

SUPSI

Audio analogico vs. audio digitale

Fondamenti di Multimedia Processing

Tiziano Leidi

28.09.2018

Audio analogico vs. audio digitale

Oggi, praticamente tutta l'industria audio (sia professionale che consumer) è passata al digitale.

La transizione è in atto da decine di anni, ... ma l'evoluzione non è ancora finita.

I vantaggi dell'audio digitale sono innumerevoli, in particolare in termini di qualità sonora che si può ottenere e mantenere, sia all'interno della catena di produzione sonora, sia per l'immagazzinamento e distribuzione dei dati di lavorazione e dei prodotti finiti.

Audio analogico vs. audio digitale

Le soluzioni audio analogiche non sono però completamente scomparse.

Alcune parti della sequenza di produzione (soprattutto quelle vicine ai microfoni e agli altoparlanti) mantengono, per necessità o per vantaggio, la loro natura analogica.

Audio analogico - cenni storici

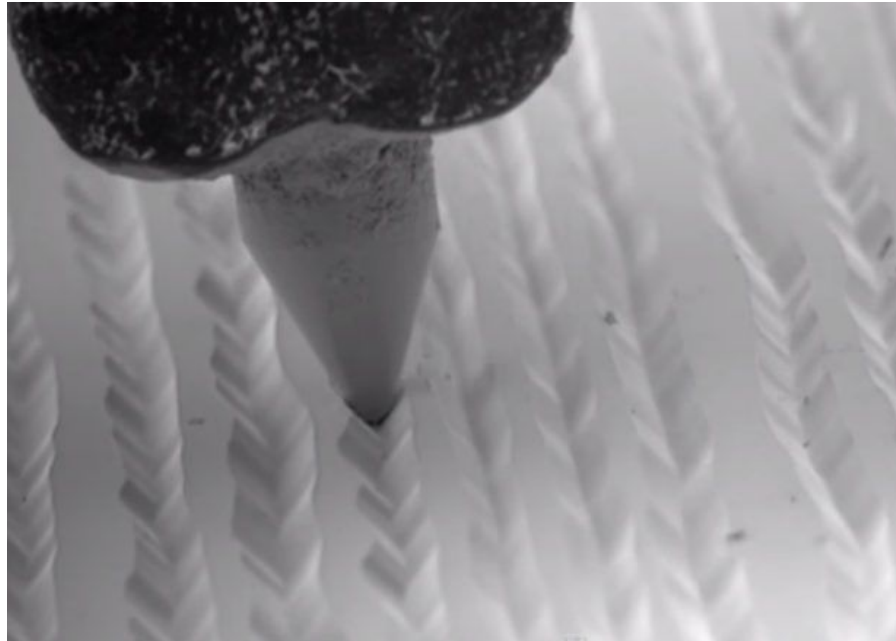
La tecnologia audio analogica ha cominciato a svilupparsi (principalmente per la registrazione sonora) nella seconda metà del 19esimo secolo con l'invenzione di sistemi meccanici come il fonautografo e il fonografo.

Attorno agli anni '20 sono state poi introdotte le prime soluzioni elettroniche, ad esempio per la registrazione con microfoni e con nastri.



Audio analogico

I metodi di registrazione audio analogica immagazzinano il segnale audio direttamente all'interno (o sulla superficie) dei supporti di registrazione, modificandone la struttura fisica.



Audio analogico

Le soluzioni elettroniche, si avvalgono di:

- trasduttori in grado di convertire i segnali sonori in segnali elettrici (microfoni)
- mezzi per la memorizzazione dei segnali elettrici (ad esempio supporti magnetici)
- trasduttori in grado di convertire i segnali elettrici in suoni (altoparlanti)

L'elettronica ha anche permesso la realizzazione di sintetizzatori sonori programmabili in grado di generare onde acustiche di diversa intensità, frequenza e timbro.

