

# Capitolo 4

## Compressione sorgenti dinamiche

- Audio

- DPCM
- APCM e ADPCM
- LPC
- CELP
- MPEG-1
- MPEG-2
- Dolby

- Video

- MJPEG
- Frame I,P,B
- Stima moto e compensazione
- H.261
- H.263
- MPEG
- Prestazioni
- Codificatori avanzati
  - H.264
  - FGS

# Audio

- **LOSSY**
- 3 categorie di codifiche
  - **Dominio del tempo**
  - **Dominio delle frequenze**
  - **Codifiche per modelli**
- Prestazioni
  - MOS
- **Riferimenti**
  - **G.711**, PCM a 64Kbps → **MOS=4.1**
  - **G.711u** e **G.711a**, PCM enfasi u-law e A-law → **MOS=4.4**

# Audio - DPCM

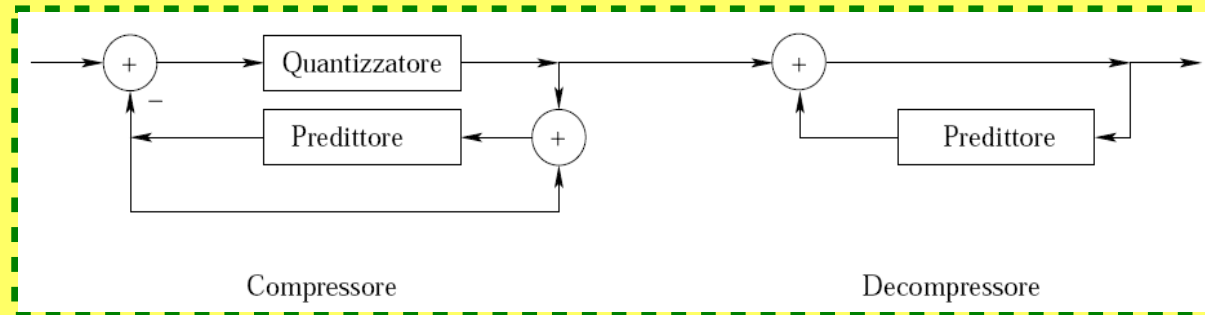
- Differential Pulse Code Modulation
- **Dominio tempo**
- Proprietà segnali audio: **range delle differenze di ampiezza tra campioni successivi è inferiore al range di campioni distinti**
  - Concetto ampliato con **p** campioni

$$\hat{x}_n = \sum_{k=1}^p \alpha_k x_{n-k}$$

$$e_n = x_n - \hat{x}_n$$

$$E_p = \frac{1}{N} \sum_{n=1}^N (x_n - \hat{x}_n)^2$$

$$\{\alpha_k\} = \operatorname{argmin} E_p$$



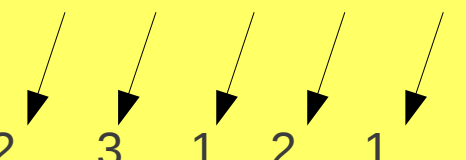
$$\begin{aligned} E_p &= E[(x_n - \hat{x}_n)^2] \\ &= E[x_n^2] - 2 \sum_{k=1}^p \alpha_k E[x_n x_{n-k}] + \sum_{k=1}^p \sum_{h=1}^p \alpha_k \alpha_h E[x_{n-k} x_{n-h}] \\ &= \phi(0) - 2 \sum_{k=1}^p \alpha_k \phi(k) + \sum_{k=1}^p \sum_{h=1}^p \alpha_k \alpha_h \phi(k-h) \end{aligned}$$

$$\hat{\tilde{x}}_n = \sum_{k=1}^p \alpha_k \tilde{x}_{n-k}$$

$$\tilde{e}_n = x_n - \hat{\tilde{x}}_n$$

# Audio - DPCM

- Concetto ampliato con  $p$  campioni
- Esempio con  $p=2$

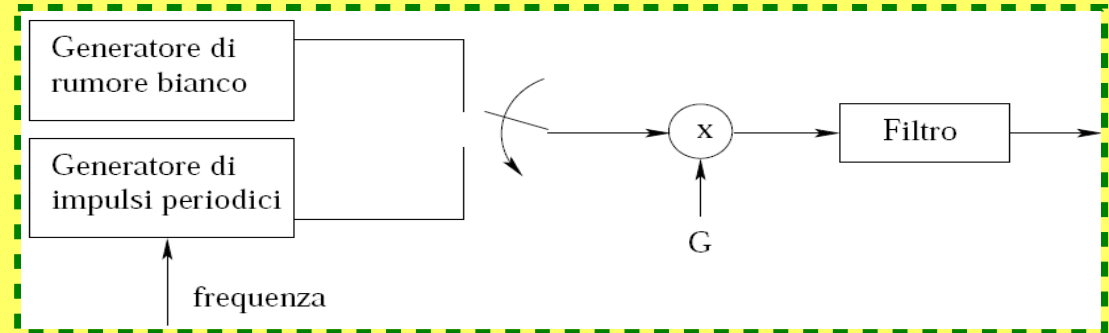
- 16 18 21 20 18 19
  - 2 3 1 2 1
- 
- |    |    |    |    |    |    |
|----|----|----|----|----|----|
| 16 | 18 | 21 | 20 | 18 | 19 |
| ↓  | ↓  | ↓  | ↓  | ↓  |    |
| 2  | 3  | 1  | 2  | 1  |    |

# Audio – APCM e ADPCM

- PCM e DPCM assumono segnale stazionario, ipotesi non reale
- **APCM**
  - Quantizzazione adattativa, con passo variabile dipendente dalla varianza del segnale
- **DPCM**
  - Variare i coefficienti del predittore sulla base delle caratteristiche attuali del segnale
  - Overhead per la trasmissione dei coefficienti
- **ADPCM**
  - Forzare il decompressore a ricalcolare i coefficienti sulla base dei campioni ricostruiti, evitando al codificatore di doverli trasmettere
- **G.721** (ADPCM), banda 32Kbps, **MOS=4.0**
- **G.726** (ADPCM), 16, 24, 32, 64 Kbps, **MOS=4.1** con 32 Kbps

# Audio – LPC

- **Linear Predictive Coding**
- Codificatore per modello: modello in grado di simulare l'**APPARATO VOCALE, TRASDUTTORE** del suono
  - Suoni sordi, non vocalizzati
  - Suoni sonori, vocalizzati
- **Trasmissione del segnale** vocale avviene indicando i **parametri del modello**:
  - Tipo di sorgente (1 bit)
  - Eventuale frequenza, per suoni vocalizzati (6 bit)
  - Coefficienti del filtro, per tratto vocale (da 6 a 10 bit)
  - Guadagno G, per intensità (5 bit)
- **LPC10, 2.4Kbps, MOS=2.3**

