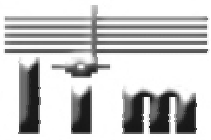
GIANCARLO VERCELLESI
giancarlo.vercellesi@dico.unimi.it

MPEG / AUDIO TUTORIAL



L.I.M. - Laboratorio di Informatica Musicale
DICo - Dipartimento di Informatica e Comunicazione
Università degli Studi di Milano
Via Comelico, 39/41
I-20135 Milano (Italy)

DICo

Ultimo aggiornamento: 09 Marzo 2005

Indice

1. La Compressione Audio
2. MPEG / Audio
   - 2.1. Struttura Generale di un sistema audio MPEG-1 ed MPEG-2
   - 2.2. MPEG-1, MPEG-2 ed i Layer
   - 2.3. MPEG Layer 3 ed MP3
   - 2.4. Caratteristiche del Formato di Codifica MPEG-1 Audio (ISO-IEC 11172-3)
   - 2.5. Caratteristiche del Formato di Codifica MPEG-2 Audio (ISO-IEC 13818-3)
   - 2.6. Formato del File MP3
   - 2.7. Effetto pre-echo: tecnica del Bit Reservoir e Short-Long blocks
   - 2.8. AAC (ISO-IEC 13818-7): la Codifica Audio ad Alta Qualità
   - 2.9. MPEG-4 (ISO-IEC 14496)
   - 2.10. MPEG e patent licence: normative vs informative
   - 2.11. mp3PRO, WMA ed OGG: struttura dei formati e patent licence
   - 2.12. Principi di Psicoacustica
     - 2.12.1. Cenni di fisiologia dell'orecchio: proprietà dei nervi uditivi
3. Valutazione oggettiva e soggettiva della qualità audio percepita
   - 3.1. Valutazione oggettiva tramite segnali multitonali
   - 3.2. Valutazione oggettiva e soggettiva secondo le recommendations ITU-R
     - ITU-R BS.1116, ITU-R BS.1254: Metodi di Test Soggettivi
     - ITU-R BS.1387: Metodi di Test Oggettivi
     - MP3 vs AAC
4. Metadati Audio
   - 4.1. ID3: metadati audio per MP3 ed AAC
   - 4.2. MPEG-7 (ISO/IEC 15938) - "Multimedia Content Description Interface"
5. MPEG-21 "Multimedia Framework"
6. Analisi e Manipolazione Diretta di Formati Compressi MP3 ed AAC

1. La Compressione Audio

Uno dei campi di maggiore importanta nel signal processing è la compressione audio. L'avvento di Internet, il potenziamento dei PC ed il continuo sviluppo di applicazioni e piattaforme multimediali integrate ai servizi di rete, hanno reso questo ambito di fondamentale importanza pratica.
Comprimere un segnale audio permette di minimizzare la quantità di risorse necessarie per la codifica, aumentando così le velocità di trasmissione dell'informazione e diminuendo, conseguentemente, i costi e l'utilizzo di banda. Anche nello storage la compressione ha introdotto notevoli vantaggi, riducendo drasticamente le spese di immagazzinamento dei dati.

Qualunque compressore audio deve trovare il giusto compromesso tra i seguenti parametri:

- **Bitrate:** si misura in bit/sec e rappresenta il numero di bit necessari per trasmettere un secondo d'informazione audio. Minore è questo valore rispetto al bitrate della corrispondente informazione non compressa, maggiore sarà il **tasso di compressione**.
- **Processing delay**: rappresenta la somma dei tempi impiegati da encoder e decoder per eseguire le rispettive operazioni di codifica (compressione) e decodifica (decompressione).
- **Signal quality:** indica la bontà del suono percepito dopo la decodifica di un segnale compresso. In genere tale valore si ricava soggettivamente con opportuni test d'ascolto (ITU-R BS.1116 e BS.1254); è comunque possibile eseguire le medesime valutazioni tramite metodi oggettivi (ITU-R BS.1387).

In genere, encoder complessi hanno tempi d'esecuzione elevati ma forniscono segnali di qualità, anche a bassi bitrate (bitstream di dimensioni ridotte). D'altro canto, semplificando gli algoritmi di codifica compressa, si ottengono codec computazionalmente efficienti, in grado però di generare audio di qualità solo a bitrate molto elevati (bitstream di grosse dimensioni).
A parità di algoritmo di codifica, se si riduce il numero di bit disponibili per ogni secondo di segnale (bitrate), sia la qualità audio percepita che la dimensione del bitstream ottenuto diminuiranno proporzionalmente. E' dunque fondamentale, per una buona scelta dell'encoder da utilizzare, tenere ben presente quali dei parametri sopra elencati ha peso maggiore rispetto alle proprie esigenze.

Un segnale audio è per sua natura continuo (variazione continua della pressione nel tempo). Per poter essere compresso con tecniche numeriche deve essere necessariamente discretizzzato, subendo perciò una prima operazione di campionamento (frequenza di campionamento) ed una

successiva di quantizzazione (Bit di Quantizzazione - SNR). Il risultato finale è un segnale digitalizzato non compresso, in codifica PCM (Pulse Code Modulation).

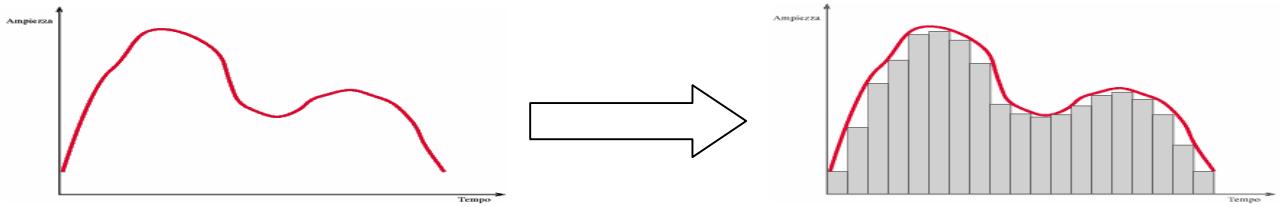


Figure 1: Conversione Analogico-Digitale

Tra le varie tipologie di segnali audio, quelli di maggiore interesse nell'ambito della compressione audio sono quelli vocali e quelli musicali. Di questi, andiamo ora ad analizzare il bitrate necessario per poterli rappresentare in formato digitalizzato non compresso (PCM):

- **Segnale Musicale**: in ambito musicale è diventato uno standard *de facto* il Compact Disc (CD) le cui caratteristiche vengono oramai prese come riferimento per il cosiddetto "audio ad alta qualità". I CD utilizzano codifiche stereo (due canali) con frequenza di campionamento 44.1Khz e parole di quantizzazione a 16bit. Dunque, il bitrate di un tale segnale è pari a circa:

  2 canali * 44100 Hz * 16 bit = **1.4MBit/sec**

- **Segnale Vocale**: in ambito vocale, come riferimento viene generalmente preso il sistema telefonico digitale dove i segnali sono campionati con frequenza di campionamento 8Khz e quantizzati con parole da 8bit. Dunque, il bitrate di un tale segnale è pari a:

  2 canali * 8000Hz * 8 bit = **128Kbit/sec**

Confrontiamo ora i bitrate appena calcolati con quelli ottenibili da due tipici metodi di compressione audio vocale e musicale:

- GSM: il bitrate netto (ossia senza overhead) di un segnale audio GSM è di 9,6Kbit/sec. Dunque, si ha un tasso di compressione di circa 1:14
- MP3: un tipico file MP3 viene codificato a 128Kbit/sec. Con tale bitrate si ha un tasso di compressione 1:10. Se si volesse avere la massima qualità possibile, si deve utilizzare un valore di bitrate pari a 320Kbit/sec; tale valore equivale ad un tasso di compressione pari a circa 1:5. In quest'ultimo caso la differenza tra file original e file compresso è in molti casi impercettibile.

Lo stato dell'arte fornisce una innumerevole quantità di tecniche per la compressione audio. La loro struttura algoritmica dipende fortemente dal compromesso scelto tra i parametri sopra citati e dal tipo di segnale su cui essi vanno ad agire (vocale o musicale).

Gli algoritmi di compressione sono generalmente suddivisi in due grosse categorie: codifiche con perdita d'informazione e codifiche senza perdita d'informazione. Quelle con perdita possono essere

ulteriormente suddivise in codifiche per segnali vocali (*per modelli*) e codifiche general pourpose (*dominio frequenziale*), generalmente indicate per la compressione di segnali musicali.

- **Codifiche lossy per modelli**: qui troviamo algoritmi con perdita d'informazione quali LPC, CELP e GSM. Lavorano generalmente su segnali vocali da cui estraggono i tipici parametri identificativi della voce (pitch e/o formanti) in fase di analisi (compressione), per poi sintetizzarli in fase di sintesi (decompressione). Comprimono fino ad un fattore 26 e sono principalmente impiegati nella telefonia mobile e nello streaming vocale over IP.
- **Codifiche lossy nel dominio delle frequenze:** sono algoritmi con perdita d'informazione che lavorano sullo spettro del segnale. Tale spettro viene codificato con una quantizzazione non lineare generalmente guidata da un modello psicoacustico il quale - sfruttando le teorie psicoacustiche - ha il compito di eliminare informazioni frequenziali "inutili", in quanto ritenute non percepite dall'orecchio umano. Tali tecniche permettono di ottenere un tasso di compressione più elevato rispetto agli algoritmi *lossless* a scapito però di una qualità audio mediamente inferiore all'originale. Inoltre la complessità algoritmica è generalmente maggiore, e ciò causa un incremento del *processing delay*. Appartengono a questa categoria gli algoritmi MPEG/Audio (trattati nel prossimo capitolo), lo standard ASPEC (Audio Spectro-Perceptual Entropy Coding) "padre" di MPEG Layer 3, lo standard MUSICAM (Masking Pattern Adapted Universal Subband Integrated Coding and Multiplexing) "padre" di MPEG Layer 2 e formato di riferimento del sistema DAB (Digital Audio Broadcasting), gli standard Dolby AC-x, PAC (Perceptual Audio Coder), ATC (Adaptive Transform Coding) ed ATRAC (Sony's Adaptive Transform Acoustic Coding).

Algoritmi lossless: sono algoritmi senza perdita d'informazione in cui il segnale audio decompresso risulta essere identico, bit a bit, a quello d'origine. Si basano sul concetto di entropia ed il loro obiettivo è quello di riconoscere ed eliminare le ridondanze numeriche presenti nel segnale attraverso opportune inferenze statistiche. La qualità audio percepita dopo la compressione è identica all'originale ma i tassi di compressione forniti sono generalmente molto più bassi di quelli previsti dai codec *lossy*. La **codifica Huffman** rappresenta un tipico esempio di compressione *lossless*: essa associa pochi bit alle parole più frequenti e tanti bit a quelle con minor probabilità di presentarsi. Nell'esempio 1 (sotto), la parola originale è lunga 60 bit, la parola compressa invece è lunga 19 bit; in questo semplice caso particolare si è ottenuto un risparmio di ben 41 bit.

Esempio: occorre codificare la seguente sequenza di valori digitali

100011	101010	100011	000110	101010	011100	100011	100011	000110	101010

Analizziamo la frequenza (probabilità di presentarsi) delle parole ed associamo loro il corrispondente codice Huffman:

Parola digitale	frequenza	Codifica Huffman
100011	4	0
101010	3	10
000110	2	110
011100	1	111

La corrispondente sequenza codificata in Huffman è:

0	10	0	110	10	111	0	0	110	10

Esempio 1: esempio di una sequenza numerica codificata in Huffman

La codifica Huffman è impiegata nei codec MPEG Layer-3 per comprimere ulteriormente la sequenza numerica ottenuta dopo la quantizzazione non lineare.
Altri esempi significativi di codifiche *lossless* sono: FLAC, MLP (impiegato nei DVD-Audio), SHL e LPAC (impiegato in MPEG-4)

2. MPEG / Audio

MPEG (Moving Picture Experts Group) è un gruppo di lavoro gestito dalla ISO/IEC (ISO/IEC JTC1 SC 29 WG11) che si occupa dello sviluppo di standard per la codifica audio-video digitale.
Il primo progetto MPEG inizio nel 1988 e terminò nel 1992 con l'uscita dello standard internazionale MPEG-1 (ISO-IEC 11172), principalmente utilizzato nei prodotti basati su Video CD ed MP3.
Successivamente vennero prodotti altri tre standard internazionali, MPEG-2 (ISO-IEC 13818), che trova il suo maggiore utilizzo nelle codifiche televisive e satellitari, MPEG-4 (ISO-IEC 14496) ed MPEG-7 "Multimedia Content Description Interface". MPEG-21 "Multimedia Framework"invece è l'ultimo progetto, iniziato nel 2000, al quale MPEG sta attualmente lavorando.

2.1 Struttura generale di un sistema audio MPEG-1 ed MPEG-2

Il sistema di codifica MPEG è costituito da tre entità fondamentali in relazione tra loro secondo lo schema in figura 2.1:
Formato di codifica: insieme di regole definite dagli standard MPEG (esempi: ISO-IEC 11172-3 → MPEG-1 Layer 3; ISO-IEC 13818 → MPEG-2 Layer 3) che specificano come deve essere codificata e strutturata l'informazione audio compressa.
Encoder: blocco software che ha il compito di prendere in input un file non compresso PCM (es: WAV o AIFF) e trasformarlo in formato compresso, secondo lo standard di codifica MPEG scelto dall'utente.
Decoder: blocco software che prende in input un formato di codifica compresso MPEG e lo riporta nel formato non compresso PCM (es: WAV o AIFF).

Il sistema di encoding-decoding è tale per cui gran parte della complessità algoritmica è posta nell'encoder, in modo da rendere il più semplice e veloce possibile la fase di decoding.
Questo perchè, è compito di chi gestisce piattaforme multimediali creare con l'encoder file MP3 di massima qualità ed elevati tassi di compressione, mentre l'utente finale deve solamente utilizzare il decoder per ascoltare musica (o audio generico) avendo a disposizione un software che occupi poco in termini di spazio (byte) e sfrutti al minimo il processore del PC.

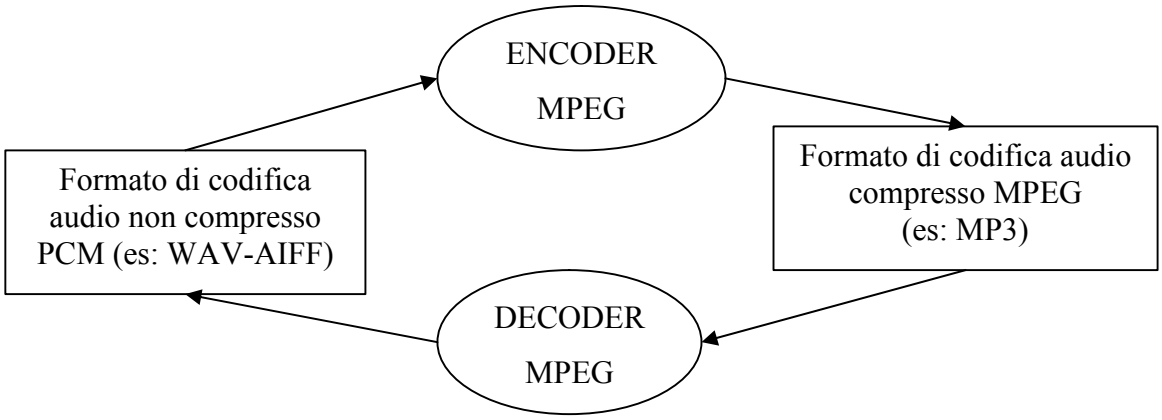


Figura 2.1: schema generico di un sistema audio MPEG-1 ed MPEG-2

Analizziamo ora la struttura di un Encoder Audio MPEG-1 ed MPEG-2 (Figura 2.2).
L'encoder riceve in input un segnale PCM e lo legge a blocchi di 384, 576 o 1152 sample, in funzione del formato MPEG e Layer utilizzati. Per ognuno di questi blocchi effettua le seguenti operazioni:

1. **Banco di Filtri Ibrido**: questa fase ha il compito di convertire i campioni PCM nel corrispondente dominio frequenziale, utilizzando un "Banco di Filtri Polifasico seguito da una Trasformata Coseno Modificata (MDCT)". In altre parole, questo blocco prende il segnale rappresentato nel dominio del tempo e lo trasforma nella corrispondente rappresentazione nel dominio delle frequenze (spettro).
2. **Modello Psicoacustico**: questo blocco rappresenta "il cuore" dell'encoder e di tutto il sistema MPEG/Audio. Il suo compito è di analizzare lo spettro del segnale (calcolato con la Trasformata di Fuorier) e definire il livello di soglia di udibilità SMR (Signal to Mask Ratio) sfruttando i principi psicoacutici dell'apparato uditivo umano (vedi Capitolo 4). In pratica, il modello psicoacustico determina quali sono le sole informazioni che il nostro orecchio è in grado di percepire e quali no, e fornisce questa informazione al blocco "**Quantizzatore non Lineare**" che la gestirà opportunamente.
3. **Quantizzatore non Lineare:** compito di questo blocco è di codificare numericamente lo spettro ricevuto dal blocco "**Banco di Filtri Ibrido**" in funzione dell'importanza di ogni banda di frequenze: se il blocco "**Modello Psicoacustico"** indica che una particolare banda di frequenze è percepita poco, essa verrà codificata con pochi bit; viceversa, se il blocco **"Modello Psicoacustico",** indica che una particolare banda di frequenze è percepita molto, essa verrà codificata con tanti bit. L'obiettivo finale è quello di ottenere una quantizzazione dello spettro tale per cui il rumore di quantizzazione introdotto si trovi al di sotto della soglia di udibilità (SMR) fornita dal modello psicoacustico.

