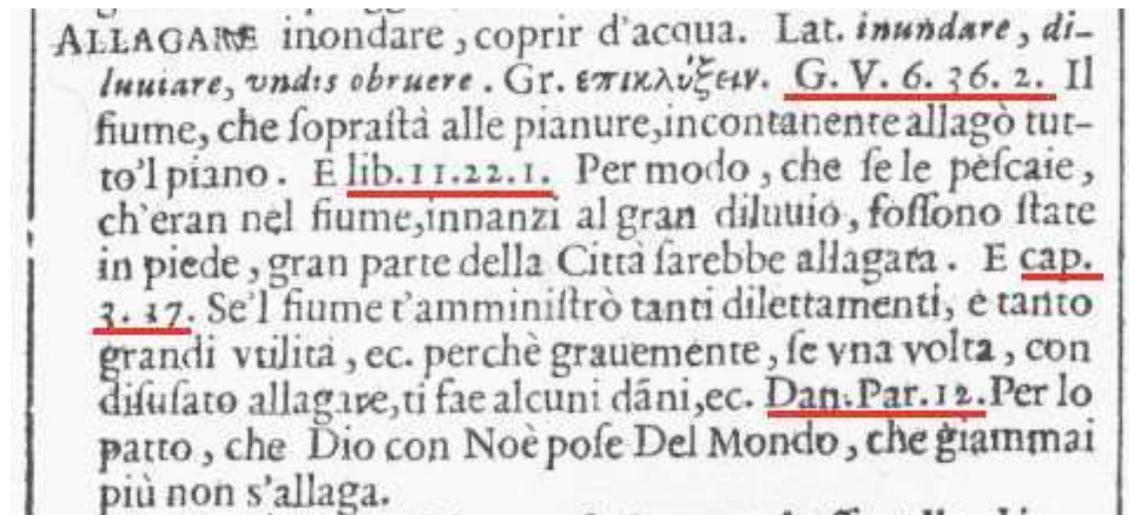


13

La compressione audio

Che cos'è la compressione software

I primi esempi di compressione dei dati risalgono agli albori della scrittura, quando uno scalpellino si trovò a combattere con una lastra di pietra troppo piccola e un testo troppo lungo e fu costretto ad abbreviare le parole. Un esempio più recente, ma sempre stagionato, è rappresentato dal **Vocabolario degli Accademici della Crusca** che nella edizione del **1612** presenta un'intera serie di abbreviazioni bibliografiche tutte diverse tra loro usate con lo scopo di far tornare le lunghezze delle righe (pazientemente composte a mano).



ALLAGARE inondare, coprir d'acqua. Lat. *inundare*, *diluviare*, *undis obruere*. Gr. *επικλυζειν*. G. V. 6. 36. 2. Il fiume, che sopraffà alle pianure, incontanente allagò tutto'l piano. E lib. 11. 22. 1. Per modo, che se le pescaie, ch'eran nel fiume, innanzi al gran diluuiò, fossero state in piede, gran parte della Città farebbe allagata. E cap. 3. 17. Se'l fiume t'amministrò tanti diletamenti, e tanto grandi vtilità, ec. perchè grauemente, se vna volta, con difusato allagare, ti fae alcuni dāni, ec. Dan. Par. 12. Per lo patto, che Dio con Noè pose Del Mondo, che giammai più non s'allaga.

È con l'avvento della memorizzazione elettronica dei dati che il problema del supporto piccolo e dei dati ingombranti si propone in tutta la sua drammaticità e la vendita di compressori diviene un ottimo affare.

A seconda della utilizzazione si distinguono tra tecniche di memorizzazione compressa (i ben noti *zip* e *rar*) e tecniche di trasmissione compressa (ovvero le tecniche usate nei **Modem** per aumentare l'efficienza di trasmissione a parità di costo). Ma la vera distinzione è da fare tra compressione **lossless** (priva di perdite) e compressione **lossy** (con perdita di informazione).

