

8

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).

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 "Zippatori") 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 con perdita di informazione.

Il primo tipo di compressione, che deve permettere la ricostituzione senza errori dei dati originali, si basa soprattutto sulla eliminazione delle ridondanze. I limiti di questa tecnica sono dettati dal teorema di Shannon (o della codifica in assenza di rumore) che individua nell'entropia di una sorgente il limite alla sua compressibilità. Questa codifica è indispensabile per memorizzare o trasmettere programmi, testi e ogni tipo di informazione che non può essere alterata senza danni (in genere qualunque tipo di informazione originariamente digitale).

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 i dati di natura

analogica (suoni ed immagini) che per la loro natura sono soggetti ad una inevitabile perdita di informazione (causata dal rumore ambientale, la risoluzione dei trasduttori, ecc.)

Nel caso delle immagini vi sono due formati classici che incarnano le due tecniche, il formato **GIF**, privo di perdite e adatto alle immagini piccole, e il formato **JPEG**, capace di ridurre maggiormente le dimensioni, più adatto a conservare immagini di grande formato introducendo però una degradazione della qualità.

Per quanto riguarda la musica il segnale analogico originario (le vibrazioni dell'aria intorno agli esecutori) viene immediatamente degradato dal microfono e dalle apparecchiature analogiche di registrazione a causa delle inevitabili distorsioni e alla presenza di rumore ambientale e termico.

Una volta digitalizzato (nello standard **CD 16bit/44.1Khz** o in qualcuno degli standard più moderni ad esempio **24bit/96Khz**) viene introdotto un ulteriore degrado (inferiore ai micidiali tagli dei supporti analogici su nastro o su vinile) e in questa forma viene memorizzato o trasmesso senza ulteriori perdite di informazione. Infatti, astraendo da possibili deterioramenti dei supporti combattuti con codifiche ridondanti, la copia digitale di un originale è ancora un originale.

Facendo un rapido conto lo standard **CD** su due canali richiede un *bit-rate* di **$16 \times 44100 \times 2 = 1411.2$ kbit/sec** ovvero circa **10 Megabyte/min**. Questo *bit-rate* limita a poco più di un ora di programma il contenuto di un **CD** tradizionale e rende praticamente impossibile la trasmissione su Internet di audio digitale non compresso.

I formati audio compressi sono stati introdotti di recente per risolvere il problema del trasporto di materiale musicale. Gli algoritmi si basano su considerazioni di psicoacustica che aiutano ad individuare e sopprimere nel materiale registrato le informazioni non udibili.

La giungla dei formati audio

Come nell'ambiente grafico, dove le immagini sono trattate e memorizzate in centinaia di modi diversi così in ambito audio vi sono moltissimi formati e molti modi di trattarli. Vediamone una rapida panoramica.

Formati non compressi.

Segnale analogico: viaggia, non digitalizzato e non compresso, su due cavi schermati, uno per canale (per esempio dal giradischi all'amplificatore). Può essere introdotto in un computer attraverso la presa **LINE-IN** della scheda audio per essere ascoltato, digitalizzato e memorizzato.

Segnale audio digitale (SPDIF): viaggia su un solo cavo schermato, o su una fibra ottica, (per esempio dal lettore **CD** ad un convertitore **D/A**). Può essere introdotto in un computer attraverso la presa **DIGITAL-IN** della scheda audio (di solito poco usata e spesso collocata all'interno del *case*) per essere ascoltato e memorizzato, senza bisogno di essere digitalizzato.

CD musicale. I segnali sono digitalizzati e messi in una forma ridondante e “mescolata” che dovrebbe ridurre i problemi di lettura (per esempio per i dischi rigati o sporchi di marmellata). Inserito in un lettore **CD** il segnale viene trasformato in uno dei formati di cui sopra. Invece in un lettore di computer il segnale per essere trattato deve essere trattato da appositi programmi di *ripping* che lo trasformano in un formato memorizzabile su *hard-disk*. L'operazione di *ripping* è molto delicata e l'integrità dei dati non è garantita.

