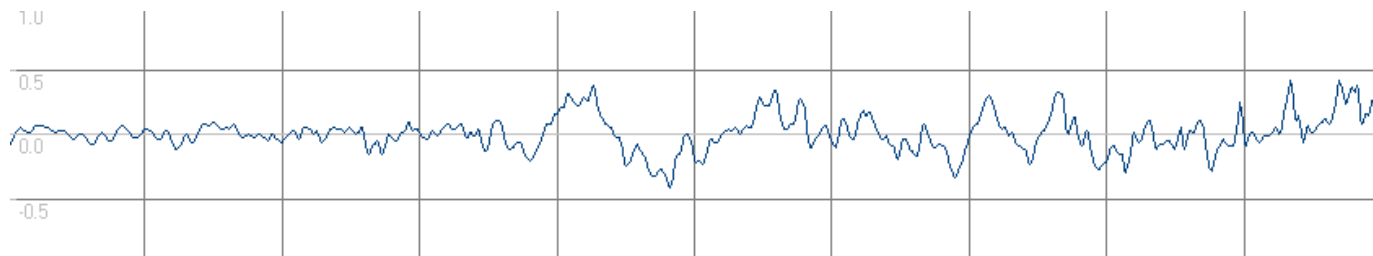
Commentary (C) : Comentários opcionais ao Programa Visual; pode ser fornecido como 1 canal a ser combinado ao serviço ME;

Emergency (E) : Canal de maior prioridade, admite inserção de mensagens de emergência, sobrepondo-se aos outros serviços;

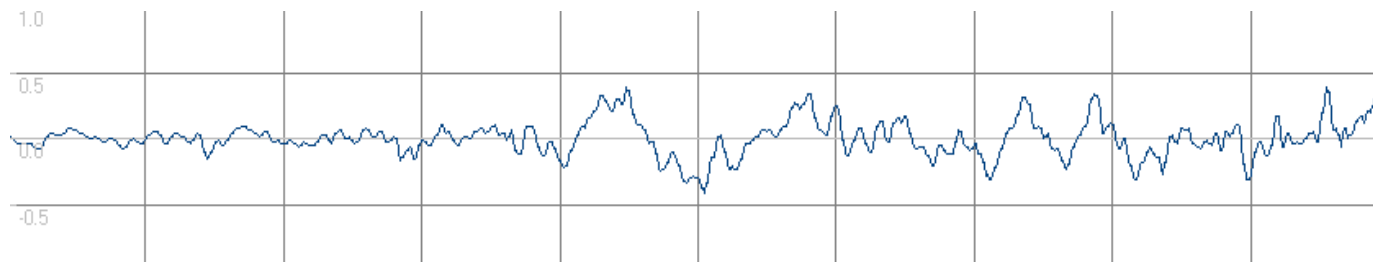
Voice-Over (VO) : 1 canal de áudio, que permite inserções de voz a serem superpostas a um serviço já codificado, sem a necessidade de que o áudio original seja decodificado , misturado com a inserção e re-codificado novamente.

Padrão de Compressão	Canais	Taxa Média de Compressão	Taxa Original	Taxa Comprimida	Atraso A/D/A
MPEG-1 / I	2	4:1	1,41 Mb/s	384 kb/s	19 ms
ATRAC	2	5:1	1,41 Mb/s	292 kb/s	
MPEG-1 / II (MUSICAM)	2	6:1	1,54 Mb/s	256 kb/s	>40 ms
MPEG-1 / III	2	8:1	1,41 Mb/s	128 kb/s	>80 ms
MPEG-2 / III	5.1	10:1	3,53 Mb/s	384 kb/s	
Dolby AC-3	5.1	12:1	3,53 Mb/s	290 kb/s	45 ms

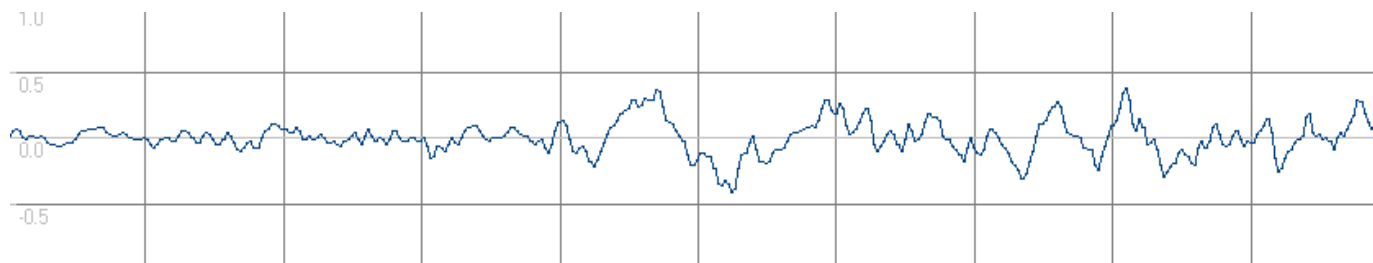




MP3
48 kbps
24 kHz



MP3
16 kbps
12 kHz



MP3
8 kbps
8 kHz

