TECNICHE DI CODIFICA MUSICALE

nota per il corso di *Elaborazione Numerica dei Segnali* a cura di

Ing. F. Benedetto – Prof. G. Giunta

Digital Signal Processing, Multimedia, and Optical Communications Lab.

Dip. Elettronica Applicata - Università degli Studi Roma TRE

1. Introduzione

Negli ultimi anni abbiamo assistito all'esplosivo sviluppo di tecnologie e tecniche che hanno reso la comunicazione a livello globale alla portata di chiunque. L'esempio più classico di tale enorme crescita tecnologica è costituito dalla rete Internet. Definire in poche parole Internet non è cosa facile: un buon termine riassuntivo dei suoi vari aspetti può essere quello di Rete delle Reti, intendendo con ciò un'infrastruttura che consente a reti distanti anche migliaia di Km e che adottano protocolli di comunicazione distinti, di entrare in contatto fra di loro e di scambiarsi informazioni. E' evidente che in una struttura così complessa è richiesto un ingente insieme di risorse per garantire una minima qualità di servizio e la disponibilità di banda diventa un nodo cruciale in quest'ottica. Se l'informazione viaggia lungo un doppino telefonico oppure attraverso una fibra ottica risulta chiaro che i tempi di transito siano assai diversi. Un modo per ovviare alla scarsità di banda è la compressione dei dati. Ovvero immagazzinare la medesima quantità di informazione nel minor numero di bit (o Byte) possibile. Un modem, per citare un esempio, prima di trasmettere i dati sul doppino telefonico, li elabora utilizzando diversi algoritmi di compressione, inclusi nel protocollo V90.

Tale introduzione è servita a dare un'idea dei vantaggi che la compressione dei dati offre, sia ai fini del risparmio di spazio per l'immagazzinamento, sia per il risparmio dei tempi di attraversamento nelle reti . È bene chiarire sin da subito che *non esiste* un algoritmo definitivo, unico e perfetto per la compressione di tutti i tipi di dati . Per fare un esempio, non si può comprimere un'immagine a colori abbastanza efficientemente se si usa un algoritmo pensato per comprimere file binari generici. Nel corso di questa trattazione, ci soffermeremo sulla compressione dei segnali audio, esaminando la basi della conversione analogico-digitale, le tecniche di compressione più semplici approfondendo infine le codifiche psico-acustiche che hanno portato alla realizzazione dell'ormai famosissimo MP3.

2. Il Segnale Audio

Il segnale audio è per sua natura un segnale analogico, ovvero un segnale che varia in modo continuo nel tempo. Se volessimo rappresentarlo graficamente, riportando su un diagramma la variazione dell'intensità sonora nel tempo, otterremmo un andamento di questo tipo:

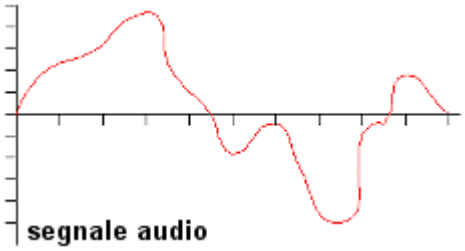

Fig. 1: Rappresentazione del segnale audio (analogico) nel tempo

Da notare che non ci sono salti bruschi. Inoltre l'ampiezza può assumere una infinità di valori passando da un valore minimo ad uno massimo. La larghezza temporale di questa forma d'onda è molto breve, tipicamente un millisecondo o due. Un segnale digitale ha invece una forma di questo tipo:

Fig. 2: Rappresentazione di un generico segnale digitale nel tempo

Riuscire a rappresentare un segnale audio, per sua natura analogico con un segnale digitale preservando l'informazione è un argomento che è alla base della teoria dell'informazione. La risposta sta nel campionare il segnale, ossia prelevare, ad intervalli regolari, il valore del segnale audio, che si presenta sotto forma di un segnale elettrico che varia nel tempo. L'idea è quella di approssimare la funzione analogica con una funzione fatta a rettangoli.

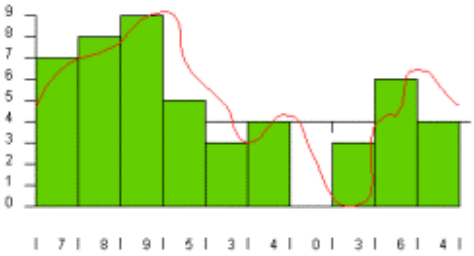

Fig. 3: Esempio di campionamento di un segnale analogico

In figura 3 abbiamo supposto di dividere la porzione del segnale in 10 parti, ognuna di durata T. Se la durata complessiva è di 1ms, il passo di campionamento sarà di 0.1 ms, ovvero la frequenza di campionamento è di 1/(0.1 ms) cioè 10KHz . L'errore che si commette con l'approssimazione è notevole. Nelle seguenti figure vediamo la *spezzata* che approssima la funzione (è stata costruita prendendo i valori delle altezze dei rettangoli) e il risultato di un campionamento a frequenza maggiore. Se aumentiamo la risoluzione, cioè il numero di rettangoli, l'errore di conversione diminuisce:

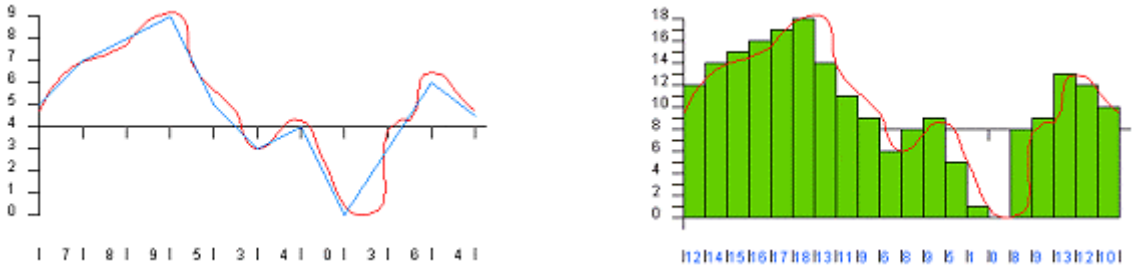

Fig. 4: Relazione tra passo di campionamento ed errore di conversione (risoluzione)

Ora, notiamo che l'altezza della funzione al variare di *t* varia in modo continuo. Nella figura in alto si è supposto di poter dividere l'altezza in 18 parti. Se il segnale originale cade nell'istante di campionamento fra i valori, diciamo, 12 e 13, si decide arbitrariamente di assegnargli il valore 12. Si è quindi quantizzata l'ampiezza che potrà assumere solo valori discreti, compresi fra 0 e 17. In informatica si lavora con le potenze del due, quindi supponiamo di quantizzare l'ampiezza con 16 livelli: dunque essa potrà assumere solo valori compresi fra 0 e 15. Il meccanismo di campionamento e la successiva quantizzazione dell'ampiezza costituiscono le basi della conversione di un segnale da analogico a digitale.

Passiamo ad un esempio concreto: le tracce audio digitali di un CD sono immagazzinate in file binari esattamente seguendo questo principio, solo che in ogni secondo vengono prelevati ben 44100 campioni e ogni campione è quantizzato con 16 bit. Considerando due canali, sinistro e destro, si avrà un bit-rate complessivo pari a: 44100 * 16 *2= 176400 bytes/secondo valore che corrisponde alla velocità di lettura "1x" dei normali lettori CD audio. Tale bit-rate, a partire dalla codifica finora discussa, porta alla qualità sonora nota come "qualità CD", che è lo standard di riferimento quando si valutano i risultati prodotti dagli algoritmi di compressione.

Per quale motivo si è scelto di utilizzare un rate o tasso di campionamento di 44100 campioni al secondo? Perchè è noto che l'orecchio umano arriva idealmente a

percepire frequenze entro i 20 KHz. Oltre questo limite l'uomo non ode nulla, così come sotto i 20 Hz. Dal noto teorema di Nyquist, si deve scegliere una frequenza di campionamento che sia almeno pari alla massima frequenza di interesse moltiplicata per due, se non si vogliono avvertire imperfezioni nell'ascolto. In breve, 20mila per due fa 40mila. Un valore leggermente superiore (44100) permette però di avere un pò di tolleranza al rumore nella costruzione dei filtri digitali preposti al trattamento del segnale. Questa discussione "qualitativa" è servita ad introdurre i concetti di base dell'audio digitale. Lo standard PCM, tipicamente usato per maneggiare un segnale audio in forma digitale non compressa, prevede proprio l'impiego di un rate di campionamento pari a 44100 campioni al secondo quantizzati a 16 bit per canale. Un formato audio non compresso molto noto agli utenti Windows è il WAV; WAV organizza i dati campionati in modo analogo al PCM. È facilmente calcolabile che un brano di 5 minuti trattato in questo modo produce un file di una cinquantina di Megabyte (10MByte/minuto). L' aumento progressivo di potenza dei calcolatori ha favorito, negli ultimi decenni, lo sviluppo e la diffusione dei sistemi di compressione dei dati audio che permettono, come vedremo, una forte riduzione delle dimensioni dei file.