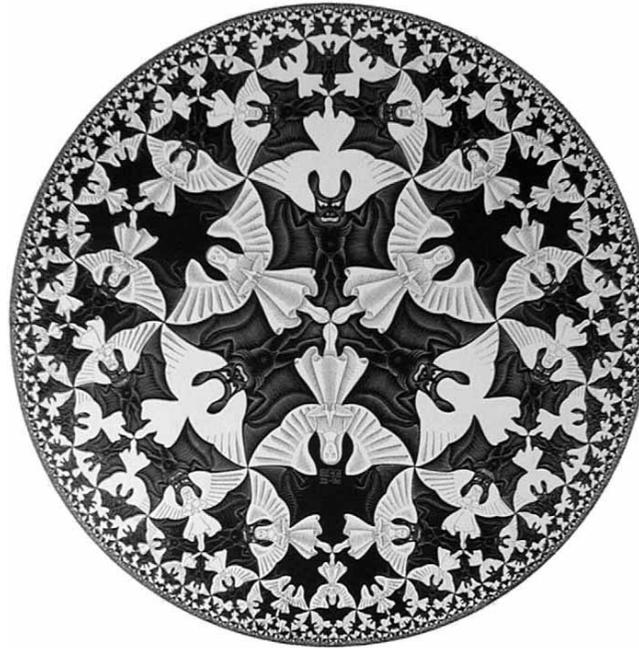
- Il primo tipo di compressione, che deve permettere la ricostruzione senza errori dei dati originali, si basa soprattutto sulla eliminazione delle ridondanze e per questo presenta limiti ben precisi. L'uso di questo tipo di compressione è indispensabile per memorizzare o trasmettere programmi, testi e ogni tipo di informazione che non può essere alterata senza danni. Vi è sempre un limite alla compressione **lossless** perché

Le stringhe corte sono poche, quelle lunghe sono tante!!!

- Nel secondo caso si accetta di perdere informazione avendo in cambio il vantaggio di non avere limiti al tasso di compressione. Questa codifica si presta bene ad essere usata per dati come suoni e immagini che per la loro natura sono comunque soggetti ad una inevitabile perdita di informazione (causata dal rumore ambientale, la risoluzione dei trasduttori, ecc.)

Per comprendere bene la compressione **lossless**, è necessaria una digressione sulla ridondanza.

Informazione e Ridondanza



Conticuere omnes intentique ora tenebant.
inde toro pater Aeneas sic orsus ab alto:
infandum, regina, iubes renovare dolorem

(Virgilio)

La **ridondanza** si può definire informalmente come un eccesso di informazione: qualcosa che si può togliere senza perdere i contenuti fondamentali che si intendono comunicare. È importante sottolineare che la ridondanza **NON** è mera **ripetizione** anche se spesso esistono trasformazioni dei dati che possono evidenziare le **ripetizioni**. In genere la presenza di una **struttura** implica sempre una **ridondanza**. Vediamo qualche esempio.

Ridondanza nelle lingue

Ogni linguaggio naturale ha una forte dose di ridondanza, è questa che permette di comprendere un testo corrotto o un parlato in presenza di rumore ambientale.

È interessante come sia facile riconoscere la lingua che si sta ascoltando, spesso anche senza comprendere ciò che viene detto (come il brano di **Virgilio** nella pagina precedente).

Quelle che seguono sono frasi **senza significato** generate al computer seguendo la frequenza delle triple tipica di alcune lingue. Il lettore è invitato ad indovinare la lingua da cui è tratta la statistica delle **triple**. Gli esempi sono tratti dal testo, ormai classico, **Abramson, Information Theory and Coding, McGraw-Hill, 1963**.

- Ianks can ou ang rler thatted of to shor of to havem a i mand and but whissitably thervereer eights takillis ta
- Jou mouplas de monnernaissains dem us vreh bre tu de toucheur dimere ll es mar balme re a ver douvents so
- Bet ereiner sommeit sinach gan turhatt er aum wie best alliebder taussichelle laufurcht er bleindeseit uber konn
- rame de lla el guia imo sus condias su e uncondadado dea mare to buerballi a nuae y hararsin de se sus suparoceda
- Et ligercum siteci libemus acererlen te vicaescerum pe non sun minus uterne ut in ario popomin se inqueneque ira

Ed ecco una approssimazione dell'**inglese** di ordine **5** (ovvero seguendo la statistica delle quintuple)

The head and in frontal attack on an english writer that the character of this point is therefore another method for the letters that the time of who ever told the problem for an unexpected

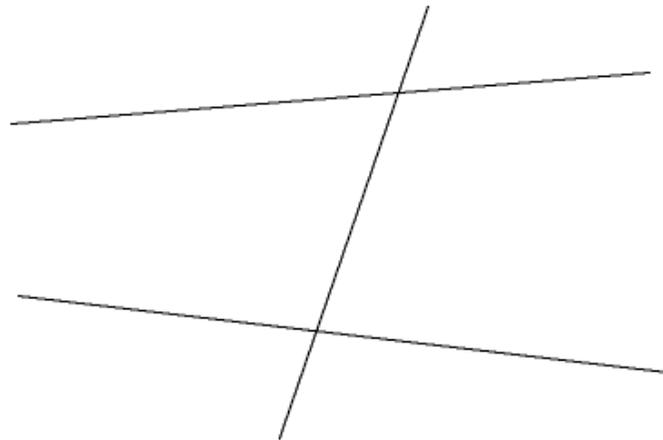
Si noti che tutte le parole appartengono al lessico inglese anche se questa caratteristica è dimostrabile solo per quelle di lunghezza minore di **5**.