4. **Impacchettamento:** compito di questo blocco è prendere la codifica numerica dello spettro frequenziale generato dal blocco "**Quantizzatore non Lineare"** ed impacchettarla secondo la sintassi dello standard MPEG utilizzato. In questa fase, il layer 3 prevede un ulteriore compressione con l'algoritmo di Huffman.

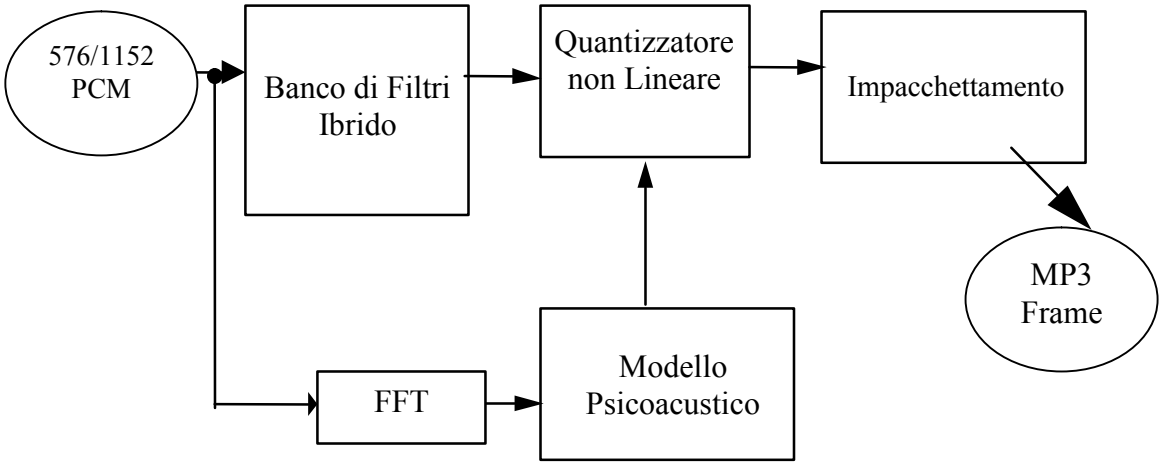


Figura 2.2: schema generico di un Encoder MPEG-1 ed MPEG-2

Analizziamo ora la struttura di un Decoder Audio MPEG-1 ed MPEG-2 (Figura 2.3).
Il decoder riceve in input uno streaming MPEG / Audio valido e per ogni frame effettua le seguenti operazioni:

1. **Spacchettamento:** compito di questo blocco è quello di reperire i frame, leggerne tutte le informazioni codificate (secondo la sintassi MPEG) ed estrarne lo spettro. Nel caso di codifiche in formato MP3, questa fase prevede anche una decodifica Huffman.
2. **Banco di Filtri Ibrido Inverso:** questo blocco ha il compito di prendere lo spettro del segnale e generare i corrispondenti campioni PCM (384, 576 o 1152) da dare in pasto al DAC (Digital to Analog Convert) della scheda audio o da scrivere su file WAV / AIFF.

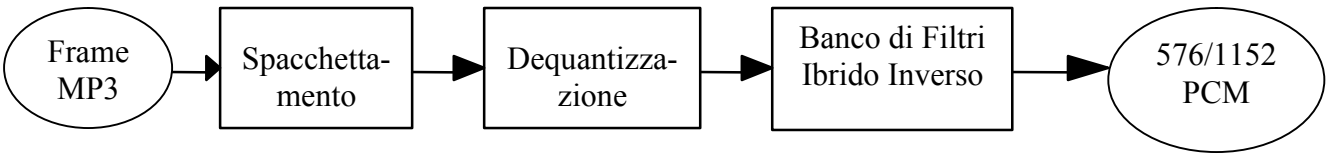


Figure 2.3: schema generico di un Decoder MPEG-1 ed MPEG-2

2.2 MPEG-1, MPEG-2 ed i Layer

Qualunque encoder MPEG/Audio è in grado di comprimere un segnale PCM con diversi algoritmi di compressione. Per quanto riguarda MPEG-1 ed MPEG-2, gli algoritmi esistenti sono tre e vengono identificati da un "Layer" di appartenenza

- Layer 1: è l'algoritmo più semplice dei tre e raggiunge buoni risultati con un bitrate pari a 384Kbit/sec per un segnale stereo. Esso associa ad un frame 384 campioni PCM per frame. Il formato di file associato è l'MP1.
- Layer 2: più complesso del primo in quanto associa ad un frame 1152 campioni PCM; è adatto per codifiche a bitrate intorno ai 192-256Kbit/sec per un segnale stereo. Il formato di file associato è l'MP2.
- Layer 3: è il più complesso dei tre ed è anche quello che raggiunge le migliori prestazioni. Il formato MPEG-1 associa ad ogni frame 1152 campioni PCM mentre MPEG-2 ne associa solo 576, aumentando così la risoluzione temporale. Già con bitrate tra 128-192kbit/sec si riesce ad ottenere un segnale stereo di qualità sufficientemente elevata. Il formato di file associato è MP3.
  I concetti che stanno alla base del layer 3 sono:
  - Dominio frequenziale suddiviso in funzione delle bande critiche
  - Utilizzo della codifica Huffman per l'impacchettamento finale dei dati audio
  - Introduzione della tecnica del "Bit Reservoir" che permette di migliorare la qualità audio a parità di bitrate.

Come capita spesso in questi casi, i Layer sono stati costruiti in modo da essere compatibili con quelli precedenti; un decoder per il Layer 3 è quindi in grado di decodificare anche il Layer 1 e 2.

2.3 MPEG Layer 3 ed MP3

Con la sigla MP3 si fa riferimento ad un formato di file che può contenere al suo interno tre diversi formati di codifica audio: MPEG-1 MPEG-2 ed MPEG-2.5 (capitolo 2.5) Layer 3. Per questo motivo, si usa spesso la sigla MPEG Layer 3 per identificare gli standard previsti dal formato di file MP3.

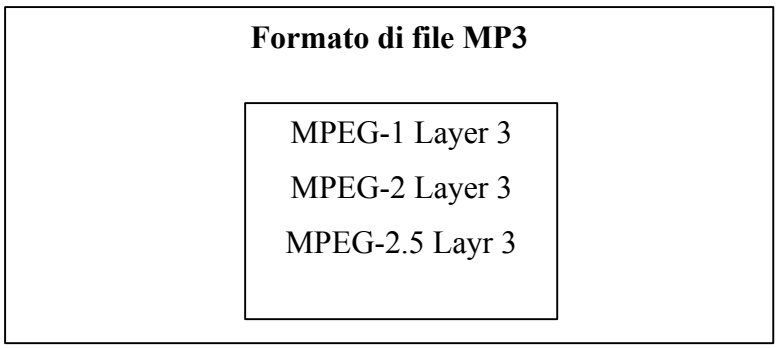


Figure 2.4: relazione tra il formato di file MP3 ed i formati di codifica audio MPEG-1, MPEG-2 ed MPEG-2.5

E' importante ricordare che il "formato di codifica audio" definisce il modo in cui vengono rappresentati i dati audio, mentre, il "formato di file" definisce il modo in cui questi dati vengono scritti su un computer (e dunque su un file).

2.4 Caratteristiche del formato di codifica MPEG-1 Audio (ISO-IEC 11172)

MPEG-1 è indicato principalmente per la compressione musicale in quanto supporta solo le alte frequenze di campionamento.

Frequenze di Campionamento:

Le possibili frequenze di campionamento assegnabili sono le seguenti:

- 32 Khz
- 44.1 Khz (qualità CD)
- 48 Khz

Valori di Bitrate:

I possibili bitrate assegnabili ai vari layer sono i seguenti (Tabella 2.1):

- Layer1: da 32 a 448 Kbit/sec con un gap tra un valore e l'altro di 32 Kbit/sec.
- Layer2: da 32 a 384 Kbit/sec
- Layer3: da 32 a 320 kbit/sec

(Kbit/Sec)		
Layer I	**Layer II**	**Layer III**
32	32	32
64	48	40
96	56	48
128	64	56
160	80	64
192	96	80
224	112	96
256	128	112
288	160	128
320	192	160
352	224	192
384	256	224
416	320	256
448	384	320

Tabella 2.1: Valori di Bitrate in kbits/s per MPEG-1

Tipi di Bitrate:

Bitrate Fisso: tutti i frame presenti nel file hanno lo stesso valore di bitrate (tra quelli presenti in tabella 2.1). Con bitrate fisso si ha la posssibilità di conoscere a priori la dimensione del file a scapito di una minore qualità audio.

Bitrate Variabile: ogni frame può avere un valore di bitrate proprio e differente dagli altri, in funzione della quantità di bit necessari per codificare l'informazione audio associata. Per esempio, la codifica di un silenzio avrà bisogno di pochi bit e dunque di un valore di bitrate basso; l'attaco di una nota invece, richiederà molti più bit per essere rappresentata e dunque di un valore di bitrate elevato. Con bitrate variabile si ha generalmente un'elevata qualità audio ed un buon tasso di compressione, ma non è possibile conoscere a priori la dimensione del file MP3 prodotto.

Bitrate *FreeFormat:* il valore di bitrate può essere diverso da quelli standard presenti nella tabella 2.1 a patto che il bitrate resti fisso ed il suo valore non superi quello massimo previsto dal Layer (esempio: in Layer 3, questo valore è 320Kbit/sec). Scarsamente supportato dai player MP3 presenti nel mercato; molti encoder danno la possibilità di salvare MP3 con valori di bitrate superiori a quello massimo.

Average Bitrate: questa è una tecnica nuova e supportata dai soli encoder di ultima generazione (come l'encoder LAME) e sfruttata anche da "MP3 Editor" per l'operazione di riallineamento dei frame. Definendo un coefficiente di qualità del segnale audio ed il bitrate medio del file MP3 da creare, l'encoder sceglie, frame per frame, il bitrate migliore per soddisfare i parametri dati in input.

Per il Layer 3 è obbligatorio supportare il bitrate variabile mentre per i Layer 1 e 2, questa funzione è facoltativa.

Codifica di Canale:

Per quanto riguarda la codifica dei canali, esistono 4 alternative:

- *Single channel:* codifica MONO
- *Dual Channel:* codifica di due canali mono distinti; questa modalità spesso si utilizza per creare file MP3 *multi-language*: nel canale destro si memorizza uno streaming audio in lingua X, nel canale sinistro invece, in lingua Y; il decoder decodificherà il solo canale sinistro o destro in funzione della lingua scelta dall'ascoltatore.
- *Stereo*: è la classica codifica stereo a due canali indipendenti
- *Joint stereo*: codifica stereo compressa, che utilizza due diversi algoritmi per eliminare le ridondanze presenti nei due canali, MS Stereo ed Intensity Stereo.
    - **Intenity Stereo**: dai principi della psicoacustica è noto che sopra i 2Khz e per ogni banda critica, il sistema uditivo umano basa la percezione dell'immagine stereo sull'inviluppo temporale del segnale audio piuttosto che sull'intensità delle singole frequenze. Perciò l'MS Stereo codifica le frequenze dei due canali stereo sopra i 2Khz, in un unico canale somma, salvando anche opportuni fattori scala, distinti per canale sinistro e canale destro. In questo modo, il segnale ricostruito in fase di decodifica, avrà lo stesso inviluppo del suono originale ma il magnitudo sarà diverso.
    - **M/S Stereo**: questa codifica memorizza nel canale sinistro la media della somma dei due canali, sul canale destro la media della loro differenza. Questa codifica è efficiente se i due canali sono molto simili tra loro in quanto, il "canale differenza" avrà bisogno di pochi bit per essere quantizzato.

2.5 Caratteristiche del formato di codifica MPEG-2 Audio (ISO-IEC 13818)

E' l'evoluzione del formato MPEG-1. Da un punto di vista concettuale non c'è nulla di nuovo rispetto allo standard precedente. Sono stati migliorati ed ottimizzati i tre algoritmi di compressione (Layer) e sono state aggiunte tre nuove frequenze di campionamento (16, 22.05, 24 Khz), più basse rispetto a quelle previste da MPEG-1. Inoltre sono presenti tassi di bitrate più bassi (da 16 a 320

Kbit/sec) ed una codifica multicanale per andare principalmente incontro alle esigenze cinematografiche.

Frequenze di Campionamento:

Oltre a quelle presenti nello standard MPEG-1, sono state aggiunte le seguenti nuove frequenze di campionamento:

- 16 Khz
- 22.05 Khz
- 24 Khz

Valori di Bitrate:

I possibili bitrate assegnabili ai vari layer sono i seguenti (Tabella 2.2):

bitrate_index	bitrate specified (kbit/s) at Fs = 16, 22,05, 24 kHz	
	Layer I	Layer II, Layer III
'0000'	free	free
'0001'	32	8
'0010'	48	16
'0011'	56	24
'0100'	64	32
'0101'	80	40
'0110'	96	48
'0111'	112	56
'1000'	128	64
'1001'	144	80
'1010'	160	96
'1011'	176	112
'1100'	192	128
'1101'	224	144
'1110'	256	160
'1111'	forbidden	forbidden

Tabella 2.2: Valori di Bitrate in Kbits/s per MPEG-2

Tipi di Bitrate:

Sono gli stessi supportati da MPEG-1.

Codifica di Canale:

Oltre alle codifiche presenti in MPEG-1, ne sono presenti di nuove che supportano le codifiche multicanale a 3, 4, 5 e 5.1 canali.

E' interessante notare come MPEG-2 sia compatibile con lo standard MPEG-1 (fatta eccezione per la codifica *joint stereo*) e per questo viene detto "*backward compatible"*. In pratica i decodificatori di MPEG-1 / Audio possono decodificare i due canali principali dal bitstream di MPEG-2 / Audio grazie ad un opportuna combinanazione di ciascuno degli '*n*' canali di MPEG-2 nei canali destro e sinistro di MPEG-1 (operazione di downmix).
In altre parole, ciò significa che è possibile leggere un file codificato in MPEG-2 Multicanale con un decoder MPEG-1.