AIFF. È un formato audio per computer introdotto dalla Apple. I dati sono scritti senza ridondanza e (tipicamente) senza compressione ma con informazioni accessorie sufficienti per conoscere il

numero dei canali, la frequenza di campionamento, il numero di bit per campione, etc. E' uno dei formati che la scheda audio può trasferire e scrivere sull'*hard-disk*.

WAV. E' simile all'**AIFF** ma più diffuso in ambiente Windows.

È possibile passare da **WAV** ad **AIFF** e viceversa senza perdita di informazione. È facile fare programmi (Pascal, C, Java etc.) che leggono questi formati e li rielaborano. I formati **WAV** e **AIFF** possono essere trasferiti su **CD-ROM** così come sono, ed in questo caso potranno essere rilette senza perdite da un lettore **CD** per computer ma non suonati su un lettore **CD** audio. In alternativa, questi formati possono essere masterizzati in formato **CD** audio, generando un prodotto che funziona (spesso ma non sempre) sui lettori audio, ma che per essere usato su un **PC** richiede un nuovo *ripping*. In questo secondo modo vengono prodotte le “famigerate” copie pirata dei **CD** commerciali che hanno causato il boom della vendita dei **CD-ROM** scrivibili.

Formati compressi

I formati compressi permettono un più agevole trasporto e immagazzinamento del materiale musicale a spese di un, più o meno avvertibile, degrado della qualità dello stesso. Il loro funzionamento si basa su considerazioni psicoacustiche: vengono rimosse le informazioni *irrilevanti* lasciando quelle necessarie alla ricostruzione del messaggio musicale. Naturalmente il concetto di “*irrilevante*” è opinabile e in pratica si ha una gamma di possibili risultati di qualità inversamente proporzionale all'entità della compressione. La cosiddetta “**Qualità CD**” dei file compressi è sicuramente una balla.

Formati compressi sono stati usati nella **DCC** (la cassetta digitale Philips e Technics), nel **MINIDISC** Sony.

Il formato più noto (soprattutto per i suoi legami con Internet) è il formato **MP3** abbreviazione di **MPEG-1 layer 3**.

MPEG sta per **Moving Pictures Experts Group** che è un gruppo di esperti di un sottocomitato **ISO/IEC** che ha formulato un certo numero di standard per la compressione Audio e Video. **MPEG-1 layer 3** prevede un certo numero di tecniche di codifica **mono** e **stereo** con *bit rate* che vanno tra **32** e **320 kbit/sec**. Una volta codificato un file **MP3** può essere memorizzato nell'*hard-disk* e suonato da programmi appositi. Il file è indipendente dalla piattaforma (Windows, Macintosh, Linux etc.). I dati possono essere trasferiti sui lettori portatili (che li memorizzano su memoria non volatile), oppure masterizzati in **CD-ROM** e letti su un altro computer o su appositi lettori (come quello in prova in questo stesso numero). Ormai defunto è il formato **MP2 (MPEG-1 layer 2)** compatibile con i lettori **MP3**, che con un algoritmo più semplice permette (a parità di qualità) una minore compressione. Il formato **MP3** permette di includere nel file dei *tag* con informazioni del tipo autore-esecutore, genere, album che facilitano la gestione di grosse collezioni di pezzi.

Un altro formato compresso è il **WMA (Windows Media Audio)** introdotto da Microsoft per la piattaforma Windows.

Il formato **MIDI** merita un posto a parte: non è propriamente un formato audio compresso, quanto una “partitura elettronica” che può facilmente essere suonata da molti player di facile reperibilità. Da una parte si perdono totalmente i veri suoni, d'altra per un musicista esperto il formato **MIDI** è di grande aiuto per “creare” la sua musica.

Le codifiche **MPEG** sono andate avanti: **MPEG-2** è la codifica usata nel **DVD**, **MPEG-4** è la codifica usata nei file video più moderni e la versione audio di **MPEG-4** è la famiglia degli **AAC (Advanced Audio Coding)**.