Esempi di testo con diversi gradi di ridondanza “semantica”

Euclide, Postulato delle parallele

Καὶ ἐὰν εἰς δύο εὐθεῖας εὐθεῖα ἐμπίπτουσα τὰς ἐντὸς καὶ ἐπὶ τὰ αὐτὰ μέρη γωνίας δύο ὀρθῶν ἐλάσσονας ποιῆ, ἐκβαλλομένης τὰς δύο εὐθεῖας ἐπ’ ἄπειρον συμπίπτειν, ἐφ’ ἃ μέρη εἰσὶν αἱ τῶν δύο ὀρθῶν ἐλάσσονες.

Se una retta, intersecando due rette, forma gli angoli interni sullo stesso lato minori di due angoli retti allora le due rette, se prolungate indefinitamente, si incontrano dalla parte in cui gli angoli sono minori di due angoli retti.



ovvero

Per un punto ad una retta passa una e una sola parallela.

Alcmane, Notturmo

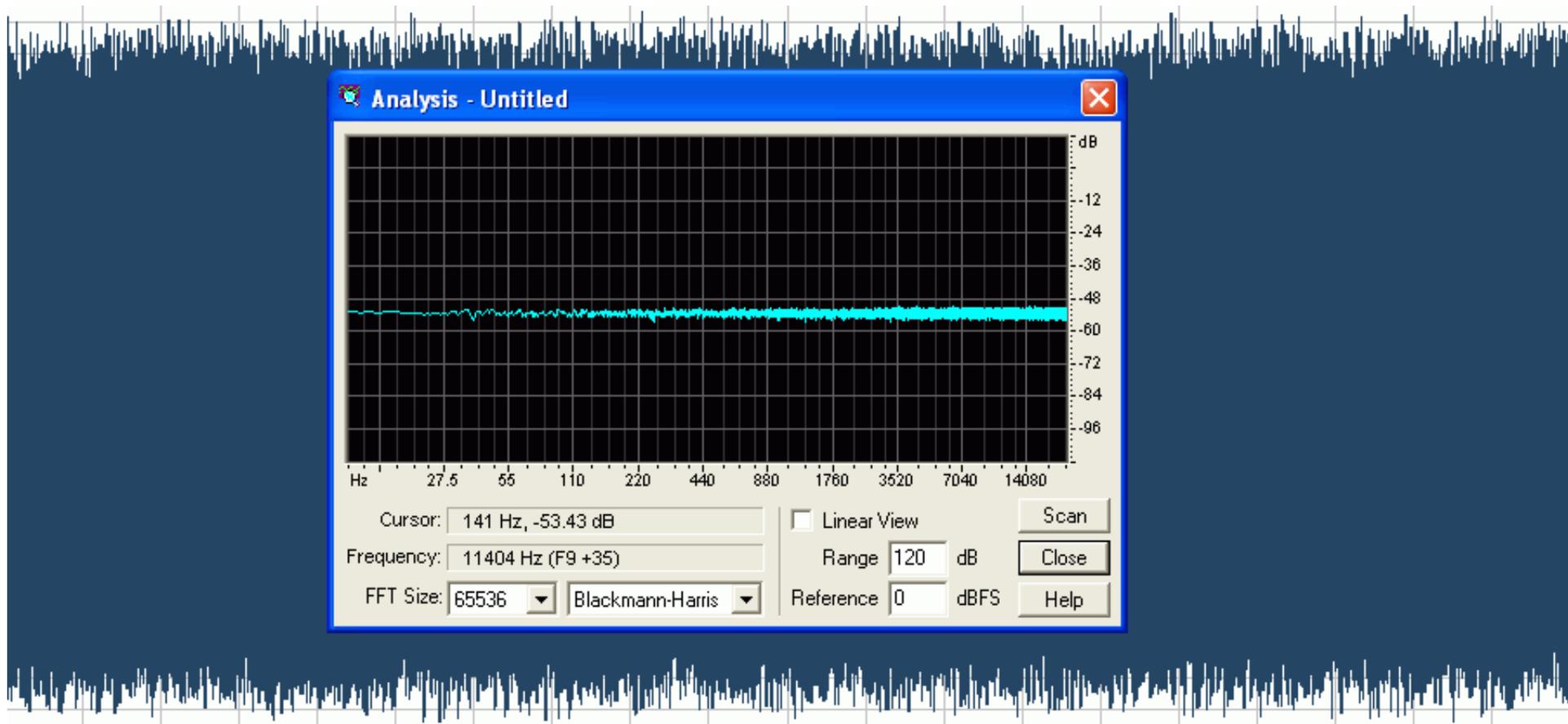
εὔδουσι δ' ὀρέων κορυφαί τε καὶ φάραγγες,
πρώονές τε καὶ χαράδραι,
ὔλα θ' ἔρπετά θ' ὅσσα τρέφει μέλαινα γαῖα,
θῆρές τ' ὄρεσκῶι καὶ γένος μελισσῶν
καὶ κνώδαλ' ἐν βένθεσσι πορφυρέας ἁλός
εὔδουσι δ' οἴνων φῦλα τανυπτερύγων.