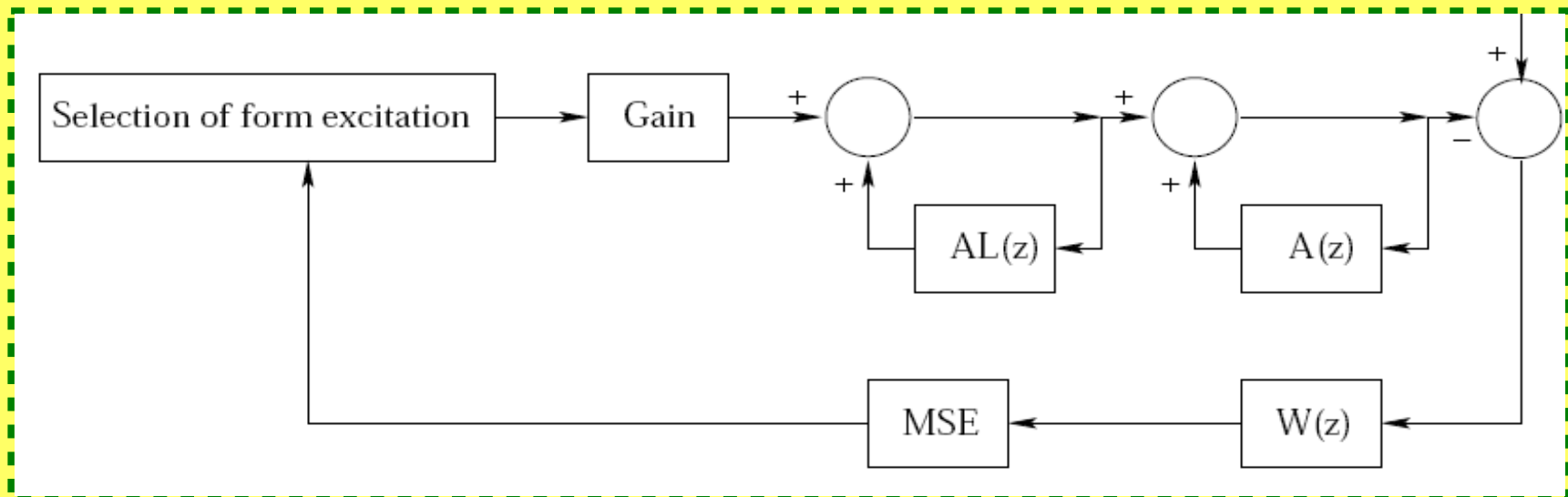


$$H(z) = \frac{G}{1 - \sum_{k=1}^p \alpha_k z^{-k}}$$

$$x_n = \sum_{k=1}^p \alpha_k x_{n-k} + Gv_n = \tilde{x}_n + Gv_n$$

# Audio – CELP

- **Code Excited Linear Predictive**
- Codificatore per modello (**TRASDUTTORE** del suono)
  - Modello sorgente più complesso di LPC
  - Funzioni di eccitazione fisse e note e funzioni adattative
  - Closed-loop per la selezione della forma d'onda di eccitazione
  - Quantizzazione vettoriale
- **G.728, 16Kbps, MOS=4.0, ritardo=0.765ms**
- **G.729, 8Kbps, MOS=3.7, ritardo=10ms**
- **G.723.1, 6.3Kbps, MOS=3.9, ritardo=30ms**



# Audio – MPEG-1

- Codificatore **PERCETTIVO**
  - Utilizza congiuntamente tre tipi di codifica: per modello, dominio del tempo e dominio della frequenza
  - modello del **SENSORE**, modello di percezione dell'orecchio umano
- Le caratteristiche dell'orecchio umano sono definite nel

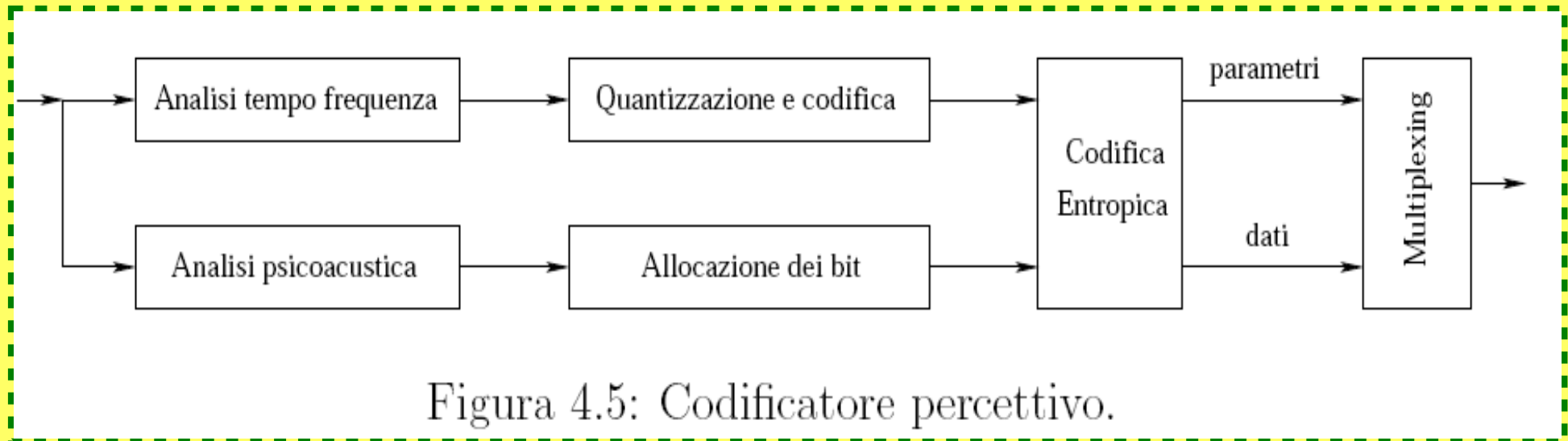
## MODELLO PSICOACUSTICO

- **Definisce il funzionamento della percezione dei suoni nell'orecchio umano.**
- Sulla base di tali considerazioni si individuano dei meccanismi che consentono di eliminare dal segnale originale quelle info che non verrebbero udite e si effettua una compressione (**LOSSY**)
  - ♦ Componenti sotto soglia di udibilità
  - ♦ Mascheramenti frequenziali e temporali



# Audio – MPEG-1

## schema di un codificatore percettivo



# Audio – MPEG-1 *modello psicoacustico*

- **Sound Pressure Level (SPL)**

- Intensità di pressione sonora ( $p$ ) in dB relativamente ad uno stimolo convenzionale ( $p_0=2*10^{-5}N/m^2$ )

$$SPL = 20 \log \frac{p}{p_0}$$

- **Soglia di Udibilità**

- Ammontare di energia di un tono sinusoidale ascoltabile da un orecchio umano in ambiente rumoroso
- Suoni con SPL superiore alla soglia di udibilità vengono uditi dall'orecchio umano, suoni con SPL inferiore alla soglia di udibilità non vengono uditi
- Suoni con SPL inferiori alla soglia di udibilità possono essere eliminati da un segnale, senza degradare la percezione sonora da parte di un orecchio umano
- Elaborata da **Fletcher** con una campagna di misure

