

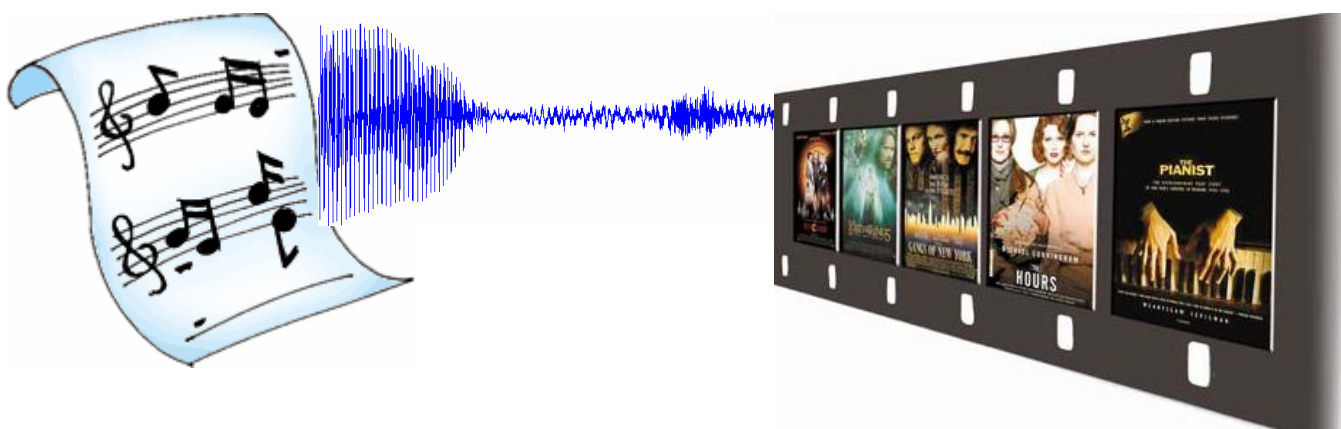


**Università degli studi di Pisa**

**Laurea specialistica in  
Tecnologie informatiche**

Anno accademico 2003 - 2004

**“I formati audio AAC e Dolby Digital (AC-3)”**



**Andrea Lorenzani**

**Laboratorio di Informatica Musicale**

**Docente  
Leonello Tarabella**

# Sommario

## CAPITOLO 1

<b>INTRODUZIONE .....</b>	<b>5</b>
1.1. MOTIVAZIONI DELLA RICERCA .....	5
1.2. ELEMENTI DI TEORIA SULLA MUSICA DIGITALE .....	5
1.3. LA COMPRESSIONE .....	6
1.3.1. <i>Tecnologie di compressione</i> .....	6
1.3.2. <i>Il mascheramento</i> .....	7
1.3.3. <i>Il codice di Huffman</i> .....	9
1.4. I FORMATI AUDIO PIÙ IMPORTANTI .....	11

## CAPITOLO 2

<b>DALL’MPEG LAYER 1 ALL’AAC.....</b>	<b>15</b>
2.1 MPEG-1 .....	15
2.1.1 <i>Origine dello Standard</i> .....	15
2.1.2 <i>Standard MPEG-1</i> .....	15
2.1.3 <i>MPEG Layer 1 e 2</i> .....	17
2.1.4 <i>L’MPEG layer 3 (MP3)</i> .....	19
2.1.5 <i>Il formato dei dati</i> .....	21
2.2 MPEG-2 .....	21
2.2.1 <i>Lo Standard MPEG-2</i> .....	21
2.2.2 <i>L’Advanced Audio Coding (AAC)</i> .....	22
2.2.3 <i>Tipi di codifica AAC</i> .....	26
2.2.3 <i>MP3 e AAC a confronto</i> .....	26
2.3 MPEG-4 .....	27
2.3.1 <i>Le caratteristiche dell’MPEG-4</i> .....	27
2.3.2 <i>Differenze tra AAC dell’MPEG-2 e dell’MPEG-4</i> .....	28
2.3.3 <i>Gli Object Types dell’AAC in MPEG-4</i> .....	29
2.4 TRATTAMENTO E GESTIONE DEGLI ERRORI .....	30
2.5 FORMATI DI FILE AAC .....	30
2.5.1 <i>Formato RAW DATA BLOCK</i> .....	31
2.5.2 <i>Formato Audio Data Interchange Format (ADIF)</i> .....	31
Campo .....	31
N. di bit.....	31
Note .....	31
2.5.3 <i>Formato Audio Data Transport Stream (ADTS)</i> .....	31
Campo .....	31
N. di bit.....	31
Note.....	31
2.5.4 <i>Formato MP4</i> .....	32

## CAPITOLO 3

<b>IL DOLBY DIGITAL.....</b>	<b>33</b>
3.1 INTRODUZIONE .....	33
3.1.1 <i>Il canale Low Frequency Effect (LFE)</i> .....	33
3.1.2 <i>La storia dell’audio multicanale</i> .....	34
3.1.2.1 <i>Le origini del suono surround</i> .....	35
3.1.2.2 <i>Lo stereo di casa e il suono quadrifonico</i> .....	36
3.1.2.3 <i>Dolby e il suono cinematografico</i> .....	37
3.1.2.4 <i>La rivoluzione video</i> .....	37
3.1.2.5 <i>Il sonoro surround entra nelle case</i> .....	38
3.1.2.6 <i>La generazione successiva: il Dolby Digital</i> .....	39
3.1.2.7 <i>Dolby Digital nelle case</i> .....	40
3.1.2.8 <i>Un futuro luminoso per la musica e l’audio multicanale</i> .....	42
3.2 APPROFONDIMENTO SUL DOLBY DIGITAL .....	42
3.2.1 <i>Codificatori e decodificatori Dolby Digital</i> .....	42
3.2.2 <i>Downmixing</i> .....	43
3.2.2.1 <i>Compatibilità al formato</i> .....	43
3.2.2.2 <i>Redirezione dei canali</i> .....	43

3.2.3	Controllo della gamma dinamica ( <i>Dynamic Range Control - DRC</i> ).....	43
3.2.4	Gestione dei bassi .....	44
3.2.5	Compatibilità con i dispositivi <i>Dolby Surround</i> esistenti.....	44
3.3	L'IMPORTANZA DELL'AMBIENTE DI ASCOLTO.....	45
3.4	L'ALGORITMO AC-3 .....	48
3.4.1	Introduzione .....	48
3.4.2	Panoramica sull'elaborazione .....	48
3.4.3	I principali blocchi di elaborazione di codifica .....	48
3.4.3.1	Buffer di ingresso.....	49
3.4.3.2	Filtraggio di ingresso .....	49
3.4.3.3	Scoperta di transienti .....	49
3.4.3.4	Banco filtri TDAC .....	49
3.4.3.5	Conversione alla virgola mobile .....	50
3.4.3.6	Precombinazione della portante .....	50
3.4.3.7	Allocazione globale dei bit .....	50
3.4.3.8	Quantizzazione .....	51
3.4.3.9	Impacchettamento dei dati.....	51
3.4.4	I principali blocchi di elaborazione di decodifica.....	51
3.4.4.1	Buffer di ingresso.....	51
3.4.4.2	Ricerca e correzione di errori.....	52
3.4.4.3	Spacchettamento di dati a formato fisso .....	52
3.4.4.4	Allocazione dei bit di decodifica.....	52
3.4.4.5	Spacchettamento dei dati a formato variabile .....	52
3.4.4.6	Conversione a virgola fissa .....	52
3.4.4.7	Reinserimento della portante a alte frequenze .....	52
3.4.4.8	Trasformata inversa .....	52
3.4.5	Altre funzionalità dell' algoritmo.....	53
3.4.6	In sintesi .....	53

## CAPITOLO 4

CONCLUSIONI.....	55
------------------	----

# Indice delle figure

<i>udibilità dell'orecchio</i>	7
<i>mascheramento in frequenza</i>	8
<i>effetto di più mascheramenti concatenati</i>	8
<i>mascheramento temporale</i>	8
<i>pre-mascheramento e post-mascheramento</i>	9
<i>soglia globale di mascheramento</i>	9
<i>esempio di compressione con l'algoritmo di Huffman</i>	10
<i>forma d'onda sinusoidale con rappresentazione DSD</i>	12
<i>decoder MPEG</i>	16
<i>encoder audio</i>	16
<i>processo di codifica e decodifica in MPEG</i>	18
<i>encoder MP3</i>	19
<i>confronto tra WAV (rosso) e MP3 (Verde)</i>	20
<i>dati sulle prestazioni della codifica MP3</i>	21
<i>codificatore MPEG-2 AAC</i>	23
<i>decodificatore MPEG-2 AAC</i>	25
<i>Tool necessari e richiesti per la codifica e decodifica AAC</i>	25
<i>esempio di risparmio di spazio tra AAC e MP3</i>	27
<i>i nuovi tools dell'AAC MPEG-4 nello schema di codifica</i>	28
<i>la linea del tempo e le nuove tecnologie</i>	34
<i>nastro magnetico 70 mm e speakers nei cinema</i>	35
<i>formato stereo ottico della Dolby</i>	37
<i>evoluzione del Surround nelle case</i>	38
<i>funzionamento del surround nelle case</i>	39
<i>Dolby Digital nelle pellicole cinematografiche</i>	40
<i>Dolby Digital nei cinema</i>	40
<i>Dolby Digital nelle case</i>	41
<i>tipi di decodificatori e compatibilità all'indietro</i>	41
<i>evoluzione del segnale sonoro sulle pellicole cinematografiche</i>	42
<i>esempio di impianto Dolby Digital</i>	42
<i>schema di codifica Dolby Surround</i>	44
<i>decodificatore Dolby Surround Pro Logic</i>	45
<i>Tipico layout di una stanza equipaggiata con decoder 5.1</i>	46
<i>stanza di medie dimensioni con decoder 4.1</i>	46
<i>Layout di una grande stanza</i>	46
<i>l'ambiente di ascolto raccomandato dall'ITU-R</i>	47
<i>encoder AC3</i>	49
<i>decoder AC-3</i>	51

# Capitolo 1

## Introduzione

### **1.1. Motivazioni della ricerca**

Ho svolto questa ricerca per descrivere in maniera più accurata possibile quale sia attualmente lo “stato dell’arte” riguardo la codifica digitale del suono. Esaminerò soprattutto due formati che ultimamente stanno acquisendo sempre maggiore importanza: lo standard AAC (*Advanced Audio Coding*) e lo standard AC3 (comunemente conosciuto come *Dolby Digital*).

Ho scelto questi due argomenti perché hanno suscitato in me particolare interesse per motivi diversi: per quanto riguarda il formato AAC se ne sente sempre più spesso parlare paragonandolo all’MP3 (Mpeg Layer III), in quanto sembrerebbe mantenere la stessa qualità audio permettendo una compressione migliore, mentre l’AC3 è interessante in quanto formato standard per i film su DVD.

Ovviamente ho ritenuto interessante documentarmi anche sulla storia di questi formati, che metterò come introduzione ai due argomenti, mentre in questo capitolo farò una breve introduzione sull’audio digitale (esaminando brevemente i formati digitali storicamente più famosi) e sulla teoria alla base della compressione audio, introducendo nozioni che verranno riprese spesso nel corso della ricerca.

### **1.2. Elementi di teoria sulla musica digitale**

Dobbiamo innanzi tutto considerare che tutti i fenomeni fisici naturali sono di natura analogica, cioè possono assumere infiniti valori in un campo continuo, e che il mondo dei PC o il mondo dei binario dei numeri invece è fatto di 1 e 0 (uno e zero) di grandezze "discrete". Quindi se considerassimo un una conversione audio da ANALOGICA a DIGITALE dovremmo prima effettuare una "discretizzazione" del segnale. Tale processo avviene in due distinti processi: il primo si chiama campionamento (o *sampling*), il secondo quantizzazione e riguarda l'ampiezza del segnale campionato. Effettuare un campionamento quindi, vuol dire considerare solo i valori che il segnale analogico assume in determinati intervalli di tempo, ossia rilevare un certo numero di campioni nell'unità di tempo. Il numero di campioni considerati nell'unità di tempo si chiama frequenza di campionamento e si esprime in Hz. Pensiamo ora che con il PC, se volessimo registrare dei suoni (ad esempio la nostra voce) dovremmo campionare a 11025Hz, mentre se volessimo la qualità CD dovremmo campionare a 44100 Hz.. Esiste una spiegazione scientifica di tutto ciò: in poche parole un segnale limitato in banda è ricostruibile in modo perfetto dai suoi campioni, purchè siano acquisiti con una frequenza almeno doppia di quella massima riproducibile di segnale (Teorema di Nyquist). Ecco spiegato perchè per fare dei CD audio è necessario campionare a 44100 Hz in modo da avere una frequenza massima di 20000 Hz. E' necessario però ancora un passaggio: la quantizzazione. La quantizzazione è direttamente dipendente dalla risoluzione che viene espressa in bit. In pratica i dati vengono raggruppati in fasce analogiche a cui ciascuna corrisponde un valore discreto digitale, quindi maggiore è il numero di bit impiegato, maggiore è l'approssimazione e quindi diciamo che la risoluzione è maggiore. Per esempio se abbiamo 8 bit si avranno 256 combinazioni, con 16 bit 65536, con 24 bit oltre i 16 milioni di

combinazioni ( $2$  elevato al numero di bit). Diciamo pure che 24bit garantiscono un'elevatissima approssimazione. I vantaggi principali del digitale sono: assenza di fruscii, mancanza di fenomeni di usura (presenti ad esempio quando si usano nastri) e riproducibilità infinita, cioè ogni copia può essere un master. Se provassimo invece a fare la copia della copia della copia con una registrazione analogica alla fine di tutti i passaggi avremo un rapporto segnale disturbo(S/N) tale da avere sul nastro solo rumore e quasi niente segnale.

### **1.3. La compressione**

#### **1.3.1. Tecnologie di compressione**

Per ridurre lo spazio occupato da un qualsiasi file su un supporto (ad esempio su un CD o su un disco fisso) si opera una compressione, attraverso software che utilizzano spesso algoritmi molto complicati. Ma tutte le compressioni non sono uguali. Vediamone le tre principali differenze:

**lossless** - compressione senza perdita di informazione (un esempio è rappresentato da un file zippato che per poter essere eseguito deve subire un processo inverso di decompressione)

**transparent**- comprime molto di più del lossless, ma sacrifica una piccola parte delle informazioni, difficilmente percepibili (un esempio è rappresentato da un file JPEG)

**lossy**- compressione con perdita sensibile di informazioni (es. si riduce fino ad un centinaio di volte la dimensione)

Due importanti concetti per la codifica percettiva dell'audio digitale sono la ridondanza e l'irrilevanza. Entrambi descrivono i motivi alla base ai quali una certa quantità di informazione audio può essere scartata senza causare scadimenti "percettibili" sulla qualità audio. L'irrilevanza è infatti una particolare caratteristica della psicoacustica attraverso la quale si possono eliminare particolari di una forma d'onda irrilevanti intesa come percezione sonora. La ridondanza invece appartiene all'informazione digitale legata al campionamento e si basa sia sul teorema di Nyquist che sull'effetto di Mascheramento, ovvero a parità di frequenza un suono minore è mascherato da uno di maggiore intensità (bisogna però stare attenti alle distorsioni armoniche).

### 1.3.2. Il mascheramento

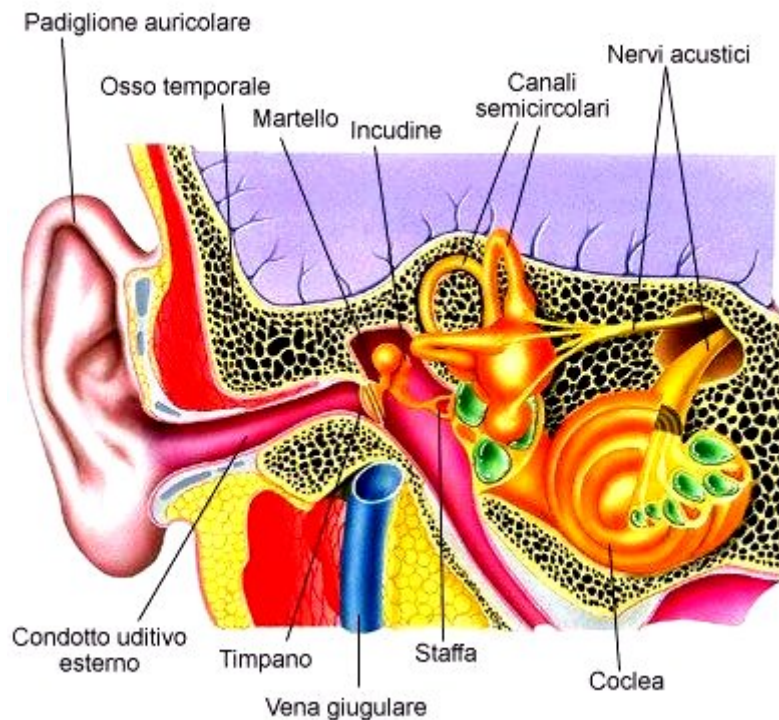


Figura 0.1: l'orecchio umano

La teoria psicoacustica si basa sull'osservazione del comportamento del sistema uditivo umano. L'orecchio ha la capacità di adattarsi al segnale che riceve, permettendo di sentire sia suoni molto deboli che molto forti con un range dinamico di circa 96 db. Se però è presente un suono molto intenso, l'orecchio non è più in grado di distinguere i suoni più deboli prodotti contemporaneamente o nei millisecondi successivi. La conoscenza di questi comportamenti è utilizzata in fase di codifica, per eliminare informazioni relative a parti del segnale non udibili. La tecnica che sfrutta l'incapacità di udire segnali deboli in presenza di segnali più intensi, a frequenze simili, è chiamato "mascheramento in frequenza". La tecnica basata sull'incapacità di sentire segnali deboli nei millisecondi successivi all'emissione di segnali intensi è chiamata "mascheramento temporale".

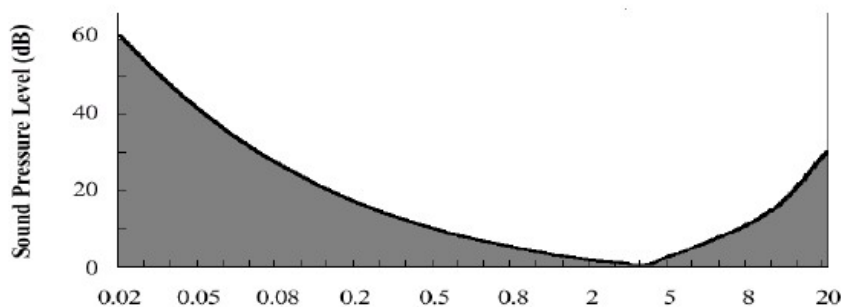


Figura 0.2: udibilità dell'orecchio

Il grafico mostra la capacità percettiva dell'orecchio umano a seconda della frequenza del segnale ricevuto. Essa è nulla al di fuori dell'intervallo 20 Hz-20 kHz, mentre è massima tra i 2 e i 4 kHz. Il grafico illustra l'attenuazione di capacità uditiva rispetto al massimo.

Se però è presente un segnale di intensità fissa (60 db) alla frequenza di 1 kHz, la capacità uditiva di tutte le frequenze vicine a 1 kHz diminuisce considerevolmente. Il picco di attenuazione attorno alla frequenza del segnale è definita "curva di mascheramento" per il mascheramento in frequenza.

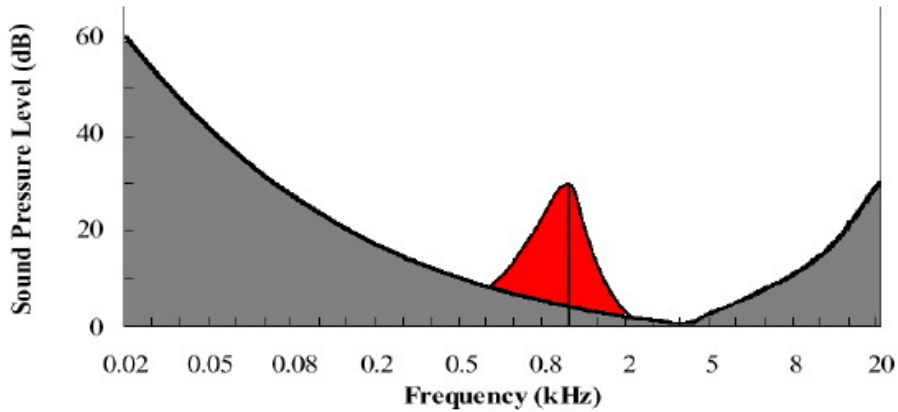


Figura 0.3: mascheramento in frequenza

Se l'esperimento è ripetuto per più frequenze, si nota che le curve di mascheramento aumentano di ampiezza con il crescere della frequenza. A frequenze inferiori ai 500 Hz la zona di mascheramento è di circa 100 Hz, oltre i 500 Hz è di circa 4 kHz. Queste zone si definiscono "bande critiche" perché al loro interno l'orecchio umano ha una risoluzione frequenziale molto limitata. Esse sono alla base della compressione audio MPEG.

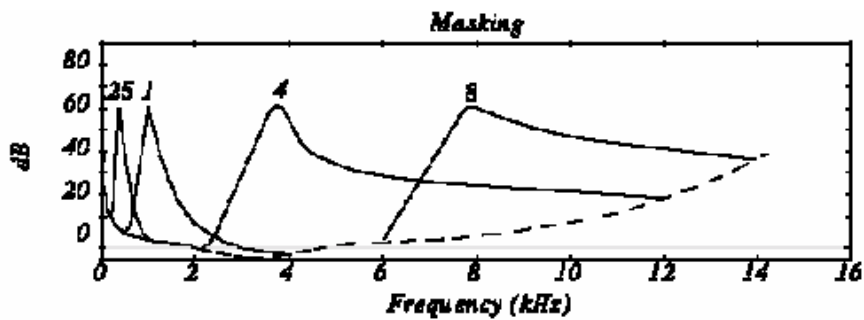


Figura 0.4: effetto di più mascheramenti concatenati

Per quanto riguarda il mascheramento temporale, quando l'orecchio umano percepisce un suono di elevata intensità per un certo intervallo di tempo, perde parte della sua capacità uditiva nei millisecondi successivi alla scomparsa del segnale. Il grafico seguente illustra l'andamento temporale della capacità uditiva in seguito alla propagazione di un segnale di 60 db di intensità della durata di 5 millisecondi. Si può notare come ci sia una notevole attenuazione della capacità uditiva nei millisecondi seguenti alla scomparsa del segnale.

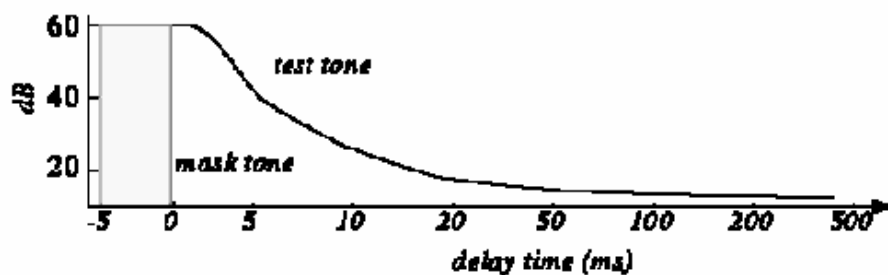


Figura 0.5: mascheramento temporale

Ci sono due tipi di mascheramento nel tempo: il post mascheramento è quello che abbiamo trattato, dato da un suono a una certa intensità che copre tutti i suoni che lo seguono per un certo tempo, ma esiste anche un pre mascheramento.

Se il post-mascheramento è facilmente comprensibile, il pre-mascheramento è meno ovvio, per averne un'idea è sufficiente ricordare che affinché un suono sia percepito è necessario che si mantenga per un certo tempo senza disturbi. Supponiamo di avere un suono collocato nella zona rossa della figura sottostante: prima che l'orecchio lo distingua chiaramente interviene un'altra sollecitazione più forte, il primo viene disturbato e non può più esser rilevato. La figura sotto fornisce anche un paragone fra le durate dei due fenomeni.