Esiste un ulteriore formato chiamato **MPEG 2.5**; questo standard è stato introdotto dalla Fraunhofer ma non è ancora stato riconosciuto dalla ISO/IEC. Si differenzia da MPEG-2 per il solo fatto di poter supportare bassissime frequenze di campionamento: 8, 11.025, 12 Khz
Schematizzando MPEG-1, MPEG-2 ed MPEG 2.5 dal punto di vista delle frequenze di campionamento, si ottiene la seguente tabella:

MPEG 1	32 Khz	44.1 Khz	48 Khz
MPEG 2	16 Khz	22.05 Khz	24 Khz
MPEG 2-5	8 Khz	11.025 Khz	12 Khz

Tabella 2.3: relazione tra frequenze di campionamento e formati MPEG

2.6 Formato del file MP3

Il contenuto di un MP3 è organizzato in *frame o bitstream*, ognuno dei quali contiene tutte le informazioni necessarie per ricostruire i corrispondenti campioni PCM, in modo indipendente da tutto il resto del file. Ciò permette di rendere utilizzabile questo formato anche in ambito streaming over IP (es. Radio Web, Telefonia, Chat vocali, ecc.) in quanto, a differenza di un formato a chunk (AIFF o RIFF-WAV per esempio), la perdita di un certo numero di byte audio non compromette la corretta decodifica del resto dell'informazione. Se per esempio vengono persi i dati di un generico frame K, il decoder è in grado di decodificare correttamente tutti gli altri generando un silenzio in luogo del frame mancante.

File MP3						
Frame 1	Frame 2	Frame 3	Frame 4	Frame 5	…	Frame N

Fig 2.5: struttura di uno streaming MP3.

Le quattro parti fondamentali che costituiscono un frame sono: *header*, *CRC*, *side information*, e *main_data,* e sono strutturate come mostrato in figura 2.6:

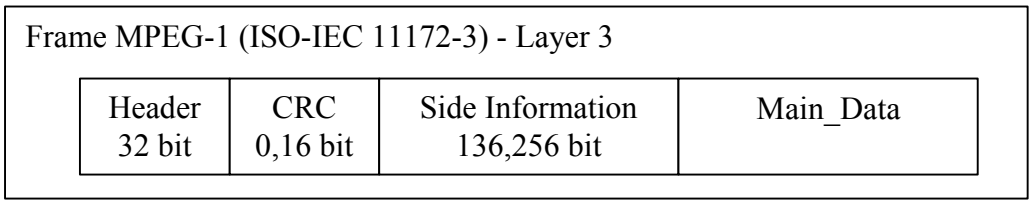


Fig 2.6: struttura di un frame MP3.

Header: contiene tutte le informazioni necessarie per descrivere il tipo di frame.

- Synch Word
- Codifica MPEG utilizzata (MPEG-1, MPEG-2, MPEG-2.5).
- Layer utlizzato (1, 2 o 3).
- Valore di Bitrate in Kbit/Sec
- Frequenza di Campionamento
- Tipo di Codifica di Canale
- Informazioni sulla presenza o assenza di diritti di copyright sul brano
- Informazione indicante il fatto che il brano è originale o una copia.

CRC: è un campo utilizzato dai decoder per validare la correttezza del frame. Questa informazione è di fondamentale imporatanza per lo streaming audio su Internet, dove la probabilità di perdita di informazione è molto elevata.

Side Information: contiene una serie di informazioni per la corretta decodifica dei dati audio veri e propri, i Main Data (puntatore all'inizio dei main_data, dimensioni delle regioni Huffman e relative tabelle utilizzate, dimensione dei fattori scala, dimensione dei main_data, ecc.).

Main Data: contengono la codifica dei dati audio veri e propri, codificati in Huffman

La dimensione fisica di un frame (o bitstream) è pari a:

bitstream_size_MPEG-1 = 144 * (bitrate / frequenza di campionamento) [*byte*]

bitstream_size_MPEG-2_2.5 = 72 * (bitrate / frequenza di campionamento) [*byte*]

2.7 Effetto pre-echo: tecnica del Bit Reservoir e Short-Long blocks

Uno degli artefatti più comuni introdotti dalle codifiche audio percettive è il cosiddetto *pre-echo*, rumore solitamente tenue e situato immediatamente prima di un attacco di segnale, generalmente causato da una ridotta risoluzione temporale dei sistemi di trasformazione tempo-frequenza. Se la finestra di analisi utilizzata dal banco di filtri contiene un attacco di segnale (per esempio un attacco di nota), in fase di resintesi l'energia trasportata dall'informazione udibile verrà sparsa in tutta la finestra, ivi compresa la zona di silenzio. Per ovviare a questo problema, lo standard MP3 prevede due tipi di finestre di analisi, una lunga *(long block)* pari ad un intero frame/granulo, ed una corta (short block) pari ad 1/3 di quella lunga. Ogni volta che il codec (più precisamente, il modello psicoacustico) trova un attacco nel segnale audio, modifica la finestra di analisi da lunga a corta, riducendo notevolmente l'effetto di pre-echo al suo interno. In figura è possibile vederne un esempio: partendo dall'alto, la prima immagine mostra il segnale originale senza pre-echo, la seconda mostra il segnale sintetizzato attraverso uno *short block*, l'ultima invece mostra il segnale sintetizzato con un *long block*.

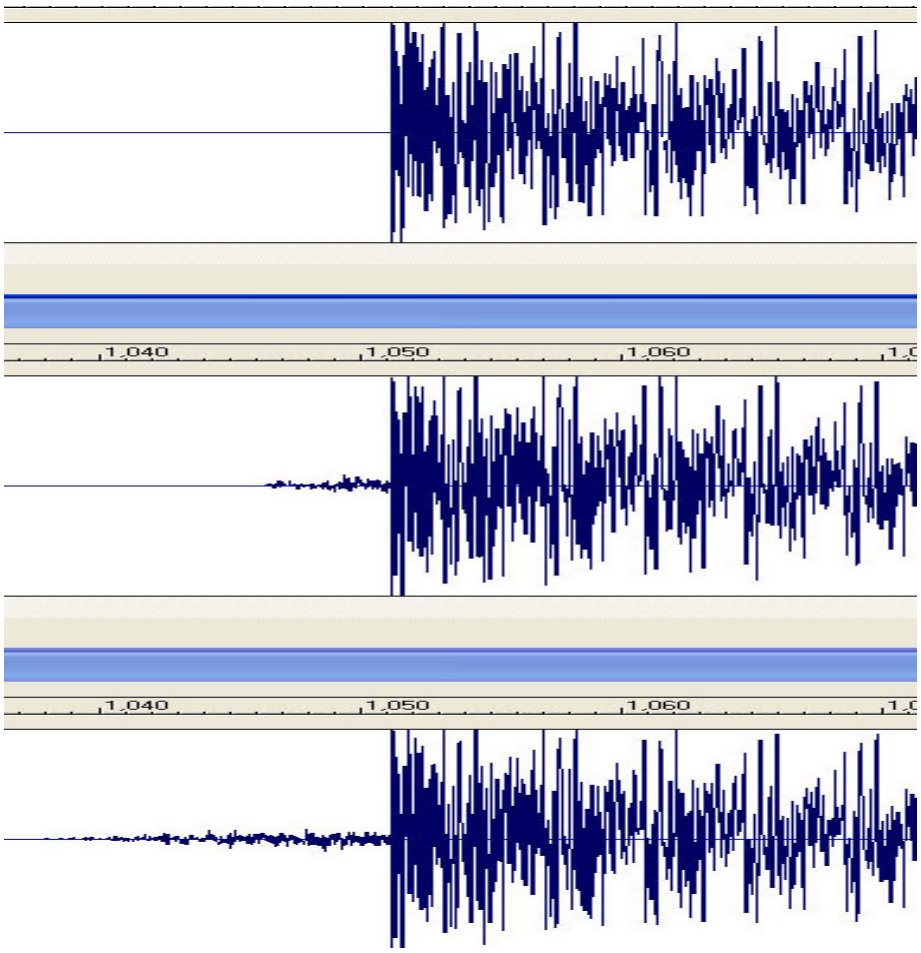


Fig 2.7: effetto pre-echo con long (centro) e short block (basso)

Oltre alla possibilità di impiegare due differenti finestrature in fasi di analisi-resintesi, MP3 mette a disposizione (facoltativamente) la tecnica del *Bit Reservoir*. Il vantaggio portato da tale tecnica è che a parità di bitrate è possibile ottenere una migliore qualità audio, in quanto frame contenenti informazioni audio "povere" (per esempio dei silenzi) mettono il loro spazio a disposizione di quei frame, la cui codifica audio necessiterebbe di un numero di bit maggiore di quelli disponibili al bitrate fissato (per esempio l'attacco di una nota).

Analizziamo ora il funzionamento di questa tecnica: ogni volta che l'encoder deve codificare un blocco di 576 o 1152 campioni PCM, determina il numero minimo di bit necessari per rappresentarlo in MP3. Se questo valore è minore della dimensione fisica del frame, i bit in esubero vengono messi in un sorta di *serbatoio* e lasciati a disposizione dei frame successivi. Questi possono quindi utilizzare, oltre ai bit messi a disposizione dal proprio bitstream, anche quelli attualmente disponibili nel *serbatoio*, la cui posizione iniziale è determinabile da una sorta di puntatore chiamato "*main_data_begin"*. Di seguito sono mostrati due esempi che mostrano la struttura dello streaming MP3 dopo l'applicazione Bit Reservoir (i "Main Info" rappresentano i dati audio veri e propri ossia i Main Data):

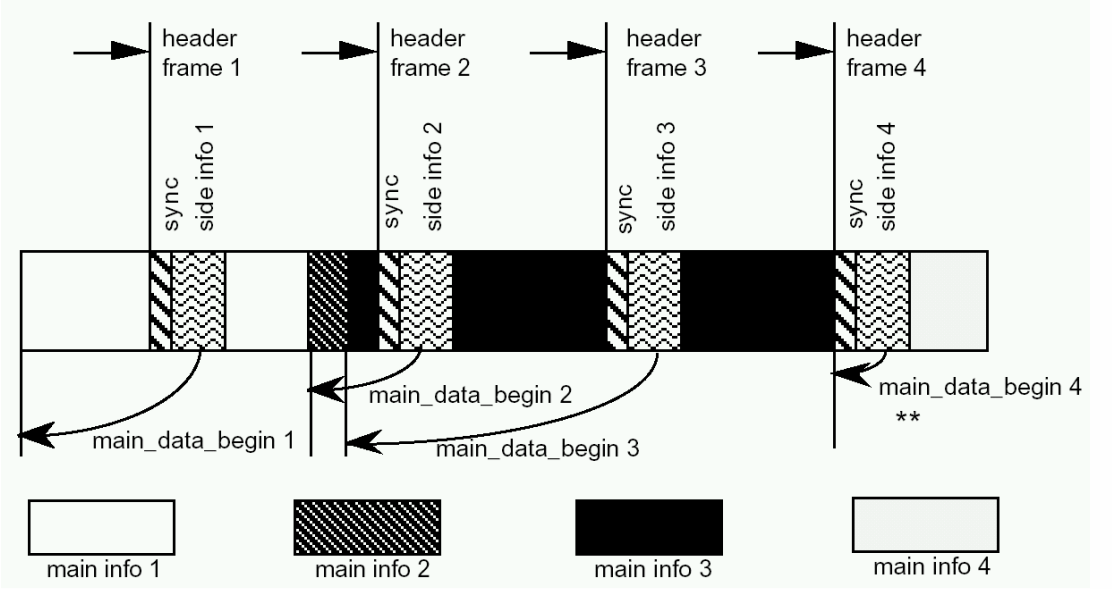


Fig 2.8: struttura dello streming MP3 dopo l'applicazione della tecnica del Bit Reservoir

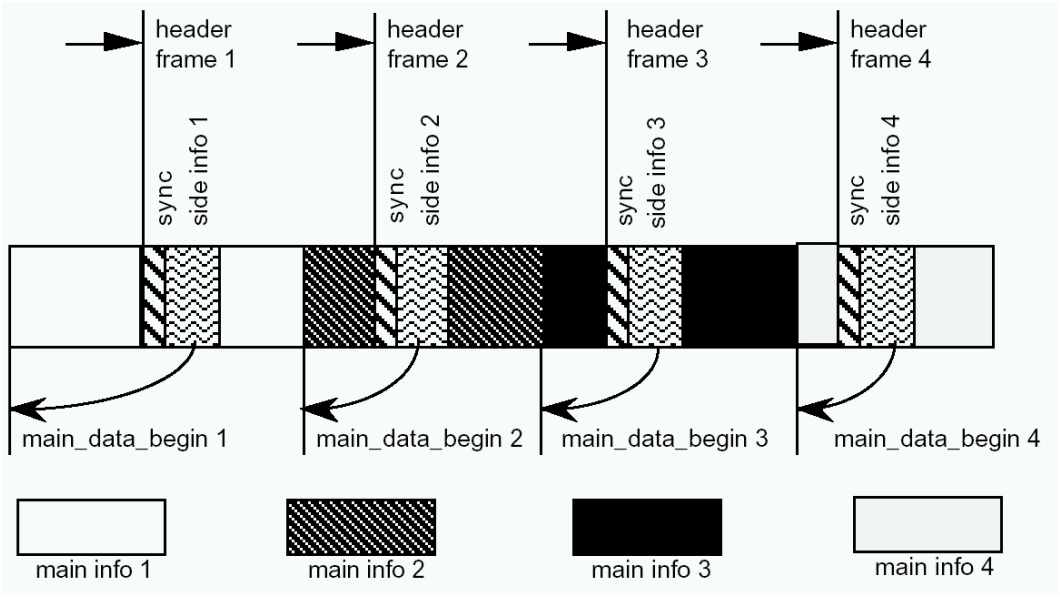


Fig 2.9: struttura dello streming MP3 dopo l'applicazione della tecnica del Bit Reservoir.

2.8 AAC (ISO-IEC 13818-7): la codifica audio ad alta qualità

I risultati di test soggettivi hanno dimostrato che, la necessità di compatibilità all'indietro compromette l'efficacia della compressione del codificatore MPEG-2 in termini di qualità audio. Di conseguenza il gruppo MPEG ha prodotto un addendum allo standard che specifica una modalità di codifica multicanale **non compatibile** all'indietro che offre migliori prestazioni in questo senso.

Tale sistema è stato standardizzato dalla ISO, e prende il nome di **Advanced Audio Coding (AAC)** (ISO-IEC 13818-7)**.** AAC è stato incorporato in MPEG-4 (MPEG-4 AAC).

E' presente un formato molto simile ad AAC, marchiato Dolby (USA) e chiamato AC-3. Questo standard viene attualemente impiegato per la codifica delle tracce audio nei DVD.

Ulteriori approfondimenti: dispensa "AAC-MP4 Overview"

2.9 MPEG-4 (ISO-IEC 14496)

Questo standard segna un'importante evoluzione nel mondo MPEG in quanto introduce il concetto di "Oggetto" nel settore Audio-Video. In sostanza ogni file multimediale è composto da diversi oggetti che, pur potendo esistere separatamente sono armonizzati per ottenere l'effetto complessivo.

Vediamo un esempio. In un film si hanno generalmente dei dialoghi vocali e delle musiche di sottofondo; queste due "entità", avendo caratteristiche fisiche completamente diverse, possono essere gestite da algoritmi di codifica dedicati ed ottimizzati: MPEG Layer 3 o AAC per la musica e HVXC o CELP per la voce. Lo stesso discorso vale per il video.