3. La Compressione

Finora abbiamo visto come si campiona e si converte un segnale da analogico a digitale. In questa operazione viene inevitabilmente introdotto del *rumore*, legato al fatto che l'approssimazione della curva analogica originale tramite una spezzata non è perfetta. Comunque, d'ora in avanti per noi il segnale da considerare perfetto è quello di qualità CD, con passo di campionamento a 44100 Hz e quantizzazione dei campioni a 16 bit (queste specifiche in realtà lasciano ancora dei margini di miglioramento, e infatti in futuro forse assisteremo a brani musicali campionati a 96 KHz con quantizzazioni a 32 bit. Chiaramente una tale risoluzione genera una mole di dati enorme che solo l'impiego dei DVD può gestire). Nel campo dei segnali vocali (es: telefonate, audio-conferenza) le richieste in termini di frequenza di campionamento e bit per campione sono più limitate di quelle sopra elencate. Infatti la voce emette frequenze tipicamente più basse di 4KHz e con dinamica limitata rispetto a quella di uno strumento musicale, quindi un segnale vocale può essere ricostruito con qualità adeguata campionando a soli 8KHz con campioni da 13bit.

Storicamente, una delle prime tecniche di compressione è stata la quantizzazione non lineare. Questa tecnica è tutt'ora utilizzata nella telefonia fissa e consiste nell'applicare una scala logaritmica in fase di campionamento e ricostruzione del segnale. Basandosi su alcune proprietà della voce e dell'orecchio si privilegiano i segnali a bassa ampiezza piuttosto che quelli ad elevata ampiezza e si riesce ad ottenere la stessa qualità dei 13bit menzionati sopra con soli 8bit. Questa codifica a 8000 campioni al secondo e 8bit per campione prende il nome di LogPCM o PCM telefonico e richiede una banda "canonica" di 64Kbit/s (guarda caso la banda di una canale telefonico ISDN) valore tipicamente usato come misura della banda necessaria per un segnale vocale.

4. Classificazione degli algoritmi di compressione

L'obiettivo delle tecniche di compressione è ovviamente quello di ridurre lo spazio necessario ad immagazzinare determinati dati o la banda necessaria per trasmetterli. Una prima classificazione delle tecniche di compressione distingue tra tecniche che mantengono perfettamente inalterate le informazioni dopo la compressione (tecniche *lossless* cioè senza perdita) e tecniche che prevedono un certo degrado delle informazioni (*lossy*). E' abbastanza ovvio che nel comprimere informazioni come testi, documenti o programmi non ci si possa permettere la perdita di nessun bit di informazione, dovremo quindi utilizzare necessariamente tecniche lossless come quelle adottate dallo *Zip*. Nel caso dell'audio, delle immagini e dei filmati, un certo livello di degradazione è un compromesso accettabile per ridurre (e di molto) l'occupazione o la banda richiesta dal file. Le tecniche di compressione audio che analizzeremo sono infatti tutte di tipo lossy.

Le codifiche di compressione dell'audio sono numerose ed utilizzano tecniche anche molto differenti l'una dall'altra. Esistono però tre categorie principali: le codifiche nel dominio del *tempo*, le codifiche per *modelli* e le codifiche nel dominio delle *frequenze*. Generalmente le prime due categorie di algoritmi vengono usate per comprimere il segnale vocale mentre alla terza appartengono algoritmi come MP3, WMA, ATRAC-3 e AAC ottimi per la compressione della musica. Esulando le prime due codifiche dall’obiettivo primario di tale nota, nei successivi paragrafi verranno presentate solamente le caratteristiche fondamentali e le definizioni delle codifiche nel dominio del tempo e per modelli. Per una più approfondita trattazione riguardo questi temi

rimandiamo a [3]. In questa breve trattazione ci soffermeremo sulle codifiche nel dominio della frequenza, approfondendo le codifiche psico-acustiche che hanno portato alla realizzazione dell'ormai famosissimo standard MPEG Layer III (MP3). Concluderemo, infine, con una breve nota riguardo il protocollo MIDI.

4.1. Codifiche nel dominio del tempo

Si tratta di algoritmi che elaborano il segnale campionato direttamente, senza estrarre le informazioni spettrali (frequenze). L'obiettivo è quello di trovare correlazioni tra i campioni e/o proprietà dalla sorgente e della destinazione che permettano di ridurre il numero di bit usati per descrivere il valore di un campione audio. Sono storicamente le prime ad essere state elaborate, hanno bassa efficienza e sono state ampiamente superate dai nuovi algoritmi. Le più importanti sono il DPCM e l'ADPCM

4.1.1.DPCM e ADPCM

La Differential Code Pulse Modulation è in sostanza una PCM "truccata". Il punto è il seguente: poichè la voce umana emette suoni che non passano da volumi bassi a volumi alti o da frequenze basse a frequenze alte in un tempo inferiore ad alcune decine di millisecondi (corrispondenti quindi a centinaia di campioni registrati), si può pensare di trasmettere non tanto il campione attuale, ma la differenza fra il campione passato e quello attuale. Facciamo un esempio per meglio comprendere il tutto: se il campione N ha ampiezza pari a 100, il campione N+1 avrà ampiezza pari a 100, 101 o 99. Non potrà avere ampiezza pari a 200 perchè fra un campione e l'altro passano solo poche centinaia di microsecondi e la gola e le corde vocali hanno un tempo di rilassamento molto maggiore. Quindi, anzichè trasmettere il valore 101 del campione N+1, trasmetto solo il valore +1. Il decoder che riceve l'informazione vede quanto è il valore di N (es: 100), legge poi che il valore del campione N+1 è pari a quello di N cui va sommato 1, e quindi assume che il valore del campione N+1 sia 101. Utilizzando questo approccio sono necessari molti meno bit per trasferire l'informazione poichè si trasmette la sola differenza fra i campioni e non i valori effettivi a 16 bit .

Qual'è il problema? Questo sistema funziona bene con la voce umana, che gode di certe proprietà, ma si presta meno bene ad essere impiegato quando siano presenti anche degli strumenti musicali.

Una tecnica molto simile al DPCM è quella Adaptative Differential PCM, in cui cioè si trasmettono sempre i bit differenza, ma tenendo conto della "storia" dei bit passati. Il meccanismo è molto più complesso poichè si cerca di capire quali saranno i campioni futuri sulla base della storia di quelli passati. Il principio è comunque lo stesso, cioè trasmettere l'informazione collegata alle differenze fra i campioni, anzichè i valori effettivi con un numero di bit che dipende dalle caratterisiche del segnale nella porzione in esame (da cui il nome "adattivo"). Questa tecnica raggiunge un rapporto di compressione di 1:2 rispetto all'originale non compresso.

4.2. Codifiche per modelli

Le codifiche per modelli sono tecniche legate ad una particolare sorgente sonora (in questo caso la voce) che si tenta di emulare tramite un modello più o meno semplificato. Le corde vocali e la gola hanno delle ben precise caratteristiche fisiche, il loro comportamento sarà quindi predicibile sulla base di un modello. Queste codifiche rappresentano una scelta ottimale per la compressione della voce, tanto che vengono utilizzate nella telefonia mobile (GSM) e anche su Internet. Le più famose sono LPC e il CELP.

4.2.1.LPC e CELP

La Linear Predictive Coding è una tecnica utilizzata esclusivamente ai fini della compressione spinta della voce (vocoder). Il vantaggio è un bit-rate bassissimo: si può arrivare a soli 2.4 Kbit al secondo (fattore di compressione: 1:26). Si è sviluppato negli anni successivi il Code Excited Linear Predictor (CELP) che fa uso della LPC ma migliora la qualità poichè trasmette anche l'informazione sull'errore associato alla codifica LPC. La qualità è buona e il bit-rate minimo raggiungibile risulta comunque molto basso: 4.8 Kbit al secondo (ratio 1:13). Questo tipo di codifica è molto usata in vari ambiti: Nei telefonini GSM sotto forma di EFR (Enhanced FullRate) da 12.2 Kbps e grazie alla sua flessibilità viene anche utilizzata spesso nella trasmissione della voce via Internet. (Audio-Video-conferenza e Voice over IP).