Come procurarsi i file compressi

Esistono due tecniche principali per procurarsi un file audio compresso : farselo o farselo dare.

Partendo da un **CD-audio** è possibile estrarre i dati convertendoli in formato **WAV** o **AIFF** e poi comprimere questi file con un opportuno encoder. Esistono encoder più evoluti che codificano direttamente a partire dal **CD**. Naturalmente è possibile comprimere anche file audio non provenienti da **CD** (ad esempio le proprie esibizioni canore, opportunamente digitalizzate). In questo caso si ha la completa libertà di scegliere l'uno o l'altro *encoder* e le varie opzioni di compressione.

In alternativa si possono prendere file compressi già pronti ricevendoli dagli amici (REATO!!) o cercandoli in rete (REATO!!).

Dal punto di vista tecnico i file trovati in rete sono dei tipi più vari: il formato più comune è **MP3 CBR 128 stereo**. Sull'**Apple Music Store** è possibile acquistare legalmente brani **AAC**.

Una volta che i file compressi sono sul nostro *hard-disk* vi sono numerose utilizzazioni possibili:

- è possibile farsi una libreria di brani con tanto di motore di ricerca per crearsi al volo ogni possibile *compilation* e ascoltarla poi sfruttando la scheda audio e le casse del **PC** (oppure un impianto stereo tradizionale collegato alla scheda audio).
- se si possiede un masterizzatore è possibile produrre **CD-ROM** con una decina di ore di musica ciascuno, sia per archiviare la musica che per ascoltarla sui numerosi dispositivi capaci di suonare questi **CD**. Attenzione: se usate un Macintosh il **CD** va masterizzato in formato **ISO9600** (quello dei **PC**) altrimenti il **CD** è leggibile solo sui Macintosh.

- è possibile scaricare via **USB** o porta parallela una compilation su un lettore portatile, per poi ascoltare musica in treno, in bici o mentre si fa *jogging*.

Le comuni tipologie di ascolto dei file compressi sono su *walkman* durante gli spostamenti, in auto, durante l'uso del computer come musica di sottofondo. Le limitazioni del formato sono certamente mascherate dalle inevitabili limitazioni degli economici apparati utilizzati, e il vantaggio di disporre in modo agevole di tutta la propria musica in un unico gigantesco *Juke-box*, può cambiare radicalmente il modo di ascoltare.

La opzioni della compressione MP3

Non vale certamente la pena di ripetere gli elementi base delle tecniche di compressione audio per personal computer, è necessario però sottolineare il concetto essenziale: gli algoritmi di compressione non sono univoci ma solo delineati e si prestano a molte diverse implementazioni (più o meno quanto accade per la conversione D/A dei CD Audio). In pratica, chiunque desidera comprimere un file musicale, può scegliere tra molti programmi diversi, ciascuno dei quali dotato di numerose opzioni. Come vedremo la scelta del programma e delle opzioni influisce nettamente sul risultato finale.

Presentiamo una panoramica delle principali opzioni.

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 $16 \times 44100 \times 2 = 1411.2$ 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**) prevede 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 (descritti nel seguito). In tutti i casi in cui si è usata la compressione **CBR** i due file sono risultati della stessa lunghezza. La compressione **VBR**, invece, è capace di sfruttare la somiglianza tra i canali per ottimizzare lo spazio. 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**.

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**.

L'Analisi Multitono fatta in casa

La tecnica di analisi proposta consiste nel generare via software dei file **AIFF** o **WAV**, comprimerli, decomprimerli e analizzare il risultato finale. Una volta generati i file di prova possono essere distribuiti, permettendo così la ripetibilità della prova.