Per quanto riguarda l'audio, i suoni vengono suddivisi in "Suono Natuarale" (acquisito da fonti di registrazioni esterne) e "Suono Sintetico" (ossia generato tramite tecniche di sintesi matematiche). Ognuno di questi poi, è ulteriormente suddiviso in suono "Parlato" e "General Audio" (ossia tutto ciò che non parlato).
Ogni oggetto ha delle sue caratteristiche particolari e dei tools ad hoc per gestirlo e manipolarlo; ciò permette di associare alle varie tipologie di suoni esistenti in natura l'oggetto più indicato.
L'architettura generale di MPEG-4 / Audio è la seguente:

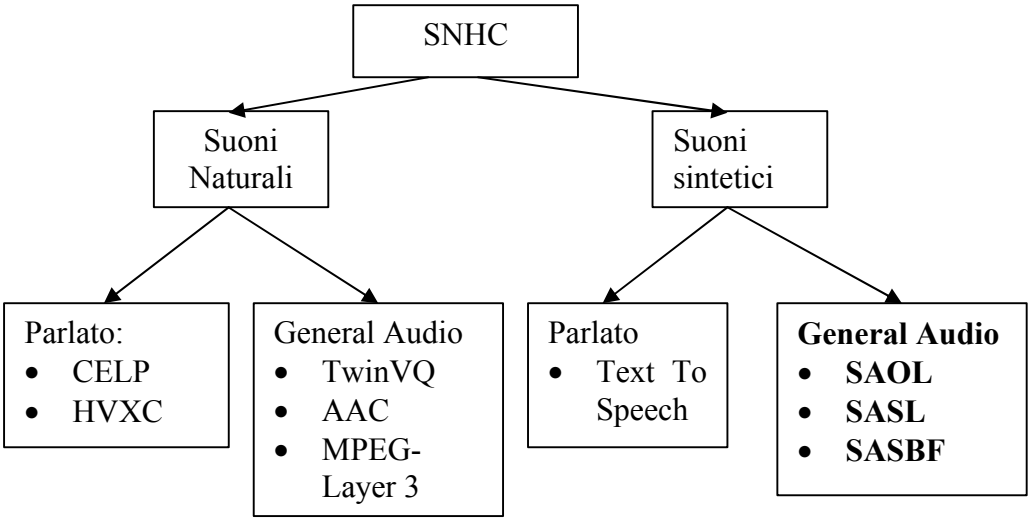


Figura 2.10: architettura ad oggetti di MPEG-4 / Audio

Alla base di tutto c'è SNCH (Synthetic-Natural Hybrid Coding), una tecnica che permette di effettuare la composizione di audio naturale ed audio sintetico sul terminale ricevente in tempo reale.
Molto interessanti sono le tecnologie create per il General Audio dei suoni sintetici che derivano direttamente dalla tecnologia CSound:

SAOL (Structured Audio Orchestra Language) è un linguaggio di sintesi ed elaborazione del segnale audio. Permette la descrizione di sintesi arbitrarie, nonché il controllo e l'utilizzo di algoritmi inclusi nei bitstream MPEG-4.

SASBF (Structured Audio Sample Bank Format) permette di trasmettere gruppi di campioni audio in formato wavetable e di descriverne gli algoritmi di gestione per la produzione di suono. SASBF coincide con lo standard DLS-2 per la gestione dei suoni sintetizzati.
SASL (Structured Audio Score Language) è un linguaggio alternativo al MIDI e descrive il modo in cui gli algoritmi sonori generati secondo il linguaggio di sintesi SAOL vengono usati per la produrre audio.
Con riferimento a CSound, possiamo dire che SAOL corrisponde al file "orchestra" (ORC) mentre SASL corrisponde al file "score" (SCO)

2.10 MPEG e patent licence: *normative vs informative*

Gli standard ufficiali MPEG-1 ed MPEG-2 sono concettualmente suddivisi in due parti distinte: una detta *normativa* ed un'altra cosiddetta *informativa*.
La prima *(normativa)* deve essere rispettata meticolosamente da chiunque (enti di ricerca, società, singoli sviluppatori, ecc.) volesse implementare codec e/o tools MP3 in quanto standard ISO.
La seconda invece rappresenta una semplice linea guida fornita dal gruppo MPEG, come dimostrazione del fatto che la tecnologia fosse realmente implementabile e funzionante. Le ottimizzazioni, i miglioramenti e le nuove invenzioni riguardanti queste parti vengono lasciate a carico di chi ha partecipato attivamente allo sviluppo dello standard, e su cui ISO/IEC ha ufficialmente dato il permesso di brevettazione.
Seppur non esplicitamente evidenziate nella documentazione ufficiale ISO/IEC 13818-7, anche AAC possiede di fatto parti *normative* e parti *informative*. E' chiaramente *normativa* la sintassi e la semantica del formato, la struttura generale del sistema di decoding, dei vari tools che lo compongono e le varie configurazioni che si possono creare. E' invece totalmente informativa la parte riguardante la fase di encoding, tant'è che non è stata nemmeno inclusa nel documento ISO/IEC 13818-7. In altre parole, lo standard specifica il formato, i tools che devono essere impiegati nelle varie configurazioni di encoding ed il loro comportamento in fase di decodifica. Tutta la parte di encoding è di fatto *non normativa*.

La ISO/IEC è l'unico proprietario delle tecnologie sviluppate e standardizzate all'interno del gruppo di lavoro MPEG. Ciò nonostante, terminata la fase di standardizzazione, si segnala la possibilità che alcune delle società partecipanti allo sviluppo dello standard (ITTF - Information Technology Task Force) possano brevettare le cosiddette parti *Informative* o comunque creare dei patent su altre tecniche che però utilizzino direttamente la parte *Normativa* dello standard MP3 - AAC. In altre

parole, ciò che è libero ed a completa disposizione di tutti è lo standard in se (di proprietà ISO/IEC), mentre è brevettabile tutto ciò che viene definito *Informativo,* e che dunque è soggetta a studi, ricerche ed investimenti da società e gruppi di ricerca operanti nel settore multimediale.

I motivi che hanno indotto ISO/IEC ad autorizzare la creazione di Patent Licence su alcune parti degli standard sviluppati all'interno del gruppo MPEG sono sostanzialmente due:

- molte parti della tecnologia sono soggette a miglioramenti continui, in funzione dell'innovazione tecnico-scientifica. Si pensi per esempio a tutte le nuove scoperte verificatesi dal 1992 -anno in cui MPEG-1 è stato standardizzato- ad oggi nell'ambito della psicoacustica e della psicofisiologia nella percezione vocale e musicale: tali risultati possono essere costantemente integrati all'interno dei modelli psicoacustici previsti dagli encoder MPEG, senza dover revisionare a livello ISO/IEC lo standard, ed allo stesso tempo migliorando notevolmente la qualità della compressione ottenuta nel prosieguo degli anni.
- le società operanti nel settore multimediale non avrebbero motivo di partecipare ad un sistema di standardizzazione come MPEG senza poter vendere i propri prodotti sotto brevetto, fomentando così la proliferazione di codifiche e tecnologie private per la conquista del mercato. Un tale sistema perciò, attraverso i Patent Licence induce le società a partecipare allo standard, ed allo stesso tempo permette di concentrare tutti gli sforzi di ricerca (privati ed accademici) su una quantità ridotta di formati, evitando così la proliferazione di decine e decine di codifiche private, di minore qualità e totalmente incompatibili tra loro.

Attualmente, le due società che detengono il monopolio *de facto* dei brevetti MP3 sono Thomson e Fraunhofer, mentre Dolby è il "licensing authority" per AAC. Ovviamente, essendo gestiti da diverse entità, anche i metodi ed i criteri scelti per la gestione dei Patent Licence è differente.
Per le tecnologie MP3, chiunque -per fini commerciali- volesse sviluppare encoder MP3-AAC (indipendentemente dall'implementazione), o utilizzare tale tecnologie all'interno di prodotti multimediali (videogiochi, ecc.) o per streaming audio, deve pagare delle royalties a coloro che detengono i brevetti su tali tecnologie, ammesso che il mercato mosso sia superiore a delle soglie note (definite nel regolamento dei Patent Licence). E' altresì necessario pagare royalties a tali società per l'utilizzo degli SDK da loro forniti. Non è invece necessario pagare, ne per utilizzi non commerciali della tecnologia, ne per gli utenti finali, diretti usufruitori di MP3.
Anche AAC precede delle royalties per la vendita di codec o di applicazioni che utilizzino tale tecnologia (*professional codec*), mentre nessuna tassa è prevista per l'uso personale o senza fini di

lucro (*consumer codec*). A differenza di Thomson-Fraunhofer, Dolby non ha messo nessuna tassa per la distribuzione di bitstream AAC in ambiti quali lo streaming, il downloading o il mercato ludico (videogiochi, ecc.), ma ha invece previsto tariffe differenziate in funzione del numero di canali impiegati dall'applicazione oggetto delle royalties. Tale differenziazione viene classificata per *output channel* mono, stereo o multicanale. Infine, esiste un'ulteriore diversificazione tariffaria tra AAC in MPEG-2 ed MPEG-4; generalmente, MPEG-4 ha un costo maggiore anche se esistono dei profili AAC in MPEG-4 equiparabili ad MPEG-2, perciò tassati alla stessa maniera (per esempio, AAC Low Complexity Profile). Occorre infine osservare che gli algoritmi di SBR impiegati in AAC+ sono di proprietà *Coding Tecnologies* e vedremo più avanti che tali royalties sono applicate anche nella tecnologia mp3PRO.

Esistono altre società che possiedono brevetti sulla tecnologia MP3 ed AAC (per esempio Philips e Sony), ma essendo meno radicate in tale mercato, sono meno attente alle eventuali violazioni. Il controllo dei brevetti e l'eventuale citazione a giudizio di coloro che violano le regole ha dei costi significativi, sia in termini di risorse economiche che umane; e spesso, tali investimenti possono risultare ben superiori al reale guadagno portato da un'eventuale vittoria in sede giudiziaria.

2.11. mp3PRO, WMA ed OGG: struttura dei formati e patent licence

Oltre ad MPEG, esistono numerosissimi altri formati e standard di codifica per l'audio compresso. In questo articolo si è ritenuto opportuno citarne solo tre di questi - mp3PRO, WMA ed OGG- in quanto attualmente valide alternative ad MP3-AAC.

mp3PRO

E' noto che a bitrate inferiori a 128Kbit/sec, MP3 degrada significativamente la qualità del segnale audio. Questo perchè l'algoritmo di compressione lavora sull'intera larghezza di banda, esaurendo subito i bit disponibili per la codifica di tutte le alte frequenze maggiormente penalizzate dal sistema di compressione. Il gruppo MPEG ha deciso che in tale situazione l'algoritmo avrebbe dovuto creare i cosiddetti "coding artefacts", ossia delle distorsioni in banda audio nel segnale audio compresso.

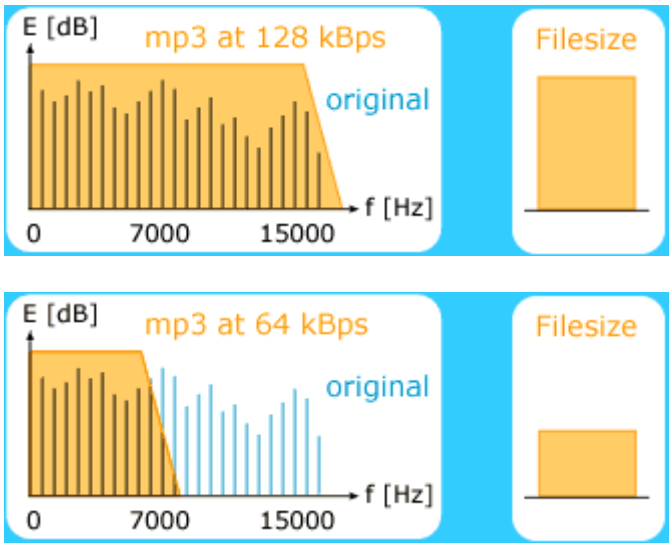


Figura 2.11: sistema di codifica MP3 a 128 e 64 Kbit/Sec

Al fine di superare questo limite alcune società leader nel settore del multimedia audio si sono unite nel gruppo *Coding Technologies* ed hanno sviluppato mp3PRO, una sorta di evoluzione dello standard MP3. mp3PRO è la combinazione di MP3 ed *SBR - Spectral Band Replicator* (impiegata anche in AAC+): l'algoritmo codifica le basse frequenze con il classico algoritmo MPEG Layer 3 mentre le alte vengono codificate separatamente con SBR (da cui deriva il nome "PRO"), tecnica in grado sintetizzare componenti in alta frequenza con un ridotto numero di bit. mp3PRO è perciò in

grado di rappresentare una traccia audio a 64Kbit/sec con una qualità prossima a quella di un MP3 a 128 Kibit/Sec.

Figura 2.12: sistema di codifica mp3PRO

Le codifiche separate di parte bassa (0 : Fc/2) e parte alta (Fc/2 : Fc) dello spettro, permettono la completa compatibilità del formato mp3PRO con il suo predecessore MP3, favorendone così il suo ingresso ed impiego nel mercato multimediale. E dunque possibile ascoltare file codificati in mp3PRO con player MP3 a patto che questi supportino le basse frequenze (16, 22.5, 24 Khz); chiaramente, i decoder classici, non essendo provvisti della tecnologia "PRO", ignoreranno la parte alta dello spettro (quella codificata con SBR) generando così segnali audio con larghezza di banda dimezzata rispetto all'originale.
Esempio: un segnale audio digitalizzato a 44.1Khz e compresso con mp3PRO, se decodificato con decoder MP3, avrà frequenza di campionamento pari a 22.05Khz (=44.1/2).

mp3PRO non è uno standard MPEG (ISO/IEC WG11 SC29), perciò è *Coding Tecnologies* proprietaria dello standard e detentrice di tutti i diritti su brevetti e patent licence. Chiaramente, sono sottoposte a royalties tutte le innovazioni riguardante la parte "PRO" dello standard, ossia gli algoritmi sviluppati per implementare la tecnica SBR - Spectral Band Replicator. Lo standard MP3 resta di proprietà ISO ed la gestione dei Patent Licence è sempre di competenza Fraunhofer-Thomson.

WMA

Windows Media Audio (WMA) è un formato proprietario completamente sviluppato e gestito da Microsoft, inglobato nel framework multimediale di Windows e nel formato di streaming ASF - Advanced System Format, utilizzabile solo tramite una serie di tools ed API dell'SDK Windows Media. Perciò, nel rispetto della politica aziendale, non esiste documentazione sulla struttura del codec, sugli algoritmi impiegati (banchi di filtri, modelli psicoacustici, ecc.) e sulla sintassi del formato; è dunque impossibile effettuare un'analisi dettagliata e comparativa di questo standard.

La necessità di pagare royalties per l'impiego della tecnologia MP3 ha indotto il gruppo di Bill Gates ha sviluppare uno proprio standard, ponendosi come obiettivo primario l'introduzione nel mercato multimediale di una codifica che potesse rappresentare una valida alternativa ad MP3 (rispetto ad MP3, WMA lavora mediamente meglio a bassi bitrate -sotto i 128Kbit/Sec-). Con l'avvento dei nuovi store musicali on-line (uno su tutti, Apple's iTunes) e la conseguente necessità di far percepire all'utente finale l'acquisto di musica ad alta qualità (in modo da giustificarne il pagamento), l'impiego di questa codifica sta via via diminuendo a favore di AAC.

Questo formato è supportato, oltre che da Windows Media Player, da una numerosa quantità di player SW ed HW attualmente presenti sul mercato, e tramite ASF, rappresenta lo standard di riferimento di molte radio web.

L'attuale programma di licenze Windows Media permette la distribuzione dei componenti in qualsiasi piattaforma, anche non Windows. Ciò significa che gli sviluppatori di applicazioni per piattaforme non attualmente supportate da Microsoft possono acquisire licenze per i componenti in formato Windows Media. Inoltre, Microsoft fornisce licenze per l'accesso a codici sorgente di specifici componenti in formato Windows Media per una determinata piattaforma, in modo da adattare e ottimizzare l'oggetto in questione.

Attualmente. esistono due differenti soluzioni di licenza, una rivolta alle piattaforme Windows ed una per qualunque altra piattaforma:

- per le soluzioni basate su Windows, viene fornito supporto integrato per i componenti in formato Windows Media (contenitori di file ASF, codec e protocolli) tramite Windows Media Format 9 Series SDK, che include Windows Media Device Manager SDK.
- per qualsiasi piattaforma (compresi i dispositivi hardware) e per solo utilizzo client, viene fornito l'accesso ai componenti in formato Windows Media (contenitori di file ASF, codec e protocolli), tramite programmi di licenze personalizzabili.