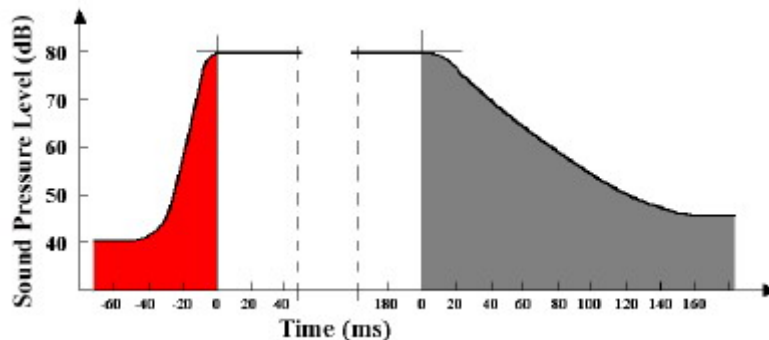


Figura 0.6: pre-mascheramento e post-mascheramento

Nel grafico seguente è riportata, a titolo di esempio, la curva di mascheramento relativa ad un segnale ad una frequenza stabilita, considerando sia il mascheramento in frequenza che nel tempo.

Tale curva di mascheramento viene generalmente chiamata Soglia Globale di Mascheramento (dall'inglese *Global Masking Threshold*) oppure soglia di udibilità dinamica.

Tutti i segnali che si trovano sotto la curva non sono udibili e quindi non è necessario codificarli.

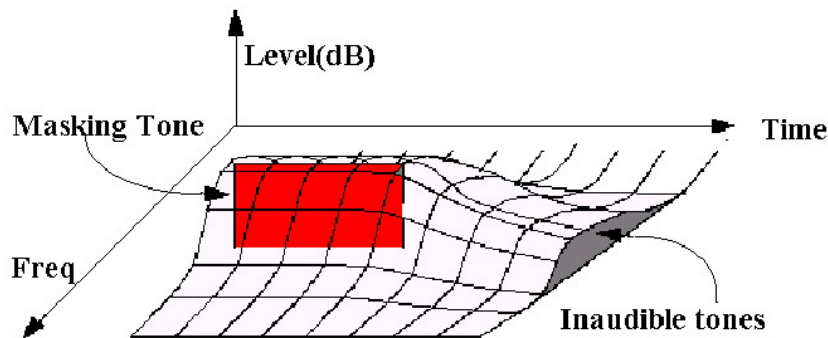
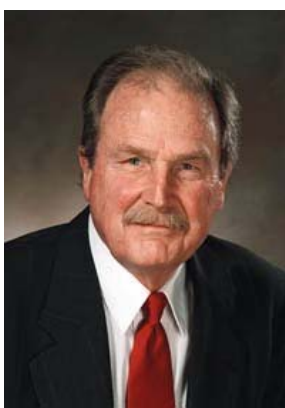


Figura 0.7: soglia globale di mascheramento

Queste osservazioni sono utilizzate negli algoritmi di compressione audio per eliminare le informazioni sulle frequenze non udibili. Gli schemi più avanzati che seguono questi principi sono il Dolby AC-3 e l'MPEG 1 (in particolare il layer 3).



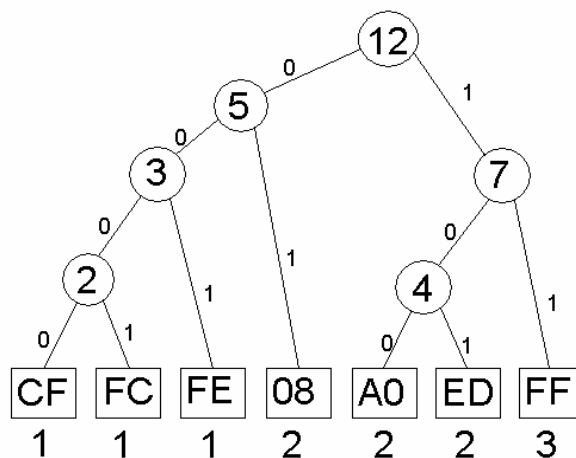
### 1.3.3. Il codice di Huffman

Gli algoritmi di Huffman forniscono un metodo per ridurre la dimensione dei dati, e sono molto diffusi, si pensi ad esempio ai programmi di compressione più conosciuti (WinZip, WinRar). La tecnica di Huffman si basa sulla sostituzione di simboli (ad 8 bit tipicamente) con una sequenza

più corta di bit (minimo 1), in funzione della frequenza del dato che compare nella sequenza da comprimere. L'algoritmo funziona tanto meglio quanto più ricorrenti sono i simboli, quindi non offre prestazioni notevoli nel caso di files PCM, ma si rivela molto più efficace se applicato a dati nel dominio delle frequenze. Prendiamo ad esempio il seguente file (rappresentato in codici esadecimali) e scriviamo in una tabella ogni singolo codice seguito dal numero delle sue ripetizioni nel file:

FF FF ED ED FF A0 A0 FE FC 08 08 CF	▶	FF (3 volte)
		ED (2 volte)
		A0 (2 volte)
		FE (1 volta)
		FC (1 volta)
		08 (2 volte)
		CF (1 volta)

La logica dell'algoritmo di Huffman è la seguente: anziché utilizzare N bit per ogni codice si assegnano meno bit per i dati che si ripetono di più, e se necessario più bit per quelli più rari. Nel nostro caso, il codice che si ripete più di tutti è FF (3 volte), per cui dovremmo aspettarci una sequenza di bit più corta rispetto agli altri. La creazione dell'albero che verrà utilizzato procede in questo modo: prima vengono messi alla base i nodi ordinati per numero di volte in cui compaiono. Poi si uniscono i due nodi con valore minimo in un nodo il cui valore è la somma dei due nodi precedenti. A questo punto si tiene in considerazione il nuovo gruppo di nodi senza i due precedentemente uniti ma con il nuovo nodo somma dei due, e si procede ricorsivamente. Alla fine si avrà un albero come in figura:



**Figura 0.8:** esempio di compressione con l'algoritmo di Huffman

Questo viene generato in modo da avere i codici con frequenza maggiore ai livelli più alti, nell'esempio FF si trova al secondo mentre gli altri sono variano dal secondo al quarto. Ora si procede ad assegnare ad ogni codice la sequenza di bit che lo rappresenterà. La procedura è semplice, partendo dalla radice si sceglie uno 0 se si procede verso sinistra, 1 se a destra. La mappatura che ne segue è la seguente:

*FF:11 Ed:101 A0:100 FE:001 FC:0010 08:01 CF:0000*

È da notare che il codice con frequenza maggiore è rappresentato con due soli bit e che il codice è univoco, non c'è possibilità d'errore. A questo punto non resta che riscrivere il file originale che diventa:

FF	FF	ED	ED	FF	A0	A0	FE	FC	08	08	CF
11	11	101	101	11	100	100	001	0010	01	01	0000

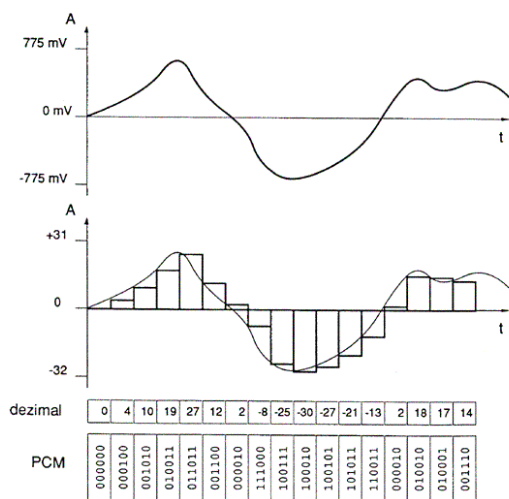
La sequenza compressa occupa 5 byte (33 bit arrotondati in eccesso) rispetto ai 12 iniziali, circa il 40% dell'originale.

In realtà i codificatori utilizzati per la compressione audio lavorano in modo leggermente diverso. Il pacchetto di dati da comprimere non è scandito per trovare le frequenze e i codici ma ci si affida a delle tabelle di frequenza appositamente redatte. Queste, che vengono scelte con particolari tecniche in modo da essere ottimizzate con il lavoro del quantizzatore, presuppongono che i bassi livelli di quantizzazione siano i più frequenti. Ovviamente al decoder dovrà esser comunicato per ogni pacchetto di dati quale tabella è stata utilizzata.

## 1.4. I formati audio più importanti

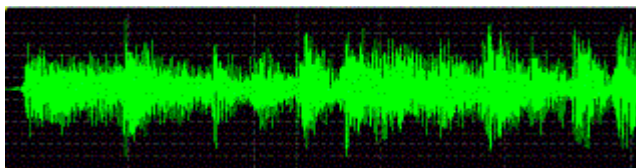
Il suono può essere trasferito dalla sua forma analogica a quella digitale in più formati, che hanno alcune differenze tra loro. In genere le prime due caratteristiche che si notano sono la qualità e l'occupazione fisica sul supporto fisico usato, ma in alcuni casi entrano in gioco anche altre caratteristiche.

I formati audio attualmente più diffusi sono:



**PCM:** acronimo di Pulse Code Modulation (modulazione a codifica numerica di impulsi). Si tratta di una tecnica utilizzata per trasformare un segnale dalla forma analogica a quella digitale. Con frequenza di campionamento di 44100Hz a 16bit stereo è il primo esempio di musica digitale utilizzato nei CD AUDIO, lo standard PCM definisce appunto la qualità CD. In pratica è un segnale audio digitale in formato non compresso. Tuttavia il Pulse Code Modulation va molto bene per i CD, campionato a 16 bit e 44.100 kHz, ma più si sale con la frequenza di campionamento più l'aumento di prestazioni diventa impercettibile, per questo motivo sono stati elaborati formati più sofisticati per i segnali audio non compressi, come ad esempio il DSD

**WAV, AIFF, AU:** Sono i formati più comuni per la memorizzazione di dati audio in formato PCM, Il formato Wave di Microsoft e IBM, gestisce frequenze di campionamento che arrivano fino a 44.1 kHz, 48 kHz e ora anche 96 e perfino 192 kHz, risoluzione fino a 32 bit lineari e offre la possibilità di memorizzare su un solo file anche segnali stereo o surround con numero illimitato di altoparlanti (che corrispondono ad altrettanti canali). Questo formato viene largamente utilizzato per creare effetti sonori, specialmente nei programmi e nei giochi o per personalizzare facilmente eventi particolari in una pagina web. Lo stesso discorso si può fare per il formato AIFF di Apple. Il formato Au gestisce anche modalità più efficienti di quantizzazione che permettono una riduzione della mole di dati anche di 4 volte il valore originale, al costo di una modesta perdita di qualità. Tutti e tre i formati non sono altro che la registrazione in digitale di suoni reali, suoni che hanno avuto origine da una fonte esterna al PC. Questi formati non sono utilizzabili per diffondere musica su internet o per memorizzare brani sul PC, a causa dell'eccessivo dispendio di risorse che il suo uso comporta.



**DSD/PDM:** (*Direct Stream Digital*) lo sviluppo di questo formato è stato voluto dalla Sony e dalla Philips, che ritenevano il formato PCM



ormai superato. Questo perché ogni sistema PCM necessita di un filtro in ingresso che blocca ogni segnale con frequenza pari o superiore alla metà di quella di campionamento. Se per esempio stiamo trattando un segnale a 96 kHz, il filtro è regolato per far passare un segnale a 47.5 kHz, ma per bloccare un segnale a 48 o 48.5 kHz. Il compito del filtro, già difficile con campionamenti a 44.1 kHz, diventa più facile se alziamo il campionamento a 96 kHz, meglio ancora a 192. Questo aumento della risoluzione non risolve però un altro problema, quello del filtro a decimazione (downsampling), utilizzato in fase di registrazione e del filtro a interpolazione (oversampling) utilizzato in riproduzione. Per chiarirci meglio le idee è bene osservare la figura sotto, che raffigura la struttura di un circuito di conversione analogico-digitale e digitale-analogico.

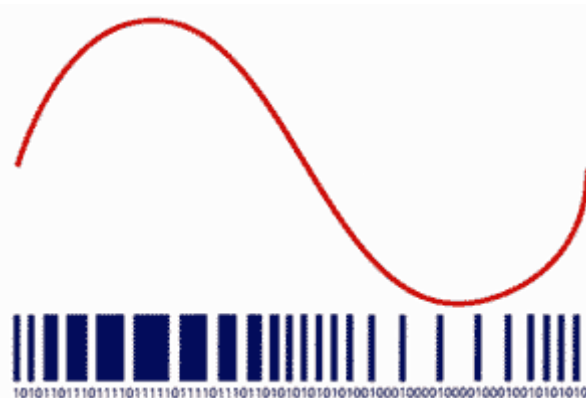
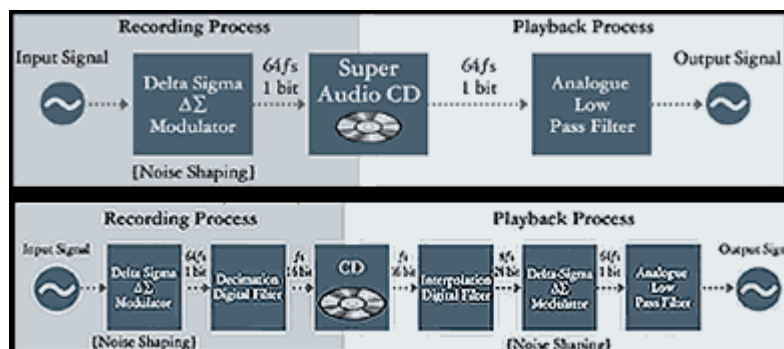


Figura 0.9: forma d'onda sinusoidale con rappresentazione DSD

La soluzione, secondo le due aziende, si chiama DSD e il principio base è l'eliminazione di questi due filtri distruttivi per il segnale. Il segnale analogico, dopo essere stato convertito in digitale ad 1 bit non viene subito decimato in una stringa multibit a 16, 18, 20 o 24 bit come accade per il PCM standard, ma viene inciso direttamente su disco come una stringa di impulsi ad un bit, ovvero un insieme di "1" e "0".



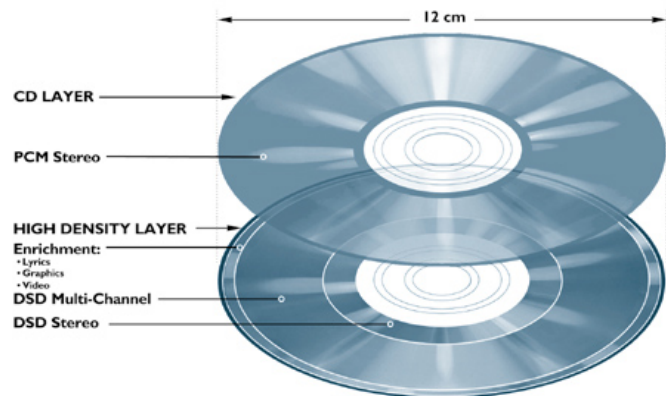
Per costruire questo treno di impulsi il convertitore utilizza un ciclo a feedback negativo che accumula la stringa di "1" e "0": se il livello di un impulso è superiore a quello

acquisito nel ciclo precedente allora viene scritto un "1", se invece il valore è inferiore viene scritto uno "0". Il risultato è ovvio: immaginando che il segnale sia un senoide perfetto, il livello massimo verrà rappresentato come una stringa di "1", lo zero verrà rappresentato come un'alternanza di "0" e di "1", il minimo come una stringa di "0". Ecco perché al posto di DSD talvolta si usa PDM, ovvero **Pulse Density Modulation**. Per un secondo di musica il DSD effettua 2.822.400 cicli. Dove sono i vantaggi? Prima di tutto il numero dei cicli è 64 volte più elevato rispetto ai 44.100 che vengono fatti per il CD audio e ai 192.000 cicli destinati ai DVD Audio. Poi con il sistema ad un bit abbiamo una più precisa ricostruzione della forma d'onda tradizionale: il treno di impulsi ci dice molto sull'effettiva ampiezza e sulla frequenza della forma d'onda originale. Per la conversione poi non servono filtri distruttivi ma basta un semplice passa basso analogico. Nel mondo reale tutto non è però così perfetto: il treno di bit è abbastanza rumoroso e per raggiungere l'elevato valore di rapporto segnale-rumore garantito dal SACD è necessario utilizzare un filtro di noise shaping che sposta il rumore a frequenze più elevate di quelle della banda audio.

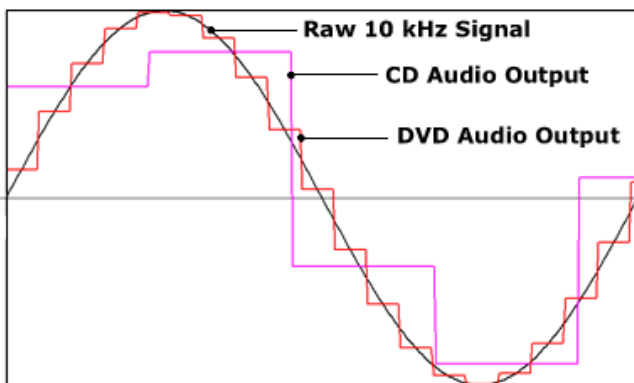
**SACD:** il “Super Audio CD” che si basa sulla tecnologia rivoluzionaria DSD. Per circa trent’anni i formati audio digitali inclusi i Compact Disc, DAT e Dash si basavano sulla tecnologia PCM “Pulse Code Modulation”.

Con l’effettivo livello di campionamento di molto superiore a quello lineare del PCM, il SACD, con la sua banda passante molto estesa permette di ottenere un suono più’ dolce, naturale nelle alte frequenze, nonché molto più’ vicino al suono analogico. Il SACD può includere audio sia in formato stereo che multi canale. ed inoltre offre il possibile vantaggio di una formattazione ibrida. Questa permette di avere un supporto a doppio strato, di cui uno SACD ed uno CD.

**Hybrid Disc Content**



**DVD Audio:** Il DVD-Audio è un formato distinto dal DVD-Video ed è progettato per migliorare notevolmente l'audio digitale e per l'ascolto multicanale ad altissima qualità. I dischi DVD-Audio sono realizzati per funzionare con tutti i lettori DVD Video, tuttavia in questo caso la qualità DVD-Audio non è reale poichè soltanto un lettore DVD-Audio è in grado di leggere la "zona DVD-Audio" separata sul disco (la directory AUDIO\_TS) che i lettori DVD-Video non riescono a vedere. Per questo se si vuole ascoltare la reale qualità del DVD-Audio, è necessario possedere un lettore DVD-Audio, o uno dei nuovi "lettori universali" che possono riprodurre entrambi i formati.



notevolmente l'audio digitale e per l'ascolto multicanale ad altissima qualità. I dischi DVD-Audio sono realizzati per funzionare con tutti i lettori DVD Video, tuttavia in questo caso la qualità DVD-Audio non è reale poichè soltanto un lettore DVD-Audio è in grado di leggere la "zona DVD-Audio" separata sul disco (la directory AUDIO\_TS) che i lettori DVD-Video non riescono a vedere. Per questo se si vuole ascoltare la reale qualità del DVD-Audio, è necessario possedere un lettore DVD-Audio, o uno dei

nuovi "lettori universali" che possono riprodurre entrambi i formati.

Inoltre il DVD Audio multicanale non adotta il sistema Dolby Digital 5.1 o il DTS, bensì l'MPL, un algoritmo 5.1 lossless senza perdite di qualità, con frequenze di campionamento di 48, 96, 192 kHz e dimensioni di campionamento di 16, 20, 24 bit. Teoricamente ciò consente una risposta in frequenza fino a 96 kHz e gamma dinamica fino a 144 dB. Il PCM multicanale è miscelabile dal lettore, anche se a 192 e a 176.4 kHz sono utilizzabili solo due canali. Le frequenze e le dimensioni di campionamento possono variare per i differenti canali usando un set predefinito di gruppi. La velocità massima di trasferimento dati è di 9,6 Mbps.

Il DVD-Audio può fornire fino a 99 immagini fisse per traccia (ai livelli tipici di compressione circa 20 immagini inserite nel buffer di 2 MB del lettore), con un set limitato di transizioni (taglio entrata/uscita, sfumatura entrata/uscita, dissolvenza e wipe). Contrariamente al DVD-Video, l'utente può visionare a volontà le immagini senza che la riproduzione audio venga interrotta: ciò viene chiamata visualizzazione di diapositive sfogliabile. La visualizzazione on-screen può essere utilizzata per la sincronizzazione dei testi e per i menu di navigazione. Può essere usata una speciale modalità di navigazione semplificata sui lettori senza un display.

Quando il DVD uscì nel 1996 non esisteva nessun formato DVD-Audio, sebbene le caratteristiche sonore del DVD-Video siano molto superiori al CD. Il Forum DVD chiese il parere delle industrie discografiche prima di definire il formato DVD-Audio. Una bozza standard vide la luce nel gennaio 1998 dal DVD Forum's Working Group 4 (WG4), e la versione 0.9 uscì a luglio. La specifica finale del DVD-Audio 1.0 (minus copy protection) è stata approvata nel febbraio 1999 e rilasciata a marzo.



Il ritardo è stato in parte causato dal lento processo di definizione dei sistemi anticopia (criptatura e filigrana digitale), con le complicazioni introdotte dal Secure Digital Music Initiative (SDMI). Le proposte offerte da Aris, Blue Spike, Cognicity, IBM, e Solana sono state valutate dalle maggiori case discografiche in collaborazione con le 4C Entity, comprendenti IBM, Intel, Matsushita e Toshiba. Aris e Solana si sono unite formando una nuova società chiamata Verance, la cui tecnologia Galaxy è stata scelta nell'agosto 1999.

I DVD-Audio solitamente non sono riproducibili su Pc, perchè esistono pochissimi computer che possano fornire un ambiente audio di alta qualità necessario per evidenziare i vantaggi della fedeltà del DVD-Audio, anche se all'inizio del 2003 la scheda Audigy 2 Sound Blaster includeva il software necessario per riprodurre DVD-Audio su un computer.

**Streaming audio: RAM, RM, ASF, ASX:** lo streaming è il trasferimento in rete dei dati audiovisivi in tempo reale; tutto questo senza tempi di attesa derivanti dal download completo del file sull'hard disk del computer. Con lo streaming, infatti, non viene scaricato l'intero file audio prima di consentirne l'ascolto, ma la riproduzione inizia per ogni blocco di due secondi di ascolto; nel frattempo viene scaricato il successivo. Si possono verificare momentanee interruzioni nella riproduzione, nel caso in cui il traffico nella rete risulti congestionato. Questo sistema di trasmissione è possibile con modem da 28.8 kbit/sec in su. In base alle capacità

di collegamento dell'utente, è possibile selezionare le diverse qualità di suono. Le due principali tecnologie di audio streaming utilizzate sono Real (attraverso il real player), e Windows Media (Windows Media player). Real Technologies ha iniziato a promuovere tecnologie di streaming audio (e successivamente video), per mezzo di piccoli files con estensione ram o rm con banda di trasmissione che arriva fino a 56Kbit/sec. Windows Media (Microsoft) utilizza una soluzione per lo streaming basata sui file Asf e Asx con una banda abbastanza ristretta (15-25 kbit/sec). La tecnologia dello streaming audio ha permesso alle principali emittenti radiofoniche di presentare sui propri siti web i loro programmi trasmessi via etere, cosicché chiunque sul globo terrestre potesse ricevere una trasmissione "radio" dal pc. Nascono anche le radio "web only", non presenti via etere, ma solo sul web. Assumono importanza maggiore quando l'impossibilità di trasmettere via etere non è solo economica ma anche politica a causa di controlli imposti da regimi autoritari.



# Capitolo 2

## Dall'MPEG Layer 1 all'AAC

### 2.1 MPEG-1

#### 2.1.1 Origine dello Standard



Nel 1988 l'organismo internazionale di standardizzazione ISO-IEC (*International Organization for Standardization and International Electrotechnical Commission*) si è assunto il compito di sviluppare uno standard per la compressione e la rappresentazione del video digitale e dell'audio ad esso associato che fosse adatto alla memorizzazione su dispositivi di memoria di massa (dischi ottici, DAT) e alla trasmissione su canale di telecomunicazione (ISDN, LAN, TV). La nascita di uno standard si era resa necessaria principalmente per due motivi: in primo luogo per assicurare l'interoperabilità tra diversi sistemi hardware e software minimizzando l'effetto degli interessi di parte; inoltre per garantire sia i fornitori che gli utenti di prodotti multimediali: dal punto di vista dei fornitori la presenza di uno standard garantisce la fruizione dei loro prodotti da parte di utenti diversi e con differenti apparati di riproduzione, mentre per gli utenti hanno il vantaggio di poter riprodurre qualsiasi opera multimediale ottenuta in svariati modi e creata da vari autori purché conforme a questo standard. Il *Moving Picture Expert Group* (MPEG) è il comitato internazionale nato in seno all'ISO per raggiungere tale obiettivo. Formalmente MPEG è il gruppo di lavoro 11 del subcomitato 29 del Joint Technical Group 1 dell'ISO-IEC (ISO-IEC/JTC1/SC29/WG11). Il lavoro di questo gruppo di esperti ha portato al rilascio di varie versioni dello standard MPEG, alcune delle quali hanno ricevuto lo status di *International Standard* (IS) mentre altre sono ancora in via di sviluppo.