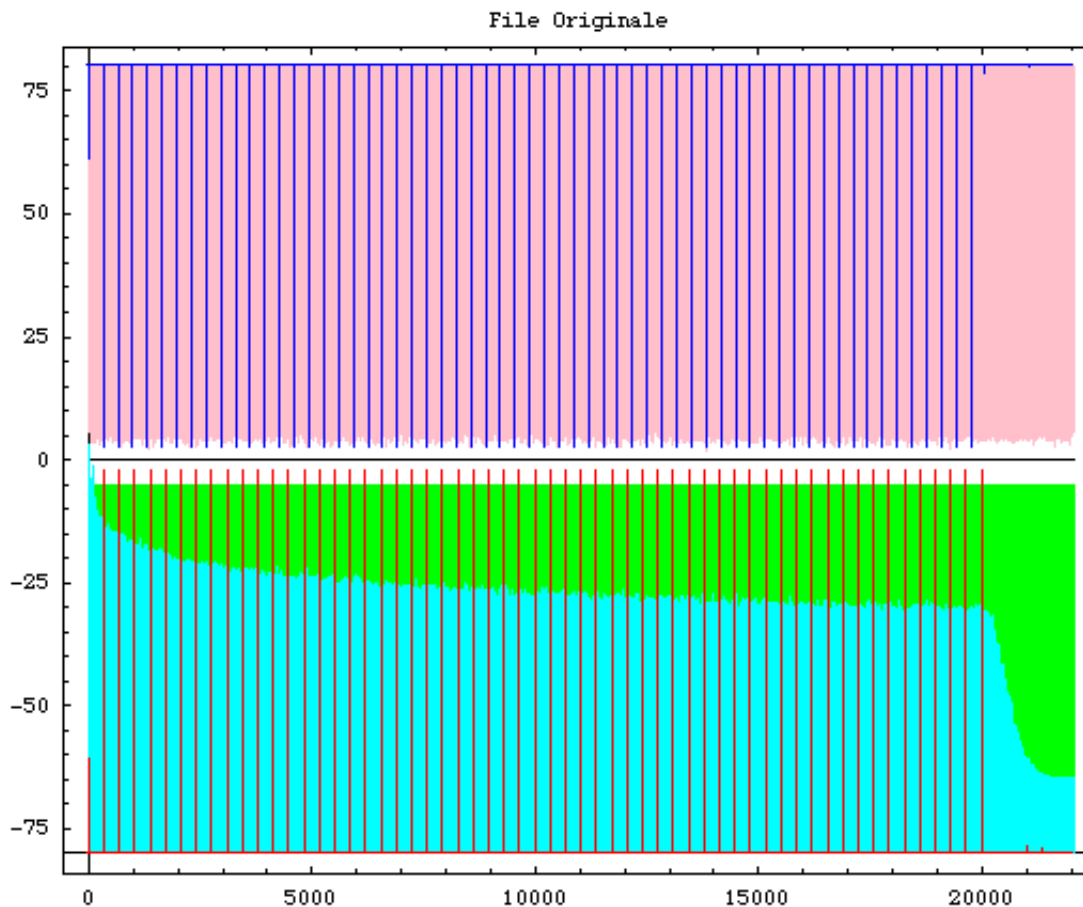
I file usati (generati con *Mathematica*) sono i seguenti:

- **58** segnali a frequenze equispaziate di **44100/128 Hz**. La scelta delle frequenze permette di fare un'analisi spettrale su **16384** campioni senza produrre artefatti.
- **60** segnali a frequenze equispaziate di **44100/134 Hz**. La scelta delle frequenze permette di fare un'analisi spettrale su **16750** campioni senza produrre artefatti.
- Un segnale consistente in **39999** campioni nulli e un campione (in posizione centrale) di valore massimo. Questa analisi permette di rilevare la risposta impulsiva del sistema. Anche se in presenza di non linearità la risposta all'impulso non permette di caratterizzare completamente il comportamento del compressore, ma fornisce comunque interessanti informazioni.
- Alcuni secondi di rumore bianco.
- Alcuni secondi di rumore rosa (rumore bianco filtrato a **6dB** per ottava).