Andiamo ora ad analizzare più dettagliatamente la parte dei codec audio. Il contratto di licenza per la distribuzione di componenti audio in formato Windows Media è rivolto ai produttori che desiderano distribuire prodotti di supporto (ad esempio SDK o altri strumenti di sviluppo) o finali (ad esempio juke-box software o lettori DVD). Questa licenza concede ai produttori il diritto di distribuire prodotti di supporto, che non vengono venduti agli utenti e non richiedono il pagamento di diritti di utilizzo, oppure prodotti finali, che vengono venduti agli utenti e potrebbero richiedere il pagamento di diritti di utilizzo. Questa licenza copre l'utilizzo di componenti in formato Windows Media ricevuti sotto forma di oggetti da un altro produttore e di componenti in formato Windows Media sviluppati dal produttore stesso. Come per AAC, le tariffe dei prodotti di encoding-decoding audio sono differenziate in funzione del numero di canali supportati dal codec in questione (stereo,

stereo ad alta qualità o multicanale) e vengono applicate per unità acquistate. Nessuna tassa è applicata sulla distribuzione di materiali audio codificati con standard WMA o ASF.

OGG

OGG Vorbis (o, più brevemente OGG) è un sistema per la compressione audio free, open source, non proprietario e libero da vincoli di Patent Licence e royalties. Obiettivo di questo progetto (sotto licenza BSD) è la creazione di uno standard mondiale per la compressione audio *lossy* ad alta qualità di pubblico dominio, in modo da svincolare completamente gli sviluppatori di applicazioni multimediali da questioni e problematiche derivanti da patent licence e royalties. Per dirla in parole loro:

"Ogg Vorbis is a fully open, non-proprietary, patent-and-royalty-free, general-purpose compressed audio format for mid to high quality (8kHz-48.0kHz, 16+ bit, polyphonic) audio and music at fixed and variable bitrates from 16 to 128 kbps/channel. This places Vorbis in the same competitive class as audio representations such as MPEG-4 (AAC), and similar to, but higher performance than MPEG-1/2 audio layer 3, MPEG-4 audio (TwinVQ), WMA and PAC." -- Xiph.Org

Vorbis è un codec basato su codifica percettiva nel dominio frequenziale molto flessibile ed in grado di operare su un ampio range di valori di bitrate, frequenze di campionamento e codifiche di canale. Esistono due differenti versioni del codec: Vorbis I e Vorbis II. La prima effettua la trasformazione tempo-frequenza tramite una Modify Discrete Trasformate Cosene (MDCT); Vorbis II invece, sfrutta un Hybrid Wavelet Filterbanks per migliorare la codifica dei transienti presenti nel segnale audio. Come accade nei codec MP3, l'encoder -essendo provvisto di modello psicoacustico- è caratterizzato da una complessità computazionale maggiore rispetto al decoder.
Infine, è possibile creare file audio OGG Vorbis definendo tipo e valore di bitrate -CBR, VBR o ABR- o indicando un *quality index* (generalmente con valori compresi tra 1 e 10) che permette di configurare automaticamente l'encoder (definendo anche il valore di bitrate) al fine di fornire in output un file audio alla qualità voluta.
Come già detto precedentemente, OGG è nato per essere completamente libero da licenze. Perciò, chiunque volesse sviluppare e vendere un codec OGG Vorbis e/o integrarlo in un'applicazione multimediale può farlo senza la necessità e la preoccupazione di pagare royalties. Stesso discorso vale per la distribuzione di materiali audio (streaming, downloading, applicazioni ludiche, ecc.) codificati in OGG Vorbis. Occorre però fare una precisazione: lo sviluppo della libreria e dell'SDK è sotto licenza BSD, i tools SW invece devono includere OGG con licenza GNU GPL.

12.2 Principi di Psicoacustica

Per la natura fisica del nostro orecchio, molte informazioni contenute in un suono non vengono percepite; per esempio, noi non sentiamo intensità al di sotto di una certa soglia in funzione della frequenza oppure percepiamo come più "forti" toni con frequenza appartenente allo spettro vocale. Ciò significa che del segnale audio originale è necessario memorizzare soltanto le informazioni effettivamente percepite, eliminando tutto il resto, ed ottenendo così un'elevata compressione in cui la perdita d'informazione c'è ma, *in linea di principio*, non si sente. *In linea di principio* perché, dato un segnale audio esiste sempre un bitrate minimo al di sotto del quale la compressione diventa percepibile (generalmente, al di sopra dei 256 Kbit/sec si ha un ottimo segnale compresso); inoltre per suoni in cui c'è una forte predominanza delle alte frequenze l'algoritmo MPEG lavora oggettivamente male.

Colui che si occupa di decidere cosa eliminare e cosa no è il **modello psicoacustico**, elemento fondamentale di MPEG / Audio che determina fortemente la bontà di compressione di un Encoder. Le tecniche utilizzate per lo sviluppo di questo blocco sono in continua evoluzione, di pari passo con gli studi psicoacustici, ma i principi che stanno alla base sono sempre gli stessi. Qui di seguito ne verrà data loro una breve descrizione.

Bande Critiche e Scala di Bark

Suddividono lo spettro frequenziale dell'orecchio umano (circa 16Hz - 20 Khz) in sottobande di dimensione diversa ed all'interno delle quali la dinamica di S/N (rapporto Segnale/Rumore) varia nel seguente modo:

"se viene emesso un tono ed un rumore bianco, l'S/N che noi percepiamo non viene calcolato considerando tutto lo spettro del rumore ma soltanto una zona particolare, situata nell'intorno del tono, detta appunto *banda critica*".

In pratica l'orecchio decide se un tono è presente o assente basandosi su un S/N calcolato solo sulla parte di spettro in cui è situata la banda critica.

La **scala di Bark** fornisce una delle possibili rappresentazioni delle bande critiche in funzione della frequenza in Hertz. Una possibile associazione tra frequenze lineari e scala di Bark può essere eseguita osservando la seguente tabella:

Band Number	F_{low}	F_{high}	F_{center}	? F
0	0	50	25	50
1	50	95	22,5	45
2	95	140	22,5	45
3	140	235	47,5	95
4	235	330	47,5	95
5	330	420	45	90
6	420	560	70	140
7	560	660	50	100
8	660	800	70	140
9	800	940	70	140
10	940	1125	92,5	185
11	1125	1265	70	140
12	1265	1500	117,5	235
13	1500	1735	117,5	235
14	1735	1970	117,5	235
15	1970	2340	185	370
16	2340	2720	190	380
17	2720	3280	280	560
18	3280	3840	280	560
19	3840	4690	425	850
20	4690	5440	375	750
21	5440	6375	467,5	935
22	6375	7690	657,5	1315
23	7690	9375	842,5	1685
24	9375	11625	1125	2250
25	11625	15375	1875	3750
26	15375	20250	2437,5	4875

Tabella 2.4: scala di Bark

Si può notare come per le basse frequenze, le dimensioni delle bande (? F) siano piccole mentre per le alte frequenze, siano molto più grandi. Ciò mostra come le basse frequenze, avendo una maggiore risoluzione frequenziale, siano trattate meglio dal nostro orecchio.

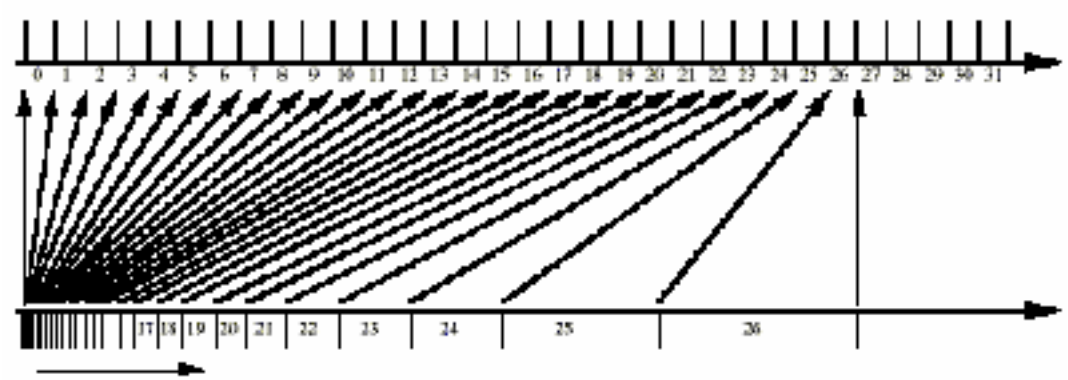


Fig 2.12: rapporto tra le frequenze rappresentate su scala lineare (sopra) e scala di Bark (sotto)

Loudness ed Absolute Threshold

L'intensità del suono percepito dal nostro orecchio (Loudness) varia in funzione della frequenza e ciò significa che a parità di Sound Pressare Level (SPL), due toni di diversa frequenza vengono percepiti con diverse intensità. In genere vengono percepiti meglio suoni la cui frequenza si trova nello spettro vocale mentre per sentire toni ad alte frequenze il nostro orecchio ha bisogno di un SPL elevato.

La curva in figura da una rappresentazione qualitativa dell'andamento della percezione del suono in funzione della frequenza in stato di quiete:

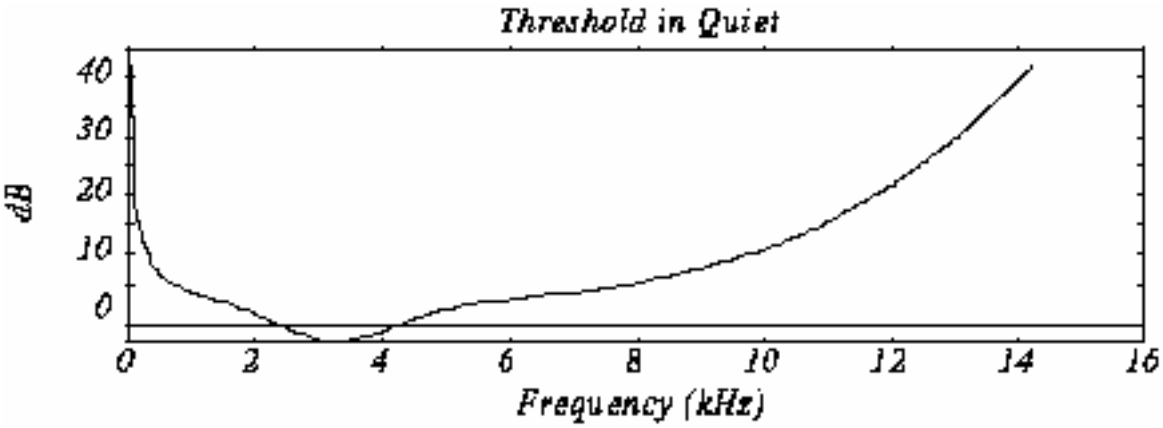


Fig 2.13: curva di percezione del suono del nostro orecchio in stato di quite

In stato di quiete, tutto ciò che sta sotto la curva (**detta Absolute Threshold o Threshold in Quiet)** viene percepito dal nostro orecchio come silenzio e quindi può essere tranquillamente eliminato in fase di codifica. Per esempio, per sentire un tono a 10 Khz è necessario che esso abbia un'ampiezza superiore a circa 10 DB.

Mascheramento

Se si è in presenza di due toni emessi nello stesso istante, quello con intensità maggiore (segnale maschera) renderà non udibile quello di intensità minore. Questo evento viene definito **mascheramento** e permette di eliminare delle componenti in funzione del particolare tipo di segnale che si sta analizzando. Esistono due tipi di mascheramento: quello frequenziale e quello temporale (pre e post).

Il mascheramento frequenziale si verifica quando si ha la presenza contemporanea di un segnale forte (lo Strong Tonal Signal in figura 2.14) ed uno debole (il Weaker Signal in figura 2.14) con frequenze molto vicine tra loro (appartenenti alla stessa banda critica o a bande critiche adiacenti). In tal caso il segnale forte produce una soglia di mascheramento al di sotto della quale il suono risulta non udibile. E' importante notare dalla figura 2.14 che in presenza di un tono maschera, le alte frequenze vengono maggiormente mascherate rispetto a quelle basse.

Il modello psicoacustico di MPEG utilizza il concetto di componente tonale e non tonale per determinare se all'interno di un banda critica è presente un segnale di maschera..

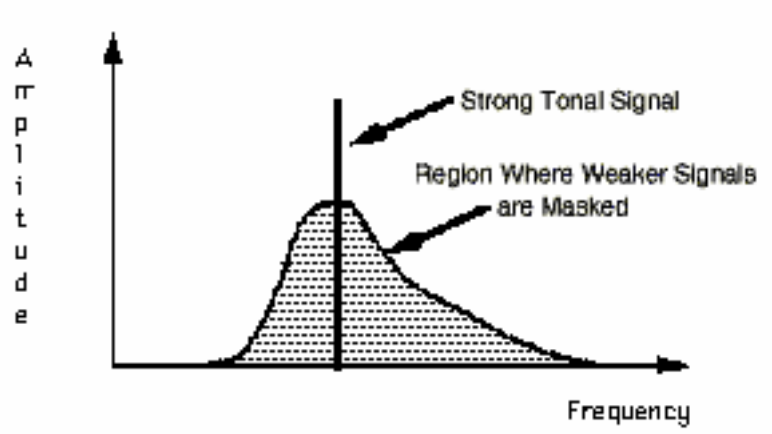


Fig 2.14: mascheramento frequenziale effettuato su una banda critica

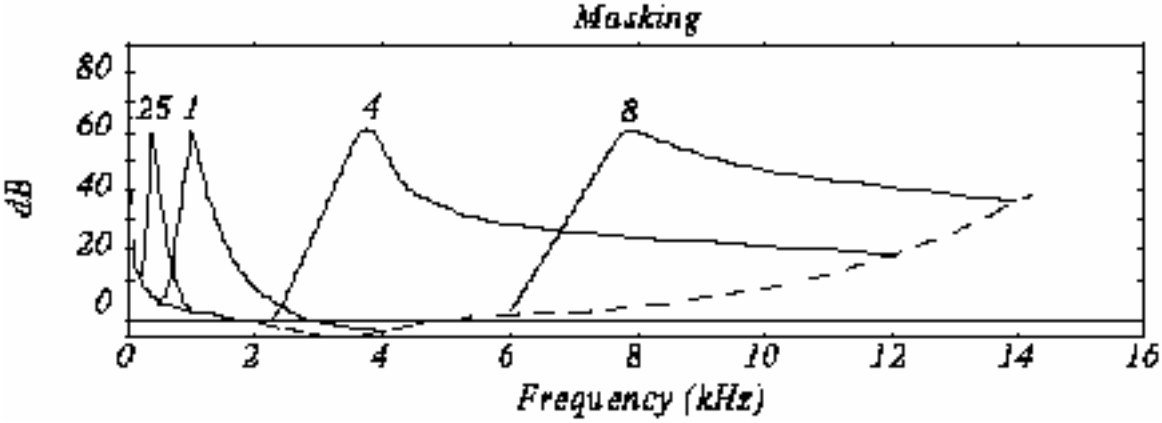


Fig 2.15: mascheramento frequenziale effettuato su più bande critiche

Il mascheramento temporale fa si che la presenza di un tono di maschera continui a rendere non udibili i toni adiacenti anche quando questo non è fisicamente presente. Esso può verificarsi sia prima che il segnale di maschera si attivi (pre-masking), sia dopo la sua disattivazione (post-masking). Il pre-masking ha una durata di pochissimi millisecondi, il post-masking invece può durare dai 20 msec (toni di maschera poco intensi) ai 500msec (toni di maschera molto intensi).
Il mascheramento temporale deriva dal fatto che le reazioni dei nervi uditivi agli stimoli esterni non sono istantanee.