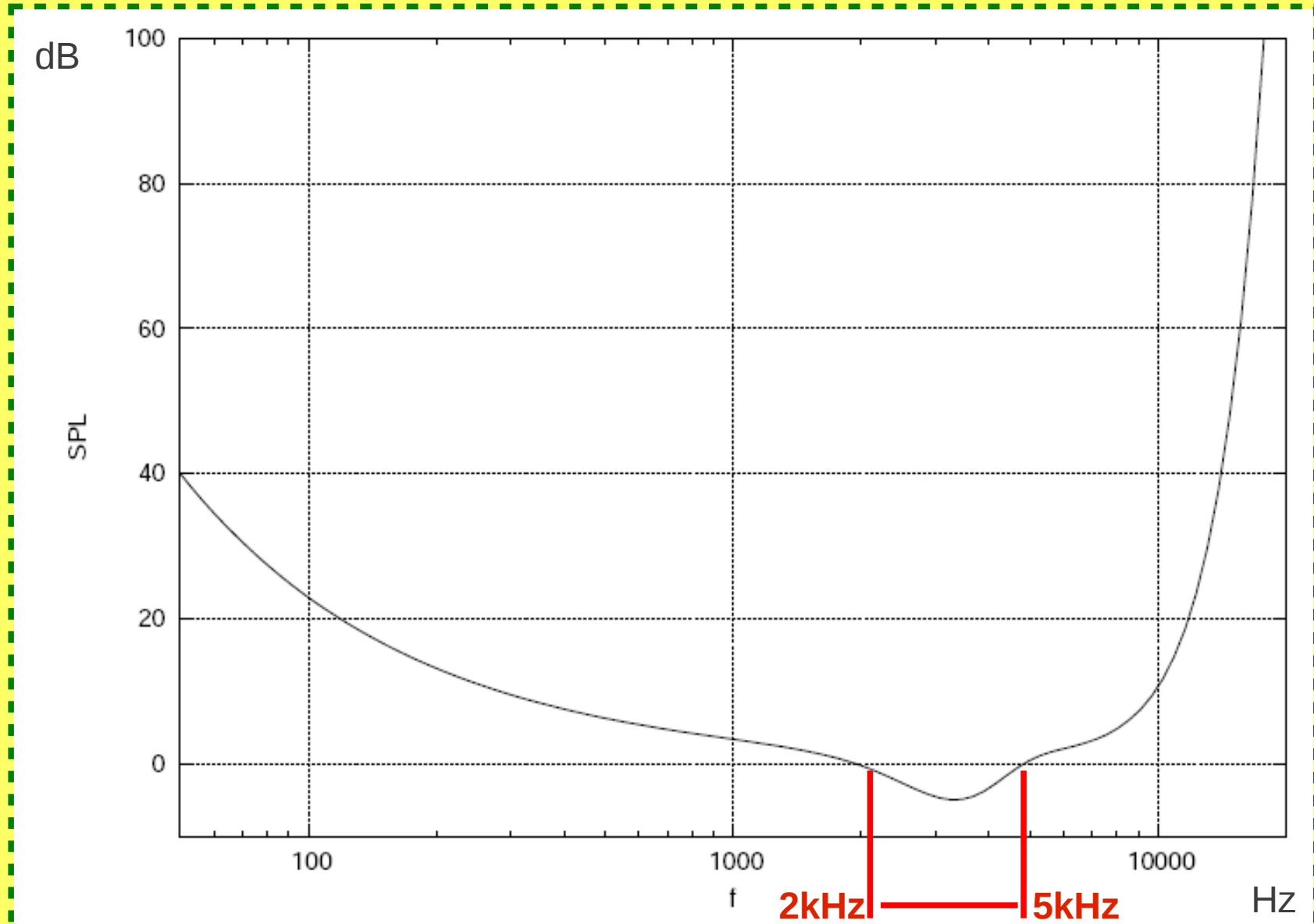
- Andamento  
non lineare

$$T(f) = 3.64 \left( \frac{f}{1000} \right)^{-0.8} - 6.5e^{-0.6\left(\frac{f}{1000}-3.3\right)^2} + 10^{-3} \left( \frac{f}{1000} \right)^4$$

- Orecchio umano ha maggiore sensibilità nel range **[2kHz-5kHz]**, ossia occorrono minori livelli di pressione sonora affinché suoni di frequenze appartenenti a tale range possano essere uditi dall'orecchio umano

# Audio – MPEG-1 *modello psicoacustico*

SOGLIA DI UDIBILITA'