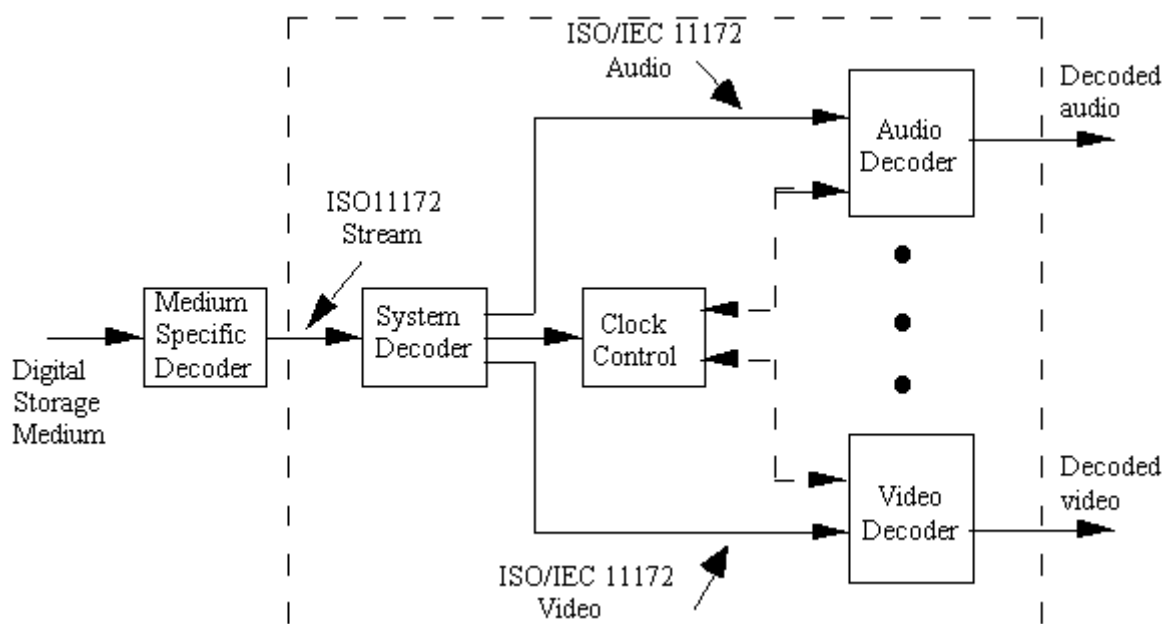
#### 2.1.2 Standard MPEG-1

E' stato progettato per la codifica in forma digitale di immagini in movimento e per l'audio ad esse associato. Sebbene la sua sintassi ammetta molti valori di bitrate e di dimensione delle immagini, esso è nato principalmente per la diffusione di contenuti multimediali tramite CD-ROM a singola velocità: infatti esiste un insieme di parametri (*constrained parameter set*) pensato appositamente per questo tipo di applicazione che presenta un bitrate di circa 1,5 Mbps.

Tra i principali limiti di questo standard esso non prevede la modalità interlacciata e tutti i flussi audio e video elementari condividono la stessa base-tempi non consentendo così l'aggregazione di più programmi tra loro indipendenti. Infine non è previsto alcun supporto alla rivelazione di errori e alla perdita di informazioni che possono avvenire su canali geografici e collegamenti radio.

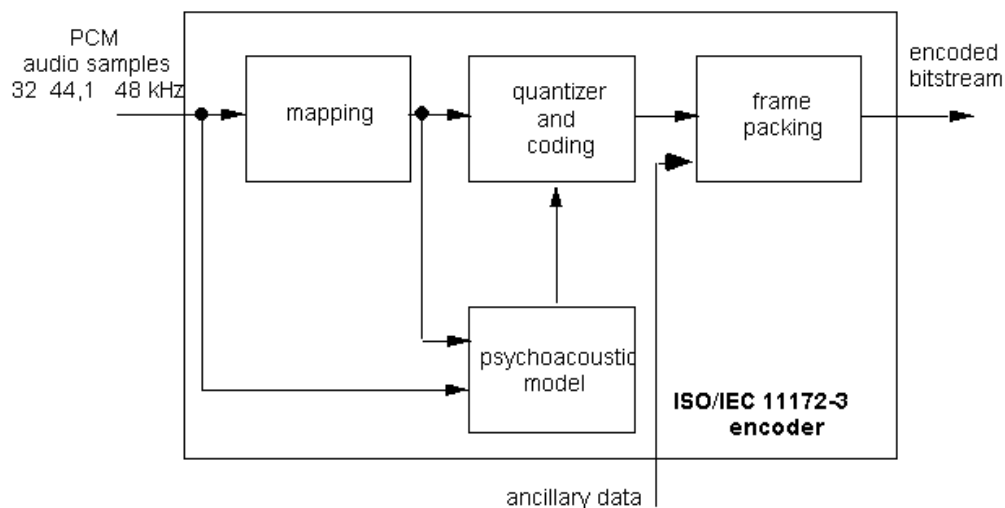
Lo standard MPEG-1 (il cui titolo ufficiale è "ISO/IEC 11172 Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s") è composto da 5 parti: le prime 3 parti sono diventate International Standard nel 1993, la parte 4 lo è diventata nel 1994, la parte 5 nel 1995.

- Parte 1 - System: descrive come sincronizzare diversi flussi audio e video e come trasportarli su canali digitali o supporti di massa;



**Figura 0.1: decoder MPEG**

- Parte 2 - Video: descrive la sintassi (header ed elementi del bitstream) e la semantica (il significato dei bit) del bitstream video. La sequenza di immagini è suddivisa in una serie annidata di livelli (*sequence, picture, slice, macroblock, block, coefficienti DCT*). La sintassi fornisce i mezzi per effettuare la sincronizzazione, l'accesso casuale, il dimensionamento dei buffer e il recupero degli errori. Al livello più alto (*sequenze layer*) è codificato il frame rate e le dimensioni dell'immagine.
- Parte 3 - Audio: descrive la sintassi e la semantica per 3 classi di metodi di compressione chiamati *layer I, II e III* e caratterizzati da una crescente efficienza nella compressione. Il layer I è simile allo standard usato da Sony MiniDisk e da Philips Digital Compact Cassette (DCC), il layer II è usato nella televisione via satellite mentre il layer III è usato su Internet e su ISDN per il suo basso bitrate (i ben noti file con estensione .mp3).



**Figura 0.2: encoder audio**

- Parte 4 - Conformità: definisce l'insieme dei test di conformità sui bitstream e sui decodificatori;
- Parte 5 - Simulazione software: contiene un esempio in linguaggio ANSI C di un codificatore e di un decodificatore software conformi allo standard e relativi alle parti audio, video e system.

### 2.1.3 MPEG Layer 1 e 2

Come detto in precedenza il gruppo MPEG ha proposto tre metodi di codifica, noti come *Livello* (*Layer* in inglese) *1*, *Livello 2* e *Livello 3*, con complessità e prestazioni crescenti.

Per ciascun livello lo standard specifica il formato del bitstream e il decodificatore. Non specifica rigidamente il codificatore per permettere diverse implementazioni future, limitandosi a fornire un codificatore per ogni livello solo a titolo esemplificativo.

Brevemente, il codificatore analizza le componenti spettrali del segnale audio per poter allocare il numero di bit che ha a disposizione, in modo da rispettare le specifiche imposte dal bitrate e dal mascheramento.

Il decodificatore è molto meno complesso, dato che il suo unico obiettivo è quello di sintetizzare un segnale audio a partire dalla codifica delle sue componenti spettrali.

Il bitstream contiene un header informativo che garantisce il rispetto della struttura gerarchica tra i livelli.

Per scendere nei dettagli i due layer eseguono i seguenti punti:

- **Analisi**

La prima necessità del codificatore è di analizzare il segnale nel campo delle frequenze. Il segnale viene scomposto in 32 sottobande tramite il banco di filtri. Da 512 campioni PCM vengono calcolati i 32 valori spettrali, sostituiti 32 campioni con tecnica FIFO e si procede nuovamente al calcolo. Ognuna di queste operazioni produce 32 campioni in uscita. Il layer I esegue un finestramento ogni 12 passi mentre il layer II ogni 36. Si hanno quindi 384 (32\*12) per finestra nel primo e 1152 (32\*36) nel secondo. Si noti che questi sono campioni spettrali, nel dominio delle frequenze e non più in quello del tempo.

- **Modello Percettivo**

Si occupa di generare per ogni pacchetto la soglia globale di mascheramento, che sarà utilizzata in fase di campionamento per eliminare le componenti frequenziali non necessarie. Questa è un'operazione alquanto onerosa, dapprima si trasforma il segnale con una FFT, poi si eseguono una serie di passaggi all'interno del modello psicoacustico. Non viene eseguita nessuna operazione di pre e post mascheramento.

- **Quantizzazione e codifica**

La quantizzazione è un'operazione che inevitabilmente introduce rumore. Durante quest'operazione i singoli campioni vengono approssimati al valore di quantizzazione più vicino, è evidente che un segnale quantizzato non potrà più esser riportato nella sua forma originale. Occorre quindi limitare il più possibile il rumore introdotto, chiamato rumore di quantizzazione ed ottenibile come differenza fra il segnale originario e quello modificato. Il numero di livelli di quantizzazione è ottenuto, per ogni componente spettrale, da un algoritmo d'allocazione dinamica dei bit. Questo sceglie il quantizzatore all'interno di quelli disponibili, lavorando in modo da ottenere la bitrate richiesta e da rispettare le imposizioni della soglia globale di mascheramento.

- **Decodifica**

Il decodificatore è sicuramente più semplice del codificatore, dovendo utilizzare meno risorse

possibili siano esse software o hardware. I dati sono dequantizzati tenendo conto del fattore di scala e dei bit allocati. Sono quindi ricostruite le sottobande, se una di queste non ha nessun bit allocato viene trascurata. A questo punto si applica il blocco di sintesi, nel caso sia realizzato via software è spesso lo stesso del codificatore fatto lavorare all'inverso, e si ricostruisce il formato audio PCM.

Qui di seguito è illustrata lo schema generale di codifica e decodifica di MPEG-1

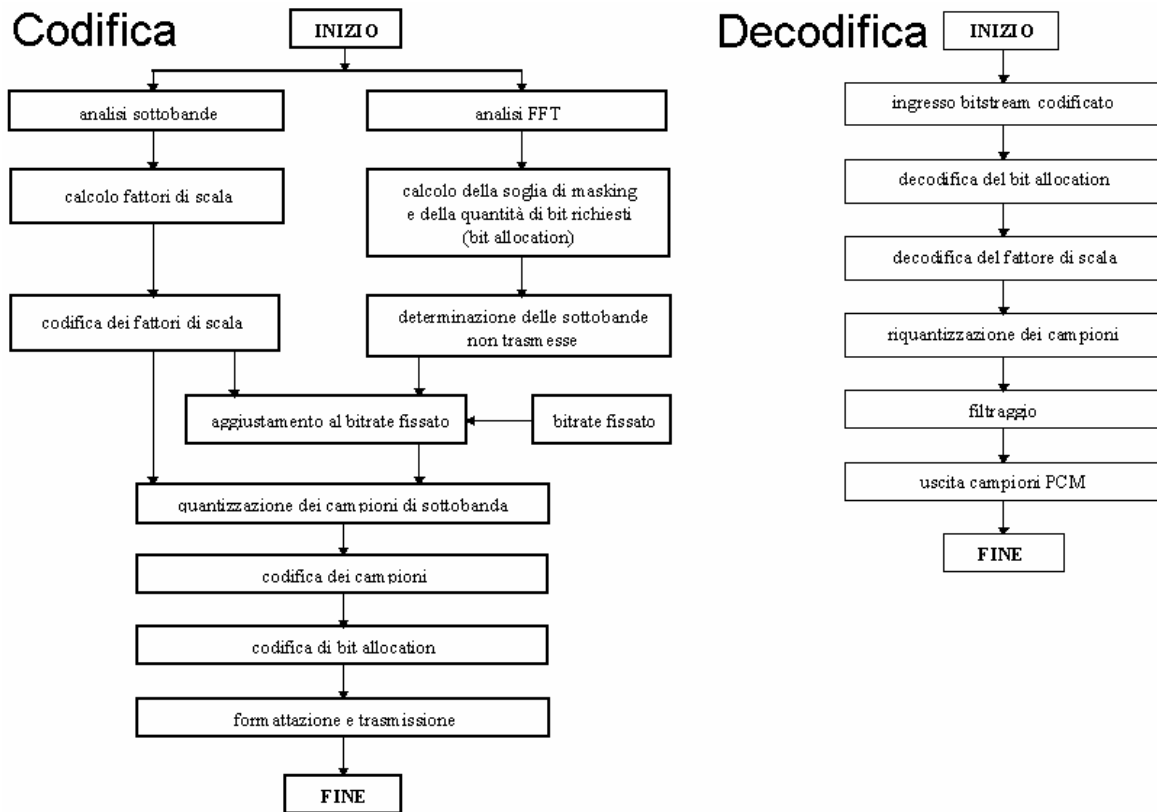


Figura 0.3: processo di codifica e decodifica in MPEG

Per quanto riguarda i due livelli, il *Livello 1* contiene il mapping di base dell'ingresso audio digitale (in 32 sottobande, divisione stabilita per la formattazione dei dati), un modello psicoacustico che determina l'allocazione ottimale dei bit.

Il *Livello 2*, oltre alle caratteristiche fondamentali del livello precedente, fornisce una codifica addizionale per l'allocazione dei bit, e una serie di fattori di scala per l'elaborazione dei campioni del segnale. La struttura del frame è diversa da quella del livello precedente e contiene un numero maggiore di campi.

## 2.1.4 L'MPEG layer 3 (MP3)



Il Layer III, più comunemente noto come MP3, introduce nuove tecniche al fine di migliorare la qualità dell'audio e mantenere accettabile la bitrate. Come si può vedere in figura lo schema di principio è molto simile a quello dei due layer precedenti, i blocchi modificati sono essenzialmente tre.

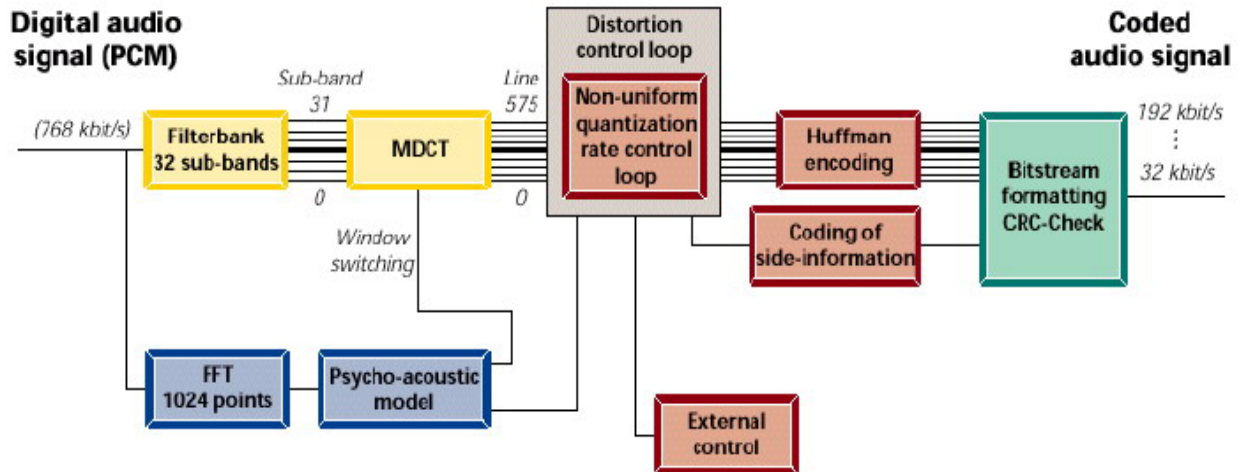


Figura 0.4: encoder MP3

- **Analisi**

Questo blocco si basa sulla stessa tecnologia dei precedenti ma introduce un blocco aggiunto per ovviare ai problemi del banco di filtri. Questi, infatti, non possono essere perfetti e introducono delle sovrapposizioni fra le sottobande (aliasing). L'aliasing può portare a fenomeni di distorsione, per evitare ciò, l'uscita del banco di filtri non è più passata direttamente al quantizzatore ma viene elaborata da un blocco MDCT (Modified Discrete Cosine Transform) a 6 o 18 punti. In realtà MDCT effettua una sovrapposizione del 50% fra finestre successive quindi in uscita ci sono ancora 1152 campioni per finestra. Il blocco a 18 punti offre una maggiore risoluzione frequenziale mentre il blocco a 6 è decisamente più rapido e viene utilizzato quando ci si aspetta un pre eco. Il passaggio da 18 a 6 e viceversa è possibile solo introducendo due particolari finestre: long-to-short e short-to-long data window.

- **Modello Percettivo**

A differenza dei primi due si effettua sia il pre sia il post mascheramento.

- **Quantizzazione e codifica**

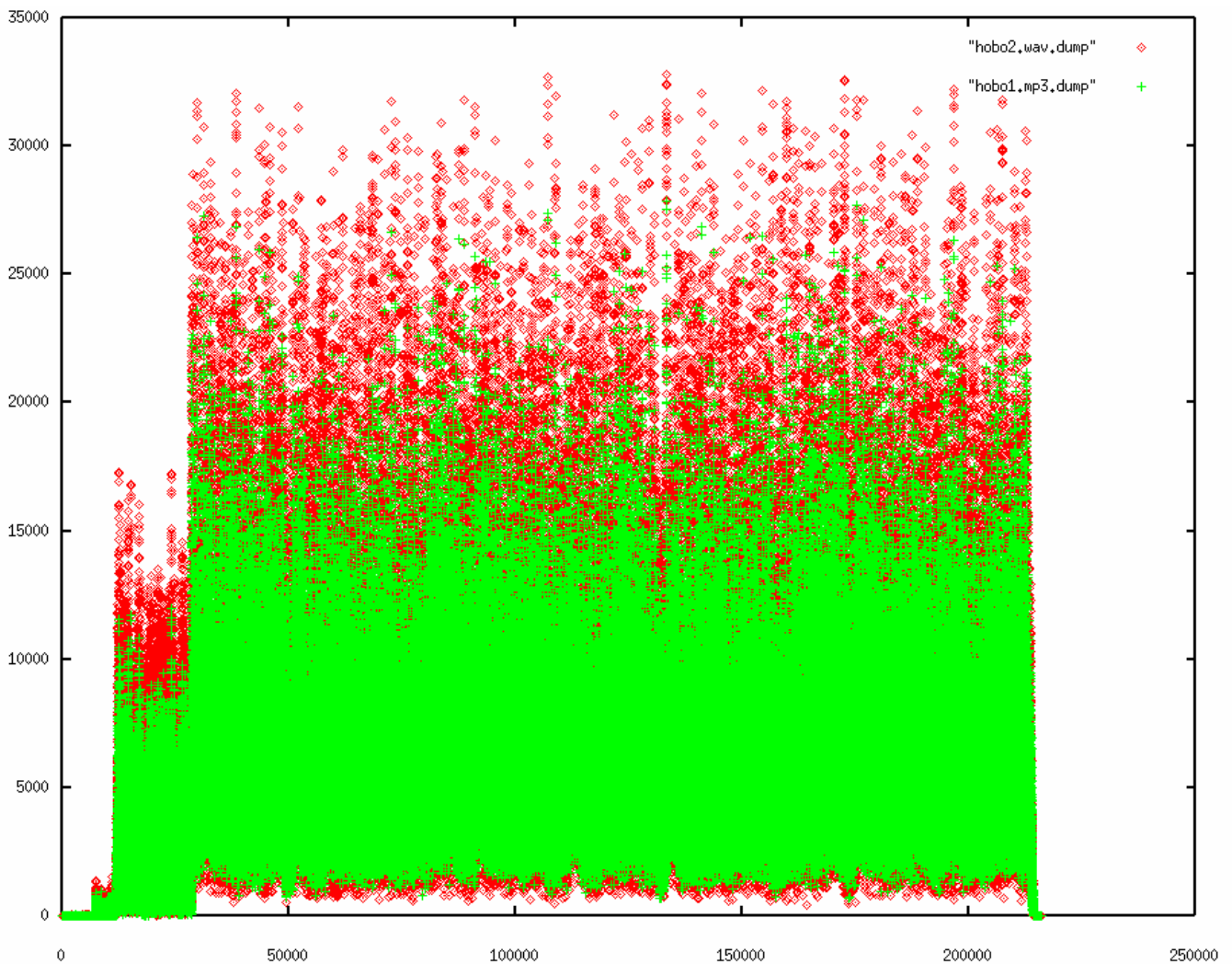
Per definire il quantizzatore occorre scegliere il valore del guadagno ed il fattore di scala. Questi due parametri devono consentire di minimizzare il rumore introdotto e di mantenere accettabili le dimensioni del blocco all'uscita del codificatore. Si procede quindi con due cicli innestati:

### *Rate loop*

Vengono assegnati i bit ad ogni codice secondo la tecnica Huffman. Se le dimensioni del blocco così codificato sono troppo grandi si interviene diminuendo il guadagno. Così facendo si diminuisce il numero di Byte necessari alla codifica. Si tenga presente che i livelli di quantizzazione piccoli sono i più frequenti e quindi codificati con un numero minore di bit.

### *Noise control loop*

Viene applicato un fattore di scala ad ogni sottobanda, valore di partenza è 1 per tutte le sottobande. Si calcola il rumore di quantizzazione e se questo supera il livello della soglia globale di mascheramento si procede ad una variazione dei fattori di scala. A questo punto è necessario ripetere il rate loop. Il ciclo procede finché non si è riduce il rumore di quantizzazione ad un livello accettabile o venga interrotto da sistemi di sicurezza per evitare loop infiniti. Questi possono capitare se il modello percettivo richiede passi di quantizzazione talmente piccoli da far sì che il rate loop sia costretto ad aumentare il guadagno.



**Figura 0.5: confronto tra WAV (rosso) e MP3 (Verde)**

Il Layer 3, grazie alla codifica qui descritta, permette una notevole riduzione sull'occupazione di spazio per un segnale audio. Nella tabella sotto sono riportati alcuni dati che mostrano le prestazioni della codifica MP3. E' da notare che i dati non sono "assoluti", la riduzione varia in funzione del suono da comprimere e dal codificatore utilizzato. I dati riportati sono stati recuperati dal Fraunhofer Institut e si riferiscono al loro codificatore.

Qualità sonora	Larghezza di banda	Modalità	Bitrate	Rapporto di compressione
Suono telefonico	2.5 kHz	mono	8 kbps	96:1
Radio a onde corte	4.5 kHz	mono	16 kbps	48:1
Migliore della radio AM	7.5 kHz	mono	32 kbps	24:1
Simile alla radio FM	11 kHz	stereo	56...64 kbps	26...24:1
Simile al CD	15 kHz	stereo	96 kbps	16:1
CD	>15 kHz	stereo	112..128kbps	14..12:1

Figura 0.6: dati sulle prestazioni della codifica MP3

### 2.1.5 Il formato dei dati


Oltre alle informazioni compresse ogni finestra deve contenere alcuni dati necessari alla decodifica. Questi, inseriti all'inizio d'ogni frame, sono:

- **Sync word o CRC:** È una “parola“ di controllo, serve per controllare che non ci siano stati errori di trasmissione. Ci si aspetta una di queste all'inizio d'ogni frame, quindi ad una distanza predeterminata l'una dall'altra, se non c'è risulta evidente l'errore.
- **Frequenza di campionamento.**
- **Bitrate.**
- **Layer:** Specifica con quale layer sono stati compressi i dati.
- **Copy protection:** Sono due bit per la SCMS (Serial Copy Management Scheme). La loro utilità non è del tutto chiara, anche perchè possono esser manipolati via software senza danneggiare il file.
- **Modalità di codifica.**

Il layer III usa un'interessante tecnica (*bit reservoir*) per risparmiare ogni singolo bit. Se la dimensione dei dati in una finestra è inferiore al numero di bit a lei assegnata, il codificatore usa questo spazio vuoto per inserire dati relativi alle finestre successive. Così può accadere che in una finestra ci siano l'header di un pacchetto ed i dati sia del pacchetto corrente sia di quello successivo.