4.3. Codifiche nel Dominio della Frequenza

Questi algoritmi sono accomunati dal fatto di esaminare e processare il segnale non nel dominio del tempo, ma nel dominio della frequenza. Ogni strumento musicale, ogni suono e anche la voce ha una propria impronta spettrale caratteristica costituita da una combinazione di frequenze contenute in uno spettro più o meno ampio. Lavorando su tale spettro è possibile comprimere il segnale in misura molto maggiore di quanto non si riesca a fare nel dominio del tempo.

Ora, si osservi la figura seguente:

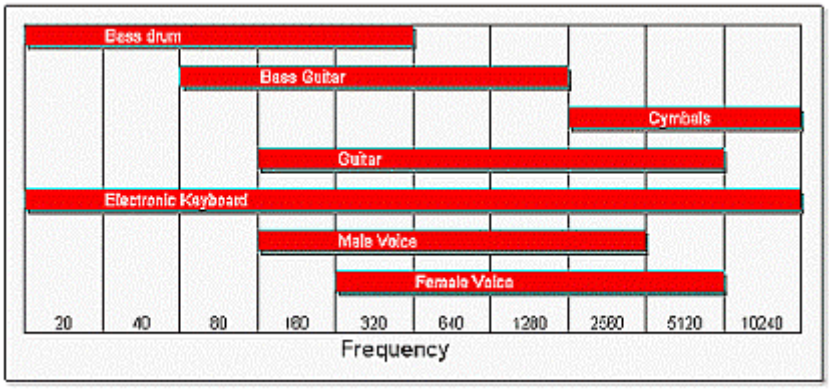

Fig. 5: Range di frequenze della voce umana e di vari strumenti musicali

Si capisce subito che la voce umana occupa solo certe frequenze mentre gli strumenti musicali spaziano secondo range diversi a seconda del tipo di strumento impiegato. La tastiera elettronica, poi, può occupare virtualmente qualsiasi frequenza. In realtà l'intervallo di frequenze occupato da uno strumento dice poco. Per esempio, come fa il nostro orecchio a distinguere fra un "do" emesso da un pianoforte e un "do" emesso da un violino , visto che la frequenza del "do" è sempre quella? Qui entra in gioco il *timbro*. Nessuno strumento emette una singola frequenza. Quando uno strumento emette, per esempio, un "la", corrispondente a 440Hz, emette in realtà molte altre frequenze multiple della fondamentale, note come armoniche (880, 1320, e 1760Hz ad esempio). È proprio la diversa distribuzione di queste frequenze, nonché la loro differente intensità, che distingue il "la" prodotto da un violino da quello prodotto da un piano. Anche se la frequenza centrale, cioè l'armonica fondamentale, sia per il violino che per il piano è sempre 440 Hz.

Oltre al timbro, un altro aspetto del suono prodotto da uno strumento o dalla nostra voce è il *pitch*. Supponiamo di avere un "la" minore e un "la" maggiore. Entrambi sono

dei "la", nel senso che la distribuzione delle varie armoniche è sempre la stessa, ma cambia la frequenza centrale di riferimento, ora leggermente più bassa, ora leggermente più alta. Si dice dunque che è cambiato il pitch, cioè la frequenza "centrale".

Detto questo, appare chiaro che tentare di comprimere una musica generica utilizzando un approccio basato su modelli (come per il CELP) risulterebbe estremamente complesso. Limitarsi alla sola voce è "facile" ma prevedere un modello per ogni strumento è, allo stato attuale della tecnologia, un impresa titanica. La soluzione è porre l'attenzione non sulla sorgente ma sulla "*destinazione*" dei suoni. In ogni caso la musica dovrà essere "ascoltata da un orecchio" quindi conoscere fino in fondo quello che un uomo medio riesce a sentire o non riesce a sentire può rivelarsi sorprendentemente utile.

5. Approccio psico-acustico alla Compressione

5.1. La sensibilità dell'orecchio umano

Il nostro orecchio *per fortuna* non è perfetto e, come vedremo, questo è un grande vantaggio. In prima analisi esso è sensibile in misura diversa alle diverse frequenze, come è possibile dedurre esaminando il grafico in basso.

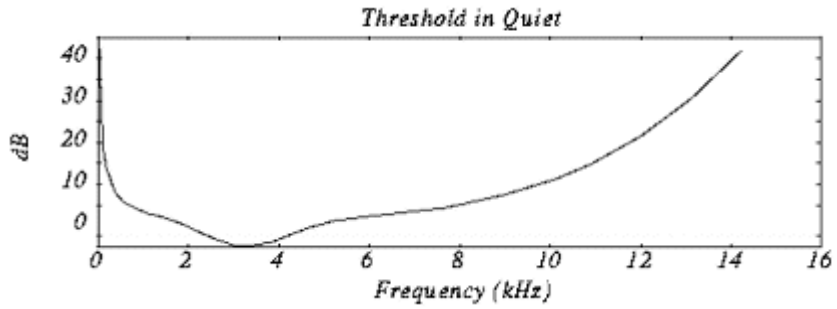

Fig. 6: Soglia di udibilità dell'orecchio umano

Dal grafico emerge che l'orecchio umano è maggiormente sensibile alle frequenze comprese fra 2 e 4 KHz. Non è un caso che l'intervallo fra i 2 e i 4 KHz sia quello massimamente usato dalla nostra voce. Ovviamente possiamo già usare questa caratteristica dell'orecchio a nostro vantaggio eliminando dallo spettro del segnale in analisi quelle componenti spettrali non udibili dall'orecchio medio. In sostanza si tagliano le alte e le bassissime frequenze. In generale, siccome l'orecchio a queste frequenze perde sensibilità e selettività, si può ridurre la quantità di informazione

trasmessa in questa parte di spettro. Questo diagramma è stato tracciato facendo variare una sola armonica, cioè un singolo tono. Ma che succede se di toni ve ne sono due? Il nostro orecchio è in grado di distinguerli sempre oppure in alcuni casi uno dei due viene mascherato dall'altro?

5.2. Il Mascheramento Audio

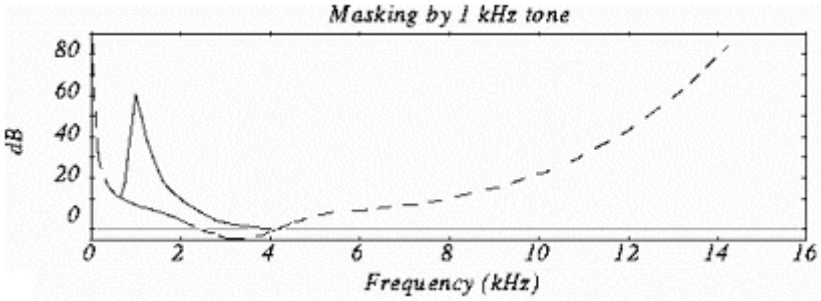

Fig. 7: Esempio di mascheramento audio di due segnali

Stavolta viene diffuso un tono alla frequenza di 1 kHz, detto tono maschera, tenuto fisso a 60 dB (si legga volume alto). Ripetiamo il discorso di prima sulla soglia di udibilità di un secondo tono, detto tono di test. Quello che emerge è che avvicinandoci sia da sinistra che da destra al tono maschera, dobbiamo alzare il volume del tono test per riuscire a distinguerlo. Oltre i 4 kHz e al di sotto degli 0.5 kHz le cose tornano a posto, però notiamo che nell'intorno di 1 kHz i due toni sono praticamente indistinguibili a meno di non alzare pesantemente il volume del tono test. In sostanza, una frequenza *debole* può essere benissimo mascherata, cioè risultare inudibile, da una frequenza anche lontana qualche centinaio di Hz, se quest'ultima è *forte*, cioè con un'intensità alta. Se abbiamo più toni tenuti fissi a volumi alti, per esempio a 60 dB e a frequenze fisse di 0.25, 1, 4 e 8 kHz, notiamo che la risoluzione del nostro orecchio peggiora sempre più, perchè per avvisare il segnale del tono test a ben 1 kHz di lontananza dal segnale tenuto fisso a 4 kHz, il segnale del tono test deve raggiungere i 40dB. Se non vi fosse stato il tono maschera a 4 kHz sarebbero bastati un paio di dB.