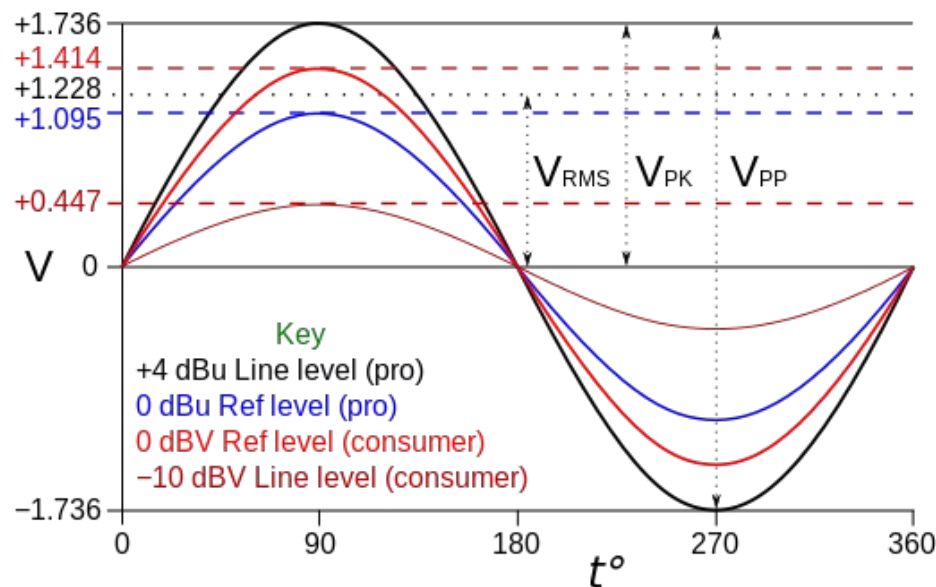
Audio analogico

Un segnale audio analogico è un segnale continuo, per cui la caratteristica variabile nel tempo, che nel caso dell'audio è solitamente il voltaggio istantaneo, rappresenta una caratteristica fisica anch'essa variabile nel tempo, come ad esempio la pressione delle onde sonore.



Audio analogico

Vengono utilizzati vari livelli di segnale elettrico per rappresentare i segnali audio. Grossa differenza esiste fra i segnali amplificati e quelli “low-power”. Inoltre, per i segnali “low-power” esistono diverse norme in ambito professionale e consumer.



V_{PK} : Picco
 V_{PP} : Picco a Picco
 V_{RMS} : Root Mean Square

dBV e dBu

dBV e dBu sono unità di misura usate per le intensità di voltaggio. Come per i dB SPL si tratta di valori relativi a specifici valori di riferimento.

dBV: voltaggio RMS relativo a 1 V

dBu: voltaggio RMS relativo a $\sqrt{0.6} \text{ V} \approx 0.7746 \text{ V} \approx -2.218 \text{ dBV}$

Range dinamico

Il range dinamico descrive il rapporto di ampiezza fra l'onda sinusoidale più forte possibile (non distorta) e il rumore di fondo (rumore rms) del sistema in uso.

Il range dinamico viene spesso confuso con il signal-to-noise ratio (SNR), che è la differenza fra il livello medio del segnale e il livello medio del rumore di fondo.

Come abbiamo visto, il range dinamico dell'orecchio umano è attorno ai 140 dB, variabile a dipendenza della frequenza. La fascia tra i 120 dB e i 140 dB è detta soglia del dolore.

Il range dinamico della musica in una sala concerti è tipicamente di 80 dB, mentre quello di un discorso è tipicamente attorno ai 40 dB.

Range dinamico di segnali analogici

Nei segnali audio analogici, il range dinamico dipende dal voltaggio massimo concesso e dal livello del rumore di fondo.

Ad esempio con un massimo di 5 V (rms) e con un rumore di fondo di 10 μ V (rms) il range dinamico è di 500000:1, quindi 114 dB.

Vari processi contribuiscono al rumore di fondo di un sistema analogico. Può provenire dal microfono, dal preamplificatore, dai cavi, etc.

Livelli e headroom

I livelli di voltaggio specificati dagli standards vengono detti livelli nominali. Sono livelli per cui gli apparecchi sono stati disegnati in modo da lavorare in maniera efficiente, sia in termini di range dinamico che di headroom disponibili.

I circuiti elettrici degli apparecchi limitano il range dinamico. A un lato vi è il segnale massimo producibile senza distorsione, all'altro lato il rumore interno prodotto dalle componenti elettroniche.