# Audio – MPEG-1

## *modello psicoacustico*

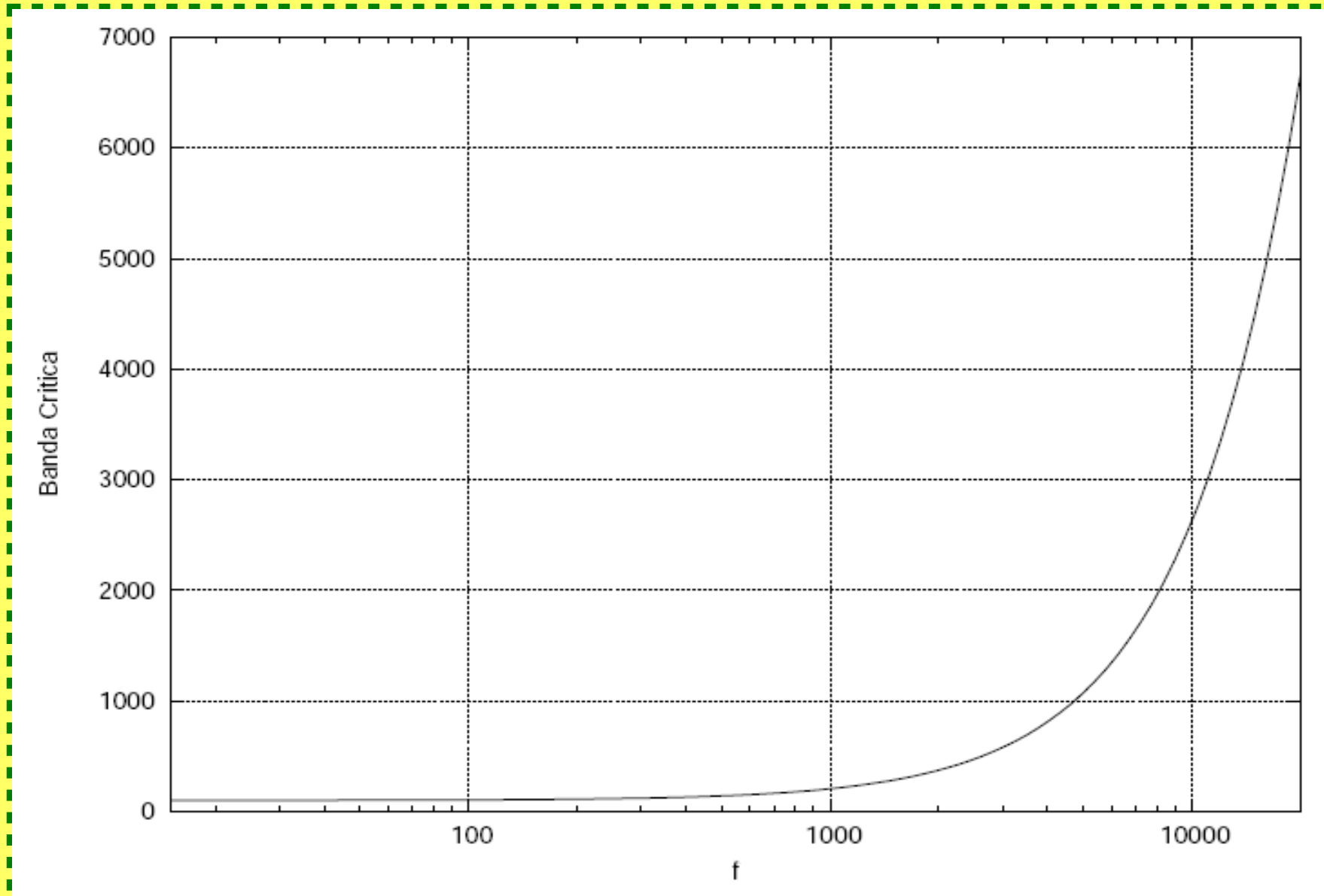
- **Banda Critica**

- Definisce come l'orecchio percepisce i suoni alle varie frequenze
- L'orecchio realizza un'analisi spettrale comportandosi come un insieme di **filtri** sovrapposti, con **banda variabile in funzione della frequenza**
- La banda di ogni filtro è chiamata banda critica
- Banda in cui due suoni a frequenze diverse vengono percepiti dall'orecchio come identici
- ***Orecchio è più sensibile alle variazioni in basse frequenze***, dove bastano alcune decine di Hz di differenza per far percepire due suoni come diversi, mentre alle alte frequenze occorrono migliaia di Hz di differenza per percepire differenze nei suoni
- Per  **$f < 500\text{Hz}$**  la banda critica è costante e pari a circa **100Hz**
- Per  **$f > 500\text{Hz}$**  si può approssimare come  **$f/5$**

# Audio – MPEG-1

## *modello psicoacustico*

Banda critica

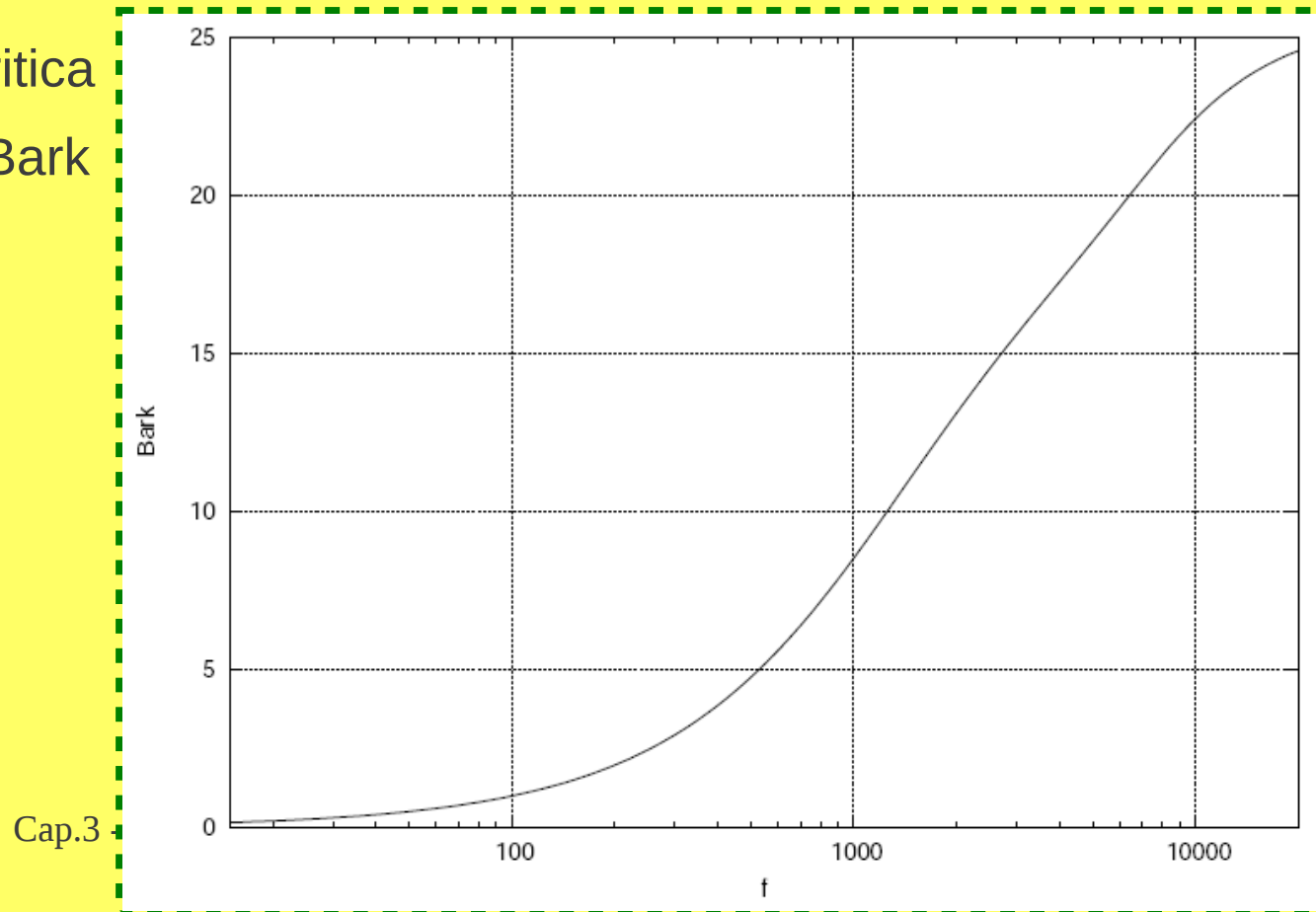


# Audio – MPEG-1

## *modello psicoacustico*

- **Bande di Bark**

- Volendo trattare l'orecchio umano come un insieme discreto e finito di filtri, si definisce la spaziatura tra tali filtri e quindi la successione delle bande.
- Tale spaziatura è definita come spaziatura di Bark
- Unità di misura
- 1 Bark = una banda critica
- Totale di 26 bande di Bark