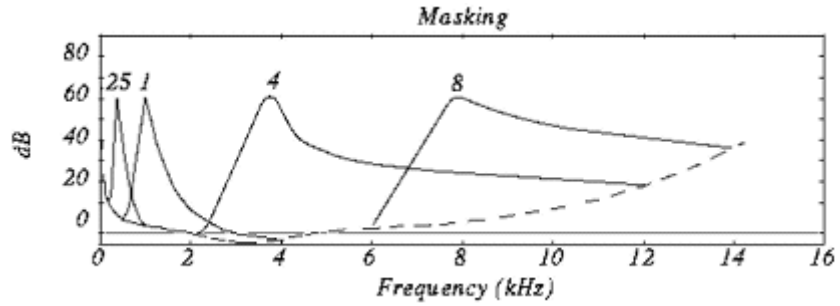

Fig. 8: La risoluzione dell'orecchio peggiora all'aumentare di toni alti

Tramite il fenomeno appena descritto è possibile eliminare componenti spettrali che, essendo troppo vicine a suoni forti, non risultano udibili all'orecchio umano.

Finora abbiamo parlato del mascheramento in *frequenza*. Esiste però un altro tipo di mascheramento, ed è quello *temporale*. Supponiamo di avere al solito due toni, uno forte e l'altro, che gli è vicino in frequenza, piuttosto debole. Dall'analisi vista prima sappiamo già che il nostro orecchio sente solo il tono più forte che si comporta così da tono maschera. Ora, se improvvisamente questo tono maschera cessa di esistere, avvertiamo subito il tono più debole o impieghiamo un pò di tempo per avvertirlo? Chiaramente la seconda, perchè la membrana del nostro timpano deve assestarsi. Il problema è: quanto tempo impieghiamo? Dipende dal volume del tono maschera e da quello del tono test. Se il tono maschera ci assorda, il nostro orecchio impiegherà un pò prima di riuscire a sentire il tono più debole anche dopo che il tono forte è morto.

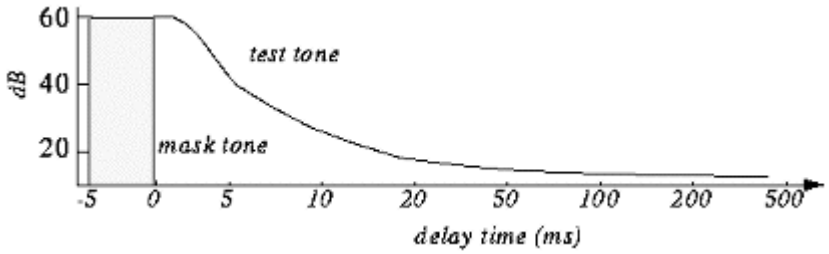

Fig. 9: Esempio di mascheramento temporale

Il tono test, di 1 kHz, viene disattivato all'istante zero: esso manteneva un valore fisso di 60dB. Se il nostro test tone ha un'ampiezza di una quarantina di dB, bastano 5 ms per avvertirlo; se è di soli 20 dB, ne occorrono quasi 20 di millisecondi perchè risulti avvertibile. Cosa succede se abbiamo più toni a diverse frequenze e alcuni di questi muoiono? in altri termini, come si fondono le maschere temporali e quelle in frequenza? La figura sottostante è molto chiarificatrice:

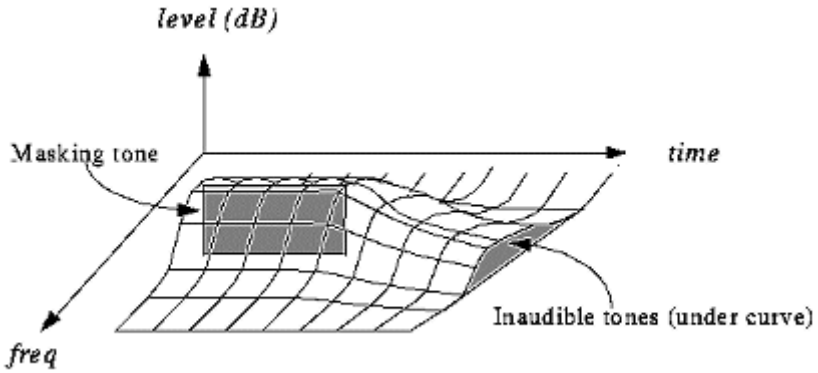

Fig. 10: Rappresentazione 3-D del mascheramento audio (in frequenza e nel tempo)

In conclusione l'effetto complessivo del mascheramento è che molti toni non saranno mai udibili perchè collocati nel dominio della frequenza e del tempo troppo vicino a toni forti. Tenendo conto della sensibilità dell'orecchio e del fenomeno del Masking Audio è quindi possibile eliminare dallo spettro del segnale una quantità molto alta di informazioni inutili, perchè non udibili dall'orecchio umano. Questi sono i fenomeni Psico-Acustici su cui si basano i moderni algoritmi di compressione audio come MP3, MP3Pro, Atrac-3, AAC, etc...

6. Panoramica sullo standard MPEG

MPEG/Audio è uno standard internazionale riconosciuto dall'ISO, che è l'organizzazione internazionale preposta all'approvazione definitiva di uno standard. È stato varato ufficialmente nel 1992 quando lo stato dell'arte dell'home computing era il 386. Esso è solo una branca degli standard di cui si occupa il comitato MPEG, come si vede dalla figura qui in basso:

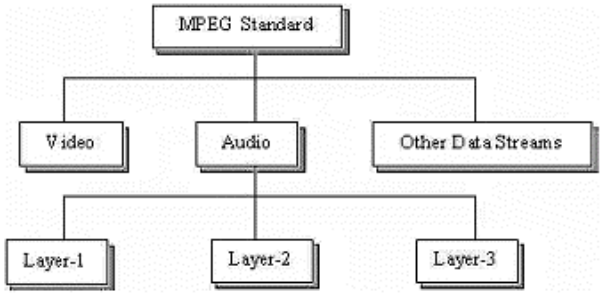

Fig. 11: Compiti del comitato MPEG e standard studiati

L'MPEG (Moving Picture Experts Group) è un gruppo di lavoro gestito dalla ISO/IEC che si occupa dello sviluppo di standard per la codifica audio-video digitale. Il primo progetto MPEG iniziò nel 1988 per terminare nel 1992 con l'uscita dello standard internazionale MPEG-1 (ISO/IEC 11172), principalmente utilizzato nei prodotti basati su Video-CD e 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.

Il layer-2 e il layer-3 sono versioni migliorate del livello 1. In linea di massima, offrono una migliore compressione e qualità audio appoggiandosi a codificatori e a

modelli psico-acustici che richiedono maggiori risorse di elaborazione. Oggi le risorse di calcolo disponibili sono sprecate e per effettuare un conversione di un file WAV in un file MP3 impostando la massima qualità occorrono al massimo un paio di minuti. MP3 sta per MPEG layer 3, ed era fino a poco tempo fa il sistema di codifica che realizzava il miglior rapporto qualità/dimensione dei file audio. Una versione recente e migliorata è l'MP3pro. I file codificati con il layer 2 e il layer 1 sono leggibili dai lettori per il layer 3 ma non viceversa. Vale cioè la retrocompatibilità.

6.1. Come funziona l'MPEG

L'algoritmo di codifica è composto di diversi steps che possono essere così riassunti: si usano dei filtri per dividere il segnale audio che è campionato con una certa frequenza, ad esempio di 44100 campioni al secondo, in 32 sottobande , per ognuna delle quali sono noti i parametri di mascheramento nel tempo e in frequenza (si fa riferimento al modello psicoacustico introdotto prima). Per ognuna delle sottobande, viene calcolata l'entità del mascheramento causata dalle bande adiacenti. Se la potenza in una sottobanda è sotto la soglia di mascheramento, allora non viene codificata in uscita l'informazione che essa trasporta, poiché sarebbe inudibile. Altrimenti, occorre calcolare il numero di bit necessari per rappresentare l'informazione della sottobanda facendo attenzione che in questo procedimento, per sua natura approssimante e dunque rumoroso, il rumore introdotto stia sotto la soglia. Infine, si forma il flusso di bit (bitstream) in uscita.