## 2.2 MPEG-2

### 2.2.1 Lo Standard MPEG-2

 MPEG-2 è un'estensione dell'MPEG-1, standard internazionale per la compressione digitale del segnale audio e video. MPEG-2 è stato creato per la diffusione in broadcast che supportino alti bitrate, quindi è in grado di codificare il segnale televisivo a una definizione standard con bitrate a circa 3-15Mbit/s e il segnale televisivo a alta definizione con bitrate da 15 a 30 Mbit/s; fornisce algoritmi che sono strumenti extra per codificare efficacemente il video e l'audio, supportando un'ampia gamma di bitrate e fornendo la possibilità di avere più canali audio (e in grado di generare l'effetto “surround”).

Per quanto riguarda MPEG-2 (il cui titolo ufficiale è “ISO/IEC 13818 Information technology – Generic coding of moving pictures and associated audio”) esso è organizzato come il suo predecessore per quanto riguarda le prime 5 parti a cui se ne aggiungono altre 3

- Parte 6 – Digital Storage Medium Command and Control (DSM-CC): fornisce una sintassi per controllare l'esecuzione e l'accesso casuale come avviene nei videoregistratori (*fermo immagine, avanti veloce, riavvolgimento*).

- Parte 7 – Non-backward compatible audio: descrive nuove estensioni della codifica audio che, a differenza della Parte 3, non sono compatibili con MPEG-1.
- Parte 8 – 10-bit video extension: è una parte molto giovane dello standard che riguarda la codifica su 10 bit dei coefficienti DCT per una migliore qualità da usarsi in ambito professionale e per l'archiviazione.

## 2.2.2 L'Advanced Audio Coding (AAC)



AAC significa Advanced Audio Coding ed è un formato di codifica per il *general audio* ad alta qualità, principalmente indicato per la codifica multicanale (come il Dolby Digital 5.1). Dichiarato standard nel 1997 (ISO/IEC 13818-7:2003), AAC apparteneva alla settima parte dello standard MPEG-2, in quanto non compatibile all'indietro con i formati MPEG già esistenti, al contrario di quanto definito nello standard MPEG-1 e MPEG-2 BC (Backward Compatible). La motivazione della non compatibilità va ricercata negli enormi svantaggi che ci sarebbero stati a mantenersi conforme agli standard precedenti, svantaggi legati sia al tasso di compressione che alla qualità audio ottenuta.

Qui di seguito sono elencate le caratteristiche tecniche principali del formato di codifica audio AAC:

- Frequenza di campionamento da 8 a 96 kHz
- Bitrate da 8 a 512 Kbit/sec
- Fino a 48 canali full-bandwidth più 15 canali per la codifica Low Frequency Enhancement (LFE)
- Fino a 15 data stream contemporanei

AAC, come gli standard audio delle versioni precedenti di MPEG, segue uno schema di codifica percettivo eliminando tutte quelle informazioni ritenute non necessarie, in quanto non percepite dal nostro orecchio a causa del già discusso effetto di mascheramento.

Come le due precedenti versioni di MPEG, il segnale audio nel dominio delle frequenze viene suddiviso in bande e successivamente esegue una quantizzazione lineare guidata dal modello psicoacustico, associando pochi bit alle bande meno percepite.

Ciò che lo distingue dai predecessori (soprattutto l'MP3) va ricercato nella struttura dell'ENCODER:

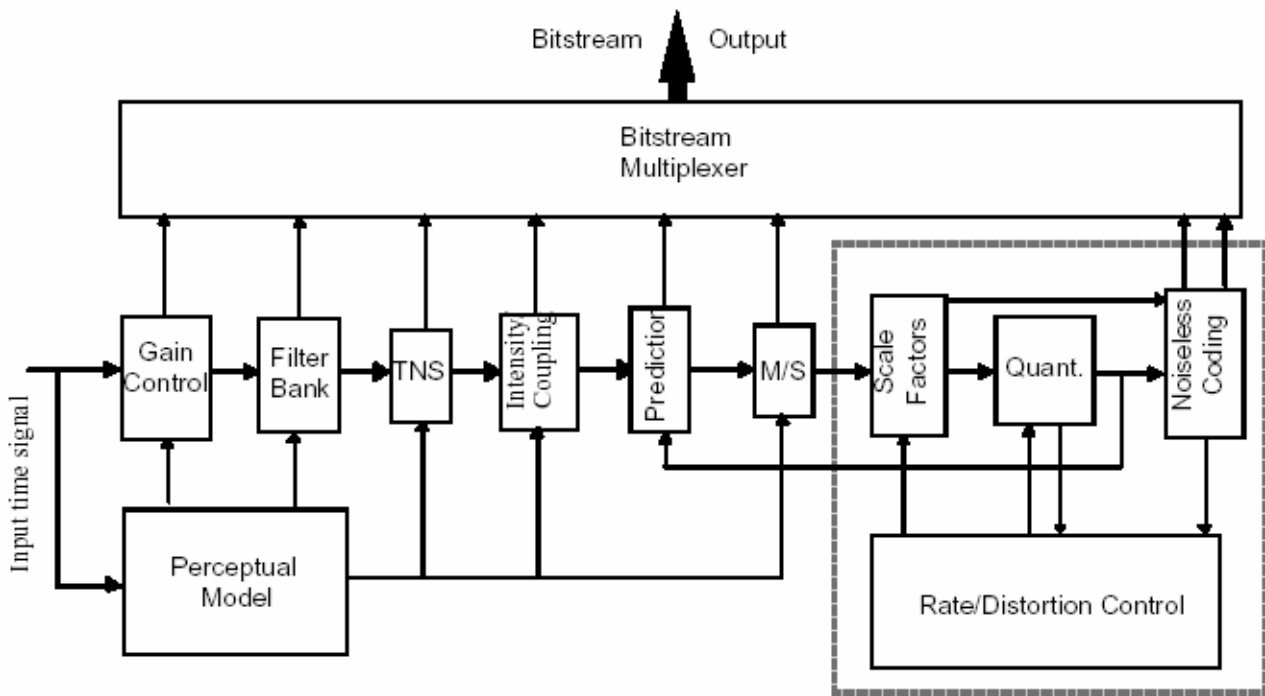


Figura 0.7: codificatore MPEG-2 AAC

Questo prende in input un segnale non compresso (PCM) e lo trasforma nel corrispondente segnale AAC attraverso i seguenti blocchi logici (chiamati *tools*):

- **Banco filtri (*Filter Bank*):** rispetto al Banco Filtri Ibrido (Hybrid Filter Bank) che sarebbe stato necessario per mantenere la compatibilità con le versioni precedenti di MPEG (e che veniva quindi utilizzato anche nell'MP3), AAC utilizza solo la Trasformata Coseno Discreta Modificata (MDCT) eliminando quindi il Banco Filtri Polifasico (Polyphase Filter Bank) che introduceva tutta una serie di imprecisioni (ineliminabili in fase di codifica) nella rappresentazione del segnale nel dominio delle frequenze.
- **Temporal Noise Shaping (*TNS*):** rappresenta una delle novità importanti introdotte negli schemi di codifica Time/Frequency: esso predice la distribuzione del rumore di quantizzazione nel tempo lavorando nel dominio delle frequenze. Permette di ottenere evidenti miglioramenti nella codifica audio vocale.
- **Intensity/Coupling:** questo tool ricade in due categorie: intensità stereo (IS) e accoppiamento. Il primo è usato per implementare la modalità Stereo Congiunto (*Joint Stereo*) fra due canali, quindi i dati in uscita da due canali sono derivati da un unico insieme di coefficienti spettrali dopo il processo inverso della quantizzazione. Questa categoria permette anche di codificare i segnali a bassa frequenza (la cui direzione non viene percepita dall'orecchio umano) in mono anche se il resto del segnale fosse in stereo. L'accoppiamento di canali invece serve per permettere una divisione dinamica di un segnale sonoro su più canali o il disaccoppiamento di una parte del segnale sonoro da un canale.
- **Prediction:** utilizza un algoritmo di predizione non presente in MPEG Layer 3; è principalmente indicato per la codifica di segnali vocali in quanto maggiormente stazionari nel tempo e quindi più facilmente predicabili

- **Quantizzazione:** il sistema di controllo per l'allocazione dei bit è stato migliorato dalla sua versione precedente e permette un utilizzo più efficiente e mirato del valore di bitrate
- **Scale Factors:** come nella versione precedente di MPEG sono presenti i fattori di scala in grado di pesare opportunamente le varie bande in funzione della loro importanza, raggruppandole in Bande Critiche
- **Mid/Side Stereo:** altri algoritmi di codifica stereo compressa e multicanale ulteriormente ottimizzati e migliorati rispetto a quelli presenti in MP3 per il supporto multicanale
- **Noiseless Coding e Bitstream Multiplex:** come in MP3 si utilizza una codifica tramite algoritmo di Huffman per minimizzare ulteriormente le ridondanze presenti nella sequenza numerica ottenuta dalla quantizzazione. Qui è però possibile costruire un bitstream a struttura variabile, ottenendo così una maggiore efficienza in fase di streaming.
- **Gain Control (GC o più spesso AGC – Automated GC):** questa unità si trova spesso nei sistemi elettronici di diversi apparecchi. Negli apparecchi audio (e quindi nell'encoder e decoder AAC) serve per avere controllo sul rapporto tra il segnale in entrata e il segnale in uscita. Questo perché il segnale, entrando nel sistema deve avere determinate caratteristiche di ampiezza (nelle radio FM è importante anche per la frequenza portante) che devono essere mantenute anche quando il segnale verrà riprodotto, per mantenerne la fedeltà.

Per quanto riguarda il DECODER, i blocchi logici (o *tools*) che lo compongono lavorano in maniera inversa rispetto al decoder, trasformando così il segnale in formato AAC nel corrispondente formato PCM

# MPEG-2 AAC Decoder overview

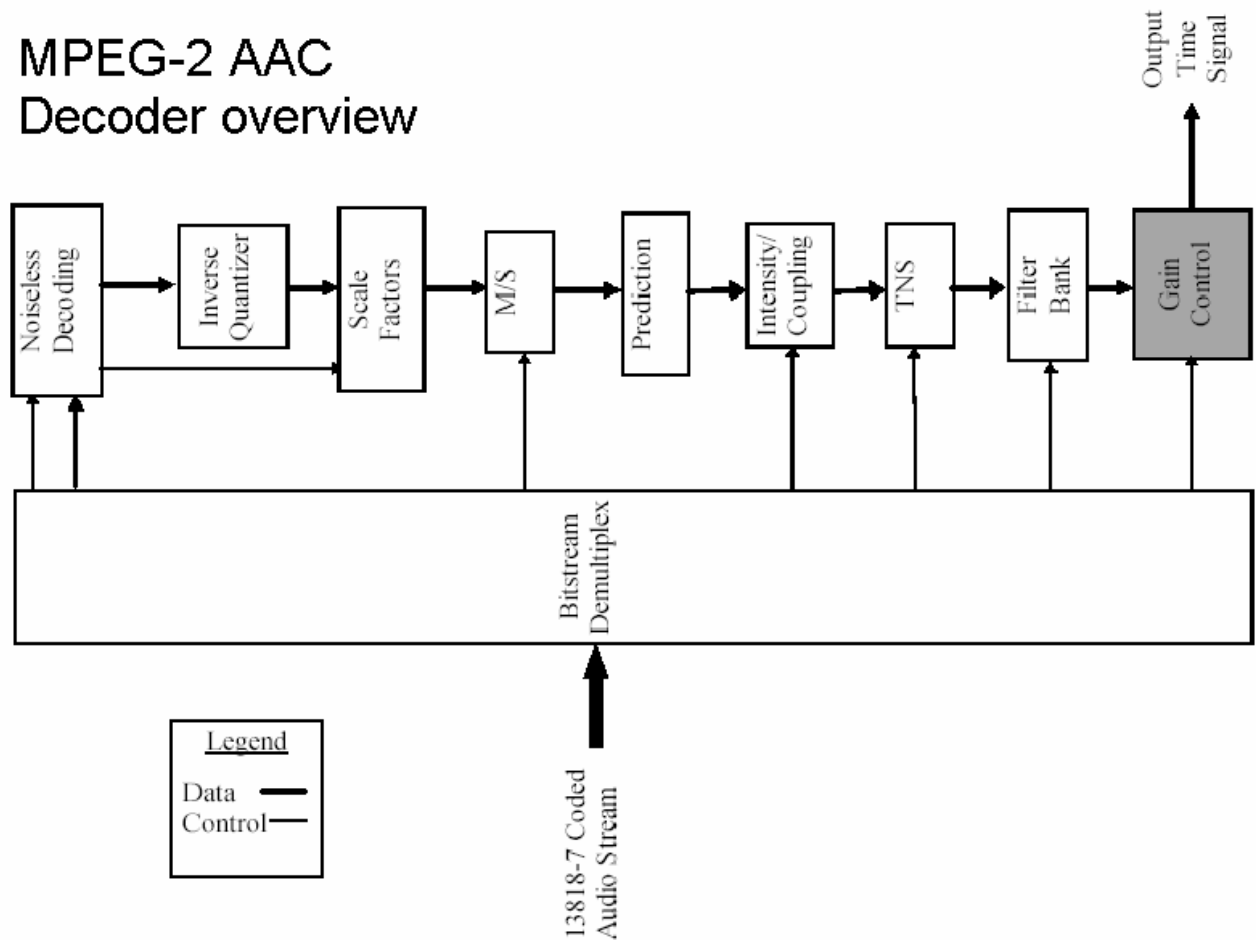


Figura 0.8: decodificatore MPEG-2 AAC

In fase di encoding-decoding non è necessario utilizzare tutti i blocchi logici (*tools*) presenti nello schema di codifica ma solo quelli ritenuti più necessari in funzione della natura e tipologia di segnale audio che si sta andando a codificare. Nella tabella sottostante sono elencati i blocchi obbligatori e opzionali

Tool Name	Required / Optional
Bitstream Formatter	Required
Noiseless Decoding	Required
Inverse Quantizer	Required
Scalefactors	Required
M/S	Optional
Prediction	Optional
Intensity/Coupling	Optional
TNS	Optional
Filterbank	Required
Gain Control	Optional

Figura 0.9: Tool necessari e richiesti per la codifica e decodifica AAC

### 2.2.3 Tipi di codifica AAC

Come abbiamo detto sia in fasi di codifica che di decodifica lo standard prevede che possono essere utilizzati solo un sottoinsieme di tutti i **tool** descritti, scegliendoli in funzione delle risorse disponibili, della tipologia del segnale o di altri parametri significativi.

Per questo motivo lo standard AAC di MPEG-2 fornisce diverse configurazioni (chiamate *Profiles*) che vado ad analizzare brevemente:

- **Profile Main:** questo tipo di profilo viene utilizzato in presenza di grandi quantità di memoria disponibile e elevate capacità di elaborazione, in quanto utilizza tutti i tools presenti nell'encoder ad eccezione del "gain control" e praticamente effettua una codifica predittiva adattiva "all'indietro" con l'obiettivo di eliminare le ridondanze tra i vari frames (tale codifica ha efficacia maggiore sui segnali stazionari). Utilizzando però la predizione "all'indietro" la codifica ottenuta è molto sensibile agli errori.
- **Profile Low Complexity (LC):** questo profilo invece viene utilizzato quando si hanno a disposizione scarse risorse. Non vengono utilizzati tutta una serie di tools tra cui il Gain Control e il Prediction, mentre il Temporal Noise Shaping (TNS) viene usato ma limitatamente. Spesso viene preferito questo profilo perché il rapporto qualità del segnale su costo computazionale in questo caso è migliore
- **Profile Scalable Sample Rate (SSR):** questo profilo necessita obbligatoriamente del Gain Control, vieta la codifica multicanale e limita l'uso del TNS. Questa tecnica introdotta dalla Sony consiste nella suddivisione del segnale in 4 bande con un banco filtri PQF (Polifase Quadrature Filter, che divide il segnale nel numero di bande equidistanti stabilito e per ogni coppia una delle bande viene inversa in frequenza, ovvero le frequenze basse diventano alte e vice versa), e ognuna delle 4 sottobande viene ulteriormente suddivisa con la trasformata coseno discreta modificata (MDCT). Il vantaggio di questa tecnica sta nel fatto di poter scegliere, in modo indipendente per ogni sottobanda, se utilizzare un blocco lungo (con una maggiore risoluzione a livello di frequenza) o uno corto (con una maggiore risoluzione per quanto riguarda la temporizzazione). Inoltre, eliminando le varie sottobande inutili ottenute col banco filtri PQF, è possibile ottenere una compressione maggiore.

### 2.2.3 MP3 e AAC a confronto

Numerosi test acustici sono stati effettuati sullo standard AAC per capire realmente se e di quanto migliorasse la qualità audio rispetto alle precedenti codifiche.

A parità di condizioni (stesso valore di bitrate e frequenza di campionamento, medesimi brani audio...) i risultati hanno dimostrato come le codifiche AAC mono, stereo e multicanale siano qualitativamente migliori rispetto alle stesse codifiche MPEG Layer 2 e Layer 3. Più precisamente è stato dimostrato come una codifica audio AAC con bitrate a 96 Kbit/sec è paragonabile, in termini di qualità, ad un MP3 a 128 Kbit/sec (guadagno del 30%) ed un MP2 (MPEG Layer 2) a 192 Kbit/sec (guadagno del 100%).

Folder	Format	Name	Size
Lost Horizons	AAC (m4a)	01 Elements.m4a	8 MB
		02 Spacewalk.m4a	6.5 MB
		03 Ramblin' Man.m4a	6.5 MB
		04 Return To Patagonia.m4a	8 MB
		05 Nice Weather For Ducks.m4a	5.6 MB
		06 Experiment No. 6.m4a	5.4 MB
		07 Closer.m4a	6.8 MB
		08 The Curse Of Ka'Zar.m4a	8.3 MB
Lost Horizons (MP3)	MP3	01 Elements.mp3	12 MB
		02 Spacewalk.mp3	9.7 MB
		03 Ramblin' Man.mp3	9.8 MB
		04 Return To Patagonia.mp3	12 MB
		05 Nice Weather For Ducks.mp3	8.5 MB
		06 Experiment No. 6.mp3	8.2 MB
		07 Closer.mp3	10.2 MB
		08 The Curse Of Ka'Zar.mp3	12.5 MB
Spirit Of The Sun	AAC (m4a)	01 The Dawn Introduction.m4a	1.3 MB
		02 The Brightness Of These Days.m4a	5.7 MB
		03 Mind Expansions.m4a	7.5 MB
		04 Deep In Your Mind.m4a	6.8 MB
		05 Stargazer.m4a	6.2 MB
		06 Eclipse.m4a	7 MB
		07 Between The Lights.m4a	6.1 MB
		08 Shine.m4a	6.6 MB
		09 Substream.m4a	6 MB
		10 Behind The Shadow.m4a	7.5 MB
		11 M E Outroduction.m4a	1.8 MB
Spirit Of The Sun (MP3)	MP3	01 The Dawn Introduction.mp3	1.9 MB
		02 The Brightness Of These Days.mp3	8.5 MB
		03 Mind Expansions.mp3	11.1 MB
		04 Deep In Your Mind.mp3	10.1 MB
		05 Stargazer.mp3	9.2 MB
		06 Eclipse.mp3	10.4 MB
		07 Between The Lights.mp3	9 MB
		08 Shine.mp3	9.8 MB
		09 Substream.mp3	9 MB
		10 Behind The Shadow.mp3	11.1 MB
		11 M E Outroduction.mp3	2.6 MB

Figura 0.10: esempio di risparmio di spazio tra AAC e MP3

## 2.3 MPEG-4

### 2.3.1 Le caratteristiche dell'MPEG-4



L'attività del gruppo MPEG prosegue e nel 1998 viene approvato un nuovo standard - MPEG-4 -che codifica un'immagine a partire dagli oggetti di cui è composta e permette, quindi, sia al creatore dell'immagine sia all'utilizzatore finale di interagire direttamente con gli elementi che la compongono.

Gli oggetti dell'immagine, infatti, mantengono una loro individualità sia nella fase di codifica che in quella di rappresentazione finale. Pensiamo, ad esempio, ad una immagine composta dai seguenti oggetti: una via cittadina ed una donna con il suo bambino che camminano lungo la strada. Una volta che le componenti dell'immagine sono state trasformate in bit non sarà necessario ritrasmettere più volte le componenti invarianti (la strada, le case, le automobili parcheggiate...), sarà sufficiente trasmettere quelle che cambiano, nella scena qui descritta basterà trasmettere i movimenti della donna e del bambino che camminano. In questo modo, MPEG-4 consente di ridurre la banda necessaria al trasporto di informazioni audio-visive di elevata qualità, garantendo la trasmissione su qualunque infrastruttura di comunicazione (satellite, cavo, fibra ottica, rete mobile) grazie alle sue caratteristiche di ottimo rapporto in termini di qualità e compressione del segnale.

Inoltre, MPEG-4 apre spazi di creatività nel campo multimediale e televisivo, consentendo di personalizzare i contenuti trasmessi. Nell'esempio visto, pensiamo alla possibilità di manipolare le immagini modificando un singolo elemento della scena senza interferire sugli altri oggetti, per esempio sostituendo un personaggio con un altro.

Questo è l'aspetto veramente innovativo di MPEG-4: qualsiasi filmato può essere arricchito di informazioni aggiuntive quali immagini fisse, videoclip, sorgenti audio, che vengono attivate grazie alla presenza di oggetti virtuali cliccabili e navigabili come su Internet. Si realizza ora pienamente il concetto di interattività: si può così far scomparire dalla scena di un film un'auto e sostituirla con un altro veicolo od un determinato personaggio per intervenire sull'evoluzione della trama. Oppure si possono selezionare alcuni elementi che si vogliono seguire con più attenzione all'interno di un contenuto multimediale, ad esempio nella trasmissione di una sinfonia si può separare un particolare strumento musicale dagli altri per ascoltare un assolo. Od ancora si possono realizzare telegiornali interattivi arricchiti in tempo reale da livelli aggiuntivi di informazioni come approfondimenti storici, geografici, linguistici, attivabili solo su richiesta dello spettatore in base agli argomenti che più lo interessano.