# Audio – MPEG-1

## *modello psicoacustico*

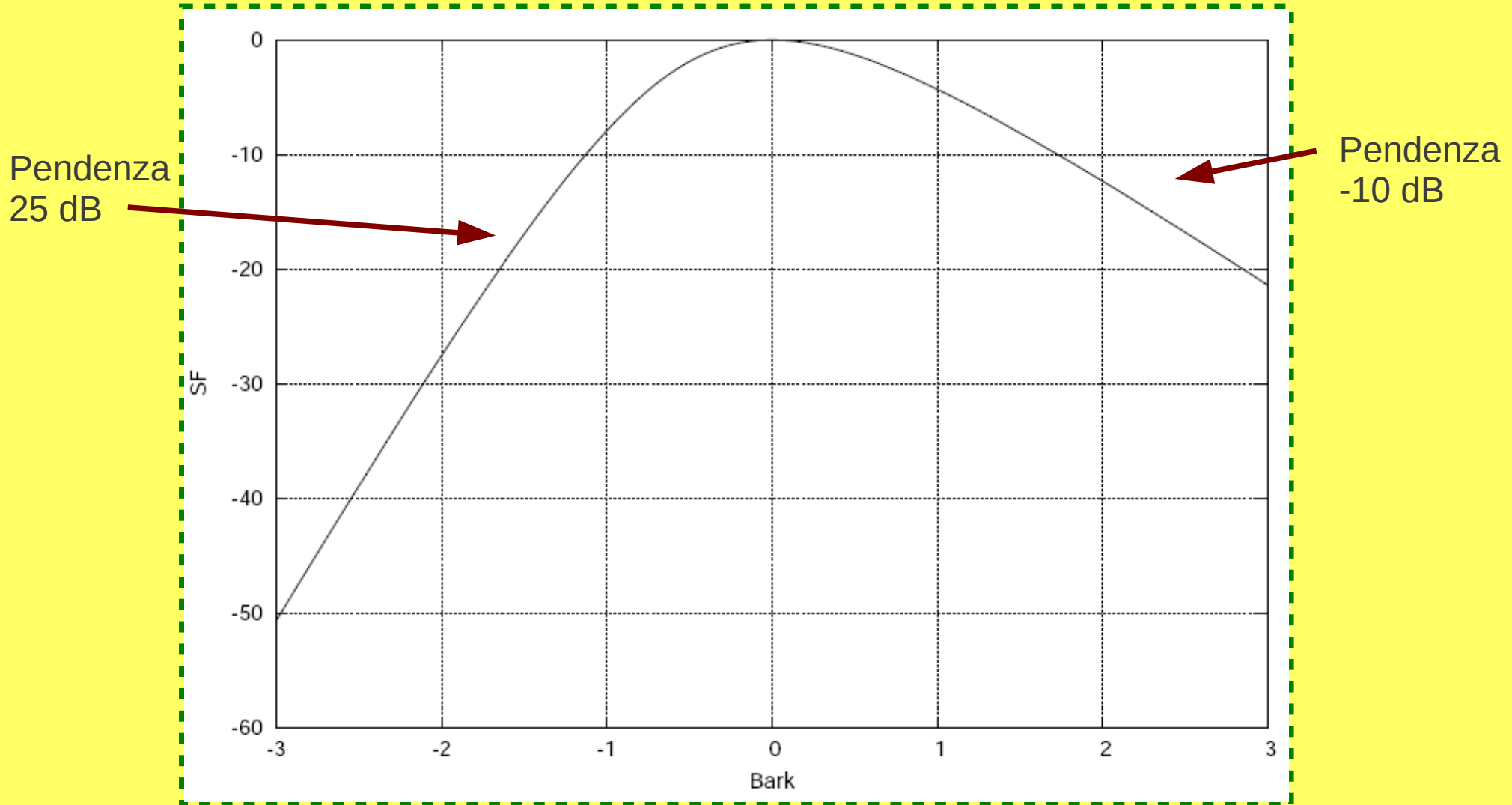
- **Mascheramenti**

- Fenomeno per cui la presenza di un suono ne rende un altro non udibile
- Una sollecitazione forte crea uno stress sulla membrana uditiva, relativamente alla banda critica, tale da non consentire l'ascolto di segnali più deboli, nell'ambito della stessa banda.
- Suoni mascherati possono essere eliminati dal segnale originale senza degradarne la percezione sonora
- Tipologie di mascheramento: frequenziale e temporale
- Frequenziale
  - Noise-Masking-Tone (NMT)
  - Tone-Masking-Noise (TMN)
  - Noise-Masking-Noise (NMN)
  - Il Signal-to-Mask Ratio (SMR) misura il rapporto tra mascherante e mascherato e definisce il limite per cui un suono è effettivamente mascherato
  - Allargamento del mascheramento