L'headroom è la differenza fra il livello nominale e il livello massimo, oltre il quale il segnale viene compromesso dalla distorsione.

VU meters

I VU meters vengono utilizzati per rappresentare i livelli di intensità sonora negli apparecchi come i mixer. L'unità di misura è il Volume Unit (VU).

In situazione normali, 0 VU permette un headroom di almeno 18 dB senza distorsione significativa.

Solitamente, negli apparecchi professionali $0 \text{ VU} = +4 \text{ dBu}$, mentre negli apparecchi consumer $0 \text{ VU} = -10 \text{ dBV}$.

Trasduttori

Un trasduttore (transducer) è una device elettronica che converte energia da una forma all'altra.

Nell'audio vengono principalmente utilizzati:

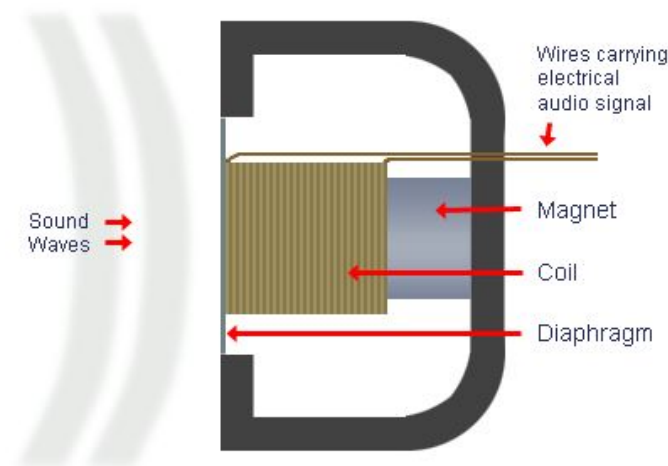
- Microfono (microphone), ovvero un trasduttore elettroacustico in input che converte un'onda sonora in un segnale elettrico.
- Altoparlante (loudspeaker), un trasduttore elettroacustico in output che esegue l'operazione opposta del microfono.

Il principio base del funzionamento di microfoni e altoparlanti è simile. Avete già provato a collegare un paio di cuffie all'ingresso di un convertitore analogico-digitale per utilizzarle come microfono?

Microfoni

Esistono più tipi di microfoni, costruiti sfruttando differenti tecnologie per l'acquisizione di segnali sonori.

L'esatto meccanismo di funzionamento varia a seconda del tipo di tecnologia utilizzata.



Avremo occasione di approfondire le conoscenze sui microfoni nel corso delle prossime lezioni.

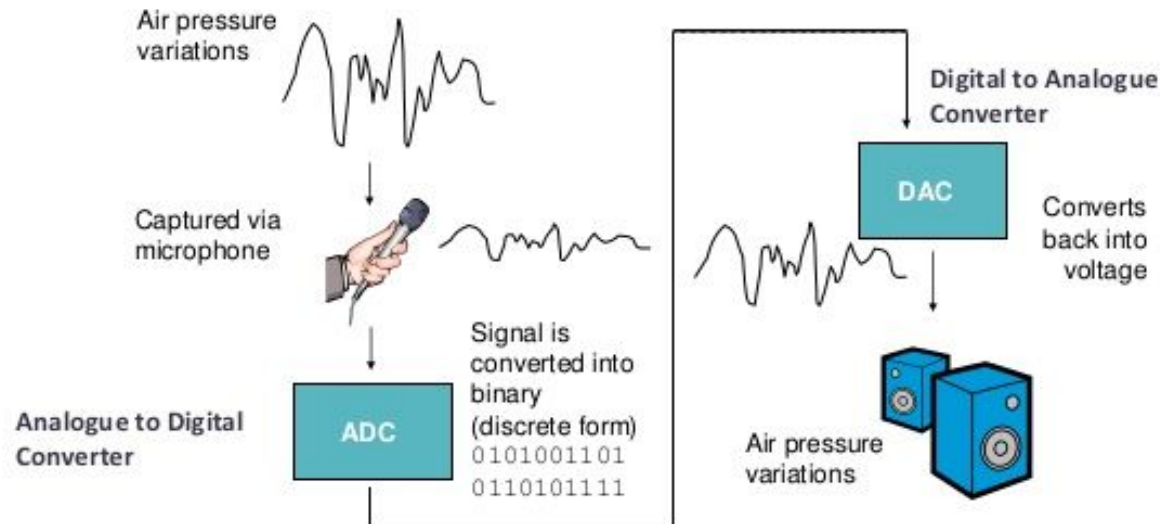
Altoparlanti

Anche le tipologie di altoparlanti sono molteplici. Anche nel caso degli altoparlanti, l'esatto meccanismo di funzionamento varia a dipendenza del tipo di tecnologia utilizzata.