Il diagramma a blocchi della codifica MP3 può essere così schematizzato:

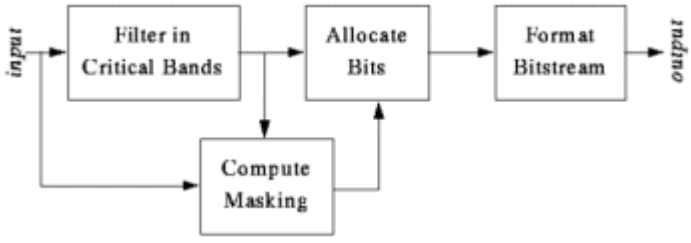

Fig. 12: Diagramma a blocchi della codifica MP3

Per capire il funzionamento facciamo un esempio: dall'esame del bitstream di ingresso, che viene suddiviso in 32 sottobande, abbiamo calcolato il livello massimo del segnale in ognuna di queste e abbiamo ottenuto una tabella di questo tipo, ove per semplicità si considerano solo 16 dei 32 intervalli:

Banda	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Livello(dB)	0	8	12	10	6	2	10	60	35	20	15	2	3	5	3	1

Fig. 13: Esempio di divisione del segnale in sotto-bande

Esaminiamo la banda numero 8. Dai dati del nostro modello psicoacustico, sappiamo che l'ottava banda, se presenta una intensità di 60dB, genera una maschera di 12dB nella settima e di 15dB nella nona. La settima ha un livello pari a 10 (<12dB), ed è perciò mascherata e tagliata via dall'uscita. La nona è a 35dB (>15) così passa in uscita . Con quanti bit si quantizza? Non si può far passare 16 bit per ogni campione. La teoria dell'informazione ci dice che per ogni bit di quantizzazione utilizzato aumento di 6dB il rapporto segnale/rumore. Siccome non devo scendere sotto il limite di mascheramento (punto 4 dell'algoritmo) che è di 15 dB, posso usare al minimo 3bit per quantizzare i dati. Se usassi meno di tre bit, otterrei un rumore di quantizzazione troppo alto che verrebbe avvertito compromettendo la qualità finale. Ora, in realtà questo era il modello di base, valido per tutti e tre i layers dell'MPEG, l'MP3 introduce invece alcune migliorie, che si pagano in termini di risorse di sistema impiegate.

6.2. L'MP3 in specifico

L'MP3 utilizza sempre il blocco dei filtri, però a differenza dei layers 1 e 2 le sottobande non sono tutte della stessa dimensione, poiche certe frequenze contengono molta più informazione e vanno trattate con maggiore dettaglio. Il layer 3 inoltre fa uso di una MDCT, cioè di una trasformata discreta coseno modificata. In breve si tratta di effettuare un'operazione che consenta di migliorare la risoluzione in frequenza per ognuna delle sottobande. Questa operazione consente di suddividere ognuna delle 32 sottobande in ulteriori 6 (short) o 18 (long) sottofrequenze, secondo un processo noto come filtraggio sottobanda (sub-band filtering).

Il modello psico-acustico lavora ulteriormente su queste sotto-sottomaschere, in particolare sui coefficienti della MDCT che le rappresentano. Il modello psico-acustico deciderà quali coefficienti devono passare in uscita e quali no, sulla base del calcolo del mascheramento temporale e sul fatto che alcuni di questi sono ridondanti giacché magari provengono dai canali sinistro e destro che spesso portano la medesima informazione. A questo punto il tutto è quasi pronto. I coefficienti "sopravvissuti" contengono le informazioni necessarie alle varie frequenze e devono ora essere

organizzati in uscita. I coefficienti vengono ordinati passando dalla frequenza più bassa a quella più alta. Poiché la massima informazione è contenuta in bassa frequenza, i coefficienti di bassa frequenza sono più numerosi di quelli in alta frequenza. L'intero intervallo viene diviso in tre parti (frequenze basse, medie e alte). Ognuno di questi intervalli viene codificato a parte secondo l'algoritmo di Huffmann, che è uno degli algoritmi basilari nella teoria della compressione. L'algoritmo è ottimizzato per ognuno dei tre intervalli. A questo punto i dati vengono inviati in uscita sotto forma di pacchetti che contengono un CRC (codice per la correzione dell'errore) per rendere il sistema più robusto agli eventuali errori che si possono presentare durante il trattamento del file. Il fattore di compressione che tipicamente si ottiene è quello di 1:11 (128Kbit/s), per cui è possibile immagazzinare un minuto di musica in poco meno di un megabyte.

6.3. Struttura generale di un sistema audio MPEG-1 ed MPEG-2

Il sistema di codifica MPEG è costituito da tre entità fondamentali:

- Formato di codifica: insieme di regole definite dagli standard MPEG (ad esempio: l'ISO-IEC 11172-2 per l'MPEG-1 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 (ad esempio WAV) 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.

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

Lo schema a blocchi di un generico Encoder Audio Mpeg-1 ed MPEG-2 è riportato nella figura seguente:

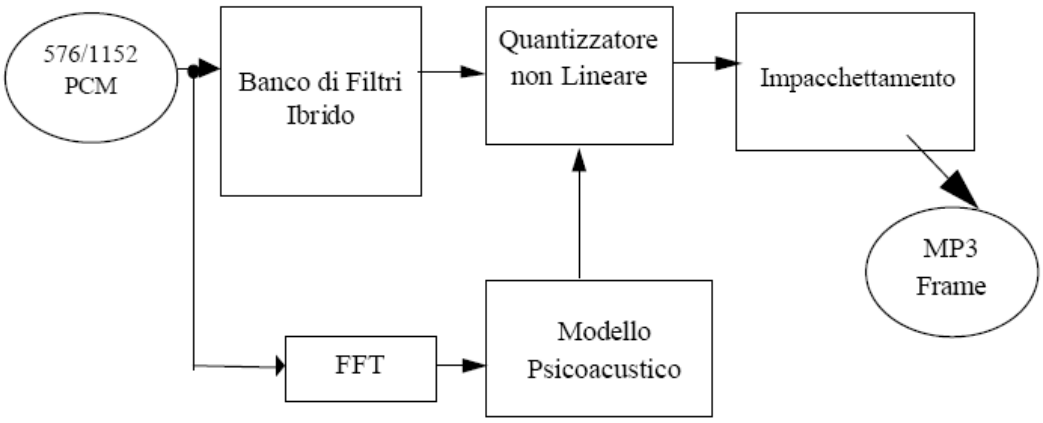

Fig. 14: Schema a blocchi dell'encoder MPEG Audio

Esso riceve un segnale PCM (segnale audio tempo-discreto con codifica di modulazione ad impulsi, Pulse Code Modulation) e lo legge a blocchi di 384, 576 o 1152 campioni, in funzione del formato MPEG e del Layer utilizzati, effettuando le seguenti operazioni nei diversi passaggi:

1. Banco di Filtri Ibrido: vengono convertiti i campioni PCM nel corrispondente dominio della frequenza (spettro) utilizzando un banco di filtri polifasico seguito da una Trasformata Coseno Modificata (MDCT).
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 Fourier) e definire il livello di soglia di udibilità SMR (Signal to Mask Ratio) sfruttando i principi psicoacustici dell'apparato uditivo umano. In pratica, il modello psicoacustico determina quali sono le informazioni che il nostro orecchio è in grado di percepire e quali no fornendo questa informazione al blocco successivo.
3. Quantizzatore non lineare: compito di questo blocco è di codificare numericamente lo spettro ricevuto dal banco di filtri ibrido in funzione dell'importanza di ogni banda di frequenze: se il modello psicoacustico indica che una particolare banda di frequenze è percepita poco, essa verrà codificata con pochi bit; viceversa se per il modello una banda di frequenze è percepita molto, la sua codifica avverrà con molti bit. L'obiettivo finale è quello di ottenere una quantizzazione dello spettro tale che il rumore di quantizzazione

introdotto si trovi al di sotto della soglia di udibilità (SMR) fornita dal modello psicoacustico.

4. Impacchettamento: si prende la codifica numerica dello spettro frequenziale e la si impacchetta secondo la sintassi dello standard MPEG utilizzato. In questa fase il layer 3 prevede un ulteriore compressione con l'algoritmo di Huffman.

Il Decoder Audio ha invece la struttura seguente:

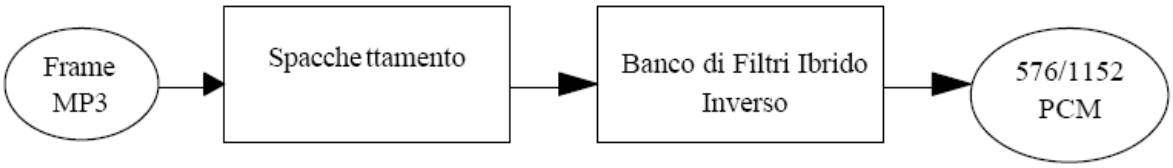

Fig. 15: Schema a blocchi del decoder MPEG Audio

Lo streaming MPEG/Audio in input viene diviso in frame per ognuno dei quali si effettua le seguenti operazioni:

1. Spacchettamento: da ogni frame si leggono tutte le informazioni codificate secondo la sintassi MPEG e si estrae lo spettro (per la codifica MP3 si ha anche la decodifica di Huffman).
2. Banco di filtri inverso: viene preso lo spettro del segnale e si generano i corrispondenti campioni PCM (384, 576 o 1152) che vengono forniti al DAC (Digital to Analog Convert) della scheda audio o da scrivere su file, ad esempio WAV.