Dormono de' monti le vette
le balze e le forre;
e le valli e le selve
e quanti gli esseri che la nera terra nutre:
le fiere montane, la specie delle api
e i mostri negli abissi del purpureo mare.
Dormono gli uccelli dalle ampie ali.

Ridondanza nell'Audio

Un suono assolutamente non ridondante è il **rumore bianco**. Il vero rumore bianco ha solo una valenza teorica, porta un'**informazione infinita** e non esiste in natura. Un segnale digitalizzato in cui ogni campione è indipendente da tutti gli altri costituisce una approssimazione realizzabile al rumore bianco e porta la **massima** informazione possibile **in quel** formato.



Da un punto di vista artistico il rumore bianco fa schifo.

Un esempio di Audio “Semanticamente Ridondante”: Bach BWV 108, 4 - coro

A rigorous piece of choral polyphony, three tersely arranged fugues in motet style (J.E. Gardiner)
Le fughe sono a 4 voci (Basso, Tenore, Alto, Soprano) ogni versetto viene ripetuto 8 volte il tutto in 2' 35"

Wenn aber jener, der Geist der Wahrheit, kommen wird, der wird euch in alle Wahrheit leiten.
Denn er wird nicht von ihm selber reden, sondern was er hören wird, das wird er reden;
und was zukünftig ist, wird er verkündigen.

Ma quando Lui, lo Spirito di verità, verrà, vi guiderà alla verità tutta intera.
Egli non parlerà da sé stesso, ma tutto quello che ascolterà, ve lo dirà;
e vi annuncerà le cose future (**Giovanni 16,13**)

Entropia informatica

La **Teoria dell'Informazione** di **Shannon** fornisce un quadro sistematico in cui i concetti di ridondanza e compressione hanno una collocazione rigorosa.

L'**incertezza** nel risultato di un esperimento con uscite x_i , ciascuna con probabilità p_i vale

$$H = - \sum_i p_i \log_2 p_i$$

si misura in **bit** e viene detta **entropia dell'esperimento**. La formula si può derivare da alcuni semplici e ragionevoli assiomi sul contenuto informativo di un **esperimento stocastico**. N.B. la formula è la stessa della espressione statistica dell'**Entropia fisica**.

- **Teorema della codifica in assenza di rumore**

Si può dimostrare che data una sorgente di informazione S la sua entropia $H(S)$ è il limite inferiore al **numero di bit necessari** per trasmettere i dati di S senza perdite. Dopo una compressione ottima nel senso di **Shannon** i bit sono **indipendenti ed equiprobabili** ovvero è stata eliminata ogni ridondanza.

Questo teorema insieme all'algoritmo di Huffman fornisce un modo ottimale per comprimere al massimo i dati che sono generati da una sorgente **di cui si conoscono le proprietà statistiche**. Purtroppo in informatica tipicamente si devono comprimere testi o brani musicali le cui proprietà statistiche sono ignote e si devono percorrere quindi altre strade.

Compressione audio lossless

L'idea di base di tutti i compressori **lossless** è quella di fare un primo trattamento del segnale per tenere conto delle correlazioni più semplici.

Dapprima si lavora sui canali cercando di sfruttare le loro somiglianze, un modo brutale è quello di codificare la somma e la differenza di due canali stereo ma vi sono tecniche molto più sofisticate.

In secondo luogo si applicano al segnale dei **filtri predittori** che permettono di trattare facilmente la parte principale dell'informazione. Gli errori di predizione non vengono ignorati ma messi da parte.