### 2.3.2 Differenze tra AAC dell'MPEG-2 e dell'MPEG-4

Come abbiamo visto l'Advanced Audio Coding è stato introdotto nello standard MPEG-2. Nell'MPEG-4 viene ripreso, ma senza introdurre sostanziali differenze rispetto al predecessore, soprattutto per quanto riguarda la qualità del suono. Infatti gli algoritmi di codifica sono gli stessi, mentre le uniche differenze riguardano:

- l'header strutturato diversamente per contenere indicazioni sul formato MPEG-4;
- il maggior numero di configurazioni per quanto riguarda l'encoder rispetto alla versione precedente
- la diversa nomenclatura per le configurazioni dell'encoder
- due tools: il PNS e il LTP

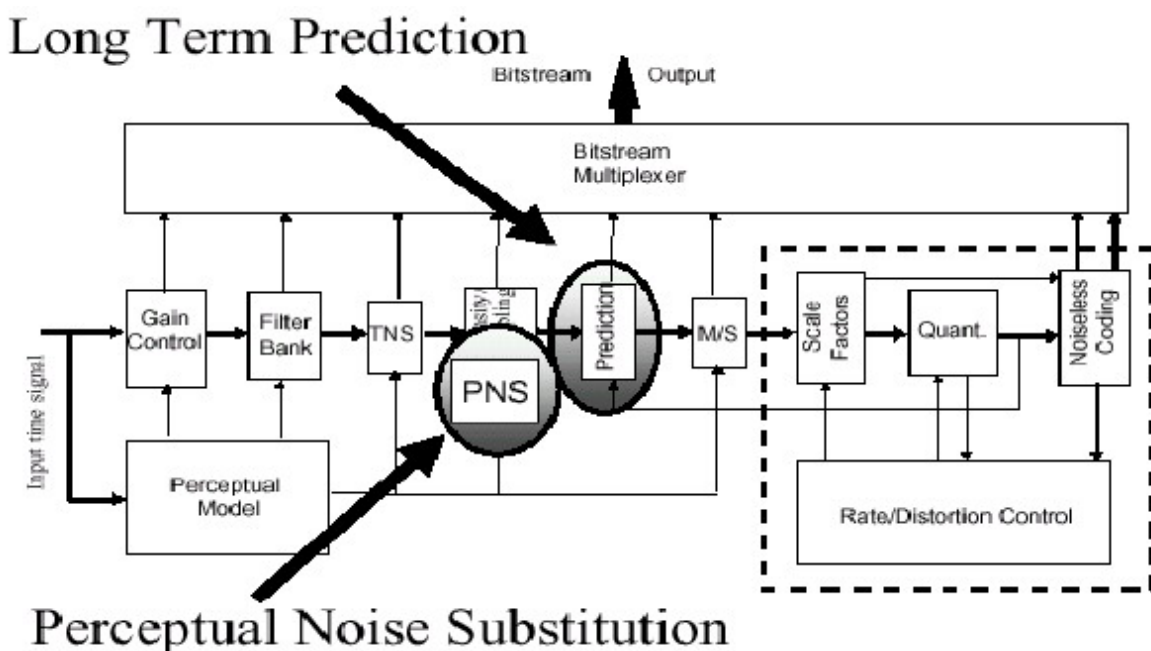


Figura 0.11: i nuovi tools dell'AAC MPEG-4 nello schema di codifica

Per quanto riguarda il terzo punto, il nome delle configurazioni dell'encoder sarà ora *Object Types*, e non più *Profiles* come nell'MPEG-2. Gli Object Types identificano anche varie tecniche utilizzate per migliorare l'efficienza della codifica e decodifica e altri tools che vennero introdotti nelle versioni successive di MPEG-4. La differenza di terminologia è legata profondamente alla caratteristica dell'MPEG-4 discussa in precedenza, ovvero la strutturazione ad oggetti.

Per il quarto punto i due nuovi tools introdotti fin dall'origine dello standard MPEG-4 sono:

- **Long Term Prediction (LTP):** ha come obiettivo quello di ridurre drasticamente le ridondanze di segnale presente tra frame successivi. Utilizza una tecnica predittiva “in avanti” e lavora molto bene a bassi bitrate e con segnali “tonali” (il cui pitch è facilmente identificabile). Riduce circa del 50% la complessità algoritmica rispetto all’Object Type Main, ed utilizzando una predizione “in avanti” è meno sensibile agli errori.
- **Perceptual Noise Substitution (PNS):** è basata sul principio “*un rumore suona come un altro*”, quindi perché codificare i suoni la cui immagine spettrale è assimilabile al rumore? Il PNS individua le regioni rumorose dello spettro le elimina ed inserisce nella bitstream l’intensità del “rumore” eliminato. Il decoder è realizzato in modo da poter ricostruire la parte eliminata dalla sua intensità e larghezza.

### 2.3.3 Gli Object Types dell’AAC in MPEG-4

Per quanto riguarda gli Object Types Main e Low Complexity non ci son state modifiche di alcun tipo riguardo ai Profiles corrispondenti nell’MPEG-2, mentre per l’SSR viene utilizzato un algoritmo lievemente modificato che permette una maggiore efficienza rispetto al corrispondente dello standard precedente.

In MPEG-4 vengono però introdotti altri Object Types. Vennero introdotti principalmente per permettere la trasmissione del segnale su reti a banda variabile (ossia dove la capacità del canale non è nota a priori e non resta costante per tutto il tempo di trasmissione) e fortemente soggette a errori di trasmissione (per esempio reti wireless). Gli Object Types sono:

- **Low Delay (LD):** permette di comprimere (e decomprimere) un segnale in formato AAC con una buona qualità audio ma ad una velocità di esecuzione molto maggiore rispetto agli altri Object Types forniti, riducendo di fatto i tempi di latenza dell’encoder e del decoder. Ciò permette la trasmissione di streaming AAC in tempo reale su reti overIP o wireless.
- **Bit Sliced Arithmetic Coding (BSAC):** implementa una delle funzionalità fornite da MPEG-4: il Bitrate Scalability. Praticamente questo sistema dà la possibilità di adattare il valore del bitrate in fase di codifica e decodifica in funzione delle esigenze di banda, memoria etc etc. In fase di codifica il bitstream viene compresso in varie fasi, ognuna delle quali codifica parti del segnale via via più importanti, utilizzando tecniche diverse. In fase di decodifica si avrà così a disposizione un bitstream suddiviso in varie sezioni, decodificabili in modo indipendente, e la scelta di quelle da decomprimere verrà fatta in funzione delle capacità di banda e di memoria, tutto ciò mantenendo elevata la qualità audio.
- **Spectral Band Replicator (SBR):** serve a ridurre il bitrate del segnale audio compresso cercando di mantenere inalterata la qualità acustica. Si cercano infatti di eliminare le alte frequenze di un segnale audio musicale e/o vocale in quanto generalmente contengono rumore bianco, le armoniche di una nota o le formanti di un parlato (che generalmente hanno fondamentale o pitch situati nelle basse frequenze). Questo viene fatto analizzando opportunamente lo spettro del segnale da trasmettere durante la codifica, e cancellando le frequenze inutili appena descritte, salvando nella Side Information (porzione del file destinato al contenimento di speciali metadati) le informazioni per ricostruire queste frequenze, ovviamente mettendo soltanto dei parametri, la cui dimensione in termini di occupazione di spazio fisico è estremamente ridotta. Quindi in fase di decodifica si ottiene normalmente la parte di spettro non eliminata, mentre quella eliminata viene sintetizzata attraverso la lettura dei parametri della Side Information. Questa tecnica, molto efficace, viene utilizzata anche in una estensione dell’ MP3 chiamata MP3PRO.
- **High Efficiency AAC (HE AAC):** è il nome dato dal comitato MPEG alla tecnica che combina la normale codifica AAC con la SBR in maniera efficiente. E’ stata resa ufficiale nel maggio 2003 ma viene anche denominata con altre terminologie non ufficiali come “AAC+”.

## 2.4 Trattamento e gestione degli errori

Uno degli ambiti per cui l'AAC è stato pensato è lo streaming audio in tempo reale. La moderna tecnologia delle reti di comunicazione si divide in "packet oriented" (come ad esempio Internet) o "stream oriented" (ne fanno parte le reti Wireless o i sistemi digitali di broadcasting). Si sa che in questi sistemi la probabilità di perdere dati o di danneggiarli è consistente, al punto che esistono sofisticate tecniche per individuare e correggere gli errori. Tuttavia al caso estremo, per quanto riguarda i dati è possibile richiedere la ritrasmissione dell'informazione danneggiata.

Ovviamente non si può fare questo quando i pacchetti compongono uno stream audio, perché il segnale risulterebbe interrotto. Per fare un esempio concreto non è possibile bloccare una telefonata perché un pacchetto audio è andato perso. In genere si preferisce sostituirlo con silenzio o rumore bianco generato sinteticamente.

Quindi in genere la ricezione di un pacchetto danneggiato introduce in fase di ascolto un suono indesiderato che riduce di molto la qualità audio. Tuttavia è sempre possibile fare affidamento alle tecniche di correzione di cui parlavamo prima, per questo si è rivelato utile costruire formati di codifica il cui obiettivo è quello di ridurre al minimo la possibilità di errori permettendo, dove possibile, la loro correzione.

A differenza di MPEG Layer 3 che forniva la sola tecnica del CRC, AAC fornisce quattro diversi approcci per la gestione degli errori:

- **Error Detection:** questa tecnica permette di rilevare la presenza di errori nello streaming aggiungendo una serie di CRC associati a parti diverse del bitstream.
- **Error Protection:** questa tecnica rappresenta una evoluzione della precedente ed è presente solo in MPEG-4. Essa suddivide lo streaming MPEG-4 in parti cosiddette "sensibili" e ad ognuna di esse vengono associati dei bit di controllo che ne permettano di rilevare errori e eventualmente correggerli. La suddivisione in parti sensibili viene fatta per ridurre al minimo l'overhead introdotto dai bit di controllo.
- **Error Concealment:** questa tecnica consiste nella sintesi delle parti di segnale andato perso. Attraverso il CRC vengono rilevati gli errori che poi vengono corretti attraverso l'aggiunta di silenzio al posto del segnale mancante o tramite la sintesi basata su modelli percettivi (Shaped Noise o Predict Harmonics) al fine di ricreare suoni il più reale possibili.
- **Error Resilience:** questa tecnica ha l'obiettivo di rendere lo stream audio più resistente agli errori. Esistono tre diversi modi definiti nell'MPEG-4:
  - **Huffman Codeword Recording (HRC):** evita la propagazione di errori all'interno dei dati spettrali
  - **Virtual Codebooks (VCB11):** serve ad identificare errori gravi all'interno dei dati spettrali.
  - **Reversible Variable Length Code (RVLC):** riduce errori di propagazione all'interno dei fattori scala

## 2.5 Formati di file AAC

Come per MP3, AAC ha un formato di file a frame ma, in aggiunta, fornisce diversi metodi di impacchettamento dei file che andrò di seguito a illustrare brevemente. A parte il formato MP4, tutti gli altri file hanno generalmente estensione \*.AAC.

L'unica differenza tra gli header dell'AAC versione MPEG-2 e quella versione MPEG-4 sono i campi che identificano il formato MPEG (2 o 4).

## 2.5.1 Formato RAW DATA BLOCK

Lo streaming consiste in una sequenza di blocchi AAC puri, ossia senza nessun tipo di header che ne descriva il contenuto. Ogni blocco corrisponde a 1024 campioni per cui la dimensione temporale di ognuno è pari a  $1024 / N$ , dove  $N$  è la frequenza secondo la legge di Nyquist. Questo formato equivale ai file audio codificati in PCM.

## 2.5.2 Formato Audio Data Interchange Format (ADIF)

ADIF consiste in uno streaming AAC in formato RAW DATA BLOCK preceduto da un unico header comune. La struttura dell'header è la seguente:

Campo	N. di bit	Note
adif_id	32	Sempre "ADIF"
copyright_id_present	1	
copyright_id	72	Solo se copyright_id_present==1
original_copy	1	
home	1	
bitstream_type	1	0: CBR, 1: VBR
bitrate	23	CBR: bitrate, VBR: peak bitrate, 0 se sconosciuto
num_program_config_elements	4	
<i>I 2 field successivi vengono ripetuti un numero di volte pari a (Num_program_config_elements+1)</i>		
buffer_fullness	20	Solo se bitstream_type==0
program_config_element	VAR	

"program\_config\_element" descrive la configurazione di canale del file AAC definendone la frequenza di campionamento, la codifica di canale (mono, stereo, mix stereo, surround etc), il tipo di formato MPEG e profile/object type utilizzato ed eventuali byte di commento.

## 2.5.3 Formato Audio Data Transport Stream (ADTS)

A differenza di ADIF, ADTS presenta un header prima di 1, 2 o 4 blocchi AAC (RAW DATA BLOCK). Questo formato di trasporto è essenzialmente lo stesso utilizzato per MPEG Layer 3; l'unica differenza sta nel fatto che l'header ADTS non presenta il bit di enfasi che presenta invece l'header dell'MP3.

Qui di seguito è riportata la struttura di un header ADTS:

Campo	N. di bit	Note
<i>Header ADTS fisso: questo non cambia da frame a frame</i>		
syncword	12	Sempre "111111111111"
id	1	0: MPEG-4, 1: MPEG-2
layer	2	Sempre: "00"
protection_absent	1	
profile	2	
sampling_frequency_index	4	
private_bit	1	
channel_configuration	3	
original/copy	1	

home	1	
<b>Header ADTS variabile: questo può cambiare da frame a frame</b>		
copyright_identification_bit	1	
copyright_identification_start	1	
aac_frame_length	13	Lunghezza del frame incluso header (in bytes)
adts_buffer_fullness	11	0x7FF indica VBR
no_raw_data_blocks_in_frame	2	
<b>Test sugli Errori dell'ADTS</b>		
crc_check	16	Solo se protection_absent==0

## 2.5.4 Formato MP4



Il formato MP4 rappresenta l'estensione di riferimento dello standard MPEG-4. Viene definito negli standard ISO/IEC 14496-1:2001 (System) e ISO/IEC 14496-14:2003; di fatto non è altro che un contenitore in grado di includere tutti i tipi di contenuti multimediali (audio naturale e sintetico,

video, grafica 2D e 3D, testo, ecc) e può essere trasmesso su qualunque rete e connessione e immagazzinato all'interno di un file per PC.

La struttura del file MP4 deriva direttamente dal formato di file Quick Time (Apple); dunque MP4 segue una struttura gerarchica ad "atoms" che permettono di inserire al proprio interno più tipologie di contenuti multimediali contemporaneamente, definendone le eventuali relazioni. Ha inoltre un proprio header che descrive completamente gli oggetti contenuti dal file.

L'estensione dei file MP4 è \*.MP4, fatta eccezione per i file creati con i software Apple che invece ne utilizzano una diversa: M4p (o M4A, M4V).

Quando si parla di MP4 si fa generalmente riferimento o alla codifica *general audio* MPEG-4 AAC o ad MPEG-2 AAC (che non differisce assolutamente col suo successore in termini di qualità). Siccome MP4 è provvisto di proprio header, non è necessario utilizzare gli header ADTS e ADIF ma è sufficiente memorizzarne lo streaming puro (RAW DATA BLOCK).

Di fatto, oltre ad AAC, MP4 può contenere qualunque tipo di codifica fornita, sia da MPEG-4 che dagli standard precedenti (per esempio MPEG Layer 1,2 e 3).

Esistono software che permettono di passare da quasi tutte le codifiche AAC al formato MP4 (eliminando, se presenti, gli header ADTS e ADIF e riscrivendone il contenuto nell'header MP4); un software che permette questo tipo di operazione è il celeberrimo Winamp 2, se provvisto di opportuno plug-in (in\_mp4.dll).

# Capitolo 3

## Il Dolby Digital

### 3.1 Introduzione



Ultimamente si sente sempre più spesso parlare di un nuovo formato audio: l' AC3. In realtà questo è il nome che viene dato al nuovo sistema digitale multicanale noto come Dolby Digital Surround, in quanto questo utilizza un algoritmo che si chiama

appunto AC-3.

Trattare in maniera approfondita questa tematica richiederebbe la stesura di un libro. In effetti la maggior parte della documentazione raccolta in questa mia relazione sarà stata presa direttamente dal sito della *Dolby Laboratories* ([www.dolby.com](http://www.dolby.com)), riassunta e tradotta in italiano.

In questo primo paragrafo parlerò del canale Low Frequency Effect (LFE) e tratterò la storia dell'audio multicanale. In quelli successivi discuterò delle tematiche legate al Dolby Digital e infine tratterò l'algoritmo AC-3.

#### 3.1.1 Il canale Low Frequency Effect (LFE)

L'audio a 5.1 canali consiste in cinque canali principali a piena banda (Sinistro Frontale, Centrale, Destro Frontale, Sinistro Posteriore e Destro Posteriore), più un canale opzionale di banda limitata chiamato appunto Low Frequency Effect (LFE).

In contrasto ai canali principali, il canale LFE porta informazioni di frequenze basse (minori di 120Hz) e non ha effetto diretto sulla percezione della direzione del suono riprodotto. Il suo scopo è di aggiungere contenuto di basse frequenze al totale del programma o di alleviare il carico degli altri canali. Il canale LFE fu originariamente creato per le produzioni cinematografica a 70 mm per trasportare un segnale separato basso a uno o più subwoofer addizionali piazzati dietro lo schermo del film. Questo permise di aggiungere effetti profondamente bassi alla colonna sonora cinematografica senza dover aggiornare gli speakers e gli amplificatori esistenti nei tre canali principali dello schermo. Questo significava inoltre che il margine di altezza delle registrazioni audio magnetiche dei 70 mm non dovevano essere ridotte alle basse frequenze, togliendo la capacità del loro rumore anche alle medie e alte frequenze. Per finire il segnale audio non aveva più bisogno di essere elaborato nei vari canali per calcolare la parte che doveva essere inviata ai subwoofer. Quindi aggiungere un canale fu ritenuto essere il metodo più diretto, conveniente e economico, vista la disponibilità di canali nelle pellicole a 70mm, per supportare la capacità di basse frequenze all'interno delle colonne sonore dei film.

Per mantenere la compatibilità con i cinema esistenti, il formato per i film Dolby Digital includono un canale separato LFE. Quando i film con questo formato raggiunsero il mercato, vennero usate le stesse colonne sonore prodotte originariamente, incluso il segnale LFE se era presente. I prodotti per i consumatori che riproducevano suono multicanale Dolby Digital dovevano combinare il canale LFE nel mix acustico col rapporto più adatto in base ai bassi riprodotti negli altri canali.

Le tracce Dolby Digital possono includere il canale di soli bassi LFE, ma questo canale non corrisponde direttamente a un segnale di uscita dei subwoofers. E' possibile che una traccia contenga un canale LFE, ma un decoder può non rilasciare alcun segnale per il subwoofer in quanto tutte le informazioni dei bassi, incluso il canale LFE, può essere riprodotto dagli speakers principali. E' vero anche l'opposto: è possibile per una traccia non contenere un canale LFE, e tuttavia il decoder può rilasciare un segnale dal subwoofer poiché alcuni o tutti gli speakers principali non sono in grado di riprodurre l'informazione dei bassi. La differenza tra il canale LFE e il segnale del subwoofer è che il canale LFE è usato per trasportare informazioni sui bassi aggiuntivi nel programma Dolby Digital mentre il segnale del subwoofer rappresenta come alcune o tutte le informazioni sui bassi vengono riprodotte.

Quindi il canale LFE contiene informazione addizionale sui bassi per supportare l'informazione proveniente dai canali principali. Il segnale del canale LFE è calibrato durante la produzione della colonna sonora in modo da contribuire con un segnale 10 dB più alto dei segnali bassi degli altri canali frontali. Anche se tutti e tre i canali frontali fossero attivi, il canale LFE potrebbe portare abbastanza bassi da solo da portare il subwoofer della sala in bilanciamento acustico con i canali frontali. Questo permette ai produttori cinematografici di non appesantire i principali canali istradando i bassi più forti al canale separato LFE, se necessario. Sotto le più domandate condizioni sonore, dove i bassi caricano pienamente i canali sinistro, destro e centrale, il canale LFE può incrementare l'intensità del basso di più di 6 dB.

Il segnale di uscita del subwoofer, d'altra parte, riceve informazioni anche da tutti e sei i canali, se sono stati selezionati per essere riprodotti da un subwoofer. La specifica combinazione di informazioni nell'uscita del subwoofer è determinata dalle caratteristiche di amministrazione dei bassi scelte per quel particolare sistema di speaker. Per esempio, in aggiunta al segnale dal canale LFE, l'uscita del subwoofer può includere l'informazione sui bassi derivati dal canale centrale e posteriore quando questi speaker sono inadatti a riprodurre adeguatamente le frequenze basse.

Come si vede da questa spiegazione, i termini LFE e subwoofer non sono intercambiabili, e le distinzioni tra i due termini è molto importante. Bisogna prestare attenzione per evitare confusione nell'uso appropriato di questi termini.

### 3.1.2 La storia dell'audio multicanale

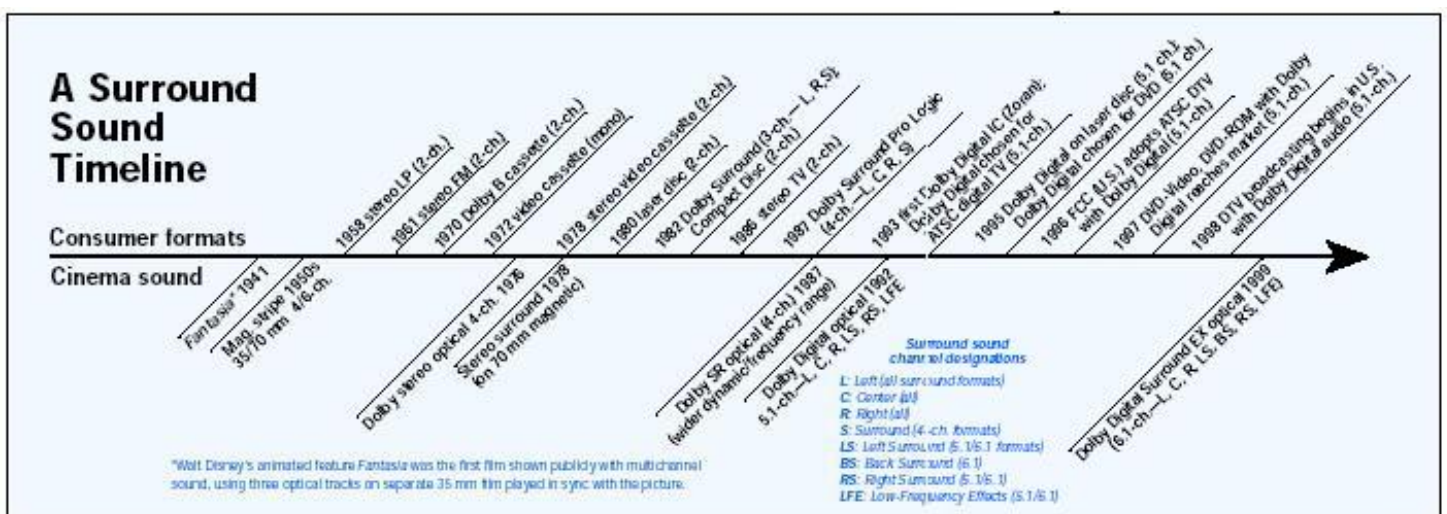


Figura 0.1: la linea del tempo e le nuove tecnologie

### 3.1.2.1 Le origini del suono surround

Il sonoro cinematografico, l'audio della televisione e il formato di riproduzione della musica erano soliti essere prodotti distintamente differenti di industrie che spesso lavoravano isolatamente. Negli anni recenti tuttavia ciò è cambiato. La popolarità del suono surround ha portato a unire assieme queste industrie e i loro formati sonori. E ora la tecnologia digitale multicanale sta avvicinandosi sempre più alla riproduzione sonora, alleviando la fatica sia ai consumatori che ai produttori e permettendo una fedeltà impareggiabile non solo per la tonalità del sonoro, ma anche per la sua spazialità.

Il primo formato sonoro commerciale e multicanale che ebbe successo fu sviluppato nei lontani anni '50 per il cinema. A quel tempo il suono stereofonico, un concetto nuovo al pubblico, fu pesantemente promosso assieme con nuovi formati di maxi schermo da una industria cinematografica che si sentiva minacciata dalla rapida crescita della televisione. A differenza del formato a due canali adottato in seguito per lo stereo di casa, il suono stereo cinematografico iniziò e continuò a usare un minimo di quattro canali.

Con formati cinematografici quali i CinemaScope (35 mm) a quattro tracce e i Todd-AO (70 mm) a sei tracce, i canali audio multipli erano registrati in strisce di materiale magnetico applicato a ogni pellicola di distribuzione. Per suonare queste strisce i proiettori erano attrezzati di testine di riproduzione come quelle di un registratore di cassette, e i cinema erano equipaggiati con amplificatori addizionali e sistemi di speakers.

Inoltre il sonoro cinematografico aveva diversi canali di fronte più almeno un canale suonato dagli speakers nel retro del cinema. All'inizio il canale posteriore era conosciuto come il canale degli effetti, e era riservato per effetti drammatici occasionali come voci eteree nei film epico religiosi, per esempio. Alcuni formati eliminarono questo canale perché causava dei sibili.

Anche se lo stereo cinematografico perse importanza negli anni '60 e '70 a causa degli alti costi dei formati magnetici e di un forte calo nel business cinematografico, i professionisti del suono continuarono gli esperimenti con il canale degli effetti. I formati come quello magnetico da 70 mm a sei tracce (vedi immagine sotto) permettevano un consistente rapporto segnale rumore su tutti i canali, così gli addetti all'audio potevano usare il canale degli effetti per racchiudere gli spettatori in un continuo suono ambientale di basso livello.