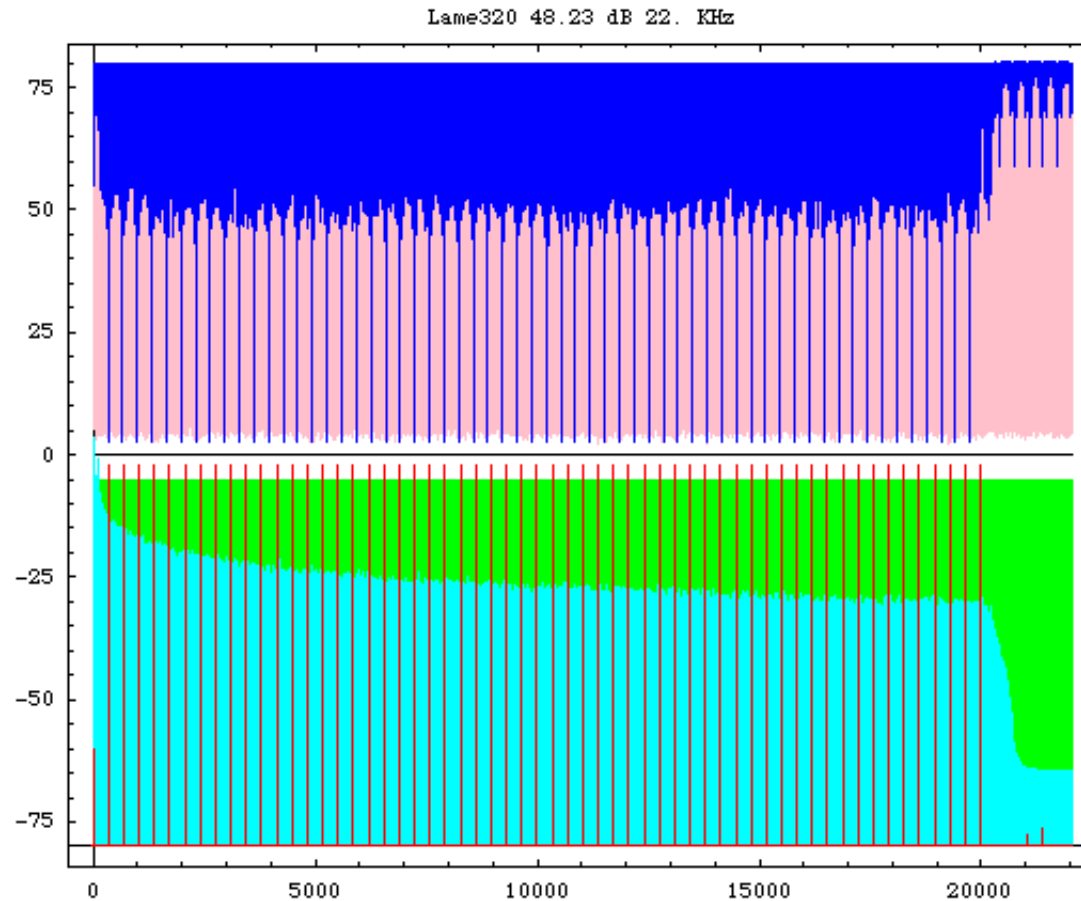
L'analisi dei risultati viene effettuata attraverso la Trasformata di Fourier. L'idea della presentazione consiste nel mettere in uno stesso grafico tutte e **5** le risposte in frequenza. La figura è divisa in due parti. Le ordinate contengono valori in **dB** (da **-80 a 0** quella inferiore e da **0 a 80** quella superiore). Nella parte bassa trovano posto tre grafici, sullo sfondo c'è la risposta impulsiva (in **verde**) con il bordo superiore a **-5dB**); sopra di essa c'è la risposta al rumore rosa (in colore

celeste) portata a circa **-20dB**; Infine, nello strato più esterno, ci sono i **58** multitono (in **rosso**) portati a circa **-2dB**. Nella parte alta sono collocati gli ultimi 2 grafici: sullo sfondo la risposta al rumore bianco (in colore **rosa!**) portata a circa **-5dB** e nello strato più esterno, ci sono i **60** multitono (in **blu**) portati a circa **-2dB**.