# Audio – MPEG-1

## *modello psicoacustico*

Mascheramenti frequenziali: allargamento del mascheramento

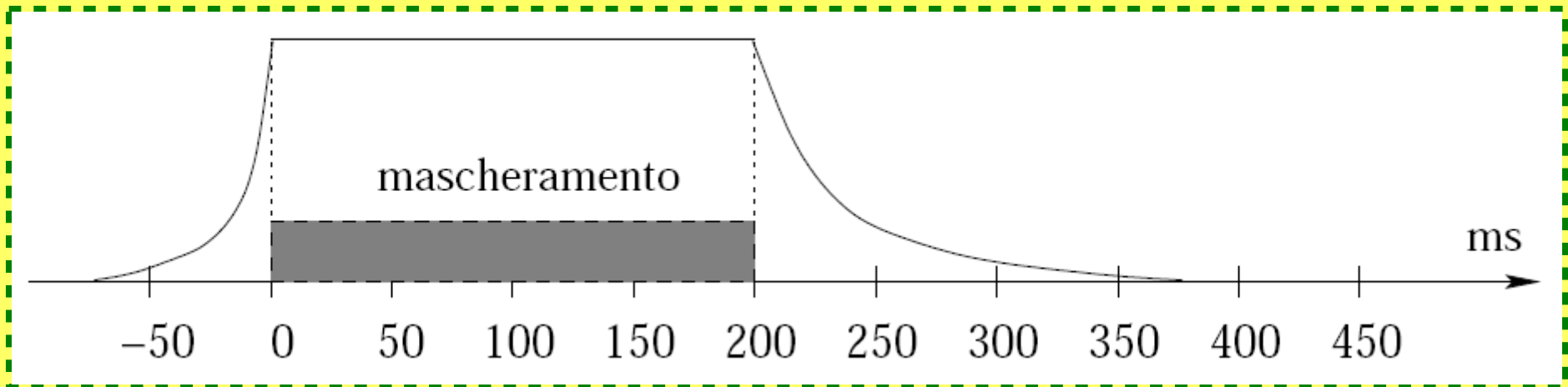




# Audio – MPEG-1

## *modello psicoacustico*

### Mascheramento temporale



# Audio – MPEG-1 *algoritmo*

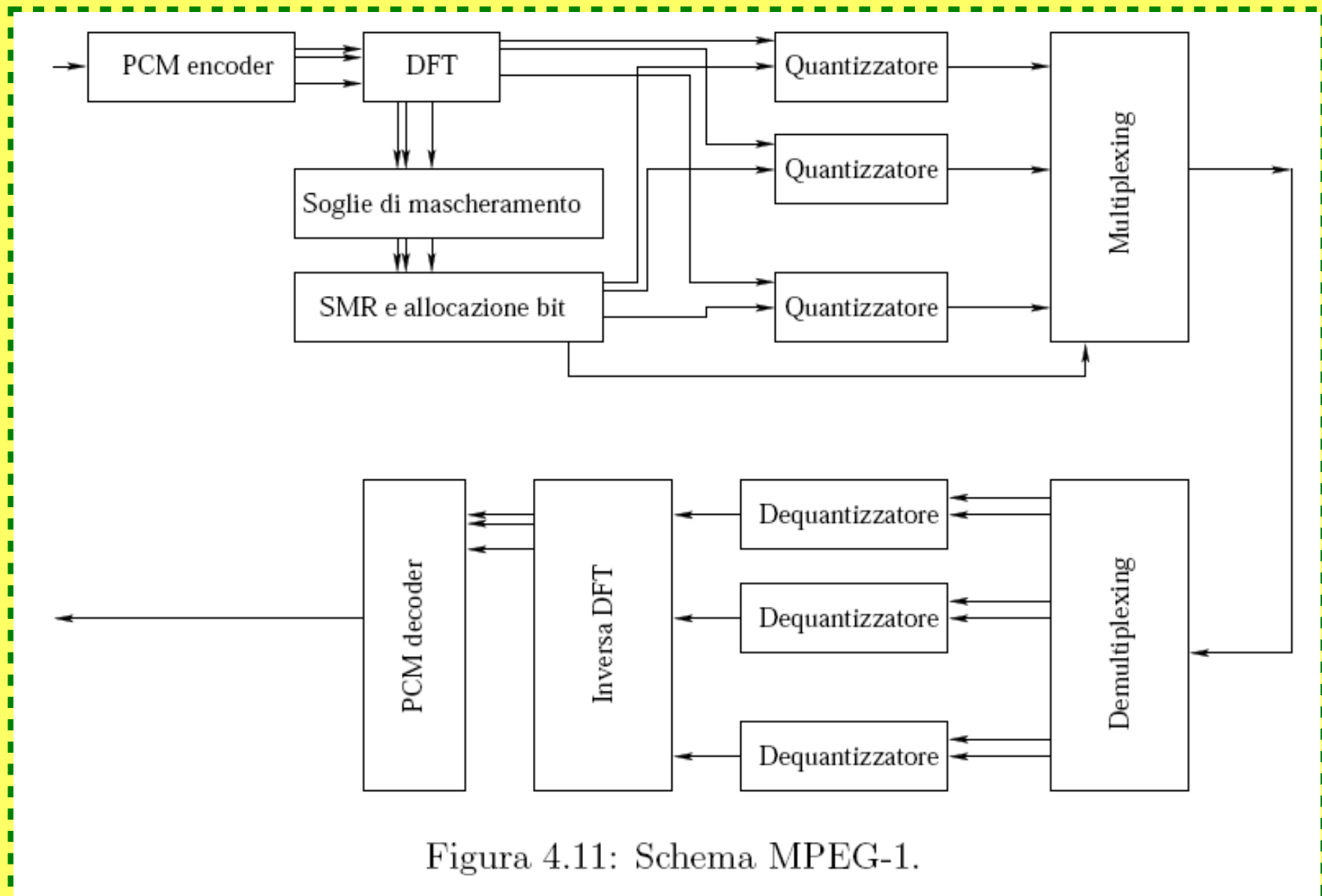


Figura 4.11: Schema MPEG-1.

# Audio – MPEG-1

## *algoritmo*

- **Campionamento e quantizzazione** del segnale in ingresso (schema **PCM**)
- **Filtraggio digitale** su un banco di filtri equispaziato, caratterizzato da 32 bande. (Se  $f_c$  è la frequenza di campionamento, il sistema è in grado di trattare un segnale con  $f_M = f_c/2$  e la banda di ogni filtro  $B = f_c/64$ ).
  - *Si osserva che queste 32 bande a dimensione fissata non rispettano le 26 bande di Bark a dimensione variabili → semplicità computazionale*
- Il banco dei filtri opera direttamente su 32 campioni del codificatore PCM e restituisce 32 campioni in uscita che rappresentano i valori trasformati secondo **Fourier**. Ognuno dei campioni di uscita è relativo ad una diversa banda.
- I campioni vengono aggregati in blocchi di 12 o 36 campioni, realizzando un frame di  $384=12*32$  o  $1152=36*32$  campioni.
- L'uscita del banco dei filtri alimenta sia il modello psicoacustico che un banco di quantizzatori.
- Il **modello psicoacustico** definisce il numero dei bit necessari per ogni sottobanda e quindi definisce il livello di compressione istantaneo. Per il modello psicoacustico si seguono le operazioni descritte nelle slides successive.

# Audio – MPEG-1

## *algoritmo modello psicoacustico*

- **PASSO 1)**

- **Calcolo delle SPL e analisi FFT**

$$P(k) = PN + 10 \log \left[ \sum_{n=0}^{N-1} w_n x_n e^{-j \frac{2\pi kn}{N}} \right]^2$$

- FFT su N punti
- P(k) espressa in dB
- PN=90.302 dB
- $w_n$  finestra di Hanning

$$\frac{1}{2} \left[ 1 - \cos \frac{2\pi n}{N-1} \right]$$

# Audio – MPEG-1

## *algoritmo modello psicoacustico*

- PASSO 2)

- Calcolo delle componenti tonali e di rumore

- **Tonali:** componenti dello spettro che rappresentano un massimo locale in grado di superare le componenti vicine (rispetto ad una certa distanza di Bark) di almeno 7dB.

$$S_T = \{P(k) | P(k) > P(k \pm 1), P(k) > P(k \pm \Delta_k) + 7dB\}$$

$$\Delta_k \in \begin{cases} 2 & 2 < k < 63 & f \in [0.17, 5.5] \text{KHz} \\ [2,3] & 63 \leq k < 127 & f \in [5.5, 11] \text{KHz} \\ [2,6] & 127 \leq k \leq 256 & f \in [11, 20] \text{KHz} \end{cases}$$

- Per ogni elemento di  $S_T$  si calcola **energia** sulla base dei due adiacenti.

$$P_{TM}(k) = 10 \log \sum_{j=-1}^1 10^{0.1P(k+j)}$$

- **Energia di rumore:** valutata con le componenti che non sono tonali

$$P_{NM}(\beta) = 10 \log \sum_{P(j) \notin \{P_{TM}(k), P_{TM}(k \pm 1), P_{TM}(k \pm \Delta_k)\}} 10^{0.1P(j)}$$

# Audio – MPEG-1

## *algoritmo modello psicoacustico*

- **PASSO 3)**
  - **Decimazione e riorganizzazione dei marker, seguendo due criteri**
  - Primo criterio: eliminare i marker tonali o di rumore che sono inferiori alla soglia di udibilità.
  - Secondo criterio: finestra mobile di 0.5 Bark, utilizzata per sostituire ogni coppia di marker a distanza inferiore a 0.5 Bark con il più forte dei due.

# Audio – MPEG-1

## *algoritmo modello psicoacustico*

- **PASSO 4)**
  - **Calcolo delle soglie di mascheramento**

$$T_{NM}(i, j) = P_{NM}(j) - 0.175z(j) + SF_{NM}(i, j) - 2.025$$

$$T_{NM}(i, j) = P_{NM}(j) - 0.175z(j) + SF_{NM}(i, j) - 2.025$$

$$SF_{TM}(i, j) = \begin{cases} 17\Delta_z - 0.4P_{TM}(j) + 11 & -3 \leq \Delta_z < -1 \\ (0.4P_{TM}(j) + 6)\Delta_z & -1 \leq \Delta_z < 0 \\ -17\Delta_z & 0 \leq \Delta_z < 1 \\ (0.15P_{TM}(j) - 17)\Delta_z - 0.15P_{TM}(j) & 1 \leq \Delta_z < 8 \end{cases}$$

# Audio – MPEG-1

## *algoritmo modello psicoacustico*

- **PASSO 5)**

- **Calcolo della soglia complessiva di mascheramento**

$$T_g(i) = 10 \log \left[ 10^{0.1T(i)} + \sum_{j=1}^{nT} 10^{0.1T_{TM}(i,j)} + \sum_{j=1}^{nM} 10^{0.1T_{NM}(i,j)} \right]$$