<https://www.youtube.com/watch?v=ln7RMmtb4gw>

Audio digitale - cenni storici

La tecnologia per l'audio digitale ha cominciato a svilupparsi negli anni '70 ed '80, principalmente con l'avvento dei PC. A partire dagli anni '90 ha poi rimpiazzato buona parte della tecnologia audio analogica per la registrazione sonora, ma anche per la produzione sonora, l'audio engineering e le telecomunicazioni.



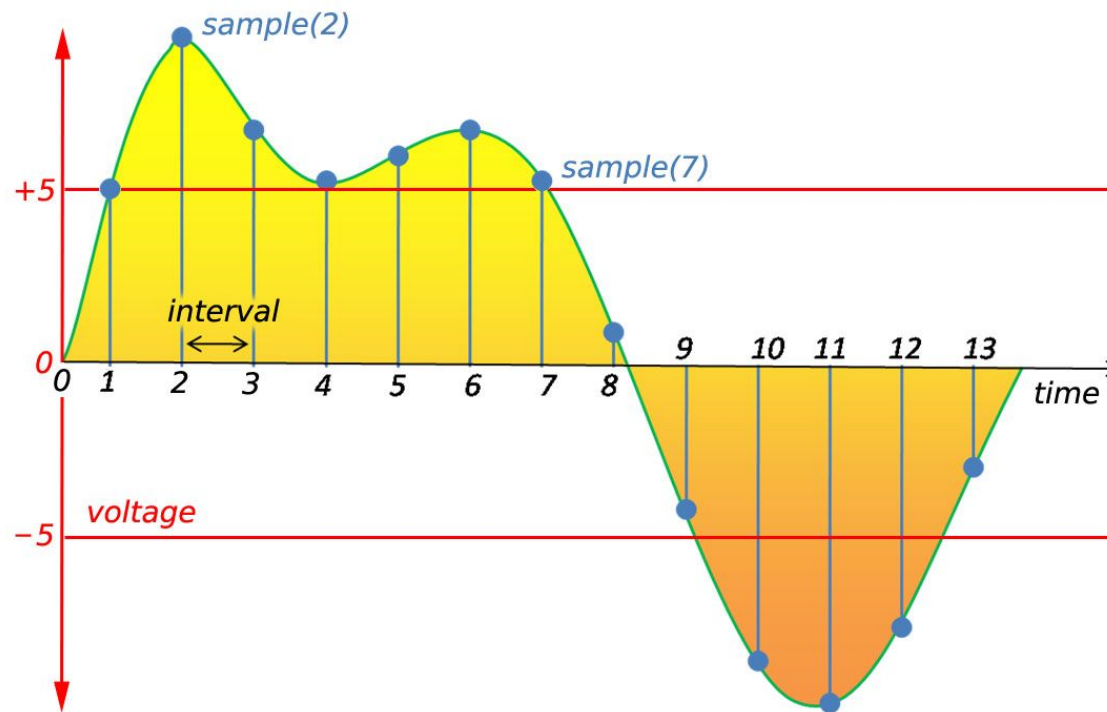
ADC and DAC

ADC (Analog to Digital Converter): è un sistema che converte un segnale analogico in un segnale digitale. Di norma, il segnale digitale prodotto riflette l'andamento del voltaggio della corrente in entrata. La conversione include il processo di quantizzazione (introduce di conseguenza piccoli errori).

DAC (Digital to Analog Converter): è un sistema che esegue l'operazione inversa dell'ADC. Il processo di conversione può degradare il segnale. In particolare, il problema che può intervenire è quello della distorsione armonica.