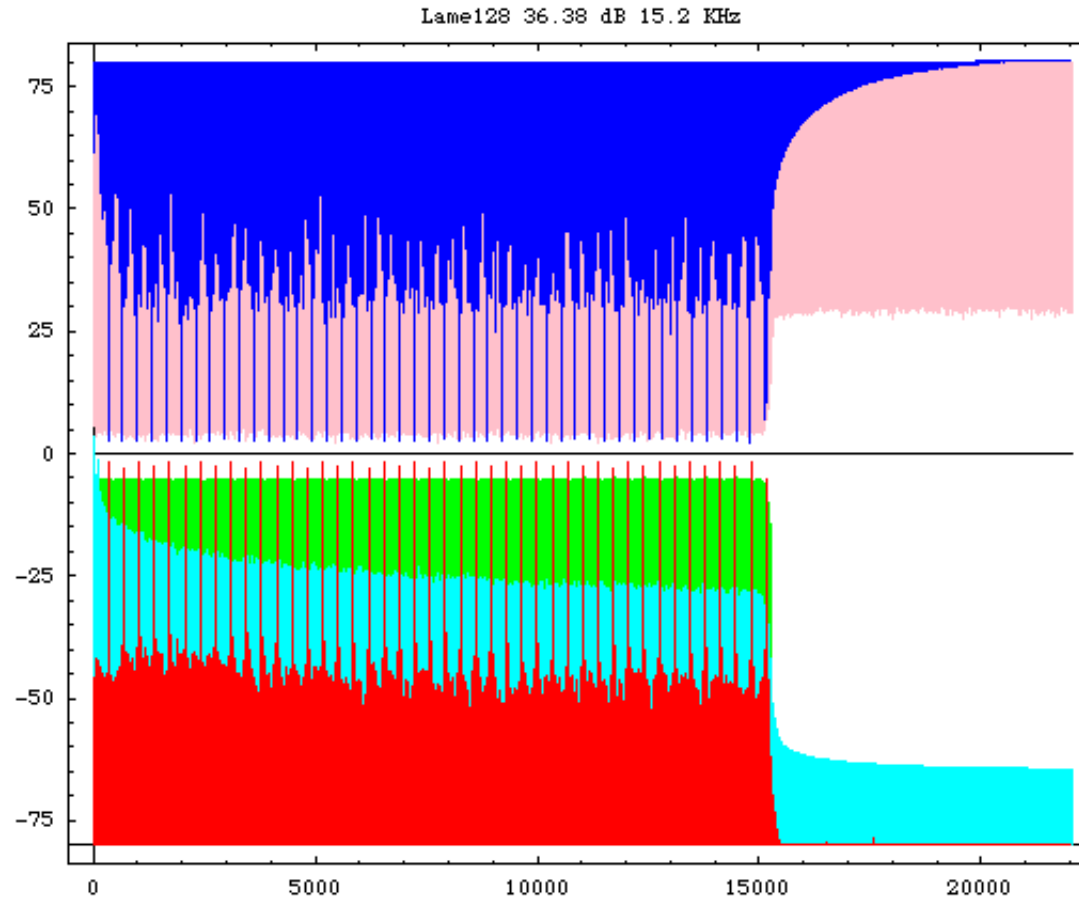
Vediamo il grafico di uscita del segnale originale non compresso