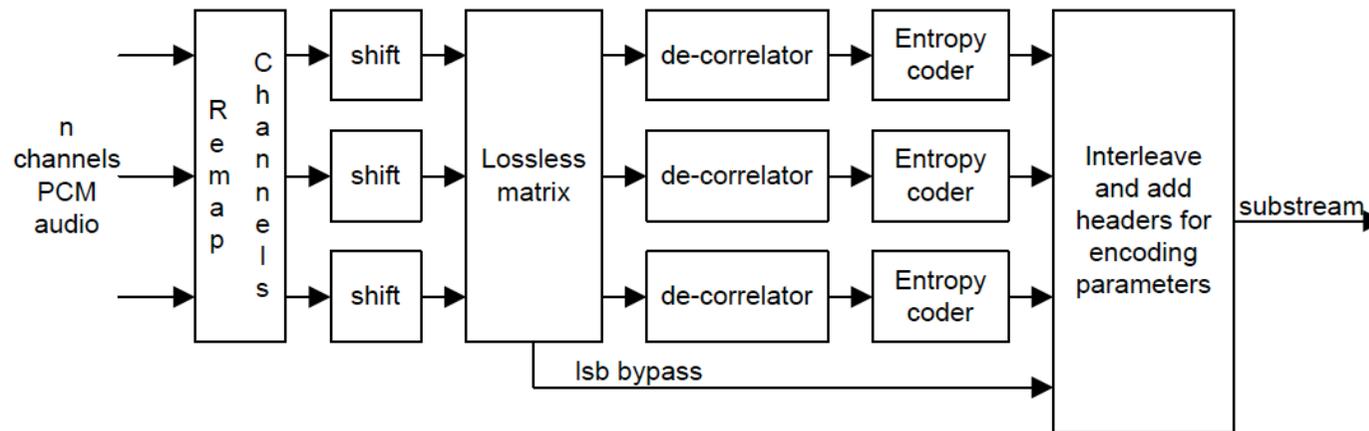
In pratica questo gruppo di trattamenti riduce la ridondanza trasformando il segnale musicale in un insieme di dati parzialmente scorrelati che possono venire ulteriormente compressi con tecniche standard.

In fase di decompressione si effettuano le operazioni inverse e si ottiene di nuovo il file originale. In alternativa si può lavorare in tempo reale suonando il file compresso invece che trasformandolo e anche in questo caso non vi è perdita di qualità purché vi sia sufficiente potenza di calcolo per eseguire correttamente l'algoritmo senza far mai svuotare il buffer di uscita.

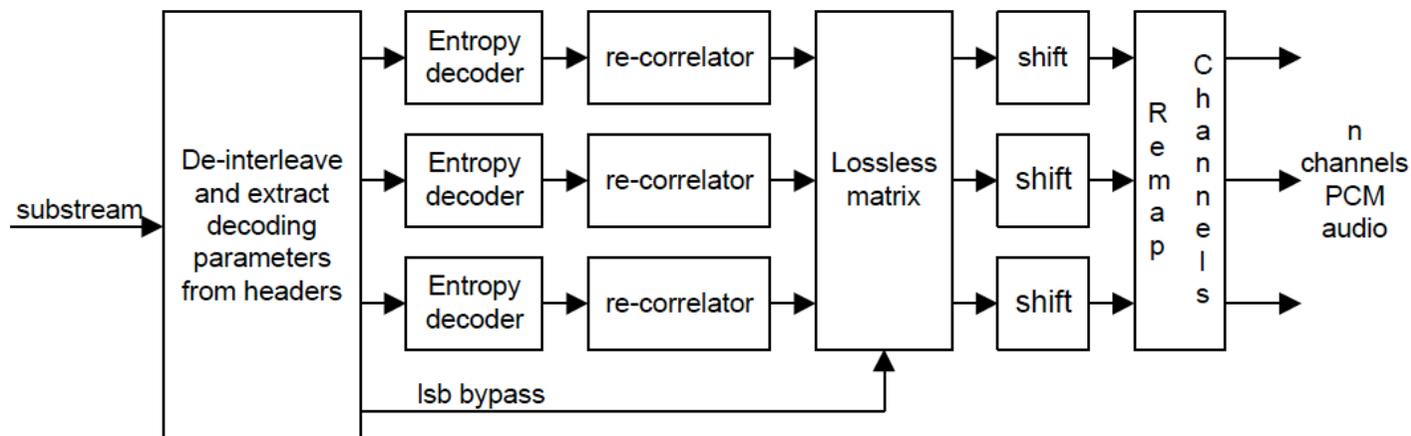
Il risparmio sulla lunghezza del file va da un 20% ad un 60% a seconda del tipo di musica. Vediamo ora una selezione dei compressori **lossless** più usati

Meridian Lossless Packing MLP

È il più *ufficiale* di tutti i compressori **lossless** in quanto il relativo decoder **deve** essere presente in tutti i lettori DVD Audio. Nelle figure seguenti sono mostrati gli schemi del compressore.



e del decompressore



In un **DVD Audio** da **4.76 GByte**, usando **MLP**, si possono ottenere le seguenti durate massime del programma musicale.

5+1 canali 96 kHz 24 bit	100 minuti
6 canali 96 kHz 24 bit	86 minuti
2 canali 96 kHz 24 bit	4 ore
2 canali 192 kHz 24 bit	2 ore
2 canali 44.1 kHz 16 bit	12 ore
1 canale 44.1 kHz 16 bit	25 ore (libro parlante)

FLAC (Free Lossless Audio Codec)

È un compressore **open-source** molto diffuso particolarmente indicato per tenere la musica compressa e suonarla senza decomprimerla, infatti è stato studiato per rendere più semplice e veloce la fase di decompressione.