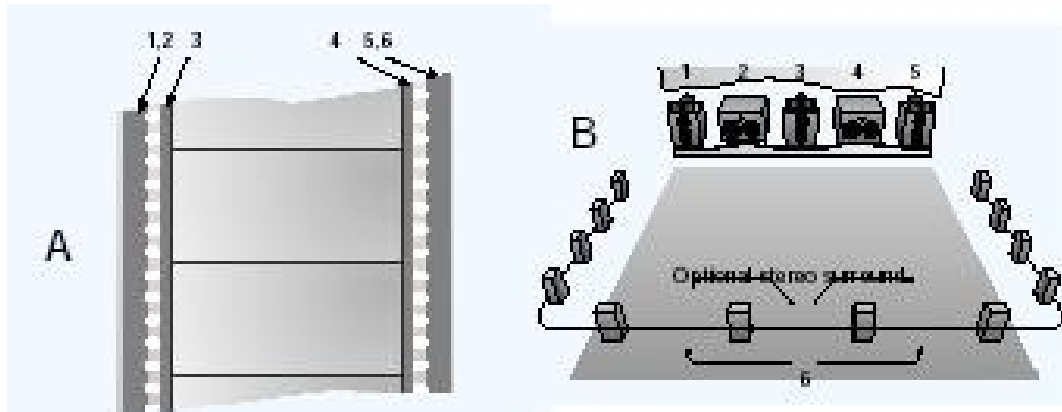


Figura 0.2: nastro magnetico 70 mm e speakers nei cinema

Nell'immagine A possiamo vedere una rappresentazione della pellicola di un nastro magnetico a 70 mm che aveva sei tracce originariamente configurate come cinque identici canali frontali più un canale surround ai lati e nel retro dell'auditorium (immagine B). Quando gli schermi diventarono

più piccoli, riducendo la necessità di cinque canali frontali per trasportare il suono del parlato, i Dolby Laboratories proposero la configurazione mostrata sopra, che è standard dai tardi anni '70: canale principale sinistro, centrale e destro (punti 1,3 e 5 della figura), un canale surround (punto 6) e due canali con le frequenze basse (punti 2 e 4).

Dolby sviluppò anche una tecnica che permise alle pellicole a 70 mm di portare due canali surround, invece del solito canale unico, a scelta del produttore cinematografico. Questo formato "stereo surround" divenne il padre del moderno canale "5.1".

Per tornare al canale degli effetti, questo portò grande realismo sonico dappertutto, non solo per creare effetti drammatici. Questa applicazione espansa e più naturale venne conosciuta come suono surround, e il canale degli effetti come canale surround. Gli speakers extra nel retro, e ora anche nei lati del cinema per creare un più diffuso campo sonoro, divennero conosciuti nel settore come "il surround"

### **3.1.2.2 Lo stereo di casa e il suono quadrifonico**

Il famoso esperimento della Bell Laboratories con il suono stereo negli anni '30 usava tre canali. Lo stereo del cinema intorno agli anni '50 usava non meno di quattro canali, e poteva arrivare a sette. Quando il suono stereo raggiunse finalmente le case nel 1958, erano usati solo due canali. Ciò non era stato fatto a causa di preferenze degli ascoltatori, o per qualche predisposizione da parte dell'audio professionale. Invece due canali (destro e sinistro) era tutto ciò che l'allora prevalente registrazione fonografa LP poteva raggiungere.

La riproduzione musicale con due canali stereo fu una drastica svolta rispetto al mono ed era abbastanza facile da implementare, così diventò presto la norma, anche se i produttori cinematografici continuarono a indicare che almeno quattro canali erano il minimo necessario per creare un campo sonoro realistico. Qualche anno più tardi fu introdotta la trasmissione FM a due canali stereo, introducendo ulteriormente il sonoro a due canali nella mente dei consumatori. Invece, con lo scorrere del tempo, soltanto alcuni patiti dei cinema erano al corrente che lo stereo in questi continuava a usare più di due canali.

Così come lo stereo cresceva in popolarità, i costruttori di attrezzature iniziarono a cercare un modo per espandere il loro mercato. Questa fu una motivazione dietro allo sfortunato formato stereo casalingo a quattro canali, o quadrofónico ("quad"), promosso nei primi anni '70. Richiedeva due speakers aggiuntivi negli angoli posteriori della stanza d'ascolto, per riprodurre due canali extra da programmi sorgente codificati in maniera particolare.

Visto però che i media esistenti avevano un effetto sonoro stereo di soli due canali e era difficile aggiungerne altri, furono sviluppati alcuni schemi per codificare i due canali base con informazioni sonore aggiuntive. La maggior parte di questi erano basati su tecniche matriciali che permettevano di immettere canali extra nei due canali regolari riproducendoli con differente fase relativa.

Quad fallì nel suo intento di catturare una fetta del mercato. Da una parte c'erano diversi sistemi di codifica/decodifica incompatibili, che crearono confusione nei consumatori. Dall'altra produttori e ingegneri del suono non erano d'accordo su come usare al meglio i canali extra. Ma più significativa di tutte, pochi consumatori ricevevano reali vantaggi dal formato. Quad non fu mai associato col formato stereo multicanale usato nei cinema, e il termine "suono surround" ancora non era stato legato a nulla al di fuori dell'industria cinematografica. Dopo tutto, il solo e più popolare media a quel tempo era la TV, che produceva solo suono mono di bassa qualità. Per la decade seguente lo stereo di casa, quello dei cinema e il sonoro della TV rimarranno entità separate supportate da industrie isolate.

### 3.1.2.3 Dolby e il suono cinematografico

A metà anni '70 i Dolby Laboratories introdussero una nuova tecnologia sonora per le pellicole a 35 mm originariamente chiamata Dolby Stereo. Invece di essere basata su strisce magnetiche, usava la tecnologia ottica o fotografica per la colonna sonora, tecnica usata per mettere suono monocolore nei film già dal 1930.

Per permettere la compatibilità di riproduzione nei cinema monocolore, era necessario far stare la colonna sonora stereo nello stesso spazio nella pellicola occupato dalle tracce mono tradizionali. Alcuni esperimenti mostrarono che due tracce, trattate con la riduzione del rumore della Dolby, potevano garantire una eccellente fedeltà. Ma tentare di comprimere nello stesso spazio più di due tracce avrebbe innalzato il rumore a un livello inaccettabile, anche con la riduzione.

Due canali, tuttavia, non erano abbastanza per lo stereo dei film. Gli schermi del cinema erano così larghi che oltre al canale destro e sinistro c'era assoluto bisogno di un canale centrale per permettere agli spettatori seduti nei posti più laterali di localizzare i dialoghi. In più "stereo" e "surround" erano diventati sinonimi per la maggior parte delle industrie cinematografiche, quindi un canale surround era considerata una cosa che doveva esserci. Così si dovette trovare un modo per codificare almeno due tracce fisiche sulle pellicole cinematografiche che avevano quattro canali di informazione: sinistra, centrale, destra e surround.

La soluzione fu trovata nella tecnica matriciale provata per la prima volta nello stereo casalingo quad, ma con i canali di riproduzione configurati con le norme stereo del cinema (canale sinistro, centrale, destro e surround) e con l'aggiunta di nuove tecniche sofisticate di direzionamento elettronico.

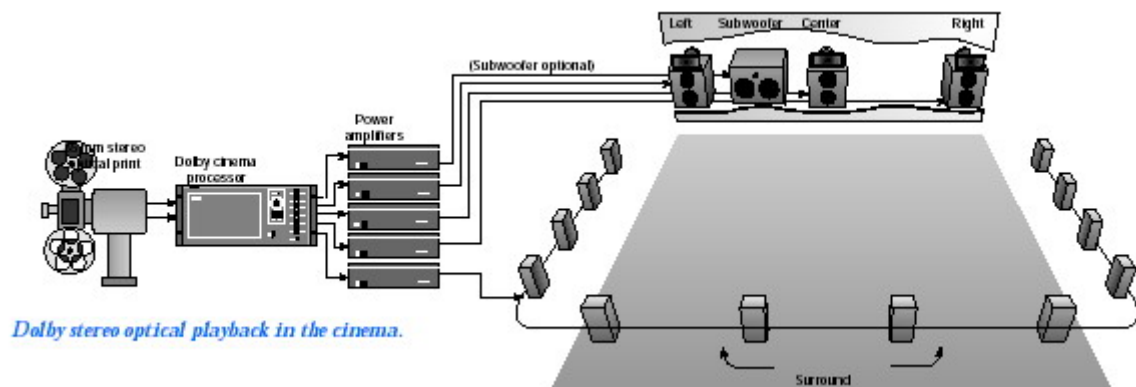


Figura 0.3: formato stereo ottico della Dolby

Il formato stereo ottico della Dolby si dimostrò essere così pratico che oggi ci sono decine di migliaia di cinema in tutto il mondo equipaggiate col processore Dolby, e per molti anni praticamente tutti i titoli più famosi sono stati rilasciati in questo formato (anche oggi, le pellicole video con colonna sonora digitale hanno anche una colonna sonora analogica Dolby, per assicurare la compatibilità in tutti i cinema). In più l'industria cinematografica passò una sorta di periodo rinascimentale negli anni '80, aiutato da un nuovo impegno per la presentazione di alta qualità con l'adozione del Dolby Stereo. Questa rinascita fu assistita anche dalla "rivoluzione video"

### 3.1.2.4 La rivoluzione video

I primi registratori di videocassette furono introdotti nei primi anni '70. Furono pubblicizzati all'inizio come dispositivi di traslazione di tempo per registrare trasmissioni TV che potevano essere visti in periodi più convenienti.

Poco dopo le videocassette acquisirono un'altra, grandiosa e popolare funzione: la visione a casa di pellicole teatrali. Questo creò un appetito vorace di articoli ad esso collegati, appetito che venne soddisfatto dall'industria cinematografica con molta felicità. Nacque un intero nuovo mercato legato alla produzione, vendita e al noleggio di video cassette preregistrate.

Mentre la crescente penetrazione delle trasmissioni televisive ha contribuito all'impoverimento dell'industria cinematografica negli anni '50 e '60, le videocassette fecero l'esatto opposto. La presenza di film aumentò con la crescita delle videocassette preregistrate. I film che si dimostrarono popolari nelle proiezioni cinematografiche si dimostravano altrettanto popolari nelle visioni casalinghe, e in alcuni casi, film che non ebbero successo ai botteghini ebbero comunque un discreto successo nelle uscite in videocassetta.

Le videocassette non furono il solo elemento della rivoluzione video. Più case che mai furono connesse a sistemi via cavo, che garantivano agli spettatori più programmazioni (e l'industria cinematografica ebbe ancora più possibilità dall'aumento di queste). Furono introdotti i laser disc, la fonte di diffusione con miglior qualità mai prodotta. I produttori di televisioni iniziarono a offrire prodotti con ciò che fu definito qualità di immagine a alta fedeltà, e i consumatori li comprarono. Così la "televisione" diventò il "video", e la TV di casa diventò il "monitor video" per mostrare una grande varietà di fonti di programmi video, così come i sistemi stereo suonavano musica da una grande varietà di fonti.

Quando la rivoluzione video era in corso, i consumatori non erano solo abituati al suono stereo di alta qualità, ma stavano diventando abituati anche al suono stereo a alta qualità nel cinema non appena il formato ottico multicanale della Dolby si diffuse attraverso l'industria cinematografica. Mentre all'inizio le videocassette offrivano solo suono monocanale di fedeltà relativamente bassa, vennero presto offerte le macchine e i nastri preregistrati con suono stereo, prima con colonne sonore lineari e poi con la tecnologia "Hi-Fi". I laser disc sono stati concepiti dall'inizio per fornire i suoni stereo di alta qualità.

In più, fu presto adottato un sistema stereo di telediffusione via cavo o via radio. Così il formato a due canali stereo a lungo adottato per la riproduzione di musica casalinga fu adottato per tutte le trasmissioni video.

### 3.1.2.5 Il sonoro surround entra nelle case

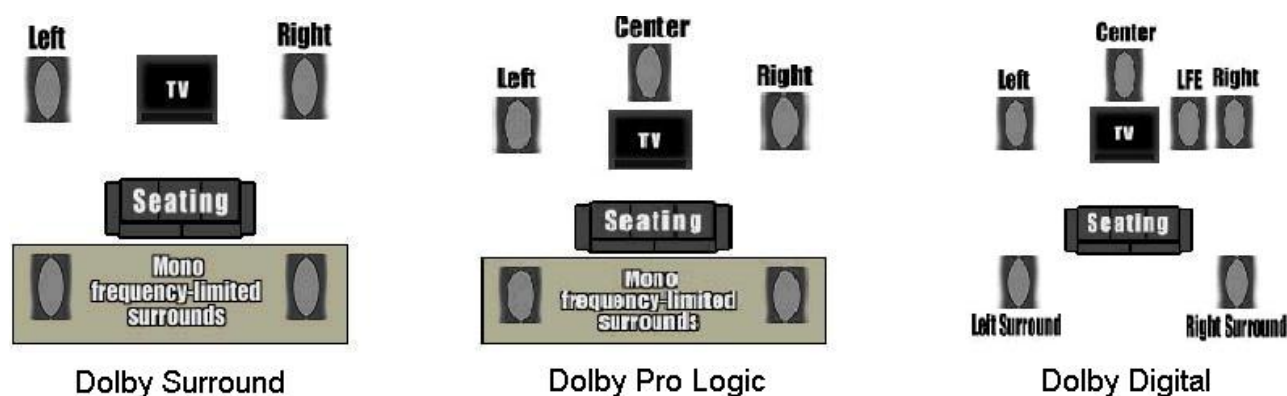


Figura 0.4: evoluzione del Surround nelle case

Dagli inizi del 1980, i sistemi di musica stereo a alte prestazioni erano la norma. L'audio così come il video si erano evoluti a comprendere nuovi generi di fonte (audio cassette e CD). Lo stereo nelle macchine era ormai in età matura, e i nuovi dispositivi portatili stavano creando un nuovo modo di ascoltare musica. Una generazione è cresciuta ascoltando musica rock che era dipendente dalla tecnologia audio quanto essere musicisti, la stessa generazione che era ora regolarmente esposta al

suono surround multicanale nei cinema. Opposto a questo background, Dolby Surround fu introdotto tardi nel 1982 per riprodurre film da cinema originariamente prodotti con colonna sonora con codifica Dolby.

I quattro canali originari della codifica Dolby della loro colonna sonora rimanevano intatti quando tali film venivano trasferiti nelle videocassette stereo e nei laser disc, o trasmesse nelle TV stereo. All'inizio, semplici decoder Dolby Surround rendevano possibile decodificare il canale surround nelle case. Presto decoder più sofisticati come il Dolby Surround Pro Logic resero possibile decodificare anche il canale centrale, e utilizzarono il circuito avanzato di direzionamento sviluppato originariamente per le proiezioni al cinema.

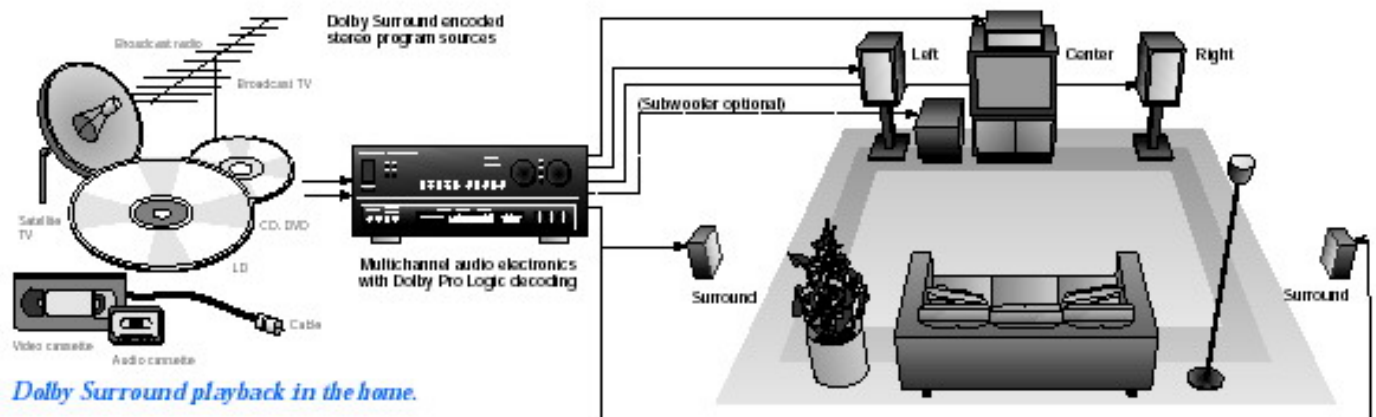


Figura 0.5: funzionamento del surround nelle case

A differenza di Quad, Dolby Surround ottenne, e continua a ottenere, approvazione da una parte considerevole del mercato. Da una parte la configurazione dei canali e il suo miglior uso fu fermamente stabilito all'interno di una industria (cinematografica) a vantaggio della sua introduzione in un'altra (elettronica di consumo). Da un'altra, fu sviluppata con un obiettivo specifico: migliorare l'esperienza visiva. E terzo, gli standard software e hardware per le industrie sia cinematografiche che dell'elettronica di consumo furono definiti da una organizzazione, i Dolby Laboratories, che è indipendente e ha acquisito la fiducia di entrambe. Come risultato milioni di consumatori hanno ritenuto che l'effetto Dolby Surround valesse l'investimento.

Oggi il Dolby Surround include le trasmissioni televisive, e non solo film con tracce sonore con codifica Dolby, ma anche le serie regolari, gli eventi speciali e sportivi. E sebbene il Dolby Surround sia stato sviluppato come un formato audio con video, compagnie di registrazione come la Delos, RCA Victor/BMG Classic e Concord Jazz rilasciano CD audio e musicassette codificate col Dolby Surround. E' stato possibile anche includere più eccitamento e coinvolgimento visivo per un numero crescente di videogames e altre applicazioni multimediali.

In altre parole il suono surround ha superato i problemi dei formati tradizionali, a vantaggio non solo dei consumatori, ma anche dei film, dei software preregistrati, dell'elettronica di consumo e dell'industria dei PC.

### 3.1.2.6 La generazione successiva: il Dolby Digital

Nei tardi anni '80, i Dolby Laboratories hanno intrapreso l'applicazione della tecnologia audio digitale al sonoro dei film a 35 mm in risposta al crescente interesse nell'industria cinematografica. Per mantenere una traccia analogica in modo da permettere alle pellicole uscite di continuare a essere proiettate in ogni cinema, venne deciso che una nuova traccia Dolby Digital ottica separata venisse messa tra i fori di ruota dentata.

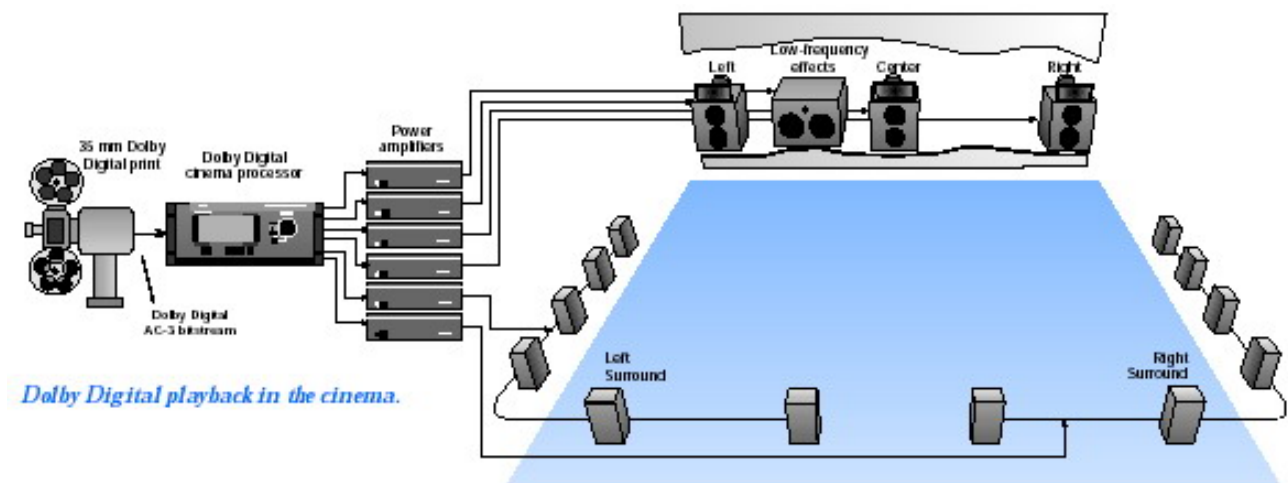


**Figura 0.6: Dolby Digital nelle pellicole cinematografiche**

Venne inoltre deciso che fossero forniti sei canali distinti nella configurazione "5.1" che, a partire da questo periodo, fu valutata da vari gruppi di industrie cinematografiche come quella che meglio soddisfaceva i requisiti di presentazione di film cinematografici.

La configurazione 5.1 garantiva cinque canali discreti a piena banda (destro, sinistro, centrale, sinistro surround e destro surround) più un sesto canale per quegli effetti potenti a bassa frequenza (LFE) che erano sentiti più che uditi nei cinema.

Il Dolby Digital debuttò nei cinema nel 1992, e oggi è il leader tra i formati sonoro digitali per i film.



**Figura 0.7: Dolby Digital nei cinema**

Come il formato audio analogico Dolby, il Dolby Digital nei cinema è stato il trampolino da lancio per il formato per i consumatori con audio Dolby Digital, iniziato coi laser disc nel 1995 e presto seguito dai DVD, dalla TV via cavo e dai sistemi DBS, dalla trasmissione della TV digitale e da una grande varietà di applicazioni multimediali.

### 3.1.2.7 Dolby Digital nelle case

L'equivalente per il pubblico del Dolby Digital dei cinema forma il collegamento finale tra i produttori di programmi multicanale e gli ascoltatori di casa. Come il formato cinematografico, forniva canali sinistro, destro, centrale, sinistro surround e destro surround così come il canale per effetti a bassa frequenza.

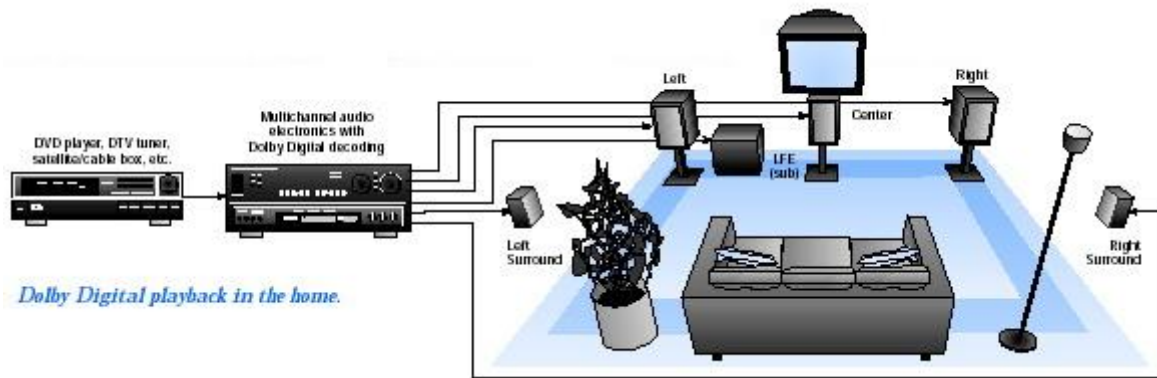


Figura 0.8: Dolby Digital nelle case

A differenza del Dolby Surround col suo canale surround singolo a banda limitata (di solito suonato su due speakers), Dolby Digital forniva due canali surround completamente indipendenti, ognuno di qualità pari ai tre canali anteriori. Come risultato, potevano essere realizzati veri effetti surround che permettevano un senso espanso di profondità, localizzazione e un realismo generale.

Dolby Digital incorpora speciali funzioni per soddisfare gli ascoltatori a partire da piccoli prodotti monofonici fino ai più grandi home theaters.