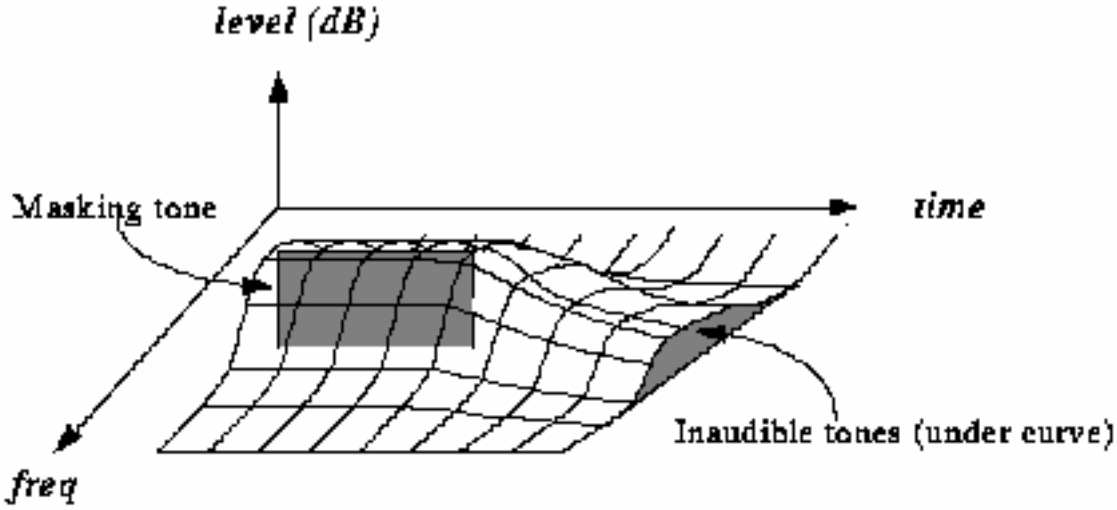


Fig 2.16: mascheramento frequenziale e temporale

Componente Tonale / non Tonale

Ci sono vari modi per definire questo concetto. Uno dei più semplice consiste nel suddividere lo spettro in bande di 1/3 d'ottava; ora, se il livello di pressione (SPL) di una o più bande supera per più di 5 Db quello delle due bande adiacenti si è in presenza una **Componente Tonale**; in caso contrario si ha una **Componente non Tonale**.
I modelli psicoacustici utilizzano questo concetto per determinare se, in una particolare banda critica, è presente (**Componente Tonale**) o assente (**Componente non Tonale)** un tono di maschera.

12.2.1 Cenni di fisiologia dell'orecchio: proprietà dei nervi uditivi

Le teorie psicoacustiche finora citate hanno trovato riscontro nello studio *fisiologico dell'orecchio*.

I nervi uditivi, stimolati dalla coclea, hanno delle proprietà molto interessanti che di fatto giustificano e dimostrano la validità degli esperimenti psicoacustici:

- **Adattamento:** il nervo appena stimolato emette un'enorme quantità di impulsi elettrici per poi diminuirli fino ad un valore stazionario (*steady tone)*.
- **Tuning**: i nervi hanno una frequenza in cui sono più sensibili (*Frequenza Caratteristica)* e questo giustifica le bande critiche e l'uso dei banchi di filtri.
- **Sincronismo**: un nervo stimolato con lo stesso tono più volte consecutive, si adatta ogni volta (*proprietà di adattamento*) sincronizzandosi con gli stimoli.

Le quattro proprietà sottoelencate hanno in comune la **non linearità:**

- **Saturation:** il numero di impulsi (*picchi*) che un nervo può generare in un certo istante di tempo non può superare un valore massimo. Ciò giustifica come il nostro orecchio abbia una dinamica limitata.
- **Soppressione tra due toni:** se ad un tono ne aggiungo un altro, il nervo diminuisce il numero di impulsi generati, facendo così percepire meno il primo tono emesso.
- **Mascheramento di un tono da rumore:** un nervo reagisce maggiormente se si è in presenza di un tono puro rispetto ad un tono con presenza di rumore.
- **Combination Tones:** se due toni hanno frequenze F1 ed F2 lontane da *frequenza caratteristica* nervo, ciò che si percepisce è una combinazione di queste due frequenze.

3. Valutazione oggettiva e soggettiva della qualità audio percepita

I sistemi per la valutazione della qualità audio percepita da sistemi di codifica lossy possono essere suddivisi in due categorie: metodi oggettivi e metodi soggettivi. I primi si basano su modelli acustici e psicoacustici volti a fornire una serie di parametri numerici circa la bontà del suono percepito; i secondi invece si basano sulle osservazioni fornite da un campione di persone sottoposte ad una serie di prove d'ascolto, effettuate nel rispetto di un set di regole ben definite.
Qui di seguito analizzeremo un metodo oggettivo, molto semplice, basato sulla risposta dei sistemi di compressione sollecitati da un segnale multitonale; successivamente tratteremo le recommendations ITU-R BS.1387, BS.1116 e BS.1254, sistema di riferimento per la valutazione oggettiva e soggettiva, sia per MPEG che per tutta l'industria operante nel settore del multimedia.

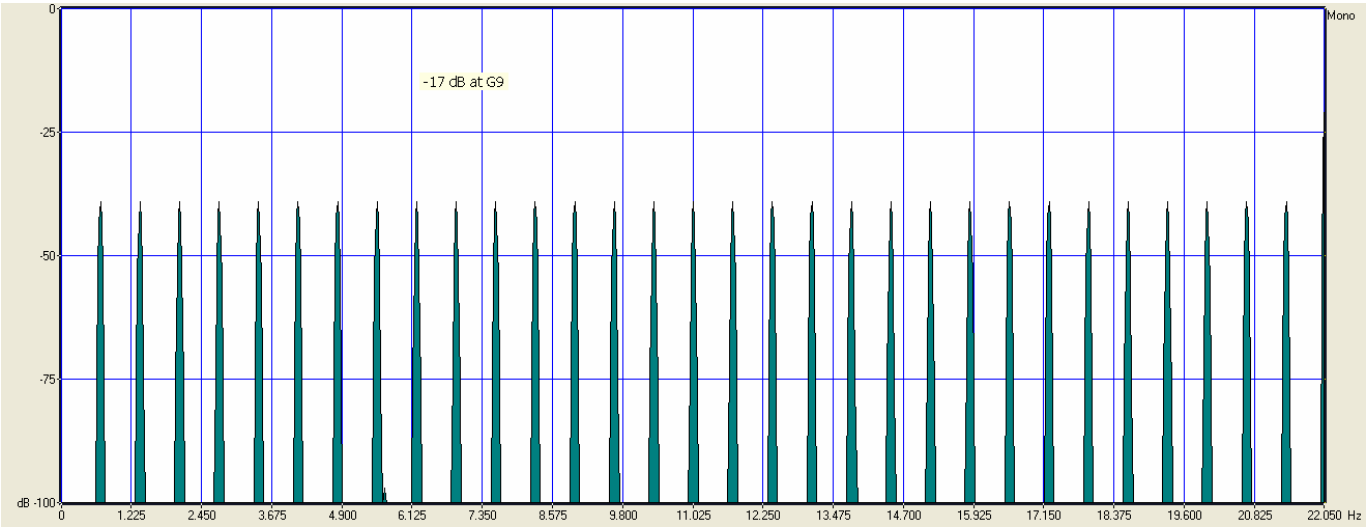
3.1 Valutazione oggettiva tramite segnali multitonali

Per la valutazione del comportamento di un sistema lineare tempo-invariante si fa generalmente uso di un segnale impulsivo. L'impulso viene dato in input al sistema ed in base alla risposta data in ouptut se ne valuta il comportamento.
I codec percettivi non sono sistemi lineari; perciò è necessario trovare un differente segnale la cui analisi, prima e dopo la compressione permetta di estrarre informazioni significative. Il segnale multitonale -toni equispaziati nello spettro- rappresenta un metodo molto semplice da applicare, che permette di effettuare alcune semplici osservazioni sulla bontà di un compressore audio lossy.
L'esperimento consiste nel creare sinteticamente un segnale multitonale a toni equispaziati, salvarlo in formato WAV (PCM) ed analizzarne lo spettro. Successivamente, il segnale viene compresso-decompresso, e rianalizzato nel dominio delle frequenze. Infine, viene effettuato un confronto tra lo spettro del segnale multitonale originale e quello ottenuto dalla compressione-decompressione, traendo così le conclusioni.
Qui di seguito è fornito il codice CSound (del solo file ORC) per la generazione del nostro segnale multitonale PCM, mono, 16 bit a 44,1Khz.

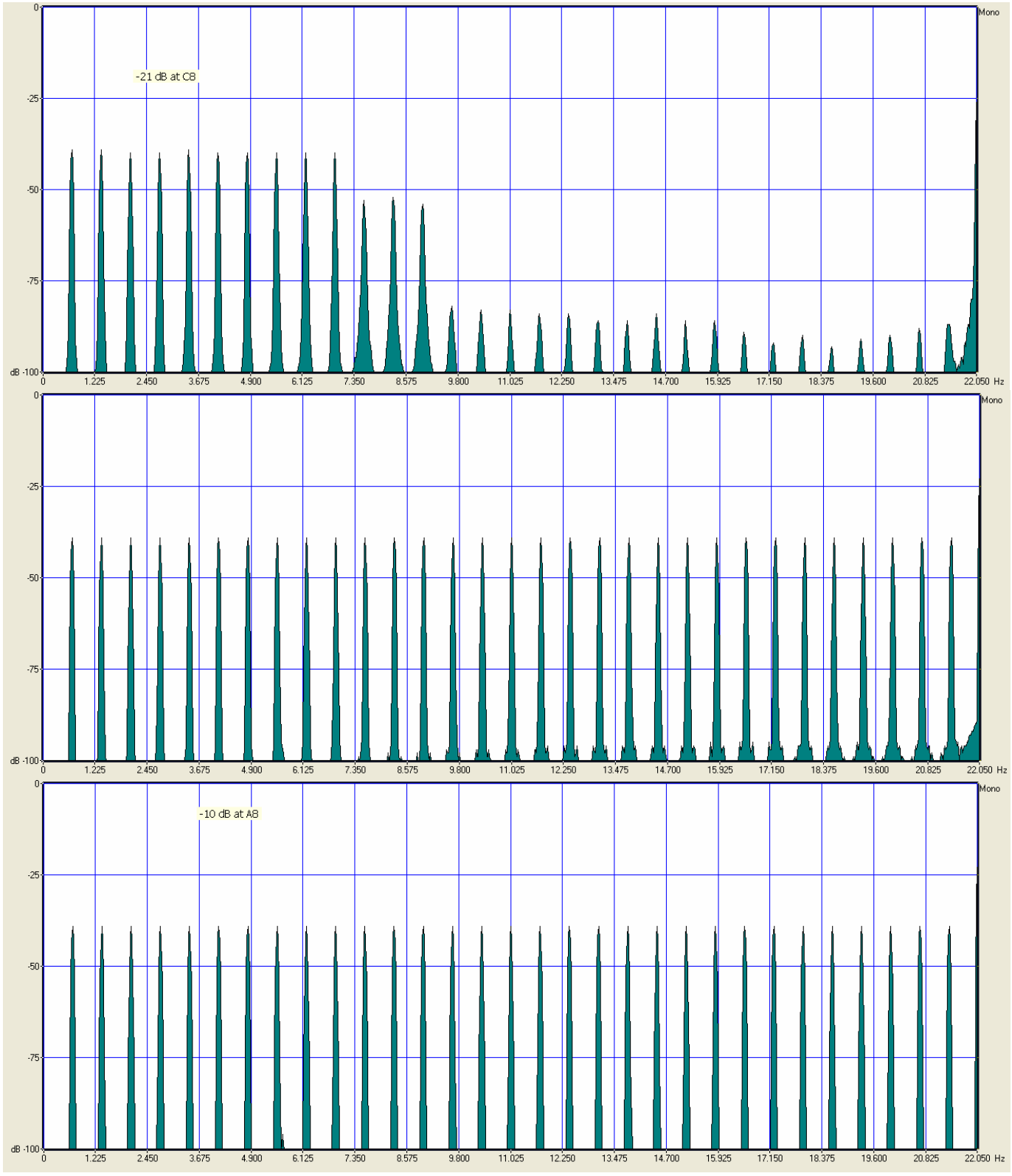
```
sr = 44100
kr = 4410
```

```
      ksmps = 10
      nchnls = 1

      instr 1
a1    oscil 1000, 689.0625, 1
a2    oscil 1000, 689.0625 * 2, 1
.................................
.................................
a32   oscil 10000, 689.0625 * 32, 1

      out a1 + a2 + ..... + a32
      endin
```

In figura lo spettro del segnale multitonale considerato. Esso è caratterizzato da 32 toni equispaziati posizionati ai bordi delle bande del banco di filtri polifasico MP3, ossia in rapporto 1 / 689,0625Hz. Ciò permette di valutare sommariamente la risposta del sistema nelle varie sottobande, analizzandone il rumore di quantizzazione introdotto.



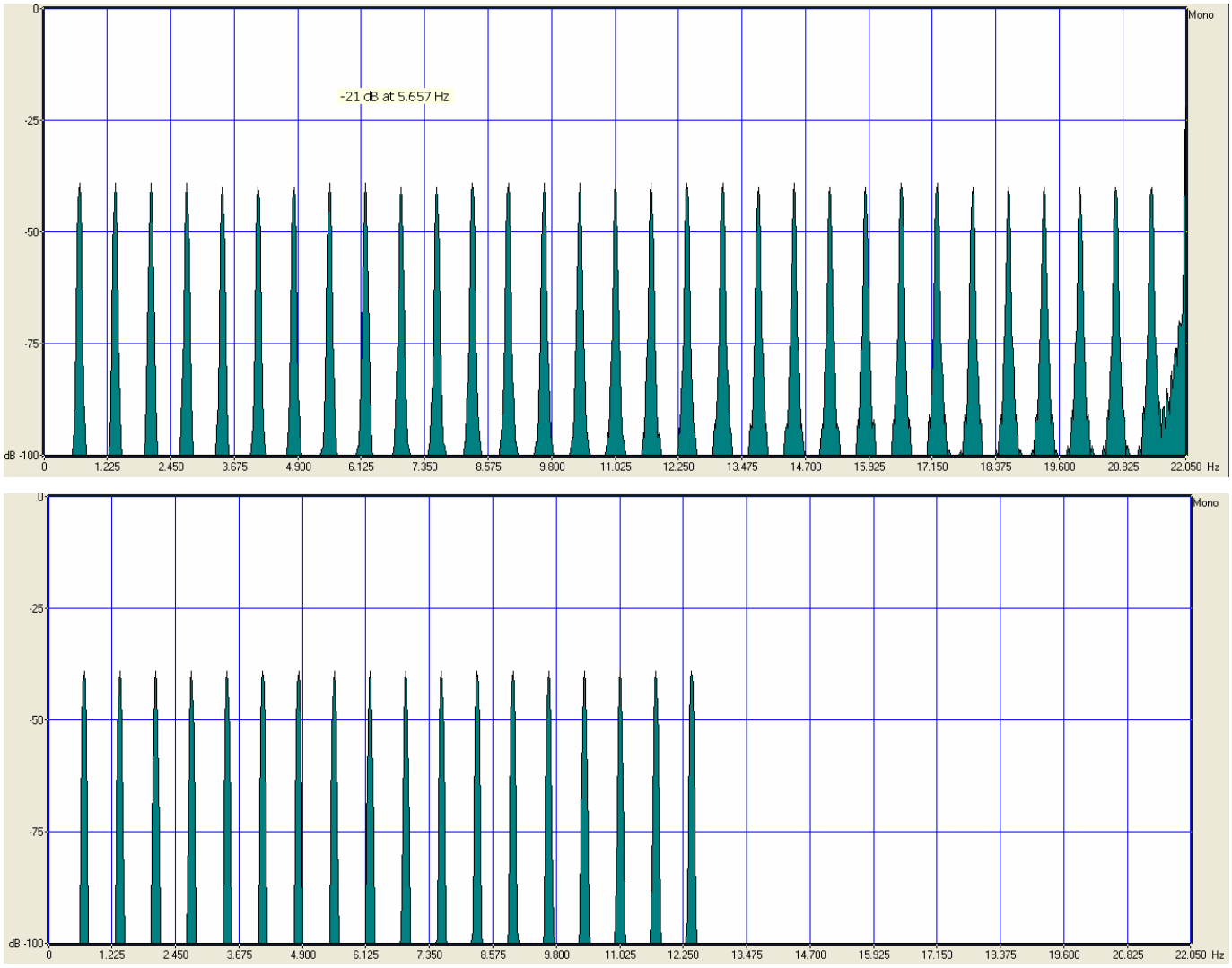
Come primo esempio, vediamo come si comporta il compressore MP3 (winLAME per Win32) cambiando come parametro di input il solo valore di bitrate: 32Kbit/sec (prima immagine), 128Kbit/sec (seconda immagine) e 320Kbit/sec (terza immagine).