Vediamo uno schema dei passaggi della codifica **FLAC** (informazioni tratte da **Wikipedia**)

- **Suddivisione in blocchi:** l'input viene diviso in più parti contigue.
- **Compattamento del flusso multicanale:** in questo passo l'encoder **FLAC** si occupa di calcolare, nel caso di input stereo e surround, la media dei canali e la loro differenza.

- **Previsione:** partendo dal primo blocco, avviene la previsione di quale possa essere il successivo con degli algoritmi matematici che tentano di ricostruire il segnale. In questo caso, verranno codificati solo i parametri, dato che gli algoritmi sono presenti sia sull'**encoder** che sul **decoder**.
- **Codifica residua:** la codifica residua permette a **FLAC** di essere effettivamente un **codec lossless**. Si codifica senza perdita tutta la parte di segnale che si differenzia dalla ricostruzione matematica di predizione, e il risultato viene incorporato nel file in uscita.

Monkey APE

È un compressore che raggiunge tassi di compressione migliori di **FLAC** anche se l'algoritmo di decompressione è più complesso e meno adatto al tempo reale. È molto usato per trasferire file musicali via Internet (anche illegalmente visto che non prevede protezioni anticopia). Il codice è pubblico ma non **open-source**.

ALAC (Apple Lossless Compression)

È il compressore **lossless** di Apple e come tale può essere suonato su gli **iPod** e con **iTunes**. Il codice è proprietario e riservato.

Altri compressori **lossless** meno usati sono **Shorten**, **MPEG-4 Audio Lossless Coding (ALC)**, **TTA**.

Vediamo ora un semplice esempio di quanto si riesce a comprimere **lossless**. Tanti sono i programmi di compressione e tante le opzioni che questi dati sono **puramente indicativi**.

È stato usato il programma **MAX** su Apple Macintosh con tre formati di uscita e per confronto si è anche effettuata una compressione standard nel formato **Zip** (che oltre a non tenere conto del fatto che siamo in presenza di segnali musicali non è neppure suonabile in tempo reale). Sono stati scelti due brani classici e uno pop. Quella che è indicata è la percentuale di spazio risparmiato (i valori più alti sono i migliori).

	APE Monkey	FLAC	ALAC	Zip
Battisti	44%	41%	40%	7.6%
Bach	64%	61%	61%	17%
Grieg	67%	64%	63%	14%

NB la qualità di riproduzione dei file lossless è sempre la stessa degli equivalenti non compressi, se si sentono differenze si è in presenza di guasti hardware o di software difettoso o di errori di configurazione o, più spesso, di fenomeni di autosuggestione.

Compressione audio lossy

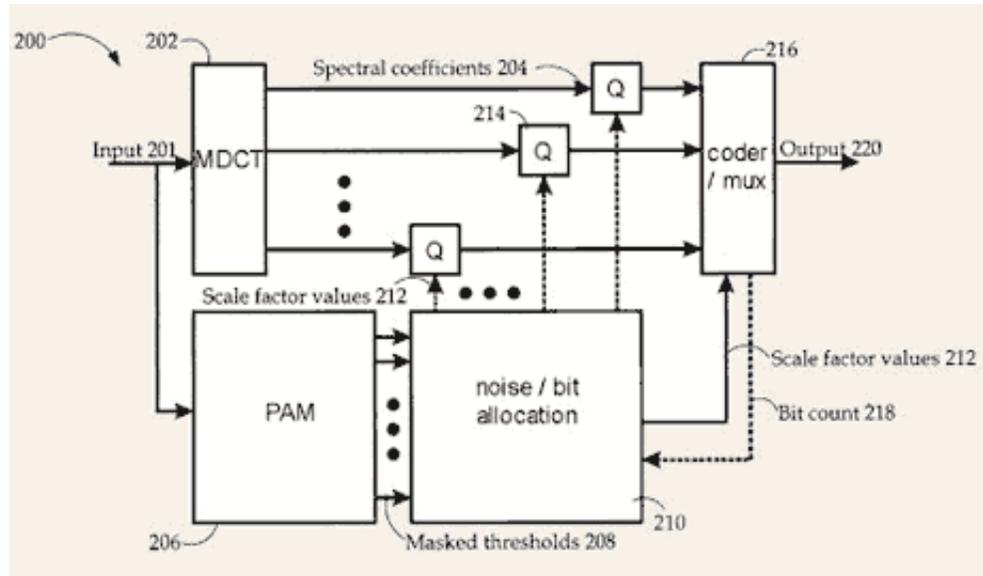
Completamente diverse sono le basi teoriche della compressione **lossy** (con perdita di informazione). L'idea è di buttare via tutto ciò che in qualche modo è inutile nel segnale originario. La compressione **lossy** è usata con successo per la compressione di immagini e filmati tenendo in particolare conto della risoluzione del dispositivo di visualizzazione mentre non è assolutamente applicabile per dati e programmi in cui non si può perdere neppure un byte senza invalidare l'intero file. Nel momento in cui si accetta di perdere informazione si ottiene il vantaggio di non avere limiti al tasso di compressione anche se, ovviamente, non si può andare troppo avanti, per questa strada come mostra l'esempio (grafico) che segue.