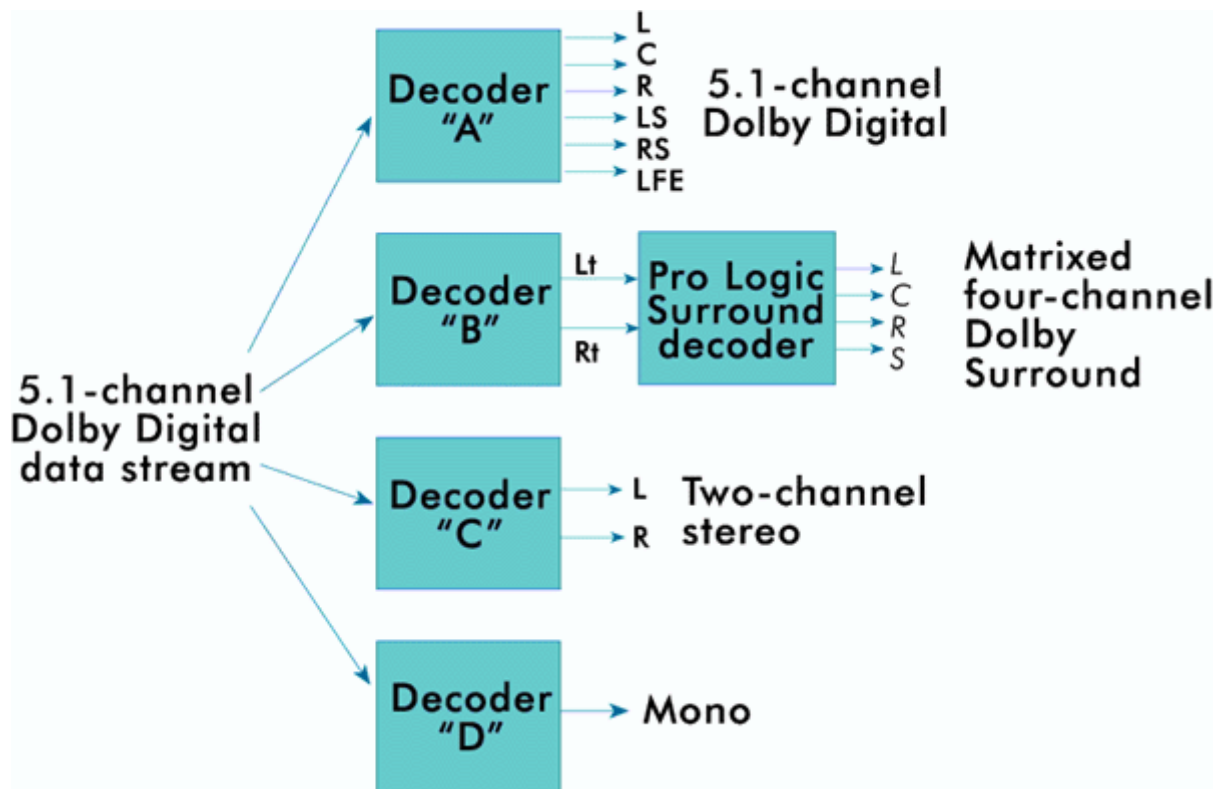


Figura 0.9: tipi di decodificatori e compatibilità all'indietro

Mentre in tutti i sistemi si trasmette il solito contenuto multicanale esteso, il decoder può ottimizzare il sonoro per il particolare ascoltatore. In più, il sistema può applicare compressione dinamica per preservare i contenuti a basso livello e prevenire drammatici passaggi da silenzio a frastuono o, se necessario, per mantenere il volume generale di riproduzione basso. In più l'ascoltatore può programmare il decoder Dolby Digital per indirizzare i suoni bassi non direzionali solo verso quei canali nel sistema che hanno ampia gamma di speakers o subwoofer.

### 3.1.2.8 Un futuro luminoso per la musica e l'audio multicanale

Alla lunga, la tecnologia digitale che ha avuto il più diretto beneficio per gli ascoltatori casalinghi è stato il compact disc. Sia per tutte le sue ergonomiche virtù che per i vantaggi rispetto agli altri formati analogici. Questo è cambiato dall'arrivo dei dischi DVD-Audio.

Questi nuovi dischi avevano dimensione, convenienza e longevità dei CD più suono surround multicanale e qualità sonora migliore. L'inclusione della versione codificata Dolby Digital permetteva a questi dischi di essere compatibili con tutti i lettori DVD-Video e DVD-ROM. L'audio multicanale attraverso i dischi DVD-Audio è sulla strada giusta per portare a una vera rivoluzione riguardante il modo in cui la musica è riprodotta e ascoltata nella casa.

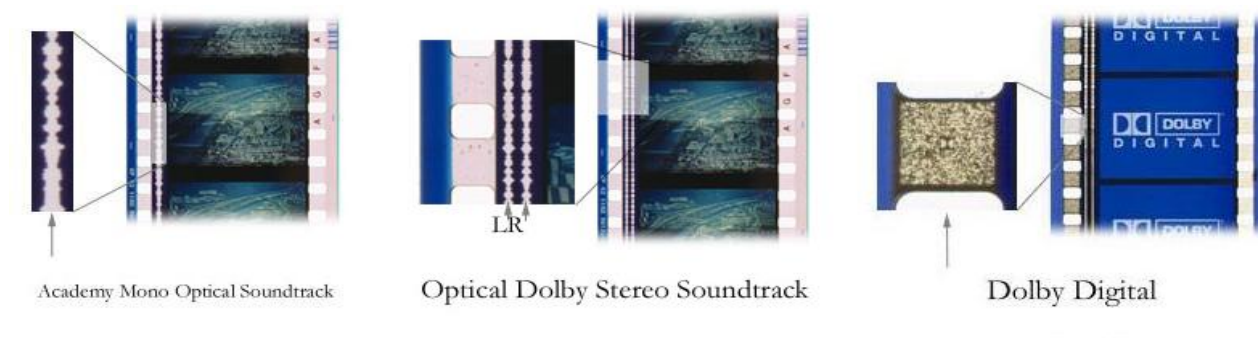


Figura 0.10: evoluzione del segnale sonoro sulle pellicole cinematografiche

## 3.2 Approfondimento sul Dolby Digital

### 3.2.1 Codificatori e decodificatori Dolby Digital



Figura 0.11: esempio di impianto Dolby Digital

In un sistema Dolby Digital i 5.1 canali discreti vengono codificati e decodificati e rimangono 5.1 canali discreti. È importante che vengano monitorati e calibrati per bene tutti i canali in modo da avere la migliore resa possibile quando il segnale verrà decodificato.

Monitorare attraverso un codificatore e un decodificatore è importante per il *Downmixing* e per il *Dinamic Range Control* (DRC). Dolby Digital mantiene diverse funzioni per consentire compatibilità all'indietro così come la facoltà di personalizzare l'audio in modo da renderlo adeguato all'ambiente in cui verrà riprodotto. Per ottenere il miglior risultato, durante la creazione di un contenuto in questo formato bisogna prestare particolare attenzione a funzioni come il

Downmixing, il DRC e la gestione dei bassi, e il prodotto va testato per vedere se soddisfa il creatore così come le necessità del consumatore.

Ci sono prodotti marchiati Dolby Laboratories (tipo il codificatore modello DP569 e il decodificatore professionale DP562) che hanno capacità di monitoraggio per i parametri di cui sopra e in più sono utili per simulare quasi ogni tipo di ambiente di ascolto.

### 3.2.2 Downmixing



Il Downmixing (ovvero la conversione dal formato digitale in uno dei formati meno tecnologicamente avanzati) ha due applicazioni frequentemente correlate: la compatibilità al formato e la redirezione dei canali, che andrò a descrivere brevemente.

#### 3.2.2.1 Compatibilità al formato

I mix mono, stereo e compatibili Dolby Surround sono spesso creati quando al materiale è stato applicato un downmix a meno canali. E' importante controllare un numero di aspetti di ogni downmix per essere sicuri che la traduzione abbia lasciato inalterati gli intenti del mix originale.

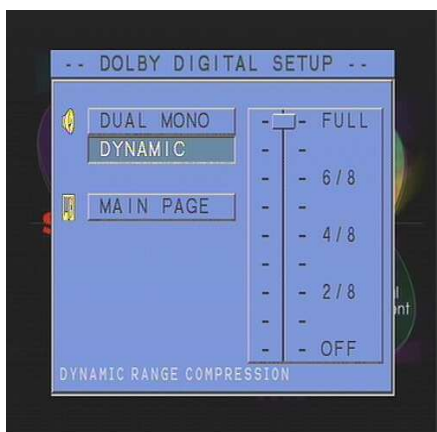
Ci sono molti consumatori che ascoltano sorgenti Dolby Digital come DVD o DTV senza avere un sistema di riproduzione propriamente a 5.1 canali e Dolby Digital. Questi consumatori ascolteranno una uscita a due canali analogica o PCM dal loro lettore DVD o DTV attraverso un sistema stereo o Dolby Surround Pro Logic. Tutti i lettori DVD o comunque i dispositivi che trasmettono segnale Dolby Digital hanno la facoltà di convertirlo a un formato Dolby Surround o stereo. In alcuni dispositivi è possibile scegliere di ascoltare il segnale eseguendo un downmix non necessario.

#### 3.2.2.2 Redirezione dei canali

L'abilità di redirezione delle informazioni di un canale permette di trarre vantaggio dal design e dal numero di speaker nell'ambiente di ascolto.

Ci potrebbero essere dei consumatori che non possono usare tutti gli speaker 5.1 col loro decoder Dolby Digital. Tali decoder hanno comunque la possibilità di reindirizzare il segnale audio a altri speaker, o comunque permettono di effettuare downmix del segnale decodificato multicanale, in modo da usare un numero inferiore di speakers

### 3.2.3 Controllo della gamma dinamica (*Dynamic Range Control - DRC*)



Dolby Digital incorpora sia la compressione della gamma dinamica acustica sia la protezione contro il sovraccaricamento del decoder, che può accadere in caso di downmix. Alcuni produttori permettono agli utenti di scegliere una gamma dinamica piena o ridotta quando ascoltano una colonna sonora multicanale Dolby Digital. Quando è in funzione il downmixing, viene applicata automaticamente la protezione dal sovraccaricamento. E' spesso possibile monitorare le

informazioni della compressione della gamma dinamica codificate nel bitstream Dolby Digital. La funzione di DRC è utilizzata anche per la normalizzazione del volume dei dialoghi.

### 3.2.4 Gestione dei bassi

SPEAKER CONFIGURATION			
FRONT	LARGE	SMALL	
CENTER	LARGE	SMALL	NONE
SIDE SURROUND	LARGE	SMALL	NONE
REAR SURROUND	LARGE	SMALL	NONE
SUBWOOFER	LARGE	SMALL	

La funzionalità di gestione dei bassi permette all'utente di ridirezionare le informazioni a bassa frequenza da uno qualsiasi dei cinque speaker principali al subwoofer, oppure, se non ci sono subwoofer, l'informazione del canale LFE può essere ridirezionata al canale destro e sinistro. Questa funzionalità è importante in quanto la maggior parte dei sistemi di speaker home theatre richiede un certo grado di controllo dei bassi visto che tipicamente nessuno dei cinque speaker principali è stato creato per riprodurre suoni inferiori agli 80Hz.

Anche quando si sta monitorando con delle casse a piena gamma che non richiedono gestione di bassi, questa funzione è utile per testare come le basse frequenze ridirezionate da qualsiasi dei canali principali può influenzare le informazioni del canale LFE.

### 3.2.5 Compatibilità con i dispositivi Dolby Surround esistenti

Nel mondo dell'audio digitale 5.1 multicanale come il Dolby Digital, è importante ricordare che continuano a esistere formati stereo come le trasmissioni televisive, le VHS e i cd. Tuttavia con il Dolby Surround, il formato stereo ha trasportato audio multicanale fin dal 1982. Il Dolby Surround porta il sonoro a quattro canali (Sinistro, Centrale, Destro e Surround: LCRS dall'inglese) su due, attraverso il codificatore, trasmettendolo attraverso i media stereo. La figura sottostante mostra un codificatore Dolby Surround:

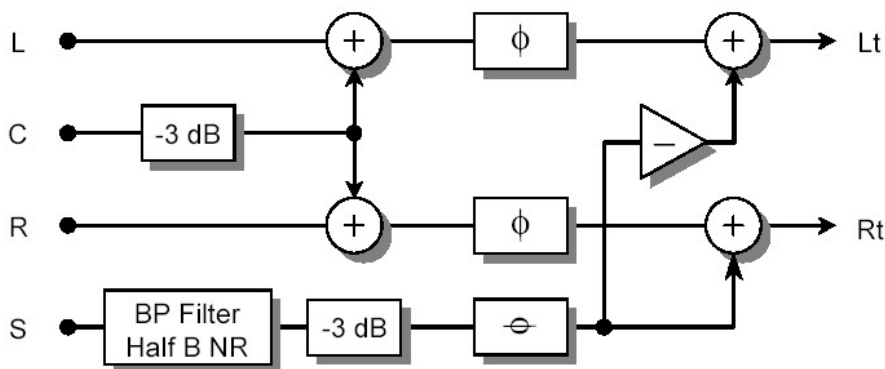


Figura 0.12: schema di codifica Dolby Surround

La colonna sonora codificata in Dolby Surround e compatibile con gli apparecchi stereo viene chiamata "Left total, Right total" oppure Lt/Rt. Sia nel mondo analogico che digitale, Dolby Surround esiste come media, ad esempio VHS Hi-Fi, trasmissioni, CD, laser disc etc. Il Dolby Digital ha compatibilità all'indietro col materiale codificato Dolby Surround.

Tutti i decoder multicanale Dolby Digital, come un ricevitore A/V, contengono una implementazione digitale del decoder Dolby Surround Pro Logic. I decoder Dolby Digital permettono al materiale stereo codificato col Dolby Surround (Lt/Rt) derivante da sorgenti digitali (come laser disc, DVD, DBS, cavi digitali, etc) di essere ri decodificati con una uscita a quattro canali (LCRS).

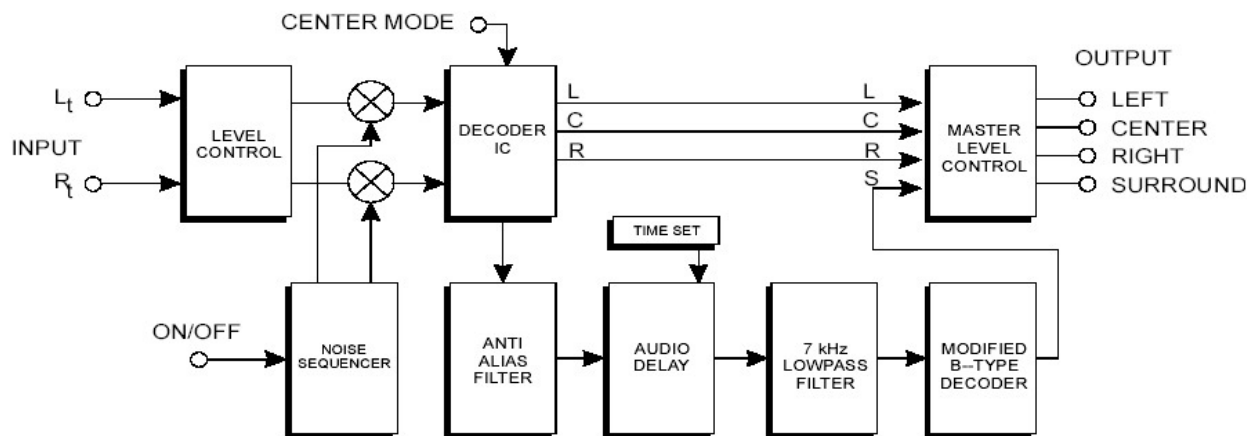


Figura 0.13: decodificatore Dolby Surround Pro Logic

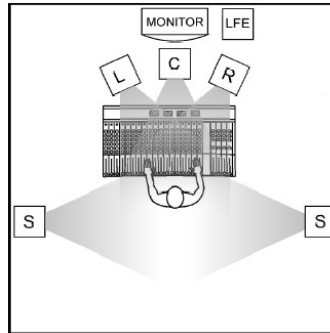
### 3.3 L'importanza dell'ambiente di ascolto

Mi soffermerò brevemente a parlare dell'importanza di creare un ottimo ambiente di ascolto quando si utilizzano dispositivi multicanale Dolby Digital. In realtà, a differenza delle tecnologie audio precedenti, ricreare un ambiente sonoro adatto a questa tecnologia è una sfida per i progettisti e i costruttori di ambienti di ascolto.

Alcuni aspetti importanti da tenere in considerazione sono:

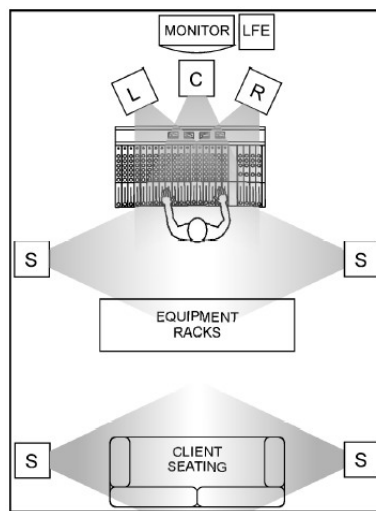
- **il design dell'ambiente:** ci sono fattori di base da tenere in considerazione per quanto riguarda questo fattore. Tra questi le necessità di equipaggiamento, la disponibilità di prese per l'elettricità, l'illuminazione e l'ergonomia.
- **la forma e la dimensione:** dipendentemente dall'applicazione, la dimensione e la forma possono cambiare notevolmente. Ad esempio se la stanza è grande e deve contenere un numero elevato di ascoltatori sarà necessario aumentare il numero di speakers e di apparecchiature per il monitoraggio.
- **l'acustica:** bisogna prestare attenzione a questo fattore. Nel caso di ambienti costruiti appositamente da zero, si può fare in modo che l'ambiente sia adatto alle esigenze, ma quando si modifica un ambiente esistente introducendo nuovi apparecchi o altro materiale può essere necessario modificare la posizione di speakers o altro per eliminare anomalie acustiche

L'ambiente minimo richiede lo spazio per posizionare i cinque speaker dei cinque canali principali (tre frontali e due Surround) più quello del canale LFE.



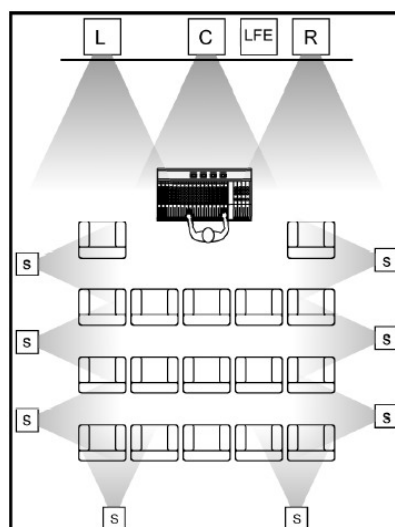
**Figura 0.14: Tipico layout di una stanza equipaggiata con decoder 5.1**

Ovviamente se lo spazio dovesse crescere si avrà bisogno di un numero più elevato di speakers. In una situazione intermedia potremmo aver bisogno solo di un paio di speakers surround in più.



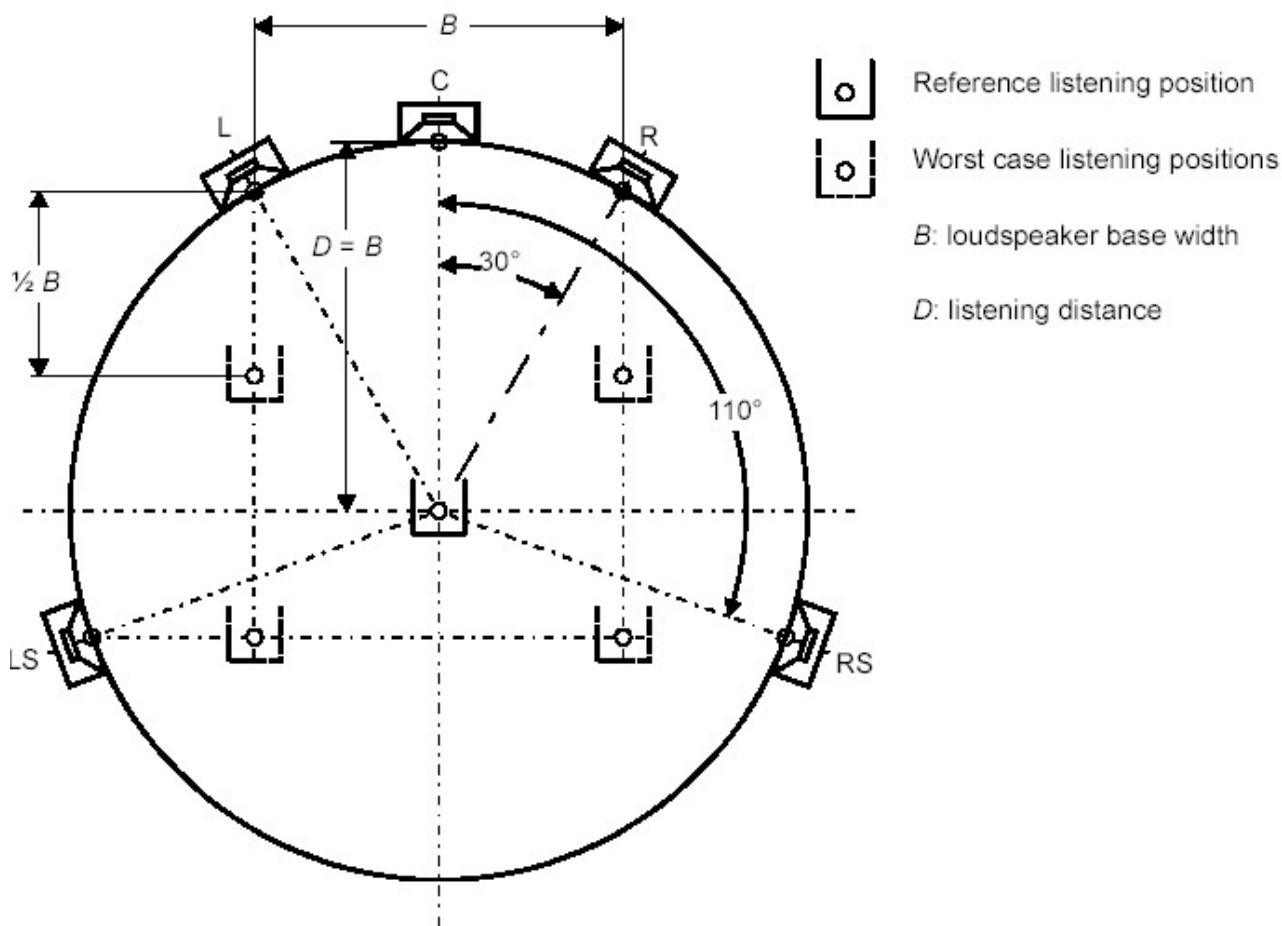
**Figura 0.15: stanza di medie dimensioni con decoder 4.1**

Se però la stanza dovesse essere molto più grossa bisognerebbe valutare un direzionamento delle casse diverso, l'aggiunta di un numero elevato di speaker surround e l'aggiunta di attrezzatura di monitoraggio.



**Figura 0.16: Layout di una grande stanza**

La ITU-R (*International Telecommunication Union- Radiocommunication Sector*) ha dato una specifica per il layout di una stanza per l'ascolto di sonoro multicanale. Queste raccomandazioni sono un buon punto di partenza per creare un buon ambiente sonoro. Senza tener conto dell'allineamento del segnale, è stata descritta una specifica geometria. Lo speaker centrale deve essere di fronte all'ascoltatore, quello destro e sinistro a  $30^\circ$  da quello centrale (formando un angolo di  $60^\circ$ ) e i due surround a  $110^\circ$  dal centro. Per semplicità, la seguente figura mostra il layout appena descritto.



**Figura 0.17: l'ambiente di ascolto raccomandato dall'ITU-R**

Ci sono molte altre caratteristiche importanti da considerare per ricreare un perfetto ambiente sonoro, come ad esempio il ritardo del suono, la gestione dei bassi, la calibrazione dei livelli, le apparecchiature di monitoraggio, la gestione del volume, i subwoofer, la loro calibrazione, la redirezione dei bassi e tanto altro. Tuttavia non basterebbe un libro per descriverle tutte, ed esulano un po' dalla trattazione informatica di questo argomento.

Rinuncio quindi a parlarne, rimandando ancora al sito della Dolby ([www.dolby.com](http://www.dolby.com)) per maggiori informazioni.

## **3.4 L'algoritmo AC-3**

### **3.4.1 Introduzione**

AC-3 è un codificatore audio a alta qualità e bassa complessità. Codificando una molteplicità di canali come una singola entità, è capace di operare a un data rate più basso data una certa qualità audio che non un insieme di codificatori a singolo canale equivalenti.