6.4. 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 quanti riguarda MPEG-1 ed MPEG-2, gli algoritmi esistenti sono 3, identificati dal Layer di appartenenza.

-Layer 1: è l'algoritmo più semplice e dà buoni risultati buoni risultati con un bitrate pari a 384 Kbit/sec per un segnale stereo. Associa 384 campioni PCM per frame. Il formato di file associato è l'MP1

-Layer 2: più complesso in quanto associa ad un frame 1152 campioni. E' adatto a codifiche a bitrate intorno ai 192-256 Kbit/sec per un segnale stereo. Formato associato: MP2.

-Layer 3: è il più complesso dei tre e quello con migliori prestazioni. Il formato MPEG-1 associa ad ogni frame 152 campioni PCM mentre l'MPEG-2 ne associa solo 576, aumentando la risoluzione temporale. Già con bitrate tra 128 e 192 kbit/sec si riesce ad ottenere un segnale stereo di qualità sufficiente,mente elevata. Il formato di file associato è il famigerato MP3, che ha come concetti base: il dominio frequenziale suddiviso in funzione delle bande critiche, l'utilizzo della codifica di Huffman per l'impacchettamento finale dei dati audio e l'introduzione della tecnica del "Bit Reservoir" che permette di migliorare la qualità audio a parità di bitrate.

I Layer sono stati concepiti per essere compatibili con le precedenti versioni.

6.5. MPEG Layer 3

La sigla MP3 fa dunque 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 Layer 3. E' importante notare che il "formato della 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 sul computer (e dunque su un file).

Il formato di codifica MPEG-1 è indicato principalmente per la compressione musicale in quanto supporta solo le alte frequenze di campionamento (quelle possibili sono 32, 44.1 –qualità CD- e 48 kHz) mentre i valori di bitrate assegnati al layer 3 vanno da 32 a 320 kbit/sec. La codifica dei canali permette 4 alternative:

- single channel (ossia mono);
- dual channel (due canali mono distinti, ad es. con lingua differente);
- stereo (due canali indipendenti);
- joint stereo (codifica stereo compressa che utilizza due diversi algoritmi per eliminare le ridondanze presenti nei due canali, MS Stereo ed Intensità Stereo).

Si possono poi avere diversi tipi di bitrate:

- *Bitrate fisso*. Tutti i frame presenti nel file hanno lo stesso valore di bitrate, si ha così la possibilità 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 del silenzio avrà

bisogno di pochi bit e dunque di un valore di bitrate basso, viceversa per l'attacco di una nota. Con bitrate variabile si ha in genere un'elevata qualità audio ed un buon tasso di compressione ma non è possibile conoscere a priori la dimensione del file che si va a creare.

- *Bitrate Free Format*. Il valore di bitrate può essere diverso da quelli standard a patto che il bitrate resti fisso e il suo valore non superi il valore massimo previsto dal Layer. E' scarsamente supportato dagli Mp3 player presenti sul mercato.
- *Average Bitrate*. Tecnica nuova e supportata dai soli encoder di ultima generazione sfruttatata 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 di input. Per il Layer 3 è obbligatorio supportare il bitrate variabile, funzione facoltativa per i primi 2.

Il formato di codifica MPEG-2 è l'evoluzione dell'MPEG-1. Concettualmente non apporta grosse novità, sono stati migliorati ed ottimizzati gli algoritmi di compressione (Layer) e sono state aggiunte nuove frequenze di campionamento più basse (per arrivare sino a 16 kHz). Sono inoltre presenti bitrate più bassi (fino a 32 kbit/sec). I tipi di bitrate sono gli stessi supportati dall'MPEG-1 mentre l'MPEG-2 è in grado di supportare anche le nuove codifiche multicanale a 3, 4, 5 e 5.1 canali.

E' interessante notare come l'MPEG-2 sia compatibile con lo standard MPEG-2 e viene detto per questo "backward compatible".

Esiste poi un ulteriore formato, l'MPEG 2.5, non ancora riconosciuto dalla ISO/IEC. La sua unica differenza con l'MPEG-2 è che può supportare frequenze di campionamento bassissime (8, 11.025, 12 kHz).

6.6. 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 di conoscere tutte quelle informazioni di fondamentale importanza per l'identificazione e la catalogazione di brani audio (titolo, genere, autore...).

La risposta è stata la nascita dello standard ID3 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 bit-stream 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 loro: è possibile ometterle, oppure utilizzarne solamente una, o ancora usarle entrmbe. La strututra 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)

6.7. Prestazioni di un codificatore MP3

Nella seguente figura 16, sono riportate le prestazioni di un codificatore MP3 a differenti bitrate, in modalità mono e/o stereo, indicando il rapporto di compressione che si riesce ad ottenere in uscita.

Sound quality	Bandwidth	Mode	Bitrate	Reduction ratio
better than short-wave	4.5 kHz	mono	16 kbit/s	48:1
better than AM radio	7.5 kHz	mono	32 kbit/s	24:1
similar to FM radio	11 kHz	stereo	56..64 kbit/s	26..24:1
near-CD	15 kHz	stereo	96 kbit/s	16:1
CD	>15 kHz	stereo	112..128 kbit/s	14..12:1

Fig. 16: Prestazioni di un codificatore MP3 a diversi bitrate.

I dati non sono "assoluti", la riduzione varia in funzione del suono da comprimere e del codificatore utilizzato. I dati riportati sono stati recuperati dal Fraunhofer Institut e si riferiscono al loro codificatore.

In figura 17, è invece riportato il confronto tra la codifica MPEG e varie altre codifiche audio-musicali.

Applications	Format	Sampling rate	Audio bitrate	Overhead bitrate	Total Bitrate
Compact Disc (CD)	PCM	44.1 kHz	1.41 Mbit/s	2.91 Mbit/s	4.32 Mbit/s
Digital Audio Tape (DAT)	PCM	44.1 kHz	1.41 Mbit /s	1.67 Mbit/s	3.08 Mbit/s
Philips Digital Compact Cassette (DCC)	MPEG-1 Layer I	48 kHz	384 kbit/s	384 kbit/s	768 kbit/s
Sony Mini Disc (MD)	ATRAC	44.1 kHz	292 kbit/s	718 kbit/s	1.01 kbit/s
European Digital Audio Broadcast *	MPEG-1 Layer II	48 kHz	256 kbit/s	256 kbit/s	512 kbit/s

Fig. 17: Codificatori a confronto.

Sono evidenti le migliori performance raggiunte dai codificatori MPEG.

[*] *Il sistema European DAB usa MPEG-1 Layer II*

6.8. MPEG 2

Con il passare del tempo è divenuto chiaro che due canali non sono sufficienti a coprire tutte le applicazioni e per permettere la codifica multi-canale è stato sviluppato il secondo standard. L'impulso principale è stato dato dalla necessità di codificare l'audio che rispondesse alle richieste del settore cinematografico-televisivo, con almeno 5 canali: destro, sinistro, centrale, destro e sinistro surround. Il progetto MPEG-2 si articola in due fasi successive, nella prima si è privilegiata la compatibilità con gli standard precedenti. Nella seconda questa è stata sacrificata, si era compreso che abbandonare la compatibilità era l'unica via per ottenere un notevole incremento nelle prestazioni.

I codificatori (vedi figura 18) non presentano novità concettuali rispetto ai precedenti. I codificatori di seconda generazione aggiungono altre frequenze di campionamento: a 16 kHz, a 22.05 e a 24 kHz. Per quanto riguarda il bitrate, questo può variare da un minimo di 8 kbit/s fino a 160 kbit/s per canale.