- **PASSO 6)**

- **Scaling dei valori ed allocazione dei bit**
- Rescaling del picco in modo da poterlo rappresentare con un numero inferiore di bit
- Si fissa una bit rate target e per ogni blocco di 12 campioni si allocano i bit in modo da ottimizzare la resa
- Il numero di bit decisi per quantizzare ogni singola uscita dei 32 filtri deve essere specificato al decodificatore



# Audio – MPEG-1

## *tre tipi di codificatori*

- **layer 1 (MP1)**
  - Cassette
  - Blocchi di 12 campioni, FFT 512 punti, fino a 448 Kbps
  - Modello psicoacustico senza mascheramento temporale
- **layer 2 (MP2)**
  - Applicazioni broadcast
  - Blocchi di 36 campioni, FFT a 1024 punti, fino 384 Kbps
  - Modello psicoacustico con mascheramento temporale
- **layer 3 (MP3)**
  - Musica con qualità analoga a CD ma bassa bit rate
  - Blocchi di 36 campioni, FFT a 1024 punti, da 8 a 320 Kbps
  - Utilizza una compressione entropica Huffman e sfrutta ridondanza segnale stereo

# Audio – MPEG-1 formato file

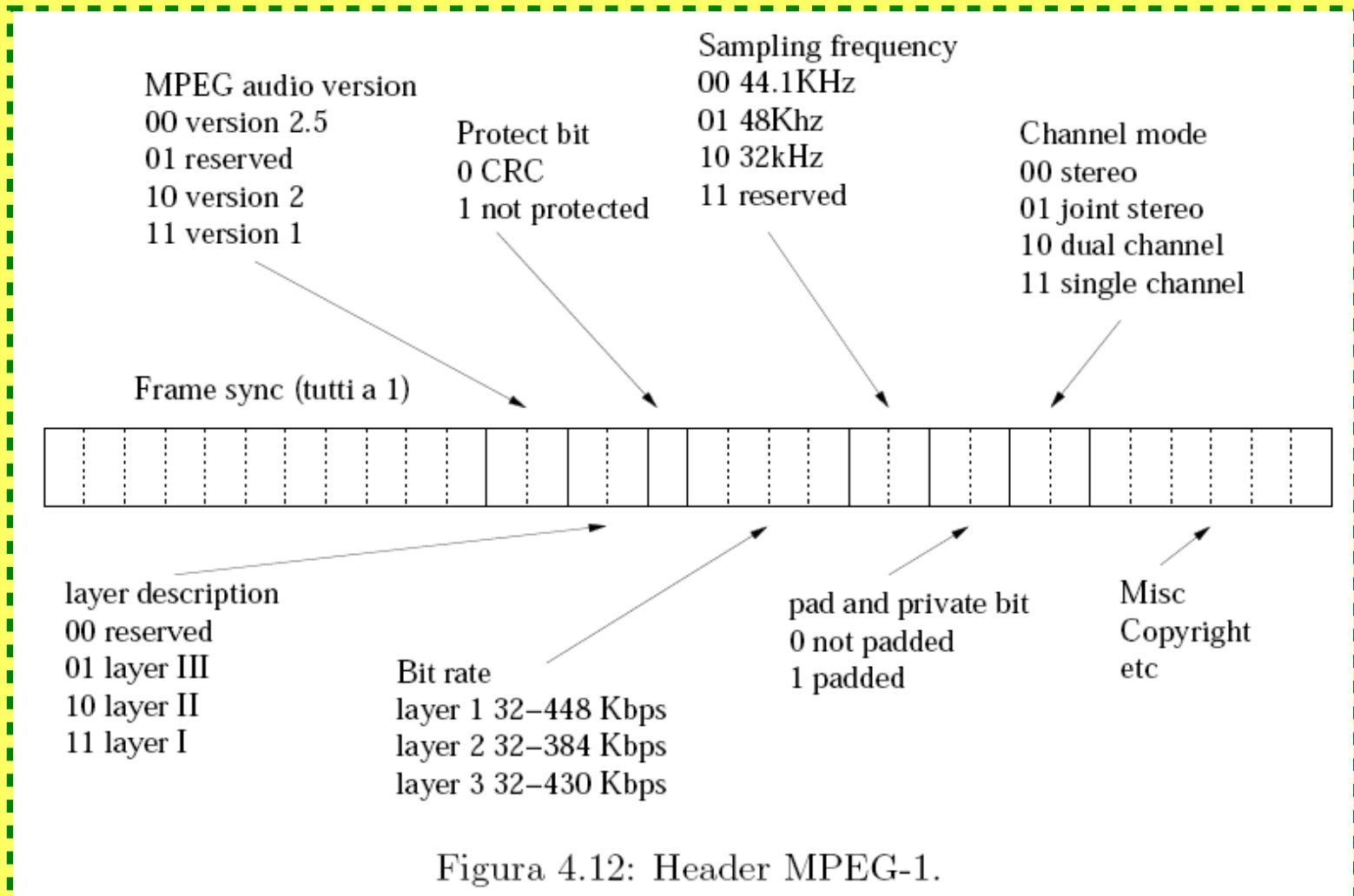


Figura 4.12: Header MPEG-1.

16	16	4x32=128	6xN=192 max	12xN
Sync	Framelen	Bit allocation	Scale factors	Audio data

Figura 4.13: Organizzazione dei campioni.

# Audio – MPEG-1

## *prestazioni*

Compressore	Complessità codificatore	Complessità decodificatore
MP1	1.5	1.0
MP2	3	1.25
MP3	7.5	2.5

Tabella 4.1: Complessità MPEG-1.

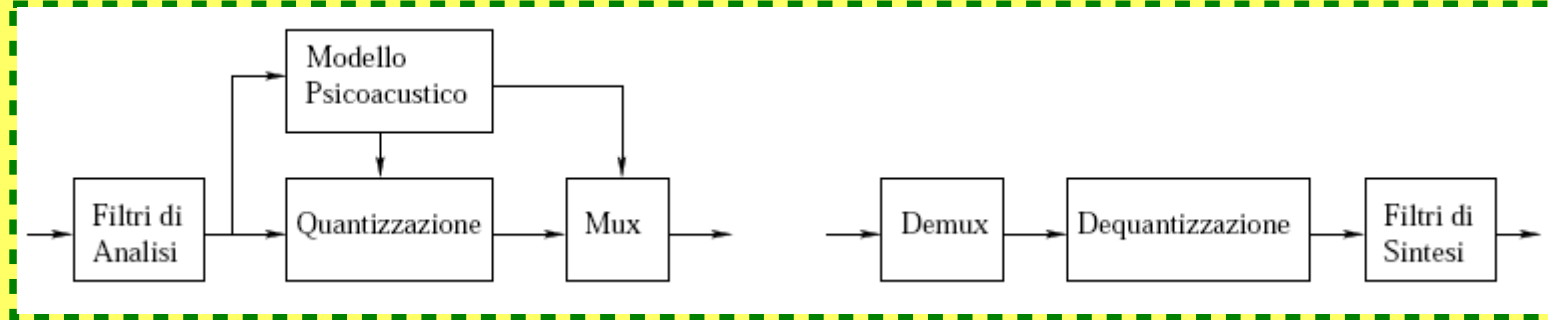
- **MP2**
  - **64 Kbps → MOS = 2.4**
  - **128 Kbps → MOS = 4.0**
- **MP3**
  - **64 Kbps → MOS = 3.7**

# Audio – MPEG-2

- Evoluzione di MPEG-1
- **Dimezzamento frequenze di campionamento**
- **Codificatore Advance Audio Coding (AAC)**: miglioramenti nel filtraggio, calcolo trasformato, correzione dati con tecniche FEC.
- Supporto di un numero di **canali** superiore a due (MPEG-1 solo mono e stereo)
  - **5.1**: capacità di riprodurre audio su 6 diversi altoparlanti
    - tre frontali (centrale e a 30°)
    - due dietro (con angolo da 90° a 110°)
    - uno per suoni bassi (sub-woofer) con posizione qualunque.