Benché gli algoritmi AC-3 siano indipendenti dal numero di canali codificati, l'implementazione corrente è basata sul 5.1: 5 canali principali e il canale LFE. AC-3 trasporta questo arrangiamento di canali con un alto grado di trasparenza a data rate bassi, come 320 Kbps.

Le seguenti sezioni descrivono l'algoritmo di codifica AC-3, identificando i suoi elementi di elaborazione principali e discutendo i meccanismi psicoacustici sottostanti che vengono implicati.

### **3.4.2 Panoramica sull'elaborazione**

Come la tecnologia di codifica a singolo canale AC-2 da cui deriva, AC-3 è fondamentalmente un codificatore adattivo basato su trasformazioni che usa un banco filtri lineare in frequenza basato sulla tecnica TDAC (*Time Domain Alias Cancellation*) di Princen – Bradley, che è una variante della Trasformata Coseno Discreta Modificata (MDTC). Il vantaggio primario di una codifica basata su un banco filtri è che le componenti del segnale e le corrispondenti componenti di rumore della quantizzazione (che vogliono essere mascherate) sono tenute all'interno delle bande critiche, e quindi si ottiene il massimo beneficio dalle caratteristiche di mascheramento dell'orecchio umano e si minimizzano i data rate risultanti dalla percezione della codifica a cui viene escluso il rumore indesiderato.

In presenza di transienti, la lunghezza del blocco di trasformazione del banco filtri può essere ridotta dinamicamente per contenere il disturbo della quantizzazione in una piccola regione temporale intorno al transiente.

Una tecnica di allocazione flessibile di bit valuta la molteplicità dei canali per assegnare bit alle frequenze e ai canali come necessario, in base all'effetto di mascheramento che può esserci tra e al di fuori dei canali. Un ulteriore guadagno di codifica è realizzato come necessario separando e codificando indipendentemente le portanti a alte frequenze e le informazioni di avvolgimento.

### **3.4.3 I principali blocchi di elaborazione di codifica**

AC-3 è un codificatore strutturato a blocchi. Sotto è riportato uno schema dei blocchi in cui è suddiviso, e di cui darò una breve descrizione.

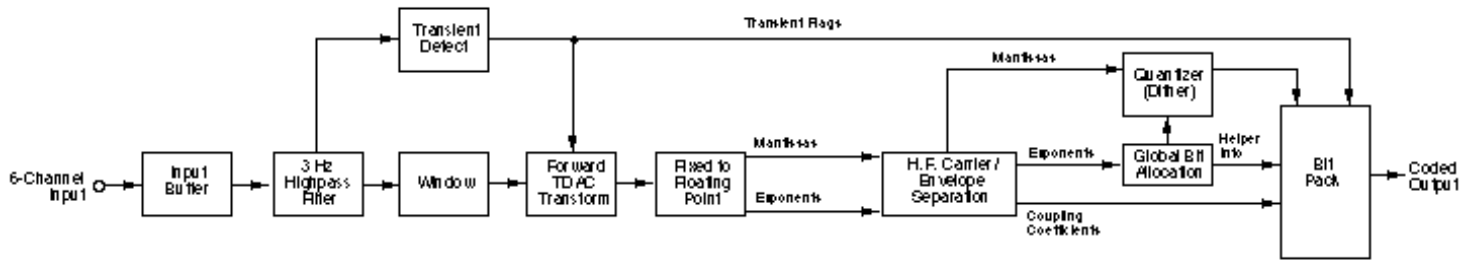


Figura 0.18: encoder AC3

### 3.4.3.1 Buffer di ingresso

Il segnale è diviso in blocchi da 512 campioni per blocco. Uno o più di questi sono riuniti in un buffer di ingresso da ogni canale prima di procedere con altri tipi di elaborazioni.

### 3.4.3.2 Filtraggio di ingresso

I segnali di ingresso del codificatore sono filtrati da un filtro passa alto, per rimuovere il “DC offset”. Questa anomalia si presenta quando nel file ci sono troppe componenti di corrente continua (DC). L'elaborazione e la resa finale del file potrebbe essere disturbata (lo "zero" della forma d'onda non è centrato bene rispetto all'asse Y del grafico), quindi è buona norma, prima di compiere altre modifiche, controllare se il file ha questo piccolo problema e correggerlo. In questo blocco si filtra anche con un filtro passabasso a 120 Hz il segnale del subwoofer.

### 3.4.3.3 Scoperta di transienti

I segnali filtrati e a piena larghezza di banda sono analizzati con un filtro passabanda a alte frequenze, per scoprire transienti. Questa informazione è usata per aggiustare la dimensione del blocco del banco filtri TDAC, restringendo così il rumore di quantizzazione associato al transiente all'interno di una piccola regione temporale in cui tale fenomeno si è verificato.

### 3.4.3.4 Banco filtri TDAC

Il segnale di ciascun canale che è stato filtrato dal filtro passa alto, è ulteriormente filtrato da un banco filtri di analisi basato su TDAC. La trasformazione TDAC è praticamente una FFT (Fast Fourier Transforms) seguita da uno stadio di ‘piccola rotazione’. Ha dalla sua una complessità computazionale bassa con una buona selettività di frequenze, una implementazione a basso costo per quanto riguarda l'hardware e una performance di segnale eccellente.

Ulteriori elaborazioni sono portate avanti dall'insieme dei coefficienti della trasformazione TDAC applicata ai 5.1 canali, trattato come una entità unica (solo i primissimi coefficienti trasformati del canale subwoofer sono mandati oltre senza ulteriori modifiche). Alla fine il decoder riporta il segnale nel dominio del tempo dai coefficienti della trasformazione ricostruita usando un banco filtri basato sulla trasformazione inversa della TDAC. Sia il banco filtri di analisi che di sintesi strumenti critici, ma che possono permettere una esatta ricostruzione del segnale, talvolta senza alcuna perdita di informazione. In aggiunta il banco filtri permette una sovrapposizione del 50% dei blocchi, evitando discontinuità ai margini del blocco. Il guadagno della codifica è ottenuto principalmente dalla quantizzazione selettiva dei coefficienti trasformati, che comprimono gli elementi principali dell'informazione codificata. Finché la codifica non elimina o aggiunge informazione udibile, il sonoro in uscita dal decoder suonerà come quello in ingresso dall'encoder.

### 3.4.3.5 Conversione alla virgola mobile

anche quando è implementato su chip DSP a virgola fissa, i coefficienti trasformati dal TDAC AC-3 sono convertiti nella rappresentazione virgola mobile per le successive elaborazioni, con la mantissa che ha una gamma di grandezza che varia tra 0.5 e 1 e esponenti corrispondenti. Questo assicura che le elaborazioni intermedie non debbano imporre pratiche limitazioni alla gamma dinamica. Larghe gamme dinamiche vengono mantenute tali usando uno stadio di scalatura se necessario. Come risultato, AC-3 preserva i benefici sonici di convertitori A-D e D-A a alte risoluzioni (18 – 22 bit). La rappresentazione a virgola mobile, e in particolare la presenza dell'esponente, serve anche per aiuto computazionale nei processi logaritmici quali l'allocazione dei bit.

### 3.4.3.6 Precombinazione della portante

in generale, la richiesta media di bit da parte di canali multipli sembra essere approssimativamente proporzionale alla radice quadrata del numero di canali. Se servono 128Kbps per codificare un singolo canale, in media 5.1 canali richiedono  $128 \cdot \sqrt{5.1} = 289$  Kbps, comodamente sotto i 320 Kbps di data rate minimo usato dall'AC3. Questo comporta che più canali possono essere codificati usando la flessibilità della tecnica di allocazione globale di bit. Per i segnali a alta richiesta, la precombinazione selettiva di componenti delle frequenze portanti alte è usata per garantire ulteriore guadagno dalla codifica.

Questa tecnica elimina le informazioni ridondanti di localizzazione delle alte frequenze, basato sul fenomeno psicoacustico del sistema uditivo che a alte frequenze localizza il suono basato principalmente sullo sviluppo che raggiunge l'orecchio della versione del segnale filtrato nelle bande critiche, invece che del segnale stesso. Questo comportamento è sfruttato in AC-3 con la separazione delle sottobande a alte frequenze del segnale in componenti di involuppo e di portante, generalmente codificando le informazioni di involuppo con più grande precisione di quello delle portanti e, se necessario, combinando selettivamente (*accoppiamento* – vedi paragrafo 2.2.2) le componenti portanti lungo i canali. Questo ha impatto udibile minimo, visto che l'informazione sulla localizzazione è preservata nei dati di involuppo e le portanti verrebbero in ogni caso combinate acusticamente nell'orecchio dell'ascoltatore, producendo un risultato equivalente.

Le informazioni della portante codificata aumentano gli array dell'esponente e della mantissa, mentre l'informazione di involuppo è trasportata come un array di coefficienti di accoppiamento. Questa tecnica preserva accuratamente la resa spaziale e le altre caratteristiche sonore.

### 3.4.3.7 Allocazione globale dei bit

il vantaggio principale di una codifica multicanale unificata è probabilmente la possibilità da parte della routine di allocazione di far fare la spola ai bit di quantizzazione attraverso canali e frequenze, se necessario, per venire incontro dinamicamente alle richieste dei segnali. L'allocatore di bit AC-3 analizza i coefficienti TDAC rispetto al loro effetto di mascheramento (reciproco) e in relazione alla soglia assoluta di ascolto, per calcolare l'accuratezza della quantizzazione (ovvero il numero di bit) richiesta per codificare ogni mantissa. Il calcolo è compiuto globalmente sull'insieme di canali come una entità unica, usando un singolo fondo comune di bit.

Tuttavia l'abilità di un segnale in un canale di mascherare rumore in un altro canale è limitata, e varia in base alla posizione dell'ascoltatore, così l'effetto globale sull'allocazione calcolata di bit è mantenuto piccolo, e in alcuni casi è trattenuto inducendo in alcuni canali carenza di bit udibili.

Sia l'encoder che il decoder eseguono lo stesso nucleo della routine di allocazione dei bit, lavorando principalmente sull'informazione dell'esponente della TDAC; tuttavia il processo di decodifica è reso più semplice dal decoder che preserva e trasmette i risultati intermedi dal suo calcolo di allocazione. Questo elimina la necessità di ripetere lo stesso calcolo nel decoder.

### 3.4.3.8 Quantizzazione

I risultati del calcolo della allocazione dei bit vengono usati per quantizzare i dati della mantissa del TDAC. Invece che inviare i bit più significativi, il valore è scalato per ottenere un livello di quantizzazione centrato sullo zero, di uguale grandezza e simmetrico (quantizzazione simmetrica dispari), per minimizzare la distorsione e per facilitare l'uso di dither sottrattivi. Un processo di compensazione è presente nel decoder per compattare il codice e ottenere di nuovo i valori della mantissa.

Il dither può essere utilizzato in fase di quantizzazione sui dati della mantissa, nel qual caso ci sono dei bit di informazione che lo segnalano e che garantiscono informazioni di sincronizzazione, così il decoder può sottrarre esattamente gli stessi dati del dither nella ricostruzione della mantissa. Il generatore di pseudo numeri è stato implementato non solo per avere buone proprietà di casualità, ma anche per produrre risultati identici a prescindere dall'hardware.

### 3.4.3.9 Impacchettamento dei dati

I processi di cui ho parlato sopra convertono ogni blocco di segnali temporali di 6 canali in una serie di array derivati e valori scalari, tra cui gli esponenti TDAC e le mantisse quantizzate, informazioni sull'allocazione dei bit, coefficienti di accoppiamento e flag di dither. Nell'ultimo passo del processo di codifica, questa informazione è impacchettata in un blocco unico, assieme alle informazioni di sincronizzazione, un header e altre cose opzionali e informazioni sulla correzione degli errori.

## 3.4.4 I principali blocchi di elaborazione di decodifica

Come per il codificatore anche il decodificatore è diviso in blocchi, in modo da stabilire e mantenere sincronizzazione con lo stream dati in arrivo. I blocchi di elaborazione sono nella figura sottostante.

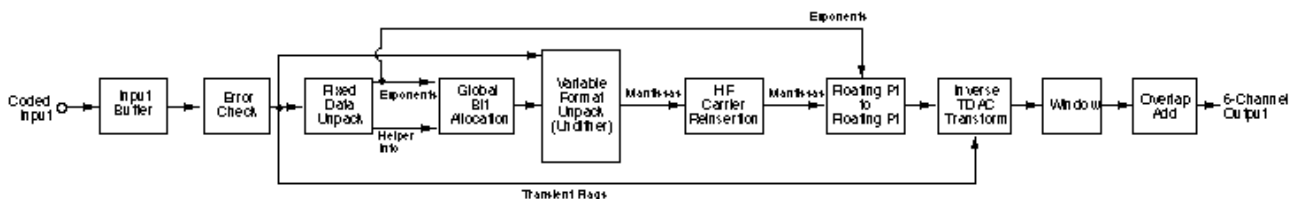


Figura 0.19: decoder AC-3

### 3.4.4.1 Buffer di ingresso

Come per il codificatore, il decodificatore ha un apposito buffer in ingresso che riunisce un intero blocco di codice prima di procedere con la decodifica

#### **3.4.4.2 Ricerca e correzione di errori**

Ogni blocco di data in ingresso nel decoder è testato per verificarne la consistenza interna, così come per verificare la presenza di informazioni di stato attraverso un processore esterno opzionale dedicato alla correzione degli errori. Se è indicata una condizione di errore incorruttibile, il decoder può usare l'ultimo blocco in ingresso riconosciuto valido al posto di quello corrente per mascherare l'errore. La natura di sovrapposizione/aggiunta del processo di ricostruzione del segnale rende questa forma di mascheramento relativamente buona, e un blocco buono può essere ripetuto diverse volte, se necessario, prima che una condizione di errore estesa richieda al decoder di mandare silenzio o, in applicazioni quali le colonne sonore dei film, dover passare a usare il segnale analogico.

#### **3.4.4.3 Spacchettamento di dati a formato fisso**

Lo spacchettamento dei dati è compiuto in due fasi. Nella prima i dati a formato fisso sono spacchettati, assieme agli esponenti, coefficienti di accoppiamento e flag di modalità. Porzioni rilevanti di questi dati sono in seguito usati dal decoder per recuperare l'allocazione dei bit, in modo da spacchettare la parte di dati a formato variabile, principalmente gli array delle mantisse TDAC.

#### **3.4.4.4 Allocazione dei bit di decodifica**

La routine di allocazione di bit in decodifica è quasi identica a quella di allocazione di bit in codifica, ad eccezione del fatto che utilizza i risultati intermedi trasmessi per guadagnare tempo, e opzionalmente per modificare l'allocazione di bit derivata come stabilito dal codificatore. Questo arrangiamento permette anche al decoder di calcolare l'allocazione dei bit un canale alla volta, riducendo le richieste di memoria.

Questa allocazione dei bit deve combaciare perfettamente quella dell'encoder, in modo da poter spacchettare in maniera giusta i dati a formato variabile, o possono essere introdotti stridii nel segnale di uscita.

#### **3.4.4.5 Spacchettamento dei dati a formato variabile**

L'allocazione dei bit di decodifica ottenuta, che specifica la dimensione di quantizzazione di ogni mantissa, è usata per spacchettare i dati in formato variabile dallo stream di bit codificati.

#### **3.4.4.6 Conversione a virgola fissa**

In preparazione alla trasformazione inversa della TDAC, i dati della mantissa e dell'esponente sono combinati per ricostruire i coefficienti a virgola fissa TDAC. Se è in uso il dither, questo è il punto della computazione in cui viene sottratto.

#### **3.4.4.7 Reinserimento della portante a alte frequenze**

I coefficienti a alte frequenze che sono stati codificati come informazione portante di involuppo sono ricostruiti attraverso portanti combinate con corrispondenti coefficienti di accoppiamento.

#### **3.4.4.8 Trasformata inversa**

I coefficienti ottenuti dalla trasformazione TDAC di tutti i canali sono ritrasformati nel dominio del tempo e sono sovrapposti/aggiunti per produrre il segnale di uscita digitale. I coefficienti del

subwoofer sono riempiti di zeri a medie e alte frequenze prima della trasformazione, in modo che l'uscita nel dominio del tempo del subwoofer sia a piena frequenza di campionamento.

### 3.4.5 Altre funzionalità dell'algoritmo

Oltre a descrivere gli elementi che compongono il codificatore AC-3, è probabilmente utile soffermarsi su un elemento omesso, ovvero l'uso della tecnica matriciale (aggiunta e eliminazione di canali) nell'esecuzione della codifica e decodifica. Nel contesto dei codificatori a basso bitrate, modalità matriciale e altri processi a sottrazione di canali hanno la caratteristica indesiderabile che il rumore di quantizzazione associato a un particolare segnale può essere ridiretto dalla matrice decodificata a un differente canale che non quello del segnale, quindi può permettere al rumore di non essere più mascherato. Il processo di codifica AC-3 preserva inerentemente la codirezionalità del segnale e il rumore di quantizzazione corrispondente, per mantenere il mascheramento.

Sia la sintassi del bitstream che le operazioni del decoder sono state create in maniera flessibile per permettere che futuri miglioramenti negli algoritmi di codifica portino a un codice compatibile con l'hardware dei decoder esistenti.

Il formato dei dati e le operazioni generali del sistema sono state concepite per permettere la modifica a livello di blocchi senza la ricodifica

Anche se questa parte ha trattato degli algoritmi AC-3, altri processi possono essere inclusi in una implementazione completa.

### 3.4.6 In sintesi

AC-3 sta per "*Audio Code Number 3*": questo è il codice che i tecnici della Dolby hanno usato per sviluppare la tecnica di compressione e impacchettamento multicanale utilizzata per il Dolby Digital, un metodo di codifica audio basato su un algoritmo di compressione con perdita di informazioni che sfrutta l'incapacità del nostro sistema uditivo di percepire determinati suoni e frequenze se mascherati da altri (l'AC-3 è in sostanza uno steamer di dati che entra nel convertitore di segnale tramite un ingresso ottico, coassiale o in radiofrequenza e viene decodificato da una scheda digitale che riporta allo stato originale il contenuto del programma).

Spesso in relazione ai dvd si sente parlare di "dolby digital 5.1": con questo termine si intende un sistema attraverso il quale si codificano, in una sola traccia audio, sei canali indipendenti l'uno dall'altro (sinistro, centrale, destro, sinistro surround, destro surround), ed un altro sesto canale (LFE, Low Frequency Effects) usato per gli effetti a bassa frequenza, tra i 20 ed i 120 Hz.

Ogni canale può trasportare un segnale completamente diverso da quello degli altri canali. L'ultimo canale richiede un decimo di banda rispetto agli altri, ed è indicato come ".1". Spesso viene anche impropriamente detto canale "subwoofer".

Impropriamente perchè in verità il subwoofer di un impianto Dolby Digital potrebbe essere collegato anche a qualsiasi dei 5 canali "principali", che non sono limitati in banda verso il basso. Questo è il motivo per cui molti dvd riportano l'indicazione "5.1-channel Dolby Digital". Non necessariamente i dvd contengono tracce audio 5.1.

Supponiamo che ad un certo punto, il canale anteriore destro resti pressoché silenzioso (ad esempio durante una panoramica) ; nello stesso istante, il canale posteriore sinistro (per dirne uno qualsiasi), debba riprodurre un suono particolarmente complesso e difficilmente comprimibile : gli servono

dunque più bit di quelli di cui potrebbe disporre realmente; interviene allora il codificatore AC-3, che sottrae quanti più bit può dall'anteriore destro, e li presta al posteriore sinistro ! così facendo, considerando il fatto che per le colonne sonore dei film di solito gli effetti sonori sono di tipo transitorio da un fronte verso un altro, e che quindi in ogni istante vi saranno dei canali liberi rispetto ad altri, è possibile ridistribuire efficacemente i pochi bit a disposizione. Il processo è però distruttivo: cioè i segnali scartati, non possono essere più recuperati ma andranno rigenerati artificialmente, in modo più o meno rigoroso, dal decodificatore.

Codificando più canali come se fossero un'entità sola, AC-3 è in grado di ottenere una grande efficienza in codifica che non attraverso gli algoritmi equivalenti che codificano un canale solo. AC-3 è stato implementato usando l'hardware DSP disponibile tra quelli disponibili e redditizi, ed è studiato per essere facilmente portabile su nuove piattaforme DSP.

Relativamente alla compressione distruttiva (criticata per il motivo che peggiora la qualità finale, spegnendo da una parte la brillantezza del segnale, e dall'altro la tridimensionalità del suono soprattutto in gamma bassa), per il Dolby Digital, questa pecca viene attenuata dalla particolare configurazione del sistema riproduttivo: in caso infatti di semplice stereofonia su due canali la compressione distruttiva fa sentire la sua presenza; nel momento in cui, però, utilizzando ben cinque canali che riproducono un suono a 360°, e quindi tridimensionale, ed un canale solo per i bassi del subwoofer, che ridà energia a questa zona dello spettro acustico, la compressione è virtualmente inudibile.

L'audio può anche essere "mono", "stereo" o "dolby surround". Ci sono anche dei dvd in cui la colonna sonora in una lingua è a sei canali, quella in una lingua diversa è un in Dolby Surround. In ogni caso, le tracce audio dei dvd sono memorizzate in formato ac3. In alcuni DVD l'audio può essere codificato con lo standard EAX (Environmental Audio eXtensions). Questa tecnologia, sviluppata da Creative Labs, si basa su un algoritmo per l'audio posizionale.

# Capitolo 4

## Conclusioni

La ricerca che ho svolto ha preso in considerazione due tecnologie per la compressione e la codifica di un segnale sonoro che sono in qualche modo legate al mercato cinematografico. Infatti AC-3 è un algoritmo per l'audio surround, ma è nato prevalentemente per l'industria cinematografica, mentre l'AAC serve più come compressore audio, quindi il suo ambito potrebbe essere più vicino alla diffusione su rete che non alla diffusione come colonna sonora di un film, ma sempre più spesso si vedono filmati o film su rete che usano i formati audio del comitato MPEG, e quindi soprattutto l'MP3 e l'AAC, come compressione sonora.

Le differenze tra AAC e AC-3 sono ovvie da quanto detto in questa relazione, ma non si dovrebbero paragonare questi due formati, in quanto le loro applicazioni sono diverse. Come detto l'AC-3 non può scendere al di sotto dei 320Kbps, mentre abbiamo visto che con un file AAC possiamo avere tracce stereo praticamente identiche all'originale solo con 96Kbps. Solo che AC-3 codifica audio multicanale (meglio, 5.1) in maniera ottimale.

Proprio per questo motivo dicevo che le applicazioni di questi due formati sono diversi: l'AAC è il tipico formato di compressione da usare per la trasmissione di audio sulla rete, e quindi viene associato ai film solo quando questi vengono compressi, AC-3 è lo standard dei dvd video, in cui il formato del supporto permette di avere una spesa in termini di spazio alta, ma che è mirata a una qualità eccellente.

Il legame di queste due tecnologie alle applicazioni video mi ha spinto a svolgere questa ricerca. Ho avuto modo, nel tentativo di capire a fondo il significato di alcune scelte implementative, di farmi una cultura sulle problematiche della trasmissione del segnale audio, sulle possibilità percettive dell'apparato uditivo umano e sulle tecniche programmatiche che stanno alla base dell'informatica musicale.

Questa ricerca mi ha aperto gli occhi sulla complessità dell'ottimizzazione dell'audio che ascoltiamo ogni giorno attraverso tutte le fonti che abbiamo disponibili. E' stato per me molto interessante anche scoprire lo sviluppo che hanno avuto queste tecnologie (sia l'MPEG in generale che le tecnologie per il cinema).

Ringrazio quindi tutti i siti italiani, ma soprattutto esteri, in cui ho potuto trovare informazioni per documentarmi, che non cito perché sono veramente troppi. Spesso ho dovuto fare ricorso anche a siti che trattano di elettronica per cercare il significato o la motivazione di alcuni fenomeni che possono verificarsi nella trasmissione dei segnali, fenomeni che spesso, se legati all'audio, generano rumori di disturbo o altri effetti indesiderati, e che quindi le moderne tecnologie tentano, in un modo o nell'altro, di mascherare.

Infine ringrazio il professore Leonello Taraballa, del corso "Laboratorio di Informatica Musicale", che con il suo modo di spiegare semplice e efficace ha saputo farmi piacere una materia che non aveva mai attirato la mia attenzione, e in più mi ha dato le basi per la stesura di questa relazione. Grazie ancora.

*-Andrea Lorenzani-*