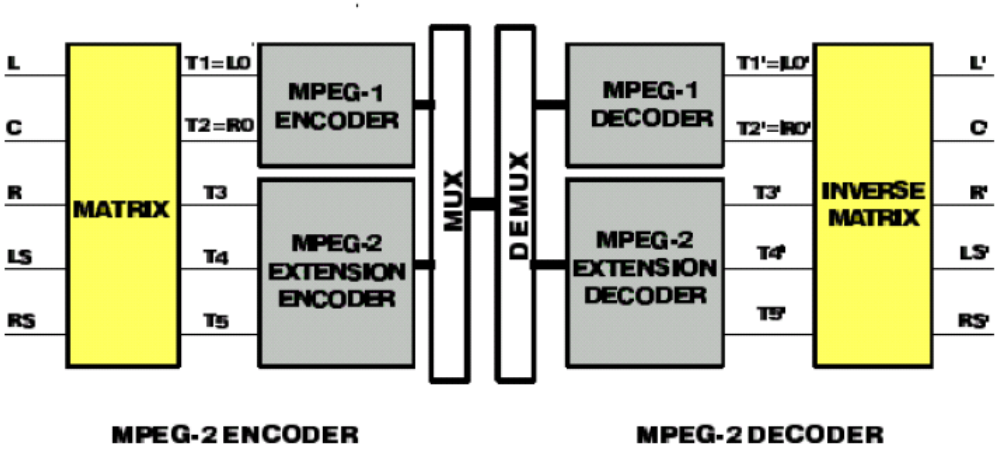

Fig. 18: MPEG-2 coder e decoder.

La seconda fase di MPEG-2, denominata AAC (Advanced Audio Coding), parte con l'obiettivo di sviluppare un tool per migliorare le prestazioni nella codifica multicanale.

Il risultato ottenuto è notevole, si possono codificare fino a 48 canali con una frequenza di campionamento che va da 8 a 96 kHz per canale. Per comprendere il livello raggiunto si tenga presente che alcuni test dimostrano come, a parità di qualità e nel caso di codifica a 5 canali, l'AAC in pratica dimezza la bitrate rispetto a MPEG-2 Layer II.

Com'è possibile notare dalla fig. 19, l'AAC segue gli stessi principi del layer III, oltre a migliorare il funzionamento d'alcuni punti, introduce una serie di blocchi che migliorano la compressione e riducono il bitrate.

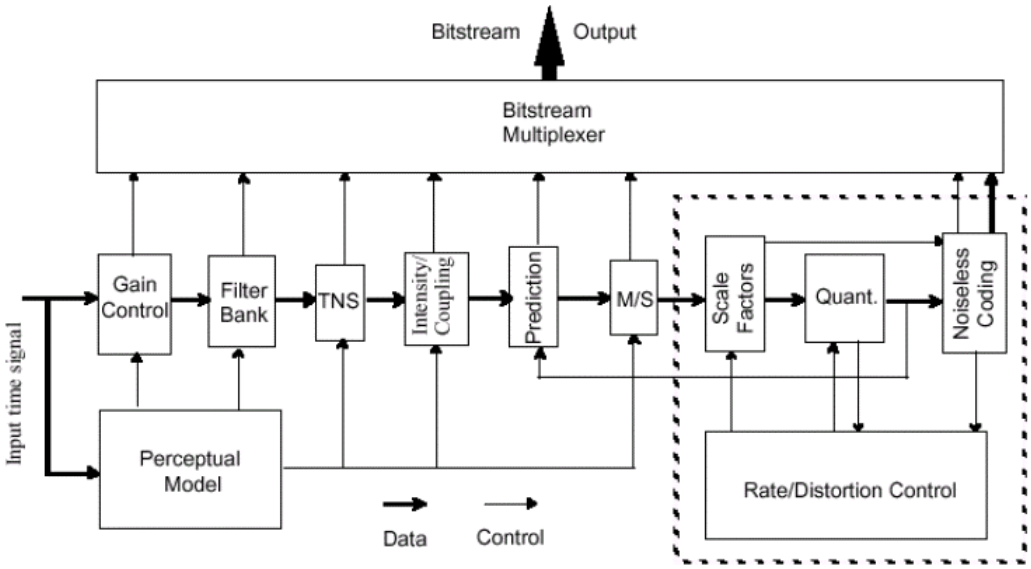

Fig. 19: Schema a blocchi dell'encoder AAC.

Fra le parti migliorate ci sono:

- **Risoluzione di frequenza**: il numero di linee di frequenza in uscita dal Filter Bank passa da 576 a 1024 (potenza di 2 e più facile da gestire), la dimensione delle finestre quindi diventa di 2048 campioni;
- **Codifica stereo congiunta, M/S**: anziché codificare i due canali separatamente si codifica la loro media e la loro differenza (left-right), si ricordi che codificare piccoli valori richiede pochi bit;
- **Codifica di Huffman**: si utilizzano un maggior numero di tabelle, che oltre a migliorare la codifica facilitano il lavoro del quantizzatore.

6.9. MPEG 4

Seguendo l'evoluzione informatica, MPEG-4 introduce il concetto di "oggetto" nel settore audiovisivo. Ogni file multimediale è una combinazione di diversi oggetti che, pur potendo esistere separatamente sono armonizzati per ottenere l'effetto complessivo.

Con specifico riferimento alla figura 20:

- Il parlato può essere realizzato in lingue diverse e codificato tramite un tool dedicato che offre buona qualità ad un bitrate modesto.
- Per l'annuncio non è indispensabile una qualità eccelsa, è codificato come parlato sintetico ottenendo l'effetto desiderato ad un bitrate molto basso (fino a 200 bit/s).
- Il rumore del treno è codificato con 8 canali, in modo da ottenere un perfetto effetto stereo.
- La musica di sottofondo è codificata con l'AAC.

Fig. 20. Esempio di diverse codifiche in MPEG 4.

Infine lo standard MPEG-4 è dotato din un'interessante proprietà, la Synthetic-Natural Hybrid Coding (SNHC): questa tecnica permette la composizione d'audio naturale (compresso) e suono sintetico creato in real-time sul terminale ricevente.

7. Estensione .MID

Nato come standard per la comunicazione tra strumenti musicali, si è in seguito evoluto in una maggiore standardizzazione denominata General MIDI. In un file Wave può essere memorizzato, ad esempio, il suono di un pianoforte che esegue una nota per una certa durata di tempo. In un file MIDI, viceversa, sono contenute istruzioni che comunicano alla scheda audio di modulare la frequenza in modo da produrre una particolare nota che abbia una timbrica simile a quella di un pianoforte e che duri una

certa quantità di tempo. In un file Wave ci sono suoni, in un file MIDI ci sono solo comandi. I file MIDI sono enormemente più piccoli dei file WAV. Un intero brano musicale, con tutte le parti, della durata di svariati minuti, può occupare qualche decina di Kbyte, dal momento che ogni singolo evento MIDI occupa soltanto 11 byte.

Nonostante siano passati quasi vent'anni, il protocollo midi rimane indiscutibilmente il linguaggio informatico che ha più cambiato il modo di eseguire e produrre musica.

7.3. La Storia

Il midi, acronimo di *musical instruments digital interface*, nasce principalmente da due fattori: primo, la ricerca di compatibilità tra strumenti diversi e di diverse marche; secondo, la crescente disponibilità di tecnologia digitale. Fino agli anni '80 il collegamento tra sintetizzatori, batterie elettroniche e altri strumenti musicali elettronici era affidato al *gate* o *trigger* (un impulso di tensione fornito da una macchina all'altra, che generalmente determinava la scansione delle note o il sincronismo) e al *VC* (*Voltage Control*, che determinava la frequenza delle note). Ogni costruttore ottemperava a proprie specifiche, nella definizione delle caratteristiche elettriche, che quasi mai corrispondevano a quelle definite da altri costruttori.

Il problema era particolarmente avvertito negli Stati Uniti, tanto che già nel 1981, fu presentato un progetto di Universal Synthesizer interface. Nel corso dei mesi successivi vennero perfezionate le caratteristiche tecniche di quello che doveva diventare il MIDI, e da parte della Sequential Circus, da sempre principale promotrice dello sviluppo del sistema (ironia della sorte ora fallita), nel dicembre del 1982 fu posto in commercio il Prophet 600, primo synth al mondo dotato del nuovo modo di dialogare in musica. Tra il 1983 e il 1984 i sintetizzatori digitali Midi ebbero una prima diffusione di massa con la commercializzazione di prodotti come lo Yamaha DX7 e il Korg Po1y800. Sorsero due comitati: l'americano MMA (Midi Manufacturers Association) e il giapponese JMSC (Japanese Midi Standards Commitee) che diedero il via a periodici incontri tesi a perfezionare il Midi e a farne rispettare le direttive. Dal marzo 1984 tutte le aziende che dotano i propri prodotti di porte Midi sono obbligate a rispettare alcune caratteristiche base, che garantiscono la compatibilità fra gli strumenti delle diverse marche. La IMA (International Midi Association) nel settembre 1985 pubblicò le specifiche dettagliate dello standard Midi 1.0, debitamente riviste e corrette. Con il rapido diffondersi di nuove apparecchiature Midi e di applicazioni

inizialmente non previste, il codice Midi è stato più volte rivisto e soprattutto espanso. Le appendici più significative sono state pubblicate nell'autunno 1986 e nella primavera 1987, ma ancora adesso si sta lavorando alla definizione di nuovi messaggi destinati a consentire lo scambio universale di file di dati MIDI.