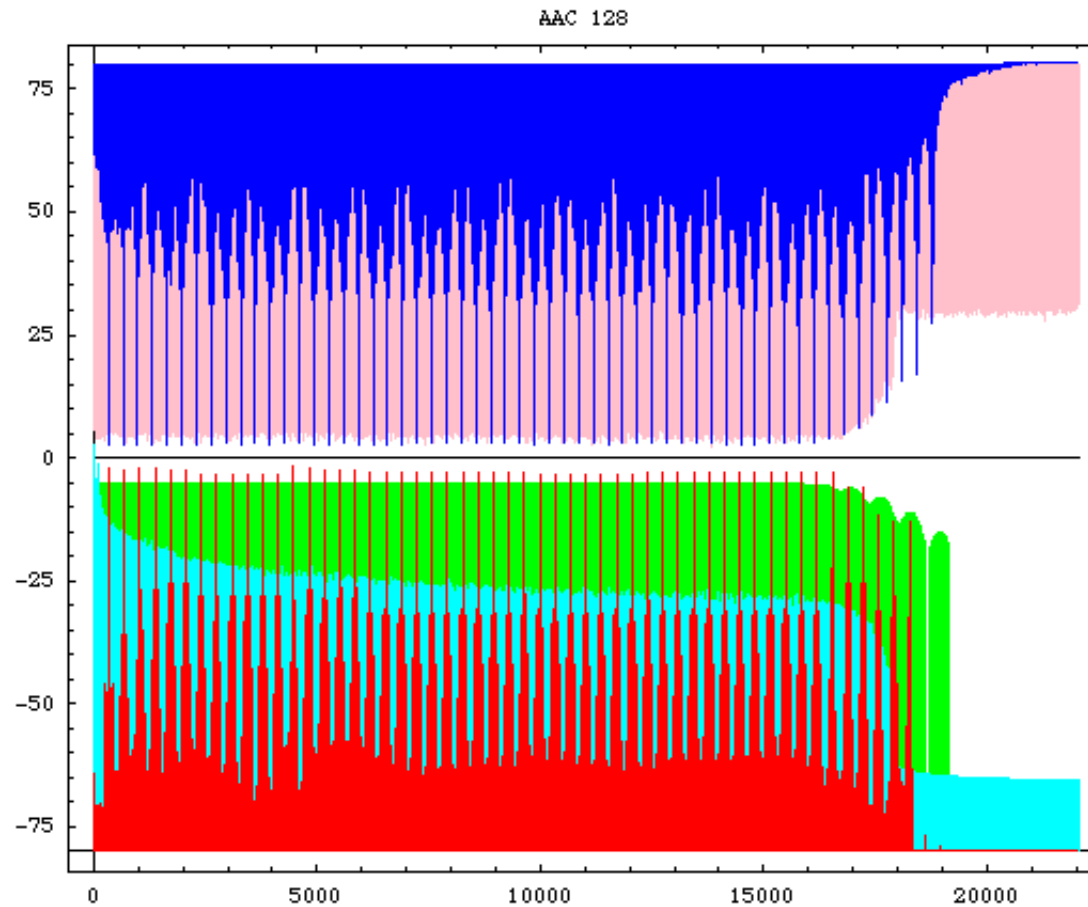
Vediamo il grafico ora di uscita con un ottimo compressore (il migliore che abbia mai provato: **LAME 320**). La risposta con la finestra di analisi inferiore (lunghezza del buffer potenza di 2) è praticamente perfetta, se ne deduce che i prodotti di intermodulazione visibili nella finestra al piano superiore dipendono dal fatto che in quel caso la finestra di analisi non ha una lunghezza potenza di due e (probabilmente) interferisce con le dimensioni delle finestre di elaborazione.



Passando ad una compressione più spinta (**LAME 128**) si ottiene ancora un grafico molto pulito ma si nota che i prodotti di intermodulazione compaiono in entrambe le finestre (stavolta sono legati alla compressione) è anche evidente il taglio a **15KHz** operato dal compressore.



E' interessante confrontare **LAME 128** con un compressore di nuova generazione: **AAC 128** nella versione di **iTunes**.



Il taglio ad alta frequenza è più dolce e anche la risposta sembra più regolare.

Una compressione a **64Kbit/sec** evidenzia drammaticamente il massacro operato sul segnale con un taglio a **6Khz**. La pulizia dei pochi multitono rimasti fa però sperare che all'ascolto (in metropolitana) ci sia ancora un po' di musica.

Ecco infine una compressione **VBR** (compressore **N2MP3** per MAC).