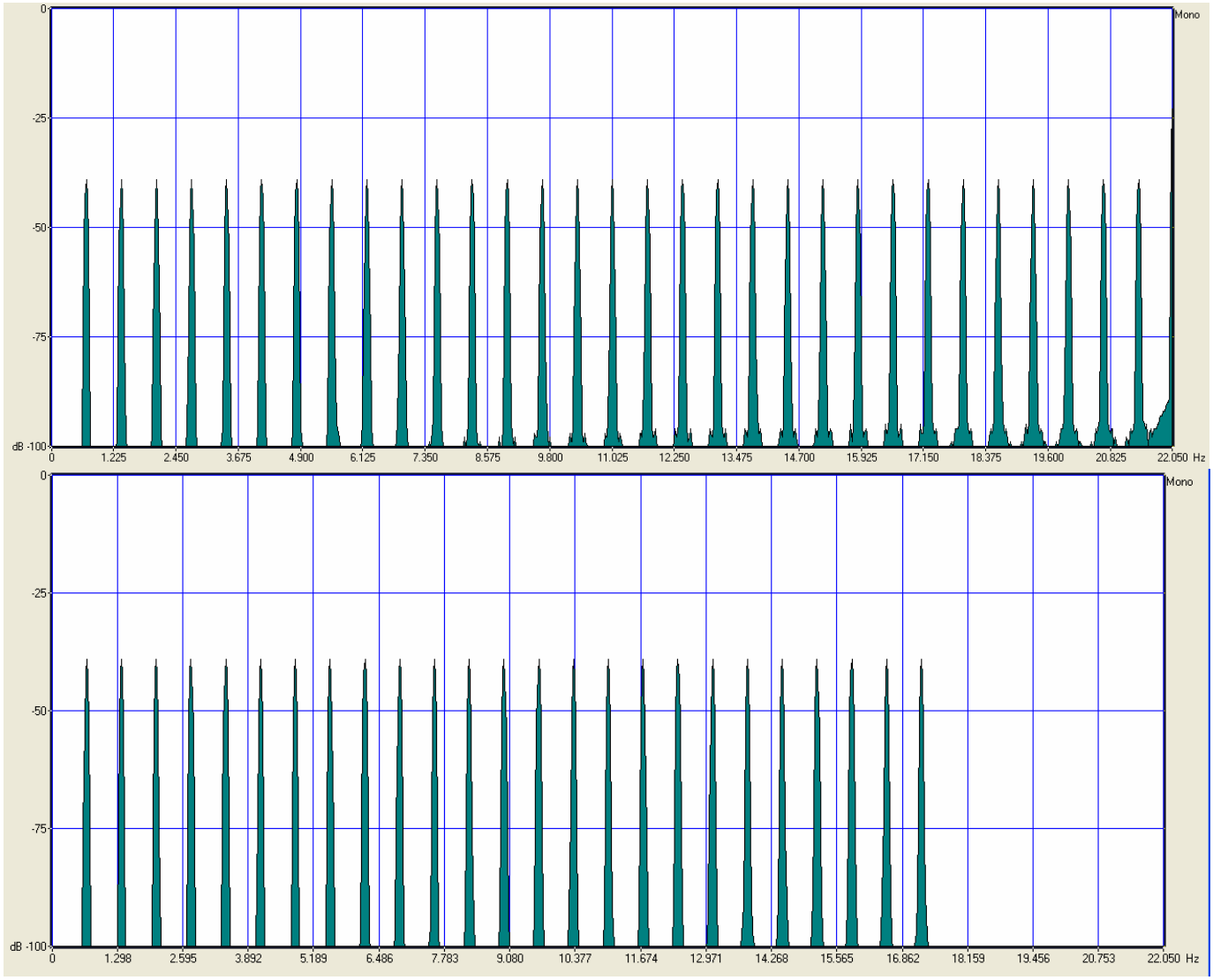
E' possibile notare come il segnale rimanga intatto a 320, inizi ad avere dei piccoli (ma già visibili) rumori di quantizzazione (maggiormente accentuate nelle altre frequenze) a 128, e come venga totalmente distorto a 32.

Da questo semplice ma efficace esempio, si evince come l'algoritmo MP3 tenda a privilegiare fortemente le basse frequenze a scapito di quelle alte.

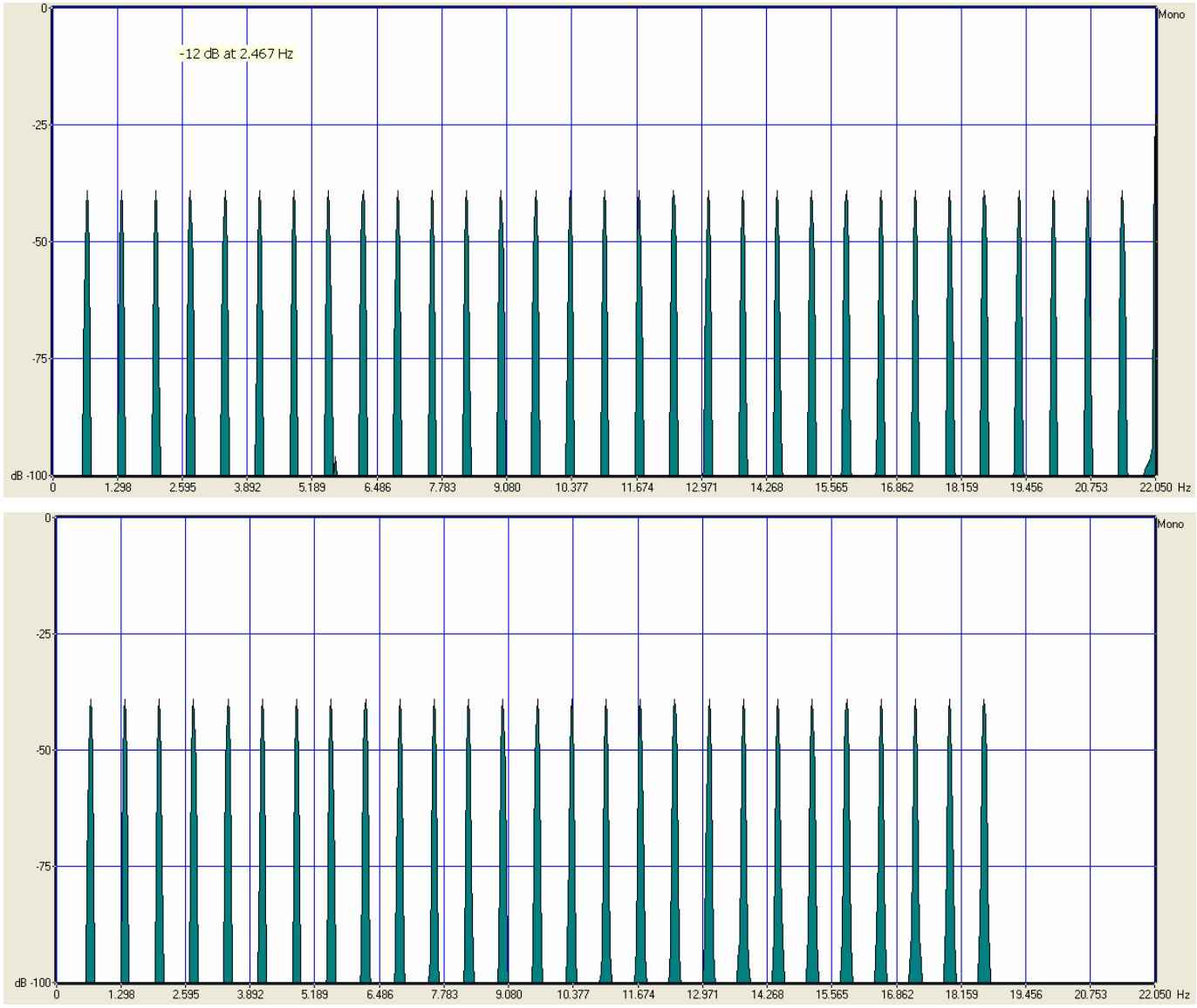
Analizziamo ora un altro esempio che confronta i comportamenti di due differenti algoritmi di compressione: MP3 e WMA. Come compressore MP3 è stato impiegato il solito winLAME; per WMA è stato utilizzato il sistema Windows Media 9 fornito da SoundForge 6.0.



La figura in alto mostra la risposta ottenuta con codec MP3 a 64 Kbit/sec. La seconda invece mostra la risposta del codec WMA sempre a 64 Kbit/sec.



La figura in alto mostra la risposta ottenuta con codec MP3 a 128 Kbit/sec. La seconda invece mostra la risposta del codec WMA sempre a 128 Kbit/sec.



La figura in alto mostra la risposta ottenuta con codec MP3 a 160 Kbit/sec. La seconda invece mostra la risposta del codec WMA sempre a 160 Kbit/sec.

Analizzando semplicemente gli spettri mostrati qui sopra, si può constatare come il modello psicoacustico presente in WMA elimini senza nessun problema le alte frequenze ritenute irrilevanti, indipendentemente dal bitrate; tale situazione non si verifica in MP3. A 64 Kbit/sec, MP3 introduce un rumore di quantizzazione maggiore di WMA; con l'aumento del valore di bitrate il rumore di quantizzazione introdotto da MP3 diminuisce fino quasi a scomparire a 160Kbit/sec. WMA invece, pur non introducendo rumore significativo, continua a non considerare le alte frequenze, modificando di fatto la natura del segnale. Si può perciò concludere che a bassi bitrate WMA funziona leggermente meglio; con l'aumento del bitrate, MP3 vince piuttosto nettamente.

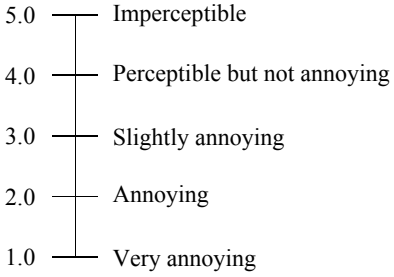
3.2 Valutazione oggettiva e soggettiva secondo le recommendations ITU-R

ITU (International Telcommunication Union) è un organo internazionale che gestisce progetti ed iniziative tecnologie nell'ambito delle telecomunicazioni. ITU si suddivide in tre parti distinte: ITU-D (ITU Development), ITU-L (ITU Standardization) ed ITU-R (ITU Recommendation).
Le specifiche riguardanti i test soggettivi ed oggettivi dei codec audio rientrano nelle Recommendation (ITU-R) in quanto, per le industrie operanti nel settore, esse non definiscono dei metodi standard da dover applicare meticolosamente, ma delle accurate linee guida da seguire (con eventuali esempi di implementazione e test) durante la progettazione e lo sviluppo di metodi per i test oggettivi e soggettivi della qualità audio percepita da sistemi audio.

ITU-R BS.1116, ITU-R BS.1254: Metodi di Test Soggettivi

ITU-R BS.1116 (Method for the Subjective Assessment of Small Impairments in Audio System Including Multichannel Sound System) ed ITU-R BS.1254 (General Methods for the Subjective Assessment of Sound Quality) definiscono delle linee guida sui test d'acolto e la valutazione soggettiva della qualità audio percepita.
Un tipico test d'acolto può essere così brevemente descritto: l'ascoltatore ha la possibilità di selezionare tre differenti sorgenti audio: "A", "B" e "C". Il *reference signal* (segnale originale) è sempre disponibile dalla sorgente "A" mentre alle altre due sorgenti vengono assegnati "casualmente" il *reference signal* ed il *signal under test* (per esempio, il segnale in uscita da un codec audio lossy). All'ascoltare viene chiesto di valutare l'indebolimento tra i segnali presenti nelle sorgenti "A - B" ed "A - C", in accordo con una scala nota come per esempio quella mostrata qui sotto.

The ITU-R five-grade impairment scale



L'output dei risultati d'analisi così ottenuti viene generalmente espresso con un indice detto di SDG (Subjective Difference Grade), definito come segue:

$$SDG = Grade_{Signal\ Under\ Test} - Grade_{Reference\ Signal}$$

Tale indice dovrebbe idealmente variare tra 0 e -4, dove lo 0 indica una differenza impercettibile tra i due segnali.

Un ultimo importante aspetto da considerare riguarda la scelta dei reference signals da utilizzare durante i test. La recommendation ITU ne segnala e fornisce alcuni, suddivisi tra segnali naturali e segnali sintetici. Tali segnali sono pensati in modo tale da presentare una serie di artefatti audio dopo la compressione:

- Transienti: presenza o meno di effetti di pre-cho
- Struttura tonale: rudezza del suono, sensibilità al rumore.
- Natural Speech: qualità degli attacchi, sensibilità alla distorsione, intelliggibilità del parlato
- Complex Sound: valutazione del sistema sotto stress
- High Bandwidth: con sistema sotto stress, perdita o meno della larghezza di banda del segnale, con particolare attenzione alle alte frequenze.

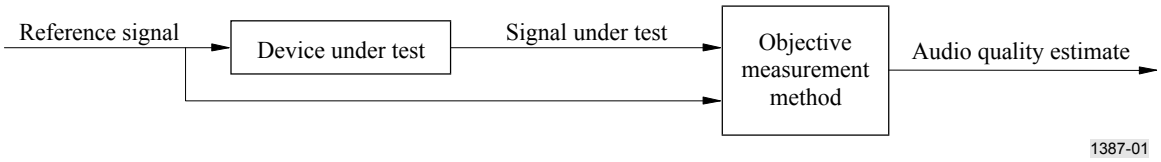
Infine, tali segnali dovrebbero avere durata massima di 10-20, con una breve parte critica.

E' disponibile un CDDA contenente tutte le tracce fornite ed utilizzate dalla EDU (European Broadcasting Union), all'interno del progetto SQAM (Sound Quality Assesstment Audio Material) per la valutazione qualitativa dei codec audio MPEG. Alcune di queste tracce sono scaricabili dal link http://www.tnt.uni-hannover.de/project/mpeg/audio/sqam/

ITU-R BS.1387: Metodi di Test Oggettivi

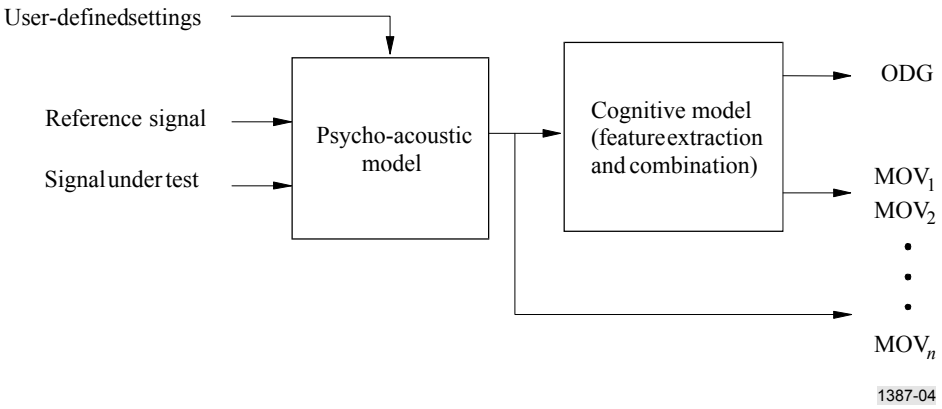
ITU-R BS.1387 (Method for Objective Measurement of Perceived Audio Quality) specifica un metodo per la valutazione oggettiva della qualità audio percepita da sistemi audio sotto stress (esempio: codec audio, catene audio digitali, ecc.) che va a sostituire i classici metodi del Signal to Noise Ratio (SNR), e del Total Harmonic Distortion (THD). Generalmente, il set di segnali audio utilizzato è il medesimo di quello impiegato nei test soggettivi. In figura è mostrata lo schema a blocchi di un tipico sistema per la misurazione oggettiva della qualità audio percepita.