Così come lo vediamo oggi, il Midi quindi è il risultato di due distinte proposte, una made in USA, e l'altra giapponese: superate le traversie operative del primo anno di vita, dal marzo 1984 esiste un accordo ulteriore, in base al quale tutte le macchine Midi prodotte ottempereranno alle medesime specifiche, almeno quelle di base.

7.2. Il protocollo MIDI

Abbiamo visto come il MIDI è stato istituito come protocollo, quindi come standard riconosciuto in tutto il mondo da tutti i costruttori, non solo di strumenti musicali, ma anche d'apparecchiature informatiche e audio/video. Il protocollo MIDI quindi stabilisce le specifiche sia hardware (interfaccia, cavi, connettori), sia software (linguaggio informatico, modalità di trasmissione, tipologia dei messaggi) che ogni apparecchiatura deve "capire" se vuole essere veramente MIDI compatibile. Allo stato attuale delle cose tutte le apparecchiature presenti sul mercato sono tranquillamente compatibili tra loro almeno nelle funzioni principali.

Lo scopo del MIDI è di trasmettere comandi: quando sulla nostra tastiera premiamo un tasto, dalla presa MIDI out uscirà un comando digitale che indicherà ad un'apparecchiatura in ricezione che è stato premuto quel determinato tasto (numero di nota) con una determinata forza (velocity). Quindi una cosa fondamentale da capire è che il MIDI non trasmette nessun tipo di suono ma unicamente comandi che verranno rieseguiti dall'apparecchiatura in ricezione. Ogni movimento fatto dall'esecutore su di una tastiera MIDI verrà codificato in un modo univoco secondo il protocollo: abbassare o rilasciare un tasto e con che forza, il muovere una rotella o un altro controllo, il cambiare timbro ecc… Questo sistema permette di comandare con una sola tastiera più generatori sonori collegati o la possibilità di memorizzazione attraverso computer o sequencer dedicato dei dati di esecuzione di una partitura suonata su una tastiera MIDI. Il primo caso è tipico di una situazione concertistica, dove il musicista suona dal vivo uno o più generatori sonori (expander) collegati ad una tastiera di comando (master keyboard) mentre il secondo è normalmente più proprio di una situazione di studio e compositiva dove, con l'ausilio di computer o

sequencer, è possibile registrare, correggere, modificare, sovrapporre, riascoltare situazioni musicali precedentemente eseguite. Lo scopo del codice Midi è quello di trasformare in numeri ogni azione compiuta da un musicista nell'eseguire un brano, tanto da poter eventualmente permetterne una riesecuzione elettronica automatica.

Ciò permette di memorizzare e manipolare digitalmente un'esecuzione musicale, svincolandola dal timbro di un particolare strumento, dalle possibilità tecniche dell'esecutore, e dalle imperfezioni accidentali tipiche dell'uomo. Midi dunque è un sistema di comunicazione dati, che consente ad uno strumento musicale o altro dispositivo (detto master) di controllarne un altro (detto slave), e anche più di uno, in modo da suonare insieme le stesse note, cambiare i timbri nello stesso momento, iniziare (da capo o dal segno) i brani memorizzati, mantenere la sincronizzazione. In alcuni casi (che vedremo più avanti), è anche possibile, tramite la tastiera che funge da 'master', controllare altri parametri, come modulation wheel, pitch bender, e anche i valori di inviluppo, filtri e volume.

7.3. Il MIDI e il suo Hardware

Il componente principale è l'UART (universal asynchronous receiver transmit) un microprocessore appositamente costruito alla comprensione e decodificazione dei dati MIDI che lavora in maniera asincrona cioè solo quando qualcosa appare alle porte MIDI. Le prese MIDI sono tre: un ingresso (Midi In, dati in ricezione) , una uscita (Midi Out, dati in uscita dalla macchina) e un terzo connettore, denominato Midi Through, che si limita a ripetere fedelmente in uscita tutti i dati che appaiono all'ingresso Midi In.

Per il sistema MIDI, anche se la trasmissone parallela è molto più veloce, è stato scelto il tipo di trasmissione seriale, per semplificare i collegamenti e aumentarne l'affidabilità. Ciascun byte dunque può assumere diversi significati, a seconda della sequenza logica di volta in volta proposta: il protocollo di trasmissione si assume il compito di determinare univocamente il significato di una sequenza di byte.

Il vantaggio della trasmissione seriale è che basta un solo filo per trasmettere l'informazione, per cui il collegamento risulta economico, pratico e affidabile. Perchè la trasmissione seriale risulti efficiente, la velocità della trasmissione deve essere abbastanza alta. Per questo motivo i bit del codice Midi vengono trasmessi alla velocità di 31.25 Kbaud, cioè 31.250 bit al secondo. La trasmissione è asincrona e ciò

significa che l'inizio e la fine di ogni byte devono essere "annunciati" ogni volta da due bit speciali, lo start bit e lo stop bit, posti rispettivamente davanti e dietro gli otto bit del byte Midi da trasmettere. Dunque è necessario trasmettere 10 bit per inviare un byte Midi, quindi tale byte richiederà 320 msec prima di essere ricevuto, ovvero l'impercettibile ritardo che il Midi impone per sua natura tra l'azione del musicista e l'effettiva esecuzione sonora.

8. Considerazioni conclusive

Abbiamo visto che esistono tecniche ottime per la compressione della voce (il CELP, ad esempio, che è ottimo per il trasferimento del segnale vocale su Internet) e della musica (MP3). Entrambe sono tecniche lossy, e fanno uso del medesimo principio per il quale l'apparato uditivo umano è incapace di rivelare certi suoni per via del mascheramento temporale e frequenziale. Il CELP è inutilizzabile per codificare la musica, poiché produrrebbe risultati scadentissimi, mentre l'MP3 è eccessivamente articolato per essere utilizzato efficacemente nella codifica in tempo reale della voce. Un nuovo standard, che incorpora sia la codifica della voce che quella della musica, è l'MPEG4, varato nel 1998. Per avere una idea della complessità che sta dietro questa tecnologia, basta menzionare il fatto che per comprimere il parlato abbina le tecniche CELP e HXVC (harmonic vector excitation coding), mentre per codificare la musica usa il TWIN-VQ e l'AAC. Interessante anche l'evoluzione diretta dell'MP3, l'MP3Pro che permette un fattore di compressione quasi doppio a parità di qualità sonora. Infine, è stato brevemente discusso il protocollo per lo scambio di dati seriali tra apparecchiature MIDI.

9. Bibliografia

[1]. P. Noll: “Wideband Speech Audio Coding”, *IEEE Audio Coding Communication Magazine*, Vol. 31. No 11, Nov 1993.

[2]. D. Pan: “A tutorial on MPEG/Audio compression”, *IEEE Trans. on Multimedia*, vol. 2, no. 2, 1995, pp. 60-74.

[3]. G. Giunta: “Codifica Audio”, Università degli Studi “ROMA TRE”, Facoltà di Ingegneria, dispense per il corso di Elaborazione Numerica dei Segnali, (scaricabile gratuitamente dal sito www.comlab.uniroma3.it/ens.html).

[4]. G. Vercellesi: “MPEG audio tutorial”, Università degli Studi di Milano, Dipartimento di Informatica e Comunicazione.

[5]. F. Visciotti: “Tecniche di Compressione Audio: Evoluzione dello Standard MPEG”, Università degli Studi di Bologna, Facoltà di Ingegneria.

[6]. Zwicker, E., e H. Fastl, Psychoacoustics: Facts and Models, 2nd ed., Heidelberg: Springer-Verlag, 1999.

[7]. ISO/IEC Joint Technical Committee 1 Subcommittee 29 Working Group 11, Information Technology—Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s Part 3: Audio, ISO/IEC 11172-3, 1993.

[8]. ITU-R Recommendation BS.1387, “Method for Objective Measurement of Perceived Audio Quality”, Dec.1998.

[9]. P. Kabal, “An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation if Audio Quality”, Department of Electrical & Computer Engineering, McGill University, Dec. 2003.

[10]. P. Supurovic, “MPEG Audio Frame Header”, Dec. 1999.