Conversione analogico-digitale

Nella conversione analogico digitale, viene utilizzato il processo di campionatura per acquisire, a intervalli regolari, dei campioni (samples) del segnale elettrico.



Pulse code modulation

La tecnica di conversione più utilizzata è quella di pulse code modulation (PCM): ad essere campionata a intervalli regolari è l'ampiezza del segnale analogico. Ogni campione viene quantizzato al valore più vicino all'interno del range dei possibili valori digitali.

Esistono quindi due caratteristiche principali di un segnale audio:

- La quantità di bits utilizzati (bit depth): influenza la qualità sonora in termini di dinamica massima acquisibile.
- La frequenza di campionamento (sample rate): influenza la qualità sonora in termini di frequenza massima acquisibile.

Range dinamico di segnali digitali

Nei segnali digitali il range dinamico massimo ottenibile è limitato dell'errore di quantizzazione e quindi dipende dal numero di bits utilizzati nella conversione.

bits	8	12	16	18	20	24
dB	48	72	96	108	120	144

Di conseguenza la dinamica massima ottenibile con un segnale audio in qualità CD è di 96 dB.

Esistono però algoritmi di noise shaping (dithering) capaci di spostare la percezione del range dinamico a 120 dB, approfittando della risposta in frequenza dell'orecchio umano.

Range dinamico di segnali digitali

In informatica 16 bits (fixed point) equivalgono al tipo di dato “short”. Il range di valori è quindi tra $-32'768$ e $32'767$.

Oggi, 16 bits o 24 bits (fixed point), vengono utilizzati principalmente per l'immagazzinamento dei dati su files.

Per il processing, il formato più utilizzato è quello dei 32 bits virgola mobile, che permette range dinamici più ampi, avvalendosi della rappresentazione con mantissa ed esponente (ad esempio 0.1×10^{-7}).

Decibels full-scale

Per misurare l'ampiezza dei segnali digitali viene utilizzata l'unità di misura dB_{FS} .

È un'unità di misura che mette in relazione l'ampiezza del segnale con l'ampiezza del picco massimo permesso da quel sistema (detto il full-scale).

Oltre alla soglia del full-scale interviene il clipping, una forma di distorsione molto aggressiva.

Al full-scale è assegnato il valore di 0 dB_{FS} , quindi tutti i valori più piccoli del massimo sono negativi.

Ad esempio, un segnale che ha metà dell'ampiezza massima, avrà un'ampiezza di $-6 \text{ dB}_{\text{FS}}$.

VU meters

Negli apparecchi digitali, 0 VU ha normalmente un valore di -18 dB_{FS} (18 dB sotto al full scale).

Per le registrazioni a 24-bit (soprattutto per i master) l'allineamento è solitamente a -24 dB_{FS} per permettere un headroom aggiuntivo, da sfruttare come margine di manovra nel momento della conversione al formato di distribuzione.

Sample rate

Il sample rate è la frequenza di acquisizione dei campioni. Viene misurato in Hz ed è proporzionale alla durata dell'intervallo che intercorre tra l'acquisizione di un campione e quello successivo. Convertire con un sample-rate di 44.1 KHz significa acquisire 44100 campioni ogni secondo.



Sample rate

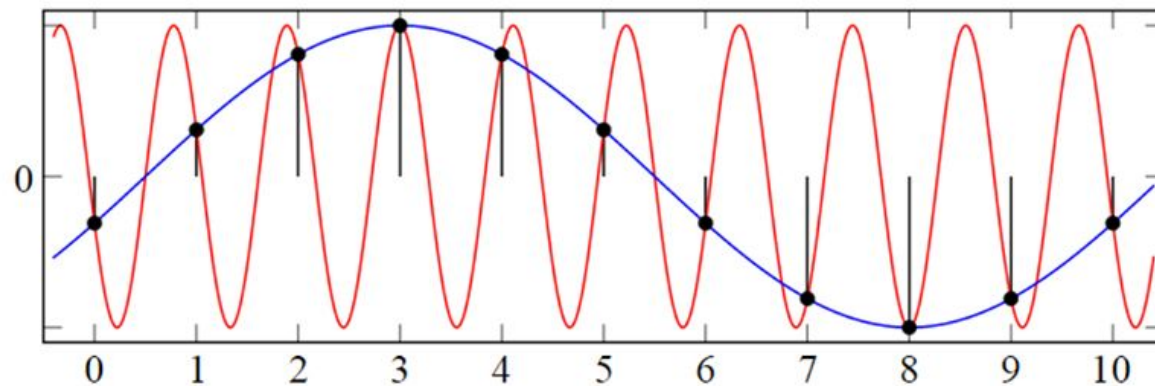
La frequenza massima rappresentabile con un segnale digitale è limitata dal sample rate utilizzato nella la conversione.