Nel caso della compressione di dati audio si usano **modelli psicoacustici** che cercano di individuare quali porzioni di segnale possono essere eliminate perché non udibili. Tipicamente si effettua una **Trasformata di Fourier** nel campo dei **numeri reali** (la **Modified Discrete Cosine Transform, MDCT**) e, osservando il segnale nel dominio della frequenza, si eliminano quelle bande che non contengono segnali significativi o che sono mascherate da segnali vicini più forti.

La figura che segue, presa da una richiesta di brevetto Apple, dà un'idea della complessità di un procedimento di questo tipo.



NB facendo così si butta via sempre *un po' di bambino insieme all'acqua sporca* e la compressione lossy ha senso solo se si ascolta in ambienti rumorosi (treno, autobus, metropolitana) o con impianti di bassa qualità.

Vediamo ora un elenco dei più diffusi formati **lossy** dei giorni nostri.

MP3

Il comitato tecnico **MPEG (Moving Picture Experts Group)** ha codificato nel tempo vari standard di compressione di filmati in cui era presente anche la parte che trattava la compressione audio. Per fare un esempio tutti i **DVD Video** in commercio seguono le standard **MPEG 2**.

Durante lo sviluppo del primo standard **MPEG 1** furono studiati diversi algoritmi di compressione audio denominati **Layer 1**, **Layer 2** e **Layer 3**. Il più sofisticato di questi, il **Layer 3**, una volta che la potenza di calcolo è stata sufficiente ha avuto una diffusione tale che la sua abbreviazione (**MPEG 1 Layer 3** è infatti divenuto **MP3**) è comunemente il sinonimo di compressione Audio.

L'algoritmo originario è stato sviluppato dal **Fraunhofer Institute for Integrated Circuits** che ne detiene il brevetto, ne esistono però anche versioni **open-source** tra cui per esempio **LAME**.

La diffusione di **MP3** fa sì che non solo sia compatibile con tutti i lettori portatili ma anche che decoder integrati esistono in molte apparecchiature stereo consumer (alimentabili attraverso **CD** masterizzati o penne **USB**). Il bit rate più usato è **128 Kbit/sec** (circa un dodicesimo dell'originale), il massimo possibile **320 Kbit/sec** (circa un quinto dell'originale), a questo rate alcuni parlano di “**qualità CD**” ma questa affermazione è sostenibile solo in presenza di impianti di ascolto che non permettano di rilevare i dettagli più fini.

Bisogna anche notare che **MP3** è tuttora il principale veicolo di diffusione **illegale** via **Internet** di musica coperta da diritti di autore.

WMA

Questo formato è stato sviluppato da Microsoft per non pagare al **Fraunhofer Institute** i diritti sul codificatore **MP3** e ha una diffusione abbastanza limitata, si tratta una tecnologia proprietaria che fa parte del **Windows Media Framework**.

Ogg Vorbis

Ogg è un formato “contenitore” **open-source** per il trasporto di suono ed immagini. Il **Vorbis** è il **codec lossy** associato a questo contenitore (come **codec lossless** è stato adottato **FLAC**). La struttura di **Vorbis** è simile a quella di **MP3**. Il formato è molto usato tra i cultori dell'**open-source** specie in ambiente **Linux** ma è meno diffuso sulle apparecchiature elettroniche commerciali. Una caratteristica interessante del formato **Ogg** è che la semplice concatenazione di due file **.ogg** è ancora un file **.ogg** correttamente decodificabile.

AAC (Advanced Audio Coding)

Il formato **AAC** è il formato di compressione audio associata allo standard **MPEG 4**, rappresenta una versione più moderna di **MP3** e a parità di bit rate offre una migliore qualità di ascolto, una frequenza di campionamento fino a **96 KHz**, un maggior numero di canali e molto altro ancora.

AAC è il formato audio standard per **Apple iTunes** per gli apparecchi **Apple iPhone**, **iPod**, **iPad**, **Nintendo DSi**, **Sony's PlayStation 3** e molti altri.

Apple ha messo in commercio tracce musicali compressi con **AAC** a **128 Kbit/s** insieme ad un algoritmo di **DRM (Digital Rights Management)** che ne rende impossibile la diffusione ulteriore su larga scala. Da qualche tempo sono in vendita a prezzo lievemente maggiorato tracce **AAC** a **256 Kbit/s** senza protezione **DRM**.