# Audio – Dolby

- Nei codificatori MPEG il modello psicoacustico è noto solo al codificatore → necessario inviare al decodificatore informazioni sull'allocatione dei bit → overhead



- Dolby hanno lo scopo di rendere più efficace tale meccanismo
- **AC-1 (Acoustic Coder 1)**
- Utilizza una strategia di allocatione di bit fissa per ogni sottobanda, scelta sulla base di parametri medi

- Non serve trasmettere l'allocatione dei bit

- Banda  
512 Kbps

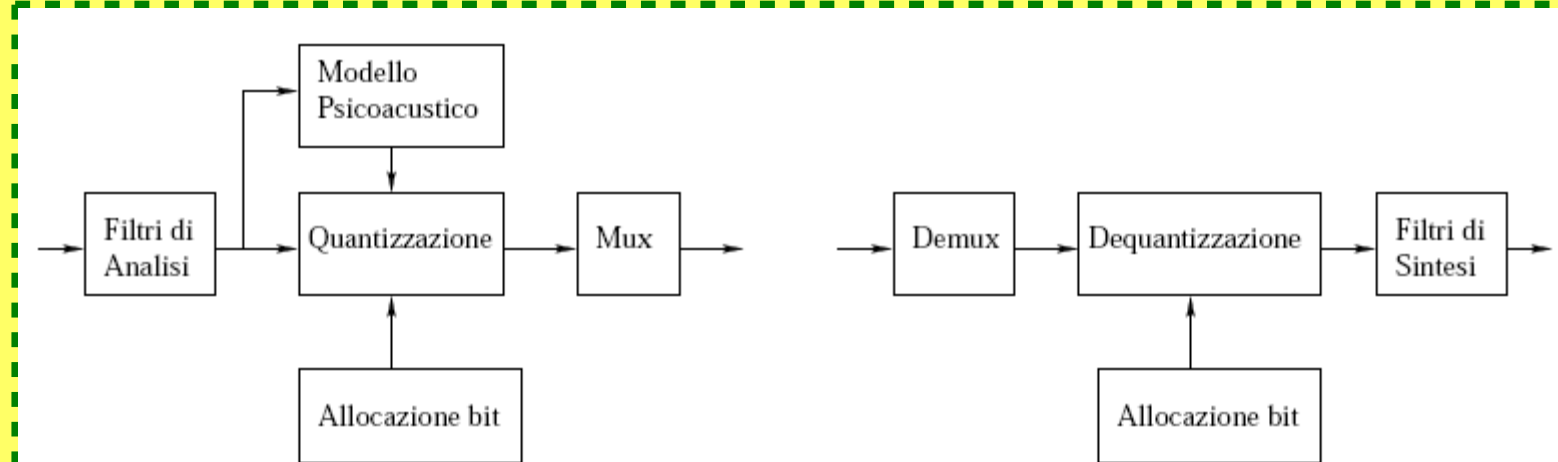
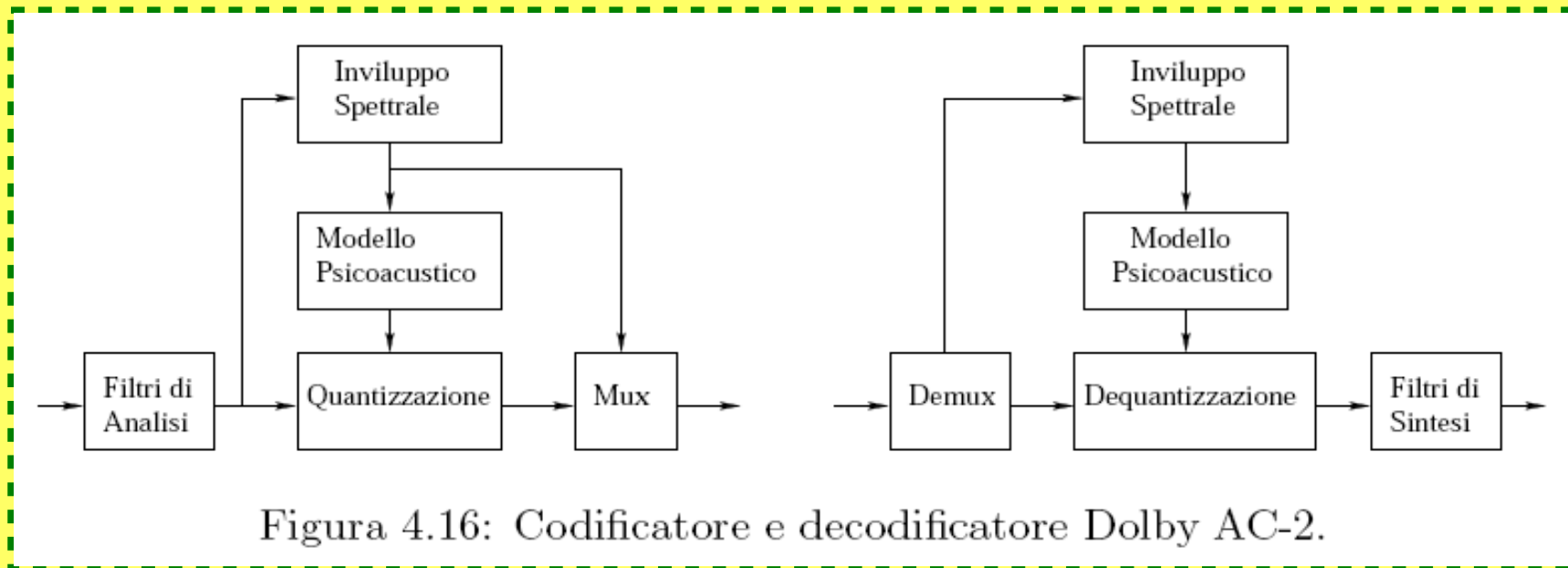


Figura 4.15: Codificatore e decodificatore Dolby AC-1.

# Audio – Dolby

- **AC-2 (Acoustic Coder 2)**
- Allocazione dinamica dei bit, non inviata dal codificatore al decodificatore
- Il decodificatore ricava le informazioni relative al numero di bit allocati nella quantizzazione applicando nuovamente il modello psicoacustico che aveva applicato il codificatore
- Banda 256 Kbps



# Audio – Dolby

- **AC-3 (Acoustic Coder 3)**

- L'analisi dell'involuppo spettrale di AC-2 produce un'uscita con un primo livello di quantizzazione quindi il modello psicoacustico lavora su campioni già affetti da rumore
- AC-3 risolve questo problema utilizzando due diversi modelli psicoacustici