Il massimo raggiungibile è la frequenza equivalente alla metà del sample rate. Questa frequenza è della frequenza di Nyquist (in onore di Harry Nyquist).

Ad esempio, con un sample rate di 44.1 KHz è possibile campionare segnali con frequenze fino ad un massimo di 22050 Hz.

Il teorema di Nyquist-Shannon

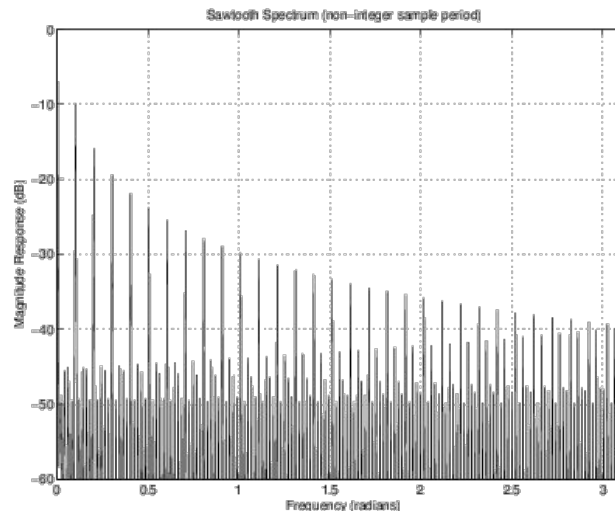
Il teorema di Nyquist-Shannon detto anche sampling theorem, stabilisce le condizioni sufficienti di sample-rate (il doppio della frequenza massima che è necessario rappresentare) per permettere ad una sequenza discreta di campioni di catturare l'informazione di un segnale continuo con una larghezza di banda definita.



Aliasing

Se si campionano frequenze oltre alla frequenza di Nyquist interviene un fenomeno detto aliasing.

Si osserva un effetto di folding. Le frequenze oltre la frequenza massima vengono riflesse all'interno dello spettro, generando frequenze inesistenti.



È possibile evitare l'aliasing applicando un filtro (passa bassa) al segnale. Va applicato prima della conversione.

Jitter

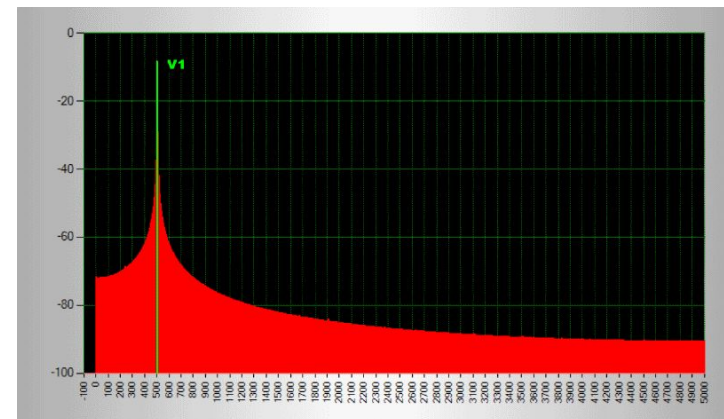
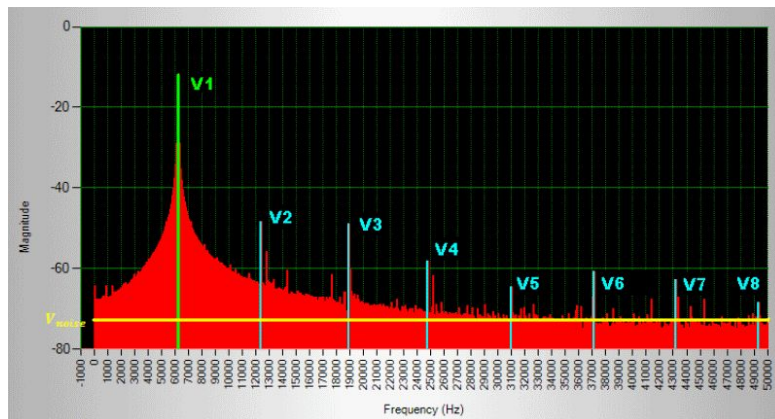
Nel processo di conversione esiste un ulteriore fattore che determina la qualità del segnale risultante.