Basic concept for making objective measurements



Il *reference signal* ed il *signal under test* vengono analizzati in parallelo tramite banchi di filtri ed FFT per poi modellare le tipiche proprietà acustiche e psicoacustiche del sistema uditivo tramite modello percettivi. Al termine del processo viene fornito un insieme di valori detti MOV (Model Output Variable) ed una loro sintesi -derivata sia dall'analisi in FFT che itramite banchi di filtri- detta ODG (Objective Difference Grade).

Stages of processing implemented in the model



In tabella, un esempio di alcuni dei parametri forniti in output dal sistema ITU-R BS. 1387 (MOV).

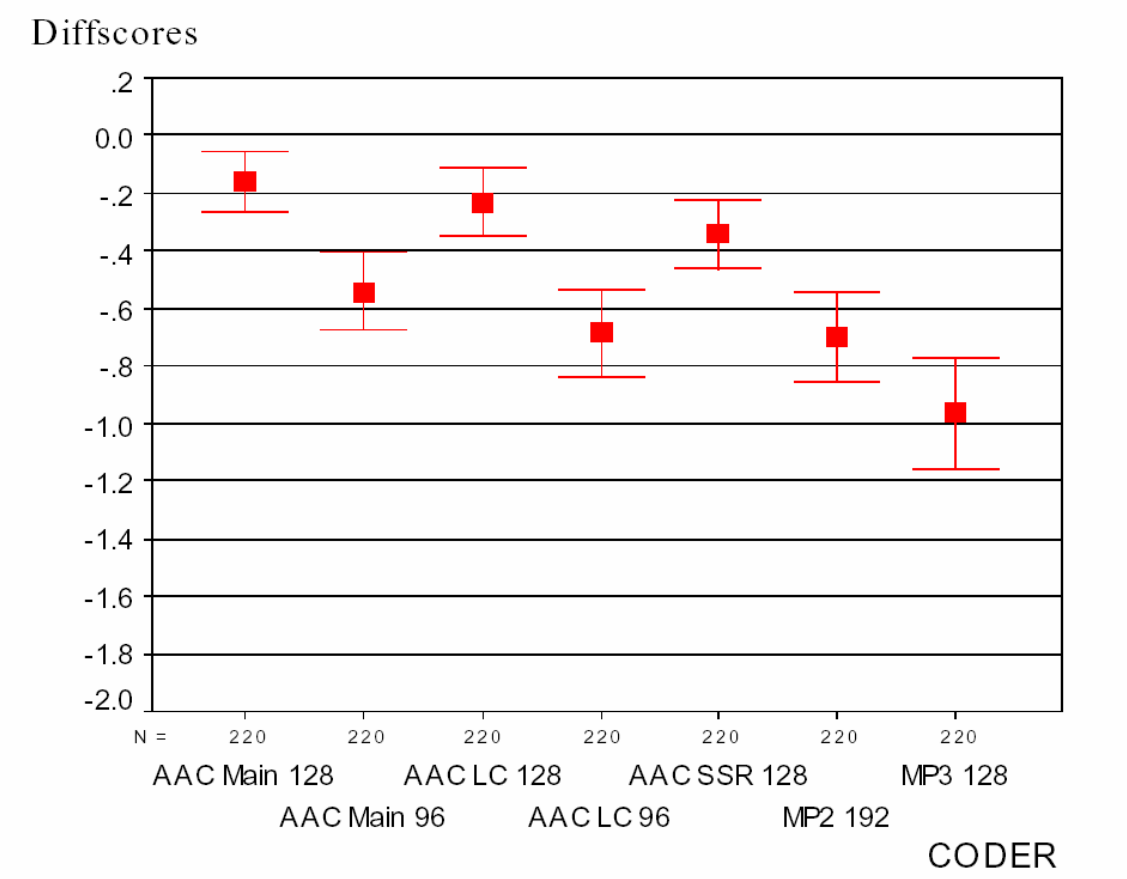
Model Output Variable	Description
.....	.....
$RmsModDiff_A$	*Rms value of the modulation difference*
$RmsMissingComponents_A$	*Rms value of the noise loudness of missing frequency components, (used in* $RmsNoiseLoudAsym_A$*)*
$RmsNoiseLoud_B$	*Rms value of the averaged noise loudness with emphasis on introduced components*
$AvgLinDist_A$	*A measure for the average linear distortions*
$BandwidthRef_B$	*Bandwidth of the Reference Signal*
$BandwidthTest_B$	*Bandwidth of the output signal of the device under test*
$TotNMR_B$	*logarithm of the averaged Total Noise to Mask Ratio*
.....	.....

L'indice ODG è un valore numerico decimale (ad una cifra) che fornisce un tasso di qualità del segnale audio percepito. Esso è dunque facilmente confrontabile con SDR, valore ottenuto dai test soggettivi.

E' importante sottolineare che i test oggettivi non sostituiscono quelli soggettivi ma, piuttosto, essi ne rappresentano una loro estensione, permettendo così una valutazione più corretta e completa dei sistemi audio. Essendo ODG ed SDG facilmente confrontabili, è molto importante calcolarne la correlazione: maggiore sarà questo valore, più veritieri saranno i test oggettivi e soggettivi.

MP3 vs AAC

Tra i numerosi test acustici basati sulle recommendation ITU, qui ne riportiamo uno dei più significativi rispetto al mondo MPEG: MP3 vs AAC. Molte prove sono stati effettuate sullo standard AAC per capire realmente, se e di quanto migliorasse la qualità audio rispetto alle precedenti codifiche MPEG-1 ed MPEG-2. A parità di condizioni (stesso valore di bitrate e frequenza di campionamento, medesimi brani audio, uguali caratteristiche dell'ambiente, stessi tester, ecc.), 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 ed un MP2 a 192 Kbit/sec.


Diffscores
.2
0.0
-.2
-.4
-.6
-.8
-1.0
-1.2
-1.4
-1.6
-1.8
-2.0
N =
220
220
220
220
220
220
220
AAC Main 128
AAC LC 128
AAC SSR 128
MP3 128
AAC Main 96
AAC LC 96
MP2 192
CODER

4. Metadati Audio

La descrizione testuale di informazione audio musicale o, più in generale, di realtà multimediali, è di fondamentale importanza nei vari settori del multimedia ove la ricerca ed il reperimento dell'informazione giochi un ruolo primario. Si pensi per esempio alle interrogazioni in DB multimediali, oppure ai sistemi per il music on-demand o ancora, ai sistemi P2P per il downloading. In tutti questi ambiti è di fondamentale importanza la descrizione dell'informazione multimediale in fornato alfanumerico, col fine di facilitarne la ricerca ed il reperimento. Qui di seguito vedremo tre sistemi per la rappresentazione dei metadati multimediali: ID3, MPEG-7 ed MX. Il primo è un sistema molto semplice e pensato per i formati MP3 ed AAC; il secondo invece è molto più ampio e sviluppato per la descrizione di una generica realtà multimediale. Infine MX, è un linguaggio XML per la rappresetazione multistrato dell'informazione musicale, attualmente in fase di sviluppo presso il LIM - Laboratorio di Informatica Musicale.

4.1 ID3: metadati audio per MP3 ed AAC

Una delle pecche dello standard MPEG/Audio Layer 3 ed AAC è sicuramente la totale mancanza di strutture dati testuali contenenti informazioni riguardanti il contenuto di un file MP3 / AAC. Infatti tutto ciò che è stato incluso a riguardo, sono due bit indicanti la presenza di Copyright e l'originalità del pezzo audio. Con il proliferarsi del formato e la conseguente necessità di catalogare file MP3 /AAC all'interno di Database, è stato necessario includere un qualcosa che permettesse, sia ai DBMS che agli utenti veri e propri, di conoscere tutte quelle informazioni di fondamentale importanza per l'identificazione e la catalogazione di brani audio (Titolo, Autore, Genere, ecc.).
La risposta è stata la nascita dello standard ID3 (sviluppata da *Eric Kemp)* la cui prima versione (ID3 V1) permette di salvare il nome dell'autore e del brano, la data di pubblicazione, ecc. negli ultimi 128 byte di un file MP3 (è stata posta in fondo per evitare problemi di compatibilità con i decoder che, nel periodo in cui questo standard uscì, si aspettavano come primo byte, l'inizio del primo frame).

La struttura completa di ID3 V1 è la seguente:

Song title	30 characters
Artist	30 characters
Album	30 characters
Year	4 characters
Comment	30 characters
Genre	1 byte

L'ovvia evoluzione (128 byte non erano sufficienti) fu ID3 V2; essa è anteposta al bitstream MP3, ha una dimensione variabile ed è strutturata a *chunk* ognuno dei quali permette di inglobare tutte le informazioni contenute in ID3 V1 più ulteriori campi come il nome dell'encoder utilizzato, informazioni sui diritti di copyright, informazioni sull'artista, un eventuale sito web di riferimento, ecc..

Le due strutture sono completamente indipendenti tra di loro: è possibile ometterle, metterne solo una, oppure entrambe.

La struttura di un file MP3 con l'aggiunta di ID3 Tag è la seguente:

ID3 V2 (Dimensione Variabile)
Streaming Audio MPEG Layer 3
ID3 V1 (128 byte)

4.2 MPEG-7 (ISO/IEC 15938) - "Multimedia Content Description Interface"

Lo standard MPEG-7 [ISO/IEC 15938], formalmente chiamato "*Multimedia Content Description Interface*", fornisce un set di Description Tools per la descrizione di contenuti multimediali audio-video (AV) a livello simbolico e di metadato.

A differenza degli standard MPEG precedenti -i quali hanno avuto come obiettivo lo sviluppo di algoritmi di compressione (MPEG-1 ed MPEG-2) e l'organizzazione in *oggetti* di realtà multimediali eterogenee (MPEG-4)- MPEG-7 si pone come fine ultimo la descrizione di informazione multimediale attraverso una rappresentazione testuale (XML) che ne permetta una

semplice ed immediata ricerca e navigazione rispetto ai contenuti, e non in funzione della propria struttura fisica (per esempio, un insieme di numeri rappresentanti la forma d'onda o lo spettro di un segnale audio).
Ulteriori approfondimenti: dispensa "MPEG-7 Tutorial"

5. MPEG-21 "Multimedia Framework"

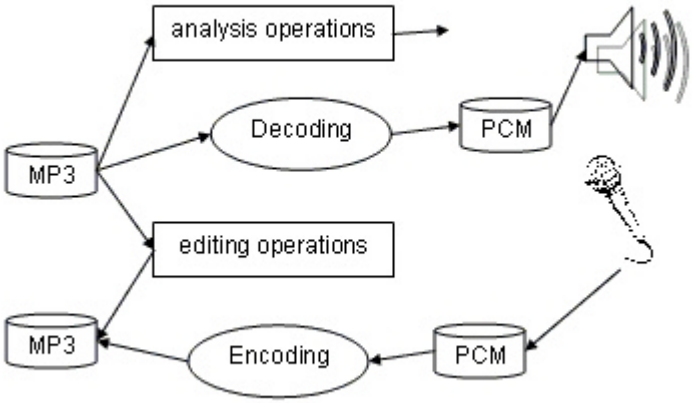
Gli standard MPEG analizzati fin'ora, trattano i contenuti multimediali soltanto da un punto di vista fisico (MPEG-1, MPEG-2, MPEG-4) e semantico (MPEG-7) mentre tutte le problematiche inerenti la distribuzione dei contenuti in funzione del proprietario (diritti, copyright, ecc.) non vengono mai prese in considerazione.
Lo standard MPEG-21, partito nel Giugno 2000, ha come obiettivo quello di risolvere questi aspetti con lo sviluppo di un framework multimediale (*Multimedia Framework*) che fornisca all'utente un supporto per lo scambio, l'accesso, il consumo, il commercio ed ogni altro tipo di operazione inerente il multimedia, che sia efficiente, trasparente ed indipendente dalla piattaforma Hw/Sw utilizzata.
MPEG-21 si basa su due concetti fondamentali:

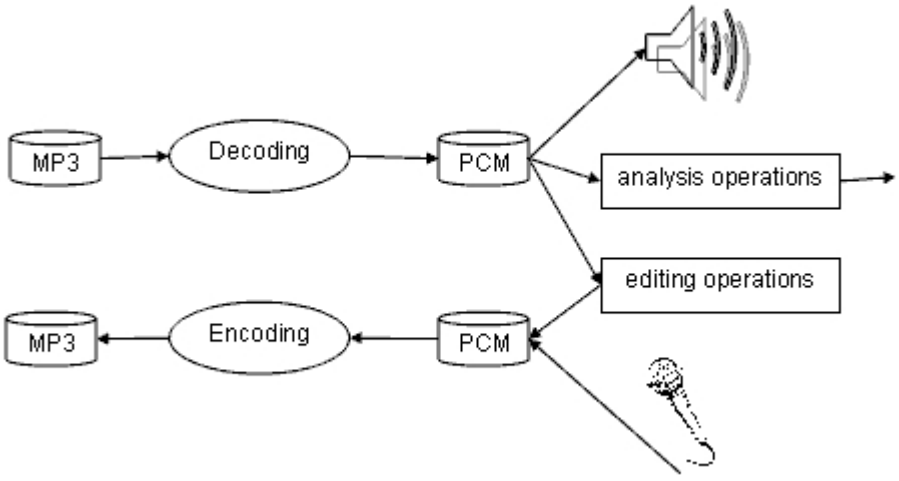
1. **Digital Item**: entità che rappresenta l'unità fondamentale per la distribuzione e la transazione; essa viene modellata attraverso il DID (Digital Item Declaration), un insieme di specifiche che permettono di descriverlo da un punto di vista astratto e concettuale.
2. ***Users****:* entità che interagisce con i *Digital Items*.

In sostanza, i **Digital Items** possono essere considerati come elementi del *Multimedia Framework* (collezione di video, album musicali, ecc.) mentre lo **User** (singoli individui, società, enti governativi, comunità, consorzi, ecc.) è colui che li utilizza. La manipolazione di questi **Digital Items** è governata da una serie di meccanismi, regole e relazioni che ne tutelino il loro trattamento in funzione del fatto che lo **User** abbia privilegi sufficienti per poter eseguire una determinata operazione.

6. Analisi e Manipolazione Diretta di Formati compressi MP3 ed AAC

Obiettivo di questa ricerca (in fase di svolgimento presso il LIM - Laboratorio di Informatica Musicale, Università degli Studi di Milano) è la progettazione di algoritmi per l'analisi e la manipolazione diretta di segnali audio compressi MPEG Layer 3 (MP3) ed Audio Advanced Coding (AAC). Trattare direttamente formati compressi MP3/AAC, significa non dover decomprimere i file in fase di apertura, e soprattutto ricodificarli durante il salvataggio.





Questo tipo di approccio elimina tre aspetti negativi

- Riduzione del *processing delay* introdotto dai tempi di encoding e decoding;
- Riduzione della quantità di memoria utilizzata, permettendo così di manipolare file MP3 di elevate dimensioni;
- Evitare la perdita d'informazione in fase di salvataggio (l'encoder MPEG/Audio utilizza un algoritmo di compressione con perdita d'informazione), mantenendo di fatto la qualità del segnale manipolato.

Queste tecniche possono essere applicate in sistemi DSP e real-time o integrate in sistemi per lo streaming, ambienti MIR, editor audio, sistemi autore ed applicazioni per la composizione musicale remota (esempi: radio web, AAP - Audio Adaptive Playout, indicizzazione di DB multimediali, plug-in per editor audio e sistemi autore, ecc.)

I prototipi software sviluppati presso il LIM (Laboratorio di Informatica Musicale) sono liberamente scaricabili dalla pagina: http://www.lim.dico.unimi.it/demo/MP3DirectEdit

Proposte di tesi di laurea e stage sugli argomenti di analisi ed editing diretto e valutazione della qualità audio percepita: http://www.lim.dico.unimi.it/didatt/dispo.html