La opzioni di un compressore lossy

Con qualunque algoritmo di compressione, chi desidera comprimere un file musicale, può scegliere tra numerose opzioni, vediamo una panoramica delle principali di quelle presenti su **MP3** ma adottate, con varianti, da molti altri programmi.

- **Il bit-rate**

È questa la scelta che influisce maggiormente sulla dimensione del file compresso. Un segnale originale stereo in standard **CD** ha un **bit rate** di $16 \times 44100 \times 2 = 1411.2 \text{ Kbit/s}$, mentre il programma di compressione di solito offre una gamma da **32 Kbit/s** a **320 Kbit/s**. Se si sceglie l'opzione **CBR** (**Constant Bit Rate**), il codificatore cerca di tenere costante il **bit rate** selezionato indipendentemente dal contenuto del segnale musicale. Questo consente di stimare in modo accurato la lunghezza del file risultante e garantisce una certa stabilità delle operazioni di taglio che dovrebbe non dispiacere ai puristi. Tipicamente l'opzione **CBR** usa rate compresi tra **128** e **320 KBit/s**. L'opzione **VBR** (**Variable Bit Rate**) permette al compressore di “risparmiare bit” quando il segnale musicale è “povero” in modo da utilizzare lo spazio guadagnato per perdere meno informazioni nei passaggi più difficili. Questa capacità adattiva, a detta degli informatici, dovrebbe migliorare la qualità globale a parità di spazio, ma la lunghezza del file compresso è più difficile da prevedere. Verosimilmente poi viene introdotto un certo “pompaggio” delle microinformazioni che può essere avvertito da un orecchio esperto. In genere oltre alla scelta **VBR** è possibile selezionare una soglia minima di compressione e un livello di qualità che aiutano il codificatore nelle sue scelte.

- **Modo Stereo**

La scelta di privilegiare alcune parti del segnale a scapito delle altre può essere applicata anche ai due canali. Sono spesso disponibili alcune opzioni che consentono di specificare le modalità di trattamento dei due canali:

- **Dual Mono** (talvolta semplicemente **Stereo** o **Normal**): i due canali sono codificati indipendentemente e il risultato impacchettato;
- **Joint Stereo**: il codificatore è autorizzato a elaborare congiuntamente i due canali per risparmiare spazio o migliorare la qualità.

Il codificatore del **Fraunhofer Institute for Integrated Circuits** (ente proprietario dei diritti della codifica **MP3**) prevedeva altri due tipi di codifica:

- **MS Stereo**: sono codificate la somma e la differenza dei due canali (può darsi che questa opzione coincida con **Joint Stereo** almeno per alcuni codificatori).
- **Intensity Stereo**: le alte frequenze sono mixate in mono insieme e viene registrata una informazione sulla direzione di provenienza; questa codifica distrugge le informazioni di fase e non è compatibile con segnali di qualità.

Si noti che nel caso limite di segnali identici sui due canali, un compressore “furbo” in **Joint Stereo** può disporre di un **bit rate** effettivo quasi doppio rispetto a un compressore in **Dual Mono**. Per provare il comportamento di questa opzione sono stati creati due file di **10** secondi di **rumore rosa**, il file **S** con i due canali completamente scorrelati e il file **M** con i due canali identici. I file sono

stati compressi utilizzando tre codificatori MP3 *d'annata*. In tutti i casi in cui si è usata la compressione **CBR**. Nella tabella seguente sono riportate le lunghezze dei file compressi.

Codificatore	Sound Jam	Sound Jam	Audio Catalyst	Audio Catalyst	N2MP3	N2MP3
Opzione	Joint Stereo	Normal	Joint Stereo	Stereo	Joint Stereo	Stereo
file S	160 KByte	160 KByte	280 KByte	288 KByte	240 KByte	200 KByte
file M	120 KByte	160 KByte	148 KByte	288 KByte	128 KByte	200 KByte

In genere conviene scegliere l'opzione **Joint Stereo** a meno che non si abbia a che fare con due tracce mono distinte (come ad esempio una traccia parlata in italiano e una in inglese)

- **Sampling Rate**

Un altro modo per ridurre l'occupazione del file compresso quello di abbassare la frequenza di campionamento, questo uccide le alte frequenze (e l'alta fedeltà del risultato) ma è consigliabile se si vuole comprimere pesantemente del materiale parlato o già originariamente a banda limitata.

- **Trattamento delle frequenze estreme**

Un alternativa meno drastica per ridurre l'occupazione salvando un po' di qualità consiste nel filtrare via le frequenze estreme durante la fase di compressione. Molti programmi per *default* tagliano sopra i **16 KHz**.