Il sampling jitter è una lieve variazione nella precisione dei clock dei convertitori ADC o DAC. Questa lieve alterazione può modificare lo spettro delle fasi del segnale, riducendo conseguentemente la qualità di riproduzione, in particolare per quanto concerne i transienti del segnale.

Total Harmonic Distortion (THD)

La Total Harmonic Distortion (THD), è il valore RMS degli armonici prodotti dai convertitori (ADC o DAC) relativi al livello RMS di un segnale input sinusoidale vicino al full-scale. Ad esempio, assumendo un segnale input di frequenza f , la non linearità del convertitore produce le frequenze armoniche $2f$, $3f$, $4f$, ... non presenti nel segnale originale. Generalmente queste frequenze armoniche distorcono il segnale di output.

THD+N è la THD con aggiunto il rumore di fondo (noise).



Formati audio per files

I formati audio per files permettono di immagazzinare dati audio su hard-disks e altre memorie persistenti. Vengono utilizzati anche per lo streaming di dati su internet.

Il formato di audio coding è la soluzione utilizzata per immagazzinare i dati. Può essere di tipo RAW (vengono immagazzinati direttamente i dati in formato PCM) oppure compresso per ridurre la taglia del file.

Anche se è possibile immagazzinare i dati RAW senza ulteriore informazioni, solitamente vengono utilizzati formati container o formati specifici audio, con metadati contenenti informazioni sul contenuto del file (come il numero di bits e il sample-rate dei dati PCM).

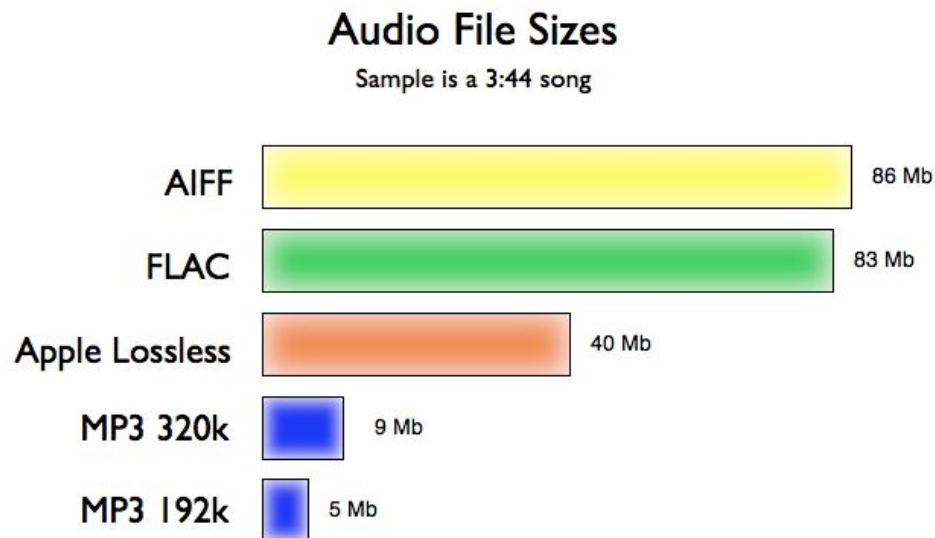
Formati audio per files

Ci sono tre maggiori tipologie di formati audio per files:

- Formati non compressi come .wav, .aiff o .au. Contengono direttamente i dati in formato PCM, preceduti da un header. In alcuni casi possono contenere direttamente i dati RAW senza alcun header.
- Formati con compressione lossless, come .flac, .ape (Monkey's Audio) o .wv (WavPack). Contengono i dati PCM, ma compressi con strumenti simili al .zip.
- Formati con compressione lossy, come .mp3, .aac, .opus. Sfruttano forme di compressione psicoacustica per ridurre la dimensione dei files.

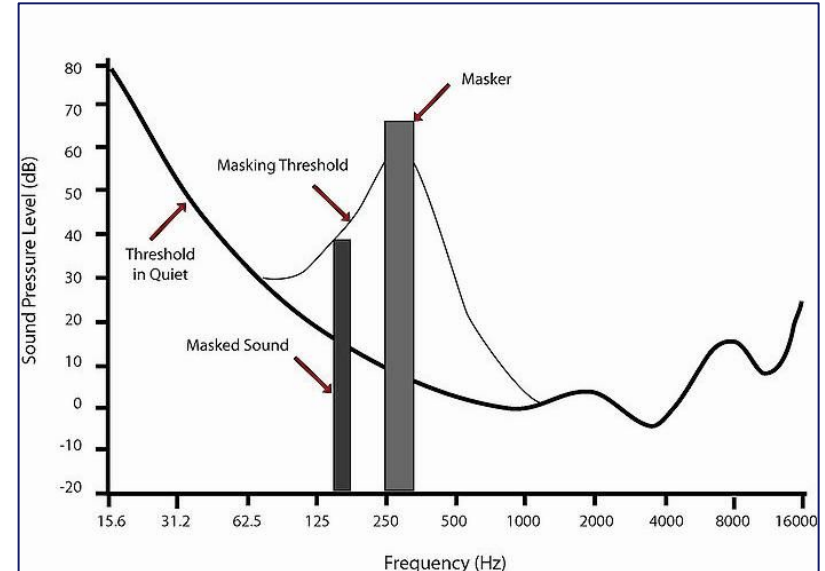
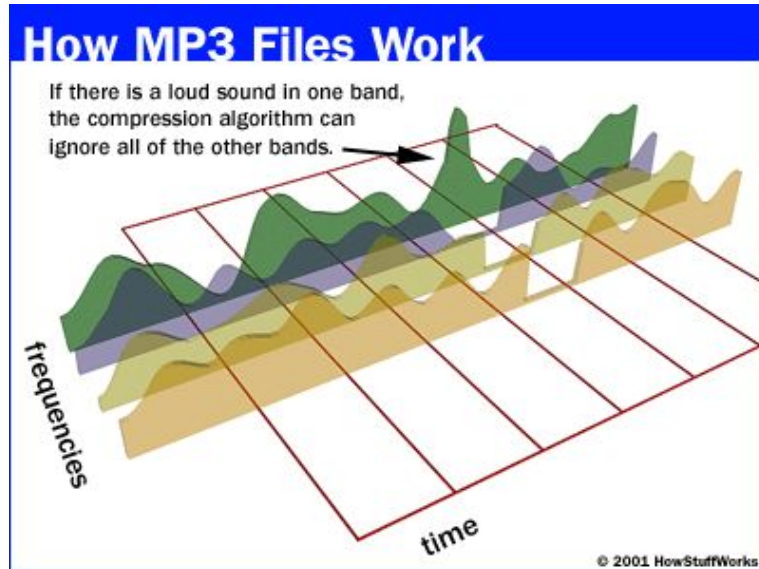
Formati audio per files

- PCM, 16 bits, 44.1 KHz, mono: 16 bits per campione x 44'100 campioni al secondo = 705'600 bits per secondo.
- PCM, 16 bits, 44.1 KHz, mono: 16 bits per campione x 44'100 campioni al secondo x 2 canali = 1'411'200 bits per secondo.
- MP3 256: comprime in al massimo 256'000 bits per secondo, indipendentemente dal numero di canali.



Compressione psicoacustica

La compressione psicoacustica (anche chiamata perceptual coding) lavora riducendo o approssimando la precisione di alcune componenti sonore considerate sotto la capacità di udito delle persone.



Compressione psicoacustica

I difetti che vengono attribuiti alla compressione psicoacustica sono: forme di distorsione, timbro metallico, perdita di profondità e di immagine 3D, riduzione della ricchezza sonora e suono complessivamente impastato.

